Improving assessment of Pandalus stocks using a seasonal, size-structured assessment model with environmental variables: Part II: Model evaluation and simulation

<table>
<thead>
<tr>
<th>Journal:</th>
<th>Canadian Journal of Fisheries and Aquatic Sciences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manuscript ID</td>
<td>cjfas-2016-0021.R1</td>
</tr>
<tr>
<td>Manuscript Type:</td>
<td>Article</td>
</tr>
<tr>
<td>Date Submitted by the Author:</td>
<td>16-Jul-2016</td>
</tr>
<tr>
<td>Complete List of Authors:</td>
<td>Cao, Jie; University of Maine, school of marine sciences Chen, Yong; University of Maine, Richards, R. Anne; Northeast Fisheries Science Center, National Marine Fisheries Service</td>
</tr>
<tr>
<td>Keyword:</td>
<td>stock assessment model, simulation, model misspecification, environment, RECRUITMENT &lt; General</td>
</tr>
</tbody>
</table>
Improving assessment of *Pandalus* stocks using a seasonal, size-structured assessment model with environmental variables: Part II: Model evaluation and simulation

Jie Cao\(^1\)*

Yong Chen\(^1\)

R. Anne Richards\(^2\)

\(^1\)School of Marine Sciences, University of Maine, Orono, Maine 04469, USA

\(^2\)Northeast Fisheries Science Center, 166 Water St., Woods Hole, MA 02543

Correspondence Email: jie.cao@maine.edu
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.

Keywords: stock assessment model; simulation; model misspecification; environment; recruitment; and seasonality
INTRODUCTION

Integrated size-structured stock assessment models are now being used widely for assessment and management of hard-to-age species (Punt et al. 2013). These include lobsters (Chen et al. 2005; ASMFC 2009; Punt et al. 2012), crabs (Zheng and Siddeek 2011; Turnock and Rugulo 2011), sea scallop (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. 2013). 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. 2010; Thorson 2011), selectivity (Linton and Bence 2011; Sampson and Scott 2012) and natural mortality (Punt et al. 2012; 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 mis-specification. A seasonal, size-structured assessment model for hermaphroditic Pandalid shrimp was developed to improve Pandalid stock assessment (Cao et al., 2016). 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, 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; Bergstrom 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 (Beverton and Holt 1957; Ricker 1975; Zheng et al. 1995; 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. 2013; 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 which 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; Bergstrom 1992)
and/or annually due to environmental changes (Apollonio et al. 1986; Bergstrom 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 stock (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. 2016) 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. 2016). It has a seasonal time step which 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. (2016) 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. (2016). 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 2012). 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 von Bertalanffy model, i.e., $L_\infty$, $K$, standard deviation of $L_\infty$ ($\sigma_{L_\infty}$), standard deviation of $K$ ($\sigma_K$), and the correlation between $L_\infty$ and $K$ ($\rho$) (Chen et al. 2003). Depending upon scenarios $L_\infty$ and $K$ either depending on season or not and either varying 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

---

**Table 1**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$L_\infty$</td>
<td>Asymptotic length</td>
</tr>
<tr>
<td>$K$</td>
<td>Growth rate</td>
</tr>
<tr>
<td>$\sigma_{L_\infty}$</td>
<td>Standard deviation of $L_\infty$</td>
</tr>
<tr>
<td>$\sigma_K$</td>
<td>Standard deviation of $K$</td>
</tr>
<tr>
<td>$\rho$</td>
<td>Correlation between $L_\infty$ and $K$</td>
</tr>
</tbody>
</table>

---

**Table 2**

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seasonal</td>
<td>Growth parameters vary by season</td>
</tr>
<tr>
<td>Interannual</td>
<td>Growth parameters vary by year</td>
</tr>
<tr>
<td>Both</td>
<td>Growth parameters vary by season and year</td>
</tr>
</tbody>
</table>

---

https://mc06.manuscriptcentral.com/cjfas-pubs
second season were assumed to be higher than those of the first season because of warmer

temperature (Haynes and Wigley 1969; Bergstrom 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)
recruitment 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 average 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 catchability, 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 or/and 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 ignores the environmental effects on recruitments. Scenario 16 aims 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:

\[
RE_n = \frac{Est_n - Sim_n}{Sim_n}
\]

\[
ARE = \frac{1}{N} \sum_{n=1}^{N} \frac{Est_n - Sim_n}{Sim_n}
\]
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.

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. 2E3). $F$ 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 $RE$s and $ARE$s 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 times (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 $ARE$s 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 due to the large number of parameters, i.e., annual von Bertalanffy growth parameters. Ignoring seasonally varying growth (in a scenario with no interannual variation, scenario 7) affected the estimation much less than ignoring annually varying growth (in a scenario with no seasonal variation, 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).
expected, there was no estimation bias for $SSB$, $R$ and $F$ in scenario 13 when informative
evironment 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
$RE$s of $SSB$ and $R$ in the recent years 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 OMs 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 is most likely to represent 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. In order to understand that to what extend the results depend on the fishing history we repeated all the 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 (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 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. 2013; 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 to accounting for time-varying $M$. 

https://mc06 manuscriptcentral.com/cjfas-pubs
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/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 were 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 inter-annual 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 2010). This can be problematic because the estimated recruitment variability is always less than the true variability among recruitments (Methot and Taylor 2010). 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 inclusion 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 mis-specification, including scenarios that mis-specified 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 diagnostic remains 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 the 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 improve an early version of the manuscript, for which we are thankful.

Appendix A. Supplementary data
REFERENCES:


Hilborn, R. 2003. The state of the art in stock assessment: Where we are and where we are going. 

Haltuch, M.A., Punt, A.E., 2011. The promises and pitfalls of including decadal-scale climate 

Cunningham, C. J. et al. 2015. Time-varying natural mortality in fisheries stock assessment 

mortality within a fisheries stock assessment model: an evaluation using simulation analysis 

Linton, B. C., and Bence, J. R. 2011. Catch-at-age assessment in the face of time-varying 

Lorenzen, K. 1996a. The relationship between body weight and natural mortality in juvenile and 

Maunder, M. N., Starr, P. J., 2001. Bayesian assessment of the SNA1 snapper (Pagrus auratus) 


time series into stock assessment models: model description, simulation testing, and example. 

Maunder, M. N., P. R. Crone, J. L. Valero, B. X. Semmens. 2016. Growth: Theory, estimation, 

Methot, R. D., Taylor, I. G., 2011. Adjusting for bias due to variability of estimated recruitments 


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

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

<table>
<thead>
<tr>
<th>Scenarios</th>
<th>Seasonal fishing pattern</th>
<th>Seasonal growth</th>
<th>Inter-annual variation in growth</th>
<th>Environmental driven $R$</th>
<th>Environmental driven $M$</th>
<th>Seasonal fishing pattern</th>
<th>Seasonal growth</th>
<th>Inter-annual variation in growth</th>
<th>Environmental driven $R$</th>
<th>Environmental driven $M$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>✓</td>
<td>–</td>
<td>–</td>
<td>✓</td>
<td>✓</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>✓</td>
</tr>
<tr>
<td>2</td>
<td>✓</td>
<td>–</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>–</td>
<td>✓</td>
<td>–</td>
<td>✓</td>
</tr>
<tr>
<td>3</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>4</td>
<td>✓</td>
<td>✓</td>
<td>–</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>–</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>5</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>–</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>6</td>
<td>✓</td>
<td>✓</td>
<td>–</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>7</td>
<td>✓</td>
<td>✓</td>
<td>–</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>–</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>8</td>
<td>✓</td>
<td>–</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>–</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>9</td>
<td>✓</td>
<td>✓</td>
<td>–</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>10</td>
<td>✓</td>
<td>✓</td>
<td>–</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>11</td>
<td>✓</td>
<td>✓</td>
<td>–</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>12</td>
<td>✓</td>
<td>✓</td>
<td>–</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>13</td>
<td>✓</td>
<td>–</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>–</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>14</td>
<td>✓</td>
<td>–</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>–</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>15</td>
<td>✓</td>
<td>–</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>–</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>16</td>
<td>✓</td>
<td>–</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>–</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>17</td>
<td>✓</td>
<td>–</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>–</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>18</td>
<td>✓</td>
<td>–</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>–</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>19</td>
<td>✓</td>
<td>–</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>–</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>20</td>
<td>✓</td>
<td>–</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>–</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

* the environmental data used in scenario 15 have large random noise (CV of 2.0)
Table 3. Mean, median, and standard error (sd) of the percentage averaged relative error ($ARE$) for two hundred estimation runs for estimated spawning stock biomass (SSB), recruitment ($R$) and fishing mortality ($F$) for the 20 scenarios under consideration.

<table>
<thead>
<tr>
<th>Scenarios</th>
<th>SSB Mean</th>
<th>Median</th>
<th>sd</th>
<th>R Mean</th>
<th>Median</th>
<th>sd</th>
<th>F Mean</th>
<th>Median</th>
<th>sd</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>117.03</td>
<td>115.50</td>
<td>22.61</td>
<td>144.55</td>
<td>142.47</td>
<td>22.14</td>
<td>-48.09</td>
<td>-48.30</td>
<td>5.61</td>
</tr>
<tr>
<td>2</td>
<td>0.49</td>
<td>0.34</td>
<td>12.65</td>
<td>0.25</td>
<td>-0.47</td>
<td>9.30</td>
<td>4.27</td>
<td>2.73</td>
<td>16.44</td>
</tr>
<tr>
<td>3</td>
<td>-10.00</td>
<td>-12.28</td>
<td>14.79</td>
<td>-6.07</td>
<td>-8.19</td>
<td>15.34</td>
<td>229.67</td>
<td>207.01</td>
<td>111.61</td>
</tr>
<tr>
<td>5</td>
<td>-10.16</td>
<td>-10.06</td>
<td>9.09</td>
<td>0.18</td>
<td>0.00</td>
<td>7.32</td>
<td>21.16</td>
<td>18.39</td>
<td>15.98</td>
</tr>
<tr>
<td>7</td>
<td>-1.90</td>
<td>-4.10</td>
<td>12.28</td>
<td>-0.45</td>
<td>-1.29</td>
<td>8.66</td>
<td>17.71</td>
<td>17.34</td>
<td>18.89</td>
</tr>
<tr>
<td>8</td>
<td>-34.84</td>
<td>-37.52</td>
<td>11.06</td>
<td>13.28</td>
<td>12.11</td>
<td>11.39</td>
<td>176.32</td>
<td>173.10</td>
<td>70.37</td>
</tr>
<tr>
<td>9</td>
<td>4.51</td>
<td>3.68</td>
<td>10.32</td>
<td>4.42</td>
<td>3.88</td>
<td>7.52</td>
<td>-0.52</td>
<td>-0.89</td>
<td>11.36</td>
</tr>
<tr>
<td>10</td>
<td>-58.72</td>
<td>-58.78</td>
<td>2.33</td>
<td>-91.86</td>
<td>-91.84</td>
<td>0.30</td>
<td>304.09</td>
<td>302.51</td>
<td>34.39</td>
</tr>
<tr>
<td>11</td>
<td>-57.87</td>
<td>-57.91</td>
<td>2.51</td>
<td>-90.16</td>
<td>-90.16</td>
<td>0.41</td>
<td>366.00</td>
<td>363.22</td>
<td>38.20</td>
</tr>
<tr>
<td>12</td>
<td>3.43</td>
<td>2.54</td>
<td>11.55</td>
<td>4.15</td>
<td>3.42</td>
<td>8.20</td>
<td>-2.21</td>
<td>-2.25</td>
<td>12.98</td>
</tr>
<tr>
<td>13</td>
<td>4.40</td>
<td>2.93</td>
<td>12.17</td>
<td>3.44</td>
<td>3.36</td>
<td>8.65</td>
<td>-0.93</td>
<td>-1.20</td>
<td>12.91</td>
</tr>
<tr>
<td>15</td>
<td>2.62</td>
<td>1.87</td>
<td>10.99</td>
<td>2.96</td>
<td>2.66</td>
<td>8.11</td>
<td>0.34</td>
<td>-1.12</td>
<td>12.25</td>
</tr>
<tr>
<td>16</td>
<td>-56.35</td>
<td>-56.49</td>
<td>2.44</td>
<td>-89.86</td>
<td>-89.86</td>
<td>0.38</td>
<td>363.53</td>
<td>362.28</td>
<td>34.01</td>
</tr>
<tr>
<td>17</td>
<td>-58.12</td>
<td>-58.19</td>
<td>2.45</td>
<td>-90.21</td>
<td>-90.22</td>
<td>0.40</td>
<td>367.85</td>
<td>369.20</td>
<td>30.25</td>
</tr>
<tr>
<td>18</td>
<td>-56.26</td>
<td>-56.32</td>
<td>2.54</td>
<td>-89.83</td>
<td>-89.84</td>
<td>0.41</td>
<td>361.28</td>
<td>360.52</td>
<td>29.35</td>
</tr>
<tr>
<td>19</td>
<td>-57.85</td>
<td>-57.72</td>
<td>2.65</td>
<td>-91.66</td>
<td>-91.66</td>
<td>0.31</td>
<td>323.02</td>
<td>317.40</td>
<td>39.90</td>
</tr>
<tr>
<td>20</td>
<td>-33.64</td>
<td>-37.98</td>
<td>13.42</td>
<td>14.96</td>
<td>11.72</td>
<td>13.14</td>
<td>234.50</td>
<td>233.03</td>
<td>107.87</td>
</tr>
</tbody>
</table>
FIGURE CAPTIONS

Figure 1. (a) Time-varying natural mortality (black line) and time-varying natural mortality containing random noise (red line) and (b) U-shaped natural mortality scaler used in the operating model.

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

Figure 3. Relative error ($RE$) of recruitment ($R$) across years for the twenty scenarios considered in this study.

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

Figure 5. Percent averaged relative error ($ARE$) of recruitment ($R$) for the twenty scenarios considered in this study.

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

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

Figure 8. Histogram of estimated natural mortality from scenario 12.
Figure 1. (a) Time-varying natural mortality (black line) and time-varying natural mortality containing random noise (red line) and (b) U-shaped natural mortality scaler used in the operating model.
Figure 2. Relative error ($RE$) of spawning stock biomass ($SSB$) across years for the twenty scenarios considered in this study.
Figure 3. Relative error ($RE$) of recruitment ($R$) across years for the twenty scenarios considered in this study.
Figure 4. Relative error (RE) of fishing mortality (F) across years for the twenty scenarios considered in this study.
Figure 5. Percent averaged relative error (ARE) of recruitment ($R$) for the twenty scenarios considered in this study.
Figure 6. Percent averaged relative error (ARE) of spawning stock biomass (SSB) for the twenty scenarios considered in this study.
Figure 7. Percent averaged relative error (ARE) of fishing mortality (F) for the twenty scenarios considered in this study.
Figure 8. Histogram of estimated natural mortality from scenario 12.
Figure S1. —Relative error ($RE$) of spawning stock biomass ($SSB$) across years for the twenty scenarios considered in this study using relatively constant fishing mortality time series.

Figure S2. —Relative error ($RE$) of recruitment ($R$) across years for the twenty scenarios considered in this study using relatively constant fishing mortality time series.

Figure S3. —Relative error ($RE$) of fishing mortality ($F$) across years for the twenty scenarios considered in this study using relatively constant fishing mortality time series.

Figure S4. —Percent averaged relative error ($ARE$) of recruitment ($R$) for the twenty scenarios considered in this study using relatively constant fishing mortality time series.

Figure S5. —Percent averaged relative error ($ARE$) of spawning stock biomass ($SSB$) for the twenty scenarios considered in this study using relatively constant fishing mortality time series.

Figure S6. —Percent averaged relative error ($ARE$) of fishing mortality ($F$) for the twenty scenarios considered in this study using relatively constant fishing mortality time series.
Fig. S1.
Fig. S2.
Fig. S3.
Fig. S4.
Fig. S5.
Fig. S6.