NRC Research Press

ARTICLE

Improving stock assessments through data prioritization

Kate I. Siegfried, Erik H. Williams, Kyle W. Shertzer, and Lewis G. Coggins

Abstract: The need for "better data" is a common response of stakeholders and managers when confronted with the uncertainty of advice resulting from quantitative stock assessments. Most contemporary stock assessments are based on an integrated analysis of multiple data types, each with their associated cost to collect. Data collection resources are inevitably limited; therefore, it is important to quantify the relative value of increased sampling for alternative data types in terms of improving stock assessments. We approached this universal problem using a simulation study of a hypothetical, amalgam species developed from eight separate stock assessments conducted for species found in southeastern US Atlantic waters. We simulated a population and a stock assessment from the amalgam species and then individually improved alternative data types (indices, age compositions, landings, and discards) by increasing either precision or sample size. We also simulated the effects of increased sampling for alternative groupings of data that might be collected in concert (e.g., commercial, recreational, or survey). Our results show that for the snapper–grouper complex we modeled, age composition data have the largest effect on the accuracy of assessments, with commercial age compositions being the most influential. This is due in part to the relative paucity of age composition data for many southeast US marine stocks, so that modest increases in collection efforts have relatively high benefits for age-based assessment models currently in use for the region. Though this study used data from a particular region of the US, our investigative framework is broadly applicable for quantitatively evaluating the benefits of improved data collection in terms of the precision of stock assessments in any region.

Résumé : La nécessité d'obtenir de « meilleures données » est une réponse fréquente des parties prenantes et gestionnaires aux prises avec le caractère incertain des avis découlant d'évaluations quantitatives de stock. La plupart des évaluations de stock actuelles reposent sur l'analyse intégrée de différents types de données, chacun ayant un coût de collecte associé. Les ressources pour la collecte de données sont inévitablement limitées, d'où l'importance de quantifier la valeur relative, pour ce qui est d'améliorer les évaluations de stock, d'élargir l'échantillonnage pour obtenir différents types de données. Nous avons abordé ce problème universel en utilisant une étude de simulation d'une espèce amalgamée hypothétique définie à partir de huit évaluations de stock distinctes réalisées dans les eaux atlantiques du sud-est des États-Unis. Nous avons simulé une population et une évaluation de stock pour l'espèce amalgamée, puis avons amélioré différents types de données (indices, compositions par âge, débarquements et rejets) séparément, en accroissant la précision ou la taille de l'échantillon. Nous avons également simulé les effets d'un échantillonnage élargi pour différents groupes de données qui pourraient être recueillies ensemble (p. ex. commerciales, les pêches sportives ou enquêtes). Nos résultats montrent que, pour le complexe des vivaneaux–mérous que nous avons modélisé, les données de composition par âge ont le plus grand effet sur l'exactitude des évaluations, les compositions par âge tirées de la pêche commerciale ayant la plus grande influence. Cela est en partie le fait de la rareté relative de données de composition par âge pour de nombreux stocks marins du sud-est des États-Unis, qui fait que de petites augmentations des efforts de collecte produisent des avantages relativement importants pour les modèles d'évaluation reposant sur l'âge utilisés actuellement pour la région. Bien que l'étude utilise des données pour une région précise des États-Unis, notre cadre d'étude peut s'appliquer à l'évaluation quantitative des avantages d'une collecte de données améliorée pour ce qui est d'accroître la précision des évaluations de stock dans n'importe quelle région. [Traduit par la Rédaction]

Introduction

Fisheries management often depends on stock assessments to inform decisions about potential catch limits relative to the dynamics of the stock. Uncertainty, or a lack of data, may necessitate a more precautionary approach to management, which often leads to lower available yields. Thus, more certainty about the data included in a stock assessment is highly desirable. When the data are available, stock assessments are an integrated analysis of those multiple data sources, including landings and discards, life history information, size and age compositions, and indices of abundance, each of which has an associated level of uncertainty (Maunder and Punt 2013). For example, fishery catches and discards may be imprecise because of low observer coverage or sporadic port sampling. Life history data used to inform natural mortality, growth, and reproductive output may be sparse or borrowed from other species (Punt et al. 2011). Age composition data require considerable postcollection processing that varies greatly across laboratories. Fishery-independent indices require extensive field survey efforts and consistent funding to maintain the integrity of the relative abundance time series. Research recommendations compiled during assessments inevitably contain requests for more of each of these data types, yet few efforts have focused on how best to prioritize data collection efforts. Even in the face of an increasing demand for assessments, only minimal improvements to data collection may be economically feasible. Thus, there is a pressing need to develop approaches for prioritizing data in terms of its utility for stock assessments (Magnusson and Hilborn 2007).

Received 21 August 2015. Accepted 18 January 2016.

K.I. Siegfried, E.H. Williams, and K.W. Shertzer. NOAA–National Marine Fisheries Service, 101 Pivers Island Rd., Beaufort, NC 28516, USA.
L.G. Coggins. US Fish and Wildlife Service, P.O. Box 346, Bethel, AK 99559, USA.
Corresponding author: Kate I. Siegfried (email: Kate.Siegfried@noaa.gov).

Table 1. The compiled parameters for the amalgam species.

	VS	RP	BSB	GAJ	RS	RG	GAG	SG	AVG
Start year	1946	1972	1978	1946	1955	1976	1962	1974	1963
L_∞	506	528	496	1194	902	848	1091	1065	792
k	0.12	0.20	0.18	0.34	0.24	0.21	0.17	0.09	0.18
t_0	–3.50	–1.53	–0.92	–0.45	–0.03	–0.67	–1.31	–2.88	–1.36
Steepness	0.71	0.41	0.48	0.74	0.85	0.92	0.95	0.84	0.74
MSST	0.22	0.225	0.38	0.23	0.08	0.14	0.115	0.12	0.19
A_{mat}	1	1	1	1	2	3	3	6	2.25
Max. age	12	14	11	10	20	16	20	25	16
Years of commercial compositions	19	13	12	7	8	5	15	14	12
Years of recreational compositions	27	8	11	1	20	11	12	3	12
Years of survey compositions	10	22	23	0	0	11	0	5	9

Note: These parameters represent species in the Atlantic waters of the southeast US: vermilion snapper (VS), red porgy (RP), black sea bass (BSB), red snapper (RS), greater amberjack (GAJ), red grouper (RG), snowy grouper (SG), and gag (GAG). AVG represents the value used in the simulated population. In most cases, we used an arithmetic mean. However, we excluded GAJ growth parameters, as they seemed to deviate from the snapper–grouper complex. MSST is the minimum stock size threshold, which is used to determine whether a stock is considered overfished. A_{mat} is age at 50% maturity.

When faced with the question of which data type to collect more of, no clear priority exists. Scientists want the assessment to reflect the population and characterize uncertainty as well as possible, but there have only been a few studies that address the effects of various data sources on assessment results (Powers and Restrepo 1993; Yin and Sampson 2004; Szuwalski and Punt 2012).

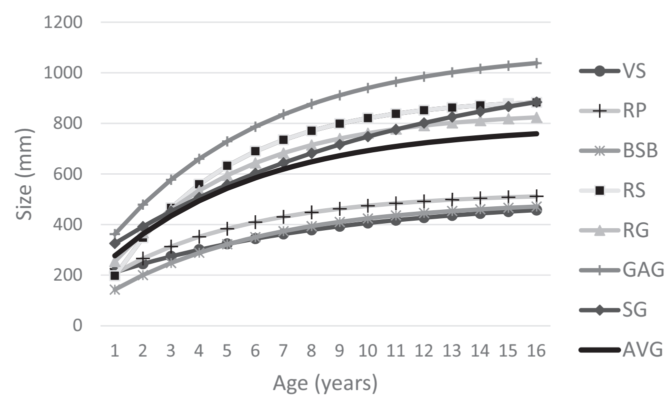
Data collection efforts, as well as the perceived needs for additional data types, vary considerably across geographic regions. In regions where catches are hardly characterized (e.g., the Caribbean), data collection priorities are clearer. However, regions that have conducted multiple age-structured stock assessments have various data deficits that need to be filled. As an example, fishery scientists working on species in the Atlantic waters of the southeastern US repeatedly encounter a lack of fishery-independent survey data, discards are difficult to characterize, and annual age composition data may be insufficient. In addition, a high proportion of the landings and discards are taken by the recreational fleet, which is inherently difficult to monitor. Nevertheless, age-structured assessments are regularly used to inform fishery management decisions. The goal of this work is to develop a framework for providing objective guidance about prioritizing data sources for fish stock assessments. Using simulation, we evaluate which data types are most effective at improving assessment accuracy by modeling the effect of improving each data source. Although our study design is applied to data from a single region, the approach is general and could be applied in other regions.

Methods

We simulated a hypothetical population (i.e., an amalgam species) based on information available for actual stocks from the snapper–grouper complex that have been individually assessed in the Atlantic waters off the southeastern US. We reviewed the assessed stocks in the region and averaged the results of eight stocks: vermilion snapper (*Rhomboplites aurorubens*), red porgy (*Pagrus pagrus*), black sea bass (*Centropristis striata*), red snapper (*Lutjanus campechanus*), greater amberjack (*Seriola dumerili*), red grouper (*Epinephelus morio*), snowy grouper (*Hyporthodus niveatus*), and gag (*Mycteroperca microlepis*) (SEDAR 2008, 2009, 2010, 2012*a*, 2012*b*, 2013*a*, 2013*b*, 2014). The combined results from the eight assessments provide an amalgam stock that represents the snapper–grouper complex from 1963 to 2012 (Table 1).

We modeled ages 1 to 16 years using annual time steps, with a mean maximum age of 16. We used an averaged von Bertalanffy (von Bertalanffy 1938) growth curve (L_∞ = 792 mm, k = 0.18, t_0 = –1.36; Fig. 1) and a mean age of 50% maturity of 2.25 years. For the growth curve, we excluded greater amberjack, as its growth pattern seemed to diverge from the other species. However, its other life history parameters were similar to the other species in the complex. We modeled two fleets (recreational and commercial)

Fig. 1. The von Bertalanffy growth curve for the amalgam species plotted with the curves for each of the species included in the amalgam. Refer to Table 1 for species abbreviations.

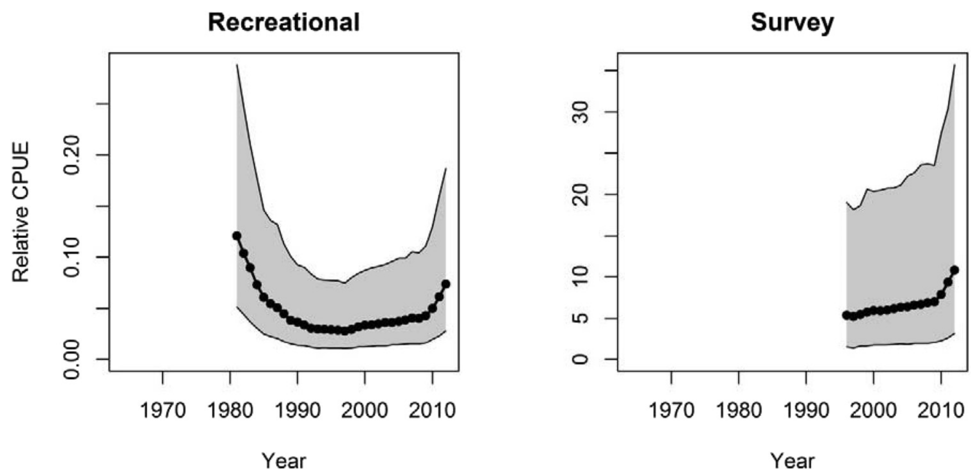


and two relative abundance indices (one fishery-dependent and one fishery-independent), each based on the averaged trends of the eight stocks.

According to the amalgam species data, the recreational index data were typically available beginning in 1981 and ran continuously through the terminal year. The fishery-independent index began in 1996 and also ran through the terminal year of the model (Fig. 2). The commercial fishery landings were assumed to be known very well, with a coefficient of variation (CV) of 0.05. The CV for the recreational fishery was related to the percent standard error provided by a Marine Recreational Information Program catch data query (personal communication from the National Marine Fisheries Service, Fisheries Statistics Division, November 2014). Discards for each fishery were available annually since 1992. Age composition data were compiled from the survey (9 years), commercial fleet (12 years), and the recreational fleet (12 years). We did not simulate length composition data, as it has been shown to provide relatively little information to estimate population dynamics for snapper–grouper species in the region (SEDAR 2012*b*).

We use a catch-age model called the Beaufort Assessment Model (BAM) that is commonly used for assessments in the southeastern US as the operating model for the simulations (Williams and Shertzer 2015). Briefly, BAM is a statistical catch-age model implemented with the AD Model Builder software (Fournier et al. 2012) and offers the flexibility of a customizable source code while providing the basic constructs and functions of most other age-based assessment models (e.g., Beverton–Holt stock–recruitment relationship, a variety of age-based selectivity options, weighting options for the likelihood components, penalties to constrain

Fig. 2. Recreational CPUE and survey index used in the base model run. The points represent the model inputs, and the grey area represents the 5th and 95th quantiles of the model fits to the data.



parameter spaces, etc.). We only altered the operating model between runs to change either the precision or the sample size of each data type (Table 2).

We used the mean of the parameters from each of the stock assessments (see Table 2) to create a set of input parameters to simulate a "true" population with the operating model. Each true population data set was then fit with the same operating model using optimization.

Our algorithm for the simulation was as follows:

Step 1: Configure the operating model with fixed parameters to create a simulated data set using BAM by inputting the mean of each of the following parameters from our amalgam stock: start year, maximum age, years of recruitment deviations, catch per unit effort time series (CPUEs), landings time series, von Bertalanffy growth parameters, steepness, selectivities, catchabilities, size at age, maturity at age, and fishing mortality trends (Figs. 3–4; also see the online Supplementary Material[1] for the data input file). We used 2012 as the terminal year, which was the last assessment year in our amalgam stock. The mean selectivity for each sector was used for the amalgam species: commercial, recreational, survey, and discards (Fig. 5). We generated age compositions with a multinomial distribution using the mean sample size and the number of years that composition data were available for our amalgam species. We then allow for random recruitment assuming lognormal deviations (Fogarty et al. 1991).

Step 2: Add observation error to mimic actual assessments. Observation error was added to the recreational CPUE and survey indices based on the mean CV of these two data types across the eight prior assessments. Error was also added to the composition data by increasing or decreasing the sample sizes used in the multinomial distribution.

Step 3: Create the operating data set using all stochastic components from step 2 and then use BAM to estimate model parameters of the stochastic, simulated data set.

Step 4: Collect the results from the estimation in step 3 and repeat for the desired number (N = 2000) of bootstrap iterations.

We tracked model estimates of virgin recruitment, steepness, acceptable biological catch (ABC), and status indicators (spawning stock biomass sufficient to produce maximum sustainable yield (SSB_{MSY}), fishing mortality required to harvest MSY (F_{MSY}), and minimum stock size threshold (MSST)). ABC was calculated for a single year following the terminal year of the assessment using terminal biomass and estimates of fishing mortality from each fishery. MSST is used as an indicator by the South Atlantic Fishery Management Council of an overfished stock and was calculated as $0.5 \times SSB_{MSY}$.

We focused on data types considered to be the best candidates for improved collection: landings, discards, age compositions, and fishery-independent data (i.e., CPUE and age composition). The relative cost of each type of data is known, and programs are currently in place that could improve their quality. For example, a larger budget for the fishery-independent survey would likely translate into more sea days and more samples taken, which would translate directly into the type of precision and sample size improvements modeled here. After initial simulations improved each individual data type separately (i.e., commercial discards, commercial age compositions, recreational discards, recreational landings, recreational age compositions, recreational CPUE, survey precision, survey age compositions), we conducted additional simulations where data types were grouped as "commercial", "recreational", and "survey". In each of the groupings, the CVs of landings, discards, and indices were decreased by fourfold and the sample sizes of age compositions were increased by fivefold. The values fourfold and fivefold were somewhat arbitrary but were chosen based on the extremes of what is considered practical for data collection efforts in the southeastern US. For example, because the amount of age data available in prior assessments was relatively small, we deemed it feasible to collect five times more otoliths above current efforts. But given current efforts to monitor landings and discards in our region, improving the precision of these data types by more than fourfold was considered impractical. The separate date types were grouped as follows: recreational landings and all discards, all age compositions, all survey data, all commercial data (except landings), all commercial data, and all indices (see Table 2 for the particulars of each scenario).

For each scenario, we generated N = 2000 bootstrap replicates, which was sufficient for the means of all estimated parameters to converge to stable values. However, one parameter, steepness, proved difficult to estimate in ~7%–12% of replicates and was estimated at its upper bound of 0.99, a situation common in stock assessment (Conn et al. 2010). These runs were culled before summarizing the results, because in practice steepness hitting an upper bound would not be considered a valid run for assessment purposes. However, we report the proportion of runs culled for

[1]Supplementary data are available with the article through the journal Web site at http://nrcresearchpress.com/doi/suppl/10.1139/cjfas-2015-0398.

Table 2. Scenarios considered for each of the data improvements.

	Survey CPUE CV	Survey composition sample size	Recreational landings CV	Recreational discards CV	Recreational CPUE CV	Recreational composition sample size	Commercial landings CV	Commercial discards CV	Commercial composition sample size
Base	0.75	25	0.75	1	0.5	10	0.05	1	30
Commercial composition sample size	0.75	25	0.75	1	0.5	10	0.05	1	**150**
Commercial discards CV	0.75	25	0.75	1	0.5	10	0.05	**0.25**	30
Recreational composition sample size	0.75	25	0.75	1	0.5	**50**	0.05	1	30
Recreational CPUE CV	0.75	25	0.75	1	**0.125**	10	0.05	1	30
Recreational discards CV	0.75	25	0.75	**0.25**	0.5	10	0.05	1	30
Recreational landings CV	0.75	25	**0.1875**	1	0.5	10	0.05	1	30
Survey CPUE CV	**0.1875**	25	0.75	1	0.5	10	0.05	1	30
Survey composition sample size	0.75	**125**	0.75	1	0.5	10	0.05	1	30
Survey CPUE CV + composition sample size	**0.1875**	**125**	0.75	1	0.5	10	0.05	1	30
Commercial discards CV + composition sample size	0.75	25	0.75	1	0.5	10	0.05	**0.25**	**150**
All recreational components	0.75	25	**0.1875**	**0.25**	**0.125**	**50**	0.05	1	30
All composition sample sizes	0.75	**125**	0.75	1	0.5	**50**	0.05	1	**150**
All indices CVs	**0.1875**	25	**0.1875**	1	**0.125**	10	0.05	1	30
Recreational landings CV + all discards CVs	0.75	25	**0.1875**	**0.25**	0.5	10	0.05	**0.25**	30

Note: The rows are each model run, either the base model or a run named after the data source that was improved. The columns are either CVs (in the case of indices, landings, or discards) or sample size (in the case of composition data). The base case contains no data improvements, and each data improvement across columns is in bold.

each scenario (Table 3), as we consider it one indicator regarding the value of improved data.

Steps 2 through 4 were repeated as each data type was improved either by decreasing the CV or increasing the age composition sample sizes. Parameter estimates and management quantities of each assessment fitted to improved data were compared with those of the initial base model estimates using the mean relative error across all bootstrap replicates, $\mathrm{RE}_i = (\hat{a}_i/a_i) - 1$, where $\hat{a}_i$ is the estimated value of parameter i, and a_i is the true value of parameter i.

Results

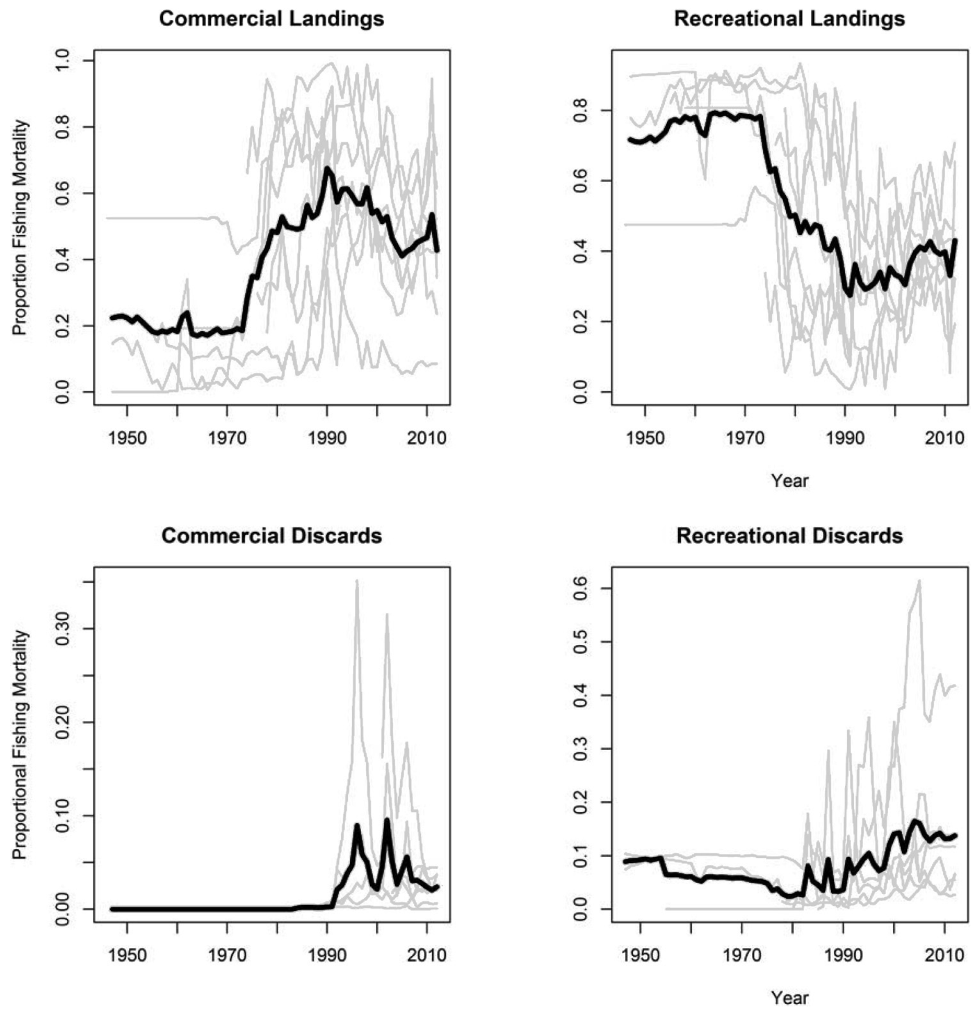
The base model run (i.e., without data improvements; see row 1 of Table 3) overestimated the biomass benchmark (RE = 0.38) and underestimated the fishing status benchmark (RE = −0.67). The estimates of the reference points show a positive bias for the base case. The bias is the result of the variation put into the data and the sample sizes used. If the variation is reduced and sample size increased prior to the data improvements, the model fits the inputs exactly. Steepness was well-estimated in the base case, as was the ABC (RE = −0.01 and 0.05, respectively). Time series of predicted landings and discards by fishery replicated the general patterns observed across the eight assessed species (Fig. 6). Simulations with improvements to the individual data types reduced the RE of model estimates by 10%–49% (Table 3). Improvements (i.e., increased effective sample sizes) in both the recreational and the commercial age compositions had the greatest effect on model accuracy (23%–68%). For example, improved commercial age compositions reduced the estimation error of the biomass benchmarks and R_0 by a considerable amount (12.5%–47%), while improved recreational age compositions had the most effect on the estimates of R_0, MSY, and ABC (12.5%–60%).

Overall, the grouped components reduced the RE of model estimates by 22%–70%. Specifically, the grouped composition data improvements had the largest effect on model accuracy overall, which is consistent with the results from the improvement of individual components. The effects were approximately double that of the individual component runs. However, the survey components grouping showed an effect that was nearly absent in the runs of their individual components. The improvement of all survey components had a large effect on both the biomass and fishing benchmark estimates (RE decreased by 37% and 70%, respectively), surpassing the effect of the all-composition data grouping for those two reference points. In general, probability density functions of key management quantities derived from the assessment illustrate how improved data collection can influence fishery management advice (Fig. 7; also see Supplementary Material[1]). For example, when all age composition data were improved, the RE in the fishing benchmark (F/F_{MSY}) and the biomass benchmark ($\mathrm{SSB}/\mathrm{SSB}_{\mathrm{MSY}}$) decreased by 58% and 37%, respectively, while the RE in the estimated ABC was near zero. For the fishing benchmark in particular, improvements in the age composition data altered the qualitative determination of stock status from overfishing to not overfishing (Fig. 7). For the amalgam species considered here, the estimated ABC was about 10% larger for models simulating improvements in the age composition data. Further improvements in the age composition data decreased the presence of outliers and resulted in a narrower distribution about the median (i.e., greater precision). The probability density functions for all scenarios can be found in the online Supplementary Material[1].

Discussion

Our study provides a framework to generate specific and objective guidance to fishery scientists and managers who must decide how to allocate limited monitoring and research resources towards improving stock assessments. Improving accuracy would provide several benefits, including better characterizing the stock

Fig. 3. Proportion of total fishing mortality attributed to commercial landings, recreational landings, commercial discards, and recreational discards used in the base model. Each of the eight stocks is plotted in grey, with the mean trend over all the stocks in black.



dynamics and reducing the uncertainty in management regulations. In US fisheries management, reduced uncertainty in the assessment generally results in larger catch limits, providing obvious manager and stakeholder benefits.

Our results are interpreted within the framework of the way assessments are conducted and the current data availability in the southeastern US. For example, the data could have been improved further (e.g., index CVs down to 5% or samples sizes in the 1000s per year as is the case in other regions); however, it would be an impractical result. It would not be possible to expand the current otolith sampling, survey, or observer programs within a reasonable time frame to achieve those improvements. Therefore, we used our judgement, based on the evidence provided in over a decade of snapper–grouper assessments, to improve the data by realistic increments for the region. The feasibility of data improvement will differ across regions depending on the current data collection programs.

Assessments conducted in the southeastern US are dominated by fishery-dependent data. Commercial landings data are treated as known with high precision because there are often both trip ticket and logbook programs in place. There is also a need to anchor the model at the reported landings values, though the historical reconstruction of landings time series are inherently less certain. Commercial discards are much less certain and often estimated using logbook or observer data where available. General recreational landings are assumed less precise and are estimated using an angler intercept survey. The angler intercepts are variable and bring substantial uncertainty to the model. Recreational effort is estimated using a phone survey (Andrews et al. 2014), which has historically suffered from a low response rate. This has motivated the Marine Recreational Information Program to switch to a mail survey to decrease uncertainty around the landings and effort estimates. However, our study showed that the most influential portion of recreational data for assessment purposes in the region is the composition data. The improvement of the recreational discards also had an effect on the fishing benchmark estimate. Both recreational and commercial discards are estimated with very large CVs and are rarely verified with observer data. For our simulation, we used the same selectivity for the recreational and commercial fisheries discards because of a lack of information to differentiate the two. Owing to the fact that much of the recreational fishing for most of these species is done using similar gear and in similar depths, there are few possibilities for recreational fisherman to discard different age classes. However, we did not model only undersized fish as being discarded. The dome-shaped selectivity we used takes into account both the closures and undersized fish. The difficulty specifying discard selectivity may have a very small impact on the assessment, as our study suggests that improving the commercial discard data had almost no effect on the model accuracy.

Fish stocks in the Atlantic waters of the southeastern US have not been adequately sampled by a fishery-independent survey in

Fig. 4. Total fishing mortality applied to the stock in the base model (shown in black). Each of the eight stocks used to inform the mean trend are shown in grey for reference.

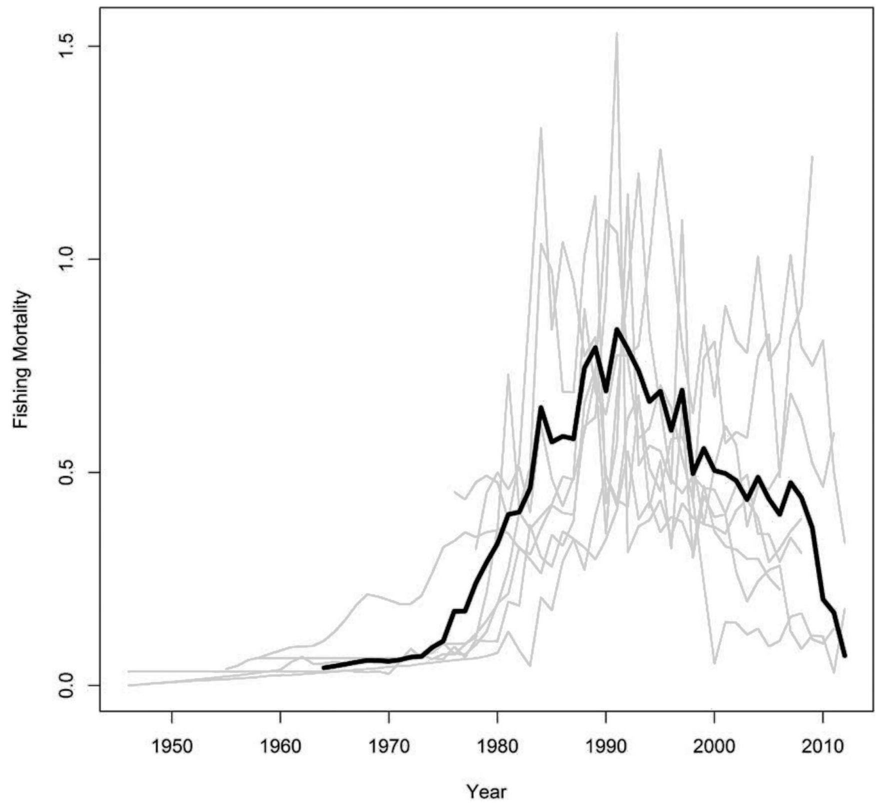


Fig. 5. Proportional selectivity by age for each fleet and survey. The commercial, recreational, and survey selectivities are each means of the various assessment model fits from the eight species. The discard selectivity is for both the commercial and recreational fleet.

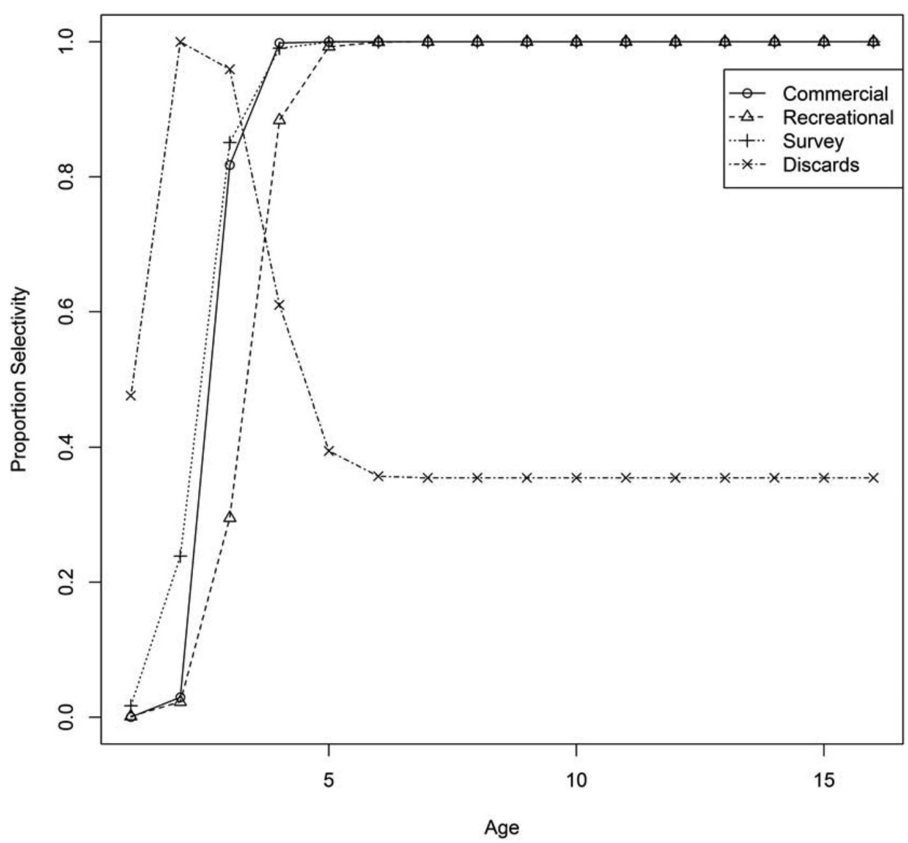


Table 3. The mean relative errors of the base assessment and each assessment of an improved data set.

	Steepness	Culled	R_0	ABC	MSY	F_{MSY}	F/F_{MSY}	SSB_{MSY}	SSB/SSB_{MSY}
Base	−0.01	0.116	−0.32	0.05	−0.22	−0.15	−0.67	−0.35	0.38
All commercial components	−0.01	0.095	−0.28	−0.01	−0.21	−0.11	−0.37	−0.31	0.29
All recreational components	−0.01	0.102	−0.29	**0.01**	−0.22	−0.11	−0.48	−0.32	0.35
Landings and discards	**0**	0.113	−0.32	0.05	−0.23	−0.12	−0.72	−0.36	0.39
All survey components	−0.01	0.075	−0.28	−0.06	−0.22	**−0.09**	**−0.20**	−0.31	**0.24**
All indices	−0.02	0.114	−0.26	0.05	−0.19	−0.17	−0.58	−0.27	0.36
All compositions	−0.01	0.088	**−0.22**	−0.03	**−0.17**	−0.10	−0.28	**−0.23**	**0.24**
Recreational age compositions	**0**	0.111	**−0.28**	**0.02**	**−0.20**	−0.13	−0.53	−0.32	0.35
Survey age compositions	−0.01	0.107	−0.31	−0.02	−0.23	**−0.11**	−0.41	−0.34	0.33
Commercial age compositions	−0.01	0.100	**−0.28**	−0.02	−0.21	−0.12	−0.36	**−0.30**	**0.29**
Recreational index	−0.02	0.125	−0.29	0.05	−0.21	−0.16	−0.58	−0.31	0.38
Survey index	**0**	0.110	−0.31	0.06	−0.22	−0.14	−0.65	−0.35	0.40
Recreational landings	−0.01	0.120	−0.30	0.06	**−0.2**	−0.15	−0.69	−0.34	0.39
Commercial discards	−0.01	0.123	−0.31	0.05	−0.22	−0.14	−0.65	−0.35	0.38
Recreational discards	−0.01	0.114	−0.30	0.05	−0.22	−0.13	**−0.34**	−0.69	0.38

Note: The values and category names in bold are the components or categories that contributed most to the accuracy of the assessment. The top section (first seven rows, excluding base case) is the grouped data improvements, and the bottom section contains the individual data component improvements. Culled is the proportion of runs culled owing to steepness hitting the upper bound.

Fig. 6. Model fits to the recreational (column one) and commercial (column two) landings (row one) and discards (row two). The points represent the model inputs, and the grey area represents the 5th and 95th quantiles of the model fits to the data.

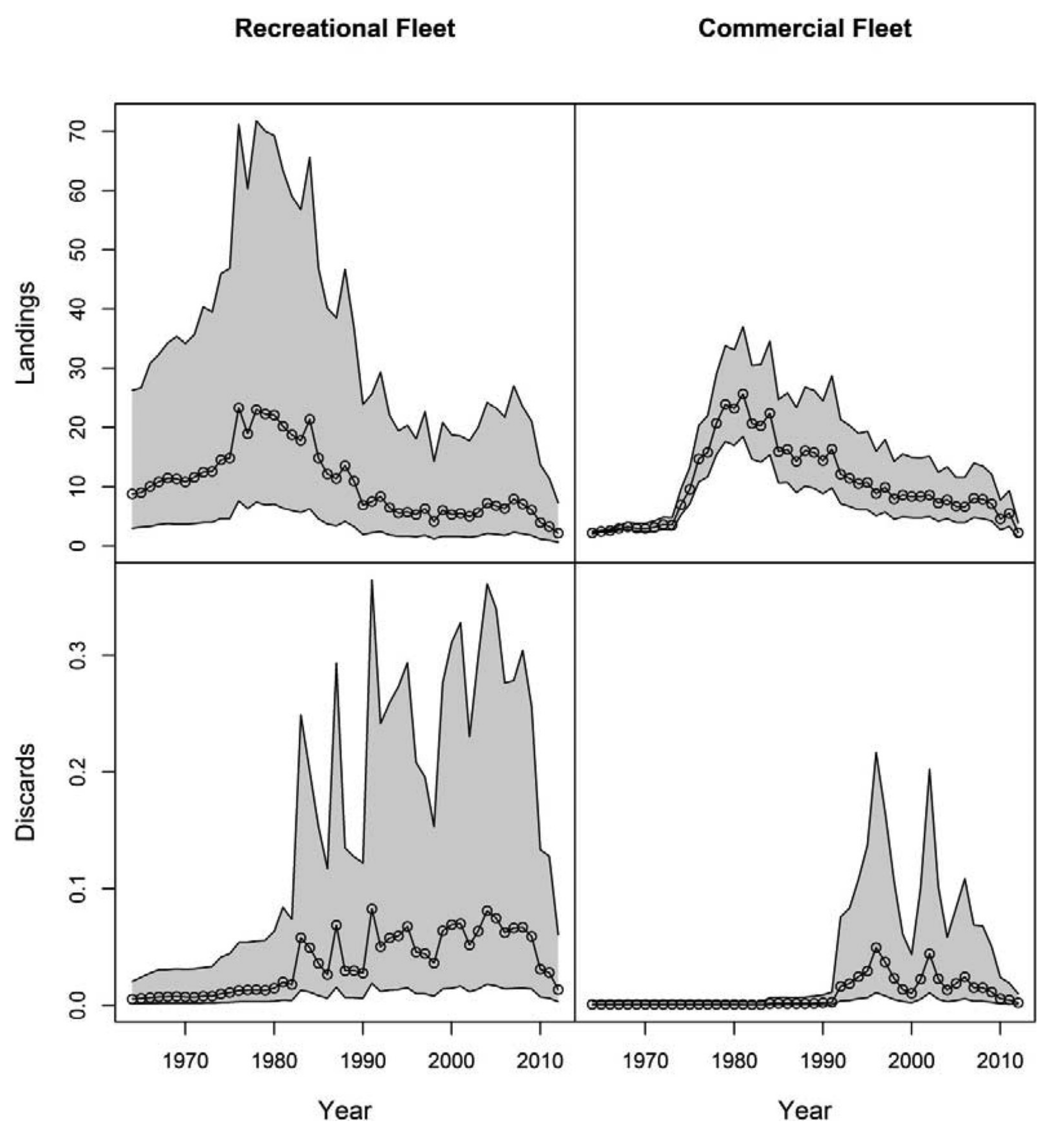
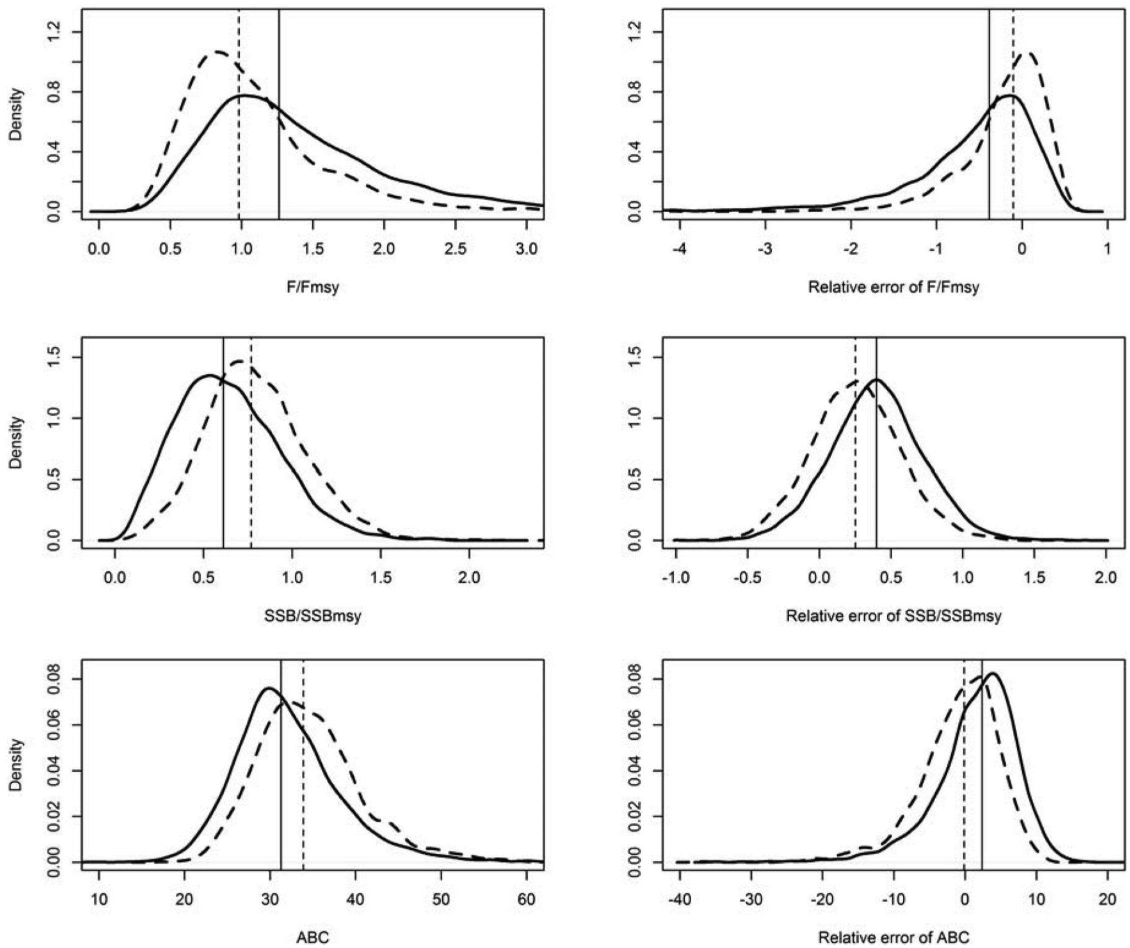


Fig. 7. Estimates of F/F_{MSY}, SSB/SSB_{MSY}, and ABC plotted alongside the relative error of the base data (solid line) and the improved data (dashed line) assessments. The vertical lines are medians of the bootstrap replicates.



the past (e.g., Rindone et al. 2015). When such data are used in assessments, corresponding age compositions are typically available but noisy. In practice, we use length compositions only when age compositions are absent because they are often too noisy to provide informative demographic information to the model. Therefore, we did not attempt to model length compositions in this study. None of the survey components has a noticeable effect individually, but the survey grouping was the second most influential data source in our study. While the composition grouping improved the biomass benchmarks, virgin recruitment, and MSY estimates, the survey grouping improved the fishing benchmark. It should be noted that the component grouping also had an effect on the fishing benchmark, but the relative error improvement was slightly less (58% versus 70%).

A few other studies have investigated the prioritization of data collection or improvements in different ways. Powers and Restrepo (1993) investigated how increasing precision in some of the data inputs changed the outcome of the assessment model used for king mackerel (*Scomberomorus cavalla*) management in the Gulf of Mexico. Their study did not contain a simulation component, so it cannot be known whether the changes in the outcomes were any closer to the truth. If the true population dynamics are not known, results cannot be measured in terms of accuracy, but rather the desirability of the assessment outcome. Szuwalski and Punt (2012) used a simulation framework to determine research priorities in light of multiple model uncertainties. Using eastern Bering Sea snow crab (*Chionoecetes opilio*), they increased the precision on the survey CPUE and the number of samples to develop the growth curve. This improved the estimability of some parameters; however, their results are much more applicable to a length-structured model. The current study is most similar to Yin and Sampson's (2004), with a few key differences. Their study used a fractional factorial design to evaluate how input data errors and the characteristics of the stock (with a focus on Pacific coast groundfish using the Stock Synthesis program; Methot 2000) affected the bias and precision of the model outcomes. We used the BAM model and formed an amalgam species data set to represent species in the Atlantic waters of the southeastern US. We used sample sizes and CVs that are reasonable for our region, which are lower than Yin and Sampson's (2004) composition sample size, and we used a larger range of CVs than theirs for the survey CPUE. Nonetheless, we share one key result; age composition sample size is very influential on the accuracy of assessment outcomes.

For our region, we demonstrated that the individual components as well as the landings and discards grouping had very little impact on model accuracy. The composition grouping had the biggest effect on the accuracy of the assessment, followed by all survey components. Of the individual component runs, the commercial composition data was most influential followed by the recreational compositions. Interpreting the results in light of the cost of improving the data is encouraging. Composition data are relatively inexpensive when an aging lab is already established and operational. All otoliths have a similar cost to process but not necessarily to collect. A recreational otolith may be more expensive to collect because the process by which a port sampler gathers otoliths is typically opportunistic. Consequently, rather than additions to the daily sampling, data improvement may require additional days of sampling. Commercial composition data are easier to gather through standard sampling procedures at the dealers or fish houses. Compared with the cost of increasing a recreational effort sampling program, increasing fishery observer coverage, or increasing sampling days of a fishery-independent survey, composition data require the smallest investment with the biggest return.

Our results come with a variety of caveats. Most importantly, the results are dependent on the combination of data sources used for the simulation. In other words, the history of removals, the biology of the stock and the availability of the different data sources through time will be unique by region and stock and will likely result in different data prioritization results. As an example, our amalgam stock had a very small proportion of the biomass removed through discarding. Other stocks' main source of removals may be discards, and improvements in the accuracy of discard reporting would likely factor in more heavily than the precision of catches or composition sample sizes. Stocks in regions where fishery-independent surveys have covered the stock well may not show a large effect if the survey CVs were reduced. Also, we examined how improvements in the existing data would affect accuracy of the assessment. We did not look at the effect of adding a new data source, such as a survey, or expanding the spatial coverage of the fleets or surveys, though that is a future research goal.

Our input data were compiled using a number of reconstructed data sets (e.g., recreational landings before 1981), but we did not look into whether those methods of reconstruction had an effect on our results. We did this for two reasons: we were interested in the effect of improving data collection now, and the methods for historical reconstruction were too varied (ranging from linear interpolation between no catches and the first year of reported data to applying a catch rate to effort estimates back in time). Changing the uncertainty of those catches back in time would not address our main objective of informing current data collection efforts.

Steepness is difficult to estimate and is often fixed in practice. However, the contrast in our data likely allowed for the parameter to be estimated in a large proportion of our model runs (Conn et al. 2010). In the model runs that remained after we culled the bounded runs (as described in the Methods section), the estimates of steepness were consistently accurate. If our amalgam species demonstrated a different exploitation pattern, steepness may not have been estimable.

We did not attempt to address bias in any of the data sets. We have little information about which direction bias might occur in the actual data. For example, we only have speculative information about whether landings may be over- or underreported or whether discards are over- or underestimated. It is known that an increase in sample size cannot overcome a known bias (Coggins and Quinn 1998). However, without better evidence of the direction of a potential bias, we decided to focus our study on precision.

Finally, we did not initially incorporate aging error in our simulation. The assessments conducted in our region rarely show an effect when the aging error matrix is used for a sensitivity analysis. However, our results beg the question of whether substantial aging error would dampen our results and change the data prioritization advice. Again we considered precision rather than bias and ran alternative scenarios with an aging error matrix that was larger than any ever used for our region (i.e., the readers agreed only half the time) before we saw a change in the results. The amount of aging error normally reported for stocks in our region did not cause a change in the data prioritization advice. In the extreme scenario we had to use to perturb our initial results, one might argue that an age-structured model should be avoided if the aging is so uncertain.

Acknowledgements

We thank the NOAA Assessment Methods Working Group for the grant that funded this project. We appreciate the helpful comments and criticisms about the manuscript from Kevin Craig and Amy Schueller, and we thank the anonymous reviewers for their helpful suggestions.

References

Andrews, R., Brick, J.M., and Mathiowetz, N.A. 2014. Development and testing of recreational fishing effort surveys: testing a mail survey design. *Edited by* NMFS.

Coggins, L.G., and Quinn, T.J., II. 1998. A simulation study of the effects of aging error sample size on sustained yield estimates. *In* Fishery stock assessment models. *Edited by* F. Funk, T.J. Quinn, II, J. Heifetz, J.N. Ianelli, J.E. Powers, J.F. Schweighert, P.J. Sullivan, and C.-I. Zhang. Alaska Sea Grant College, Anchorage, Alaska. AK-SG-98-01. pp. 955–976.

Conn, P.B., Williams, E.H., and Shertzer, K.W. 2010. When can we reliably estimate the productivity of fish stocks? Can. J. Fish. Aquat. Sci. **67**(3): 511–523. doi:10.1139/F09-194.

Fogarty, M.J., Sissenwine, M.P., and Cohen, E.B. 1991. Recruitment variability and the dynamics of exploited marine populations. TREE, **6**(8): 241–246. PMID: 21232469.

Fournier, D.A., Skaug, H.J., Ancheta, J., Ianelli, J.N., Magnusson, A., Maunder, M.N., Nielsen, A., and Sibert, J.R. 2012. AD Model Builder: using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. Opt. Methods Softw. **27**: 233–249. doi:10.1080/10556788.2011.597854.

Magnusson, A., and Hilborn, R. 2007. What makes fisheries data informative? Fish Fish. **8**(4): 337–358. doi:10.1111/j.1467-2979.2007.00258.x.

Maunder, M.N., and Punt, A.E. 2013. A review of integrated analysis in fisheries stock assessment. Fish. Res. **142**(0): 61–74. doi:10.1016/j.fishres.2012.07.025.

Methot, R.D. 2000. Technical description of the stock synthesis assessment program. NOAA Technical Memorandum. NMFS-NWFSC-43.

Powers, J.E., and Restrepo, V.R. 1993. Evaluation of stock assessment research for Gulf of Mexico King Mackerel: benefits and costs to management. N. Am. J. Fish. Manage. **13**(1): 15–26. doi:10.1577/1548-8675(1993)013<0015:EOSARF>2.3.CO;2.

Punt, A.E., Smith, D.C., and Smith, A.D.M. 2011. Among-stock comparisons for improving stock assessments of data-poor stocks: the "Robin Hood" approach. ICES J. Mar. Sci. **68**(5): 972–981. doi:10.1093/icesjms/fsr039.

Rindone, R.R., Kellison, G.T., and Bortone, S.A. 2015. Data availability for Red Snapper in Gulf of Mexico and southeastern U.S. Atlantic Ocean waters. N. Am. J. Fish. Manage. **35**(2): 191–204. doi:10.1080/02755947.2014.992559.

SEDAR. 2008. SEDAR 15 — Stock Assessment Report 2 (SAR 2): South Atlantic Greater Amberjack. SEDAR, North Charleston, S.C.

SEDAR. 2009. SEDAR 19 — Stock Assessment Report: South Atlantic Red Grouper. SEDAR, North Charleston, S.C.

SEDAR. 2010. SEDAR 24 — Stock Assessment Report: South Atlantic Red Snapper. SEDAR, North Charleston, S.C.

SEDAR. 2012*a*. SEDAR 1 Update: Stock Assessment of Red Porgy off the Southeastern United States. SEDAR, North Charleston, S.C.

SEDAR. 2012*b*. SEDAR 17 Update: Stock Assessment of Vermilion Snapper off the Southeastern United States. SEDAR, North Charleston, S.C.

SEDAR. 2013*a*. SEDAR 25 Update: Stock Assessment of Black Sea Bass off the Southeastern United States. SEDAR, North Charleston, S.C.

SEDAR. 2013*b*. SEDAR 36 — Stock Assessment Report: South Atlantic Snowy Grouper. SEDAR, North Charleston, S.C.

SEDAR. 2014. SEDAR 10 Update: Stock Assessment of Gag off the Southeastern United States. SEDAR, North Charleston, S.C.

Szuwalski, C., and Punt, A.E. 2012. Identifying research priorities for management under uncertainty: the estimation ability of the stock assessment method used for eastern Bering Sea snow crab (*Chionoecetes opilio*). Fish Res. **134–136**: 82–94. doi:10.1016/j.fishres.2012.08.007.

von Bertalanffy, L. 1938. A quantitative theory of organic growth (inquiries on growth laws. II). Hum. Biol. **10**(2): 181–213.

Williams, E.H., and Shertzer, K.W. 2015. Technical documentation of the Beaufort Assessment Model (BAM). US Department of Commerce, NOAA Technical Memorandum NMFS-SEFSC-671. doi:10.7289/V57M05W6.

Yin, Y., and Sampson, D.B. 2004. Bias and precision of estimates from an age-structured stock assessment program in relation to stock and data characteristics. N. Am. J. Fish. Manage. **24**(3): 865–879. doi:10.1577/M03-107.1.