ICES Journal of
Marine Science

ICES Journal of Marine Science (2020), 77(1), 162–173. doi:10.1093/icesjms/fsz191

Original Article

Comparisons of mean length-based mortality estimators and age-structured models for six southeastern US stocks

Quang C. Huynh [1]*, Nancie J. Cummings[2], and John M. Hoenig[1]

[1]Virginia Institute of Marine Science, William & Mary, P.O. Box 1346, Gloucester Point, VA 23062, USA
[2]Southeast Fisheries Science Center, 75 Virginia Beach Drive, Miami, FL 33149, USA

*Corresponding author: tel: +1 604 822 2731; e-mail: q.huynh@oceans.ubc.ca.

Huynh, Q. C., Cummings, N. J., and Hoenig, J. M. Comparisons of mean length-based mortality estimators and age-structured models for six southeastern US stocks. – ICES Journal of Marine Science, 77: 162–173.

Received 15 April 2019; revised 17 September 2019; accepted 19 September 2019; advance access publication 25 October 2019.

Length-based mortality estimators have been developed as alternative assessment methods for data-limited stocks. We compared mortality estimates from three methodologically related mean length-based methods to those from an age-structured model (ASM). We estimated fishing mortality and determined overfishing status, i.e. if $F/F_{MSY} > 1$, for six stocks which support important recreational and commercial fisheries in the southeastern United States. The similarities in historical fishing mortality between the length-based methods and the most recent assessments varied among the case studies, but the classification of overfishing status in the terminal year did not differ based on the choice of model for all six stocks. There was also high agreement in the number of overfishing years within different historical periods. Applications of length-based methods can be consistent with the results that might be obtained from an ASM. In one case, diagnostics were used to identify the problems with the length-based estimators. The potential for determining overfishing status from these methods can encourage data collection programmes for unassessed stocks.

Keywords: biological reference points, data-limited, data-poor, fisheries management, overfishing

Introduction

Simpler, alternative stock assessment methods for exploited stocks are generally desirable when a more complex age-structured stock assessment model may not be viable or practical from a management perspective (Chrysafi and Kuparinen, 2016). Simple methods are largely used in "data-limited" situations, where the data available for an assessment may be restrictive, for example, due to lack of sampling resources (Bentley, 2015). In these cases, tractable assessment methods typically make necessary simplifying assumptions regarding the population. However, a more comprehensive stock assessment model, such as an age-structured model (ASM), is typically used in "data-rich" scenarios where ageing information and multiple sources of data exist (Dichmont *et al.*, 2016). In both data-limited and data-rich scenarios, analytical methods are used to estimate historical trends in fishing mortality (F), biomass (B), or both. Model output, including forecasts and reference points, from such methods can be used to provide short-term management advice.

In data-limited situations, length-based assessment methods are appealing because they are easy to use and length information is easily collected for many fisheries. In conjunction with growth parameters, simple methods typically estimate mortality from a single size composition or mean length, often with equilibrium assumptions (Beverton and Holt, 1956; Hordyk *et al.*, 2015, 2016; Kokkalis *et al.*, 2015).

Recently, four related methods have been developed to analyse time series of mean length. These methods expanded the estimator of Beverton and Holt (1956), which estimates total mortality (Z) from a single observation of mean length. Development of these methods was motivated by the ability to relax the equilibrium assumptions of the Beverton–Holt method. Gedamke and Hoenig (2006) developed a non-equilibrium method for estimating total mortality (in this paper, we denote this model as "ML" to contrast from the mean length data). Changes in mortality over time are characterized by stepwise changes, and the non-equilibrium method accounts for the gradual change in mean length that arises following such a change. From a time series of mean length, a historical series of mortality rates and the timing of the changes in mortality are estimated. This method has also been used with proxy reference points for MSY to determine overfishing status, i.e. if $F/F_{MSY} > 1$ (Huynh, 2016).

Subsequent extensions of Gedamke and Hoenig (2006) incorporate additional data types with mean lengths to relax additional assumptions and evaluate goodness of fit. The approach can be expanded to incorporate either recruitment indices to relax the constant recruitment assumption (Gedamke *et al.*, 2008), effort to provide year-specific mortality estimates (Then *et al.*, 2018), or both (ICES, 2017). Indices of abundance also contain information on mortality and can be used with mean lengths to estimate mortality (Huynh *et al.*, 2017). Here, these models will be collectively referred to as ML-based models.

To evaluate how simpler, data-limited methods may perform relative to ASMs, the former can be applied to data sets from stocks for which there are age-structured assessments (e.g. Dick and MacCall, 2011; Kokkalis *et al.*, 2017). Synchrony in the results among models, i.e. whether or not the historical stock trends are in agreement, can be a form of endorsement for the data-limited methods. While there is no guarantee that the ASM is correct nor that it produces precise and accurate estimates, benchmark assessments undergo a thorough peer-review process and the results of the ASMs usually represent our best knowledge of the stock (Dichmont *et al.*, 2016). If similar results are obtained among models, then the use of simpler models is inconsequential for classifying overfishing status. Use of the simpler models could also be advantageous for management agencies to allocate resources to stocks that have not been previously assessed.

In this study, we use three multi-year, ML-based methods to estimate historical fishing mortality for six stocks in the southeastern United States. These stocks are of interest because they have been assessed using ASMs. The stocks are Gulf of Mexico (GOM) greater amberjack *Seriola dumerili*, GOM Spanish mackerel *Scomberomorus maculatus*, GOM cobia *Rachycentron canadum*, Atlantic (ATL) cobia, GOM king mackerel *Scomberomorus cavalla*, and ATL king mackerel. The Beaufort Assessment Model (BAM; Williams and Shertzer, 2015) was used for ATL cobia, while Stock Synthesis (SS; Methot and Wetzel, 2013) was used for all others.

For these stocks, length composition data were used in the age-structured assessments which were accepted as the basis for management advice by NOAA (National Oceanic and Atmospheric Administration) Fisheries. The length data from these assessments were then obtained for analysis with the ML-based methods. In an ASM, length data potentially contain information on recruitment strength, mortality, and selectivity. While these data primarily inform mortality with fixed assumptions regarding recruitment and selectivity in the ML-based models, a common subset of data allows for comparison of historical mortality rates between these two types of models. We compared the trends in historical fishing mortality and the classification of overfishing status using F/F_{MSY} estimates between the ML-based models and the age-structured assessments. We also used model diagnostics, i.e. residuals, for the ML-based models to explain whether the methods were suitable for the particular stocks.

Methods

Stocks of interest and their assessments

Greater amberjack is managed under the Reef Fish Fishery Management Plan, and Spanish mackerel, cobia, and king mackerel are managed under the Coastal Migratory Pelagic Fishery Management Plan of the Gulf of Mexico Fishery Management Council and South Atlantic Fishery Management Council. Each of the four species is considered to be separate GOM and ATL stocks for management purposes.

Over time, these stocks have been managed with seasonal closures, bag limits, minimum size limits, and catch limits, i.e. quotas. Size limits, i.e. minimum retention sizes, have generally increased over time for the recreational sector (Table 1). The recreational sector includes the charterboat/private fleet and the headboat fleet. The charterboat/private fleet consists of boats rented by day (or half day) for a small group of recreational anglers, whereas headboats charge on a per-person, per-trip basis, and typically have more anglers than charterboats per fishing trip.

Benchmark assessments for these stocks were conducted in 2013–2014 (SEDAR, 2013a, b, c, 2014a, b, c). Data inputs for these ASM included landings, discards, standardized indices of abundance, length composition, and length-at-age observations from commercial and recreational sectors. Fishery-dependent indices were derived from fishery catch-per-unit-effort (CPUE). Fishery-independent indices and length compositions from surveys were also included in the assessments, although the time series are shorter than for fishery-dependent data. For some assessments, the charterboat/private and headboat fleets were treated as a single recreational fleet if both are thought to behave similarly in targeting the stock (Table 2).

In addition to fishing mortality, ASM assessments estimated selectivity (specified to be either logistic or dome-shaped), annual recruitment, and growth parameters. The number of growth parameters varied among assessments. For example, K was estimated with L_∞ fixed for GOM greater amberjack, whereas both were fixed for ATL cobia and all growth parameters were estimated for GOM cobia, GOM Spanish mackerel, GOM king mackerel, and ATL king mackerel. For all stocks, estimates from growth studies were available prior to the assessment (Table 3). Natural mortality (M) varied by age in the assessment, using the parameterization from Lorenzen (1996) and subsequently rescaled such that the mean value was equal to that obtained from Hoenig (1983) using maximum observed age.

ML-based mortality estimators

Three mean length-based methods were used to estimate mortality: (i) the non-equilibrium ML estimator of Gedamke and Hoenig (2006); (ii) the mean length with catch rate (MLCR) estimator of Huynh *et al.* (2017); and (iii) the mean length with effort (MLeffort) estimator of Then *et al.* (2018). A technical

Table 1. Summary of size regulations from the recreational fishery (in terms of fork length).

Stock	Minimum legal size limit (cm)	Years
GOM greater amberjack	28 in (71.1)	1990–2007
	30 in (76.2)	2008–2012
GOM Spanish mackerel	12 in (30.5)	1993–2011
GOM and ATL cobia	33 in (83.8)	1985–2011
GOM and ATL king mackerel	12 in (30.5)	1990–1991
	20 in (50.8)	1992–1999
	24 in (61.0)	2000–2012

Only years preceding the year of the assessment are considered. Size regulations were obtained from the assessment documents (Table 2). Size regulations are published in inches.

Table 2. Summary of assessment models and the length composition and index of abundance for the length-based mortality estimators.

Stock	Assessment model	Fleet for length analyses	Length time series	Index time series	References
Gulf of Mexico greater amberjack	SS	Charter/private	1981–2012	1986–2012	SEDAR (2014a)
Gulf of Mexico Spanish mackerel	SS	Recreational	1981–2011	1981–2011	SEDAR (2013b)
Gulf of Mexico cobia	SS	Recreational	1979–2011	1986–2011	SEDAR (2013a)
Atlantic cobia	BAM	Recreational	1982–2011	1985–2011	SEDAR (2013c)
Gulf of Mexico king mackerel	SS	Charter/private	1985–2012	1986–2012	SEDAR (2014b)
Atlantic king mackerel	SS	Charter/private	1978–2012	1980–2012	SEDAR (2014c)

The Recreational fleet combines the data from both the Charter/private and the Headboat fleets.

Table 3. Life history parameters used in the analyses for the length-based mortality estimators.

Stock	L_∞ (cm)	K (year^{-1})	t_0 (year)	L_c (cm)	L_{mat} (cm)	α	b	t_{max} (yr)	M (year^{-1})	Sources
Gulf of Mexico greater amberjack	143.6	0.18	−0.95	77.5	90	7.0e-5	2.63	15	0.28	SEDAR (2014a) and Murie and Parkyn (2008)
Gulf of Mexico Spanish mackerel	56.0	0.61	−0.50	39	31	1.5e-5	2.86	11	0.38	SEDAR (2013b)
Gulf of Mexico cobia	128.1	0.42	−0.53	88	70	9.6e-6	3.03	11	0.38	SEDAR (2013a)
Atlantic cobia	132.4	0.27	−0.47	95	70	2.0e-9	3.28	16	0.26	SEDAR (2013b)
Gulf of Mexico king mackerel	128.9	0.12	−4.08	80	58	7.3e-6	3.01	24	0.17	SEDAR (2014b) and Lombardi (2014)
Atlantic king mackerel	121.1	0.15	−3.73	80	58	7.3e-6	3.01	26	0.16	SEDAR (2014c) and Lombardi (2014)

Parameters are defined in Supplementary Table B.1.

description of the three methods is provided in Supplementary Materials A.

The analyses were based on the data from the recreational sector. This sector was chosen because it is believed that this sector has been most informative for inference on stock trends in the benchmark assessments (Sagarese *et al.*, 2016). In the southeastern United States, the largest targeted fishing effort has historically come from the recreational sector (Siegfried *et al.*, 2016). The indices from the recreational sector have also generally had the lowest root mean square error in the age-structured assessments (Sagarese *et al.*, 2016). In cases where the two recreational fleets are distinct units in the assessment, data from the larger charterboat/private fleet were used for the ML-based methods. The length compositions, standardized indices of abundance, and the landings corresponding to the index were obtained directly from the assessments (Table 2).

In contrast to the ASM which accommodates and estimates the parameters of various selectivity functions, all ML-based methods assume knife-edge selectivity and require an estimate of the length at full selection (L_c) to be determined prior to the analysis. The mode of the length composition compiled for all years was chosen to be the L_c, which was larger than the minimum retention size for all stocks (Figure 1). There was generally no trend in the modal length over most years for the six stocks. The annual mean length of animals larger than L_c was calculated, and von Bertalanffy asymptotic length (L_∞) and growth coefficient (K) were obtained from growth studies presented during the benchmark assessment (Table 3).

First, the ML estimator was used to estimate mortality. From annual observations of mean length, the time series is partitioned into stanzas of constant mortality. The total mortality rates and the duration of each stanza are then estimated. Total mortality is modelled as a stepwise change from one stanza to another, and the predicted mean length changes gradually depending on previous mortality rates and elapsed time since mortality changed.

Second, the index of abundance was used in conjunction with the mean length time series with MLCR. In this model, both the mean length and the index are predicted to decrease gradually after a stepwise increase in mortality and, similarly, to increase after a decrease in mortality. This allows for an evaluation of the consistency between the length and index data for mortality estimation.

The ML and MLCR models were systematically fitted by varying the number of stanzas and Akaike Information Criterion (AIC) was used to select the best fitting model, i.e. the model with the lowest AIC score. To avoid overfitting, models with more parameters were accepted only if the reduction in AIC was >2 (Burnham and Anderson, 2002). Models were fitted assuming zero, one, or two change points in mortality (additional analyses with >2 change points were not supported by AIC).

While ML and MLCR estimate Z, we assume, as many ASMs do, that M is constant over time. Thus, changes in Z examined here are assumed to arise solely from changes in F. From total mortality estimates, fishing mortality F was obtained by subtracting the value of M assumed in the benchmark assessments (Table 3). Since ML-based models also assume constant mortality across all selected ages, the age-invariant M obtained from the Hoenig (1983) method was used.

Third, year-specific mortality rates were estimated from mean lengths and estimates of effort, using the MLeffort model (Then *et al.*, 2018). In this method, fishing mortality F is proportional to fishing effort f via the estimated catchability coefficient q. Total mortality Z in year y of the model is $Z_y = qf_y + M$, where f_y is the effort and M was fixed in the model (to the same value used in ML and MCLR). This formulation precludes the need to estimate mortality in time stanzas. The effort time series was obtained by taking the ratio of the landings (thousands of fish) and index of abundance (CPUE, number of fish per angler hour). Since the model requires a full time series of effort, the first year of the

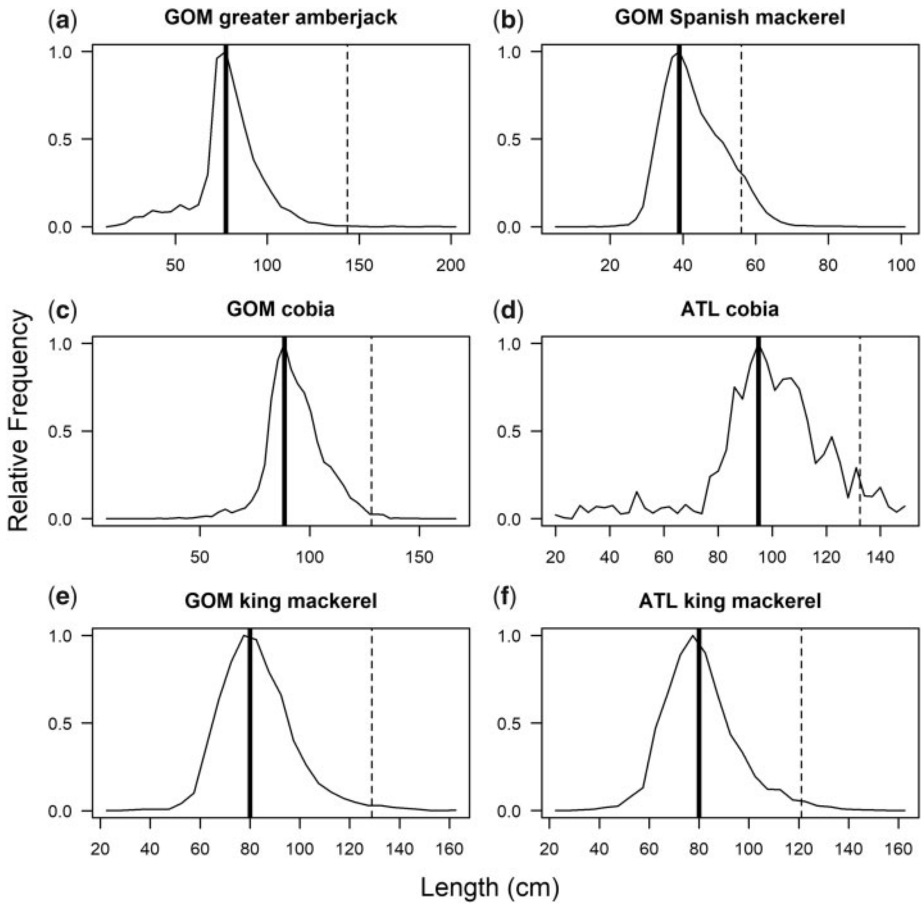

Figure 1. Summary length compositions summed across all available years of data for the six stocks (a–f) for the ML-based mortality estimators. Solid vertical line indicates L_c and dashed vertical line indicates L_∞.

model was set to the first year with available indices of abundance. The equilibrium effort prior to the first year of the model was set equal to the effort in the first year.

All three models were fit using maximum likelihood. Visual analysis of standardized residuals, calculated by subtracting the predicted value from the observed and then dividing by the estimated standard deviation, was used to indicate the quality of fit in the respective model. Residuals in mean lengths were calculated for all methods, with additional residuals in the indices of abundance also calculated for the MLCR model.

Comparison among models

Two sets of quantities were used to facilitate comparison among the ASM, ML, MLCR, and MLeffort. First, the absolute magnitude of the F estimates from the all four models was used. Annual estimates of F from the ASM were obtained from assessment reports (SEDAR, 2013a, b, c, 2014a, b, c). Only estimates since the first year of length composition data were considered here (Table 2).

Second, the annual F estimates were divided by F_{MSY} (relative F). The F/F_{MSY} ratio is often relevant to management for classification of historical and current overfishing status. Proxy reference points are often used instead of directly estimating F_{MSY}. In the benchmark assessments, $F_{30\%}$, the fishing mortality rate that reduces the spawning potential ratio (SPR) to 0.3, was generally used as the proxy for F_{MSY}. The exception was in the case of ATL cobia, where F_{MSY} was reported for the ASM instead of a proxy (SEDAR, 2013c).

The calculation of the value of the proxy reference point should be consistent with the assumptions of the method used to estimate F. As a result, two separate calculations of SPR were used. For the ASM, the value of $F_{30\%}$ was obtained from the assessment documents, while for the ML-based methods, a separate value was calculated for $F_{30\%}$ assuming knife-edge selectivity and constant M with age (Supplementary Materials B). Values of $F_{30\%}$ were identical for ML, MLCR, and MLeffort because the selectivity and M assumptions among them were identical.

To evaluate the synchrony of relative F among models, the proportion of years in which overfishing is estimated to occur was calculated for four time periods: (i) pre-1995 (approximately the first half of the time series for the six stocks), (ii) post-1995 (approximately the second half of the time series), (iii) the last 5 years, and (iv) the terminal year of the time series.

All analyses were performed in the R statistical environment using the MLZ package, which is publicly available on the CRAN repository (R Core Team, 2017; Huynh, 2018).

Results

For most stocks analysed here, all methods generally indicated high mortality in the 1980s–1990s followed by a reduction in

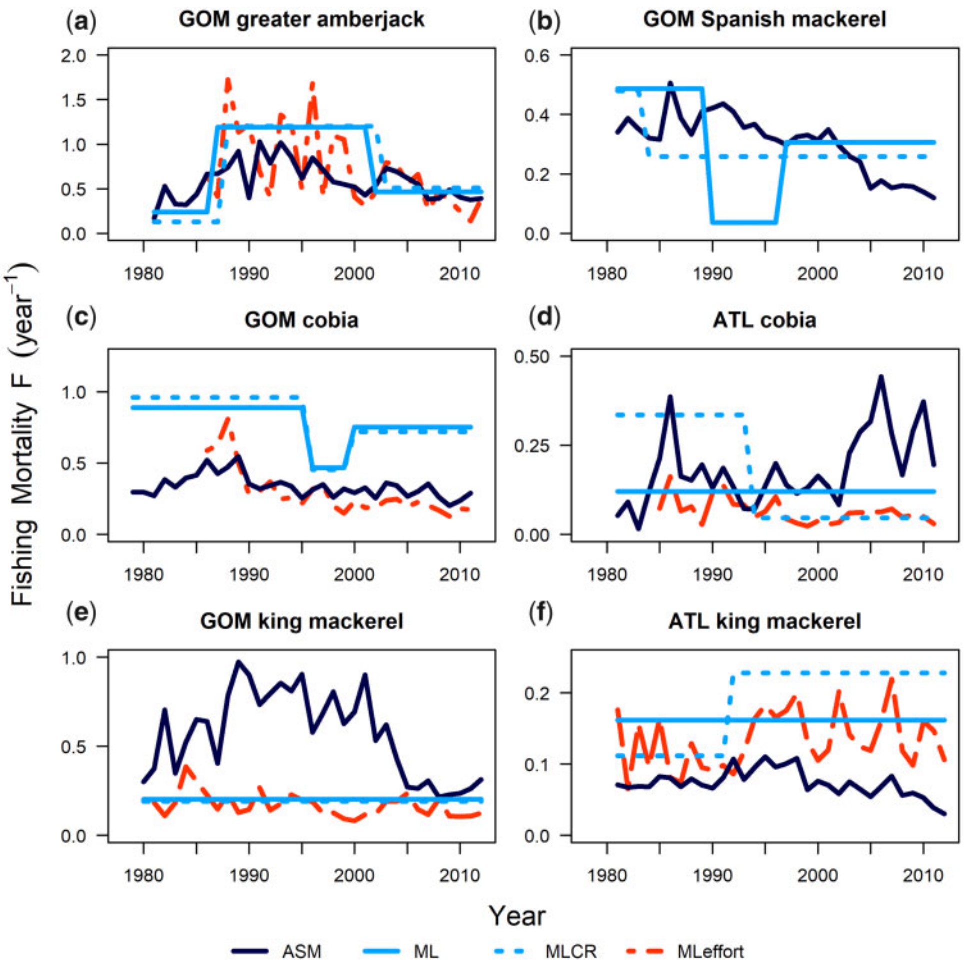

Figure 2. Annual estimates of F from the four models (ASM, age-structured model; ML, mean length; MLCR, mean length with catch rate; MLeffort, mean length with effort) for the six case studies (a–f). The MLeffort model did not converge for GOM Spanish mackerel. The ASM was the BAM for ATL cobia and SS for all other stocks.

mortality since then (Figure 2). For all six stocks, the four models agreed in the overfishing status in the terminal year of the time series, i.e. $F/F_{\mathrm{MSY}} > 1$ for GOM greater amberjack and $F/F_{\mathrm{MSY}} < 1$ for the other five stocks (Figures 3–4).

GOM greater amberjack

There was strong agreement in the mortality estimates over time in both terms of trend and magnitude (Figure 2a). Both the ASM and MLeffort models showed an increase in F from 1981 to 1993 followed by a gradual decrease from 1993 to 2012, with higher inter-annual variability in F from MLeffort. Both models suggested very similar declines in mortality. The ML and MLCR models showed two changes in mortality, an initial increase to an extended plateau in mortality during the 1990s, corresponding to the time period surrounding the peak in the ASM and MLeffort models, followed by a reduction in the 2000s. The F from ML and MLCR during the 1990s were higher compared to estimates from the ASM.

Furthermore, all models showed that overfishing was occurring in 2012, the terminal year of the time series (Figure 3a). The magnitudes of relative F, i.e. F/F_{MSY}, over time were very similar among the four models, with a very large relative F in the late 1980s and 1990s coinciding with large observed catches (SEDAR, 2014a). Although a reduction in relative F followed, overfishing was still occurring in 2012. In addition, the four models generally agreed on the extent of overfishing within the four time periods (Figure 4). While a lower proportion of overfishing years was inferred in the most recent 5 years for the MLeffort model compared to the other three models, this appeared to be a result of the high inter-annual variability in relative F.

GOM Spanish mackerel

The ASM, ML, and MLCR models all showed a general reduction in mortality over time, although the trends and timing differ (Figure 2b). The MLeffort model did not converge. The ASM showed relatively high F in the 1980s and early 1990s followed by a gradual reduction in F afterwards. The reduction started in the late 1990s coincident with the gillnet ban in Florida, although mortality from all sectors (commercial, recreational, and bycatch) has since reduced (SEDAR, 2013b). The trend from the ML model is markedly different compared to the ASM and MLCR. Two changes in mortality were indicated, with a decrease in mortality to a very low level during the early 1990s from the initial mortality rate. This was caused by the large increase in mean length from 1990 to 1995 (Figure 5b). Afterwards, a modest increase to an intermediate

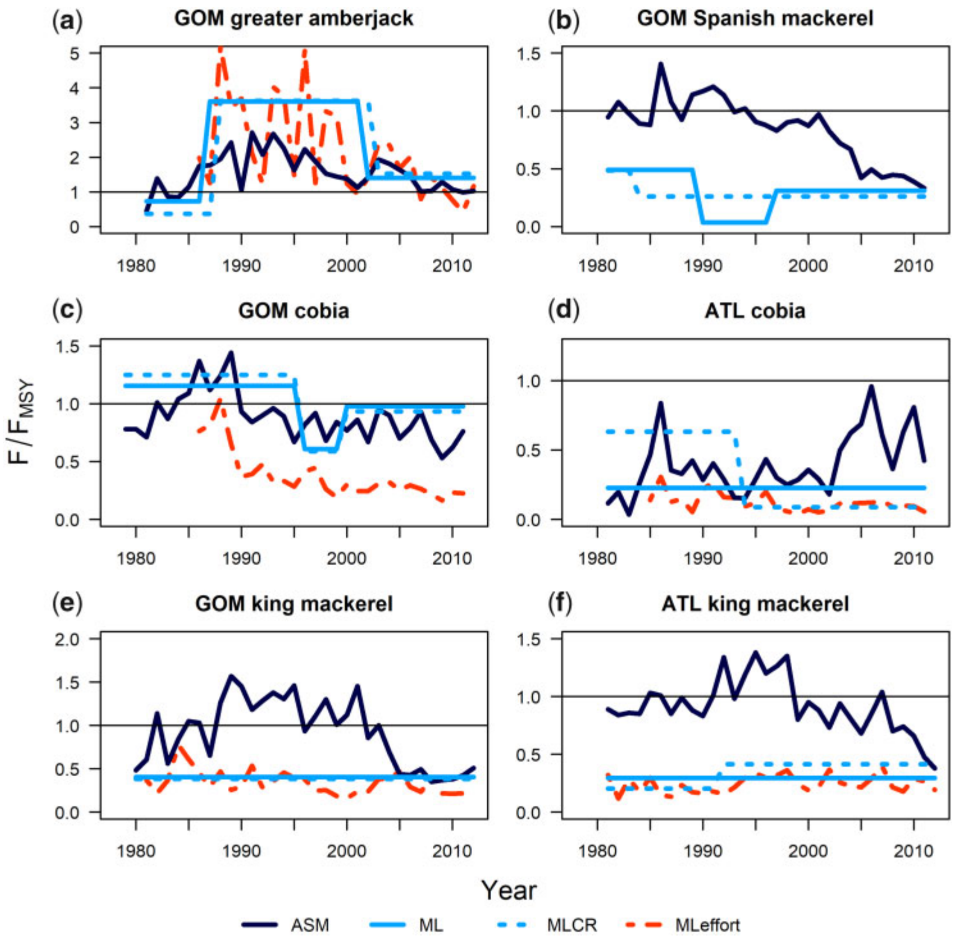

Figure 3. Annual estimates of F/F_{MSY} (relative F) from the four models for the six case studies (a–f). F_{MSY} was reported from the ASM for ATL cobia while for all other methods, the F_{MSY} proxy is $F_{30\%}$. Separate calculations of $F_{30\%}$ were used for the ASM and ML methods.

mortality rate until the present time was estimated. The trends in the index, however, did not support two changes in mortality (Figure 5b). Thus, only one change in mortality, a modest decrease over the time series, was inferred in the MLCR model (Figure 2b).

Compared to the ML and MLCR models, the ASM showed more contrast in relative F, with overfishing occurring in 8 out of 14 years (57%) in the pre-1995 period (Figure 4). The ML and MLCR models showed that overfishing has not occurred (Figure 3b). All three models agreed that overfishing has not occurred post-1995.

GOM cobia

All four models indicated a reduction in mortality since the 1990s (Figure 2c). The ASM showed an initial upward ramp in mortality followed by a gradual decrease after 1990. The MLeffort model showed a large decrease prior to 1986–1990 (effort data were not available prior to 1986), but after 1990, the mortality trend closely mimicked that inferred in the ASM in magnitude over time. The ML and MLCR models both estimated two changes in mortality, with a temporary decrease in the late 1990s followed by a modest increase to a mortality rate that is less than the initial estimated mortality rate. This pattern was inferred from the synchronous increase and decrease in the mean length and index in the late 1990s (Figure 5c). The ML and MLCR models estimate much higher F than the other two models (Figure 2c).

The relative F in MLeffort was lower over time than in the other three models. Pre-1995, an increase and decrease in relative F corresponded to overfishing in 1 out of 9 years (11%) in the MLeffort model, but 7 out of 16 years (44%) in the ASM (Figure 4). During the same time period, the ML and MLCR estimated a plateau in mortality which indicated overfishing in all included years. Post-1995, overfishing has not occurred based on all four models (Figure 3c).

ATL cobia

Differing trends in mortality were inferred among the four models (Figure 2d). While there were trends in the mean length over time, the ML model indicated zero changes in mortality based on AIC. However, the MLCR model indicated a decrease in mortality, largely based on the increase in the index after 1995 (Figure 5d). The MLeffort model showed a gradual decrease in mortality over time. The ML-based models estimate lower F than the ASM in recent years, although there is high inter-annual variability in F estimates in the latter without a clear trend over time. Based on the relative F from all four models, overfishing has not occurred (Figures 3d and 4).

GOM king mackerel

Differing trends in mortality were estimated among the models (Figure 2e). The stability in mean lengths over time resulted in

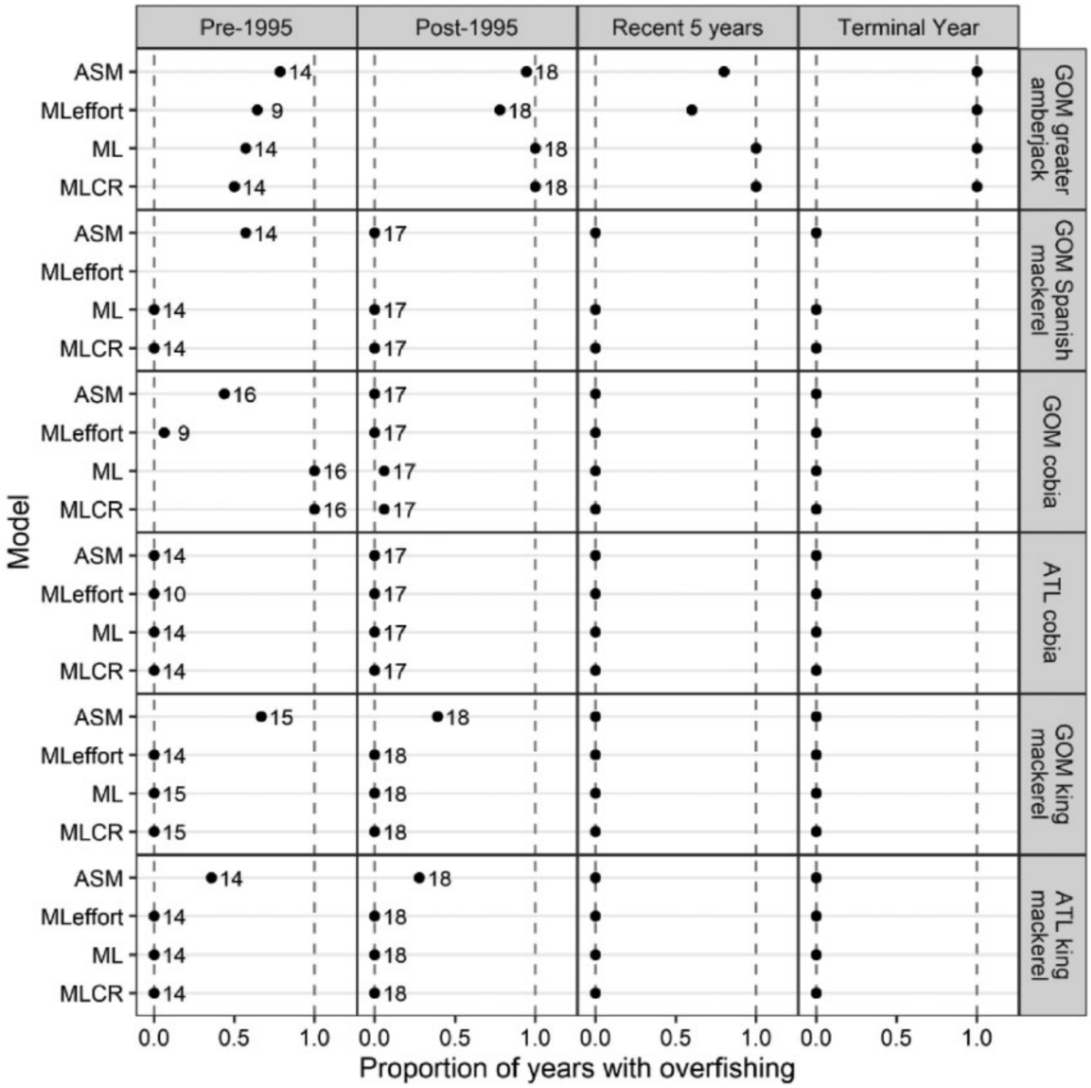

Figure 4. The proportion of years with overfishing as estimated with the four models within the respective time periods for the six stocks. The MLeffort model did not converge for GOM Spanish mackerel. For pre-1995 and post-1995, numbers indicate the number of years in the assessment for the respective time period.

estimates of constant F over the entire time series from the ML and MLCR models. The trend in F in the MLeffort model was relatively flat as well. Fishing mortality was much higher in the ASM than in the ML-based models from the 1980s to the mid-2000s, although the difference decreased with a pronounced drop in F in the ASM the late 2000s.

The ASM showed that overfishing was occurring over much of the pre-1995 period, contrary to the other three models which showed no overfishing in the same time period (Figure 3e). In the early part of the post-1995 period, the ASM showed that overfishing was occurring (20–40% of years post-1995) until mortality was reduced shortly after 2000. The ML-based models indicated no historical overfishing.

ATL king mackerel

The F trend in the ASM is relatively flat with a slight decrease in the recent years (Figure 2f). The ML and MLeffort models produce relatively stable F over time as well, although the magnitude is higher in these models than in the ASM. The MLCR model produces a pronounced stepwise increase in F in the mid-1990s due to the pronounced decrease in the index at this time (Figure 5f).

The ASMs indicated that overfishing occurred in 29% (5 out of 17 years) of pre-1995 years (Figure 4). The ML-based models here also did not indicate overfishing in the stock history (Figure 3f).

Residual analysis

For each of the ML-based models, residuals were analysed visually to examine goodness of fit (Supplementary Materials C). The model selection procedure with the ML model generally selected the model which minimized residual trends except in the case of ATL cobia (Supplementary Figure C.1). In the MLCR model, an extensive trend of positive and negative residuals of the mean lengths and index, respectively, was observed over time for GOM Spanish mackerel (Supplementary Figure C.2). Similarly, negatively correlated residuals were also present for ATL king mackerel in the most recent years of the analysis. In the

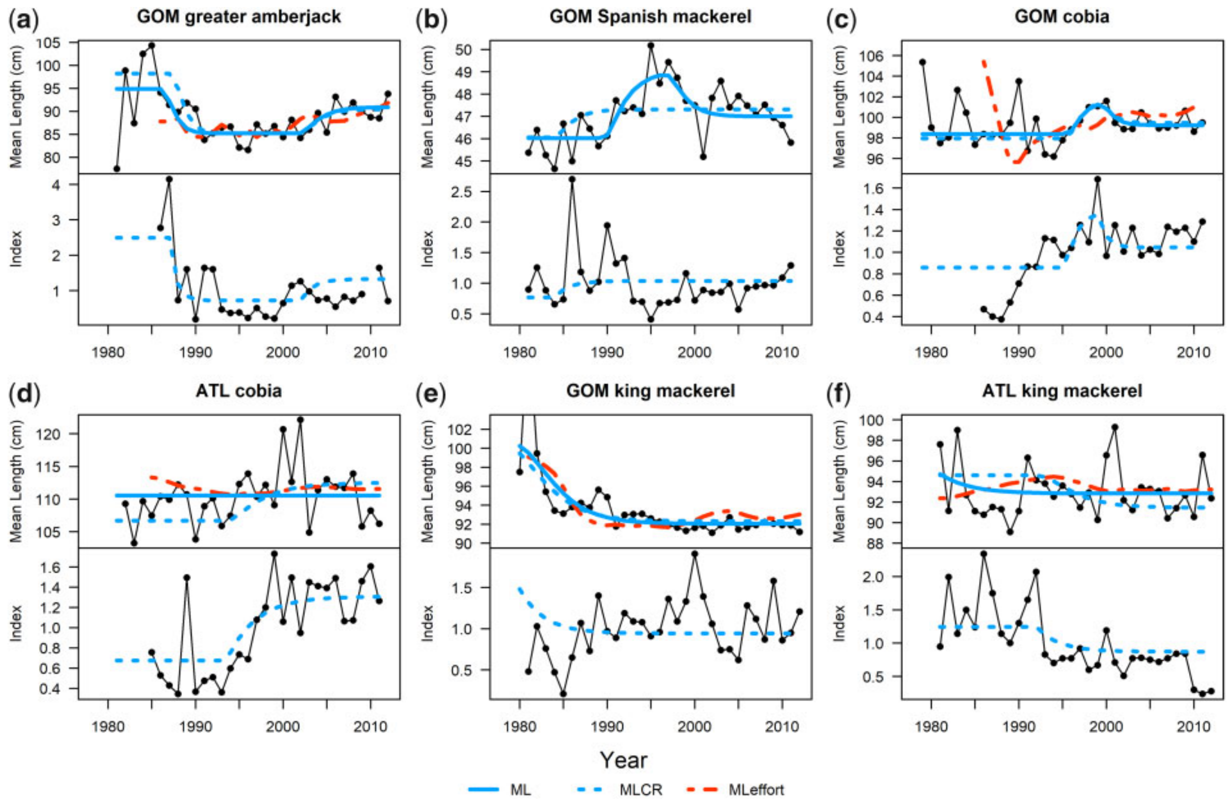

Figure 5. Observed (connected points) and predicted mean lengths (coloured lines) from the three length-based mortality estimators, and observed and predicted index for the MLCR model for the six case studies (a–f).

MLeffort model, there were trends in residuals over the course of the entire time series for both GOM and ATL king mackerel (Supplementary Figure C.3).

Discussion

The historical mortality pattern observed here, high mortality in the 1980s–1990s followed by a reduction, is common for south-eastern US stocks that were targeted by fisheries that were unregulated during these decades (Siegfried *et al.*, 2016). Although differences in the magnitude of F/F_{MSY} varied for the terminal year of the analyses, which potentially affect the management advice, there was agreement in the stock perception, i.e. overfishing vs. not overfishing, among the ML-based models and the ASMs for the six case studies.

For data-limited situations, there is potential to use ML-based models to explore historical changes in mortality over time. Despite only using a subset of the data, the results are likely to be consistent with what might be obtained from an ASM. The ML and MLCR models provide a series of historical mortality rates, although it is recognized that the changes in mortality over time will be coarser than in models with year-specific mortality rates. This is due to the stepwise, time stanza structure of the ML and MLCR models. The MLeffort model can provide year-specific mortality rates, and F estimates could be smoothed *post hoc* to describe the trend over time if there is high inter-annual variability.

Trends in recruitment to the recreational sector

The assumption of constant recruitment to length L_c was likely violated for GOM Spanish mackerel due to the changes in the dynamics of the shrimp fleet over time which affected bycatch of smaller animals. In the ASM assessment, the shrimp fleet was the highest source of fishing mortality (with 100% discard mortality assumed) until the late 1990s, when effort subsequently decreased (SEDAR, 2014b; Figure 6). This reduction increased survival and recruitment to size L_c (39 cm in this study), which could have caused the decrease in the observed mean length from the recreational fleet (Figure 5b).

For the MLeffort model, non-convergence for GOM Spanish mackerel was caused by the data conflict where the recreational effort and mean length concurrently decreased (Figures 5b and 6). An increase would have been expected based on the observed effort trend alone in the mean length. Concurrently, the gradual increase in the index of abundance with the decrease in ML since the mid-1990s would support the hypothesis of increased recruitment to the recreational fishery (Huynh *et al.*, 2017). Fewer change points were inferred with the MLCR model compared to the ML model. If there are trends in recruitment, MLCR can avoid overfitting spurious trends in the mean length data. The observed trends in the paired residuals of mean length and the abundance index in the MLCR model were also consistent with hypothesized increased recruitment (Supplementary Figure C.2). Indeed, the ASM corroborates this hypothesis since it estimated an increase in abundance of animals recruiting to the 39-cm length class during the same time period (Figure 6).

While trends in mortality are affected by factors external to the recreational sector, the analysis of residuals in the MLCR model and non-convergence of the MLeffort model allowed us to diagnose issues in the application of the ML-based models for GOM

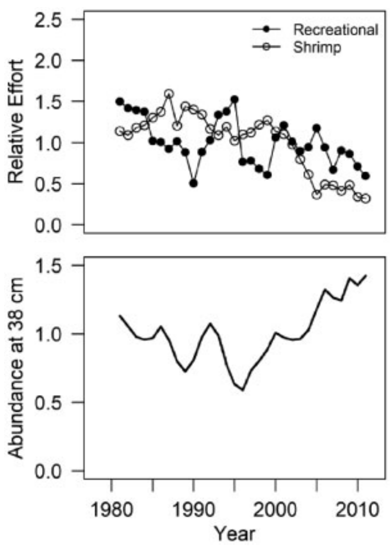

Figure 6. Upper: Estimates of relative effort for GOM Spanish mackerel from the recreational fleet, obtained as the ratio of the recreational catch and index of abundance, and the shrimp bycatch fleet, estimated as described in Linton (2012). Estimates are scaled so that the time series mean is one. Lower: Relative abundance at the 38-cm length bin (relative to the time series mean) estimated from the ASM. This length bin corresponds to the presumed length of recruitment (39 cm) to the recreational fleet in the ML-based models. Increased recruitment to the recreational fleet from decreased shrimp bycatch mortality is hypothesized to decrease the mean length despite the decrease in recreational effort.

Spanish mackerel without external information. With the ASM, we can corroborate that bycatch mortality may have been the primary driver of the historical stock dynamics. In isolation, the length composition from the recreational fleet may not provide sufficient information on the stock history, i.e. reductions in *F*. This is evident in the contrasting trends in mortality in the ML model and ASM since the mid-1990s (Figure 2b). Overall, the general presence of large animals in the length composition relative to L_∞ would indicate that the GOM Spanish mackerel stock is in generally good shape (Figure 1b).

The impact of bycatch mortality from the shrimp fleet would not be as noticeable in the length-based analysis for GOM and ATL king mackerel, since shrimp bycatch is a minor source of mortality relative to the recreational fleet. Nevertheless, for ATL king mackerel, large residuals in the mean lengths and index were observed in the most recent years of the MLCR models (Supplementary Figure C.2). The increasing mean length and decreasing index since 2007 would be consistent with decreased recruitment (animals of length L_c). The ASM for ATL king mackerel estimated a decreasing trend in recruitment of age-0 animals since 2003. The qualitative information about recruitment trends from the MLCR model are further supported by the reduced recruitment estimates from the ASM after accounting for the time lag from age 0 to the age of full selection to the recreational fishery (SEDAR, 2014c).

Management actions may need to be more precautious when presented with information about recent reduced recruitment. Overall, the GOM Spanish mackerel and ATL king mackerel case studies highlight the benefit of indices of recruitment in a length-based analysis. Such information can be incorporated into the analysis to account for variable recruitment (Gedamke *et al.*, 2008).

These two case studies also highlight the fact that ASMs should not be replaced by simpler methods without cautious considerations. ASMs provide more modelling options to accommodate multiple drivers of fishing mortality and productivity, as well as more diagnostic tools to evaluate the quality of the assessment. Nevertheless, in data-rich scenarios, the ML-based methods can be used as a diagnostic to evaluate and explain how the mean length has changed over time (through fishing mortality or other causes) (e.g. SEDAR, 2013c). When there are conflicting results, diagnostic procedures can provide additional insight on the causes of model or data conflict. Models which incorporate multiple data types are advantageous, because the agreement (or lack thereof) between data types can be used to determine whether the chosen model is appropriate for the stock of interest.

Life history parameters

The ML-based models and their corresponding reference point proxies require simpler life history assumptions than the ASM. With ASMs, growth incorporates variability in size at age and parameters may be estimable within the model (Francis, 2016). In contrast, growth is fixed and assumed to be deterministic with age in the ML-based models, although simulations have suggested robustness of the ML-based models to this assumption (Then *et al.*, 2015; Huynh *et al.*, 2018).

In many ASMs, including those presented here, natural mortality was parameterized to asymptotically decline with age. Age-varying *M* would violate the assumption of age-constant *Z*, especially for the youngest age classes which may experience much higher *M* than older ones (Lorenzen, 1996). If selectivity were restricted to the oldest age classes, then the violation of this assumption could be minimal as *M* is similar among these ages. Simulation studies can be used to evaluate the bias, if any, in mortality estimates from the ML-based methods arising from age-varying *M*.

Errors in growth and natural mortality have similar effects on mortality estimates in both length-based methods and ASMs. An overestimate of asymptotic length leads to the perception of an overly truncated size composition and smaller mean length, resulting in an overestimate of fishing mortality. Since length data contain information on total mortality, an overestimate of natural mortality would result in an underestimate of fishing mortality. Simulation studies and sensitivity analyses have largely confirmed these trends (Clark 1999; Hordyk *et al.*, 2015; Huynh *et al.*, 2017). Further work is needed to evaluate whether mortality estimates from a length-based method is more sensitive to errors than those from ASMs.

In data-limited situations, uncertainty in mortality estimates can be evaluated in several ways. While confidence intervals can be obtained from the Hessian matrix of maximum likelihood models, the intervals are conditional on the assumptions of the model, including that life history parameters are known correctly. Alternatively, Monte Carlo sampling of life history parameters from parametric distributions (Huynh *et al.*, 2017; Nadon, 2017) and sensitivity analyses of alternative parameter values (Gedamke and Hoenig, 2006) have been employed to characterize uncertainty of mortality estimates. Bayesian methods that employ life history priors can also be used to make probabilistic statements regarding the mortality estimate and overfishing status (Harford *et al.*, 2015).

Such methods can be employed in the ML-based models here to calculate confidence intervals or posterior intervals in F and F/F_{MSY}.

Notably, confidence intervals and posterior intervals are conditional on the assumptions of the model. The length-based methods used here assume constant recruitment, but only the MLCR model allows for evaluation of this assumption (Huynh *et al.*, 2017). Even for the MLCR model, confidence intervals for mortality estimates would not include the effect of failure of this assumption when in fact there is a trend in recruitment over time.

Selectivity and retention behaviour

Complex fishing behaviour can be modelled in ASMs, albeit at the cost of estimating many, sometimes confounding, selectivity parameters. Multiple fishing fleets with disparate selectivity patterns and fishing behaviours are typically modelled separately, and there may be enough information to model logistic and dome-shaped selectivity functions. Length composition of discarded and retained catch allow for estimation of the vulnerability and retention functions, the product of which would be the effective selectivity of the gear for retained catch. Finally, changes in size regulations can be modelled with time-varying features of the ASM (Methot and Wetzel, 2013). For the ML-based models, knife-edge selectivity is assumed at length L_c. Thus, the analysis uses a subset of the length composition data so that only animals assumed to be fully selected are included in the calculation of the mean length.

Application of the data-limited models should consider if changes in mean length occurred due a change in retention behaviour as opposed to a change in mortality. We chose values of L_c that were larger than any implemented minimum retention size for the stocks in this study. In this way, all lengths larger than L_c would have the same presumed selectivity to minimize the effect of the management regulations. However, to the extent that there has been variable fishing over time on fish smaller than L_c, the assumption of constant recruitment is violated by being confounded with fishing mortality. Changes in bag limits could alter discard and retention behaviour; for example, the implementation of a bag limit may increase discarding of smaller animals in favour of retaining larger ones. To account for this, one would need to evaluate whether there were significant changes in the length distribution of retained catch once those regulations were implemented.

The age-structured assessments estimated dome-shaped selectivity for the recreational fleet for three of the six stocks, these being GOM greater amberjack and both GOM and ATL stocks of king mackerel. This contrasts with the knife-edge selectivity assumption made with the ML-based models. If the selectivity of the fleets were dome-shaped, then it is presumed that mortality would be overestimated by the length-based models. However, there was no consistent discrepancy for these three stocks in this study. Mortality estimates for GOM greater amberjack did not substantially differ between those in the ASM and from ML-based models. However, mortality estimates from ML-based models were higher than those in the ASM for ATL king mackerel but lower for GOM king mackerel. Certainly the degree of doming could affect the magnitude of the discrepancy. High F, such as those seen in GOM greater amberjack, would decrease the influence of dome selectivity in the bias of mortality estimates, since fewer animals would survive to the larger size classes affected by the dome selectivity.

Uncertainty in catch and effort

In any assessment, the quality of the data and their representativeness to the underlying population dynamics should be evaluated. For example, since discard estimates had generally large coefficients of variation (Siegfried *et al.*, 2016). In data-limited situations, discard data may not be available. However, in a management context, it is still important to consider the magnitude of discard mortality and whether it can be reduced. As another example, expert judgement is needed to decide if the CPUE can serve as index of abundance. Spanish mackerel and cobia are reported to be opportunistically caught by the recreational fleet, resulting in high percentages of zero catch (Bryan and Saul, 2012). This may degrade the quality of the CPUE as an index of abundance. Such uncertainties can be addressed through improved data collection programmes. In this case study, continued investment in fishery-independent surveys will produce a long-time series sufficient for inferring changes in mortality over time.

One must obtain length compositions from multiple years for application of the ML-based models used in this study. In this study, the recreational sector data were obtained from MRFSS (Marine Recreational Fisheries Statistics Survey) and its successor MRIP (Marine Recreational Information Program), which are design-based sampling programmes for the charter and private boat fleet, and from SRHS (Southeast Region Headboat Survey), which strives to be a census of all headboats in the region. We followed the decision of the assessment team in regards to combining or separating the data from these two programmes.

Data from multiple fleets or sectors could be combined if the fleets are believed to operate similarly temporally and spatially. Otherwise, mortality estimates can be confounded by the contrasting fishing effort and selectivity of the different fleets. For example, a multimodal length composition that arises from using two very different gears would not be easily accommodated by the assumptions of the ML-based models. Uncertainty in the composition data could be evaluated by comparing the length data from the different gear sectors separately. Differences in mortality estimates would be attributable to, among other factors, disparate selectivity patterns and sampling among gears. In these cases, mortality estimates are likely to have low precision (Pons *et al.*, 2019).

The MLeffort model provides year-specific mortality rates, but the fit to the mean lengths varies from good in the case of GOM greater amberjack to poor, as in the case of GOM king mackerel (Figure 3). For multispecies fisheries, nominal effort such as days fished may not be an indicator of targeted effort due to switches in targeting. As effort in the recreational fisheries examined here is not allocated on a species-specific basis, methods such as the so-called "guild" approach, where a subset of fishing trips that are believed to have targeted the stock of interest are identified based on catch of associated species, are used to develop indices for these fleets (SEDAR, 2011; Smith *et al.*, 2015). Poor estimates of recreational effort could have contributed to poor performance of the MLeffort model for GOM and ATL king mackerel. Formal statistical tests of model residuals, e.g. tests of normality or runs test, could be used to accept or reject a model.

Conclusion

The goal of this paper was to evaluate whether length-based methods could perform reasonably well and indicate when there are problems in the analysis. We did not intend to evaluate whether length-based methods could replace ASMs. Overall, ML-based methods can provide similar results, i.e. mortality trends and classifying overfishing status, as those of age-structured assessments. Such case studies have important ramifications for fishery managers who manage many stocks. Simple methods can be used to determine the overfishing status for stocks that are being assessed for the first time. If managers desire to use length-based methods, then such analyses can prompt allocation of more resources for data collection to improve mortality estimates. As a large majority of stocks worldwide do not and will not likely have fully age-structured assessments in the near future, fishery managers can use studies such as this in elucidating likely results from ML-based mortality estimators.

Supplementary data

Supplementary material is available at the *ICESJMS* online version of the manuscript.

Acknowledgements

We thank John Walter for his constructive comments on an earlier draft of this manuscript. We also thank the editor and two anonymous reviewers for their comments. This paper is Contribution No. 3849 of the Virginia Institute of Marine Science, William & Mary.

Funding

Q.C. Huynh was funded by a National Marine Fisheries Service (NMFS)/Sea Grant Population and Ecosystem Dynamics fellowship (NA15OAR4170184).

References

Bentley, N. 2015. Data and time poverty in fisheries estimation: potential approaches and solutions. ICES Journal of Marine Science, 72: 186–193.

Beverton, R. J. H., and Holt, S. J. 1956. A review of methods for estimating mortality rates in fish populations, with special reference to sources of bias in catch sampling. Rapports et Procès-verbaux des Reunions. Conséil International Pour L'Exploration de la Mer, 140: 67–83.

Bryan, M., and Saul, S. 2012. Recreational Indices for Cobia and Spanish Mackerel in the Gulf of Mexico. SEDAR28-DW-22. SEDAR, North Charleston, SC. 44 pp.

Burnham, K. P., and Anderson, D. R. 2002. Model Selection and Multimodel Inference: A Practical Information-theoretic Approach, 2nd edn. Springer-Verlag, New York, NY.

Chrysafi, A., and Kuparinen, A. 2016. Assessing abundance of populations with limited data: lessons learned from data-poor fisheries stock assessment. Environmental Reviews, 24: 25–38.

Clark, W. G. 1999. Effects of an erroneous natural mortality rate on a simple age-structured stock assessment. Canadian Journal of Fisheries and Aquatic Science, 56: 1721–1731.

Dichmont, C. M., Deng, R. A., Punt, A. E., Brodziak, J., Chang, Y.-J., Cope, J. M., Ianelli, J. N. *et al.* 2016. A review of stock assessment packages in the United States. Fisheries Research, 183: 447–460.

Dick, E. J., and MacCall, A. D. 2011. Depletion-based stock reduction analysis: a catch-based method for determining sustainable yields for data-poor fish stocks. Fisheries Research, 110: 331–341.

Francis, R. I. C. C. 2016. Growth in age-structured stock assessment models. Fisheries Research, 180: 77–86.

Gedamke, T., and Hoenig, J. M. 2006. Estimating mortality from mean length data in nonequilibrium situations, with application to the assessment of Goosefish. Transactions of the American Fisheries Society, 135: 476–487.

Gedamke, T., Hoenig, J. M., DuPaul, W., and Musick, J. A. 2008. Total mortality rates of the barndoor skate, *Dipturus laevis*, from the Gulf of Maine and Georges Bank, United States, 1963–2005. Fisheries Research, 89: 17–25.

Harford, W., Bryan, M., and Babcock, E. A. 2015. Probabilistic Assessment of Fishery Status Using Data-limited Methods. SEDAR46-DW-03. SEDAR, North Charleston, SC. 5 pp.

Hoenig, J. M. 1983. Empirical use of longevity data to estimate mortality rates. Fishery Bulletin, 82: 898–903.

Hordyk, A., Ono, K., Valencia, S., Loneragan, N., and Prince, J. 2015. A novel length-based empirical estimation method of spawning potential ratio (SPR), and tests of its performance, for small-scale, data-poor fisheries. ICES Journal of Marine Science, 72: 217–231.

Hordyk, A. R., Ono, K., Prince, J. D., and Walters, C. J. 2016. A simple length-structured model based on life history ratios and incorporating size-dependent selectivity: application to spawning potential ratios for data-poor stocks. Canadian Journal of Fisheries and Aquatic Science, 73: 1787–1799.

Huynh, Q. C. 2016. Estimating Total Mortality Rates and Calculating Overfishing Limits From Length Observations for Six U.S. Caribbean Stocks. SEDAR46-RW-01. SEDAR, North Charleston, SC. 19 pp.

Huynh, Q. C. 2018. MLZ: Mean Length-based Estimators of Mortality. R Package Version 0.1.0. https://cran.r-project.org/package=MLZ

Huynh, Q. C., Beckensteiner, J., Carleton, L. M., Marcek, B. J., Nepal KC, V., Peterson, C. P., Wood, M. A. *et al.* 2018. Comparative performance of three length-based mortality estimators. Marine and Coastal Fisheries, 10: 298–313.

Huynh, Q. C., Gedamke, T., Porch, C. E., Hoenig, J. M., Walter, J. F., Bryan, M., and Brodziak, J. 2017. Estimating total mortality rates of mutton snapper from mean lengths and aggregate catch rates in a non-equilibrium situation. Transactions of the American Fisheries Society, 146: 803–815.

ICES. 2017. Report of the Benchmark Workshop on North Sea Stocks (WKNSEA), 14–18 March 2016, Copenhagen, Denmark. ICES CM 2016/ACOM: 37. 698 pp.

Kokkalis, A., Eikeset, A. M., Thygesen, U. H., Steingrund, P., and Andersen, K. H. 2017. Estimating uncertainty of data limited stock assessments. ICES Journal of Marine Science, 74: 69–77.

Kokkalis, A., Thygesen, U. H., Nielsen, A., and Andersen, K. H. 2015. Limits to the reliability of size-based fishing status estimation for data-poor stocks. Fisheries Research, 171: 4–11.

Linton, B. 2012. Methods for Estimating Shrimp Bycatch of Gulf of Mexico Spanish Mackerel and Cobia. SEDAR28-DW-06. SEDAR, North Charleston, SC. 14 pp.

Lombardi, L. 2014. Growth Models for King Mackerel From the South Atlantic and Gulf of Mexico. SEDAR, North Charleston, SC. SEDAR38-AW-01. 62 pp.

Lorenzen, K. 1996. The relationship between body weight and natural mortality in juvenile and adult fish: a comparison of natural ecosystems and aquaculture. Journal of Fish Biology, 49: 627–642.

Methot, Jr, R. D. , and Wetzel, C. R. 2013. Stock Synthesis: a biological and statistical framework for fish stock assessment and fishery management. Fisheries Research, 142: 86–99.

Murie, D. J., and Parkyn, D. C. 2008. Age, Growth and Sex Maturity of Greater Amberjack (*Seriola dumerili*) in the Gulf of Mexico. SEDAR33-RD-13. SEDAR, North Charleston, SC. 41 pp.

Nadon, M. O. 2017. Stock Assessment of the Coral Reef Fishes of Hawaii, 2016. U.S. Department of Commerce, NOAA Technical Memorandum, NOAA-TM-NMFS-PIFSC-60. 212 pp.

Pons, M., Kell, L., Rudd, M. B., Cope, J. M., and Fredou, F. L. 2019. Performance of length-based data-limited methods in a multi-fleet context: application to small tunas, mackerels, and bonitos in the Atlantic Ocean. ICES Journal of Marine Science, 76: 960–973.

R Core Team, 2017. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Sagarese, S. R., Walter, III, J. F. , Bryan, M. D., and Carruthers, T. R. 2016. Evaluating methods for setting catch limits for gag grouper: data-rich versus data-limited. *In* Assessing and Managing Data-limited Fish Stocks. Ed. by T. J. Quinn , J. L. Armstrong, M. R. Baker, J. D. Heifetz, and D. Witherell. Alaska Sea Grant, University of Alaska Fairbanks, Alaska, USA.

SEDAR. 2011. SEDAR 9: Stock Assessment Update Report, Gulf of Mexico Greater Amberjack. SEDAR, North Charleston, SC.

SEDAR. 2013a. SEDAR 28: Gulf of Mexico Cobia Stock Assessment Report. SEDAR, North Charleston, SC.

SEDAR. 2013b. SEDAR 28: Gulf of Mexico Spanish Mackerel Stock Assessment Report. SEDAR, North Charleston, SC.

SEDAR. 2013c. SEDAR 28: South Atlantic Cobia Stock Assessment Report. SEDAR, North Charleston, SC.

SEDAR. 2014a. SEDAR 33: Gulf of Mexico Greater Amberjack Stock Assessment Report. SEDAR, North Charleston, SC.

SEDAR. 2014b. SEDAR 38: Gulf of Mexico King Mackerel Stock Assessment Report. SEDAR, North Charleston, SC.

SEDAR. 2014c. SEDAR 38: South Atlantic King Mackerel Stock Assessment Report. SEDAR, North Charleston, SC.

Siegfried, K. I., Williams, E. H., Shertzer, K. W., and Coggins, L. G. 2016. Improving stock assessments through data prioritization. Canadian Journal of Fisheries and Aquatic Science, 73: 1703–1711.

Smith, M. W., Goethel, D., Rios, A., and Isley, J. 2015. Standardized Catch Rate Indices for Gulf of Mexico Gray Triggerfish (*Balistes capriscus*) Landed During 1986–2013 by the Headboat Fishery. SEDAR43-WP-06. SEDAR, North Charleston, SC. 18 pp.

Then, A. Y., Hoenig, J. M., Gedamke, T., and Ault, J. S. 2015. Comparison of two length-based estimators of total mortality: a simulation approach. Transactions of the American Fisheries Society, 144: 1206–1219.

Then, A. Y., Hoenig, J. M., and Huynh, Q. C. 2018. Estimating fishing and natural mortality rates, and catchability coefficient, from a series of observations on mean length and fishing effort. ICES Journal of Marine Science, 75: 610–620.

Williams, E. H., and Shertzer, K. W. 2015. Technical Documentation of the Beaufort Assessment Model (BAM). NOAA Technical Memorandum NMFS-SEFSC-671, U.S. Department of Commerce. 43 pp.

Handling editor: Jan Jaap Poos