

Improving assessment of *Pandalus* stocks using a seasonal, size-structured assessment model with environmental variables.

Part II: Model evaluation and simulation

Jie Cao, Yong Chen, and R. Anne Richards

Abstract: Integrated, size-structured stock assessment models are now being used widely for assessment and management of hard-to-age species. However, few studies have attempted to evaluate their performance. A seasonal, size-structured assessment model with environmental covariates has been developed for hermaphroditic Pandalidae. We conducted simulations to evaluate its sensitivity to model configuration and performance with various misspecifications. Ignoring the seasonal fishing pattern (half-year closure) led to risk-prone assessment results of overestimating spawning stock biomass (SSB) and recruitment (R) and underestimating fishing mortality (F). Failure to incorporate environmental signals when the recruitment dynamics was environmentally driven led to bias in recent estimates of SSB, R , and F in the simulation. Ignoring annual variability in growth resulted in large estimation bias. Failing to account for time-varying natural mortality (M) led to strong biases; however, misspecifying size-specific M produced even stronger estimation bias. This result may depend on the variation of M among size classes. Under no model misspecifications, an unbiased estimate of M could be obtained by taking advantage of the seasonal fishery closure. Annual growth parameters were also estimable, but the large number of parameters with annual growth made it difficult for the model to converge.

Résumé : Si des modèles intégrés d'évaluation des stocks structurés selon la taille sont maintenant largement utilisés pour l'évaluation et la gestion d'espèces dont l'âge est difficile à déterminer, peu d'études ont toutefois tenté d'en évaluer la performance. Un modèle d'évaluation saisonnier structuré selon la taille intégrant des covariables environnementales a été développé pour les pandalidés hermaphrodites. Nous avons effectué des simulations pour évaluer la sensibilité de ce modèle à sa configuration, ainsi que sa performance étant donnés divers paramètres erronés. La non-prise en compte du caractère saisonnier de la pêche (fermeture durant la moitié de l'année) menait à des résultats risquant de se traduire par une surestimation de la biomasse du stock reproducteur (BSR) et du recrutement (R) et la sous-estimation de la mortalité par pêche (F). La non-incorporation de signaux environnementaux dans les cas où la dynamique du recrutement était modulée par des facteurs environnementaux menait à un biais dans les estimations récentes de la BSR, du R et de la F dans la simulation. La non-prise en compte de la variabilité annuelle de la croissance se traduisait par un fort biais d'estimation. La non-prise en compte de la mortalité naturelle variable dans le temps (M) menait à de forts biais; cependant, l'intégration d'une M selon la taille erronée produisait un biais d'estimation encore plus grand. Ce résultat pourrait dépendre de la variation de M selon les classes de taille. En l'absence de paramètres erronés du modèle, une estimation non biaisée de M pouvait être obtenue en tirant parti de la fermeture saisonnière de la pêche. Les paramètres de croissance annuelle pouvaient également être estimés, mais le grand nombre de paramètres découlant du fait d'intégrer la croissance annuelle faisait en sorte qu'il était difficile pour le modèle de converger. [Traduit par la Rédaction]

Introduction

Integrated size-structured stock assessment models are now being used widely for assessment and management of hard-to-age species (Punt et al. 2013b). These include lobsters (Chen et al. 2005; ASMFC 2009; Punt et al. 2013a), crabs (Zheng and Siddeek 2011; Turnock and Rugulo 2009), sea scallops (NEFSC 2004, 2010), and shrimps (Fu and Quinn 2000; Drouineau et al. 2012). Size-structured models are preferred because they can be directly fitted to length-structured data, overcoming the length-to-age conversion problem (Punt et al. 2013b). However, much less testing and evaluation has been done for performance of size-structured assessment models than for age-structured models. The effects of model misspecification have been extensively studied for age-structured

models over the past two decades, focusing especially on catchability (Wilberg et al. 2009; Thorson 2011), selectivity (Linton and Bence 2011; Sampson and Scott 2012), and natural mortality (Punt et al. 2013a; Deroba and Schueller 2013; Johnson et al. 2015).

Similar studies are needed to evaluate the performance of size-structured models and how they behave under model misspecification. A seasonal, size-structured assessment model for hermaphroditic pandalid shrimp was developed to improve pandalid stock assessment (Cao et al. 2017). The model is designed to be flexible with an aim to incorporate biological realism. The built-in complexity of the model makes it possible to account for some important features of pandalid biology and fisheries: (i) sex change, which has important implications for fishery management (Fu et al. 2001); (ii) vulnerability to environmental changes

Received 15 January 2016. Accepted 19 July 2016.

J. Cao and Y. Chen. School of Marine Sciences, University of Maine, Orono, ME 04469, USA.

R.A. Richards. Northeast Fisheries Science Center, 166 Water St., Woods Hole, MA 02543, USA.

Corresponding author: Jie Cao (email: jie.cao@maine.edu).

Yong Chen currently serves as an Editor; peer review and editorial decisions regarding this manuscript were handled by co-Editor Keith Tierney.

Copyright remains with the author(s) or their institution(s). Permission for reuse (free in most cases) can be obtained from [RightsLink](https://www.nrcresearchpress.com/cjfas).

(i.e., the effect of temperature on recruitment dynamics; Richards et al. 2012); and (iii) seasonality in growth patterns (Haynes and Wigley 1969; Pauly and Gaschutz 1978; Bergström 1992) and fisheries (NEFSC 2014). These features cannot be explicitly taken into account by traditional assessment models applied to pandalid stocks, such as Catch–Survey–Analysis (Cadrin et al. 1999; Cadrin 2000) and biomass dynamic models (Cadrin et al. 2004). However, the added model complexity requires more input data, and the model's behavior might be more difficult to understand (Geromont and Butterworth 2015). It is therefore worthwhile to evaluate how much model performance would be enhanced by explicitly accounting for these features and conversely the consequences of ignoring these features. Also, the robustness of the more complex model to misspecification needs to be evaluated. For instance, it is important to understand how the model would perform if the environmental signal is noisy or even wrong or if interannual variation in growth is ignored. Exploring such questions using simulation studies helps define and understand the trade-offs between model complexity and simplicity.

Individual fish experience natural mortality (M) that varies over the life span and can also be highly variable over time (Lorenzen 1996; Chu et al. 2008; McCoy and Gillooly 2008). However, fisheries stock assessments often assume that M is known and constant for all ages or sizes and across time (Johnson et al. 2015). The primary reason is that M is often confounded with other parameters in the assessment model (e.g., catchability, fishing mortality, and recruitment), which makes it very difficult, if not impossible, to independently estimate M within a stock assessment model (Quinn and Deriso 1999). Nevertheless, M may be estimable within a stock assessment model under certain circumstances (e.g., with appropriate data or no misspecification in the model; Fu and Quinn 2000; Lee et al. 2011). Seasonal periods with no fishing mortality is a characteristic of many pandalid fisheries in the North Atlantic, due either to ice cover or fishery regulations. It is possible to take advantage of this to estimate M (Xu et al. 1995; Drouineau et al. 2012); however, the accuracy and precision of the estimate need to be evaluated. Misspecifying time-varying M when it is present would lead to biases in stock assessment outcomes (Punt et al. 2013b; Johnson et al. 2015), while ignoring age-specific M may result in smaller biases compared with ignoring time-varying M (Deroba and Schueller 2013). However, whether these results can be generalized to size-structured models remains unclear, and further study is needed.

The growth model underlying a size-structured assessment model is important because it determines the transition between size bins and calculates the expected length composition (Chen et al. 2003). Thus, specification of growth is particularly important for size-structured assessment models that fit to length composition data directly. Like other parameters in the model, growth parameters tend to be time-dependent (Maunder et al. 2016). Temporal variation in growth of pandalid stocks can be manifested seasonally (Pauly and Gaschutz 1978; Bergström 1992) and (or) annually because of environmental changes (Apollonio et al. 1986; Bergström 1992). However, this temporal variability in growth is rarely accounted for in fisheries stock assessment models. More importantly, the consequences of misspecifying time-varying growth within size-structured models have only rarely been examined (Maunder et al. 2016).

There is growing interest in including environmental covariates in the assessment model (Maunder and Starr 2001; Maunder and Watters 2003; Schirripa et al. 2009; Methot and Wetzel 2013) as we recognize the importance of environmentally induced changes in abundance, such as environmentally driven recruitment found in pandalid stocks (Richards et al. 2012). However, there have been debates about the benefits of integrating environmental data into stock assessments (Hilborn 2003; Schirripa et al. 2009; Haltuch and Punt 2011). In the newly developed size-structured model for pandalid stocks, environmental covariates

were added into the recruitment model following Schirripa et al. (2009). This allows environmental signals to inform recruitment estimates. However, the performance of size-structured models with environmental covariates added to recruitment dynamics needs to be evaluated.

In this paper we extend the development of our size-structured model for pandalid stocks (Cao et al. 2017) by using simulation to evaluate model performance and robustness under different scenarios of model complexity and parameter misspecification. The goals of this study are to (i) evaluate the performance of the size-structure model in the face of growth and M misspecifications; (ii) evaluate the reliability of M estimated within the model by taking advantage of the seasonal fishing pattern; and (iii) demonstrate the importance of accounting for features such as seasonality and environmental effects on recruitment.

Materials and methods

Overview of size-structured model and simulation framework

A seasonal, size-structured assessment model was developed for hermaphroditic Pandalidae in a separate study (Cao et al. 2017). It has a seasonal time step that accounts for seasonal variations in biological processes and fishing patterns and incorporates sub-models for changes in length at sex change and environmental effects on recruitment dynamics. Simulated true population dynamics regarding these features were constructed (operating model; OM). Monte Carlo simulations were used to evaluate the performance of the Cao et al. (2017) size-structured model under various misspecifications regarding temporal fishing pattern, growth, recruitment, and natural mortality (estimation methods; EM). A scenario defined in this study refers to a combination of OM and EM, and each scenario was tested using the following steps: (i) simulate “true” population dynamics according to the OM and generate data with observation errors, (ii) apply the EM to the data simulated from the OM, and (iii) compare estimates of spawning stock biomass (SSB), recruitment (R), and fishing mortality (F) from the EM with their true values defined in the OM. These steps were repeated 200 times to capture the variability.

True underlying dynamics

The OM is formulated identically to the size-structured model developed by Cao et al. (2017). The underlying population was simulated to represent the northern shrimp stock and fishery in the Gulf of Maine (GOM) (Table 1). The reasons for selecting this stock are (i) northern shrimp in the GOM is at the southern limit of the species' boreal distribution, and recruitment dynamics here are partially regulated by temperature (Richards et al. 2012); (ii) the northern shrimp fishery in the GOM has extensive seasonal closures (6 months or more) with large interannual variation in landings; (iii) temporal variations in growth are seen in this stock (NEFSC 2014). These properties have not been accounted for in previous assessments (NEFSC 2014). Five configurations of the OM were considered in this study. They differ in patterns of temporal growth, recruitment dynamics, and natural mortality (Table 2).

The OM included two seasons, with two fisheries operating during the first (December to the following May) open season and the fisheries closed during the second. Both fisheries had asymptotic selectivity following logistic functions of length (Table 1), and fishing mortality followed the same temporal pattern in each simulation, based on actual estimated rates from the assessment (ASMFC 2009). Growth was modeled with potential seasonal and interannual variations, and the growth model was used to define the size-transition matrices that were applied in a given year for each season. For a given season and year, the expected growth increments followed a von Bertalanffy growth model. The probabilities of a shrimp growing from one size class to others (i.e., size-transition matrix) are determined by the five parameters of

Table 1. Summary of information used in the operating model (OM).

Item	Descriptor	Note
Years covered	1984–2013	
Seasons	2	December–May and June–November
No. of sexes	2	Female or non-female
Lengths	10–35 mm	
Length bins	1 mm	
Commercial fleets	2	Operated only in the first season
Commercial selectivity at length	Fleet 1	Logistic ($k = 0.32$; $L_{50} = 22.4$)
	Fleet 2	Logistic ($k = 0.22$; $L_{50} = 20.4$)
Fishing mortality	Instantaneous rates	Estimates from stock assessment report of ASMFC (2009)
Surveys	2	Spring (May) and fall (October)
Survey selectivity at length	Survey 1	Logistic ($k = 0.32$; $L_{50} = 22.4$)
	Survey 2	Logistic ($k = 0.22$; $L_{50} = 20.4$)
Natural mortality	Constant	0.25 per season
	Time-varying	NEFSC 2014 (Fig. 1a)
	Size-specific	NEFSC 2014 (Fig. 1b)
Size at sex change	Proportion female at size	Constant across years
Weight at size	Exponential function	$\ln(\text{weight}) = -7.4302 + 3.007(\text{length})$ from Haynes and Wigley (1969)
Recruitment dynamics	Environmentally driven recruitment	Two ocean temperature series at lags were used to derive recruitment deviations (Richards et al. 2012)
	No functional relationship with SSB	
Recruitment lengths	10–12 mm	$\lambda_k = c(0.8, 0.15, 0.05)$
Growth	Seasonal growth	Season 1: $L_{\infty} = 35$; $K = 0.1$, $SD(L_{\infty}) = 0.2$; $SD(K) = 0.1$; $\rho = -0.9$
		Season 2: $L_{\infty} = 36$; $K = 0.2$, $SD(L_{\infty}) = 0.2$; $SD(K) = 0.1$; $\rho = -0.9$
		$L_{\infty} = 35$; $K = 0.1$, $SD(L_{\infty}) = 0.2$; $SD(K) = 0.1$; $\rho = -0.9$
	Constant seasonal growth	Random walk, $y_t = y_{t-1} + \varepsilon_t$, y is the parameter, the mean of ε is 0, and the standard deviation of ε is 1 for L_{∞} and 0.01 for other four parameters; y_1 is given the above values
	Interannual variation	

Table 2. Simulation scenarios considered in this study.

Scenario	Operating model (OM)						Estimation method (EM)					
	Seasonal fishing pattern	Seasonal growth	Interannual variation in growth	Environmental driven R	Time-varying M	Size-specific M	Seasonal fishing pattern	Seasonal growth	Interannual variation in growth	Environmental driven R	Time-varying M	Size-specific M
1	✓	—	—	—	—	—	×	—	—	—	—	—
2	✓	—	✓	—	✓	✓	✓	—	✓	—	✓	✓
3	✓	—	✓	—	✓	✓	✓	—	—	—	✓	✓
4	✓	✓	✓	—	✓	✓	✓	✓	—	—	✓	✓
5	✓	✓	✓	—	✓	✓	✓	—	—	—	✓	✓
6	✓	✓	✓	—	✓	✓	✓	—	✓	—	✓	✓
7	✓	✓	—	—	✓	✓	✓	—	—	—	✓	✓
8	✓	✓	—	—	✓	✓	✓	✓	—	—	—	✓
9	✓	✓	—	—	✓	✓	✓	✓	—	—	×	✓
10	✓	✓	—	—	✓	✓	✓	✓	—	—	✓	—
11	✓	✓	—	—	✓	✓	✓	✓	—	—	—	—
12	✓	✓	—	—	✓	✓	✓	✓	—	—	✓	✓
13	✓	✓	—	✓	✓	✓	✓	✓	—	✓	✓	✓
14	✓	✓	—	✓	✓	✓	✓	✓	—	—	✓	✓
15	✓	✓	—	✓	✓	✓	✓	✓	—	✓*	✓	✓
16	✓	✓	—	✓	✓	✓	✓	—	—	—	—	—
17	✓	✓	—	✓	✓	✓	✓	✓	—	—	—	—
18	✓	✓	—	✓	✓	✓	✓	—	—	✓	—	—
19	✓	✓	—	✓	✓	✓	✓	—	—	—	✓	—
20	✓	✓	—	✓	✓	✓	✓	—	—	—	—	✓

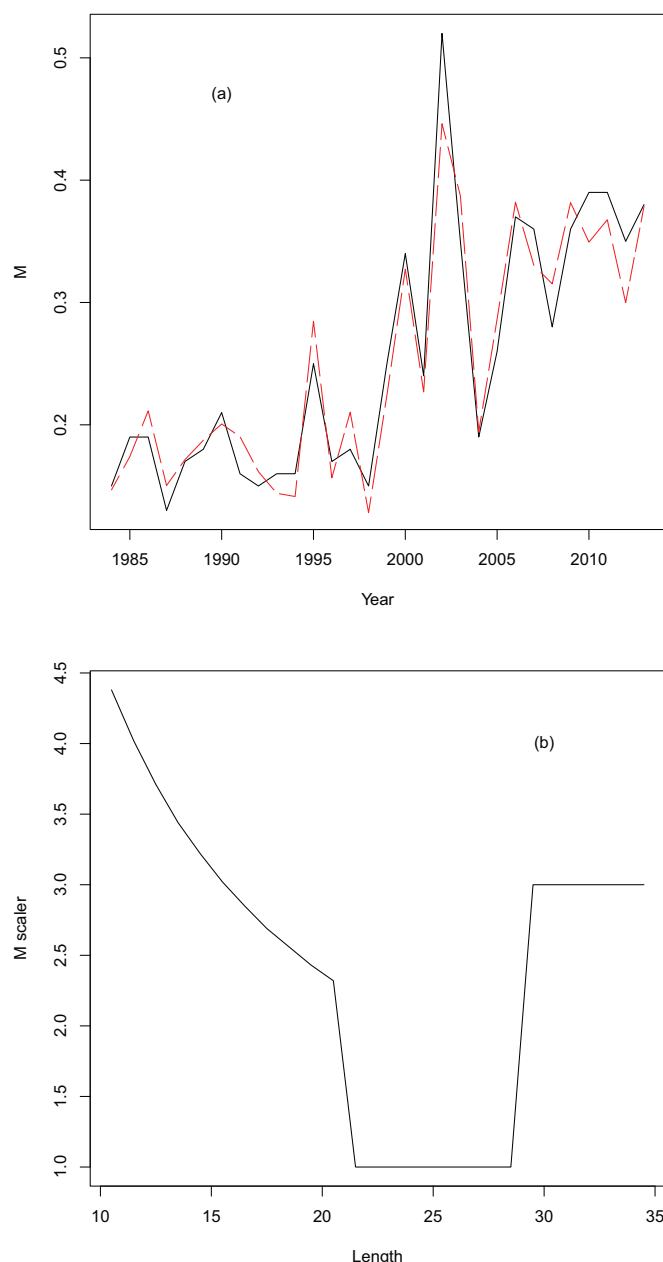
*The environmental data used in scenario 15 have large random noise (CV of 2.0).

von Bertalanffy model: L_{∞} , K , standard deviation of L_{∞} ($\sigma_{L_{\infty}}$), standard deviation of K (σ_K), and the correlation between L_{∞} and K (ρ) ([Chen et al. 2003](#)). Depending upon scenarios, L_{∞} and K either depend on season or not and either vary over years or not. When interannual variations occurred, these were modeled by independent variations in the five parameters following random walk models, and when these occurred in conjunction with seasonal differences, the same random walk perturbations applied to both the seasons (preserving the relative differences in growth parameters between seasons over years). Growth rates in the second season were assumed to be higher than those of the first season

because of warmer temperature ([Haynes and Wigley 1969](#); [Bergström 1992](#)).

The natural mortality was assumed separable into a year and size effect. Time-varying M was derived from an annual predation pressure index ([NEFSC 2014](#); [Fig. 1a](#)), which incorporated the occurrence of pandalids in fish stomachs and changes in predator biomass indices. Size-specific M was assumed to be U-shaped over the life span of the shrimp ([NEFSC 2014](#); [Fig. 1b](#)) rather than constant throughout the life history, declining with size in early life stages and increasing again at large size. Annual recruitments were determined by multiplying the predicted (mean) recruit-

Fig. 1. (a) Time-varying natural mortality (black line) and time-varying natural mortality containing random noise (dashed red line) and (b) U-shaped natural mortality scaler used in the operating model. [Colour online.]



ment by a time series of recruitment deviations from the mean recruitment. Environmentally driven recruitments were simulated by assuming that recruitment deviations were negatively related to an environmental signal, which was the mean of 1-year lagged spring bottom and surface temperature anomalies, estimated from NEFSC spring surveys in offshore northern shrimp habitat (Richards et al. 2012). To make the results comparable, the same recruitment series was used in the simulation for each scenario, except when environmentally driven recruitment was assumed. The environmentally driven recruitment dynamics did not vary among the corresponding scenarios.

Data generation

Data inputs to the size-structured assessment model were generated from the OM for 30 years (i.e., 1984–2013). Two time series

of fishery-independent survey abundance indices, which were assumed to occur in spring and fall each year, were simulated. Survey selectivity was assumed to be asymptotic following logistic functions of length (Table 1). The spring and fall surveys have different catchabilities (i.e., 0.01 and 0.02, respectively). The data simulated also included length composition of the catch from two commercial fishing fleets and length and sex composition from two fishery-independent surveys. Both fishery-dependent and fishery-independent length composition data were generated using a multinomial distribution with sample size of 100. The random errors added into the catch and survey indices had coefficients of variation (CV) of 0.2 and 0.15, respectively. Random errors were added to the environmental index used for tuning the recruitment deviations with CV of 0.2 and 2.0 depending upon scenarios.

Estimation methods and scenarios

The size-structured model was configured so that the combinations of OM and EM (i.e., scenarios) allowed us to examine the model performance under a variety of circumstances. Various EMs were explored, leading to a total of 20 scenarios (Table 2). All of the EMs were configured with the correct effective sample size for the multinomial samples (i.e., 100).

Scenario 1 was designed to examine the effects of ignoring the seasonal fishing pattern in the EM when it actually occurred in the OM (i.e., assessing a shrimp stock with only a seasonal fishery using an assessment model with an annual time step). Scenario 2 was used to evaluate the estimability of annual von Bertalanffy growth parameters within the size-structured model with no model misspecifications. Scenarios 3–7 were developed to evaluate the model performance with misspecified seasonal and (or) time-varying growth. Scenarios 8–11 were designed to investigate the effects of misspecified M (i.e., ignoring time-varying M and (or) size-specific M or incorrectly specifying the values of M). For scenario 9, we added a small amount of random noise to the time-varying M , which resulted in incorrectly specified values for M in the EM (Fig. 1a). Scenario 12 was used to evaluate the estimability of M within the size-structured model by taking advantage of the fact that there is no fishing activity in the second model season. Scenarios 13–15 were designed to demonstrate the benefits of linking environmental variables to recruitment within the model. The environmental data with small and large random noises (CV of 0.2 and 2.0) were used to fit the recruitment deviations in scenarios 13 and 15, respectively. Scenario 14 ignored the environmental effects on recruitments. Scenario 16 aimed to represent the real case of the northern shrimp stock in the GOM in which the time-varying growth, size-, and time-, and age-varying M s, and environmentally driven recruitment were not accounted for in the assessment. Scenarios 17–20 were used to investigate which features are more important to take into account in the size-structured model.

Model performance and output metrics

Estimation performance was evaluated using measures of relative error (RE) and average relative error (ARE), which are calculated as

$$(1) \quad RE_n = \frac{Est_n - Sim_n}{Sim_n}$$

$$(2) \quad ARE = \frac{1}{N} \sum_{n=1}^N \frac{Est_n - Sim_n}{Sim_n}$$

where Est_n is the point estimate of the quantity in year n from the EM, Sim_n is the simulated quantity for that year from the OM, and N is the number of years.

Table 3. Mean, median, and standard error (SD) of the percentage averaged relative error (ARE) for 200 estimation runs for estimated spawning stock biomass (SSB), recruitment (R), and fishing mortality (F) for the 20 scenarios under consideration.

Scenario	SSB			R			F		
	Mean	Median	SD	Mean	Median	SD	Mean	Median	SD
1	117.03	115.50	22.61	144.55	142.47	22.14	-48.09	-48.30	5.61
2	0.49	0.34	12.65	0.25	-0.47	9.30	4.27	2.73	16.44
3	-10.00	-12.28	14.79	-6.07	-8.19	15.34	229.67	207.01	111.61
4	-14.14	-14.51	7.96	-4.12	-3.99	6.63	29.16	28.22	15.55
5	-10.16	-10.06	9.09	0.18	0.00	7.32	21.16	18.39	15.98
6	14.25	13.16	14.83	11.42	9.81	11.64	-10.71	-10.32	12.66
7	-1.90	-4.10	12.28	-0.45	-1.29	8.66	17.71	17.34	18.89
8	-34.84	-37.52	11.06	13.28	12.11	11.39	176.32	173.10	70.37
9	4.51	3.68	10.32	4.42	3.88	7.52	-0.52	-0.89	11.36
10	-58.72	-58.78	2.33	-91.86	-91.84	0.30	304.09	302.51	34.39
11	-57.87	-57.91	2.51	-90.16	-90.16	0.41	366.00	363.22	38.20
12	3.43	2.54	11.55	4.15	3.42	8.20	-2.21	-2.25	12.98
13	4.40	2.93	12.17	3.44	3.36	8.65	-0.93	-1.20	12.91
14	12.98	11.56	12.93	19.26	18.70	9.99	-7.70	-8.13	12.20
15	2.62	1.87	10.99	2.96	2.66	8.11	0.34	-1.12	12.25
16	-56.35	-56.49	2.44	-89.86	-89.86	0.38	363.53	362.28	34.01
17	-58.12	-58.19	2.45	-90.21	-90.22	0.40	367.85	369.20	30.25
18	-56.26	-56.32	2.54	-89.83	-89.84	0.41	361.28	360.52	29.35
19	-57.85	-57.72	2.65	-91.66	-91.66	0.31	323.02	317.40	39.90
20	-33.64	-37.98	13.42	14.96	11.72	13.14	234.50	233.03	107.87

The ARE was used to measure the disparity between the OM and the EM for a particular simulation replicate. Boxplots were used to summarize the results for each year as indicated by RE. The quantities we considered in this study were SSB, R, and F. All the metrics were based on converged model results. The simulation procedure was repeated when a replicate failed to converge.

Results

Misspecifying seasonal fishing pattern

Scenario 1 was designed specifically for investigating the impact of misspecifying seasonal fishing patterns in the EM (i.e., annual time step was used in the EM) when the true fisheries were seasonal. The result showed large bias in estimates of SSB and R (i.e., overestimation of both for the entire model period; Table 3; Figs. 2–3). F was underestimated with median ARE of -48.3%. Compared with SSB and R, F was relatively robust to the misspecification of seasonal fishing pattern. However, such misspecification in the EM would lead to risk-prone assessment results. Therefore, it is important to take seasonal fishing pattern into account within the assessment model.

Estimates of annual growth parameters and M

There were no model misspecifications in scenarios 2 and 12 (Table 2), which were used to examine the reliability of estimated annual growth parameters and M, respectively. As expected, the estimates of SSB, R, and F were unbiased, as the REs and AREs were all centered at zero (Figs. 2–7). The histogram of M estimates in scenario 12 showed that out of 200 simulation runs, M was estimated without errors approximately 90% of the time (Fig. 8). This suggests that annual growth parameters and M are estimable in cases when the data are informative and there are no model misspecifications.

Misspecifying time-varying growth

Ignoring annually varying growth when the true growth varied annually but not seasonally (scenario 3) resulted in underestimation of SSB and R in the early years and overestimation in recent years (Figs. 2–3), while F was overestimated for the whole time period (Fig. 4). The AREs of scenarios 4, 5, and 6, in which the OM had growth varying both seasonally and annually, were relatively small (Table 3), which suggests that misspecification of seasonal

and (or) annual growth in these cases would not severely affect the assessment results. Scenario 4, in which annually varying growth was ignored, performed better than scenario 3, especially in the estimation of F. However, scenario 6, which included annual growth variation, generally produced an opposite bias pattern compared with scenarios 4 and 5. It was difficult to produce converged estimates for scenario 6 because of the large number of parameters (i.e., annual von Bertalanffy growth parameters). Ignoring seasonally varying growth (in a scenario with no interannual variation; i.e., scenario 7) affected the estimation much less than ignoring annually varying growth (in a scenario with no seasonal variation; i.e., scenario 3), suggesting that accounting for annually varying growth when it actually occurs is more important.

Misspecifying M

Ignoring time-varying M resulted in underestimated SSB and overestimated F (scenario 8; Table 3); the overestimation of F was significant in the recent decade (Fig. 4). The ARE of R was relatively small because it was averaged out by the reversed error pattern (overestimation in the early years and underestimation in recent years; Fig. 5). The EM performed well when time-varying M was taken into account but the values were slightly misspecified (scenario 9; Table 3). The EM produced large bias for SSB, R, and F when size-specific M was not accounted for (scenario 10; Table 3); SSB and R were greatly underestimated and F greatly overestimated for the whole model time period (Figs. 2–4). The bias pattern when both size- and time-varying M was ignored in the EM (scenario 11) was similar to scenario 10 in which only time-varying M was not accounted for in the EM. These comparisons indicate that ignoring size-specific M produced much larger bias in estimated SSB, R, and F than ignoring time-varying M.

Environmentally driven recruitment

Scenarios 13–15 investigated the benefits of incorporating environmental data within the size-structured model and evaluated the robustness of EMs with environmental inputs (Table 2). As expected, there was no estimation bias for SSB, R, and F in scenario 13 when informative environment data were used to tune recruitment estimates. However, when the environmental data were not used to fit the environmentally driven recruitment deviations (scenario 14), the REs of SSB and R in the recent years

Fig. 2. Relative error (RE) of spawning stock biomass (SSB) across years for the 20 scenarios considered in this study.

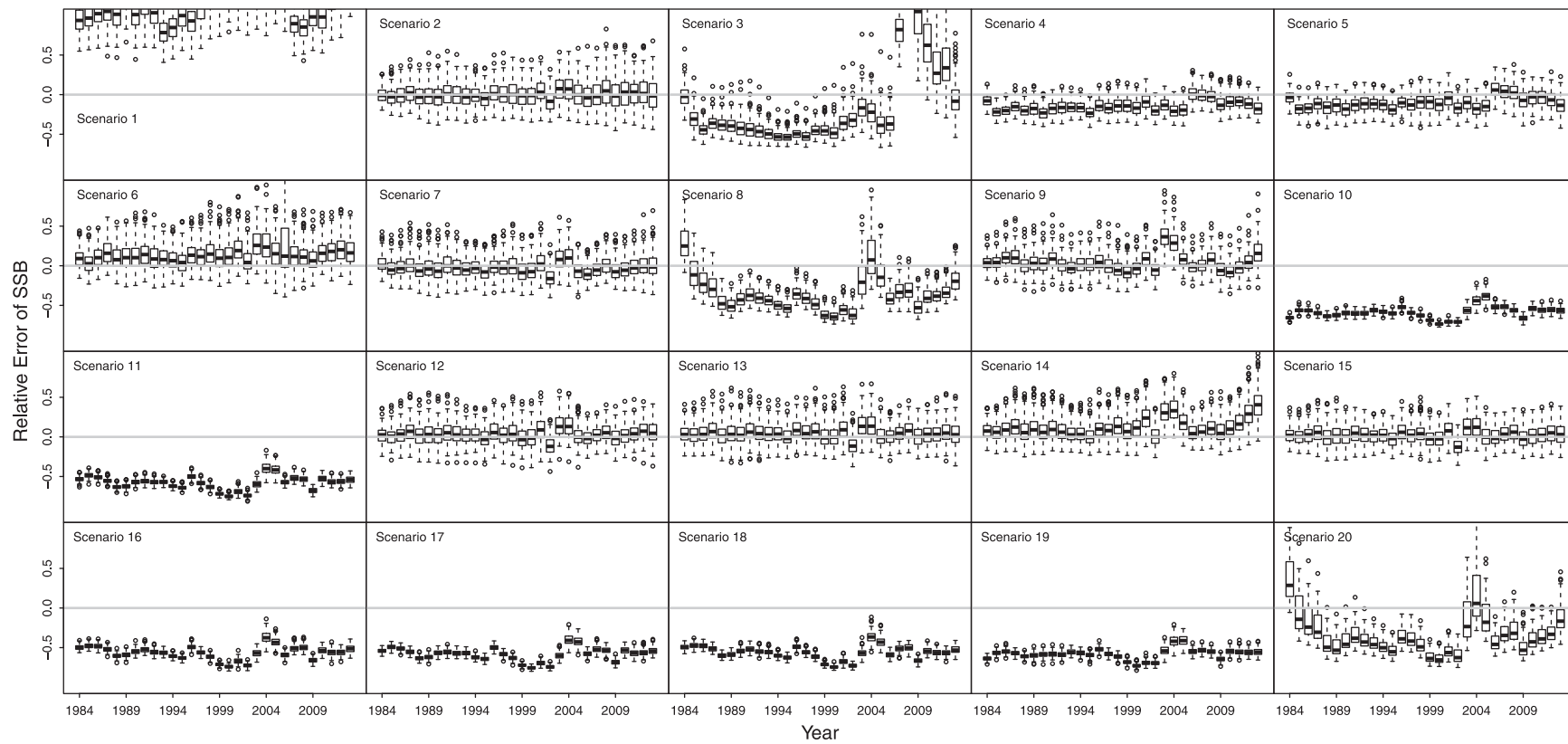


Fig. 3. Relative error (RE) of recruitment (R) across years for the 20 scenarios considered in this study.

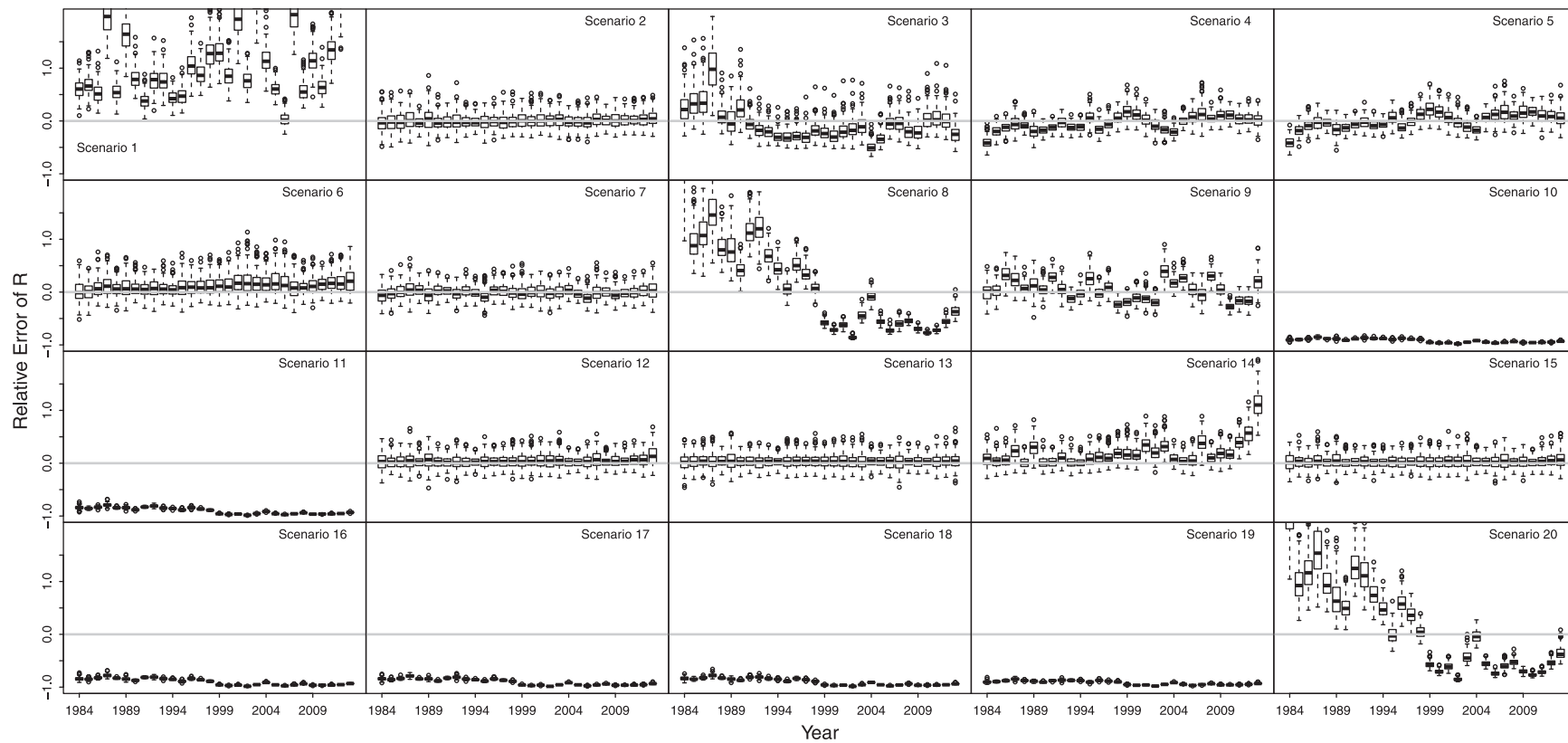


Fig. 4. Relative error (RE) of fishing mortality (F) across years for the 20 scenarios considered in this study.

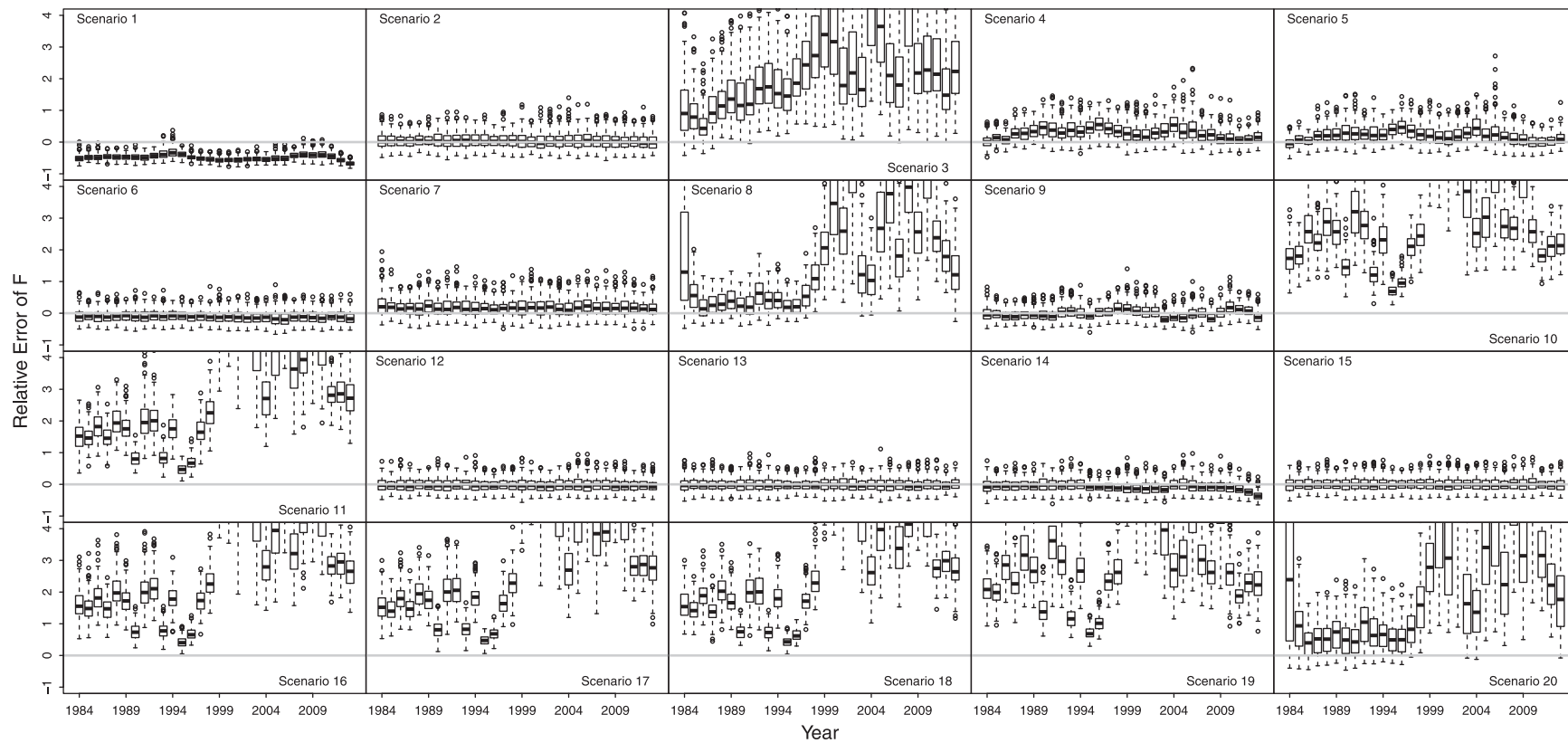


Fig. 5. Percent averaged relative error (ARE) of recruitment (R) for the 20 scenarios considered in this study.

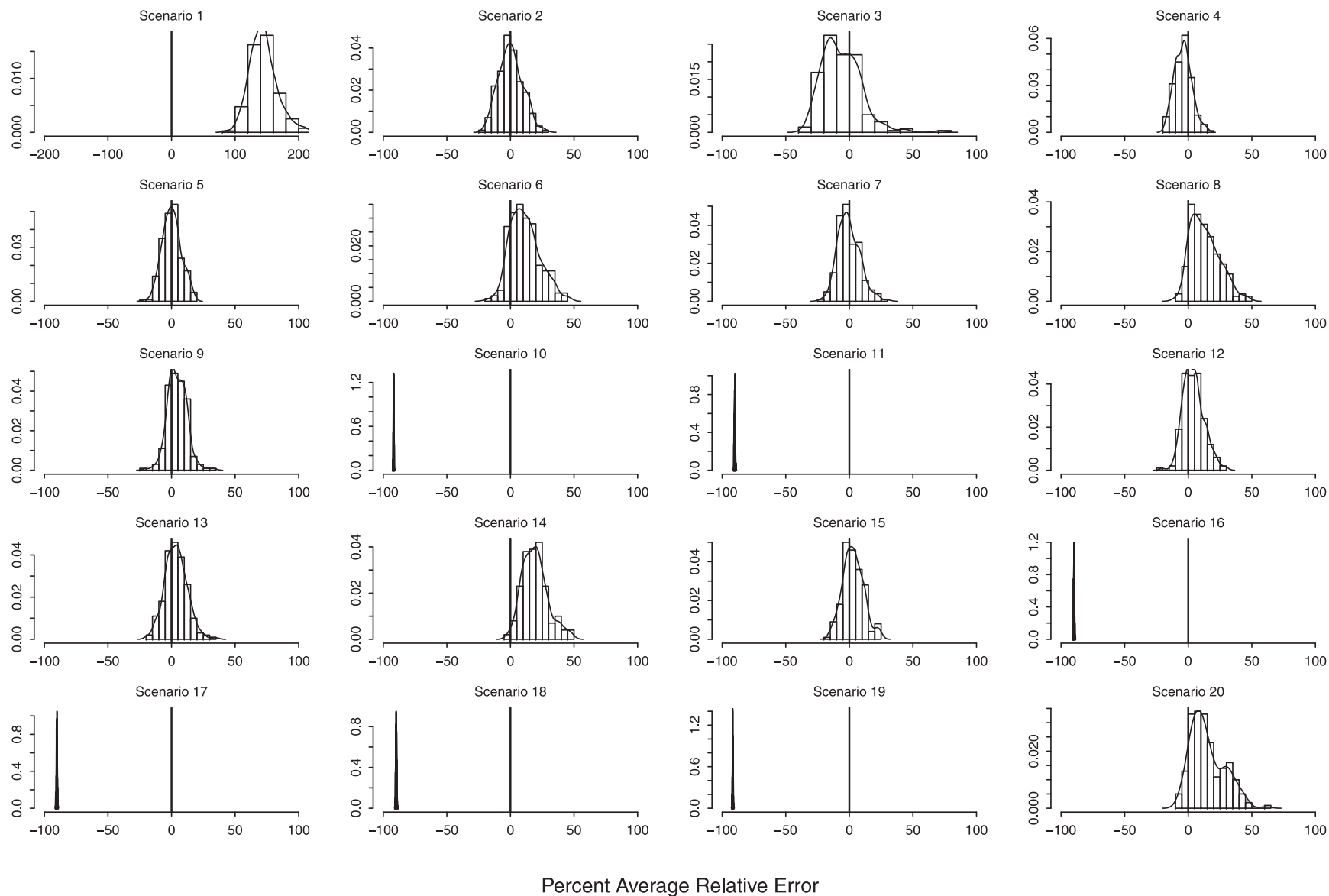


Fig. 6. Percent averaged relative error (ARE) of spawning stock biomass (SSB) for the 20 scenarios considered in this study.

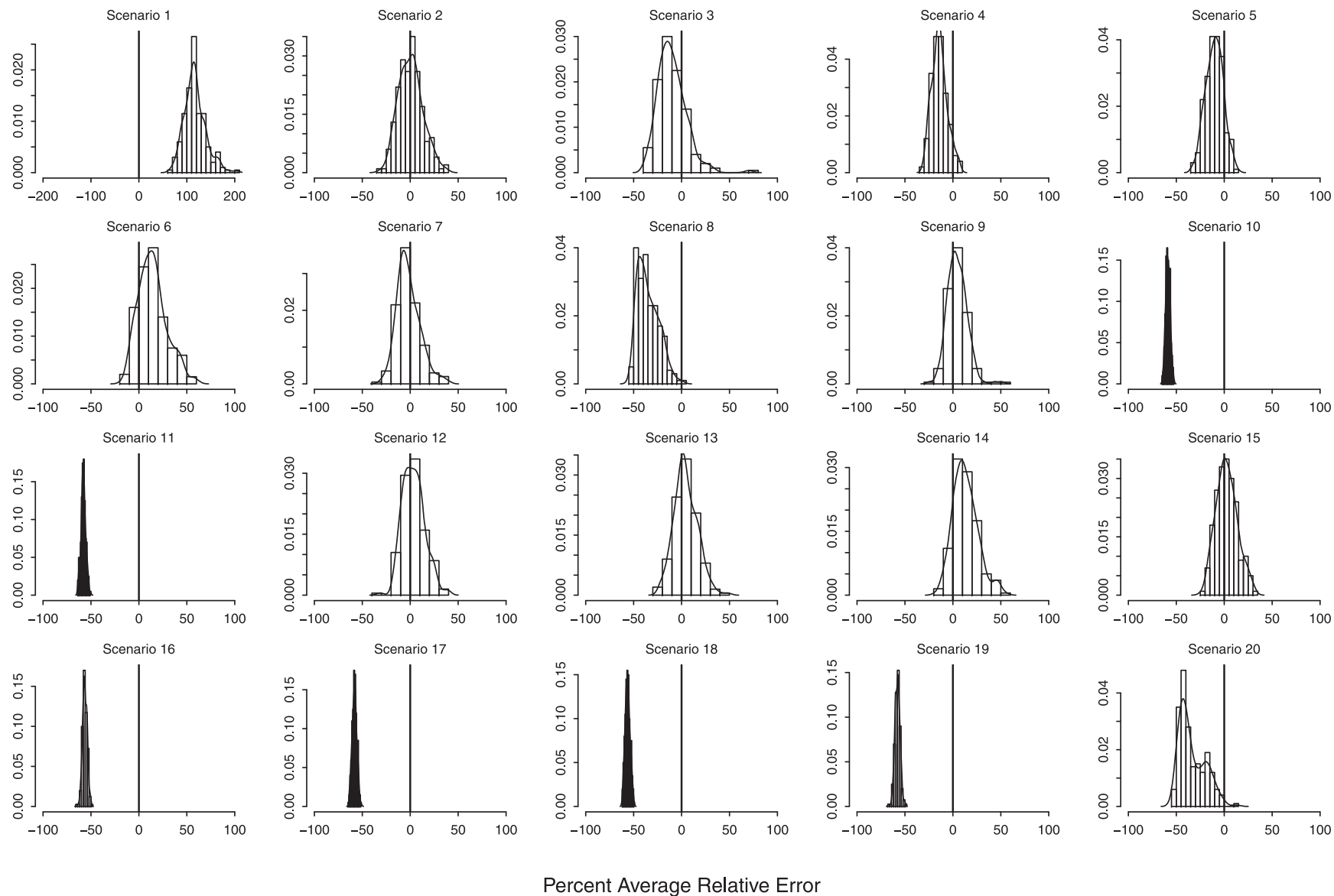


Fig. 7. Percent averaged relative error (ARE) of fishing mortality (F) for the 20 scenarios considered in this study.

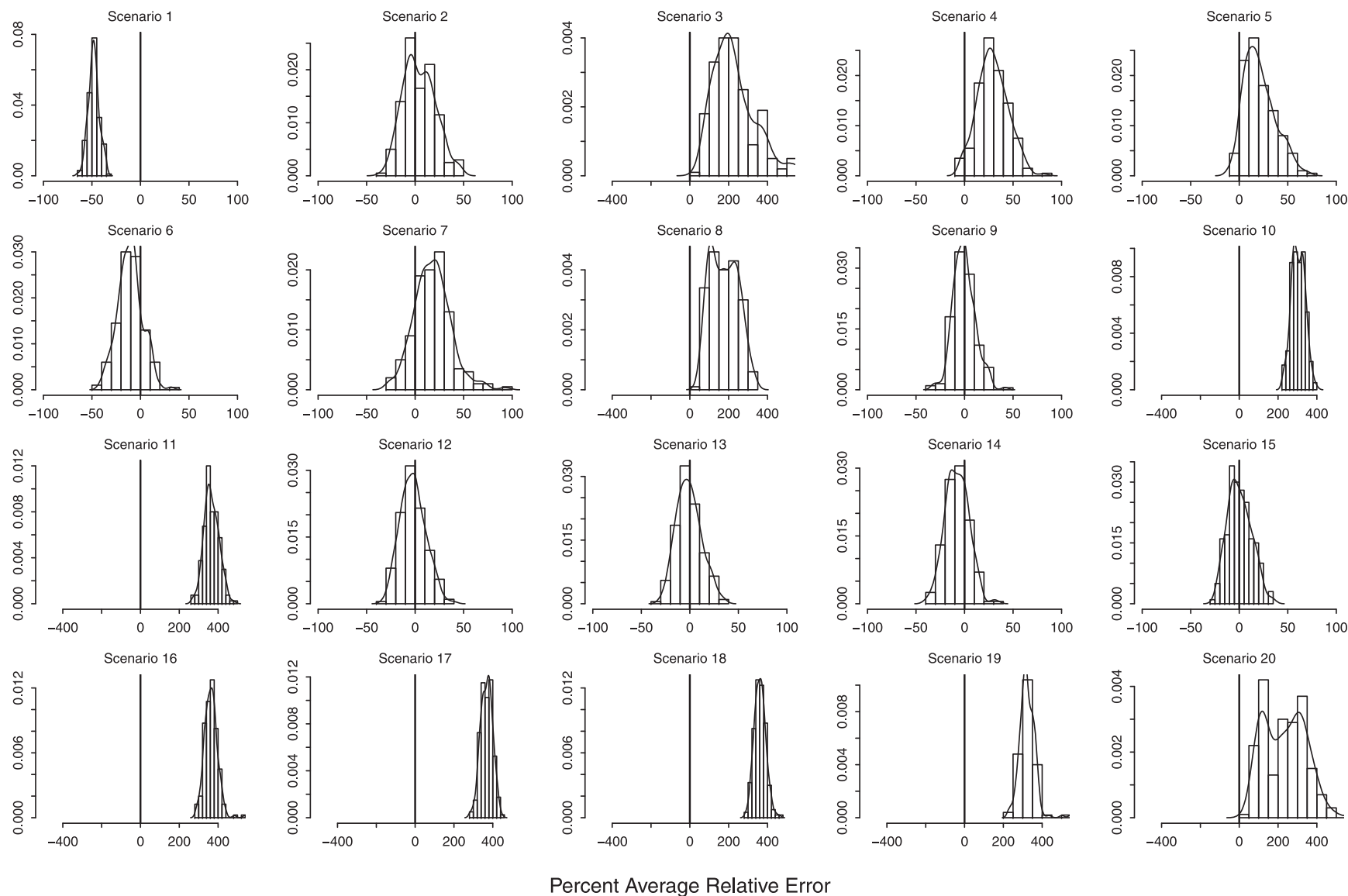
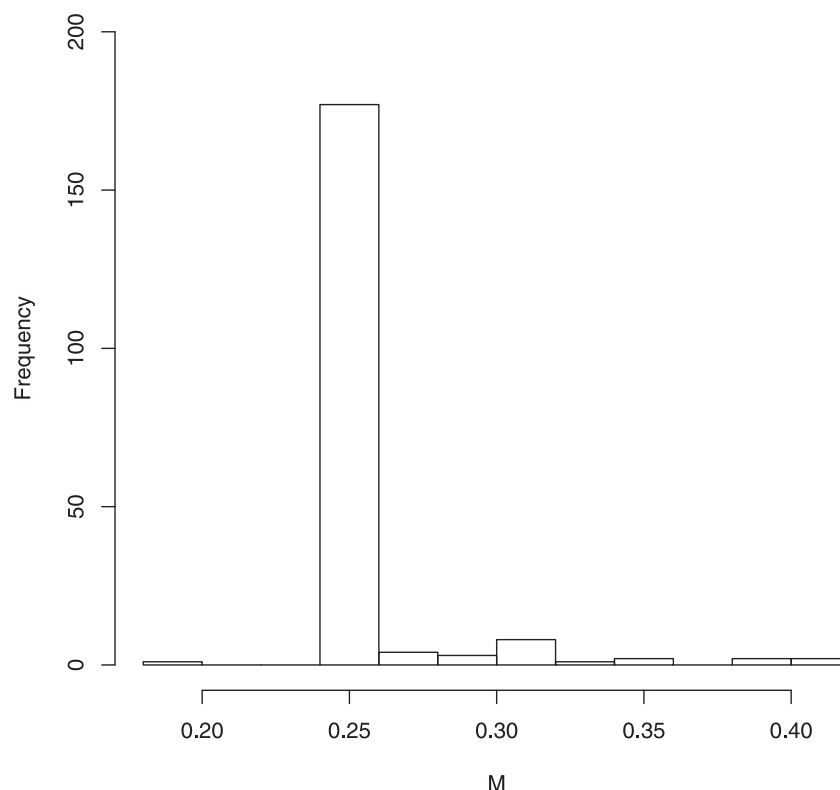


Fig. 8. Histogram of estimated natural mortality (M) from scenario 12.

were significantly above zero (Figs. 2 and 3), which indicates that they were overestimated. Also, F was slightly underestimated for the recent years in this scenario (Fig. 4). Overall, these results suggest that using environmental data can increase the accuracy of environmentally driven recruitments and other quantities within the model. When the environmental data contain large random noises (scenario 15), the EM still performed better than an EM with no environmental inputs (scenario 14; Table 3) when the OM had an environmental signal.

Comparison between misspecifications

The OM of scenarios 16–20 were the same (i.e., the underlying population had a seasonal fishing pattern, seasonal growth, environmentally driven recruitment, and size- and time-varying M). This OM most likely represents the real case (e.g., northern shrimp in the GOM). Comparing the results of these scenarios allows us to identify the importance of accounting for each feature in the EM. The results of scenarios 16–19 were more or less the same, and the bias pattern was similar to that of scenario 10 in which size-specific M was ignored in the EM. When size-specific M was accounted for in the EM (scenario 20), the bias pattern changed. This suggests that ignoring size-specific M with or without other misspecifications in the EM would lead to unique bias patterns of SSB, R , and F . Among all the scenarios considered in this study, scenario 1 (incorporating only a seasonal fishing pattern) produced the largest bias in estimation of SSB and R (Table 3), and scenarios with misspecified size-specific M produced the largest biases in estimation of F . These results indicate the importance of accounting for seasonality in the fishing pattern and size-specific M when the underlying population has these features.

Discussion

The model misspecifications that had the greatest impact in this study were model time step and M (Table 3). Results from the size-structured model evaluated in this study were sensitive to the model dimension (i.e., time step) when the fisheries showed strong seasonality. A broader time step (e.g., annual) led to risk-prone assessment results of overestimated SSB and R and underestimated F . One possible explanation for such an outcome has to do with the assumption of growth (Maunder et al. 2016). Size composition is assumed to be unchanged within a year for the annual time-step model; however, seasonal growth would alter the size composition within a year for the seasonal time-step model. The misspecified size composition could then result in biased estimation of selectivity. To our knowledge, there is no published study that specifically focuses on consideration of the time step in integrated fisheries stock assessment models. Finer time steps may increase biological realism, but increase model complexity and require more data, which can be obstacles. However, when a fishery is strongly seasonal, for example with seasonal closures, it is important to use a finer time step to adequately describe the fishing pattern.

The fishing mortality used in the simulation has large variation over time. To understand to what extend the results depend on the fishing history, we repeated all 20 scenarios assuming alternative fishing mortality in the simulation (i.e., relatively constant fishing pressure over time). The results indicate that the conclusions are fairly robust to the fishing mortality (see online supplementary material¹, Figs. S1–S6).

One of the most influential parameters in stock assessment models is M (Johnson et al. 2015). Usually it is fixed at a certain

¹Supplementary data are available with the article through the journal Web site at <http://nrcresearchpress.com/doi/suppl/10.1139/cjfas-2016-0021>.

value within the assessment model, and sensitivity analysis is conducted to examine the impact of different values of M on the assessment outcomes. Estimating M within the assessment model would be beneficial but is very difficult (Quinn and Deriso 1999). In this study, M could be estimated without errors by taking advantage of a seasonal fishery closure. This reduced the uncertainty brought about by assigning M values in the assessment model. However, the success on estimating M in this study might also depend on the fact that the model was given adequate samples (e.g., effective sample size) and initial values of parameters. Like many other studies (Fu and Quinn 2000; Deroba and Schueller 2013; Punt et al. 2013b; Johnson et al. 2015), our study also suggests that it is important to account for time-varying effects of M when present. However, in practicality, estimating time-varying M is expected to occur with uncertainty. Our results show that slightly misspecifying time-varying M would still produce better estimates than ignoring the time-varying effect (scenario 9). We also found that ignoring size-specific M led to more pronounced biases than ignoring time-varying M . This contrasts with Deroba and Schueller (2013), who found that it was less important to account for age-specific M within their assessment model compared with accounting for time-varying M . However, the variation of M among size classes in this study is much larger than the variation of M among age classes in Deroba and Schueller (2013). We therefore suspect that variation of M among age or size classes or years may largely determines how important it should be accounted for in the assessment. When the misspecifications involved size-specific M , the bias patterns were very similar (scenarios 16–19). This indicates that misspecifying size-specific M has great influence on the assessment results. Our results are based on U-shaped, size-dependent M , and the results undoubtedly would be different if a different M function was used.

Despite the fact that temporal variability in growth is widely acknowledged for many exploited fish species (Clark et al. 1999; Francis 1997; Arnekleiv et al. 2006; Feltrim and Ernst 2010), it is rarely accounted for in practical stock assessment models. Whitten et al. (2013) found that accounting for variable growth in the assessment model led to a better fit to the data and produced significantly different population estimates. However, the consequences of not considering temporal growth variation when it actually occurs are not well understood, especially for size-structured models where growth has a large impact on the transition of size classes. In the current study, the impacts of ignoring seasonal and (or) annual variation in growth on the assessment results were investigated through five scenarios (scenarios 3–7). Major biases were evident only when the underlying population showed interannual growth variation that was ignored (scenario 3). Ignoring seasonal growth in this study appeared to be not important. However, this conclusion might depend on the contrast between different seasons and number and duration of seasons. There are different ways to account for time-varying growth in the EM (Methot and Wetzel 2013). In this study, annual von Bertalanffy growth parameters were estimated to account for time-varying growth. This method resulted in a large number of parameters within the assessment model. The model had more difficulty converging in this case, but was able to estimate the parameters within the model (scenario 2). Alternative methods (e.g., random walk for growth deviation and time blocks for time periods with similar growth) are worth investigating further.

Integrated analysis models such as the one we present in this study are able to produce recruitment estimates without direct recruitment observations. Recruitment in the model is informed by other information that reflects individual year class strengths (e.g., length composition data in this case). Also, a penalized likelihood component is often added to the total likelihood to penalize log(recruitment) for deviating from zero (Maunder and Deriso 2003). Therefore, the estimation of recruitment is always a compromise between the information in the data and the penalty

term in the objective function (Methot and Taylor 2011). This can be problematic because the estimated recruitment variability is always less than the true variability among recruitments (Methot and Taylor 2011). When the recruitment deviations are caused by environmental factors, conventional fishery and survey data may not be able to fully capture the variabilities (Schirripa et al. 2009). In this case, inclusion of environmental data would be beneficial. Additionally, because there are no data available to inform the terminal year recruitment, the estimate usually contains larger uncertainty. In our study, the estimates of terminal year recruitments were positively biased in the absence of environmental inputs (scenario 14; Fig. 3). Environmental data could improve the estimates of recent recruitments, which are usually of most interest to fishery managers, when there is correlation between environmental variability and recruitment deviations. This is demonstrated in this study (comparing scenario 13 with 14; Fig. 3). The robustness of including environmental data was also examined. The results suggest that even when there was large noise in the environmental data, the model still produced better estimates (scenario 15).

Studies of model misspecification usually deliberately isolate a certain effect by correctly specifying all parameters in the model except one (e.g., M ; Deroba and Schueller 2013; Johnson et al. 2015). However, in actuality, more than one parameter is likely to be misspecified. In this study, we used scenarios that considered many potential sources of misspecification, including scenarios that misspecified more than one parameter. A previous study revealed that estimates could be unbiased when more than one parameter is misspecified in the model if the misspecification acts in countering ways (Tyler et al. 1989). In our study, all the scenarios with more than one misspecified parameter produced biased estimates.

Integrated size-structured stock assessment models have been applied to a variety of economically important species and form the basis for the management. However, model behavior and diagnostics remain very complicated, especially when misspecifications occur. Model misspecifications can be more serious in some circumstances than others as shown in our study. Therefore, simulation study should be conducted to better understand the model robustness to various misspecifications so that priorities can be identified for data collection and methods development.

Climate change is already having impacts on various aspects of marine species. Incorporating environmental covariates into the assessment model is not a new idea; however, it is rarely applied in practical situations. Inclusion of environmental data is beneficial for estimating recruitment in our case. It could also be helpful for forecasting of possible future stock conditions. In practical situations, when environmental covariates are added into the assessment model, then the appropriate management should also consider the environment changes.

Acknowledgements

Financial support for this study was provided by the Maine Sea Grant Program, the Maine Department of Marine Resources, and the NOAA FATE program. We thank Margaret Hunter and the Atlantic States Marine Fisheries Commission Northern Shrimp Technical Committee for informative discussions on the northern shrimp stock assessment and for providing the data on which this study is based. Comments from the two anonymous reviewers improved an early version of the manuscript, for which we are thankful.

References

- Apollonio, S., Stevenson, D., and Dunton, E.E., Jr. 1986. Effects of temperature on the biology of the northern shrimp, *Pandalus borealis*, in the Gulf of Maine. NOAA Tech. Rep. NMFS 42.
- Arnekleiv, J.V., Finstad, A.G., and Ronning, I. 2006. Temporal and spatial variation in growth of juvenile Atlantic salmon. J. Fish Biol. 68(4): 1062–1076. doi:10.1111/j.0022-1112.2006.00986.x.

- ASMFC. 2009. American Lobster Stock Assessment Report for Peer Review. Stock Assessment Report No. 09-01 (Supplement), Atlantic States Marine Fisheries Commission, Washington, D.C.
- Bergström, B. 1992. Growth, growth modelling and age determination of *Pandalus borealis*. Mar. Ecol. Prog. Ser. **83**(2): 167–183. doi:10.3354/meps083167.
- Cadrin, S.X. 2000. Evaluating two assessment methods for Gulf of Maine northern shrimp based on simulations. J. Northw. Atl. Fish. Sci. **27**: 119–132. doi:10.2960/j.v27.a11.
- Cadrin, S.X., Clark, S.H., Schick, D.F., Armstrong, M.P., McCarron, D., and Smith, B. 1999. Application of catch–survey models to the northern shrimp fishery in the Gulf of Maine. N. Am. J. Fish. Manage. **19**(2): 551–568. doi:10.1577/1548-8675(1999)019<0551:AOCST>2.0.CO;2.
- Cadrin, S.X., Boutillier, J.A., and Idoine, J.S. 2004. A hierarchical approach to determining reference points for Pandalid shrimp. Can. J. Fish. Aquat. Sci. **61**(8): 1373–1391. doi:10.1139/f04-133.
- Cao, J., Chen, Y., and Richards, R.A. 2017. Improving assessment of *Pandalus* stocks using a seasonal, size-structured assessment model with environmental variables. Part I: Model description and application. Can. J. Fish. Aquat. Sci. [In press.] doi:10.1139/cjfas-2016-0020.
- Chen, Y., Hunter, M., Vadas, R., and Beal, B. 2003. Developing a growth-transition matrix for the stock assessment of the green sea urchin (*Strongylocentrotus droebachiensis*) off Maine. Fish. Bull. **101**(4): 737–744.
- Chen, Y., Kanaiwa, M., and Wilson, C. 2005. Developing and evaluating a size-structured stock assessment model for the American lobster, *Homarus americanus*, fishery. N.Z. J. Mar. Freshw. Res. **39**: 645–660. doi:10.1080/00288330.2005.9517342.
- Chu, C.Y., Chien, H.K., and Lee, R.D. 2008. Explaining the optimality of U-shaped age-specific mortality. Theor. Popul. Biol. **73**: 171–180. doi:10.1016/j.tpb.2007.11.005. PMID:18178233.
- Clark, W.G., Hare, S.R., Parma, A.M., Sullivan, P.J., and Trumble, R.J. 1999. Decadal changes in growth and recruitment of Pacific halibut (*Hippoglossus stenolepis*). Can. J. Fish. Aquat. Sci. **56**(2): 242–252. doi:10.1139/f98-163.
- Deroba, J.J., and Schueller, A.M. 2013. Performance of stock assessments with misspecified age- and time-varying natural mortality. Fish. Res. **146**: 27–40. doi:10.1016/j.fishres.2013.03.015.
- Drouineau, H., Savard, L., Desgagnés, M., and Duplisea, D. 2012. SPAM (Sex-Structured *Pandalus* Assessment Model): a stock assessment model for *Pandalus* stocks. Can. J. Fish. Aquat. Sci. **69**(4): 770–783. doi:10.1139/f2012-011.
- Feltrim, M., and Ernst, B. 2010. Inter-cohort growth variability and its implication for fishery management of the common sardine (*Strangomera bentincki*) stock off the coast of south-central Chile. Fish. Res. **106**(3): 368–377. doi:10.1016/j.fishres.2010.09.006.
- Francis, M.P. 1997. Spatial and temporal variation in the growth rate of elephantfish (*Callorhynchus milii*). N.Z. J. Mar. Freshw. Res. **31**(1): 9–23. doi:10.1080/00288330.1997.9516741.
- Fu, C., and Quinn, T.J., II. 2000. Estimability of natural mortality and other population parameters in a length-based model: *Pandalus borealis* in Kachemak Bay, Alaska. Can. J. Fish. Aquat. Sci. **57**(12): 2420–2432. doi:10.1139/f00-220.
- Fu, C., Quinn, T.J., and Shirley, T.C. 2001. The role of sex change, growth and mortality in *Pandalus* population dynamics and management. ICES J. Mar. Sci. **58**(3): 607–621. doi:10.1006/jmsc.2001.1050.
- Geromont, H.F., and Butterworth, D.S. 2015. Complex assessments or simple management procedures for efficient fisheries management: a comparative study. ICES J. Mar. Sci. **72**(1): 262–274. doi:10.1093/icesjms/fsu017.
- Haltuch, M.A., and Punt, A.E. 2011. The promises and pitfalls of including decadal-scale climate forcing of recruitment in groundfish stock assessment. Can. J. Fish. Aquat. Sci. **68**(5): 912–926. doi:10.1139/f2011-030.
- Haynes, E.B., and Wigley, R.L. 1969. Biology of the northern shrimp, *Pandalus borealis*, in the Gulf of Maine. Trans. Am. Fish. Soc. **98**(1): 60–76. doi:10.1577/1548-8659(1969)98[60:BOTNSP]2.0.CO;2.
- Hilborn, R. 2003. The state of the art in stock assessment: Where we are and where we are going. Sci. Mar. **67**(S1): 15–20. doi:10.3989/scimar.2003.67s115.
- Johnson, K.F., Monnahan, C.C., McGilliard, C.R., Vert-pre, K.A., Anderson, S.C., Cunningham, C.J., et al. 2015. Time-varying natural mortality in fisheries stock assessment models: identifying a default approach. ICES J. Mar. Sci. **72**(1): 137–150. doi:10.1093/icesjms/fsu055.
- Lee, H-H, Maunder, M.N., Piner, K.R., and Methot, R.D., Jr. 2011. Estimating natural mortality within a fisheries stock assessment model: an evaluation using simulation analysis based on twelve stock assessments. Fish. Res. **109**: 89–94. doi:10.1016/j.fishres.2011.01.021.
- Linton, B.C., and Bence, J.R. 2011. Catch-at-age assessment in the face of time-varying selectivity. ICES J. Mar. Sci. **68**(3): 611–625. doi:10.1093/icesjms/fsq173.
- Lorenzen, K. 1996. The relationship between body weight and natural mortality in juvenile and adult fish: a comparison of natural ecosystems and aquaculture. J. Fish Biol. **49**: 627–642. doi:10.1111/j.1095-8649.1996.tb00060.x.
- Maunder, M.N., and Deriso, R.B. 2003. Estimation of recruitment in catch-at-age models. Can. J. Fish. Aquat. Sci. **60**(10): 1204–1216. doi:10.1139/f03-104.
- Maunder, M.N., and Starr, P.J. 2001. Bayesian assessment of the SNA1 snapper (*Pagrus auratus*) stock on the north-east coast of New Zealand. N.Z. J. Mar. Freshw. Res. **35**: 87–110. doi:10.1080/00288330.2001.9516980.
- Maunder, M.N., and Watters, G.M. 2003. A general framework for integrating environmental time series into stock assessment models: model description, simulation testing, and example. Fish. Bull. **101**: 89–99.
- Maunder, M.N., Crone, P.R., Valero, J.L., and Semmens, B.X. 2016. Growth: Theory, estimation, and application in fishery stock assessment models. Fish. Res. **180**: 1–3. doi:10.1016/j.fishres.2016.03.005.
- McCoy, M.W., and Gillooly, J.F. 2008. Predicting natural mortality rates of plants and animals. Ecol. Lett. **11**: 710–716. doi:10.1111/j.1461-0248.2008.01190.x. PMID:18422635.
- Methot, R.D., and Taylor, I.G. 2011. Adjusting for bias due to variability of estimated recruitments in fishery assessment models. Can. J. Fish. Aquat. Sci. **68**(10): 1744–1760. doi:10.1139/f2011-092.
- Methot, R.D., and Wetzel, C.R. 2013. Stock Synthesis: a biological and statistical framework for fish stock assessment and fishery management. Fish. Res. **142**: 86–99. doi:10.1016/j.fishres.2012.10.012.
- NEFSC. 2004. 39th Northeast Regional Stock Assessment Workshop (39th SAW) assessment summary report. US Dept. Commer., Northeast Fish. Sci. Cent. Ref. Doc. 04-10a. Available from National Marine Fisheries Service, 166 Water Street, Woods Hole, MA 02543-1026, USA.
- NEFSC. 2010. 50th Northeast Regional Stock Assessment Workshop (50th SAW) Assessment Report [online]. US Dept. Commer., Northeast Fish. Sci. Cent. Ref. Doc. 10-09. Available from <http://www.nefsc.noaa.gov/publications/crd/crd1017/> [accessed November 2015].
- NEFSC. 2014. 58th Northeast Regional Stock Assessment Workshop (58th SAW) Assessment Report. US Dept. Commer., Northeast Fish. Sci. Cent. Ref. Doc. 14-04. Available from National Marine Fisheries Service, 166 Water Street, Woods Hole, MA 02543-1026, USA, or online at <http://nefsc.noaa.gov/publications/> [accessed November 2015].
- Pauly, D., Gaschutz, G. 1978. A simple method of fitting oscillating length growth data with a programme for pocket calculations. Comm. Meet. Int. Coun. Explor. Sea C.M.-ICES/G: 24.
- Punt, A.E., Trinnie, F., Walker, T.I., McGarvey, R., Feenstra, J., Linnane, A., and Hartmann, K. 2013a. The performance of a management procedure for rock lobsters, *Jasus edwardsii*, off western Victoria, Australia in the face of non-stationary dynamics. Freshw. Res. **137**: 116–128.
- Punt, A.E., Huang, T., and Maunder, M.N. 2013b. Review of integrated size-structured models for stock assessment of hard-to-age crustacean and mollusc species. ICES J. Mar. Sci. **70**(1): 16–33. doi:10.1093/icesjms/fss185.
- Quinn, T.J., and Deriso, R.B. 1999. Quantitative fish dynamics. Oxford University Press, New York.
- Richards, R.A., Fogarty, M.J., Mountain, D.G., and Taylor, M.H. 2012. Climate change and northern shrimp recruitment variability in the Gulf of Maine. Mar. Ecol. Prog. Ser. **464**: 167–178. doi:10.3354/meps09869.
- Sampson, D.B., and Scott, R.D. 2012. An exploration of the shapes and stability of population-selection curves. Fish. Fish. **13**(1): 89–104. doi:10.1111/j.1467-2979.2011.00417.x.
- Schirripa, M.J., Goodyear, C.P., and Methot, R.M. 2009. Testing different methods of incorporating climate data into the assessment of US West Coast sablefish. ICES J. Mar. Sci. **66**(7): 1605–1613. doi:10.1093/icesjms/fsp043.
- Thorson, J.T. 2011. Auxiliary and focal assessment models: a proof-of-concept involving time-varying catchability and fishery stock-status evaluation. ICES J. Mar. Sci. **68**(10): 2264–2276. doi:10.1093/icesjms/fsr160.
- Turnock, B.J., and Rugolo, L.J. 2009. Stock assessment of eastern Bering Sea snow crab. In Stock assessment and fishery evaluation report for the king and tanner crab fisheries of the Bering Sea and Aleutian Islands Regions. North Pacific Fishery Management Council, Anchorage, Alaska. pp. 29–123.
- Tyler, A.V., Beamish, R.J., and McFarlane, G.A. 1989. Implications of age determination errors to yield estimates. In Effects of ocean variability on recruitment and an evaluation of parameters used in stock assessment models. Edited by R.J. Beamish and G.A. McFarlane. Canadian Special Publication of Fisheries and Aquatic Sciences, Vol. 108. pp. 27–35.
- Whitten, A.R., Klaer, N.L., Tuck, G.N., and Day, R.W. 2013. Accounting for cohort-specific variable growth in fisheries stock assessments: a case study from south-eastern Australia. Fish. Res. **142**: 27–36. doi:10.1016/j.fishres.2012.06.021.
- Wilberg, M.J., Thorson, J.T., Linton, B.C., and Berkson, J. 2009. Incorporating time-varying catchability into population dynamic stock assessment models. Rev. Fish. Sci. **18**: 7–24. doi:10.1080/10641260903294647.
- Xu, X., Bishop, J.M., Mohammed, H.M.A., and Alsaffar, A.H. 1995. Estimation of the natural mortality rate of green tiger prawns *Penaeus semisulcatus* (de Hann, 1844) in Kuwait waters using relative abundance data. J. Shellfish Res. **14**: 179–184.
- Zheng, J., and Siddeek, M.S.M. 2011. Bristol Bay red king crab stock assessment in fall 2011. In Stock assessment and fishery evaluation report for the king and tanner crab fisheries of the Bering Sea and Aleutian Islands Regions. North Pacific Fishery Management Council, Anchorage, Alaska. pp. 169–283.