Correct in theory but wrong in practice: Bias caused by using a lognormal distribution to penalize annual recruitments in fish stock assessment models

Jonathan J. Deroba *, Timothy J. Miller

NOAA Fisheries, Northeast Fisheries Science Center, 166 Water Street, Woods Hole, MA 02543, United States

A R T I C L E   I N F O

Article history:
Received 2 October 2015
Received in revised form
24 November 2015
Accepted 3 December 2015

Keywords:
Likelihood
Recruitment
Assessment
Simulation
Fish

A B S T R A C T

Penalties are widely used for a range of parameters while fitting fish stock assessment models. Penalizing annual recruitments for deviating from an underlying mean recruitment is probably the most common. Assuming that recruits are log-normally distributed for the purposes of this penalty is theoretically justifiable. In practice, however, bias may be induced because this distributional assumption includes a term equal to the summation of the log observed data, which in the case of recruitment equals the summation of the log recruitment parameters that are not data. Using simulation, the potential for bias caused by assuming that recruits were log-normally distributed was explored, and results were contrasted with the assumption that log-recruitment was normally distributed, an alternative that avoids the potentially troublesome summation term. Spawning stock biomass (SSB) and recruitment were negatively biased, while fishing mortality (F) was positively biased under the assumption of log-normally distributed recruitments, and the bias worsened closer to the terminal year. The bias also worsened when the true underlying F was low relative to natural mortality, and with domed fishery selectivity. Bias in SSB, recruitment, and F was nonexistent or relatively small under the assumption that log-recruitment was normally distributed. Distributional assumptions for penalties used in assessment models should be reviewed to reduce the potential for biased estimation. These results also provide further support for simulation testing to evaluate statistical behavior of assessment models.

Published by Elsevier B.V.

1. Introduction

Penalizing the objective function is common practice for estimating parameters in fish stock assessment models (e.g., Butterworth et al., 2003; Punt et al., 2011). Penalties have been applied to dampen the degree of interannual variation of time varying parameters (e.g., selectivity, catchability, growth), prevent unrealistically large changes in selectivity among ages, and control the degree to which annual recruitments deviate from an underlying stock-recruitment curve (Janelli, 2002; Parma 2002; Maunder and Deriso, 2003; Thorson et al., 2015). At least some applications of penalized likelihood, however, can induce biased parameter estimates (Maunder and Deriso, 2003). Despite the potential for bias, little research has evaluated the performance of stock assessments under various assumptions for the penalty terms.

Perhaps the most ubiquitous application of penalized likelihood in stock assessment is for recruitment parameters (Legault and Restrepo, 1999; Brodziak, 2005; Ebener et al., 2005; Butterworth and Rademeyer, 2008; Methot and Wetzel, 2013). Fish stock assessment models commonly estimate recruitment parameters with an assumption for annual recruitment deviations based on some variation of a normal distribution:

$$\hat{R}_y = \bar{R}_Y e^{\bar{f}_Y}; e^{\bar{f}_Y} \sim N (0, \sigma^2)$$

where \(\hat{R}_y\) is recruitment in year \(y\), \(\bar{R}_Y\) is mean recruitment that may be a function of spawning stock (e.g., Beverton–Holt, Ricker), \(\bar{f}_Y\) is the annual deviation from the log-scale mean, and \(\sigma^2\) is the variance of the deviations. \(\bar{R}_Y\) is also sometimes multiplied by \(e^{-\sigma^2/2}\) as a bias correction so that the mean of the log-normally distributed \(\bar{R}_y\) equals \(\bar{R}_Y\), and this bias correction may also vary annually (Methot and Taylor, 2011).

A variety of definitions have been used for the contribution of a recruitment penalty to the overall objective function in statistical catch-at-age models (Table 1), but one important distinction is...
whether normal or log-normal likelihoods are used. The difference between log-likelihoods when specifying:

\[
\hat{R}_y \sim \text{LN} \left( \ln \left( \bar{R}_y \right), \sigma^2 \right) \tag{1}
\]

or

\[
\ln \left( \hat{R}_y \right) \sim \text{N} \left( \ln \left( \bar{R}_y \right), \sigma^2 \right) \tag{2}
\]

is \( \sum_{y=1}^{n \text{rec}} \ln \left( \hat{R}_y \right) \), which appears in the log-likelihood for the lognormal distribution, but not the normal (Table 1). Both of these distributional assumptions are equally justified theoretically, and in typical maximum likelihood estimation each option produces identical parameter estimates because this extra summation term is a constant in the objective function. For example, using these distributions for (log-) relative abundance indices would estimate identical parameter values. However, when specifying recruitment to be lognormally distributed, as in Eq. (1), the annual recruitments \( \hat{R}_y \) are estimated parameters rather than data, which may be problematic because the extra summation term is no longer a constant. More specifically, the model fit may improve (i.e., the penalized likelihood increased) by reducing the scale of the recruitment estimates despite signals from other data sources, and induce biased estimation of various population attributes. Ignoring the extra summation term in the lognormal distribution as a constant, however, is technically incorrect when Eq. (1) is assumed. Furthermore, the fact that these two distributions are both equally justified in theory, but may not perform equivalently in practice suggests that the topic of distributional assumptions for penalized maximum likelihood warrants evaluation in application to recruitment as well as any other non-normal penalties that might be used.

The objective of this manuscript was to review the potential for biased estimation of spawning stock biomass (SSB), fishing mortality (F), and recruitment caused by using the lognormal distribution to penalize annual recruitment. Using a simulation study with the alternative assumptions of Eqs. (1) and (2), we evaluated the effect of the summation term \( \sum_{y=1}^{n \text{rec}} \ln \left( \hat{R}_y \right) \) in the log-likelihood on the performance of a statistical catch-at-age (SCAA) model.

### 2. Methods

#### 2.1. Overview

The SCAA model used for all simulations was the Age Structured Assessment Program, version 3.0.8 (ASAP; Legault and Restrepo, 1999; NOAA, 2012). This version of ASAP does not include a bias correction term for recruitment deviations (i.e., \( e^{-\frac{\hat{R}_y^2}{2}} \)), but sensitivity analysis using an annually varying adjustment described by Methot and Taylor (2011) and Methot and Wetzel (2013) suggested conclusions about the relative performance of the lognormal and normal penalties were robust to this omission, but improvements in bias near the terminal years could be achieved (Appendix A).

Simulations were used to estimate bias in parameter estimates by fitting the SCAA model to pseudo-datasets. Differences in bias were contrasted for normal and lognormal penalty assumptions for recruitment by fitting models with the alternative assumptions to the same pseudo-datasets (i.e., fits with and without the summation term highlighted above; Table 1). The simulation experiment was repeated for different values of \( F \), natural mortality (\( M \)), and selectivity patterns (see below). For each simulation experiment, 100 pseudo-datasets each 40 years long were generated for use in the SCAA model.

#### 2.2. Simulations

##### 2.2.1. True underlying dynamics and pseudo-data generation

True population characteristics were based on a generic fish species with the general characteristics of groundfish in the northeast United States. Fish were approximately 50% mature at age-2 and 100% mature by age-4 (Table 2). Mean weights-at-age were time invariant and were the same for harvested and spawning fish (Table 2). Maturity and weights-at-age were constant among all simulations. We performed separate simulation experiments for multiple values of other population characteristics in a full factorial design for two levels of fully-selected \( F \) (time invariant), two levels of age- and time-invariant \( M \), and for flat topped and domed fishery selectivity (Table 2). This study design resulted in separate simulation experiments for each of 16 combinations (2 \( F \) values \( \times \) 2 \( M \) values \( \times \) 2 selectivity shapes \( \times \) 2 with normal or lognormal penalty). Numbers-at-age in the first year of the simulations equaled the deterministic equilibrium values associated

---

### Table 1

<table>
<thead>
<tr>
<th>Assumed distribution</th>
<th>Negative log-likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \hat{R}_y \sim \text{LN} \left( \ln \left( \bar{R}_y \right), \sigma^2 \right) )</td>
<td>( n_{\text{rec}} \ln(\hat{R}<em>y) + n</em>{\text{rec}} \ln(\sigma) + \frac{1}{\sigma^2} \sum_{y=1}^{n_{\text{rec}}} \left( \ln(\hat{R}_y) - \ln(\bar{R}_y) \right)^2 )</td>
</tr>
<tr>
<td>or ( \ln \left( \hat{R}_y \right) \sim \text{N} \left( \ln \left( \bar{R}_y \right), \sigma^2 \right) )</td>
<td>( n_{\text{rec}} \ln(\hat{R}<em>y) + n</em>{\text{rec}} \ln(\sigma) + \frac{1}{\sigma^2} \sum_{y=1}^{n_{\text{rec}}} \left( \hat{R}_y - \ln(\bar{R}_y) \right)^2 )</td>
</tr>
</tbody>
</table>

### Table 2

<table>
<thead>
<tr>
<th>Age</th>
<th>Low M</th>
<th>High M</th>
<th>Maturity (%)</th>
<th>Weight (kg)</th>
<th>Flat fishery selectivity</th>
<th>Domed fishery selectivity</th>
<th>Survey selectivity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.2</td>
<td>0.6</td>
<td>0</td>
<td>0.15</td>
<td>0.05</td>
<td>0.05</td>
<td>0.10</td>
</tr>
<tr>
<td>2</td>
<td>0.2</td>
<td>0.6</td>
<td>46</td>
<td>0.40</td>
<td>0.20</td>
<td>0.20</td>
<td>0.50</td>
</tr>
<tr>
<td>3</td>
<td>0.2</td>
<td>0.6</td>
<td>97</td>
<td>0.60</td>
<td>0.40</td>
<td>0.40</td>
<td>1.00</td>
</tr>
<tr>
<td>4</td>
<td>0.2</td>
<td>0.6</td>
<td>100</td>
<td>0.90</td>
<td>0.50</td>
<td>0.50</td>
<td>1.00</td>
</tr>
<tr>
<td>5</td>
<td>0.2</td>
<td>0.6</td>
<td>100</td>
<td>1.25</td>
<td>0.80</td>
<td>0.80</td>
<td>1.00</td>
</tr>
<tr>
<td>6</td>
<td>0.2</td>
<td>0.6</td>
<td>100</td>
<td>1.65</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>7</td>
<td>0.2</td>
<td>0.6</td>
<td>100</td>
<td>1.85</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>8+</td>
<td>0.2</td>
<td>0.6</td>
<td>100</td>
<td>2.20</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>
with a mortality rate that produced a population that was at 20% of unfished abundance. Sensitivity analyses suggested that results were robust to these starting conditions. Given the initial numbers at age and SSB, a single realization of the true underlying population was generated for all simulations, with numbers at age in subsequent time steps determined by the usual Baranov catch equations, exponential rates of death, and a Beverton–Holt stock-recruit relationship with steepness equal to 0.8 (Table 3). Post-recruit numbers at age were decremented deterministically whereas recruitment varied from the mean underlying Beverton–Holt relationship by an annual, lognormal deviation (no bias correction term) with a coefficient of variation (CV) equal to 1.0 (Table 3). This method of generating a time series of recruitment was structurally consistent with the SCAA model used for estimation.

Pseudo-datasets were generated using the age-based Population Simulator (PopSim, NOAA, 2013). The technical details of PopSim are available in Deroba et al. (2015), and so only a brief description was provided here. A single relative abundance index was simulated with time-invariant catchability and flat-topped selectivity (Table 2). Errors in annual indices and total fishery catch were independent and identically distributed as lognormal with a CV of 0.4 for the indices and 0.1 for the catch (Table 3). Annual samples for survey and catch age compositions were multinomial distributed with annual sample sizes of 200 (Table 3).

### 2.2.2. Stock assessment model

In order to isolate the effects of the distributional assumptions for the recruitment penalty in the objective function, all other specifications of the SCAA model were consistent with the generation of the simulated datasets. The SCAA model was provided with the correct M, maturity- and weights-at-age vectors, and used an age-8+ (plus) group (Table 2). The model was also provided the correct CVs for the annual survey indices, total fishery catch, recruitments, and annual sample sizes for the age compositions (Table 3). For flat-topped fishery selectivity, the SCAA model estimated age-specific selectivity parameters for age-1–age-5, but selectivity was fixed for other ages at the values used to simulate the data. For domed fishery selectivity, the SCAA model estimated age-specific selectivity parameters for all ages except age-6, which was fixed at the value used to simulate the data. Survey selectivity at age-1 and age-2 was estimated, but fixed for other ages at the values used to simulate the data. The stock–recruitment relationship in the SCAA model was the same Beverton–Holt functional form as used to generate the true recruitment time series, and steepness and unexploited biomass were estimated as parameters. Fully selected F in each year was estimated in log space as multiplicative deviations from the F in the previous year. The other parameters estimated by the SCAA model included numbers-at-age in the first year, annual recruitments, and survey catchability. All initial parameter values were the same as those used to simulate data.

### 2.2.3. Output metrics

For a given simulation experiment and recruitment penalty assumption, mean relative bias for F being the true annual SSB, fully selected F, or recruitment (i.e., age-1 abundance) was:

\[
\text{RelativeBias}(\theta) = \frac{1}{n} \sum_{i=1}^{n} \left( \frac{\theta_i}{\text{true}_i} - 1 \right)
\]

where \( \theta \) was the estimate of \( \theta \) (i.e., SSB, fully selected F, or recruitment) for pseudo-dataset i. Bias was considered significant if the 95% confidence intervals (mean ± 1.96 standard deviations) for the relative bias estimate did not overlap zero. Example results were provided graphically to highlight the simulation characteristics that had meaningful effects. A table of relative bias estimates for terminal year SSB, F, and recruitment, however, was provided for all simulation experiments.

### 3. Results

#### 3.1. Normal distribution for log-recruitment

With the normal penalty for log-recruitment, results for relative bias were generally consistent regardless of the other simulation characteristics (i.e., M, fishery selectivity, fully selected F). So, only example graphs with relatively high M (0.6), domed fishery selectivity, and relatively high F (0.8) were provided.

Relative bias in SSB, F, and recruitment was not significantly different from zero in most years but was significantly less than zero for SSB and recruitment, and greater than zero for F in the 3–4 nearly terminal years and terminal year (Fig. 1). Relative bias in the terminal year for SSB was not significantly different from zero except when F was relatively high (0.8), but the mean bias was 5% or less, whereas relative bias in the terminal year for F was significantly greater than zero in all cases (Table 4). Relative bias in the terminal year for recruitment was significantly less than zero in all cases (Table 4).

#### 3.2. Lognormal distribution for recruitment

With the lognormal penalty for recruitment, bias depended on various simulation characteristics. Consequently, detailed results were provided for a subset of simulations chosen to highlight the simulation characteristics that most affected bias.

Relative bias was significantly less than zero for SSB and recruitment during all or some of the time series, depending on simulation characteristics, and the degree of error worsened closer to terminal years (Figs. 2 and 4). Results for F were similar except with positive bias (Fig. 3). Bias in SSB, F, and recruitment was worse in all years and reached greater extremes in more recent years when F was relatively low than high (Figs. 2–4, panel a versus b and d versus e). Bias in SSB, F, and recruitment was also worse in all years and reached greater extremes when M was relatively high and F was relatively low than for all other combinations (Figs. 2–4, panels b and c versus others). With relatively low F, bias in SSB, F, and recruitment was also worse with domed fishery selectivity than with flat fishery selectivity (Figs. 2–4, panel b versus c and e versus f), but this effect was relatively minor compared to other simulation characteristics and diminished with relatively high F (not shown in figures).

Relative bias in the terminal year was significantly less than zero for SSB and recruitment, and significantly greater than zero for F in all cases (Table 4). Bias was always greater with the lognormal penalty than the normal penalty.

---

**Table 3**

Stock-recruit parameters used in simulations, coefficients of variation defining the degree of noise in recruitment, indices, and catch, and the sample sizes used for survey and catch age compositions.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Steepness (all simulations)</td>
<td>0.8</td>
</tr>
<tr>
<td>Unfished SSB (all simulations)</td>
<td>28,896</td>
</tr>
<tr>
<td>Unfished recruitment for high M</td>
<td>38,831</td>
</tr>
<tr>
<td>Unfished recruitment for low M</td>
<td>5023</td>
</tr>
<tr>
<td>Recruitment CV</td>
<td>1.0</td>
</tr>
<tr>
<td>Annual index CV</td>
<td>0.4</td>
</tr>
<tr>
<td>Annual catch CV</td>
<td>0.1</td>
</tr>
<tr>
<td>Catch age composition sample size</td>
<td>200</td>
</tr>
<tr>
<td>Survey age composition sample size</td>
<td>200</td>
</tr>
</tbody>
</table>
4. Discussion

Under the assumption in Eq. (1) for the recruitment penalty, inclusion of the term $\sum_{p=1}^{\nu} \ln (\hat{R}_p)$ in the log-likelihood induced negative bias in estimates of $SSB$ and recruitment, and positive bias in estimates of $F$. The degree of bias was worse when $F$ was relatively low combined with high $M$, and (or) domed fishery selectivity. The reasons for these results are unclear, but it did suggest that the magnitude of the effect of the lognormal penalty will be case specific. For example, exploratory work using actual assessments for Atlantic herring (Clupea harengus; NEFSC, 2012) and pollock (Pollachius virens; Hendrickson et al., 2015) suggested that the lognormal penalty interacts with aspects of model structure that were not evaluated as part of these simulations. Using a normal penalty for log-recruitment (instead of the lognormal penalty for recruitment) in the Atlantic herring stock assessment model increased recent estimates of spawning stock biomass to a greater extent when the assessment assumed time-varying $M$ rather than time-invariant $M$. Using the normal penalty for log-recruitment (instead of the lognormal penalty for recruitment) in the pollock assessment model increased estimates of $SSB$ by about the same extent throughout the entire time series when fishery and survey selectivity were domed, but recent estimates were affected to a greater extent than historic estimates when selectivity was flat. Conversely, several assessments produced similar results regardless of the penalty used for recruitment. The effect of including the $\sum_{p=1}^{\nu} \ln (\hat{R}_p)$ term as part of the log-likelihood should be evaluated on a case by case basis, and this may reveal important aspects of model structure.

The notion of maximum penalized likelihood estimation has existed in the statistical literature for some time. The penalized likelihood is used in a range of applications such as density estimation (Good and Gaskins, 1971), ridge regression (Hastie et al., 2009), the lasso (Tibshirani, 1996), and generalized additive models (Marx and Eilers, 1998). The penalty is typically meant to deviate from likelihood-based estimates when the likelihood provides
implausible estimates. For example, penalized likelihoods in generalized additive models limit the roughness of the smoother defining effects of continuous covariates. The use of penalized likelihood in these cases implies a lack of belief in the relationship of the observations to their expectation defined by the unpenalized model. On the other hand, the typical purpose of penalties in assessment models is to ensure model convergence or estimate stock-recruitment parameters within the assessment model. It is not that we think the modeled relationship of age composition data to recruitment is in some way incorrect. It is not clear whether using penalized likelihood for stock assessment models is desirable because of the bias of parameter estimation induced by penalizing the likelihood. A reduction in mean squared error for penalized likelihood estimation would be supportive of the approach over unpenalized estimation. However, worse bias and mean squared error were observed by de Valpine and Hilborn (2005) for some parameters using penalized likelihood rather than a marginal likelihood. Quantitative comparisons for a wide range of population attributes and
Table 4

Mean (95% confidence interval) relative bias for terminal year estimate of spawning stock biomass, fishing mortality, and recruitment from simulations using a lognormal penalty on recruitment (low F), normal, normal penalty for log-recruitment (normal), low natural mortality (low M), high natural mortality (high M), low fishing mortality (low F), and flat (upper) fishing selectivity (flat).

<table>
<thead>
<tr>
<th>Penalty</th>
<th>Low M</th>
<th>Low F</th>
<th>High M</th>
<th>High F</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Spawning stock biomass</td>
<td>Fishing mortality</td>
<td>Recruitment</td>
<td>Lognormal</td>
</tr>
<tr>
<td>Low M</td>
<td>-0.23 (0.34, -0.70)</td>
<td>-0.41 (0.29, -0.70)</td>
<td>-0.49 (0.43, -0.87)</td>
<td>-0.47 (0.56, -0.70)</td>
</tr>
<tr>
<td>Flat</td>
<td>0.00 (0.01, 0.00)</td>
<td>0.00 (0.01, 0.00)</td>
<td>0.00 (0.01, 0.00)</td>
<td>0.00 (0.01, 0.00)</td>
</tr>
<tr>
<td>Lognormal</td>
<td>0.56 (0.49, 0.63)</td>
<td>0.35 (0.29, 0.42)</td>
<td>0.31 (0.25, 0.37)</td>
<td>0.13 (0.07, 0.18)</td>
</tr>
<tr>
<td>Normal</td>
<td>0.07 (0.02, 0.12)</td>
<td>0.13 (0.07, 0.18)</td>
<td>0.10 (0.05, 0.05)</td>
<td>0.01 (0.00, 0.01)</td>
</tr>
<tr>
<td>Low F</td>
<td>0.17 (0.08, 0.26)</td>
<td>0.11 (0.06, 0.16)</td>
<td>0.07 (0.02, 0.13)</td>
<td>0.03 (0.02, 0.05)</td>
</tr>
<tr>
<td>Flat</td>
<td>-0.64 (0.09, -0.83)</td>
<td>-0.66 (0.09, -0.83)</td>
<td>-0.71 (0.09, -0.83)</td>
<td>-0.71 (0.09, -0.83)</td>
</tr>
<tr>
<td>Lognormal</td>
<td>-0.45 (0.47, -0.45)</td>
<td>-0.47 (0.49, -0.45)</td>
<td>-0.44 (0.44, -0.44)</td>
<td>-0.44 (0.44, -0.44)</td>
</tr>
<tr>
<td>Normal</td>
<td>0.08 (0.02, 0.12)</td>
<td>0.11 (0.06, 0.16)</td>
<td>0.10 (0.05, 0.05)</td>
<td>0.03 (0.02, 0.05)</td>
</tr>
</tbody>
</table>

alternative penalties for recruitment and other processes (see for example Thorsen et al., 2015) to the objective function of stock assessment models remain for future research.

The potential exists for stock assessments to use theoretically justified assumptions for a range of penalized likelihoods that may not perform as expected and induce parameter bias. Determining the extent to which this problem may occur is difficult because the technical details of many stock assessments are poorly documented. Some relatively well documented stock assessments are also unclear as to the assumptions being made, even for recruitment. For example, documentation may claim that recruitments are assumed lognormal for the purposes of the penalty, but equations suggest that log recruitment is assumed normal (e.g., Ebener et al., 2005; Butterworth and Rademeyer, 2008). We encourage the technical details of stock assessments to be more clearly and carefully documented. We also suggest simulation tests similar to the self-consistency evaluations described in Deroba et al. (2015) as standard practice for evaluating the potential for bias caused by penalized likelihood and other aspects of assessment model structure.

In this study, using a normal penalty for log-recruitment outperformed the lognormal penalty, but the normal penalty also produced some bias, particularly around the terminal year. This bias was partially due to using a single realization of the recruitment process. Sensitivity analyses using different random number seeds suggested that bias was always present, but the direction and magnitude may differ from that reported in this manuscript. Use of an annually varying bias correction eliminated relative bias for estimates near the terminal years when using a normal penalty and reduced relative bias when using a lognormal penalty (Methot and Taylor 2011; Appendix A). Relative bias was generally unaffected by the inclusion of the bias correction for other years (i.e., more than ~5 years prior to the terminal year) likely because age composition existed in all years and was relatively informative (i.e., annual samples of 200). The bias correction would have more effect if the information content of the data varied among years and estimation of recruitment began prior to the availability of age composition data (Methot and Taylor, 2011). Although the effects of inclusion of the bias correction term were relatively minor compared to the difference between the normal and lognormal penalties, an annually varying bias correction term should likely be considered when using penalized likelihood (Methot and Taylor, 2011). Bias was also eliminated when no recruitment penalty was used, but recruitment estimates became unrealistically large and variable for some datasets. Without any recruitment penalty, Mauder and Deriso (2003) also found unrealistically large and variable recruitment estimates and recommended not using maximum likelihood estimation without a recruitment penalty or penalized likelihood when estimates of the variance of recruits were needed, and only using these methods for estimation in cases with full and informative catch-at-age matrices. The results of this study are consistent with those recommendations. Thorsen et al. (2015), however, suggested an alternative that applies the Laplace approximation to penalized likelihood estimates that can provide relatively well estimated variance parameters for time varying processes.

Another consideration for future research may be to treat recruitment as the unobserved process that it is, much like in a random effects/state-space setting (Mauder and Deriso, 2003; Nielsen and Berg 2014; Thorsen et al., 2015). Under this assumption, the variance of the recruitment process error would be an estimated parameter, while the recruitments themselves would be posterior predictions given the estimates of the other parameters in the model. Using this method with lognormal recruitment, the recruitment predictions integrate out of the likelihood and so may not affect fit as they would if being treated as true parameters. Mauder and Deriso (2003) found that such methods are superior
to penalized likelihood for estimating the variance of recruitments and in cases with missing catch-at-age data. These methods would also be more consistent with the fact that direct observations of recruitment are rare, and so should be treated as an unobserved process.

**Acknowledgment**

We are grateful to several colleagues for helpful dialogues during preparation of this manuscript, including Chris Legault, Jim Bence, Rick Methot, Jim Thorson, and Doug Butterworth. Frequent strolls with Mattie Ross Deroba and Woody Miller also helped maintain our clarity of thought on this and other topics. Comments from Ian Taylor, an anonymous reviewer, and the editor (Andre Punt) also improved the manuscript.

**Appendix A.**

In sensitivity analysis using an annually varying bias correction term, recruitment estimates from ASAP were amended to:

\[
\hat{R}_y = \hat{R}_y e^{-y - b_y \frac{M}{2}}; \quad \varepsilon \sim N(0, \sigma^2)
\]

where \(\hat{R}_y\) was estimated using the Beverton–Holt stock recruit function, as described in the main text. The likelihood equations (Table 1) were also amended to include the \(b_y\) as described in Methot and Taylor (2011), although this feature is moot when the variance is not estimated as a parameter. The \(b_y\) ramped linearly up from a value of 0 at the beginning of the time series to a constant value for several years, and ramped linearly back down to 0 in the terminal year:

\[
b_y = \begin{cases} 
0 & \text{fory} \leq y^b_1 \\
\frac{y - y^b_1}{y^b_2 - y^b_1} & \text{fory}^b_1 < y \leq y^b_2 \\
\frac{y - y^b_2}{y^b_3 - y^b_2} & \text{fory}^b_2 \leq y \leq y^b_3 \\
\frac{y - y^b_3}{y^b_4 - y^b_3} & \text{fory}^b_3 < y \leq y^b_4 \\
0 & \text{fory}^b_4 \leq y 
\end{cases}
\]

where \(b_{\text{max}} = 1\), \(y^b_1\) equaled zero, \(y^b_2\) equaled the 8th year in the times series, \(y^b_3\) equaled the 32nd year in the time series, and \(y^b_4\) equaled the terminal year (Methot and Taylor, 2011).

The effect of the annually varying bias correction term was similar regardless of other model characteristics (e.g., natural mortality, selectivity), and so only example plots with relatively high \(M(0.6)\), domed fishery selectivity, and relatively high \(F(0.8)\) were provided. With the normal penalty and the annually varying bias correction, relative bias for \(SSB\), \(F\), and recruitment near the terminal years was eliminated when compared to not using the varying bias correction (Figs. A1: 1). With the lognormal penalty, relative bias for \(SSB\), \(F\), and recruitment near the terminal years was reduced compared to not using the varying bias correction (Figs. A1: 2a, 3a and 4a). Relative bias in other years was similar regardless of whether or not the annually varying bias correction was applied for both the normal and lognormal penalties.

![Fig. A1](image_url) Mean and 95% confidence intervals of relative bias for simulations using a normal penalty (top row) or lognormal penalty (bottom row) for log-recruitment using an annually varying bias correction term with high \(M(0.6)\), domed fishery selectivity, and high \(F(0.8)\).
References


