

Manuscript Number: FISH6374R2

Title: Implementing a generic method for bias correction in statistical models using random effects, with spatial and population dynamics examples

Article Type: Research Paper

Keywords: random effects; mixed-effects model; Template Model Builder (TMB); stock assessment; epsilon estimator; bias correction

Corresponding Author: Dr. James T Thorson, Ph.D.

Corresponding Author's Institution: NWFSC/NOAA

First Author: James T Thorson, Ph.D.

Order of Authors: James T Thorson, Ph.D.; Kasper Kristensen, Ph.D.

Manuscript Region of Origin: AFGHANISTAN

Dr. André Punt

Nov. 17, 2015

Enclosed, please find responses to a request for moderate revisions for a manuscript entitled “Implementing a generic method for bias correction in statistical models using random effects, with spatial and population dynamics examples.” Dr. Kristensen and I have contributed to the work, accept responsibility for it, and agree to its submission to Fisheries Research. We do not have any extended abstracts, articles, or reports submitted or published elsewhere related to it.

In these revisions, we have made all suggested editorial changes.

We continue to believe that this work will be important and likely well cited for three main reasons. First, it develops a method that is widely applicable throughout fisheries statistical modeling whenever unbiased estimates are desirable. Second, the method is distributed with Template Model Builder, a replacement for AD Model Builder that is rapidly gaining users in fisheries sciences. Third, the method presents basic research regarding statistical methods, with application in other areas of statistical analysis.

James T. Thorson
NWFSC-NOAA
James.Thorson@noaa.gov

Highlights

- Mixed-effects models are increasingly important in fisheries science and ecology
- The standard plug-in estimate of a nonlinear transformation of random effects will be biased
- This bias can be mitigated using a new “epsilon” estimator implemented in Template Model Builder
- The epsilon bias-correction estimator is demonstrated for age-structured and spatial models
- The epsilon estimator performs similar or better than previous bias-correction methods for age-structured models

Implementing a generic method for bias correction in statistical models using random effects, with spatial and population dynamics examples

James T. Thorson^{1*}, Kasper Kristensen²

¹ Fisheries Resource Assessment and Monitoring Division, Northwest Fisheries Science Center, National Marine Fisheries Service, NOAA, 2725 Montlake Blvd. E, Seattle, WA 98112

² Technical University of Denmark, Charlottenlund Slot Jægersborg Allé 1, 2920 Charlottenlund

Abstract

Statistical models play an important role in fisheries science when reconciling ecological theory with available data for wild populations or experimental studies. Ecological models increasingly include both fixed and random effects, and are often estimated using maximum likelihood techniques. Quantities of biological or management interest (“derived quantities”) are then often calculated as nonlinear functions of fixed and random effect estimates. However, the conventional “plug-in” estimator for a derived quantity in a maximum likelihood mixed-effects model will be biased whenever the estimator is calculated as a nonlinear function of random effects. We therefore describe and evaluate a new “epsilon” estimator as a generic bias-correction estimator for derived quantities. We use simulated data to compare the epsilon-method with an existing bias-correction algorithm for estimating recruitment in four configurations of an age-structured population dynamics model. This simulation experiment shows that the epsilon-method and the existing bias-correction method perform equally well in data-rich contexts, but the epsilon-method is slightly less biased in data-poor contexts. We then apply the epsilon-method to a spatial regression model when estimating an index of population abundance, and compare results with an alternative bias-correction algorithm that involves Markov-chain Monte Carlo sampling. This example shows that the epsilon-method leads to a biologically significant difference in estimates of average abundance relative to the conventional plug-in estimator, and also gives essentially identical estimates to a sample-based bias-correction estimator. The epsilon-method has been implemented by us as a generic option in the open-source Template Model Builder software, and could be adapted within other mixed-effects modeling tools such as Automatic Differentiation Model Builder for random effects. It therefore has potential to improve estimation performance for mixed-effects models throughout fisheries science.

Keywords: random effects; mixed-effects model; Template Model Builder (TMB); stock assessment; epsilon estimator; bias correction

* Corresponding author:
Email: James.Thorson@noaa.gov
Telephone: 206-302-1772
Fax: 206-860-6792

1. Introduction

Statistical models for the dynamics of wild populations are designed to reconcile available data with ecological theory, and are often used to make predictions about future, historical, or otherwise unobserved process or events (Hilborn and Mangel, 1997). For example, population models are commonly used when evaluating trade-offs arising from management decisions, e.g., determining what level of fishery catch is consistent with long-term management objectives of fishery stakeholders (Walters and Martell, 2004). Population models are also used to improve insight regarding environmental trends at large spatial scales, e.g., detecting long-term trends in abundance for populations of birds (Schaub et al., 2007).

Parameters in population dynamics models are often estimated by identifying the values of parameters that maximize the “likelihood function”. The likelihood function is defined as the probability that the observed data would arise given a hypothesized model and proposed set of parameter values. Modern population models increasingly include a mix of random and fixed effects, where random effects are assumed to arise from a random process and generally account for non-independence in available data (Thorson and Minto, 2015). In particular, random effects may be used to approximate spatial variation in population density and productivity (Kristensen et al., 2014; Thorson et al., 2015b), variation in growth rates among individuals (Shelton et al., 2013), and variation over time in survival rates (Minto et al., 2014). Random effects can also be used to approximate more complicated stochastic processes, e.g., the unknown function representing the relationship between population abundance and population growth or recruitment (Munch et al., 2005; Thorson et al., 2014b).

Ecologists are often interested in predictions arising from population models, and these predicted quantities may be a function of random effects. As one example, population abundance could be calculated as the sum of density at multiple sites, where the logarithm of density $\log(d_i)$ at each site i is modeled as a normally-distributed random effect, i.e., $\log(d_i) = \gamma_i$, where $\gamma_i \sim N(\mu, \sigma^2)$ and μ and σ^2 are the mean and variation of log-density among sites. In this example, the random effect γ_i represents log-density at each site, and γ_i is transformed via a nonlinear function (exponentiated) when calculating the derived variable of interest (site-specific density). An ecologist might be interested in different “levels” of this hierarchical model, e.g., either the average density among sites μ or the predicted density at any individual site d_i .

Whenever a random variable is transformed by a nonlinear function, the mean and variance of the variable are also transformed. This is handled smoothly when estimating parameters using Bayesian methods (i.e., given samples from the posterior probability of parameters) because the value of the derived quantity can be calculated for each realization of the posterior distribution. However, in maximum likelihood estimation, the derived quantity is frequently calculated using a “plug-in” estimator, i.e., $\hat{d}_i = \exp(\hat{\gamma}_i)$ where $\hat{\gamma}_i$ is the estimate of log-density at site i , and \hat{d}_i is the plug-in estimator for density. Given an unbiased estimate of log-density $\hat{\gamma}_i$ in this example, the plug-in estimator of density $\hat{d}_i = \exp(\hat{\gamma}_i)$ will be biased because the function $\exp(\hat{\gamma}_i)$ is nonlinear. Previous research has developed methods to calculate the expected value of a derived quantity for specific fisheries models. For example, Methot and Taylor (2011) developed an algorithm to approximate the expected value of recruitment for marine populations, because recruitment is often calculated as a nonlinear function of a random effect representing temporal variation in survival of juvenile fish. Methot and Taylor (2011) note that their algorithm is not necessary for a Bayesian model of recruitment, because the expected value can be calculated for any nonlinear function given a set of samples from the posterior distribution of parameters. However, most stock assessment models used for tactical management in the United States and

elsewhere continue to use maximum likelihood estimation methods, so the Methot and Taylor (2011) algorithm remains important for many stock assessment models.

Here we demonstrate a novel approach for approximating the expected value of a quantity derived from random effects in a mixed-effects model. We first explain the theory of derived quantities in more detail, and explain the proposed “epsilon” method. We then provide a simulation experiment comparing the epsilon method with the previous Methot and Taylor (2011) algorithm for the case of recruitment in population models for marine populations, using four levels of data availability. We conclude by applying the epsilon method when estimating expected population abundance using a spatial regression model and a real-world application. This case study shows that the bias-corrected estimate from the epsilon method (and its estimated variance) are similar to estimates from an alternative, sample-based bias-correction estimator. We have implemented the epsilon method in the new Template Model Builder software (publicly available at <https://github.com/kaskr/adcomp>), and it can be applied off-the-shelf to other mixed-effects models.

2. Methods

2.1 Defining bias in maximum likelihood estimates of mixed-effects models

Fixed effects are parameters that are estimated by identifying which values maximize the likelihood function, defined as the probability (likelihood) of obtaining the observed data given a hypothesized value and proposed values for the fixed effects. By contrast, random effects are treated as random variables, and therefore are represented by a distribution of possible values where that distribution generally has a finite mean and variance. For mixed-effects models, the marginal likelihood function is calculated by integrating across all possible values of the random effects. Random effects are included in statistical models to account for situations in which available data are not statistically independent (e.g., variation over time, space, or among individuals), and mixed-effects models represent a generic approach for calculating the covariation among data that arises from a hypothesized model (Thorson and Minto, 2015). Values for random effects in maximum likelihood models can be predicted after fixed effects have been estimated (we note that other authors sometimes refer to the task of predicting random effects based on estimated fixed effects as “estimating” the random effects). Values for random effects are generally predicted via “empirical Bayes” (de Valpine, 2009), i.e., by plugging in the maximum likelihood estimate of fixed effects and then setting random effects to the mode or mean of their conditional distribution. However, it is possible to estimate fixed effects by maximizing the likelihood function without ever predicting the value of random effects. In the following, we are concerned with providing unbiased predictions of quantities that are calculated as a nonlinear function of random effects, and this in no way affects the estimates of fixed effects that arise from maximum likelihood estimation.

Maximum likelihood estimates of fixed effects are generally “consistent”, i.e., will converge to their true value given that samples sizes increase indefinitely (given reasonable conditions and that the model is correct; see Gelman et al. (2003, pg. 107-111)). By contrast, Empirical Bayes predictions of random effects are not generally consistent. Instead, additional data will often result in an increase in the number of random effects that must be predicted, thus maintaining a constant ratio of data per random effect.

Despite providing a consistent estimate of fixed effects, maximum likelihood estimates may be biased. An estimator $\hat{\theta}$ of parameter θ is said to be biased if $E[\hat{\theta}] - \theta \neq 0$ for a given sample size, and maximum likelihood estimates are often biased given small sample size. Bias-

correction methods generally exist, and can be used to calculate a “bias-corrected” estimator $\hat{\theta}_{BC}$ that is unbiased for all sample sizes (i.e., where $E[\hat{\theta}_{BC}] - \theta = 0$ regardless of sample size). However, bias-corrected estimators generally increase the total estimation variance, i.e., $Var[\hat{\theta}_{BC} - \theta] > Var[\hat{\theta}_{PI} - \theta]$ (where the left-hand side is the variance of the bias-corrected estimator, and the right-hand side is the variance of the plug-in estimator $\hat{\theta}_{PI}$ without bias correction). There is therefore a tradeoff between bias and variance when deciding whether to use bias-correction methods when estimating parameters via maximum likelihood. This definition of bias and bias-correction applies to both estimates of fixed-effects (via the likelihood function), and predictions of random effects (conditional on estimated fixed effects). In the following, we are specifically concerned with bias-correction for predictions of random effects, and a general treatment of bias-correction for fixed-effects in nonlinear mixed-effects models is not addressed here.

2.2 A generic approach to bias correction for nonlinear functions of random effects

In the following, we seek to obtain unbiased predictions of random effects, or quantities derived from predictions of random effects. Bias-correction for random effects may increase estimation variance, but we believe that some modelling contexts will favor prediction with minimized bias over predictions that are more biased and more precise (for example, when US federal law is interpreted to require unbiased estimates of a derived quantity like stock status). We therefore discuss and demonstrate a generic bias-correction algorithm for predicting random effects or quantities derived from them when fitting mixed-effects models via maximum likelihood.

We start by noting that mixed-effects models treat random effects as a random variable, and integrate across their unknown value when calculating the likelihood function (sometimes termed the marginal likelihood, see Thorson and Minto, (2015)). When calculating a derived variable B as a nonlinear transformation of a random variable A , $B = f(A)$, the expected value of the derived variable is not simple to compute in general. However, the expected value of the derived variable, $E[B]$, can be approximated via second-order Taylor series expansion:

$$E[B] \cong f(E[A]) + \frac{f''(E[A])}{2} Var[A] \quad (1a)$$

where $E[A]$ is the expected value of the original variable A , $Var[A]$ is the variance of the original variable, $f()$ is the nonlinear function for some derived quantity (e.g., exponentiated), and $f''()$ is the second derivative of this function. By Jensen’s inequality, $E[B] \geq f(E[A])$ whenever $f()$ is convex (e.g., when $f=\exp$), and this property is reflected in the Taylor series expansion (i.e., if f is convex, then $f'' > 0$).

Now envision a situation where the random variable A is a random effect predicted conditional on data x , where the empirical Bayes estimator \hat{A} for random effect A is its prediction conditional on data and the maximum likelihood estimate of fixed effects $\hat{\theta}$, e.g., $\hat{A} = E[A; x, \hat{\theta}]$. We seek an unbiased estimator $\hat{B} = E[f(A; x, \hat{\theta})]$ of the derived quantity $B = f(A; x, \hat{\theta})$. Many common models and software packages use the plug-in estimator:

$$\hat{B} = f(\hat{A}) = f(E[A; x, \hat{\theta}]) \quad (1b)$$

However, this plug-in estimator does not include the second- or higher-order Taylor series correction (where second-order correction is included in the right-hand side of Eq. 1a). For example, the random effects module of the automatic differentiation model builder software (ADMB-RE) uses the plug-in estimator for quantities calculated via a nonlinear transformation

of random effects (Skaug and Fournier, 2006), and therefore \hat{B} will be a biased estimator for derived quantity B whenever B is a nonlinear function of random effects.

A generic approach to bias correction in population models involving random effects can be used when calculating the expected value of derived quantities, and we have implemented this approach in Template Model Builder (TMB; Kristensen et al., 2014). This method (based on Tierney et al., (1989), which we call the “epsilon” method) specifically introduces an auxiliary variable ε for each derived quantity, and uses the gradient of the likelihood function with respect to this variable to directly compute the expected value of a derived quantity. We do not know of any previous application of this method in fisheries science or ecology, and have implemented this method in TMB in response to concerns regarding bias correction in fisheries models.

Specifically, we define $\ell(\theta, u; x)$ as the natural logarithm of the joint likelihood of fixed effects θ and random effects u given data x . Fixed effects are then estimated by maximizing the likelihood function when integrating across random effects:

$$\hat{\theta} = \operatorname{argmax}_{\theta} (\log (\int \exp (\ell(\theta, u; x)) du)) \quad (2a)$$

We seek to calculate the expected value of quantity ϕ , where $\phi = f(\hat{\theta}, u|x)$ is calculated from random effects u and estimated fixed effects $\hat{\theta}$ conditional on available data. The expected value of ϕ given available data is:

$$E[\phi|x] = \frac{\int \exp (\ell(\hat{\theta}, u; x)) f(\hat{\theta}, u) du}{\int \exp (\ell(\hat{\theta}, u; x)) du} \quad (2b)$$

which is the minimum variance unbiased estimator of ϕ that we seek (it also corresponds to the mean of the predictive distribution given estimated fixed effects and data).

The expected value of ϕ given available data can be calculated by including a nuisance parameter ε , and calculating the gradient of the marginal likelihood with respect to this nuisance parameter. We proceed by defining a new function, g :

$$g(\theta, u, \varepsilon; x) = \log (\int \exp (\ell(\theta, u; x) - \varepsilon f(\theta, u)) du) \quad (2c)$$

In this case, the derivative of the new function $g(\theta, u, \varepsilon; x)$ with respect to ε is:

$$\frac{\partial}{\partial \varepsilon} (g(\theta, u, \varepsilon; x)) = \frac{\int \exp (\ell(\theta, u; x) - \varepsilon f(\theta, u)) f(\theta, u) du}{\int \exp (\ell(\theta, u; x) - \varepsilon f(\theta, u)) du} \quad (2d)$$

If this derivative of $g(\theta, u, \varepsilon; x)$ is evaluated at $\varepsilon = 0$ given estimated fixed effects $\hat{\theta}$, it yields:

$$\frac{\partial}{\partial \varepsilon} (g(\hat{\theta}, u, \varepsilon; x)) = \frac{\int \exp (\ell(\hat{\theta}, u; x)) f(\hat{\theta}, u) du}{\int \exp (\ell(\hat{\theta}, u; x)) du} = E[\phi|x] \quad (2e)$$

which is our minimum variance unbiased estimator for $\phi = f(\hat{\theta}, u|x)$. A similar calculation shows that the second-order derivative of g , evaluated at $\varepsilon = 0$, yields the conditional variance $\operatorname{Var}[\phi|x]$. This suggests the following variance estimator:

$$\operatorname{Var}[\phi|x] = \frac{\partial^2}{\partial \varepsilon^2} (g(\hat{\theta}, u, \varepsilon; x)) + \frac{\partial}{\partial \theta} \frac{\partial}{\partial \varepsilon} (g(\hat{\theta}, u, \varepsilon; x))^T \operatorname{Var}(\hat{\theta}|x) \frac{\partial}{\partial \theta} \frac{\partial}{\partial \varepsilon} (g(\hat{\theta}, u, \varepsilon; x)) \quad (2f)$$

where the first summand on the right-hand side is the typical variance estimator for quantity ϕ arising from estimating random-effects u conditional upon data x and estimated fixed effects $\hat{\theta}$, while the second summand is the variance arising from conditioning on estimates of fixed effects (where $\operatorname{Var}(\hat{\theta}|x)$ is the variance for estimated fixed effects $\hat{\theta}$). Calculating $E[\phi|x]$ and $\operatorname{Var}[\phi|x]$ both require identifying the maximum likelihood estimate of fixed effects $\hat{\theta}$, as well as

evaluating derivatives of $g(\hat{\theta}, u, \varepsilon; x)$ with respect to ε , but do not require any optimization beyond that required to estimate $\hat{\theta}$. We therefore hypothesize that computing the bias-corrected estimator will be relatively fast relative to estimating fixed effects for most fisheries models.

We note that, in the following, we approximate all integrals using the Laplace approximation, as is done by default in ADMB-RE (Skaug and Fournier, 2006), Template Model Builder (Kristensen, 2014; Kristensen et al., In press), and the “glmer” function of the *lme4* package (Bates and Maechler, 2009) in the R statistical environment (R Core Development Team, 2013). Therefore, $g(\theta, u, \varepsilon; x)$ in (2c) is approximated as following a multivariate normal distribution, and the appropriateness of this assumption will vary among applications. The estimator in Eq. 2c-2e is implemented in TMB as a generic approach for estimating the expected value of any quantity ϕ that is derived from random effects u and fixed effects θ . Importantly, the epsilon estimator accounts for the variation in random effects, and has no effect when calculating a quantity using only fixed effects (which are treated as fixed quantities), i.e., $E[f(\hat{\theta})] = f(\hat{\theta})$. This epsilon estimator is also derived without any reference to Taylor series expansion (Eq. 1a). However, the epsilon estimator relies on the Laplace approximation when approximating the integral in the numerator and denominator of Eq. 2e, and this approximation is inexact whenever the likelihood function (conditional on fixed values for fixed effects) is not multivariate normal. Using the Laplace approximation in this manner may introduce an additional source of bias, and the performance of the Laplace approximation can be evaluated using numerical integration methods, e.g., comparison with Markov-chain Monte Carlo sampling (see example #2 below). Bias may also be introduced by other potential characteristics of the model including model misspecification (e.g., Thorson et al., 2015b), measurement errors in predictor variables (e.g., Draper and Smith, 1998), and endogeneity of predictor and response variables (e.g., Walters, 1985). We therefore interpret the epsilon estimator as a bias-correction algorithm (rather than providing unbiased estimates *per se*), and note that it will only be unbiased when other sources of bias are also absent.

2.3 Example 1: Bias correction when estimating recruitment in population models

Populations of marine fishes typically exhibit considerable variation in larval and juvenile survival rates, which leads to variation about the relationship between spawning biomass and resulting recruitment (Minto et al., 2008; Thorson et al., 2014a). Population models for marine fishes therefore typically estimate recruitment variability:

$$R_t = f(SB_t) \cdot \exp\left(r_t - \frac{\sigma_r^2}{2}\right) \quad (3a)$$

where $f(SB_t)$ is the recruitment expected on average to arise from spawning biomass SB_t and we refer to $\exp\left(r_t - \frac{\sigma_r^2}{2}\right)$ as the “recruitment multiplier”, which represents lognormal variation around the stock-recruitment function $f(SB_t)$. r_t is generally called the “recruitment deviation” for year t :

$$r_t \sim \text{Normal}(0, \sigma_r^2) \quad (3b)$$

where σ_r^2 is the variance in recruitment deviations over time, and the term $\exp\left[-\frac{\sigma_r^2}{2}\right]$ in the recruitment multiplier (Eq. 3a) is included to ensure that $f(SB_t)$ represents the mean of expected recruitment:

$$E[R_t] = E \left[f(SB_t) \cdot \exp \left(r_t - \frac{\sigma_r^2}{2} \right) \right] = f(SB_t) \quad (3c)$$

where the equality in (3c) follows from the definition of a lognormal distribution. We define $f(SB_t)$ as the mean of expected recruitment (rather than the median) to aid interpretation of parameters governing $f(SB_t)$ (e.g., the parameter representing average unfished recruitment).

However, recruitment deviations r_t are treated as a random effect, and recruitment R_t is a “derived quantity” calculated via a nonlinear function of random effects r_t . Some years have more or less information to inform estimates of recruitment deviations \hat{r}_t , causing the variance of estimated recruitment deviations $SE[\hat{r}_t]^2$ to change among years. Therefore, the appropriate formula for calculating the expected value of recruitment arising from estimated recruitment deviations will change among years. Responding to this concern, Methot and Taylor (2011) proposed a bias-correction algorithm, i.e., where the population model is estimated initially to obtain standard error estimates for each recruitment deviation \hat{r}_t , $\widehat{SE}[\hat{r}_t]$, and then the population model is fitted a second time with the following formula for recruitment:

$$\hat{R}_t = f(SB_t) \cdot \exp \left(\hat{r}_t - b_t \frac{\sigma_r^2}{2} \right) \quad (4a)$$

where the bias correction fraction b_t is calculated from the estimated standard error in the initial model fit:

$$b_t = 1 - \frac{\widehat{SE}[\hat{r}_t]^2}{\sigma_r^2} \quad (4b)$$

where (4a-4b) then reduce to:

$$\hat{R}_t = f(SB_t) \cdot \exp \left[\hat{r}_t - \left(\frac{\sigma_r^2}{2} - \frac{\widehat{SE}[\hat{r}_t]^2}{2} \right) \right] \quad (4c)$$

This formula has two intuitive properties:

1. When there is no information regarding recruitment deviation \hat{r}_t in year t , the recruitment deviation is shrunk towards zero, and also has a standard error that approaches the standard deviation of recruitment, $\widehat{SE}[\hat{r}_t]^2 \rightarrow \sigma_r^2$. In this case, $b_t \rightarrow 0$ and recruitment receives no bias correction, such that $R_t = f(SB_t)$.
2. When there is abundant information regarding recruitment deviation \hat{r}_t in year t , this parameter estimate will generally approach its true value $\hat{r}_t \sim Normal(0, \sigma_r^2)$ and its standard error will approach zero, $\widehat{SE}[\hat{r}_t]^2 \rightarrow 0$. In this case, $b_t \rightarrow 1$ and recruitment receives full bias correction, such that $E(\hat{R}_t) = f(SB_t)$.

Further details can be found in Methot and Taylor (2011).

This bias-correction algorithm can be derived independently of the rationale provided by Methot and Taylor (2011), given that we seek to calculate the expected value for recruitment \hat{R}_t given variation in the random variable \hat{r}_t . In this case, recruitment is derived from a stochastic process for recruitment variation:

$$\hat{R}_t = f(SB_t) \cdot \exp \left[\hat{r}_t - \frac{\sigma_r^2}{2} \right] \quad (5a)$$

but where the expected value of estimated recruitment is calculated given that estimation error follows a lognormal distribution:

$$E[\hat{R}_t] = f(SB_t) \cdot \exp \left[\hat{r}_t - \frac{\sigma_r^2}{2} \right] \cdot \exp \left[\frac{\widehat{SE}[\hat{r}_t]^2}{2} \right] \quad (5b)$$

where (4c) and (5b) are identical. Therefore, the bias correction developed by Methot and Taylor (2011) can be interpreted as a method for calculating the expected value of recruitment given the treatment of recruitment deviation \hat{r}_t as a random variable with variance $\widehat{SE}[\hat{r}_t]^2$.

In this first example, we conduct a simulation experiment testing the performance of the epsilon estimator implemented in TMB when calculating recruitment (a derived quantity) given estimated recruitment deviations (a random effect) in an age-structured population model. In each simulation replicate, we generate data from an age-structured simulation model for a fish population that has natural mortality rate $M=0.18 \text{ yr}^{-1}$ and von Bertalanffy growth rate $k=0.1 \text{ yr}^{-1}$. We simulate fishing for a single fleet with logistic, age-based selectivity, and fishing effort is simulated to increase and subsequently decrease as a function of fish population abundance (Thorson et al., 2013). We then simulate the collection of data including total catch, an index of fishery catch rates, and age-composition samples from fishery catch (see Appendix A for details). In each replicate, we fit an estimation model that uses the same equations for age-structured dynamics. The estimation model is implemented in Template Model Builder (code is publicly available at <https://github.com/James-Thorson/CCSRA>). To evaluate performance of the epsilon estimator, we compare bias-corrected estimates of recruitment arising from the epsilon estimator, $\widehat{E}[\exp(\hat{r}_t - \sigma_r^2/2)]$, with results from the Methot-Taylor bias correction method, $\exp(\hat{r}_t - b_t \sigma_r^2/2)$, which has been tested previously. We also include recruitment estimates without bias-correction, $\exp(\hat{r}_t)$, as a reference case to illustrate the magnitude of bias arising from ignoring bias-correction entirely.

We explore four configurations of this age-structured estimation model, which include different amounts of information regarding annual variations in recruitment:

1. *Catch curve*: When the estimation model is fitted to only a single year of age-composition samples and fishing mortality rates are assumed to be constant over time, the model is similar to a conventional catch curve that estimates fishery selectivity (Thorson and Prager, 2011), although it also estimates recruitment deviations (Millar, 2014).
2. *Stock reduction analysis (SRA)*: When the estimation model is provided with only a prior on final spawning biomass relative to average unfished spawning biomass as well as a time series of catch data, the model functions as a stock reduction analysis (Kimura et al., 1984).
3. *Catch-curve stock reduction analysis (CCSRA)*: When the estimation model is provided with a single year of age-composition samples and a time series of catch data, it functions as the recently proposed “catch-curve stock reduction analysis” (Thorson and Cope, 2015).
4. *Full data*: When the estimation model is provided with age-composition samples from the fishery for all years, a time series of catch information, and also an index of relative abundance, the model resembles a typical data-rich assessment model (Maunder and Punt, 2013).

All four configurations estimate a recruitment deviation for each of 20 simulated years, as well as for ages 1 through 20 in the first simulated year (i.e., a total of 40 recruitment deviations). All estimation models also fix the variance of recruitment deviations at its true value $\hat{\sigma}_r = 0.6$ to simplify interpretation of results (i.e., to avoid any potential impact of bias in estimating σ_r). This level of recruitment variation is consistent with the annual, statistically-independent variation estimated recently via meta-analysis (Thorson et al., 2014a). We evaluate estimation performance in each of the four configurations by comparing the estimated recruitment multiplier $E[\exp(\hat{r}_t - \frac{\sigma_r^2}{2})]$ with its true value:

$$Error_t = E \left[\exp \left(\hat{r}_t - \frac{\sigma_r^2}{2} \right) \right] - \exp \left(r_t - \frac{\sigma_r^2}{2} \right) \quad (6)$$

for each of the 40 estimated recruitment deviations in each simulation replicate, where the first term on the right-hand side is the estimator for the bias-corrected recruitment multiplier, and the second term is the true recruitment multiplier. An unbiased model will have an average recruitment multiplier $E \left[\exp \left(\hat{r}_t - \frac{\sigma_r^2}{2} \right) \right] = 1.0$ when averaged across all years in a given simulation replicate, and an average error $Error_t = 0.0$ when averaging across all years and simulation replicates. The Methot-Taylor bias correction approach was initially developed and tested for a data-rich context, so we hypothesize that both the Methot-Taylor and epsilon estimators will perform well in this instance. By contrast, the stock reduction analysis setting has very little information regarding individual deviations in recruitment, so the Methot-Taylor method is not expected to perform well in this instance. The catch curve and catch-curve stock reduction analysis models represent an intermediate level of information, and are likely to have performance between that in the full data and stock-reduction analysis models.

2.4 Example 2: Estimation of population abundance in geostatistical models

We also demonstrate the use of the epsilon method with a model for which there is no previously proposed bias-correction algorithm. We therefore apply the method to a recently developed geostatistical model for index standardization (Thorson et al., 2015a). This geostatistical model estimates total population abundance B_t in year t as:

$$B_t = \sum_{j=1}^{n_j} a_j \text{logit}^{-1} \left(d_t^{(p)} + \omega_j^{(p)} + \varepsilon_{j,t}^{(p)} \right) \exp \left(d_t^{(\lambda)} + \omega_j^{(\lambda)} + \varepsilon_{j,t}^{(\lambda)} \right) \quad (7)$$

where $\text{logit}^{-1}(x)$ is the logistic transformation of value x , a_j is the area associated with polygon j , $d_t^{(p)}$ and $d_t^{(\lambda)}$ are fixed effects representing variation in encounter probabilities (p) and positive catch rates (λ) among years, $\omega_j^{(p)}$ and $\omega_{j(i)}^{(\lambda)}$ are random effects representing spatial variation, and $\varepsilon_{j,t}^{(p)}$ and $\varepsilon_{j,t}^{(\lambda)}$ are random effects representing spatiotemporal variation. We apply this model to data from the West Coast groundfish bottom trawl survey (Bradburn et al., 2011) applied to data for arrowtooth flounder, and use a relatively coarse spatial resolution to approximate spatial variation in population density ($n_j=100$, i.e., using 100 knots). We seek to calculate the expected value of total abundance $E[B_t]$ and must account for the nonlinear transformation of random effects (the code is publicly available at https://github.com/nwfsc-assess/geostatistical_delta-GLMM).

We seek to demonstrate that the epsilon bias-correction estimator propagates uncertainty in random effects when calculating a bias-corrected estimator for derived quantities. We also seek to demonstrate the estimate of variance for the bias-corrected estimator (Eq. 2f). We therefore compare the epsilon estimator with an alternative, sample-based bias correction estimator. Specifically, Markov-chain Monte Carlo (MCMC) sampling can be run when sampling across all possible values of random effects while holding fixed effects at their maximum likelihood estimates. In this manner, MCMC samples directly from the expectation of a derived variable $B = f(A)$ calculated as a function of random effects A (Eq. 1b). We specifically run Hamiltonian (a.k.a. hybrid) MCMC using a leapfrog solution to the differential equation for trajectories within each Hamiltonian sample (see Neal (2011) for details). We generate samples using a single Markov chain after 2,000 leapfrog updates per sample, where each leapfrog update has a total distance of 0.02, and using 200 samples for burn-in followed by 1,000 monitored samples. We check for evidence of non-convergence using traceplots, autocorrelation, and Geweke diagnostics for each sampled parameter (Plummer et al., 2006), and found no such

evidence. For each sample, we calculate total population abundance B_t , and the bias-corrected value $E[B_t]$ and variance $Var[B_t]$ as its average and variance for all 1,000 retained samples. This MCMC-based bias-correction estimator represents an alternative, generic approach to bias correction in population models. However, we introduce the epsilon estimator in this study because MCMC-based estimators require checking for evidence of non-convergence for each individual application (and convergence may not be feasible to achieve for some large models), whereas the epsilon estimator provides an off-the-shelf approach to bias correction for nonlinear mixed-effect models estimated using maximum likelihood techniques.

3. Results

We illustrate results in the first replicate of our simulation experiment involving the age-structured model, showing estimates of the recruitment multiplier either without bias correction $\exp(\hat{r}_t)$, using the Methot-Taylor bias correction method $\exp(\hat{r}_t - b_t \sigma_r^2/2)$, or using the epsilon bias-correction estimator $\hat{E}[\exp(\hat{r}_t - \sigma_r^2/2)]$ compared with the true recruitment multiplier $\exp(r_t - \sigma_r^2/2)$, (Fig. 1). The “Full data” scenario in particular illustrates the importance of bias correction, where the recruitment multiplier without bias correction is nearly 17% higher than the true value. In this simulation replicate, the Methot-Taylor estimator is somewhat less biased (7% higher than the true value) while the epsilon estimator is least biased (within 2% of the true value). The Methot-Taylor also splits the difference between the epsilon estimator and the estimates without bias correction for the catch-curve scenario. For the SRA scenario, however, the Methot-Taylor estimator is more biased than the estimates without bias-correction, while the epsilon-estimator has relatively little bias (<4% higher) in each case, and the Methot-Taylor and estimates without bias correction have similar bias for the CCSRA scenario.

We next summarize the error in estimates of recruitment multipliers without bias correction $\exp(\hat{r}_t) - \exp(r_t - \sigma_r^2/2)$, using the Methot-Taylor estimator $\exp(\hat{r}_t - \frac{b_t \sigma_r^2}{2}) - \exp(r_t - \sigma_r^2/2)$ or the epsilon estimator $E[\exp(\hat{r}_t - \frac{\sigma_r^2}{2})] - \exp(r_t - \sigma_r^2/2)$ (Fig. 2). This summary illustrates many of the same patterns shown previously for the first replicate of the experiment (i.e., Fig. 1), where the Methot-Taylor estimator splits the difference between the epsilon estimator and the estimate without bias correction for the catch curve and full data scenarios. The Methot-Taylor estimator on average eliminates more than half of the bias in the full data scenario, while the epsilon estimator has a residual positive bias of less than 1%. For the CCSRA and SRA scenarios, by contrast, the Methot-Taylor method has a positive bias of 10-15%, and this bias peaks in year 5. This bias is largely eliminated by the epsilon estimator, which has an average error within $0 \pm 2\%$ for all scenarios. We also illustrate error in estimates of spawning output relative to average unfished levels, which is another quantity that is derived from recruitment deviations in each model (Fig. 3). In the full-data and CCSRA scenarios, the Methot-Taylor estimator has a positive bias of nearly 3%, while the epsilon estimator again has little bias. All estimators are biased for the catch curve scenario (because fishing mortality changes over time in violation of model assumptions) and the stock-reduction analysis scenario (because the prior on relative spawning biomass is misspecified relative to the true relative spawning biomass).

Finally, we demonstrate the epsilon method when applied to a model for which a bias-correction estimator has not previously been developed. Using the geostatistical model to estimate an index of abundance of arrowtooth flounder (Fig. 4), the epsilon method resulted in an index that has almost identical trend to the plug-in estimator. However, the bias-correction

method estimates average abundance nearly 10% greater than the plug-in estimator, and this increase is also reflected in the sample-based estimator. The difference in scale between plug-in and bias-corrected estimators would be important if the abundance estimate were used as an “absolute” index of abundance, i.e., having a catchability coefficient of one. Finally, the epsilon bias-correction estimator has wider confidence intervals than the sample-based bias-correction estimator, but this difference in width is small for all years.

4. Discussion

Mixed-effects are increasingly used in population models to account for variation in demographic processes over time, space, and among individuals (Thorson and Minto, 2015). For example, variation over time in fishery selection or catchability can be included by treating selectivity parameters as random effects that follow a random walk (Nielsen and Berg, 2014; Wilberg et al., 2010). Similarly, natural mortality and individual growth rates may change over time due to changes in environmental or community interactions (Mäntyniemi et al., 2013; Thorson and Minto-Vera, In press). Furthermore, there is increasing interest in spatially-explicit analysis and monitoring of species distribution models and fishery impacts, and spatial models can generally be estimated via mixed-effects methods (Cressie and Wikle, 2011). In particular, computational advances in Bayesian estimation of mixed-effects models has led to an explosion of research regarding mixed-effects models (Kéry and Schaub, 2012; Royle and Dorazio, 2008).

In this paper, we have demonstrated that a new epsilon estimator improves upon the bias-correction algorithm previously introduced by Methot and Taylor (2011) for models that treat recruitment deviations as random effects. Specifically, we have shown that the new epsilon estimator decreases bias relative to the Methot-Taylor estimator in both data-rich and data-poor scenarios, but that the two estimators are more similar in the data-rich context. In the case of the stock reduction analysis, the Methot-Taylor method has substantial (>10%) positive bias for years -10 through 10, and the magnitude of this bias is greater than for the model without bias correction. This increase in bias relative to the scenario without bias correction arises because the standard errors during this period are greater than the standard deviation of recruitment (and therefore $b_t < 0$ in Eq. 4a). This occurs in turn because the stock-reduction analysis model has highly negative covariance between recruitment in adjacent years, and therefore the covariation among estimates of recruitment deviations is not well summarized by the standard error in each year. A similar positive bias is seen for the Methot-Taylor estimator in the CCSRA scenario, and is not present for the epsilon estimator. We therefore hypothesize that the Methot-Taylor estimator is not well suited for bias correction in instances with high covariation among random effects (i.e., when data are only informative about abundance changes over a multiyear period), and that the epsilon estimator is appropriate for these instances. Importantly, improvements in estimates of recruitment using the epsilon estimator translate to improved estimates of spawning biomass relative to average unfished levels in the CCSRA and full-data scenarios. Estimates of spawning biomass are important when determining stock status and implementing harvest control rules during fisheries management, and biased estimation therefore signals the potential importance of improvements in bias-correction methods.

We have also shown that the epsilon method results in an important difference in the absolute scale of population biomass when estimating an index of abundance using a spatial index standardization model, relative to a simple plug-in estimator for population biomass. Thorson et al. (2015a) acknowledged the importance of developing bias-correction algorithms for geostatistical models, but found that neglecting bias-correction did not greatly affect relative

trends in abundance (i.e., instances where the catchability coefficient, which scales the estimated index to estimated abundance in a population dynamics model, is freely estimated). However, bias-correction will be more important when a spatial model is used to estimate an absolute index of abundance (i.e., in instances where the catchability coefficient is fixed at 1.0), and bias-correction may account for observed differences in scale between estimates of population biomass arising from geostatistical and stratified index standardization models for West Coast groundfishes (see Fig. 3 of Thorson et al. (2015a)).

Mixed-effects models can be estimated using both Bayesian and maximum likelihood methods (Cressie and Wikle, 2011). Despite advances in Bayesian implementation of mixed-effects models, we believe that there are many modelling tasks where maximum likelihood estimation is more appropriate. In particular, “non-informative” priors may not be available for some parameters in population models and existing theory may not be sufficient to generate informative priors, which makes Bayesian estimation problematic in these cases (Schweder, 1998). In such cases, maximum likelihood mixed-effect estimation is more appropriate than Bayesian estimation methods. However, maximum likelihood mixed-effect methods have previously been hindered, due to a lack of easy-to-use tools for generating unbiased predictions of derived variables.

Bias arising from nonlinear transformation of random effects will generally have larger magnitude when the random effect variance is large or when the transformation is highly nonlinear (see Eq. 1a). Therefore bias-correction will be particularly important in these cases and will have no effect when the transformation is exactly linear. Bias-correction can in some cases increase estimation variance (i.e., the root-mean-squared error of empirical Bayes estimates relative to their true value) and decision theory will be necessary to distinguish cases where this additional variance is worse than bias. In particular, the relative importance of bias and estimation variance (e.g., when estimating a relative index of abundance) on management outcomes will depend upon the structure of population models used to provide management advice, delays between science advice and management actions, and the management actions that are used to regulate the fishery (Haltuch and Punt, 2011; Shertzer and Prager, 2007). Increased availability of bias-correction estimators in statistical software will hopefully facilitate simulation studies comparing management performance when using models with and without bias correction estimators, and we therefore recommend implementation of the epsilon estimator in other software involving the Laplace approximation (e.g., the *lme4* package in R, and ADBM-RE). However, we have designed the epsilon estimator to account only for bias arising from nonlinear transformation of the variance of random effects, and it will not account for other potential sources of bias. In particular, generic approaches to account for bias arising from nonindependence of process errors and predictors (known as “time series bias”, Walters (1985)), correlations between process errors and sampling intensity (known as “preferential sampling”, Diggle et al. (2010)), and model misspecification (“model error”, Thorson et al. (2014b)) remain important topic of generic research in mixed-effects models.

We conclude by appealing for continued research regarding improvements in estimation methods for mixed-effects models, whether implemented using Bayesian or maximum likelihood techniques. Mixed-effects models represent an intuitive procedure for eliciting the covariation among data arising from a model involving unobserved variables, and thus provide a unifying framework for many types of models in fisheries. Mixed-effects estimation is already revolutionizing quantitative methods throughout ecology (Clark and Gelfand, 2006; Cressie et al., 2009; Royle and Dorazio, 2008). We hope that computational and statistical improvements

such as the epsilon estimator facilitate continued learning about ecological processes at large spatial and temporal scales, where experimental evidence is often difficult to accumulate.

5. Conclusions

In this paper, we have introduced a new, generic bias-correction estimator for derived quantities in fisheries models that include mixed effects. This epsilon estimator is now implemented and distributed in the TMB software, and can be used generically in mixed-effect models to reduce bias resulting from nonlinear transformation of random effects. We have shown that the epsilon method improves upon a previous method for correcting bias when estimating recruitment deviations in age-structured stock assessment models. We have also shown that the epsilon method results in important differences in the estimated scale of population abundance relative to the plug-in estimator when estimating abundance indices, while providing similar results to an alternative sample-based bias-correction method. We therefore recommend that studies involving the use of maximum likelihood estimation for mixed-effects models explore using the epsilon-estimator when reporting derived quantities calculated as a nonlinear function of random effects.

6. Acknowledgements

We gratefully acknowledge comments and discussion with H. Skaug, who significantly improved this research and manuscript. We also thank R. Methot and two anonymous reviewers for comments on an earlier draft.

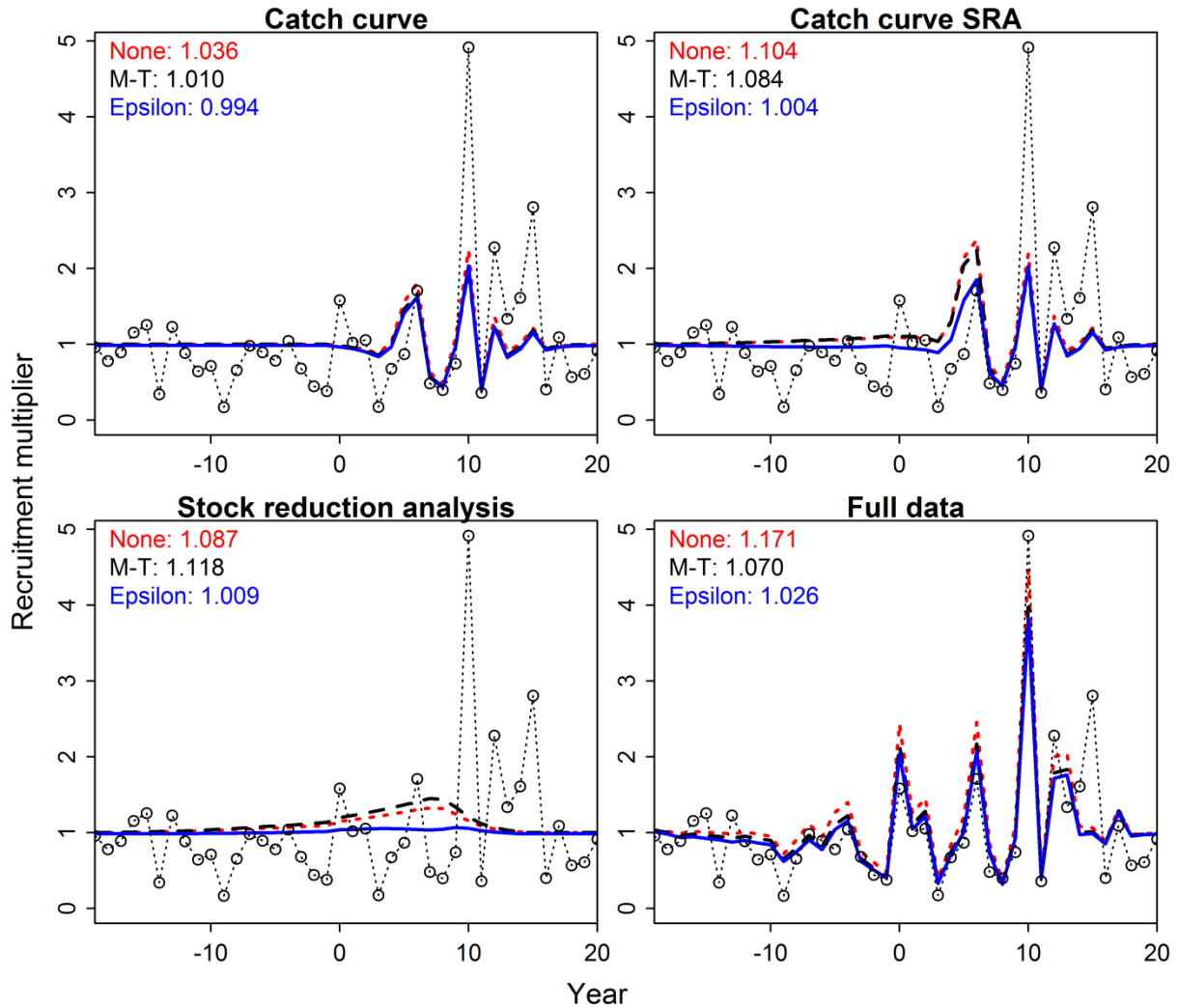
References

- Bates, D., Maechler, M., 2009. lme4: Linear mixed-effects models using S4 classes.
- Bradburn, M.J., Keller, A.A., Horness, B.H., 2011. The 2003 to 2008 US West Coast bottom trawl surveys of groundfish resources off Washington, Oregon, and California: estimates of distribution, abundance, length, and age composition (NOAA Tech. Memo. No. NMFS-NWFSC-114). US Department of Commerce, National Oceanic and Atmospheric Administration, National Marine Fisheries Service, Northwest Fisheries Science Center, Seattle, WA.
- Clark, J.S., Gelfand, A.E., 2006. A future for models and data in environmental science. *Trends Ecol. Evol.*, Twenty years of TREE - part 2 21, 375–380. doi:10.1016/j.tree.2006.03.016
- Cressie, N., Calder, C.A., Clark, J.S., Hoef, J.M.V., Wikle, C.K., 2009. Accounting for uncertainty in ecological analysis: the strengths and limitations of hierarchical statistical modeling. *Ecol. Appl.* 19, 553–570.
- Cressie, N., Wikle, C.K., 2011. *Statistics for spatio-temporal data*. John Wiley & Sons, Hoboken, New Jersey.
- de Valpine, P., 2009. Shared challenges and common ground for Bayesian and classical analysis of hierarchical statistical models. *Ecol. Appl.* 19, 584–588.
- Diggle, P.J., Menezes, R., Su, T., 2010. Geostatistical inference under preferential sampling. *J. R. Stat. Soc. Ser. C Appl. Stat.* 59, 191–232.
- Draper, N.R., Smith, H., 1998. *Applied regression analysis*, Third. ed. Wiley-Interscience.
- Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B., 2003. *Bayesian Data Analysis*, Second Edition. Chapman & Hall, Boca Raton, FL.
- Haltuch, M.A., 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, 912–926.
- Hilborn, R., Mangel, M., 1997. *The Ecological Detective: Confronting Models with Data*, 1st ed. Princeton University Press, Princeton, New Jersey.
- Kéry, M., Schaub, M., 2012. *Bayesian population analysis using WinBUGS: a hierarchical perspective*. Academic Press, Waltham, MA.
- Kimura, D.K., Balsiger, J.W., Ito, D.H., 1984. Generalized stock reduction analysis. *Can. J. Fish. Aquat. Sci.* 41, 1325–1333.
- Kristensen, K., 2014. TMB: General random effect model builder tool inspired by ADMB. R package version 1.6.2. Available from: <https://cran.r-project.org/web/packages/TMB/index.html>
- Kristensen, K., Nielsen, A., Berg, C.W., Skaug, H., In press. Template Model Builder TMB. *J. Stat. Softw.*
- Kristensen, K., Thygesen, U.H., Andersen, K.H., Beyer, J.E., 2014. Estimating spatio-temporal dynamics of size-structured populations. *Can. J. Fish. Aquat. Sci.* 71, 326–336. doi:10.1139/cjfas-2013-0151
- Mäntyniemi, S., Uusitalo, L., Peltonen, H., Haapasaari, P., Kuikka, S., 2013. Integrated, age-structured, length-based stock assessment model with uncertain process variances, structural uncertainty, and environmental covariates: case of Central Baltic herring. *Can. J. Fish. Aquat. Sci.* 70, 1317–1326. doi:10.1139/cjfas-2012-0315
- Maunder, M.N., Punt, A.E., 2013. A review of integrated analysis in fisheries stock assessment. *Fish. Res.* 142, 61–74.

- Methot, R.D., Taylor, I.G., 2011. Adjusting for bias due to variability of estimated recruitments in fishery assessment models. *Can. J. Fish. Aquat. Sci.* 68, 1744–1760.
- Millar, R.B., 2014. A better estimator of mortality rate from age-frequency data. *Can. J. Fish. Aquat. Sci.* 72, 364–375. doi:10.1139/cjfas-2014-0193
- Minto, C., Mills Flemming, J., Britten, G.L., Worm, B., Rose, K., 2014. Productivity dynamics of Atlantic cod. *Can. J. Fish. Aquat. Sci.* 71, 203–216. doi:10.1139/cjfas-2013-0161
- Minto, C., Myers, R.A., Blanchard, W., 2008. Survival variability and population density in fish populations. *Nature* 452, 344–347. doi:10.1038/nature06605
- Munch, S.B., Kottas, A., Mangel, M., 2005. Bayesian nonparametric analysis of stock-recruitment relationships. *Can. J. Fish. Aquat. Sci.* 62, 1808–1821.
- Neal, R.M., 2011. Chapter 5: MCMC using Hamiltonian dynamics, in: *Handbook of Markov Chain Monte Carlo*. Chapman & Hall/CRC Press, Boca Raton, FL.
- Nielsen, A., Berg, C.W., 2014. Estimation of time-varying selectivity in stock assessments using state-space models. *Fish. Res.* 158, 96–101.
- Plummer, M., Best, N., Cowles, K., Vines, K., 2006. CODA: Convergence diagnosis and output analysis for MCMC, *R News* 6, 7–11. Available at: https://www.r-project.org/doc/Rnews/Rnews_2006-1.pdf
- R Core Development Team, 2013. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Royle, J.A., Dorazio, R.M., 2008. *Hierarchical modeling and inference in ecology: the analysis of data from populations, metapopulations and communities*, 1st ed. Academic Press, London.
- Schaub, M., Gimenez, O., Sierro, A., Arlettaz, R., 2007. Use of integrated modeling to enhance estimates of population dynamics obtained from limited data. *Conserv. Biol.* 21, 945–955.
- Schweder, T., 1998. Fisherian or Bayesian methods of integrating diverse statistical information? *Fish. Res.* 37, 61–75.
- Shelton, A.O., Satterthwaite, W.H., Beakes, M.P., Munch, S.B., Sogard, S.M., Mangel, M., 2013. Separating intrinsic and environmental contributions to growth and their population consequences. *Am. Nat.* 181, 799–814.
- Shertzer, K.W., Prager, M.H., 2007. Delay in fishery management: diminished yield, longer rebuilding, and increased probability of stock collapse. *ICES J. Mar. Sci.* 64, 149.
- Skaug, H., Fournier, D., 2006. Automatic approximation of the marginal likelihood in non-Gaussian hierarchical models. *Comput. Stat. Data Anal.* 51, 699–709.
- Thorson, J.T., Cope, J.M., 2015. Catch curve stock-reduction analysis: An alternative solution to the catch equations. *Fish. Res.* 171, 33–41. doi:10.1016/j.fishres.2014.03.024
- Thorson, J.T., Jensen, O.P., Zipkin, E.F., 2014a. How variable is recruitment for exploited marine fishes? A hierarchical model for testing life history theory. *Can. J. Fish. Aquat. Sci.* 71, 973–983. doi:10.1139/cjfas-2013-0645
- Thorson, J.T., Minto-Vera, C., In press. Relative magnitude of cohort, age, and year effects on size at age of exploited marine fishes. *Fish. Res.*
- Thorson, J.T., Minto, C., 2015. Mixed effects: a unifying framework for statistical modelling in fisheries biology. *ICES J. Mar. Sci. J. Cons.* 72, 1245–1256. doi:10.1093/icesjms/fsu213
- Thorson, J.T., Minto, C., Minto-Vera, C., Kleisner, K., Longo, K., 2013. A new role of effort dynamics in the theory of harvest populations and data-poor stock assessment. *Can. J. Fish. Aquat. Sci.* 70, 1829–1844. doi:10.1139/cjfas-2013-0280

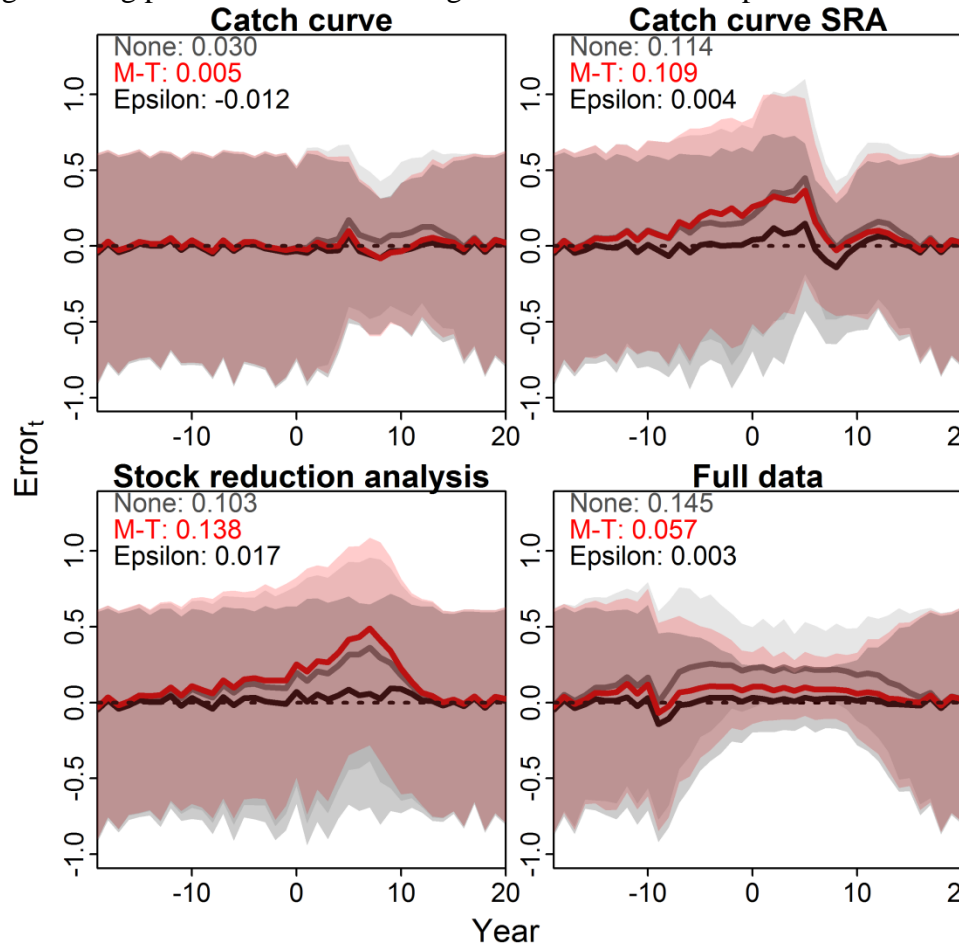
- Thorson, J.T., Ono, K., Munch, S.B., 2014b. A Bayesian approach to identifying and compensating for model misspecification in population models. *Ecology* 95, 329–341. doi:10.1890/13-0187.1
- Thorson, J.T., Prager, M.H., 2011. Better catch curves: Incorporating age-specific natural mortality and logistic selectivity. *Trans. Am. Fish. Soc.* 140, 356–366.
- Thorson, J.T., Shelton, A.O., Ward, E.J., Skaug, H.J., 2015a. Geostatistical delta-generalized linear mixed models improve precision for estimated abundance indices for West Coast groundfishes. *ICES J. Mar. Sci. J. Cons.* 72, 1297–1310. doi:10.1093/icesjms/fsu243
- Thorson, J. T., Skaug, H.J., Kristensen, K., Shelton, A.O., Ward, E.J., Harms, J.H., and Benante, J.A. 2015b. The importance of spatial models for estimating the strength of density dependence. *Ecology*, 96: 1202–1212.
- Tierney, L., Kass, R.E., Kadane, J.B., 1989. Fully exponential Laplace approximations to expectations and variances of nonpositive functions. *J. Am. Stat. Assoc.* 84, 710–716.
- Walters, C.J., 1985. Bias in the estimation of functional relationships from time series data. *Can. J. Fish. Aquat. Sci.* 42, 147–149.
- Walters, C.J., Martell, S.J.D., 2004. *Fisheries Ecology and Management*. Princeton University Press, Princeton, New Jersey.
- Wilberg, M.J., Thorson, J.T., Linton, B.C., Berkson, J., 2010. Incorporating time-varying catchability into population dynamic stock assessment models. *Rev. Fish. Sci.* 18, 7–24.

643 Fig. 1 – Results for a single replicate of the simulation experiment, showing the recruitment
 644 multiplier with no bias correction $\exp(\hat{r}_t)$ (thick dotted red line), using the Methot-Taylor bias
 645 correction method $\exp(\hat{r}_t - b_t \sigma_r^2 / 2)$ (thick dashed black line) or the epsilon method
 646 $\hat{E}[\exp(\hat{r}_t - \sigma_r^2 / 2)]$ (thick solid blue line) compared with the true recruitment multiplier
 647 $\exp(r_t - \sigma_r^2 / 2)$ (thin dotted black line and circles), where the x-axis shows the simulation year
 648 (year 0 is age-0 recruits in the first year, and year $-x$ is age x in the first year). For each scenario,
 649 the average recruitment multiplier is shown in the upper-left corner for each method (M-T
 650 signifies the Methot-Taylor method), and a well-behaved estimator will have a value near 1.000,
 651 see online for color-coded version)

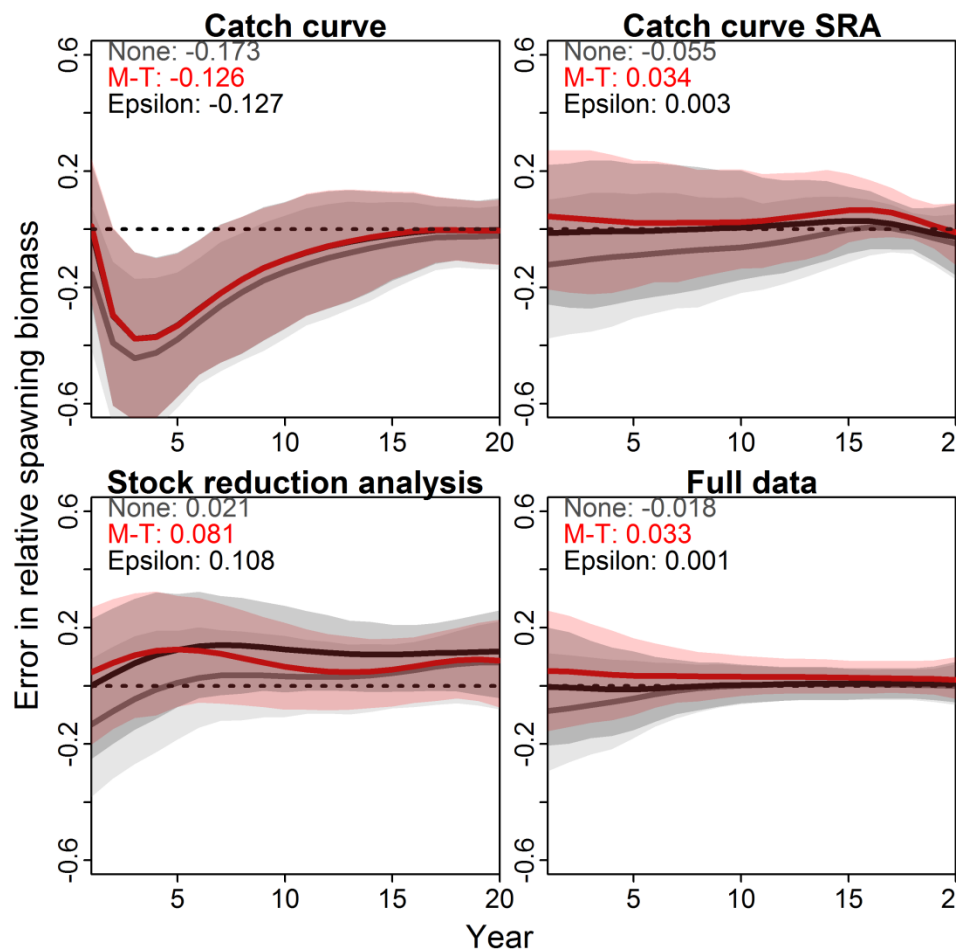


652

Fig. 2 – Median (solid line) and 80% simulation interval (shaded area) for the error in estimating recruitment multipliers (defined in Eq. 6, where the dashed line shows $Error_t=0.0$) with no bias correction (grey line and shading), using the Methot and Taylor (2011) bias-correction approach (red line and shading), or using the epsilon-method correction (black line and shading) for the four candidate models (clockwise from top-left: catch curve, catch-curve stock reduction analysis, a fully age-structured model with index and age compositional sampling data, and stock reduction analysis), where the x-axis shows the simulation year (year 0 is age-0 recruits in the first year, and year $-x$ is age x in the first year). The text in each panel gives the average error $Error_t$ when aggregating across all years and replicates, while the solid line can be used to identify individual years that have on average a positive or negative bias given the data generating process used in the design of the simulation experiment.

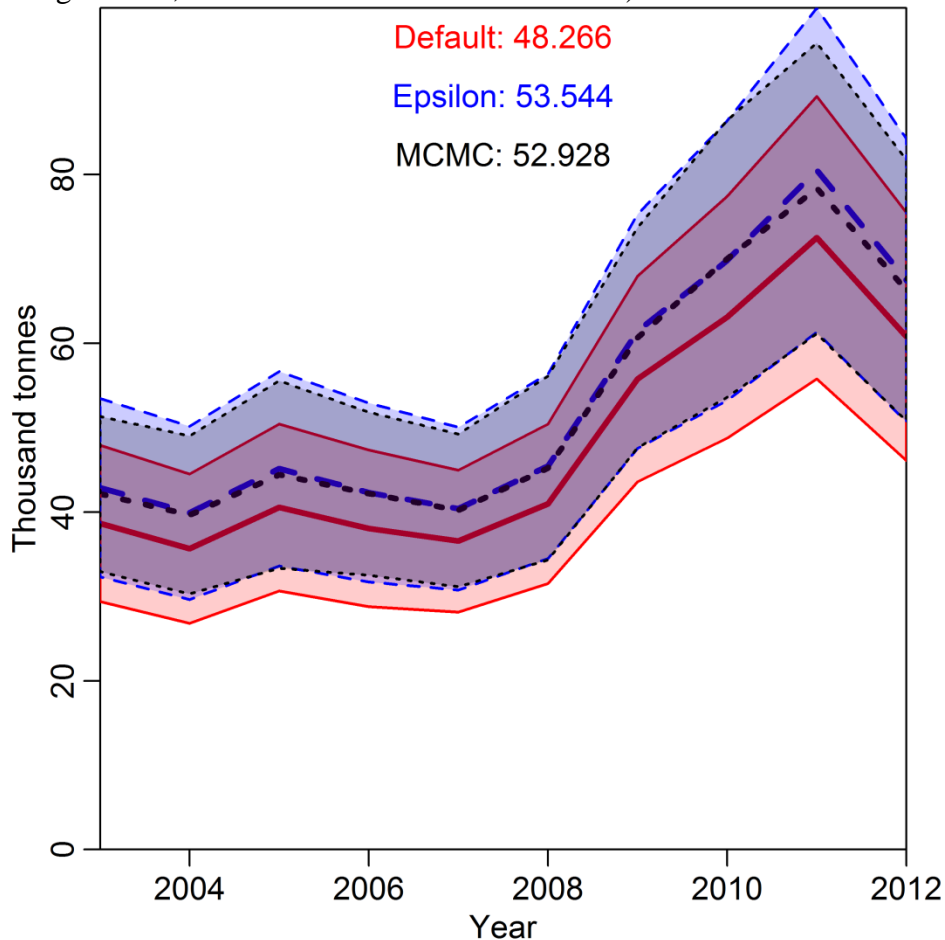


666 Fig. 3 – Median (solid line) and 80% simulation interval (shaded area) for the error in estimating
 667 spawning biomass relative to average unfished levels (see caption for Fig. 2 for details regarding
 668 figure layout)



669

Fig. 4 – Plug-in estimator for total abundance of arrowtooth flounder 2003-2012 off the U.S. West Coast (solid red lines), compared with bias-corrected estimates using the epsilon (dashed blue lines) or Markov-Chain Monte Carlo (dotted black lines) estimators when fitting a geostatistical index standardization model (the thick is the maximum likelihood estimate and the area encompassed by the thin lines is the 95% confidence interval for each model, where the text indicates the average biomass across all years 2003-2012 to assess changes in absolute scale among models, see online for color-coded version).



Supplementary material for on-line publication only

[Click here to download Supplementary material for on-line publication only: Appendices_v2-1.docx](#)