Model Stability, Model Complexity, & Ensemble Modeling

PROCEEDINGS OF THE 13th NATIONAL STOCK ASSESSMENT WORKSHOP

Hosted by the Southwest Fisheries Science Center

Irvine, CA | May 22nd-24th, 2018

NOAA FISHERIES

Edited by Melissa A. Karp, Kristan Blackhart, Patrick D. Lynch, Jon Deroba, Dana Hanselman, Vladlena Gertseva, Steve Teo, Howard Townsend, Erik Williams, Annie Yau

U.S. Department of Commerce
National Oceanic and Atmospheric Administration
National Marine Fisheries Service

NOAA Technical Memorandum NMFS-F/SPO-189
February 2019

PROCEEDINGS OF THE 13TH NATIONAL STOCK ASSESSMENT WORKSHOP

Model Complexity, Model Stability, and Ensemble Modeling

Edited by Melissa A. Karp, Kristan Blackhart, Patrick D. Lynch, Jon Deroba, Dana Hanselman, Vladlena Gertseva, Steve Teo, Howard Townsend, Erik Williams, and Annie Yau

NOAA Technical Memorandum NMFS-F/SPO-189
February 2019

U.S. Department of Commerce
Wilbur L. Ross, Jr. Secretary

National Oceanic and Atmospheric Administration
RDML Tim C. Gallaudet, PhD, USN Ret., Acting NOAA Administrator

National Marine Fisheries Service
Chris Oliver, Assistant Administrator for Fisheries

SUGGESTED CITATION:

Karp, Melissa A., Kristan Blackhart, Patrick D. Lynch, Jon Deroba, Dana Hanselman, Vladlena Gertseva, Steve Teo, Howard Townsend, Erik Williams, and Annie Yau (eds). 2018. Proceedings of the 13th National Stock Assessment Workshop: Model Complexity, Model Stability, and Ensemble Modeling. U.S. Dept. of Commerce, NOAA. NOAA Technical Memoranda NMFS-F/SPO-189, 49 p.

A COPY OF THIS REPORT CAN BE OBTAINED FROM:

Office of Science and Technology
Assessment and Monitoring Division
National Marine Fisheries Service
National Oceanic and Atmospheric Administration
1315 East-West Highway SSMC3
Silver Spring, MD 20910

OR ONLINE AT:

http://spo.nmfs.noaa.gov/tech-memos

Contents:

EXECUTIVE SUMMARY

The National Oceanic and Atmospheric Administration National Marine Fisheries Service (NOAA Fisheries) held its 13th National Stock Assessment Workshop (NSAW) from May 22 to May 24, 2018 at the National Academy of Sciences' Beckman Center in Irvine, California. Eighty participants drawn from NOAA Fisheries headquarters, science centers, and regional offices, and Fisheries Management Council staff attended, along with speakers invited from the National Weather Service, the Inter-American Tropical Tuna Commission, and the International Pacific Halibut Commission.

The NSAWs have two primary objectives:

i. Address an important or topical theme of common concern to all NMFS Science Centers,
ii. Provide a forum for interaction between the large diversity of NMFS scientists involved in conducting stock assessments, providing management advice, and related activities.

The 13th NSAW focused on challenges, lessons learned, and best practices related to two themes: model complexity and stability, and ensemble modeling. The workshop was broken into two sessions, the first focused on model complexity and stability and the second focused on ensemble modeling. Each session included oral and poster presentations, as well as two moderated breakout group discussions to delve further into specific questions related to the session theme. Each session culminated in a plenary session where representatives from each breakout group presented conclusions from their group's discussion. In combination, the presentations, breakout sessions, and plenary discussions formed the basis for recommendations and guidance from this workshop. Below is a list of the main conclusions and recommendations from the 13th NSAW.

Conclusions and Recommendations

Session 1: Model Complexity and Stability

- Increased biological realism is not always necessary to improve fisheries management; there is likely a sweet spot where realism is sufficient to achieve management objectives.
- Adding model complexity does not necessarily create computational instability in a model if the added complexity is supported by available data.
- Model complexity should be data driven and product/objective oriented.
- A range of model complexities should be explored, but analysts can use their discretion in determining an approach
- Simulation and Management Strategy Evaluations can be used to determine the appropriate level of model complexity and to evaluate net benefits and tradeoffs of building more complex models.
- Terms of Reference (TOR) should describe the minimum level of complexity needed to achieve management objectives, but they should not be overly prescriptive, such that they dictate the scientific approach.
- Decisions about model complexity should have a strong scientific basis and should not be driven by the favorability of results.
- Uncertainties associated with results from more complex models should be summarized and presented to managers to help meet their decision-making needs.

Session 2: Ensemble Modeling

- Analysts should always try to fully characterize risk, uncertainty, and different potential hypotheses in their models, and ensemble modeling (or multi-model inference) enables analysts to more fully characterize uncertainty compared with the traditional single "best" model approach.
- When deciding whether to run an ensemble model, analysts should evaluate the data available to support the ensemble approach, as well as how robust the single "best" model is, and how sensitive management decisions are to different hypotheses.
- Challenges to operationalizing ensemble modeling include: 1) communication of methods and results within the assessment community and to managers and stakeholders, 2) technical expertise, and 3) cost, time, and resource availability.
- In general, model weights should be based on the same type of evaluation techniques that are used when identifying a base model, and the weighting scheme should be reproducible.
- Analysts should clearly demonstrate to managers and stakeholders the benefits of the new approach; this will be important to get buy-in and help operationalize ensemble modeling for use in management decisions.
- Results from model ensembles should be presented as probabilistic statements included in standard assessment reports; this will likely allow for a smoother transition to using ensemble models.
- Providing ensemble modeling training opportunities to NOAA stock assessment scientists will improve the agency's ability to utilize this promising approach.

The NSAW steering committee will track progress on these recommendations and draw connections between this and future NSAWs to maintain a consistent thread of discussion between NSAWs.

INTRODUCTION

The 13th installment of NOAA Fisheries' National Stock Assessment Workshop (NSAW) was hosted by the Southwest Fisheries Science Center from May 22 to May 24, 2018, at the National Academy of Sciences' Beckman Center in Irvine, California. NOAA Fisheries holds these workshops to address important stock assessment-related themes germane to the agency's science-based sustainable fisheries management mission. The NSAWs also provide a forum for interaction and collaboration among NOAA Fisheries scientists from around the nation, particularly those who are involved in the stock assessment process. The workshop was planned by a national steering committee (Appendix A) made up of representatives from all six NOAA Fisheries Science Centers and the Office of Science and Technology. The workshop drew its 80 participants principally from NOAA Fisheries headquarters and science centers; however, there were also some participants from the regional offices and Fishery Management Council staff. In addition, invited speakers came from the National Weather Service, the Inter-American Tropical Tuna Commission, and the International Pacific Halibut Commission. Attendees are listed in Appendix B.

The 13th NSAW primarily focused on two themes. The first was model complexity and model stability, including the technical benefits and costs associated with development of increasingly complex fisheries stock assessment models. The second theme of the workshop was ensemble modeling. This is an approach commonly used in weather forecasting, for example, where the results of multiple models with alternative plausible hypotheses or a single model with differing values for fixed inputs are combined to generate scientific advice.

The primary goal of this workshop was to facilitate in-depth discussions of the trade-offs associated with increasing model complexity and use of ensemble modeling approaches in the fishery stock assessment process. Each theme was addressed in a plenary session that included a series of presentations from subject matter experts, and in breakout sessions in which smaller groups were offered a set of questions for discussion related to the theme. Breakout groups addressed thematic questions individually and were paired off with another group to compare notes. Each session ended with a facilitated plenary discussion where each pair of groups shared the key conclusions, and recommendations from their breakout discussions.

The workshop also included a brief session on general issues of national interest to NOAA Fisheries stock assessment scientists. Presentations during this session provided updates on the National Standard 1 Technical Guidance Working Group, Strategic Plan for the National Stock Assessment Program, the Next Generation Stock Assessment Improvement Plan, and the Assessment Methods Working Group (AMWG) (see Appendix C for more detailed summaries of these presentations).

A complete agenda for the workshop can be found in Appendix D and presentation and poster abstracts can be found in Appendix E.

SESSION 1: MODEL COMPLEXITY AND STABILITY

SESSION OVERVIEW

NOAA Fisheries' push toward ecosystem-based and holistic stock assessments, combined with recent technological progress, such as, advances in computer simulation and data analysis, have facilitated the development of increasingly complex fisheries stock assessment models. These models tend to have increased realism; that is, they take better account of biological interactions and other factors and sources of variation often not included in simpler assessment models. However, using complex models over simpler ones may come with tradeoffs in data requirements, model stability, time and resource costs, and even potentially the quality and timeliness of management advice. There is limited guidance on when to increase model complexity or how to determine the right level of complexity.

This session, which included presentations by NOAA stock assessment scientists and breakout discussions, provided an opportunity to share regional experiences with models of varying complexity in the context of fisheries stock assessment. Presentations addressed issues related to stability-complexity (and variance-bias) tradeoffs, different approaches to building complex models, and different types of model complexity. Presentations and discussions highlighted the fact that complexity can result from the number of parameters being estimated, the structure of the sub-model, different data inputs, and the complexity of the system being modeled. Presentations focused on each of these aspects of model complexity, from the influence of adding more parameters to the model to how to model fleet and spatial stock dynamics.

Two breakout sessions provided the opportunity for focused discussions about particular questions and issues relating to the use of increasingly complex models in fisheries management. A plenary discussion after the two breakout sessions gave each breakout group the opportunity to share the outcomes of their discussions with the rest of the workshop participants. Before each group provided the summary of its discussions, a quick poll of the audience was done to solicit opinions from all workshop participants regarding each trigger question. Specific results of these polls are provided in Appendix F and summaries of the discussions during the two breakout sessions are provided in the following two sections. Workshop participants generally agreed that:

- a range of model complexities should be explored;
- management strategy evaluations (MSEs) and/or simulations are useful tools to help select an appropriate level of model complexity;
- more complex models are not inherently less stable; and,
- complex models can account for hidden constants in simple models.

However, participants were split evenly when asked if the benefits of using increasingly complex models outweigh the potential shortcomings and time/resource costs. Additionally, polling results indicated that biological realism is more important than parsimony; however, from discussions it became clear that participants actually disagreed on this topic. While biological realism is important, increased realism is not always necessary to improve fisheries management. Analysts tend to try to account for as much realism as necessary to adequately predict aspects of the system needed in management decisions.

A common thread throughout workshop discussions was the need to explicitly consider

BOX 1: Conclusions and Recommendations

- Increased biological realism is not always necessary to improve fisheries management; there is likely a sweet spot where realism is sufficient to achieve management objectives.
- Adding model complexity does not necessarily create computational instability in a model if the added complexity is supported by available data.
- Model complexity should be data driven and product/objective oriented.
- A range of model complexities should be explored, but analysts can use their discretion in determining an approach (e.g., starting with a simple model and building in complexity versus starting with complexity and working toward simpler configurations).
- Simulation analysis can be used to determine an appropriate level of model complexity, whereas management strategy evaluations can be used to test how different approaches affect management outcomes.
- The net benefits of increasing complexity should be examined to evaluate the tradeoffs of building more complex models.
- Terms of Reference for stock assessment analysis and review should describe the minimum level of complexity needed to achieve management objectives, but they should not be overly prescriptive, such that they dictate the scientific approach.
- Decisions about model complexity should have a strong scientific basis and should not be driven by the favorability of results.
- Uncertainties associated with results from more complex models should be summarized and presented to managers to help meet their decision-making needs.

three things when determining the appropriate level of model complexity. These are: 1) the management end goals or objectives; 2) the hypotheses being tested; and 3) the available data. Engaging stakeholders early on in the process is important, particularly when identifying management goals and objectives. With regard to building complex models, two approaches were discussed: 1) starting with a simple model and building in complexity, or 2) starting with complexity and working toward simpler configurations. Although participants did not clearly identify which approach was best, they strongly supported the recommendation that a range of complexities be explored.

Main conclusions and recommendations from this session are summarized in Box 1.

BREAKOUT SESSION PART I

Following plenary presentations, participants divided into breakout groups to discuss how to choose the optimum level of model complexity. Each group was offered one of the following four trigger questions related to the overall session theme:

- How do you find the appropriate balance between realism and parsimony?
- What is the process, and what are the appropriate diagnostics, to ensure that the level of model complexity is optimal for your situation?
- What are best practices for exploring a range of model complexities?
- How should MSEs and/or simulation studies be used to select model complexity level?

Further discussion of each question occurred during the facilitated plenary session. Here we present the main discussion points and recommendations from those breakout and plenary discussions.

How do you find the appropriate balance between realism and parsimony?

Moderators: John Syslo / Jim Hastie
Rapporteurs: Josh Nowlis / Jane DiCosimo

Adding biological realism to stock assessment models requires a tradeoff. Increased biological realism can increase variance (i.e., the variance-bias tradeoff), and can create the potential for computational instability. It is therefore important to find a balance between increasing biological realism, which often means increasing model complexity, and building a more parsimonious model that will convey information necessary for management decisions using as few predictor variables as possible. The general sentiment of the group was that while biological realism is a goal, analysts usually seek to incorporate enough realism to adequately predict aspects of the system needed by managers. To that end, when trying to find the balance between biological realism and parsimony, analysts should explore a range of model complexities, from the basic stock assessment model to more complex models. The consensus was that one should build the most complex model supported by the data, and individual cases should be evaluated independently, depending on science and management objectives and data availability. Analysts should make use of statistical techniques to evaluate the utility range of model complexities (e.g., Akaike Information Criteria, Deviance Information Criteria, and Hessian determinant analysis to identify convergence potential). Nontraditional stock assessment data may be included to help improve biological realism, but this data may not improve model or management advice without the inclusion of traditional data (i.e., a minimum of catch data and a relative index of stock abundance).

What is the process, and what are the appropriate diagnostics, to ensure that the level of model complexity is optimal for your situation?

Moderators: Don Kobayashi / Kyle Shertzer
Rapporteurs: EJ Dick / Laura Solinger

Model complexity can be interpreted in various ways, such as: the number of model parameters, structure of the model in terms of form (e.g., polynomial versus exponential function, error structure), the range of data inputs and types, and complexity in the modeled system (e.g., spatial or temporal resolution, environmental factors, multispecies considerations, etc.). No matter the type of complexity being explored, the group recommended the following five steps to determine the appropriate level of model complexity:

1. **Identify ultimate goals and objectives** – the level of complexity will depend heavily on the end goals. During this first step analysts should ask:
    a) What are the questions or hypotheses available for the system being modeled?
    b) What are the end goals of the analysis (e.g., to provide a stock status determination? To provide catch advice? To conduct a management strategy evaluation? To make projections of future conditions?)?

2. **Assess data availability and quality** – the amount and quality of data available may limit the options available in terms of model complexity. The levels of complexity explored in modeling applications must be supported by the data available. Analysts should be careful not to build a model that is more complex than the data can support.

These first two steps are very important as the levels of model complexity explored hinges on both the objectives and data availability and quality.

3. **Build models with varying levels of complexity**. The choice of starting with a simple model and building up in complexity versus starting with complexity and working toward simpler configurations is less important

than ensuring candidate models cover a range of complexity.

4. **Run model diagnostics and evaluate model fit**. A variety of diagnostic tests can be used for evaluating and comparing models of varying levels of complexity (e.g., out-of-sample prediction, retrospective analysis, information criteria, perturbation analysis, randomization tests, sensitivity analysis, gradient equals zero, and cross-validation).

5. **Select final model(s) and document and justify the model(s) and model selection process**.

What are the best practices for exploring a range of model complexities?

Moderators: Jim Berkson / Patrick Lynch
Rapporteurs: Hui-Hua Lee / Jeff Vieser

Context matters in choosing model complexity. Across different regions and in different science centers, some data is inherently more complex, necessitating models that are also more complex. In addition, some Regional Fishery Management Council's Scientific and Statistical Committees (SSCs) prioritize complex models when determining assessment schedules. They push for conducting assessments of stocks that are more complicated in nature, and push to add complexity to those and to other assessments. Those requests often go beyond established terms of reference during the assessment peer-review process. However, there is no established pathway for "walking back" stock assessment complexity.

With regard to best-practices for exploring a range of model complexities, the modeling approach and range of plausible hypotheses explored in a stock assessment should be data-driven (i.e. determined by the data that are available). A complex model with bad data is not useful, while a simpler model appropriately fit to the available data can be very useful. Complexity should also be constrained by the type of stock assessment being conducted (e.g., those within the scope of research assessments versus operational assessments), and therefore by the objectives and intended outputs of the assessment. All assessment outputs need to be considered in the context of what management can tolerate – if a council only wants a single point estimate or range, providing a range of results from models with different tradeoffs and varying degrees of complexity may not be welcome or useful. Any complexity introduced should be weighed against how much uncertainty it introduces into the overall product. When determining an appropriate level of model complexity, it is important to move deliberately as it could be difficult to reduce model complexity if data issues are encountered later on in the process.

The group also raised the question of whether age-structured assessment models should be implemented throughout the agency as the default stock assessment modeling approach. Does having age-structured data necessitate the use of an age-structured model, or does lacking this data mean that an age-structured model cannot be used? Data-limited approaches (that are not age-structured) could be masking large assumptions or problems that could otherwise be brought forward and examined if an age-structured model is used. Additionally, age data might be more informative than catch-per-unit-of-effort (CPUE) or abundance indices – it might be more consistent across areas. However, the utility of age data compared with abundance indices is unclear and likely depends on the quality and/or representativeness of each data source. One suggested approach is an integrated, step-based, age-structured assessment model, in which complexity is built in only when encountering issues with the model. However, there was some debate over whether this was a good approach; therefore further discussions are needed on this issue.

How should MSE/simulation studies be used to select model complexity level?
Moderators: Cindy Tribuzio / Brian Langseth
Rapporteurs: Abigail Furnish / Sarah Margolis

Management strategy evaluations are involved processes that include stakeholder participation and many iterations. To evaluate model complexity, therefore, stock assessment scientists will most likely be using closed-loop simulations, rather than true MSEs. Nonetheless an MSE can be used to test the complexity of an assessment model, depending on the assumptions of the operating models. The operating models can test whether a simple or more complex assessment model should be applied to the data, while also considering the goals of the managers and questions that need to be addressed. The group noted, however, that although this is possible, using MSEs to test complexity may be too cumbersome a process. Therefore, participants recommended simulation studies as a good alternative to MSEs to determine model complexity levels, evaluate the effect of adding or removing data sources from the model, and examine the tradeoffs between stability and fit, or variance and bias. When their use is possible, MSEs can be a useful tool for testing the effects on management outcomes of including additional data in models.

The group raised one potential concern, however, for both simulation and MSEs. They noted that how the operating model ("truth") is set up could unduly influence the results of the study. For example, if a true population were set up with a significant trend in recruitment, then having a model that explicitly includes recruitment would likely be considered important. Therefore, the group recommended that different individuals be involved in the creation of operating and assessment models to avoid bias from a single individual modeler building both. An alternative approach, should only one individual be available for model building, is to explore multiple configurations of the truth to get at multiple scenarios to more fully understand the effects of added complexity.

Another important consideration is the choice of performance measures when determining the desired level of complexity. Outcomes that are desired by management may inform the degree of model complexity needed to calculate them. Similarly, the group recommended developing performance metrics and diagnostics that can help determine at which levels of complexity models may be breaking down. These metrics could be useful in determining the appropriate level of complexity.

Finally, MSEs likely are best used to determine model complexity when starting from a complex model and scaled down to simpler models to see if the model with fewer inputs can still have useful outputs. This approach would allow for the determination of how losing data (e.g., due to loss of sea days for a survey) would affect model and management outcomes. An MSE also can be used to evaluate a data gap or the utility of planned data collection, in which case one would start with a simple model and build in complexity. Regardless, if an MSE indicates that complex models and simple models provide similar levels of information and management performance, simple models may be preferred because they are often easier to run and communicate.

BREAKOUT SESSION PART II

The second breakout session focused on the tradeoffs associated with increasing model complexity. Each group in this session addressed one of the following four questions related to the objective of this session.

- How should you address impacts on stability from added complexity?
- How should you balance the benefits of increasingly complex models with the potential shortcomings and time/resource costs?

- How can you ensure that complex models effectively account for the hidden constants (e.g., natural mortality, growth, gear selectivity) in simple models?
- How do you ensure that model complexity does not interfere with management practicality?

Participants further discussed these issues during the facilitated plenary session to give all participants a chance to respond to each question. Here we present the main discussion points and recommendations coming out of the breakout and facilitated plenary discussion sessions.

How should you address impacts on stability from added complexity?

Moderators: John Syslo / Jim Hastie
Rapporteurs: Josh Nowlis / Jane DiCosimo

Stability, or conversely, instability, can be defined in terms of the ability to develop, maintain, and run model code (i.e., computational instability) or in terms of how results are affected as new data are included in a model, which relates to the sensitivity of the management advice (i.e., management instability). Increasing model complexity can have tradeoffs in terms of decreased stability. However, adding complexity may not cause computational instability if the data can support it. The group recognized that models with additional parameters do not always have a higher probability of being over-specified, depending on the data that are available. Instead, the effective number of parameters is what matters – too many parameters can dilute the ability to understand the system being modeled. If the use of a more complex model requires fixing some parameters due to estimation issues, analysts may underestimate the uncertainty. A potential solution would be to parameterize the model differently. As described in responses to previous questions, participants in this group also discussed how the determination of proper model complexity is case-dependent, and in most cases complexity should be added in steps with model building and evaluation proceeding hand-in-hand. This will allow analysts to evaluate the effects of iteratively adding complexity on model stability.

How should you balance the benefits of increasingly complex models with the potential shortcomings and time/resource costs?

Moderators: Don Kobayashi / Kyle Shertzer
Rapporteurs: EJ Dick / Laura Solinger

As a first step toward answering this question the discussion group reiterated the need to define the goals of the assessment, and evaluate the level of complexity as it relates to those goals (e.g., number of parameters, underlying dynamics of the system being modeled, review and communication process). Discussion then focused on identifying some of the benefits and shortcomings of increased model complexity.

The main benefit of building more complex models is their ability to better explain or describe the system being modeled. Additionally, adding more parameters to a model may provide more flexibility, and – if the model is constructed accordingly – potentially enable the model to be extended to apply to other questions.

Building more complex models is not without its shortcomings. These are mostly related to time and resource needs. More complex models require more time to build, explain, and communicate results than simpler models do. Therefore, when deciding whether to build a more complex model, analysts need to take into account the effort needed to communicate the complexity and uncertainties in the models to managers. The shortcomings of building more complex models are easier to understand and measure than the benefits. Participants suggested an evaluation of the benefits and tradeoffs of added complexity could be carried out through the examination of the marginal benefits offered by the model. This could be accomplished by looking at the deviance explained for each parameter. Another suggested approach could compare complex models to simpler ones and evaluate how much more precision is obtained, and whether there is an increase in information produced by the more complex model. Participants also suggested measuring the entropy ratio (increase or decrease relative to a simpler model) as a way of evaluating the information content produced by various model complexities. To help address the increased time required to determine whether to build a more complex model, participants suggested developing a data inventory for the various fishery management plans[1]. This would increase efficiency by providing a streamlined way to determine data available for each stock and therefore what level of complexity could be supported.

Ultimately, the benefits of more complex models may depend on stakeholder and manager buy-in. Analysts may need to qualitatively or quantitatively show that the more complex model is better, or contributes to management success. The group cautioned, however, that the choice to use a more or less complex model should have a strong scientific basis and should not be driven by the favorability of the results.

How can you ensure that complex models effectively account for the hidden constants in simple models (e.g., natural mortality, growth, gear selectivity)?
Moderators: Jim Berkson / Patrick Lynch
Rapporteurs: Hui-Hua Lee / Jeff Vieser

The group approached this question by first discussing how to determine model quality. Initial discussion centered on the importance of looking at the data. Since a complex model that lacks the data to support it is not more useful than a simple model, the quality and benefit of a more complex model depends on analysts having an independent method for validating data and a standard method for stepping through complexity. The group noted that qualitative criteria for data selection is generally lacking.

The group then discussed approaches for building and evaluating complex models. Central to the discussion was the recognition that exploring the setting of parameters or priors one at a time in logical sequence in a practiced and practical manner is a good way to ascertain the appropriate approach. To ensure that the right level of complexity is achieved and accounts for hidden constants in simpler models, analysts should start with simple models and build in parameters incrementally with in-depth diagnostics. This approach was considered preferable to starting with peak complexity or incorporating too much complexity too quickly. However, during plenary discussions the group recognized that there were also situations where starting with complexity and then stepping down to simpler models was preferred (e.g., to evaluate the effects of losing data, see Breakout Session 1).

All stock assessments should *a priori* determine: 1) how to evaluate model adequacy; 2) an objective measure of complexity and whether it is warranted; and 3) a method for determining when the overall assessment model adequately addresses the management objective. Stock assessment analysts should target an assessment model category determined to be achievable given the available data. Given that model category, analysts should begin with minimal complexity and then utilize a measured, stepwise, and data-driven process to incorporate additional complexity. Diagnostic and objective tests should be used whenever a complex model is considered, and should accompany increases in complexity.

How do you ensure that model complexity does not interfere with management practicality?
Moderators: Cindy Tribuzio / Brian Langseth

[1] The Next Generation Stock Assessment Improvement Plan recommends implementation of a similar effort (https://spo.nmfs.noaa.gov/sites/default/files/TMSPO183.pdf)

Rapporteurs: Abigail Furnish / Sarah Margolis

The discussion group began by identifying the context of, and defining, "management practicality." Practicality was generally discussed in the context of the ability of scientists to effectively communicate assessment methods and results to managers and stakeholders. It was agreed that when implementing more complex models into management, the priorities of the managers (i.e., what questions they are trying to address and desired assessment outputs) should be clearly understood by the stock assessment scientists. This will influence the type of model used and whether adding more complexity would be beneficial. Understanding management priorities requires clear communication, and the group suggested that management output needs should be clearly identified in the TOR for any stock assessment, if they exist. A TOR should describe the minimum of what managers need, but not exclude certain analyses or approaches, thus allowing the assessment scientists flexibility when building the model. Overall, adding complexity to a model should be done purposefully and should be the decision of the assessment scientist in the context of answering specific questions.

The outputs resulting from complex models need to then be summarized in a way that will be useful to managers. The importance of the uncertainties resulting from increased complexity should also be evaluated by analysts and communicated to managers. This last point led to a discussion of the variance-bias tradeoff (i.e., as model complexity increases, uncertainty [variance] increases, yet systematic error [bias] decreases). The variance-bias tradeoff can be balanced using Models of Intermediate Complexity for Ecosystems (MICE models)[2]; however, managers should be aware that more complex models may increase uncertainty relative to simpler models, which tend to underestimate uncertainty. Therefore, an approach that accounts for uncertainty around the results should be in place.

The group noted that there are regional differences in the degree to which management bodies participate in the review and acceptance of stock assessment models. For Councils, their SSCs evaluate the scientific advice and therefore need to understand the model, its complexity, and how the models are selected for use. In some regions, the assessment scientists may have to present their models to stock assessment review panels, and the council, in addition to the SSC. It is important for the assessment scientists to know their audience(s) and also how best to communicate model complexity and the results most important to managers.

[2] MICE are a category of models that are intermediate in complexity between simple single-species models and whole-of-ecosystem models in terms of number of components and processes represented. They restrict the complexity by including only those components that are needed to address specific management questions and objectives. *Plaganyi et al. 2014. Multispecies fisheries management and conservation: tactical applications using models of intermediate complexity. Fish and Fisheries 15(1): 1-22*

SESSION 2: ENSEMBLE MODELING

SESSION OVERVIEW

Fisheries science tends to employ methods that result in the selection of a single "best" model. This tactic can be problematic because there is uncertainty around the selection of the "best" model, and potential loss of relevant information in the models not selected. One may ask then, why use only a single model? Why not incorporate information from several models? Ensemble modeling is a statistical technique that does just that by combining results from multiple alternative models and data sets to generate results and projections of future system states. The use of ensemble modeling enables a more complete characterization of uncertainty in the system, particularly the structural uncertainty associated with assumptions about stock and fishery dynamics. Ensemble modeling also allows for the exploration of multiple plausible hypotheses (Figure 1). This is an especially promising technique to help NOAA as it moves toward ecosystem-based fisheries management (EBFM) and attempts to account for more tradeoffs, with potentially increased complexity and uncertainty in the system.

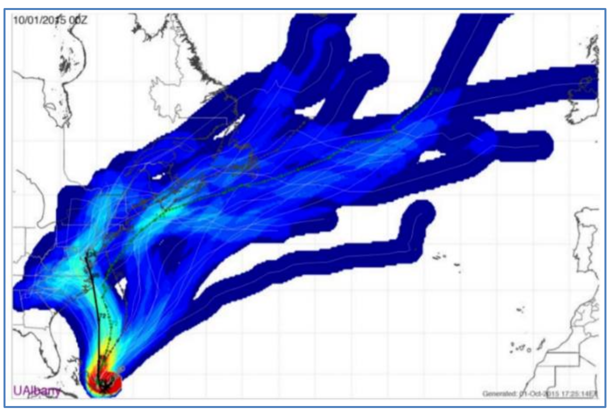


Figure 1: Illustration of an ensemble of models for forecasting a hurricane's path. Each tract represents different model structures and/or assumptions. *Image Credit:* Ian Stewart & Allen Hicks

Ensemble modeling and other multi-model inference approaches are commonly used in other fields such as weather prediction, climate science, and machine learning. Ensemble modeling in fisheries stock assessments is still nascent and faces several potential challenges. There is as yet no clear guidance for use of ensemble modeling in the context of fisheries management. The 12th NSAW[3] identified the need for more guidance on the suite of configurations that should be considered in an ensemble, how results across the ensemble should be combined to develop management advice, and how to appropriately characterize uncertainty. Therefore, the objectives of this session were to identify: 1) situations when ensemble modeling is beneficial for providing fisheries management advice; 2) challenges to operationalizing ensemble modeling in fisheries stock assessments and management; and 3) best practices for conducting ensemble modeling in fisheries stock assessments and incorporating results into management decisions.

The presentations, posters, breakouts, and plenary discussions from this session provided opportunities for NOAA stock assessment scientists from across the nation and other subject matter experts to consider the benefits, challenges, and potential uses of ensemble modeling in U.S. fisheries management. During the plenary discussions each breakout group had the opportunity to share the outcomes of its discussions with all workshop participants. As was done for session I, the audience was polled to elicit their

[3] https://www.st.nmfs.noaa.gov/Assets/stock/documents/NSAW%2012%20PROCEEDINGS%20ONLINE.pdf

opinions on each trigger question before each group provided their discussion summary. Complete results from the polling can be found in Appendix F.

In general, workshop participants agreed that ensemble modeling could be useful and appropriate for fisheries management; the main benefit to using an ensemble modeling approach was improved characterization of uncertainty; the major challenges to operationalizing ensemble models were issues with communicating results, management inertia, and time requirements. Participants also noted that improved communication, more resources, best practices guides, and staff training would help overcome the challenges. Considering that one of the main challenges to operationalizing ensemble models is communication and management inertia, it will be important to engage with stakeholders throughout the model development process to ensure that they are aware of and understand the new modeling approaches being utilized.

Common questions raised throughout presentations and during the breakouts were related to determining the candidate models in the ensemble set, deciding at what point models are removed from the ensemble, how to determine model weights, and how to effectively communicate ensemble results to managers and stakeholders. Main conclusions and recommendations are summarized in Box 2.

BOX 2: Conclusions and Recommendations

- Analysts should always try to fully characterize risk, uncertainty, and different potential hypotheses in their models; a main benefit of ensemble modeling (or multi-model inference) is that it enables analysts to more fully characterize uncertainty compared with the traditional single “best” model approach.
- When deciding whether to run an ensemble model, analysts should evaluate the data available to support the ensemble approach, as well as how robust the single “best” model is, and how sensitive management decisions are to different hypotheses.
- Challenges to operationalizing ensemble modeling include: 1) communication of methods and results within the assessment community and to managers and stakeholders, 2) technical expertise, and 3) cost, time, data, and resource availability.
- In general, model weights should be based on the same type of evaluation techniques that are used when identifying a base model, and the weighting scheme should be reproducible.
- Analysts should clearly demonstrate to managers and stakeholders the benefits of the new approach; this will be important to get buy-in and help operationalize ensemble modeling for use in management decisions.
- Results from model ensembles should be presented as probabilistic statements included in standard assessment reports; this will likely allow for a smoother transition to using ensemble models.
- Providing ensemble modeling training opportunities to NOAA stock assessment scientists will improve the agency’s ability to utilize this promising approach.

BREAKOUT SESSION PART I

The overall goal of this breakout session was to discuss the main benefits and challenges of ensemble modeling in fisheries management, and identify best practices for determining when and why to use it. Each breakout group discussed one of the following three questions related to the overall session goals:

- What are the benefits and shortcomings of utilizing ensemble modeling in fisheries stock assessment and what are the implications for management?
- When are ensemble models appropriate for fisheries?
- What are the main challenges to operationalizing ensemble modeling in stock assessments, and how do we overcome those challenges?

Further discussion of each question occurred during the facilitated plenary session. Here we present the main discussion points and recommendations coming out of those breakout and plenary discussions.

What are the benefits and shortcomings of utilizing ensemble modeling in fisheries stock assessment and what are the implications for management?

Moderators: Paul Crone / Howard Townsend
Rapporteurs: Kate Siegfried / Rob Cheshire

Two categories of benefits were identified by the group – 1) scientific and 2) management. Scientific benefits mainly centered on the ability to account for uncertainty. Management benefits focused on the quality of advice provided to managers.

Ensemble modeling can result in improved understanding of the system and sources of variability within it. The group discussed two types of ensembles models: 1) single-model (i.e., using a single model structure with different assumptions about fixed inputs or priors on estimated parameters) and 2) multi-model (i.e., using models with different structures). Single model ensembles account for parameter uncertainty while multi-model ensembles account for model structural uncertainty. Of necessity, models cannot account for all sources of uncertainty, so model choice may introduce bias into results. For example, climate patterns may episodically influence a stock. A model structure that does not include climate patterns may result in biased output if applied to a period when the climate patterns are influencing the stock. Using model ensembles enables the analyst to make the uncertainty transparent – to be explicit about what aspects of the system are well understood, and to point out the potential repercussions of knowledge gaps.

Ensemble models allow the exploration of plausible hypotheses about a system. The plausibility of various hypotheses can be evaluated by an analyst, as well as by other participants in the process. Thus, ensemble modeling may facilitate engagement in the assessment process by offering a venue for alternative ideas and hypotheses to be discussed and tested. Creating an ensemble allows analysts to carry forward multiple hypotheses about model structure or biological variability. In turn, this allows analysts to present probabilistic forecasts of key model outputs. Additionally, a divergent model among an ensemble can serve as a diagnostic tool. Generally, when analysts use multiple models they can learn more about the system and the data, which should improve management advice.

Using ensemble models should benefit management in multiple ways. Advice from ensembles provide managers with a more comprehensive characterization of uncertainty surrounding the scientific advice, so they have a better understanding of the risk associated with their decisions. One might think that explicitly accounting for uncertainty would compound uncertainty such that clear advice may not be obtained from ensemble models, but in some cases ensemble models might actually reduce uncertainty. Additionally, the group noted that running multiple models from multiple analysts may reduce the "analyst effect" – the introduction of bias based on analysts' preferred methods for processing and parameterizing data. Multiple models used in ensembles would dampen the volatility of any given model so that large swings in perceived stock dynamics would be reduced in cases where the swings are attributable to sensitive parameters rather than actual natural variability.

Finally, if most models in an ensemble converge on similar results, then the credibility associated with the assessment results would likely improve, and managers and stakeholders may be more supportive of the resultant management actions. The potential for improvements in the characterization of uncertainty, increased stakeholder engagement, and additional stability and trust in the management process provide strong justification for pursuing ensemble modeling in fishery stock assessments.

The group recognized that although there are abundant benefits to model ensembles, scientists should proceed with a systematic application of ensemble models with caution and consideration of the shortcomings and challenges. These include the cost and tractability of such an effort. Ensemble modeling may require an analyst for each type of model and each analyst would likely need to run single-model ensembles for their model type. Alternatively, one analyst may run multiple types of models, but given typical constraints on operational stock assessments, this approach could produce more errors until a routine process is in place. Use of more models would require more time for review (at least initially when new models are introduced), more time to communicate, and possibly a broader range of reviewer qualifications. Further, the group emphasized that additional research into appropriate model weighting schemes is needed.

Besides the technical and scientific challenges of ensemble modeling, it gives rise to some management challenges as well. Managers and stakeholders accustomed to a “single best model” approach would need an introduction into how ensemble modeling could be used, and how the current process would be affected. The added benefits of ensemble modeling will therefore need to be weighed against the potential costs for each specific case.

When are ensemble models appropriate for fisheries?

Moderators: Will Satterthwaite / Jon Deroba
Rapporteurs: Russ Brown / Lisa Peterson

Simply put, ensemble models are almost always appropriate. However, participants recognized that an elaborate multi-platform or even multi-structure approach is not always needed; what’s important is to try to fully characterize risk and different potential hypotheses underlying assessment models.

Although an ensemble approach may be appropriate in many situations, there are various considerations that could constrain the ability to use it. For instance, it is important that the management body is willing to accept advice from this kind of modeling approach. Effective communication of ensemble modeling methods and results to stakeholders is important to help increase understanding and acceptance of the new approach by management bodies and stakeholders. This is particularly important for the communication of uncertainty; analysts should stress the importance of characterizing uncertainty to support effective decision-making. Resource availability is also a potential constraint on the use of an ensemble modeling approach. Analysts require sufficient resources, not only to devote to the construction of an ensemble model, but also to maintain and evaluate multiple models year after year. Another consideration is the selection of models for inclusion in the ensemble, which must be carefully weighed.

In addition to considerations that may constrain the use of ensembles even when their use is appropriate, there are situations when an ensemble approach would be inappropriate. For example, the use of ensemble modeling in data-limited situations might be inappropriate if there is not enough information to support an ensemble. However, the group recognized that it may also be the case that with limited data it would be difficult to rule out a particular hypothesis. In these situations, there is actually a greater need to evaluate multiple hypotheses, even if a full ensemble approach is not possible. On the other end of the spectrum, if a single model has already been shown to be robust, or if management decisions are insensitive to different hypotheses, limited resources might be better allocated elsewhere. That being said, decisions regarding the use of ensemble modeling should be re-evaluated regularly, as changes in the available data or new modeling approaches could change the circumstances.

What are the main challenges to operationalizing ensemble modeling in stock assessments, and how do we overcome those challenges?

Moderators: Dana Hanselman / Steven Teo
Rapporteurs: Kevin Craig / Matthew Supernaw

Before delving into a discussion about the challenges to operationalizing ensemble modeling, the group reviewed the benefits to ensemble modeling for fisheries management. The main benefit identified is the ability to better characterize uncertainty, which may lead to more informed decisions about management risks. The ability to more fully characterize uncertainty as a probability distribution from an ensemble of models would help create probability-based buffers for catch recommendations that are consistent with a council's accepted risk of overfishing. Additionally, ensemble models may lead to more accurate results compared with single models that can suffer from analyst bias. Ensemble models may also offer a benefit because once one is defined, it may provide a clearer distinction between operations and research stock assessment models, allowing assessment scientists to focus on research models. Furthermore, quantitatively incorporating risk and uncertainty in decision-making could improve management in some regions by taking advantage of the more comprehensive characterization of uncertainty provided by ensemble modeling.

The discussion group covered three broad topics of challenges to operationalizing ensemble modeling: 1) communication of ensemble model methods and results to managers, stakeholders, and within the assessment community, 2) technical expertise needed to build and run the ensemble, and 3) cost, time, and resources. Several communication-related challenges and considerations were identified and discussed. One challenge is the management inertia that will need to be overcome. It has taken a substantial amount of effort to engage management and stakeholders on the details of single-species single-model results and it may be challenging to get management and stakeholders to jump on board with a new approach that might at first be perceived as substantially different. Significant resource and time commitments are required to educate both stock assessment scientists and the consumers of the results about the management benefits of moving in this direction. Stakeholders are better at understanding risk when presented as probability, so translating ensemble model results into that language (e.g., similar to hurricane forecasts – there is a 10% chance of the hurricane hitting your house) could be a helpful approach. It will be important to balance the communication of a greater range of uncertainty with the potential appearance that analysts have no confidence in their results. In cases where individual model runs from the ensemble are provided, stakeholders or managers might gravitate to a single member of the ensemble based on outcomes. To avoid this, analysts must emphasize that an ensemble is not a multitude of models to choose from, but a carefully chosen collection, all of which work in concert to present a more accurate assessment.

In addition, the group noted that moving from single "best" model management advice to ensemble modeling will involve a number of technical challenges. Determining the set of member models that comprise the ensemble will be critical; the number of models, the types of models, and the range of axes of uncertainty are all choices that will likely be subjective in nature but could be informed by expert opinion and/or external information. Establishing who determines the set of ensemble models will also be critical (e.g., whether stakeholders should be able to submit models). Methods used to combine the results of an ensemble for use in management will need to be developed. Classical model averaging techniques are likely not adequate for models of different structural classes, and information-theoretic approaches do not work well for penalized likelihood models. An additional challenge is that for fisheries models, the "truth" is never known, which is unlike weather models where the models can be tuned when "truth" is collected after the forecast. Standardized simulated data sets will have to be developed, with which ensemble models and ensemble forecasts can be tested and trained.

Implementing ensemble models into operational management advice will at least initially require an investment in terms of staff time and expertise. It will also require managing expectations regarding timeliness and the level of review required. Some of the focus of the Next Generation Stock Assessment Improvement Plan is on streamlining the stock assessment process. The additional requirements associated with maintaining a group of different models and their respective data inputs make streamlining a challenge. Some models may need to be updated in sequence, which will make timeliness difficult, particularly in regions that do annual assessments. Reviewing an ensemble model may take a wider group of experts than a traditional stock assessment review due to a potential wide range of model types (e.g., ecosystem-driven models versus traditional single-species models).

BREAKOUT SESSION PART II

The second breakout session focused on discussing best practices for conducting ensemble modeling and utilizing results in advice to managers. Each breakout group discussed one of the following three questions related to that overarching focus:

- What are best practices for selecting models to include in an ensemble?
- How should model predictions be combined to develop ensemble predictions (e.g., simple average, stacking, weighted average, boosted regression, superensemble, etc.), and how should performance of the final ensemble prediction be evaluated?
- How should results from ensemble models be communicated to, and used by, managers?

Further discussion of each question was carried out during the facilitated plenary session. Here we present the main discussion points and recommendations coming out of those breakout and plenary discussions.

What are the best practices for selecting models to include in an ensemble?
Moderator: Paul Crone / Howard Townsend
Rapporteur: Kate Siegfried / Rob Cheshire

Selecting models to include in an ensemble is an iterative process. An overall best practice for developing and implementing ensemble model approaches is therefore to take small steps and evaluate along the way. With that in mind, the group discussed other practices that could help the selection of candidate models.

The group recommended that the selection of candidate models be carried out by a working group of experts as opposed to a single analyst. A clear process for submitting new models to the model working group should be established. Similarly, a clear process of evaluating a model's utility in the ensemble should be established. When deciding what models to include in an ensemble, a working group should consider:

a. How management deals with uncertainty – what step of the management process are you trying to inform?
b. Assumptions that need to be tested – e.g., recruitment patterns, time-varying parameters, fishery productivity, spatial dynamics, fixed life history parameters, and environmental forcing.
c. Axes of uncertainties for your stock, and prioritizing which to address.

To facilitate the early stages of implementation of the ensemble, the group suggested establishing a reference model and varying factors from there (a case array). This approach should help analysts and managers clearly comprehend how different factors contribute to the uncertainty in the model. Additionally, a range of complexity is useful to consider when selecting models to include in an ensemble. In cases where the data used for models do not overlap, objective model comparison may be difficult. Using models that estimate the variance of process errors will minimize the issue of *post hoc* tuning. For the final ensemble, analysts should consider model weighting to create a model average. However, the group recognized that model weighting schemes for stock assessment have not been clearly established, and that this should be a focus of future research. The primary goal of model ensembles should be to encapsulate the variance and the covariance of each of the axes of uncertainty, so the ensemble does not have to be averaged. A final consideration is whether the ensemble put forward can be supported and implemented in future years. Analysts should take this into account when building the ensemble and only put forward ensembles that can be supported and implemented over the long term.

How should model predictions be combined to develop ensemble predictions (e.g., simple average, stacking, weighted average, boosted regression, superensemble, etc.), and how should performance of the final ensemble predictions be evaluated?

Moderators: Will Satterthwaite / Jon Deroba
Rapporteurs: Russ Brown / Lisa Peterson

Participants generally felt that they lacked the expertise to really answer this question. That being said, the group discussed some general ideas and guidelines for combining and evaluating performance of ensemble predictions.

Simple averaging was viewed as less than ideal for most cases. If the goal is to provide a final number, there is probably a better approach than simple averaging to combine hypotheses. For example, averaging to get a final point estimate from two non-overlapping distributions may give an average point estimate outside the original two distributions, and thus possibly outside the realm of plausibility. In other words, a simple average could be misrepresentative of the estimates from the different models. In situations like this, it would be especially important to report the full (summed/combined) distribution such that its multi-modal nature is apparent.

Participants also discussed different ways of weighting the models. They noted that weighting, too, can be a tricky approach to implement, because of the need to decide how to weight. It is hard to define a single best type of weighting, as weighting can be subjective and based on the analyst's experience and choice. Inverse variance weighting is a possibility, but it has drawbacks. For example, higher estimates tend to have the highest variance, so the lower estimates are often favored in this approach. Bayesian model averaging may be a good approach to determine weights. In general, model weights should be based on the same types of evaluative techniques that are used when identifying a base model. Regardless of the weighting method chosen, the group emphasized that the weighting scheme needs to be reproducible, with explicit rules for evaluating and rejecting hypotheses and then weighting the hypotheses that remain. If possible, the effect of the weighting scheme on management decisions should be tested. Expert judgement needs to play a role, no matter what technique is used, because the weighting or averaging needs to be reasonable. To that end, participants recommended establishing a process to identify the expert panel, since decisions would benefit from not being left to a single analyst.

The group noted that there are many potential ways to evaluate the performance of the model predictions. If the prediction is being used to forecast, then the forecast can be compared to what was actually observed. It may be appropriate to do a retrospective analysis, evaluate model fits, and perform other classic diagnostic techniques. With an ensemble modeling approach, there are multiple layers of models, and therefore the group recommended evaluating the performance of the individual models in the ensemble as well as the performance of the ensemble as a whole. If additional data are available, then the ensemble could be compared to an independent dataset, or analysts could use a cross-validation technique. Whatever the evaluation technique, it is important to regularly revisit the approach and the suite of models being used.

How should results from ensemble models be communicated to, and used by, managers?

Moderators: Dana Hanselman / Steven Teo
Rapporteurs: Kevin Craig / Matthew Supernaw

At least when starting to use model ensembles, the group recommended that communication of ensemble results should whenever possible be similar to traditional presentation methods. Managers and stakeholders are familiar with and relatively comfortable with traditional assessment presentations. Sudden deviations from that format can lead to mistrust and pushback; however, some situations require more complex solutions. Benefits of the new approach (e.g., reducing risk of lost harvest or overfishing) must be well demonstrated for management and stakeholder buy-in. Analysts should also stress that the model ensemble more completely accounts for uncertainty.

The discussion also focused on the pros and cons of presenting managers with results from individual model runs within an ensemble, which depend on specific assumptions. For example, a pro is that showing individual model runs might more clearly communicate some of the drivers for uncertain bifurcating model results. If there is an apparent bifurcation in the ensemble results, it may be important to highlight assumptions underlying that bifurcation. Such a bifurcation might drive future research. A con of presenting individual model runs would be that managers may then focus on results from one or more individual models, rather than looking at the whole ensemble. Converting standard assessment plots like phase or Kobe plots to include probabilistic statements likely will allow for a smoother transition to using model ensembles. Many components of fishery management, at least for the U.S. domestic process, already include probabilistic statements (e.g., p* harvest control rules, rebuilding plans, etc.). It should be relatively straightforward to use the uncertainty from a model ensemble in such components. However, some tweaks might be needed to incorporate the more complete characterization of uncertainty inherent in ensemble results.

APPENDIX A: NSAW 13 STEERING COMMITTEE

Name	**Center**
Patrick Lynch	Office of Science and Technology
Kristan Blackhart, *NSAW Program Manager*	Office of Science and Technology
Howard Townsend, *EBFM Liaison*	Office of Science and Technology
Steven Teo, *Chair*	Southwest Fisheries Science Center
Jon Deroba	Northeast Fisheries Science Center
Dana Hanselman	Alaska Fisheries Science Center
Annie Yau	Pacific Islands Fisheries Science Center
Vladlena Gertseva	Northwest Fisheries Science Center
Erik Williams	Southeast Fisheries Science Center

APPENDIX B: LIST OF PARTICIPANTS

<u>Alaska Fisheries Science Center</u>
Cindy Tribuzio
Dana Hanselman
Grant Thompson
Jim Ianelli
Kari Fenske
Kerim Aydin
Peter Hulson

<u>Northeast Fisheries Science Center</u>
Charles Perretti
Dan Hennen
Dvora Hart
Jon Deroba
Liz Brooks
Robert Gamble
Russ Brown
Sean Lucey

<u>Northwest Fisheries Science Center</u>
Isaac Kaplan
James Thorson
Jim Hastie
Josh Nowlis
Kelli Johnson
Merrill Rudd
Owen Hamel

<u>Pacific Islands Fisheries Science Center</u>
Brian Langseth
Donald Kobayashi
Felipe Carvalho
John Syslo
Jon Brodziak
Maia Kapur
Marc Nadon

<u>Southeast Fisheries Science Center</u>
Daniel Goethel
Kate Siegfried
Kevin Craig
Kyle Shertzer
Nancie Cummings
Nikolai Klibansky
Rob Cheshire
Xinsheng Zhang

<u>Southwest Fisheries Science Center</u>
Doug Kinzey
Edward Dick
Hui-Hua Lee
Kevin Hill
Kevin Piner
Kristen Koch
Nick Grunloh
Paul Crone
Steven Teo
Will Satterthwaite
Xi He

<u>Office of Science and Technology</u>
Abigail Furnish
Howard Townsend
Jane DiCosimo
Jeff Vieser
Jim Berkson
Kenric Osgood
Kristan Blackhart
Lisa Peterson
Matthew Supernaw
Melissa Karp
Ned Cyr
Patrick Lynch
Sarah Margolis

<u>Office of Sustainable Fisheries</u>
Karen Greene

<u>Alaska Regional Office</u>
Alicia Miller

West Coast Regional Office
Penny Ruvelas

North Pacific Fishery Management Council
Diana Stram

Humboldt State University
Laura Solinger

Institute of Marine Research
Knut Korsbrekke

Inter-American Tropical Tuna Commission
Mark Maunder

National Weather Service
Hendrik Tolman

International Pacific Halibut Commission
Ian Stewart

APPENDIX C: OTHER TOPICS OF NATIONAL INTEREST SUMMARIES

A special topic session during the NSAW included presentations on other topics of national interest not directly related to the two theme sessions. Presentations provided updates on the National Standard 1 (NS1) Technical Guidance Working Group, the strategic plan for the National Stock Assessment Program (NSAP), the next generation Stock Assessment Improvement Plan (next gen. SAIP), and the national Assessment Methods Working Group (AMWG).

National Standards 1 Technical Guidance Working Group

Melissa Karp, an NSAP contractor, provided an overview of the NS1 Technical Guidance Working Group. The purpose of the group, which is made up of members representing each science center, regional office, NMFS HQ, and includes Fishery Management Council staff, is to develop technical guidance in response to the 2009 and 2016 updates to the NS1 guidelines. The working group divided into three subgroups, each focusing on a particular issue from the updated NS1 guidelines that requires additional technical guidance. The three subgroups are: 1) reference points, 2) carry-over and phase-in harvest control rules, and 3) data gaps and alternative approaches for managing data-limited stocks. Ms. Karp provided an overview of each of the subgroups, focusing on the goals, expected work products from each group, and progress made to date. Drafts of key work products are anticipated by the end of 2018, and likely will be reviewed by the Council Coordinating Committee in February 2019.

Strategic Plan for the National Stock Assessment Program

Alicia Miller, on detail to the NSAP, presented the draft five-year strategic plan for the NSAP (since finalized[4]). The strategic plan describes the NSAP's work and how it supports national goals and regional stock assessment programs. The NSAP has been working to further develop and advance NOAA Fisheries next generation stock assessment enterprise since the program was formally established in 2016. The NSAP is comprised of a program lead, an administrative team, and a modeling team. The NSAP acts as a liaison to leadership in the Office of Science and Technology, Senior Science Advisors (which includes NMFS Senior Scientist, Senior Scientist for Stock Assessments, Senior Scientist for Ecosystems and Senior Scientist for Economics), NMFS HQ Policy and Sustainable Fisheries offices, and other programs. The NSAP staff develops technical and strategic guidance, administers the Expand Annual Stock Assessment (EASA) budget line to invest in regional stock assessments, provides national coordination and tracking of stock assessment activities and initiatives, and facilitates national-level collaboration. The NSAP strategic plan outlines five core goals of the program, and specific objectives and actions under each goal. The five broad goals of the program are:

(1) Support stock assessments and surveys
(2) Advance the stock assessment enterprise
(3) Support innovative science and technology
(4) Build organizational excellence
(5) Represent the enterprise nationally and internationally

[4] https://docs.google.com/document/d/1iDquv87fUqf14uiAEKm6kV5XudVdt-bsEz0UskBueI4/edit

Stock Assessment Improvement Plan: A Next Generation Stock Assessment Enterprise

Dr. Patrick Lynch, NSAP lead, provided an overview of the new strategic document, *Implementing a next generation stock assessment enterprise: An update to the NOAA Fisheries Stock Assessment Improvement Plan* (next gen. SAIP). The next gen. SAIP summarizes the current state of the NMFS stock assessment process and provides recommended actions designed to advance the process toward a next generation stock assessment enterprise that is more holistic and ecosystem-linked, uses innovative science, and is more timely, efficient, and effective. The presentation was followed by a facilitated discussion that provided participants an opportunity to offer feedback and ask questions about the overall document and the three objectives of the next gen. SAIP.

During the discussion, participants raised concerns about the funding, resources, and time necessary to research and build more holistic and ecosystem-linked assessments. Dr. Lynch pointed out that streamlining the overall assessment process will facilitate more time for research among stock assessment scientists. Discussion regarding the objective to expand assessments to be holistic and ecosystem-linked focused on the proposed three-step decision process to determine when an assessment should be expanded and how that would be implemented and utilized. Dr. Lynch clarified that the three-step decision process follows from the stock assessment prioritization process that is underway in each region, with the first step being to determine if the species is a priority for expanding the scope of the assessment. Regarding data collection and innovative science, participants generally agreed that fishermen and other stakeholders can play an important role in enhancing data collection capacity. There was general agreement that subsidizing fishermen to collect data, particularly between surveys, is a great idea and would help improve our understanding about changes in the natural environment, while also providing information that can help adapt surveys.

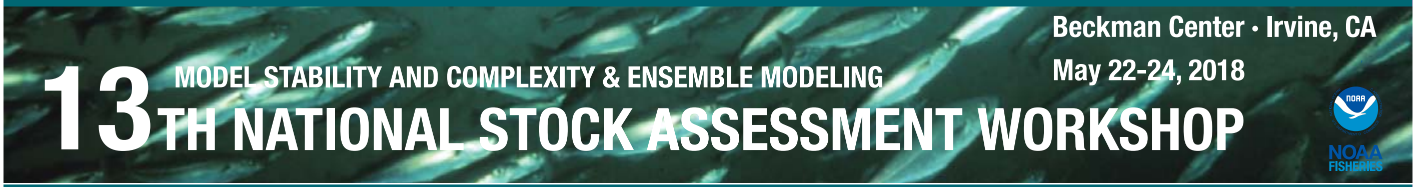
Assessment Methods Working Group

Dr. Owen Hamel, assessment scientist from the Northwest Fisheries Science Center and chair of the AMWG, provided an overview of the AMWG's mission and activities. The AMWG was established to foster development, testing, and maintenance of high-quality assessment methods, and to provide tested, well-documented models that efficiently incorporate all available data into well-understood and accurate assessments. The working group oversees the NOAA Fisheries Toolbox, supports AD Model Builder and Template Model Builder software, supports workshops and training designed to improve stock assessment methods and capacity, and develops, conducts, and supports methods projects.

The work group previously administered two internal funding opportunities, *Improve a Stock Assessment* (ISA), and *Stock Assessment Analytical Methods* (SAAM), but these funding opportunities ended at the end of FY18; instead, starting in FY19 funds will be evenly distributed to the science centers. Nonetheless, Dr. Hamel highlighted the success of those programs over the years. ISA funded 27 projects focusing on numerous topics including improving abundance measures, genetic analyses and stock structure, and ecosystem considerations. SAAM funded 55 projects also focusing on a range of topics, but the greatest number of projects dealt with software and model development, supporting workshops, spatiotemporal dynamics, and data-limited stocks and assessment methods.

Dr. Hamel provided an overview of a new methods project the group is planning to undertake with the goal to: 1) document and evaluate the various modeling approaches being used across the U.S., 2) provide recommendations on good or “best” practices, and 3) identify gaps and challenges. Questions from the audience focused on this new project and how results and conclusions from work group discussions would be made available to the general stock assessment community. He acknowledged that this is still being worked out by the group, but said that once a final version of the conclusions is drafted, the group will determine the best means of sharing it with the larger NOAA stock assessment community.

APPENDIX D: AGENDA FOR THE 13TH NATIONAL STOCK ASSESSMENT WORKSHOP



AGENDA Day 1 - Tuesday, May 22, 2018

TIME	SPEAKER	TITLE
8:00	EVERYONE	COFFEE
8:40	Kristen Koch	Welcome
8:50	Steve Teo	Opening Remarks
NATIONAL ISSUES SESSION		
9:00	Melissa Karp	NS1 Technical Guidance Workgroup Status Update
9:20	Owen Hamel	Assessment Methods Work Group Report
9:40	Alicia Miller	National Stock Assessment Program Strategic Plan Overview
10:00	EVERYONE	BREAK
10:15	Patrick Lynch	Implementing Recommendations of the Next Generation Stock Assessment Improvement Plan
11:30	EVERYONE	LUNCH (Locations within walking distance of Beckman Center: See https://food.uci.edu/dining/ and https://twitter.com/icgft; also Specialty's Cafe, 100 Innovation Drive) **Breakout Session Moderators please meet at the front of lecture hall for a brief check-in with organizers before leaving for lunch**
MODEL COMPLEXITY AND STABILITY SESSION - PART 1		
13:00	Dana Hanselman	Introduction to Session and Theme *Session Objectives: Recent technological process such as advances in computer simulations and data analysis have made it possible to develop increasingly complex fisheries stock assessment models. These models can be more realistic and account for additional potentially influential factors and sources of variation. However, using complex models over simpler ones can come with trade-offs in data requirements, model stability, time and resource costs, and even potentially the quality of management advice. During this session, participants will discuss situations when such trade-offs justify the use of more complex models. The first breakout session will focus on model structural issues, and work towards developing a list of best practices for determining the 'optimum' level of model complexity to balance realism and parsimony. The second breakout will focus on trade-offs and discuss when moving towards complexity is in the best interest of fisheries management advice.*
13:10	Mark Maunder	PLENARY: Goldilocks and the Three Assessments: This Assessment is Too Simple, This Assessment is Too Complex, and This Assessment is Just Wrong!
13:50	Brian Langseth	Transitions in Model Complexity for Assessments for the Deep7 Bottomfish Fishery in the Main Hawaiian Islands
14:10	EVERYONE	BREAK
14:30	Dan Hennen	Are Scale Free Reference Points a Solution to Model Instability? A Cautionary Tale from the Atlantic Surfclam Stock Assessment Regarding the Perils of Creative Thinking
14:50	Doug Kinzey	Using Randomized Phase Orders in the Assessment Model for Antarctic Krill
15:10	EVERYONE	BREAKOUT DISCUSSION: Best Practices for Choosing the Optimum Level of Model Complexity *First 45 minutes - moderated discussion on subtopic; last 15 minutes - exchange ideas with paired group in preparation for presenting unified recommendations and conclusions during plenary discussion* • Groups 1/5: How do you find the appropriate balance between realism and parsimony? • Groups 2/6: What is the process, and appropriate diagnostics, to ensure that the level of model complexity is optimal for your situation? • Groups 3/7: What are best practices for exploring a range of model complexities? • Groups 4/8: How should MSE/simulation studies be used to select model complexity level?
16:10	EVERYONE	ADJOURN FOR THE DAY
17:00	**EVERYONE**	**POSTER SESSION @ EMBASSY SUITES**

AGENDA Day 2 - Wednesday, May 23, 2018

TIME	SPEAKER	TITLE
8:00	EVERYONE	COFFEE
MODEL COMPLEXITY AND STABILITY SESSION - PART 2		
8:30	Isaac Kaplan	PLENARY: Of Sushi and Paella: Recipes for Ecosystem Models
~~9:10~~ **11:20**	Daniel Goethel	Accounting for Fleet Complexity and Stakeholder Objectives in the Establishment of Stock Status Determination Criteria and Annual Catch Limits: The Controversial Case of Gulf of Mexico Red Snapper
~~9:30~~	~~TBD~~ **WITHDRAWN**	~~Insights From Partial-Factorial Sensitivity Analyses Conducted to Evaluate the Importance of Data Collection Activities in the Southeastern United States~~
~~9:50~~ **10:40**	Robert Gamble	One Model - Many Parameters: A Multi-Parameter Inference Framework
~~10:10~~ **11:00**	Kari Fenske	Evolution of a Spatial Stock Assessment Model: A Moving Tale of Exploring Spatial Complexity
~~10:30~~ **10:25**	EVERYONE	BREAK
~~10:45~~ **9:10**	EVERYONE	BREAKOUT DISCUSSION: Best Practices for Using More Complex Models *First 50 minutes - moderated discussion on subtopic; last 25 minutes - paired group discussion* • Groups 1/5: How should you address impacts on stability from added complexity? • Groups 2/6: How should you balance the benefits of increasingly complex models with the potential shortcomings and time/resource costs? • Groups 3/7: How can you ensure that complex models effectively account for the hidden constants (e.g. natural mortality/growth/gear selectivity) in simple models? • Groups 4/8: How do you ensure that model complexity does not interfere with management practicality?
~~12:00~~ **11:40**	EVERYONE	LUNCH
13:30	EVERYONE	PLENARY DISCUSSION: Model Complexity and Stability *Breakout groups will reconvene together for a moderated and interactive discussion to share conclusions and best practices, share ideas, and work together to reach consensus towards the session objectives. Each pair of breakout groups should pick one spokesperson to address their subtopic from Part 1 and Part 2 (can be the same person, or two separate people). Plan to give a brief (~3 minute) summary of your group's conclusions, followed by discussion for each subtopic. Additionally, we will discuss hurdles that exist in each region to using the optimum complex models, and potential solutions.*
ENSEMBLE MODELING SESSION - PART 1		
15:00	Melissa Karp	Introduction to Session and Theme *Session Objectives: Fisheries science tends to employ methods that result in the selection of a single, 'best' model, but this can be problematic as there is uncertainty around the selection of a single model and potential loss of information in the models not selected. Ensemble modeling offers a different approach by combining results from multiple alternative models and data sets to generate projections of future system states. The use of ensemble modeling enables a more complete characterization of uncertainty in the system, particularly model structural uncertainty, in assessment outcomes. The objectives of this session are to (1) Identify benefits and challenges to using ensemble modeling in fisheries stock assessments, and (2) Identify 'best' practices for conducting ensemble modeling in fisheries stock assessment and using results for management decisions.*
15:10	Ian Stewart	PLENARY: From Concepts to Pragmatic Application of Ensemble Modelling
15:50	EVERYONE	BREAK
16:10	Liz Brooks	Exploring Ensemble Modeling with Different Model Complexities
16:30	Jon Brodziak	Forecasting Recruitment Using Model Ensembles
16:50	Kelli Johnson	Testing Algorithms for Ensemble Modeling Using Management Strategy Evaluation
17:10	Grant Thompson	Model Averaging and the Quest to Account for Full Uncertainty
17:30	EVERYONE	ADJOURN FOR THE DAY
18:30	**EVERYONE**	**SOCIAL - GUNWHALE ALES - 2960 RANDOLPH AVE., UNIT A, COSTA MESA, CA 92626**

AGENDA Day 3 - Thursday, May 24, 2018

TIME	SPEAKER	TITLE
8:00	EVERYONE	COFFEE
ENSEMBLE MODELING SESSION - PART 2		
8:30	Hendrik Tolman	PLENARY: Model Forecast Accuracy and Uncertainty as Key Elements for Decision Support: A National Weather Service Perspective
9:10	James Ianelli	Selecting the “Best” Model Ensemble of Ensembles
9:30	Jonathan Deroba	Assemble an Ensemble for Gulf of Maine/Georges Bank Atlantic Herring Stock Assessment
9:50	EVERYONE	BREAKOUT DISCUSSION: Evaluating the Benefits and Challenges of Ensemble Models in Fisheries Management *First 45 minutes - moderated discussion on subtopic; last 15 minutes - paired group discussion* • Groups 1/4: What are the benefits and shortcomings of utilizing ensemble modeling in fisheries stock assessment and implications for management? • Groups 2/5: When are ensemble models appropriate for fisheries? • Groups 3/6: What are the main challenges to operationalizing ensemble modeling in stock assessment, and how do we overcome those challenges?
10:50	EVERYONE	BREAK
11:10	James Thorson	Using Multivariate Predictions for Steepness, Natural Mortality, Growth, and Maturity Parameters to Inform Ensemble Models and Decision Tables
11:30	Merrill Rudd	Ensemble Models for Data-Poor Assessment: The Value of Life-History Information
11:50	Kerim Aydin	Comparing Single-Species and Multi-Species Reference Points Across a Range of Models - Do Alternate Viewpoints Make an Ensemble?
12:10	EVERYONE	LUNCH
13:40	EVERYONE	BREAKOUT DISCUSSION: Best Practices for Running/Conducting Ensemble Models *First 50 minutes - moderated discussion on subtopic; last 30 minutes - paired group discussion* • Groups 1/4: What are best practices for selecting models to include in an ensemble? • Groups 2/5: How should model predictions be combined to develop ensemble predictions (e.g. simple average, stacking, weighted average, boosted regression, superensemble, etc), and how should performance of the final ensemble prediction be evaluated? • Groups 3/6: How should results from ensemble models be communicated to, and used by, managers?
15:00	EVERYONE	BREAK
15:15	EVERYONE	PLENARY DISCUSSION: Ensemble Modeling *Breakout groups will reconvene together for a moderated and interactive discussion to share conclusions and best practices, share ideas, and work together to reach consensus towards the session objectives. Each pair of breakout groups should pick one spokesperson to address their subtopic from Part 1 and Part 2 (can be the same person, or two separate people). Plan to give a brief (~3 minute) summary of your group’s conclusions, followed by discussion for each subtopic.*
16:30	EVERYONE	Wrap Up • *Suggesting Themes for NSAW 2020* • *What to Expect Next*
17:00	EVERYONE	ADJOURN FOR THE DAY
18:00	**EVERYONE**	**SOCIAL - TRADE FOOD HALL - 2222 MICHELSON DR., IRVINE, CA 92612**

13TH NSAW POSTER TITLES
Tuesday, May 22, 2018 | 5:00pm

#	LEAD AUTHOR	TITLE
P-1	Felipe Carvalho	JABBA: Just Another Bayesian Biomass Assessment
P-2	Nicholas Grunloh	A Bayesian Model Averaging Approach for Improving Catch Estimation Methods in Sparsely Sampled, Mixed-Stock Fisheries
P-3	D. R. Hart	Estimation of Juvenile Natural Mortality in a Size-Structured Stock Assessment Model
P-4	Xi He	Utilities of Abundance Indices in Stock Assessment Models: Examples from Recent West Coast Stock Assessments
P-5	J. Jeffery Isely	Incorporation of Multiple Indices in Two Data Limited Assessment Models
P-6	Sean Lucey	Implementing Management Strategies Using an Ecosystem Model as an Operating Model
P-7	Josh Nowlis	A Fisheries "Theory of Everything" - The Architecture of Harvest Control Rules
P-8	Mark Smith	Data Gaps and Alternative Management Approaches: Applying National Standard 1 to Data-Limited Stocks
P-9	Laura Solinger	Developing Spatially Explicit Stock Assessment Models to Explore Spatial Variability in Recruitment of Chilipepper Rockfish (*Sebastes goodei*) and Possible Links to Environment
P-10	Steven Teo	Model Ensembles of Common Thresher Sharks in the Northeast Pacific Ocean: Comparing "Traditional" Stock Assessment and Model-Averaging Approaches, and Something with a Little Bit of Both
P-11	Desiree Tommasi	Evaluating Alternative Management Strategies for North Pacific Albacore

APPENDIX E: PRESENTATION AND POSTER ABSTRACTS

Session 1: Model Complexity and Stability

Keynote: Goldilocks and the Three Assessments: This Assessment is Too Simple, This Assessment is Too Complex, and This Assessment is Just Wrong! By *Mark Maunder*

Choosing a model for stock assessment is more complicated than simply determining the most parsimonious model based on standard statistical model selection criteria. The model should be designed for the objective of the analysis, and the complexity of the model and data used may differ among objectives. We use the age structured production model (ASPM) diagnostic applied to several tuna stocks (yellowfin, bigeye, albacore, and Bluefin tuna in the Pacific Ocean) to illustrate the issues in determining the appropriate level of complexity. Under the typical objectives of providing management advice (e.g., calculating the maximum sustainable yield, or MSY), the goal should be to find the fishing effect on population abundance as measured by an index of relative abundance. The first task should therefore be to ensure that the index represents abundance and that the catch is removed at approximately the correct age. Next, the ASPM should be applied to determine if it can reveal the impact of fishing on the index of relative abundance. Often, this model is "too simple" and estimates of recruitment are needed to appropriately extract the absolute abundance information from the index of relative abundance. Simply estimating the recruitment deviates in the ASPM often creates a model that is "too complex," because there is insufficient information to estimate the annual recruitments. Composition data generally provides the best information about recruitment. However simply including the composition data in the stock assessment model can lead to an assessment that is "just wrong" because influential processes (e.g., selectivity and growth) are misspecified and down weighting the composition data is not an appropriate solution. The model misspecification needs be fixed. The next step is essentially fine tuning our ability to reveal the effects of fishing on population abundance and may involve appropriately modeling patterns in recruitment. Finally, we provide advice on how to construct a stock assessment model.

Transitions in Model Complexity for Assessments for the Deep7 Bottomfish Fishery in the Main Hawaiian Islands, By *Brian Langseth*

The Deep7 bottomfish fishery in the Main Hawaiian Islands is assessed as a complex of seven species using a surplus production model (SPM). Sparse data on species-specific life history parameters have limited the ability to assess each species individually, while a lack of catch- or size-at-age data have limited the ability to utilize structured assessment models. However, species-specific life history data are becoming more available, and average weight of the catch by species from the fishery are available, allowing for comparisons between models of varying complexity and levels of species aggregation. Available data from the dominant species in the Deep7 complex were used within an integrated assessment model. Using a diagnostic tool, results from the integrated model were compared to results from a single species SPM and an SPM for the Deep7 complex to determine the information gained from the added model complexity. We found that for our case study, the majority of information on stock productivity was contained within the abundance index and not in the weight data, and therefore results from the single-species models were similar. Results from the single-species SPM scaled to results from the SPM for the complex. Given current data quality and trends, it appears that the SPM for the complex was sufficient to capture general trends in the stock, and that for the dominant species, greater model complexity using data that added limited new information was not warranted. Overall, our process informs a general approach for transitioning between models with varying levels of complexity to support available data.

Are Scale-Free Reference Points a Solution to Model Instability? A Cautionary Tale from the Atlantic Surfclam Stock Assessment Regarding the Perils of Creative Thinking, By *Dan Hennen*

The stock assessment for the Atlantic surfclam, a large, commercially important bivalve in the Northwest Atlantic, showed considerable model instability. Though trend was relatively well described by the model, it was highly uncertain in scale. In the 2017 benchmark assessment, trend-based reference points were developed and deemed appropriate for management by a peer review committee. These reference points, however, caused some consternation when applied by managers. The causes for instability in the model, some possible mitigations for them, the development of trend-based reference points, as well as the experiences of the lead assessment scientist in trying to explain them to managers will be discussed.

Using Randomized Phase Orders in the Assessment Model for Antarctic Krill, By *Doug Kinzey,* G.M. Watters, and C.S. Reiss

An integrated assessment model for Antarctic krill includes 108 potentially estimable parameters. We developed a procedure for sequentially estimating increasing numbers of parameters in replicate sets using different randomized phase orders assigned by an R script. Generally, as more of these parameters were estimated rather than pre-specified, the model configurations fit the data better, but configurations attempting to estimate all or nearly all of the parameters did not converge. Some configurations estimating many parameters produced an invertible Hessian matrix that allowed MCMC sampling but did not pass coda diagnostics for stationarity at up to 200 million MCMC samples. Also, some replicates in the configurations estimating many parameters produced estimates representing a local minimum rather than the lowest negative log-likelihood observed using a different ordering of the estimation phases. For models with fewer than 50 estimated parameters, the phase order in which they were estimated didn't matter and all replicates converged to the same negative log-likelihood, although possibly requiring several re-randomizations of the phase order before converging. Configurations with up to 107 estimated parameters produced invertible Hessian matrices but did not all pass coda diagnostics. The best-fitting model that was able to pass coda diagnostics had 94 estimated parameters. Twenty-four parameters remained at pre-specified values in this configuration. Some parameters, such as those for sigmar, steepness, or von-Bertalanffy growth, were not estimable even in models estimating only a few other parameters.

Keynote: Of Sushi and Paella: Recipes for Ecosystem Models, By *Isaac Kaplan*

Ecosystem modeling broadens the consideration of harvested stocks to include prey availability, predators, oceanographic effects, or other drivers of population dynamics and spatial distributions. This can be a recipe for excessive model complexity. Using examples related to sardine (Koehn et al. 2016; Punt et al. 2016; Kaplan et al. 2017) and salmon (Chasco et al. 2017), I first illustrate how pitfalls related to ecosystem model complexity can be avoided by tailoring models to the key management questions and progressively phasing in components. Second, I illustrate cross-regional comparisons (Olsen et al. 2018) that can be used to understand the relative importance of ecosystem drivers on harvested stocks, or on the ecosystem. This type of comparison, and better quantification of parameter uncertainty, has been greatly facilitated by recent advances in computing and steps toward ensemble approaches. Third, I present some recent examples where ecosystem models have been incorporated for strategic management purposes in the U.S., including alongside stock assessments (for Tier I harvest specifications) and as operating models in new management strategy evaluations (NMFS 2015; Lucey 2018). Finally, I note some convergent evolution and needs between ecosystem models and stock assessments, with respect to life-history parameters, spatial distributions, and model review.

A goal of this talk is to poll the audience and ask questions including:

- How are you harnessing new advances in computing resources?
- What cross-regional comparisons (meta-analyses) are you engaged in?
- How (by what logic) do you limit model complexity, and how would you do so if you were not estimating all parameters?
- What are convergent paths for stock assessment and ecosystem models, for instance via MSE or estimation of life-history parameters and spatial distributions?

Accounting for Fleet Complexity and Stakeholder Objectives in the Establishment of Stock Status Determination Criteria and Annual Catch Limits: the Controversial Case of Gulf of Mexico Red Snapper, By *Daniel R. Goethel*, Matthew W. Smith, Shannon L. Cass-Calay, and Clay E. Porch

Maximum sustainable yield (MSY) based reference points are prescribed by the Magnuson-Stevens Act as the basis for annual catch limits. However, identification and calculation of MSY is complicated by the existence of bycatch fleets or mortality due to discarding. Projections based on the Gulf of Mexico Red Snapper fishery, one of the most complex assessment management scenarios in the U.S., are utilized to demonstrate the various ways that MSY can be computed when multiple fleets and bycatch fisheries exist, and to illustrate the tradeoffs that occur between yield and spawning-stock biomass (SSB). Presenting the full array of alternative MSY proxies can lead to subjective decision making that may encourage the maximization of yield at the expense of maintaining stocks within safe biological limits. We propose that the spawning potential ratio (SPR) associated with SSB-MSY can be utilized as a target reference point in most fishery applications, where the corresponding yield streams used to achieve it are conditional on extant selectivity patterns and bycatch levels. Our approach maintains the inherently sustainable SSB associated with MSY as a target reference point, while explicitly accounting for current fleet dynamics and avoiding unsustainable proxies that may result when bycatch rates are high. Additionally, it adheres to the Magnuson-Stevens Act by maintaining the spawning stock at or above the level that will produce the MSY, while reducing yield streams to account for extant economic and social factors that contribute to suboptimal selection and bycatch patterns. Despite SPR-MSY providing a sustainable overfishing metric, resulting yield streams may not satisfice the sometimes-competing objectives inherent in Red Snapper fishery management decisions of extended recreational fishing seasons and high commercial yield.

One Model – Many Parameters: A Multi-Parameter Inference Framework, By *Robert J. Gamble*, Sarah K. Gaichas, Andy Beet, and Laurel A. Smith

A potential problem in more complex models, especially those with parameters that have few empirical observations to support them (e.g., competition or predation interaction terms, environmental effects in multispecies production models) is that there can be widely varying sets of parameters that provide reasonable fits to observed data. In many cases, the fitness landscape can have multiple local minima, and many traditional optimization routines are unable to find the global minima but instead are dependent on the initial parameter estimates. We evaluate the use of a genetic algorithm to find a number of potentially valid parameter estimate sets, and further apply the principles of multi-model inference. Our framework will be a relatively simple, 10-species production-estimation model that incorporates competition, predation, and density-dependent terms, with a more complicated length-based multispecies operating model that generates the test data. We will generate 30 to 50 years of data from the operating model, and fit the multispecies production model to it, holding out the last 10 years. We will use the 10 best-estimated parameter sets to forecast the last 10 years of data, and compare the results of each individual parameter set to the average forecast results from all 10 parameter sets. In so doing, we will evaluate the potential benefits of using a multi-parameter inference framework as opposed to using just the best-fitting parameter set.

Evolution of a Spatial Stock Assessment Model: a Moving Tale of Exploring Spatial Complexity, By *Kari Fenske* and Dana Hanselman

Federally managed sablefish in Alaska are highly mobile, valuable, and relatively data rich due to a time series of fishery-dependent and fishery-independent data. Sablefish are the only species in Alaska that are assessed as one stock through the Gulf of Alaska, Bering Sea and the Aleutian Islands because of their high movement rates. However, sablefish are managed under an individual fishing quota (IFQ) program with quota shares tied to management areas, so preservation of biomass in all management areas is a key concern for constituents and a spatial model was undertaken to try to develop robust estimates of regional biomass. Using externally estimated movement rates, we developed a spatial stock assessment model that estimates biomass and reference points for three spatial regions. Despite the large amount of data, we encountered challenges in estimating key spatial parameters for selectivity, catchability, and recruitment, and were forced to fix parameters across areas or share them between sexes, sometimes resulting in a compromise to fit to data in exchange for model stability. We will share our experiences with spatial model building for Alaska sablefish as we worked from a simple model with few spatially estimated parameters to a very complex model with many spatial parameters, how we decided which parameters should or should not be estimated spatially, what we learned about how much spatial complexity was needed to balance between a model that easily converges and a model that sufficiently describes the spatial dynamics of sablefish. Another important consideration is that the data requirements for spatial models increases greatly as data are parsed into smaller and smaller areas. In addition, we comment on future directions for this assessment.

Session 2: Ensemble Modeling

Keynote: From Concepts to Pragmatic Application of Ensemble Modeling, By *Ian Stewart* and Allan Hicks

This talk by representatives of the International Pacific Halibut Commission will introduce a conceptual basis for adopting an ensemble approach in support of probabilistic results for use in fisheries management decision making. Many of the impediments to the use of multiple models have been addressed in classical model selection and multimodel inference methods, but additional considerations apply for commonly used fisheries approaches. Some potential benefits of ensemble modeling, including improved interannual stability, robustness to changes in the performance of individual models and to the analysts constructing them, as well as some of the downsides of ensemble usage in increased technical overhead, challenges in effectively communicating results, and conceptual disagreements among analysts will also be discussed. The International Pacific Halibut Commission has been using an ensemble of assessment models to provide management information for the last six years. Several aspects of our specific experience and examples from other ensemble applications will be used to highlight the efficacy of the approach.

Exploring Ensemble Modeling with Different Model Complexities, By *Liz Brooks* and Jon Brodziak

Model uncertainty and model complexity are important considerations when fitting data for stock assessment and providing management advice. Model uncertainty can include structural decisions as well as alternative parameterizations, while model complexity can refer to the degree of data aggregation and statistical treatment of the data. In practice, management advice typically comes from one "best" model, with both analysts and reviewers relying on comparisons with alternative models to provide reassurance that the main trend has been captured, rather than trying to quantifying overall uncertainty. We summarize simulations that explore ensemble modeling across multiple assessment and projection models, and make comparisons between different assessment models of varying complexity on both an absolute and relative scale. For each assessment and forecast model ensemble, we explore the issues of selecting a set of candidate models, how different candidate models should be from each other, whether or not the true model is bounded by the model set, and sensitivity to the weighting factors used to combine the model results. From each assessment model in the candidate set, the current estimates of SSB and F and the associated reference points are weighted to produce ensemble stock status; similarly, ensemble catch advice is produced from the set of projection models. We compare the performance of this ensemble approach with the status quo "single best model" approach, and discuss important issues related to performance consistency (retrospective patterns) and data weighting. Comparisons between ensembles of different model complexities is an open question, and we summarize some initial thoughts.

Forecasting Recruitment Using Model Ensembles, By *Jon Brodziak* and Liz Brooks

We describe a general approach to forecasting recruitment, and associated quantities of interest (QOI) such as future spawning biomass or total allowable catch, using model ensembles. In nearly all assessments, there is uncertainty about the stock assessment model structure and about the forecast model structure conditioned on the assessment model. We use multimodel inference to choose a set of candidate assessment models. Each candidate assessment model is run for the full assessment time horizon, and model weights are calculated based on model fits to sequential subsets of the time horizon. The model weights are based on an objective measure such as the mean squared error of the model fit to a relative abundance index or to size composition data. Forecast model weights measure the predictive accuracy between projected QOI and the subsequent QOI estimates from the next assessment iteration. This two-stage process produces a model ensemble comprised of a set of credible assessment models, each of which has an associated set of forecast models along with both assessment and forecast model weights. The ensemble model with the associated model weights provides a probabilistic characterization of both assessment model and forecast model-based uncertainties for QOI. The capability to characterize forecast uncertainties is important because fisheries are actively managed for the present and future, and not the past. We illustrate this two-stage forecasting approach using recent assessment information for the Georges Bank haddock stock, including uncertainty in the magnitude and scaling of natural mortality rate with body mass.

Testing Algorithms for Ensemble Modeling Using Management Strategy Evaluation, By J. Brodziak, L. Brooks, R. Hillary, E. Jardim, K.F. Johnson, N. Klibansky, C. Legault, A. Magnusson, A. Mannini, C. Millar, C. Minto, I. Morqueira, R. Nash, E. Olsen, K. Shertzer, S. Subbey, P. Vasilakopoulos, and B. Wells

Ensemble modeling in the field of natural resource management allows for the combination of results from several models in the creation of novel products for scientific advice. A select few ensemble algorithms have been explored in fisheries science such as simple averaging, performance-based averaging, and model-selection-based averaging, but a best-practice guide for how to choose the most appropriate weighting algorithm with respect to stock assessment models is needed. The utility of several algorithms will be assessed using the Fisheries Library for R and the Assessment for All R packages that currently include the following types of stock assessment models: extended survivor analysis, statistical catch-at-age, state-space assessment, and Pella and Tomlinson. These two R packages will facilitate the quantification of ensemble-algorithm performance under alternative management strategies and procedures to multiple axes of uncertainty in a reproducible and timely manner. Currently, the proposed case studies include the following stocks: Barents Sea capelin (*Mallotus villosus*), Georges Bank haddock (*Melanogrammus aeglefinus*), Iberian hake (*Merluccius merluccius*), red grouper (*Epinephelus morio*), Pacific blue marlin (*Makaira nigricans*), and U.S. West Coast lingcod (*Ophiodon elongatus*). Deliverables will include a comparison of the robustness of estimates of stock status and management advice (e.g., total allowable catch) from ensemble methods versus status quo management, ultimately strengthening the support necessary for evidence-based policy making.

Model Averaging and the Quest to Account for Full Uncertainty, By *Grant G. Thompson*

In model averaging, a sample of models is first developed, and then the distributions from those models, for some quantity of interest, are averaged (unweighted or weighted). One of the arguments advanced in support of model averaging is the assertion that, unlike use of a single "best" model, model averaging accounts for the full uncertainty surrounding the quantity of interest. However, being based (by definition) on only a sample from the population of all possible models, this assertion is at best an approximation, as the full uncertainty is embodied only in the population of all possible models. The issue of whether it is possible to improve upon the (unweighted or weighted) averaged sample distribution is the subject of this presentation. Simulation is used to compare five candidate distributions to the true population distribution: (1) the distribution from the single best model, (2,3) the (unweighted, weighted) averaged sample distribution, and (4,5) the (unweighted, weighted) estimate of the population distribution obtained by equating moments of a flexible parametric distribution to those of the (unweighted, weighted) averaged sample distribution. The simulations consider the case in which the sample of models is drawn randomly from the population of all possible models as well as several cases in which the sample of models is biased in various ways.

Keynote: Model Forecast Accuracy and Uncertainty as Key Elements for Decision Support: A National Weather Service Perspective, By *Hendrik Tolman*

For more than 60 years, computer models for our physical environment have been the foundation of weather forecasts. As the models have become more accurate, we have started to realize that some processes are inherently chaotic, and have a limited predictability. Moreover, the predictability varies with actual conditions, and can also be predicted. Furthermore, the value of a model forecast is not in the forecast itself, but in how it directs decision making. All this is moving weather forecast models towards an ensemble approach to modeling. The presentation will address the scientific foundations of model accuracy and uncertainty, some technical aspects of how to create model ensembles that accurately describe or predict uncertainty, and how the NWS has evolved in using such approaches.

Selecting the "Best" Model Ensembles of Ensembles, By *James Ianelli*

Data weighting practices in modern stock assessment approaches typically strive to specify and improve estimates of uncertainty. However, such approaches imply that the model specified is correct and that estimated observation errors are wrong. In this paper we provide example alternative model specification approaches with different data-error feedback assumptions. In addition, the role of ensemble modeling is considered in the context of actual fishery management problems, i.e., assessment results that affect catch limits. We note that the trend towards increasingly complex model-averaging methods results in greater difficulty in reviewing and providing constructive advice. Therefore, we advocate better transparency in tested harvest control rule procedures. This involves providing data-driven advice from simpler assessments as tested against a wide-ranging set of structurally diverse model ensembles.

Assemble an Ensemble for Gulf of Maine/Georges Bank Atlantic Herring Stock Assessment, By *Jonathan Deroba*

During approximately the past 15 years, a range of assessment model types (e.g., delay difference, statistical catch-at-age, state-space) that have used different data and structural assumptions (e.g., natural mortality) have been applied to Gulf of Maine/Georges Bank Atlantic herring. The results of these assessment model fits, however, have never been synthesized to create an ensemble. The objective of this research was to create such an ensemble, and compare time series results and diagnostics among them. In particular, focus was placed on time series of spawning stock biomass (SSB), retrospective patterns, and reference points. Despite a broad range of severity in retrospective patterns and model structures, the majority of assessment models had qualitatively similar time series estimates of SSB, especially in trends among years. Some changes to model structure, however, such as natural mortality, had large effects on MSY reference points. While the presence of a retrospective pattern is likely to indicate incorrect model structure, the presence or absence of a retrospective pattern may not be informative as to the ultimate validity of an assessment model. Using reductions in retrospective patterns to justify structural changes in assessment models is also therefore not a valid practice. In cases where most assessment models in an ensemble are generally consistent, efforts should focus on identifying reference points robust to the structural uncertainties in the ensemble. Constructing ensembles can identify those models that are generally inconsistent and may therefore be seen as having greater uncertainty than other models, which should be considered when making management decisions.

Using Multivariate Predictions for Steepness, Natural Mortality, Growth, and Maturity Parameters to Inform Ensemble Models and Decision Tables, By *James Thorson*

Stock assessment models require estimating or fixing parameters representing basic demographic processes, including natural mortality, growth, sexual maturation, and recruitment of juveniles. Recent research has predicted adult life-history parameters (representing natural mortality, growth, and maturity) for all fishes worldwide based on available data, life-history correlations, and taxonomic similarity. I quickly review this research, including accessing results through R package FishLife. I then describe ongoing research to additionally predict stock-recruit parameters including steepness and the standard deviation and autocorrelation in recruitment variation. Preliminary results suggest that steepness is not correlated with major axes of life-history variation (either body size or the timing of maturation), but is clustered taxonomically (e.g., with low steepness for family *Sebastidae* and high for *Lutjanidae*). I then discuss how results can be used for stock assessment including: (1) fixing values in data-poor assessment models; (2) defining a set of life-history parameters to use an ensemble model; and (3) defining an axis of uncertainty for decision tables that represents model sensitivity to the aggregate effect of life-history variation. I conclude by recommending ongoing research to predict, evaluate, and distribute life-history predictions, as well as research to explore the likely performance of these different uses in data-poor and data-rich assessments.

Ensemble Models for Data-Poor Assessment: The Value of Life-History Information, By *Merril Rudd*, James Thorson, and Skyler Sagarese

Scientists and resource managers need to understand a population's biological parameters, such as mortality and individual growth rates, to successfully manage exploitation and other risks. In the absence of age data, growth rate estimates come from direct observation, proxies, or patterns in the length data, while natural mortality rate estimates are often assumed based on empirical relationships with the growth-rate estimates. Length composition of the catch can inform recruitment and fishing mortality, but parameter estimates depend on accurate growth and natural mortality rates. Uncertain or unreliable mortality and growth rates propagate high uncertainty in stock status estimates when relying on length data to inform vital stock assessment parameters. This study uses predictive stacking as a method of model averaging across a distribution of values for growth and mortality. We used the R package FishLife to develop distributions of life history parameters based on a multivariate model with taxonomic structure, drawing combinations of points from these distributions to integrate uncertainty in life history parameters into a data-limited, length-based stock assessment. Through simulation we demonstrate that predictive stacking leads to better assessment performance than assuming the parameter means from FishLife when the true values of life history parameters are unknown. We then applied the predictive stacking method for a U.S. Caribbean stock previously lacking accepted management advice due to debilitating uncertainty in life history parameters. This method will be applicable for stock assessments concerned with properly accounting for uncertainty in biological parameters, ranging from life-history-based to length-or age-based stock assessments.

Comparing Single-Species and Multi-Species Reference Points Across a Range of Models – Do Alternate Viewpoints Make an Ensemble? By *Kerim Aydin*, Kirstin Holsman, Jim Ianelli, Jonathan Reum, and Andy Whitehouse

The development of greater skill and techniques in multispecies and ecosystem models in recent years has produce suites of fitted, calibrated and peer-reviewed models that provide "alternate viewpoints" on management reference points, such as unfished biomass or maximum sustainable yield. For the eastern Bering Sea, several models have been constructed that include key fish stocks, including single species assessment models, multispecies assessment models, food web models, and size-structured models. These have entered the management arena; for example, the multispecies stock assessment model CEATTLE has been used and reviewed as part of the Walleye Pollock stock assessment to provide alternate views of stock status and recruitment relationships. Further, as part of the NOAA Alaska Climate Integrated Modeling (ACLIM) Project, these models have been driven forward under the Intergovernmental Panel on Climate Change (IPCC) climate-forcing scenarios to predict changing reference points which may differ between models. Here, we present the results from this modeling suite and explore whether ensemble combinations are appropriate, or whether the best information and advice is derived by examining and explaining the differences with less formal quantification.

Topics of National Interest

National Standard 1 Technical Guidance Workgroup Status Update, By Erin Schnettler, Patrick Lynch, *Melissa Karp*, and Rick Methot

In October 2016, NOAA Fisheries published revisions to the National Standard 1 (NS1) guidelines. The guidelines were last revised in 2009, but NOAA Fisheries has not produced updated technical guidance for the NS1 guidelines since 1998 (Restrepo et. al., 1998). Because the guidelines have changed significantly since 1998, updates to the existing NS1 technical guidance are needed based on new concepts that were added in the 2009 and 2016 guideline revisions. Thus, the NS1 Technical Guidance Workgroup was formed in order to develop updated NS1 technical guidance that will strengthen NS1-related management decisions. Rather than develop one comprehensive NS1 Technical Guidance document, the workgroup has selected key strategic topics of the NS1 Guidelines to address. Based on these topics, the workgroup split into three subgroups focused on reference points, harvest policies, and best practices/alternative approaches for data-limited stocks. All three subgroups have work plans that prioritize work products for development. This presentation will highlight progress on work products associated with all three subgroups, including, but not limited to: 1) Guidance on a range of, and default for, suitable proxies for F_{MSY} (i.e., the maximum rate of fishing mortality) and B_{MSY} (i.e., the biomass achieved by fishing at F_{MSY}); 2) Advice and recommendations for designing, evaluating, and implementing carry-over and phase-in provisions and associated control rules; and 3) Best practices and recommendations for when and how to use alternative approaches to meeting the annual catch limit (ACL) requirement in data-limited situations.

Assessment Methods Work Group Report, By *Owen Hamel*

The NOAA Fisheries Assessment Methods Working Group (AMWG) provides a forum for stock assessment scientists from across the nation to discuss assessment practices and challenges from the different management regions of the United States. The aim of this group is to foster development, testing and maintenance of high-quality assessment methods and provide tested, well-documented models that efficiently incorporate all available data and provide well understood and accurate assessments. The group meets regularly to discuss methods and projects, and previously served as the body reviewing funding proposals submitted to the Stock Assessment Analytical Methods and Improve a Stock Assessment RFPs. The AMWG is currently conducting a national project to comprehensively evaluate and document the various modeling approaches being used in stock assessments across the country. Where appropriate, the group is also providing good practice recommendations and identifying gaps. The AMWG chair will provide an overview of past and planned AMWG activities.

National Stock Assessment Program Strategic Plan Overview, By *Alicia Miller*

Fishery stock assessments are a fundamental component of sustainable fisheries management under the Magnuson-Stevens Fishery Conservation and Management Act. Significant advances have been made within the national stock assessment enterprise since publication of the Stock Assessment Improvement Plan in 2001. Funding increases have been key for advancing stock assessment science provided through the Expand Annual Stock Assessments budget line. Since 2014, the National Stock Assessment Program (NSAP), in collaboration with senior science advisors and contributing partners have drafted *Implementing a Next Generation Stock Assessment Enterprise: An Update to NOAA Fisheries' Stock Assessment Improvement Plan*. This document outlines the vision for the next generation of NOAA Fisheries' assessments including specific recommendations for achieving this vision. The NSAP strategic plan identifies the five-year strategic vision for the NSAP and describes how NSAP activities support NOAA Fisheries' strategic priorities. This plan is intended to be an internal document to facilitate communication within NMFS.

Implementing Recommendations of the Next Generation Stock Assessment Improvement Plan, By *Patrick Lynch*

NOAA Fisheries conducts stock assessments to provide fundamental scientific advice in support of sustainable fisheries management. Managers use the results of stock assessments, along with other information, to establish catch targets and limits that strive to maximize yield while ensuring that overfishing does not occur and stocks do not become overfished. While nationwide, NOAA Fisheries is currently completing quality assessments at a high rate, there are increasing demands and challenges facing NOAA's stock assessment programs. This document provides a summary of NOAA's Next Generation Stock Assessment (NGSA) framework, which is an update to the 2001 Stock Assessment Improvement Plan (SAIP). The NGSA strategic vision is designed to complement NOAA Fisheries' other strategic efforts in order to accomplish its mission of sustainable fisheries through resource conservation and management. The NGSA framework has three main themes:

1. Expanding the scope of stock assessments to be more holistic and ecosystem-linked
2. Using innovative science and advanced technologies to improve data and analytical methods
3. Establishing a more timely, efficient, and effective stock assessment process

The new NGSA framework acts as a road map to guide efforts by NOAA Fisheries to address these three themes. First, it advocates for expanding the scope of stock assessments to support harvest policies that are more holistic in nature. This means that more stock assessments will consider ecosystem and socioeconomic factors that affect the dynamics of fish stocks and fisheries. Such expansion aligns with the "Tier III" goal of the 2001 SAIP and it is reemphasized as a priority here, accompanied with decision trees that help determine when this information is of greatest importance and how it should be incorporated. Second, it advocates for the continued use of innovative technologies to reliably and efficiently provide data for maximizing use of advanced modeling methods. Examples of clear benefits from this emphasis include improved calibration of data collection methods, streamlined analytical processes, and establishment of robust harvest policies to manage fisheries between assessments. Finally, the plan provides a method for objectively determining stock-specific goals that create a stock assessment process that is more timely, more efficient, and more effective at optimizing available resources and delivering results to fishery managers and the public. Ultimately, the NGSA framework will achieve the best balance among the "4Ts" of stock assessment: throughput, timeliness, thoroughness, and transparency. Implementation of the NGSA framework will require strong collaboration among NOAA Fisheries, management partners, and stakeholders. Leveraging the strong data collection partnerships already in place will help achieve improvements in data collection, processing, and management. Similarly, enhancing and expanding assessments to include new data types can be accomplished through cooperation and utilization of diverse platforms that provide low-cost options and opportunities for interdisciplinary collaboration. Investments in training and retention of assessment scientists will be paramount for capitalizing on recent advancements in software and statistical modeling techniques. Finally, standardizing aspects of the assessment process, while emphasizing regional priorities through national initiatives such as classifying data inputs, setting targets for assessment level and frequency, and conducting gap analyses will focus productivity and increase communication to stakeholders and the public. The new NGSA framework helps NOAA Fisheries accomplish its mission of conserving healthy ecosystems while achieving productive and sustainable fisheries. The specific actions and recommendations under each theme facilitate the transition to an NGSA enterprise. These are provided as goals that will improve NOAA Fisheries' ability to meet its mandates. They are not prioritized or associated with specific timelines or resource requirements or reallocations. Rather, the items provide a directional framework that NOAA Fisheries can use to ensure a high quality and quantity of stock assessments that meet the growing demands of the fishery and management process.

Poster Presentations

JABBA: Just Another Bayesian Biomass Assessment, By *Felipe Carvalho*, Maia Kapur, and Henning Winker

This study presents a new, open-source modeling software entitled Just Another Bayesian Biomass Assessment (JABBA). JABBA can be used for biomass dynamic stock assessment applications, and has emerged from the development of a Bayesian State-Space Surplus Production Model framework, already applied in stock assessments of sharks, tuna, and billfishes around the world. JABBA presents a stable, flexible framework for biomass dynamic modeling, runs quickly and generates reproducible stock status estimates and diagnostic tools. Specific emphasis has been placed on allowing the user to specify alternative scenarios of varied model complexity, achieving high stability and rapid convergence rates. Default JABBA features include: 1) an integrated state-space tool for averaging and automatically fitting multiple abundance (i.e., catch per unit effort, or CPUE) time series; 2) data-weighting through estimation of additional observation variance for individual or grouped CPUE; 3) selection of Fox, Schaefer, or Pella-Tomlinson production functions; 4) options to fix or estimate process and observation variance components; 5) model diagnostic tools; 6) future projections for alternative catch regimes; and 7) a suite of inbuilt graphics illustrating model fit diagnostics and stock status results. As a case study, JABBA is applied to the 2017 assessment input data for South Atlantic swordfish (*Xiphias gladius*). We envision that JABBA will become a widely used, open-source stock assessment tool, readily improved and modified by the global scientific community.

A Bayesian Model Averaging Approach for Improving Catch Estimation Methods in Sparsely Sampled, Mixed-Stock Fisheries, By *Nicholas Grunloh*, Edward Dick, and John Field

Effective management of exploited fish populations requires accurate estimates of commercial fisheries catches to inform monitoring and assessment efforts. In California, the high degree of heterogeneity in the species composition of many groundfish fisheries, particularly those targeting rockfish (genus *Sebastes*), leads to challenges in sampling all potential strata, or species, adequately. Limited resources and increasingly complex stratification of the sampling system inevitably leads to gaps in sample data. In the presence of sampling gaps, current methods for speciating commercial landings provide ad-hoc point estimates of species compositions in unsampled strata by “borrowing” data across adjacent stratum in time and space. Due to complex interactions between biogeography and market category sorting dynamics, it is not possible to be certain about optimal *a priori* pooling strategies. Here we introduce a Bayesian Model Averaging (BMA) method for discovering quantitatively justifiable pooling strategies by averaging across exhaustive sets of spatially partitioned models. In combination with Bayesian hierarchical modeling, these methods allow us to infer pooling strategies from port sampling data. Furthermore, combining Bayesian hierarchical models with BMA allows for a complete statistical summary of the major sources of uncertainty in species composition estimates.

Estimation of Juvenile Natural Mortality in a Size-Structured Stock Assessment Model, By *Dvora Hart*, J.H. Chang, and L.D. Jacobson

Natural mortality is one of the processes that is most difficult to estimate, and yet it strongly affects model outputs such as reference points and biomass. For sea scallops (*Placopecten magellanicus*), juveniles (20-90 mm shell height) are active swimmers, which is an effective defense against sea star predation. However, their thin shells make them vulnerable to predation by decapods, most particularly Cancer crabs. The stronger shells of adults, by contrast, protect them against Cancer predation, but their lack of mobility makes them vulnerable to predation from sea stars. Thus, the natural mortality of juveniles and adults might not be the same or similar. Surveys are able to detect juvenile sea scallops about two years before they enter the fishery as adults. Because survey selectivity is known, juvenile natural mortality is estimable and not strongly confounded with fishing mortality. We estimate juvenile natural mortality within the size-structured CASA (statistical catch at size) model assuming a fixed value for adult mortality, and compare these results to model runs that assume adult and juvenile natural mortality are the same, as was done in previous assessments.

Utilities of Abundance Indices in Stock Assessment Models: Examples from Recent West Coast Stock Assessments, By *Xi He*, John Field, Edward Dick, and Melissa Monk

We examined how abundance indices have been used in recent stock assessment models from the U.S. West Coast. These indices came from a variety of sources, including a combination of different fishing gears (trawl or hook and line), and both fishery-dependent and fishery-independent sampling programs. Most surveys target juvenile and adult fish, but fishery-independent surveys also sample young-of-year fish and larval abundance (the latter is typically used as a proxy for spawning output). To evaluate the relative influence of each index, indices were sequentially removed from the assessment model or were used as the sole index in the model. Assessment outputs, including estimates of key stock assessment parameters, were compared among these models and with the original models in which all index data were utilized. We examined a total of 25 stock assessment models and found that abundance indices had very small effects on 10 assessment models, while they had very large effects on seven assessment models. We also compared index utilizations among fishing gears and sampling methods (fishery-dependent or -independent), and found there was no evidence that fishery-independent sampling had a greater influence on abundance estimates, although this could be a reflection of temporal differences in the availability of different data types. Our preliminary results also gave some indication that indices derived from trawl gear tended to be slightly more influential than those derived from hook and line sampling, and we are exploring simulations that may help to evaluate if and how the information content of abundance indices are related to qualities of other data (i.e., composition data) used in assessment models.

Incorporation of Multiple Indices in Two Data-Limited Assessment Models, By J. Jeffery Isley and *Skyler Sagarese*

Data-rich stock assessments often incorporate multiple indices. However, some data-poor stock assessment models limit inputs to a single index. As multiple indices are often available for data-poor species, the ability to incorporate multiple indices would be a valuable addition to the data-limited methods (DLM) Toolkit. Here, we introduce model changes to the I-slope and I-target models to incorporate multiple indices and evaluate these models using management strategy evaluation (MSE) techniques.

Implementing Management Strategies Using an Ecosystem Model as an Operating Model, By *Sean Lucey*, Sarah Gaichas, Kerim Aydin, Gavin Fay, Steve Cadrin, and Andre Punt

Management strategy evaluation (MSE) is an effective tool to gauge the relative performance of fishery management options. MSEs have mostly been applied to evaluate single-species management strategies. However, there is a growing desire to test management strategies in an ecosystem context. Managers and stakeholders recognize there are technical and biological interactions that are not always accounted for in single species management approaches. Ecopath with Ecosim (EwE) is a popular mass balance representation of an ecosystem that includes many features of interest for managers. Until recently, full feedback interactions between a management strategy and an EwE-based operating model were impractical. However, with the development of Rpath, an R implementation of the EwE algorithms, users now have the ability to fully customize the operating model to be conditioned on the outputs from external assessment models. This gives the user an opportunity to evaluate a range of management strategies in an ecosystem context. We demonstrate the ability to conduct an MSE using an example Georges Bank model in Rpath as the operating model. We input pseudo-data generated from the Rpath operating model to a simple surplus production model (discrete Schaefer model), and passed the results to harvest control rules that mimic characteristics of management for commercially-targeted species on Georges Bank. We evaluated benefits of particular strategies on the target species, along with impacts to other parts of the ecosystem. Using this method, the evaluation criteria for selecting a management strategy will be based on the inherent tradeoffs within the system.

A Fisheries "Theory of Everything" – the Architecture of Harvest Control Rules, By *Josh Nowlis*

Harvest control rules (HCRs) translate measures of fish stock status into management action. They are a staple tool in achieving performance objectives, like optimum yield, from fisheries. Despite recent improvements, we have lacked a fundamental understanding of how design details of HCRs translate into performance. This lack of understanding is especially problematic because of the multiple competing objectives that exist for fisheries: income/profits, conservation, income and food security, and sustainability. Using geometric analysis, this paper provides fundamental architectural principles, demonstrating how to design HCRs to achieve common objectives; illustrating constraints on performance outcomes; and identifying tradeoffs among objectives and how to achieve various performance outcomes through the design of HCRs.

Data Gaps and Alternative Management Approaches: Applying National Standards 1 to Data-Limited Stocks, By Mark Smith, *Jim Berkson*, Patrick Lynch, Erin Schnettler, Melissa Karp, Donald Kobayashi, Clay Porch, Pete Christopher, Cindy Tribuzio, Jason Cope, and Skyler Sagarese

In October of 2016, the National Marine Fisheries Service (NMFS) published final revisions of the National Standard 1 (NS1) guidelines. The guidelines have changed significantly since 1998, and numerous commenters requested additional technical guidance addressing the updated revisions to the guidelines made in 2009 and 2016. This project is being conducted by a subgroup of a national NS1 Technical Guidance Working Group, which is developing recommendations and/or best practices for addressing data gaps and alternative management approaches, particularly related to meeting the requirements of Annual Catch Limits (ACLs) in data-limited situations. In these situations, the revised guidelines state that councils may propose "alternative approaches for satisfying requirements of the MSA other than those set forth in these guidelines." Providing guidance on what these "alternative approaches" might look like, and in what situations they could be applied, is the main task set forth for this subgroup. We will present progress made on our group's objectives to: identify best practices for adhering to the ACL framework with data-limited stocks; identify situations where the current ACL framework may not be appropriate due to data limitations; and provide examples of alternative methods. This will include a comparison of the current stock assessment tier systems used by each Fishery

Management Council, including a case-study comparison of the methods used to calculate catch recommendations and limits for data-limited stocks.

Developing Spatially Explicit Stock Assessment Models to Explore Spatial Variability in Recruitment of Chilipepper Rockfish (*Sebastes goodie*) and Possible Links to Environment, By *Laura Solinger*, Eric Bjorkstedt, Andre Buchheister, Mark Henderson, and John Field

It is generally understood that variability in recruitment of West Coast rockfishes (*Sebastes* spp.) is driven in part by oceanographic conditions affecting early life history stages, yet spatial variability in both recruitment and oceanographic dynamics can confound efforts to elucidate these links at the scale of an entire stock. To address this challenge, we extend the work of Field and Ralston (2005, CJFAS) to resolve spatial and temporal variability in recruitment of Chilipepper rockfish (*Sebastes goodei*) for an additional 15 years, as a foundation for subsequent comparison to candidate oceanographic drivers of recruitment variability. Chilipepper rockfish range from Baja, California to Canada, with the core of the stock located off central California, where they represent an important fishery resource. Stock Synthesis 3 (SS3) was used to build seven spatially explicit stock assessment models for Chilipepper rockfish that represented the coast of California and the majority of the population. The assessments were constructed in Stock Synthesis (SS3) by applying the framework of the most recent formal assessment (Field et al. 2015) to fishery-dependent and -independent data partitioned by latitude into seven regional blocks associated with specific ports or port complexes along the California coast. Recruitment deviations were extracted from assessment results and subsequently analyzed using Principal Components Analysis (PCA). The leading mode of variability in recruitment, representing 70% of variability among ports, captured coherent variability in recruitment along the California coast. A second mode of variability (12.6%) captured regional variability between northern and southern regions, demarcated by San Francisco. Ongoing work seeks to link these patterns to oceanographic conditions varying at similar spatial and temporal scales, taking into account the possible influence of advection and transport in linking recruitment signals along the coast. Results from this work are intended to inform efforts to incorporate spatial structure and oceanographic influence on recruitment and population dynamics in West Coast rockfishes.

Model Ensembles of Common Thresher Sharks in the Northeast Pacific Ocean: Comparing "Traditional" Stock Assessment and Model-Averaging Approaches, and Something with a Little Bit of Both, By *Steven L. H. Teo* and Paul Crone

One major advantage of using multi-model ensembles is the integration of information from diverse, plausible models into model results, projections, and management advice. In "traditional" stock assessment approaches, a base case model is decided on and used to quantify results and associated uncertainty. In this more typical assessment approach, model structure uncertainty is evaluated through informal sensitivity runs, but often ignored during the management process itself. In a multi-model ensemble, such model structure uncertainty is a formal part of the broader umbrella of uncertainty at the ensemble level. Information from the ensemble is then combined through model averaging (e.g., unweighted, or weighted using AIC, BIC, parameter priors, cross-validation, expert opinion, democracy, etc.). Thus, model averaging reduces the decision maker's information load by highlighting the central tendency. However, this may prevent information on high-impact, low-probability events from reaching decision makers. One potential way to address this issue is to use *post-hoc* classification of ensemble results in qualitatively different groups for purposes of highlighting information from particular models in addition to model averaging. This is similar to using model sensitivity runs in “traditional” stock assessments or to use them to represent different hypotheses that are difficult to capture with explicit probabilities. A recent stock assessment of common thresher sharks in the northeast Pacific Ocean (NEPO) is used as an example here because the uncertainty surrounding biological studies associated with this species resulted in contrasting assumptions of population productivity, ranging from moderately productive ($Age_{50\%MAT}$: 5 y, Fecundity: 4 pups y-1, M: 0.179 y-1) to highly unproductive ($Age_{50\%MAT}$: ~13 y, Fecundity: ≤2 pups y-1, M: 0.04 y-1). A model averaging approach would have resulted in overall results being weighted toward configurations with moderate productivity. However, a recently conducted review panel instead recommended a “traditional” approach and a base case model that reflected a population characterized by very low productivity. In addition, we also use *post-hoc* clustering to separate models into groups to highlight qualitatively different outcomes in the ensemble.

Evaluating Alternative Management Strategies for North Pacific Albacore, By *Desiree Tommasi* and Steven L. H. Teo

North Pacific albacore tuna (*Thunnus alalunga*) is a highly migratory species whose range spans the entire North Pacific. Spawning occurs in the tropical and sub-tropical waters of the western and central Pacific Ocean (WCPO). Juvenile fish then undertake trans-Pacific migrations, with some moving from the WCPO to the eastern Pacific Ocean (EPO) to feed in the productive coastal waters of the California Current. This juvenile fish migration sustains lucrative U.S. commercial and recreational surface fisheries. In contrast, the U.S. longline fleet operating out of Hawaii largely catches adults. U.S. vessels account for ~17% of North Pacific albacore catch, while most (62%) is caught by Japan. The highly migratory nature of this stock and the large number of nations involved in its fisheries necessitates international cooperation via two regional fishery management organizations (RFMOs), the Western and Central Pacific Fisheries Commission (WCPFC) and the Inter American Tropical Tuna Commission (IATTC), to ensure effective management. The United States is a member of the International Scientific Committee for tuna and tuna-like species in the north Pacific (ISC), which is the primary science provider for this stock. The WCPFC and IATTC recently discussed the need to improve the current management framework for albacore by identifying formal harvest rules with appropriate limit and target reference points. Here we outline the management strategy evaluation (MSE) process and framework developed at the ISC in collaboration with international stakeholders and both RFMOs, which is used to examine the performance of alternative harvest strategies and reference points for North Pacific albacore given uncertainty. Potential uncertainties in mortality, growth, recruitment, and movement (via time varying age-selectivity) were considered. Results presented will provide an overview of the effectiveness of each harvest strategy in achieving specific management objectives, and will highlight trade-offs among the objectives. One benefit of the MSE process is the exploration of various model structures during the development of the operational models. If these operational models turn out to be well-conditioned, a subset of these operational models may become the foundation for “ensemblelizing” the current assessment models.

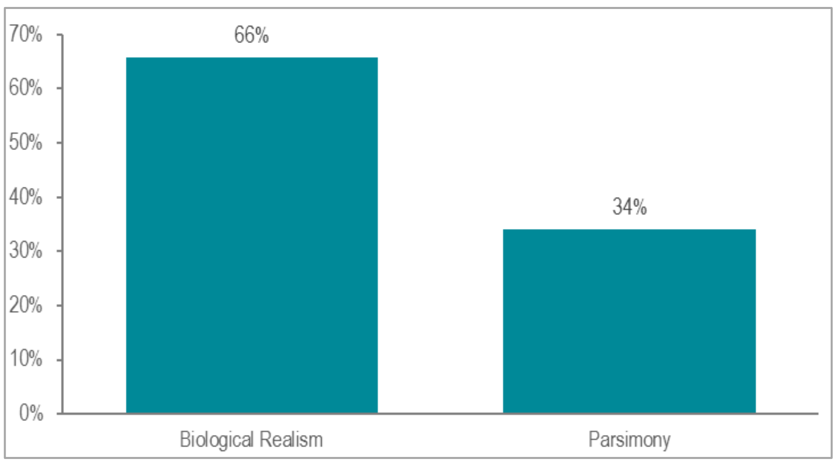
APPENDIX F: POLLING RESULTS

During the plenary discussions for each theme session the audience was polled before the presentation of each breakout group's discussion summary. The polling questions were related to the trigger questions given to the breakout groups. Polling the audience was intended to spur group discussion and give participants an opportunity to contribute opinions on each trigger question, even if they were not part of that particular breakout group. The results from these polls is found below, separated out by theme and breakout session.

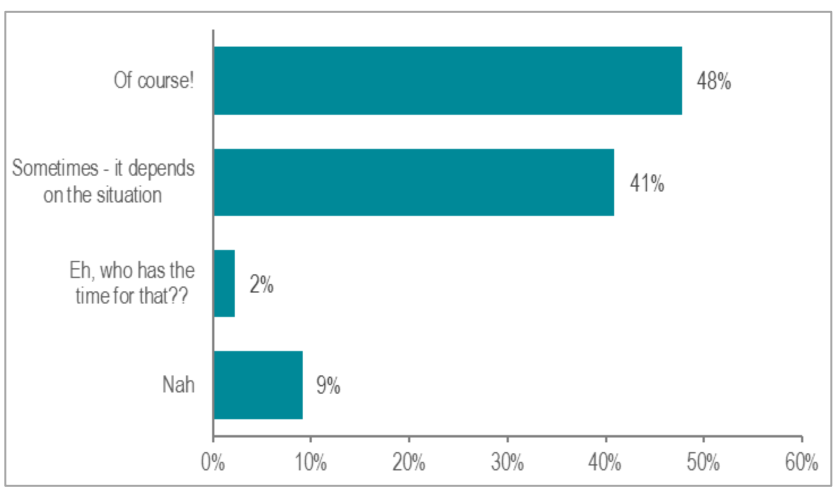
Theme 1: Model Complexity and Stability

1st Breakout Session Trigger Questions

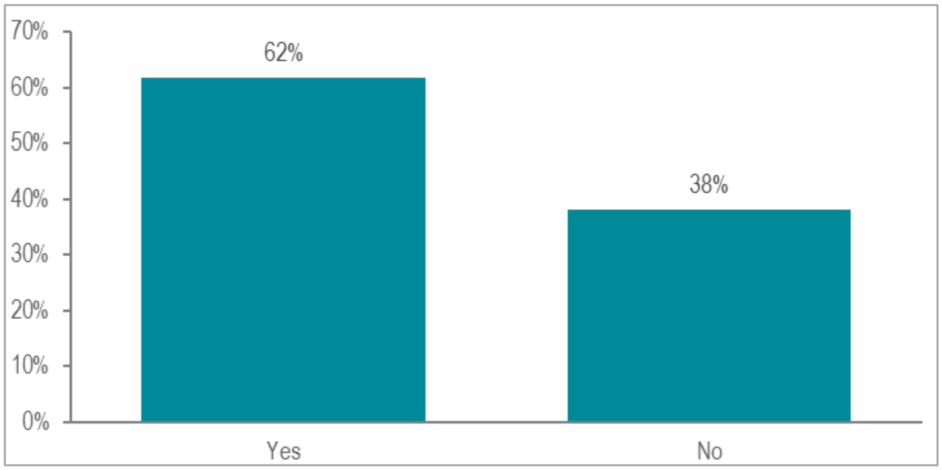
Question 1: What is more important in stock assessment models: biological realism or parsimony?



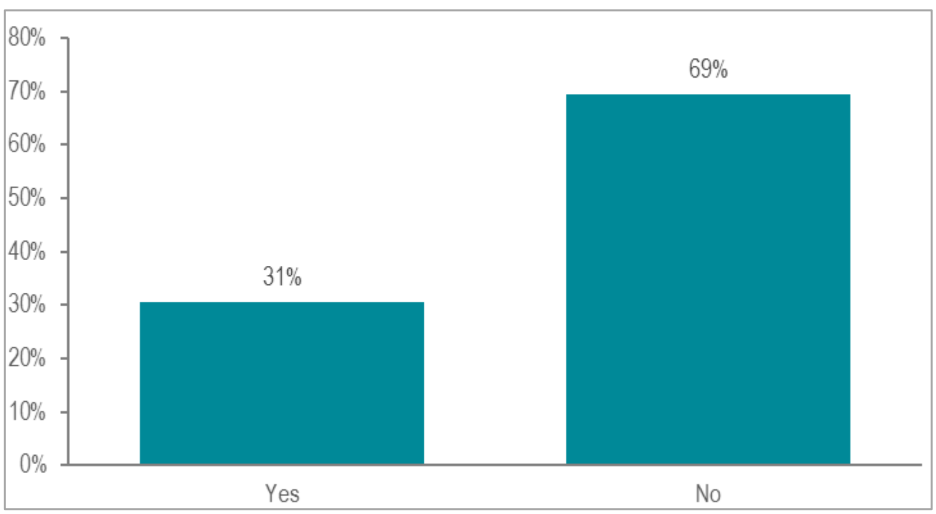
Question 2: Is it appropriate to explore a range of model complexities?



Question 3: Should MSE and/or simulation studies be used to select model complexity level?

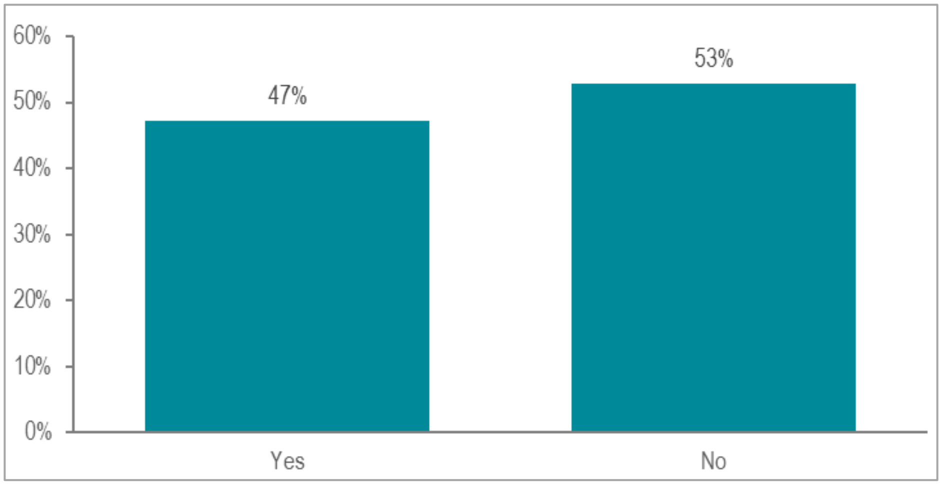


Question 4: Are complex models inherently less stable?

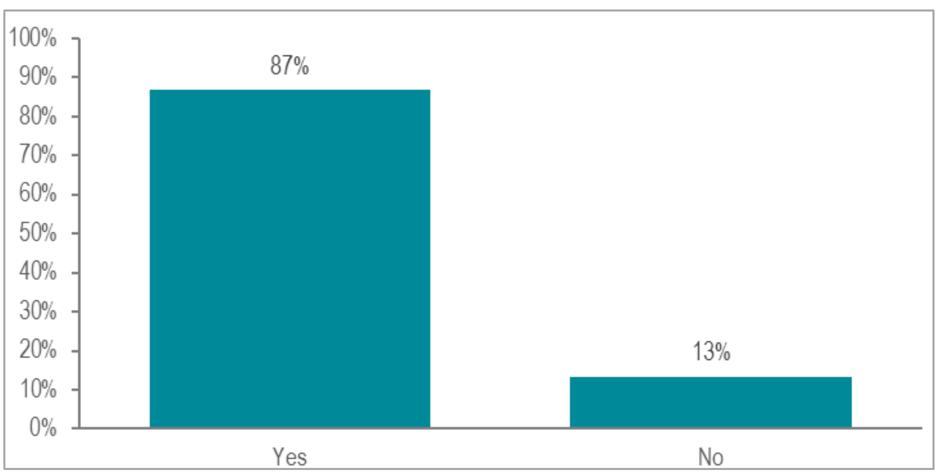


2nd Breakout Session Trigger Questions

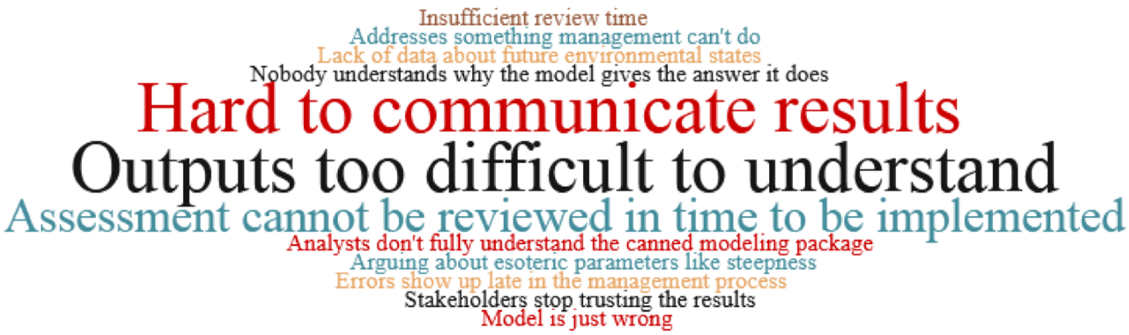
Question 1: Do the benefits of using increasingly complex models outweigh the potential shortcomings and time/resource costs?



Question 2: Can complex models account for hidden constants (e.g., natural mortality/growth/gear selectivity) in simple models?



Question 3: Briefly describe situations when model complexity could interfere with management practicality?

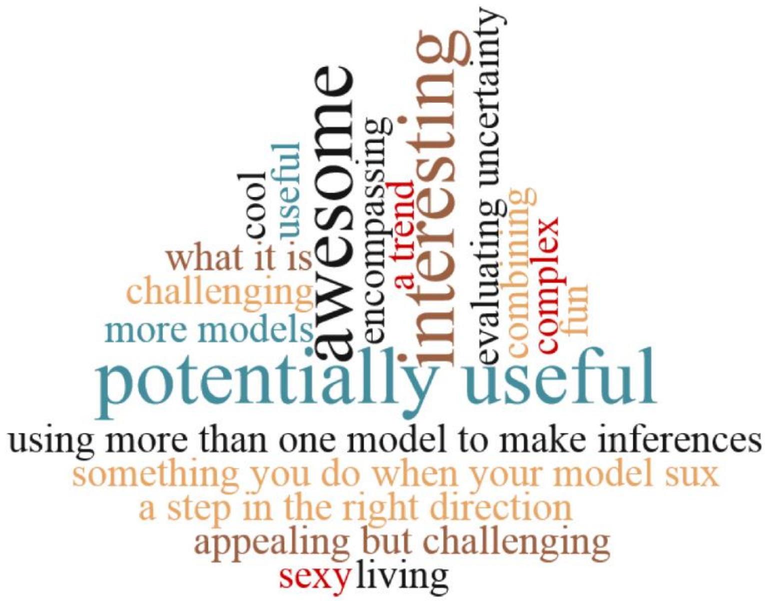


Question 4: What hurdles exist in your region to using the optimum model complexity?

Theme Session 2: Ensemble Modeling

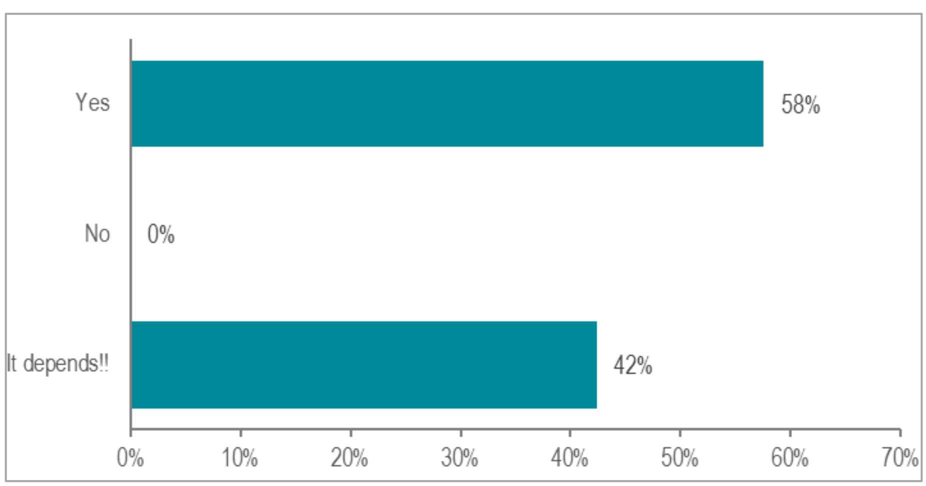
1st Breakout Session Trigger Questions

Question 1: Ensemble modeling is…?



Question 2: What are the benefits to an ensemble modeling approach?

Question 3: Is ensemble modeling appropriate for fisheries management?



Question 4: (a) What is a challenge or barrier to using ensemble modeling to provide management advice?

(b) What is needed to overcome the challenges and barriers to operationalizing ensemble modeling to provide fishery management advice?