Can autocorrelated recruitment be estimated using integrated assessment models and how does it affect population forecasts?

Kelli F. Johnson[1,*], Elizabeth Councill[1,2,§], James T. Thorson[2], Elizabeth Brooks[3], Richard D. Methot[4], André E. Punt[1]

[1]School of Aquatic and Fishery Sciences, University of Washington, Box 355020, Seattle, WA 98195-5020, USA
[2]Fishery Resource Analysis and Monitoring Division, Northwest Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, 2725 Montlake Blvd. East, Seattle, WA 98112, USA
[3]Northeast Fisheries Science Center, 166 Water Street, Woods Hole, MA 02543, USA
[4]NOAA Senior Scientist for Stock Assessments, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, 2725 Montlake Blvd. East, Seattle, WA 98112, USA

[*]Corresponding author:
telephone: +1 206 543 4270; fax: +1 206 616 8689; email: kfjohns@uw.edu
[§]Present address: Cooperative Institute for Marine and Atmospheric Studies, University of Miami, 4600 Rickenbacker Causeway, Miami, Florida 33149, USA

Abstract

The addition of juveniles to marine populations (termed "recruitment") is highly variable due to variability in the survival of fish through larval and juvenile stages. Recruitment estimates are often large or small for several years in a row (termed "autocorrelated" recruitment). Autocorrelated recruitment can be due to numerous factors, but typically is attributed to multi-year environmental drivers affecting early life survival rates. Estimating the magnitude of recruitment autocorrelation within a stock assessment model and examinations on its effect on the quality of forecasts of spawning biomass within stock assessments is uncommon. We used a simulation experiment to evaluate the estimability of autocorrelation within a stock assessment model over a range of levels of autocorrelation in recruitment deviations. The precision and accuracy of estimated autocorrelation, and the ability of an integrated age-structured stock assessment framework to forecast the true dynamics of the system, were compared for scenarios where the autocorrelation parameter within the assessment was fixed at zero, fixed at its true value, internally estimated within the integrated model, or input as a fixed value determined using an external estimation procedure that computed the sample autocorrelation of estimated recruitment deviations. Internal estimates of autocorrelation were biased toward extreme values (i.e., towards 1.0 when true autocorrelation was positive and -1.0 when true autocorrelation was negative). Estimates of autocorrelation obtained from the external estimation procedure were nearly unbiased. Forecast performance was poor (i.e., true biomass outside the predictive interval for the forecasted biomass) when autocorrelation was ignored, but was non-zero in the simulation. Applying the external estimation procedure generally improved forecast performance by decreasing forecast error and improving forecast interval coverage. However, estimates of autocorrelation were shown to degrade when fewer than 40 years of recruitment estimates were available.

Keywords: autocorrelated recruitment; integrated stock assessment model; statistical catch at age; rebuilding plan; population forecast

1. Introduction

Under the United States Magnuson-Stevens Fishery Conservation and Management Act (United States Public Law 104-297), all stocks included in United States Fishery Management Plans must have target and limit reference points and forecasts of the level of catch (annual catch limit) that will prevent overfishing. Protocols for calculating annual catch limits in a way that will prevent overfishing with a specified probability have been developed (Shertzer et al., 2008), but are dependent on the quality of forecast precision. Further, all overfished stocks must have a rebuilding plan. Rebuilding plans involve specifying management measures to rebuild the stock to a biomass associated with maximum sustainable yield (β_{MSY}) within 10 years (or, if rebuilding within 10 years is impossible, then one generation time plus the median time for rebuilding in the absence of fishing). Legally, rebuilding plans must be more likely than not to succeed, i.e., be based upon a probabilistic forecast of future population dynamics given the agreed level of fishing that implies recovery with ≥ 50% probability.

Stock assessment models represent the link between collected data and scientific advice in fisheries management. Assessments are expected to use fits to historical data and prescribed harvest policies to forecast future stock abundance and catch levels. These predicted "Acceptable Biological Catches" must account for scientific uncertainty and ensure ≤ 50% probability that overfishing will occur (Methot et al., 2013). Variability in recent recruitment to the stock is a major contribution to this scientific uncertainty. As the United States National Marine Fisheries Service (NMFS) works to reduce the number of overfished stocks, projection success is being examined more critically, and the accuracy of probabilistic forecasts in rebuilding plans is receiving increased research attention (Neubauer et al., 2013; NRC, 2013).

Reference points and rebuilding forecasts are often estimated using a stock assessment model that treats fluctuations in recruitment as a random process around a prediction derived from a presumed relationship between spawning biomass and recruits (Clark, 1993; Methot and Wetzel, 2013). Stock assessments are increasingly conducted using "integrated" population dynamics models that typically incorporate many data types, including samples of compositional data from fisheries and surveys, indices of abundance, and information regarding total fishery harvests (Maunder and Punt, 2013). These data are combined to estimate values for population productivity (parameters in the stock-recruitment relationship) and status (spawning biomass in each year relative to reference points). Probabilistic forecasts of future population dynamics can then be made given assumed fishing mortality rates.

Recent studies illustrate that recruitment for many fishes is non-random over time and includes high and low periods (Hollowed et al., 2001; Szuwalski et al., 2014; Thorson et al., 2014). These periods could be driven by environmental factors acting on recruit survival (Wilderbuer et al., 2002), adult reproductive output (Jørgensen et al., 2006), or both simultaneously (Okamoto et al., 2012; Wooster and Bailey, 1989), or changes in the abundance of predators (Bailey, 2000). Ideally, researchers can identify measureable environmental factors that are correlated with recruitment deviations or regime shifts, and which can be forecast into the future (Haltuch and Punt, 2011). If an environmental factor that helps predict future recruitment can be identified, it can then be used to inform rebuilding forecasts (Holt and Punt, 2009; Punt, 2011) and reference point calculations (Lindegren and Checkley, 2013). If an environmental factor cannot be identified, population forecasts are sometimes calculated for various "states-of-nature", where each state-of-nature depends upon a hypothetical scenario for expected future recruitment (e.g., high, average, and low productivity scenarios; Peterman and Anderson, 1999).

When correlated measurable environmental factors remain unidentified, the influence of regime shifts can still be accounted for by invoking autocorrelation in future recruitment deviations (i.e., where future recruitment deviations are greater or less than zero for many years in a sequence). Including "autocorrelated recruitment" in the population dynamics model may result in wider forecast intervals (i.e., less precise) compared with the case in which recruitment is assumed to follow a white-noise process. This wider forecast interval may, in some cases, have better statistical coverage (e.g., a 75% forecast interval that contains the true value 75% of the time) than forecasts that do not account for autocorrelation in recruitment. Well-calibrated statistical coverage is a pre-requisite of probabilistic methods used for forecasting and reference point determination (Shertzer et al., 2008).

In this study, we explore and evaluate the performance of population forecasts obtained from an integrated, age-structured assessment model when recruitment is autocorrelated. We conduct a simulation experiment using a design involving six plausible levels of autocorrelation in recruitment deviations (ρ) and four alternative configurations for estimating ρ in the assessment model. We explore estimation performance by answering two questions:

1. How well can the magnitude of autocorrelation be estimated? and
2. Does accounting for autocorrelation improve the accuracy and predictive coverage of forecasts compared with ignoring autocorrelation in recruitment deviations?

We conclude by outlining a practical strategy to test and account for autocorrelated recruitment when generating forecasts in real-world assessment models.

2. Methods

We conducted a simulation experiment using the Stock Synthesis (SS; based on version 3.24f) assessment software (Methot and Wetzel, 2013), which is widely used in the Unites States and provides an integrated framework for conducting assessment models for a broad variety of data and biological conditions. The SS software is an age-structured forward-projection single-species stock assessment framework that estimates recruitment along with other parameters related to stock productivity and trends. SS uses the C++ ADMB libraries (Fournier et al., 2012) to calculate uncertainty estimates for parameters of interest (e.g., past and future recruitments) based on the Delta method approximation. Simulations and analyses were accomplished using the *ss3sim* software package (Anderson et al., 2014a, 2014b; available at github.com/kellijohnson/AR-perf-testing) to ensure the results are reproducible.

The simulation framework consists of three components: (1) an operating model that generates the true population dynamics; (2) a sampling model that generates data from the operating model; and (3) an estimation method that is applied to the simulated data, where the parameter estimates and derived quantities (i.e., forecasted future population abundances) from the estimation method can be compared with their true values from the operating model. We use a design involving six levels of ρ and four alternative configurations of the estimation method. Additionally, a "less-informative" scenario was simulated and fitted using each estimation method while also estimating stock-recruit steepness to facilitate evaluating performance in a more realistic environment. One hundred simulation replicates were generated for each scenario, where each replicate has a different realization of process (here, recruitment deviations) and observation errors. Each replicate involves simulating population dynamics over 100 years, which we divide into three periods:

1. "Burn-in period" – Years 1-25 are simulated without any fishing;

2. "Fishing period" – Years 26-80 include a simulated fishery, with fishing mortality set to F_{MSY}, and the potential for data from the fishery and a survey, which is used to fit an assessment model in year 80; and
3. "Forecast period" – Years 81-100 are simulated without fishing, which can be compared to forecasts based on parameter estimates derived from the estimation method.

2.1 Operating model

The operating model represents a cod-like life history based on biological parameters estimated from the stock assessment for North Sea cod (*Gadus morhua*; Deroba et al., 2015) with some simplifications facilitating interpretation of the results (Table 1). Simplifications include: one fishery and one survey, combined sexes, and selectivity parameters based on the maturity ogive.

We used the steepness-parameterization of the Beverton-Holt stock-recruit function:

$$r_t = \frac{4hr_0 b_t}{b_0(1-h)+b_t(5h-1)} e^{\varepsilon_t - \sigma_r^2/2}, \qquad (1)$$

where r_t and b_t are the estimates of recruitment output and spawning biomass, respectively, in year t, h, and r_0 are estimated parameters representing steepness (the strength of recruitment compensation) and average recruitment at unfished spawning biomass b_0. The recruitment deviation ε_t is calculated as:

$$\varepsilon_t = \rho\varepsilon_{t-1} + \delta_t\sqrt{1-\rho^2}, \qquad (2)$$

where δ_t is a normally distributed coefficient representing recruitment variability after accounting for the stock recruit relationship:

$$\delta_t \sim N(0, \sigma_r^2), \qquad (3)$$

where σ_r^2 is the marginal variance of recruitment deviations and ρ is the magnitude of autocorrelation in recruitment. Eq. 1 includes the term $e^{\varepsilon_t - \sigma_r^2/2}$, which has an average value of 1.0. This term is included to ensure that r_0 is equal to mean (not median) recruitment given unfished spawning biomass.

Each replicate of the operating model involved simulating true dynamics over 100 years, where recruitment is variable each year, but the same across scenarios for a given iteration (i.e., the values of δ_t for the first replicate of the $\rho = 0.0$ scenario were the same as for the first replicate of the $\rho =$ 0.9 scenario, see Fig. 1). Years 1 through 25 had no fishing and are included to ensure that the population age-structure in year 25 had plausible deviations away from its expectation in an unfished state. In years 26-80, fully-selected fishing mortality, F, was fixed at the value that produced *MSY*. Fishery selectivity was logistic, based on fish length, and was identical to the maturity ogive. Survey selectivity was similar, except that the length at which 50% of individuals were selected by the survey was specified as 80% of the length at which 50% of individuals were mature to ensure that the survey sampled younger fish than were caught in the fishery.

We simulated data for six scenarios that differed in the value of autocorrelation used to generate recruitment: -0.25, 0, 0.25, 0.5, 0.75, and 0.9. Included levels of ρ are centered approximately around estimates from recent meta-analyses (Mueter et al., 2007; Thorson et al., 2014). An autocorrelation level of 0.5 and a marginal log-standard deviation of recruitment of 0.6 (0.2 higher than all other scenarios) was used for a "less-information" scenario.

2.2 Sampling model

Annual catch was reported without error from the start of the fishery (year 26) to the year of the assessment (year 80). Fishery and survey age-composition data were simulated every year for years 26-80, and were drawn from a multinomial distribution with an annual sample size of 100. The survey was simulated every year providing an index of relative abundance for years 26-80, and the abundance index was drawn from a lognormal distribution with log-standard deviation of 0.1 and log-mean equal to logarithm of stock biomass available to the survey in that year. Data are relatively informative to focus the results on the effects of autocorrelated recruitment deviations when estimation is theoretically possible. Data collection for the "less-information" scenario started in year 41 and the log-standard deviation of the index of abundance was 0.25.

2.3 Estimation method

An age-structured stock assessment model was fit to each simulated data set, using data generated during the "fishing period" (see Table 1 for a list of estimated parameters). Each estimation method provides forecasts of population abundance during years 81 to 100, and estimates recruitment deviations for years 1-100. For clarity of communication, we refer to recruitment deviations during the three periods:

1. Recruitment deviations for years 1-25: These recruitment deviations occur prior to the collection of any data, and are estimated so that the estimated age-structure in the first yearwith data (typically year 26) has plausible deviations away from the unfished age-distribution;
2. Recruitment deviations for years 26-80: These recruitment deviations occur during years with available data, and are generally estimated with some precision;
3. Recruitment deviations for years 81-100: These recruitment deviations occur during the forecast period, and ensure that dynamics during this period include a plausible magnitude of recruitment variation.

All estimation methods are provided no data during the forecast period (years 81-100), so recruitment deviations for years 81-100 are estimated at their expected value (i.e., zero when ρ =0, or decaying towards zero from the value of the estimated recruitment deviation in year 80 when ρ ≠ 0).

The estimation method is similar to the operating model, except it also includes annually varying bias-correction for estimated recruitment:

$$r_t = \frac{4hr_0b_t}{b_0(1-h)+b_t(5h-1)}e^{\varepsilon_t-\gamma_t\sigma_r^2/2}, \tag{4}$$

where Eq. 4 replaces Eq. 1 from the operating model, and γ_t is the fraction of bias-correction included for each year. The bias-correction term $e^{-\gamma_t\sigma_r^2/2}$ is included to ensure that r_0 is equal to mean (not median) recruitment given unfished spawning biomass. The corresponding negative log-likelihood computation is:

$$-\log(\mathcal{L}_t) = \begin{cases} \gamma_t \log\left(\sigma_r\sqrt{1-\rho^2}\right) + \frac{\varepsilon_t^2}{(1-\rho^2)\sigma_r^2} & \text{if } t = t_{first} \\ \gamma_t \log\left(\sigma_r\sqrt{1-\rho^2}\right) + \frac{(\varepsilon_t-\varepsilon_{t-1})^2}{(1-\rho^2)\sigma_r^2} & \text{if } t > t_{first}, \end{cases} \tag{5}$$

where this equation uses the conditional standard deviation, $\sigma_r\sqrt{1-\rho^2}$, as the standard deviation for each recruitment deviation, such that the input standard deviation parameter, σ_r, corresponds to the standard deviation across the entire time series and t_{first} refers to the first year that recruitment deviations are estimated. This calculation is identical to the negative log-likelihood for a normal distribution except that it ignores the additional constant of integration, $\log(2\pi)$, and multiplies the conditional standard deviation by the bias-correction term, γ_t. Exploratory analysis suggested that scaling the log of the conditional standard deviation by the bias-correction factor leads to improved estimates of recruitment variability σ_r. However, we note that it is necessary to remove γ_t from Eq. 5 when conducting mixed-effects estimation (Thorson et al., 2015b), and that an alternative bias-corrected estimator is possible using mixed-effects methods without including an explicit bias-correction term in the likelihood computation (Thorson and Kristensen, 2016). However, we use Eqs. 4-5 here, following standard practice in penalized likelihood models and SS.

We implement bias-correction for each simulation replicate following the approach in Methot and Taylor (2011) of:

1. Run the model once to identify maximum likelihood estimates and standard errors for all parameters including ε_t;
2. Calculate standard error estimates, $\widehat{SE}(\varepsilon_t)$, and estimate the bias-correction for each year, $\hat{\gamma}_t = 1 - \widehat{SE}(\varepsilon_t)^2/\sigma_r^2$
3. Fit a five-parameter bias-correction "ramp" (Methot and Taylor, 2011) to the annual bias correction estimates, $\hat{\gamma}_t$;
4. Use predictions of bias-correction, γ_t, for each year in Eq. 1, and re-run the estimation method to identify maximum likelihood estimates and standard errors for all parameters.

This bias-correction algorithm can be derived under the assumption that recruitment deviations are a random effect (Thorson and Kristensen, 2016). For estimation methods with $\rho \neq 0$, the bias correction γ_t is sometimes greater than 0.0 during the forecast period, particularly for larger levels of recruitment autocorrelation. Bias-correction is included during the forecast period because recruitment deviations at the end of the fishing period (e.g., year 80) will inform recruitment deviations during the forecast period (e.g., year 81) whenever $\rho \neq 0$. The delta-method is used for calculating uncertainty in population abundance during the forecast period. Therefore, forecast period abundance has a standard error that includes uncertainty about future recruitment deviations, and this uncertainty is a function of the level of recruitment autocorrelation.

2.3.1 Estimation method configurations

The following four estimation methods were investigated for each level of ρ:

1. "True" – an estimation method where the autocorrelation parameter was fixed at the level used to generate the recruitment deviations in the operating model. This estimation method is not plausible for any real-world assessment (given that the true value of ρ will never be known), but is included as a reference case to demonstrate model performance if the extent of autocorrelation were known exactly.
2. "Zero" – an estimation method where ρ=0. This estimation method represents the most common assumption in stock assessment models to date.
3. "Internal" – an estimation method where ρ is estimated as a fixed effect in SS. This scenario will likely result in biased estimates of ρ, given that SS implements "penalized likelihood" estimation rather than true "mixed-effect" estimation (Thorson and Minto, 2015). Previous research demonstrates that penalized likelihood estimation results in negative bias when estimating the variation in the recruitment deviations (σ_r, Thorson et al., 2014). The bias correction approach developed by Methot and Taylor (2011) is an empirical attempt to overcome this negative bias. However, its performance when estimating the magnitude of ρ has not been previously explored.

4. "External" – an estimation method where ρ is estimated externally to SS. This involves extracting estimates of recruitment deviations from the "Zero" estimation method, and then estimating the first-order autocorrelation of these estimates using the `acf` function in R (R Core Development Team, 2015). This level of autocorrelation is then set as a fixed value in SS and the bias-correction parameters are updated, and then SS is run again. This estimation method will likely have different estimation performance than the "Internal" estimation method, given that sample- and population-level estimates are often different in maximum likelihood estimates of mixed-effects models (Breslow and Clayton, 1993).

In each scenario, the marginal log-standard deviation of recruitment σ_r was fixed at the true value (Table 1). Steepness was estimated in the "less-information" scenario using a beta prior (mean = 0.65, sd = 0.147) and fixed at the true value for all other scenarios.

For each estimation method, we specified that fishing mortality was zero during the forecast period, and this matches the operating model, which has no fishing during the forecast period. Given that fishing rate is correctly specified during the forecast period, any bias or imprecision in population abundance during the forecast period arises either from (1) bias and imprecision of estimated parameters during the fishing period or (2) the impact of mis-specifying ρ during the forecast period. The correct input sample size for multinomial composition samples (N_{input} = 100) were specified in each estimation method (i.e., the estimation method had correct weighting for age-composition sampling data). Convergence of the estimation method was determined using the maximum gradient of the objective function, where models with a maximum gradient of less than 0.01 and a positive definite Hessian matrix were assumed to have converged. Models that failed to converge were removed from the analysis, and exploratory analysis confirms that results (not shown) are qualitatively similar when changing the gradient threshold used to identify model convergence.

2.3.2 Evaluating model performance

Estimation performance was evaluated using three performance statistics:

1. relative error, $RE = (\hat{\theta} - \theta)/\theta$, where $\hat{\theta}$ and θ are estimated and true parameter values, respectively and a well-performing estimation method will have a relative error close to zero for all simulation replicates;
2. average absolute relative error, $AARE = \left(\sum_{i=1}^{n_{reps}} \sum_{t_{min}}^{t_{max}} |RE_{i,t}|\right)/N$, where $RE_{i,t}$ is the relative error in spawning biomass, n_{reps} is the number of simulation replicates, t_{min} and t_{max} are years over which AARE is calculated (e.g., $t_{min}=26$ and $t_{max}=80$ when summarizing performance during the "fishing period"), and N is the total number of observations (i.e., years and replicates); and
3. yearly forecast interval coverage, defined as the proportion of simulation replicates where the forecast interval contains the true value from the operating model. A well-calibrated model will have approximately nominal forecast interval coverage, i.e., a 50% forecast interval will contain the true value in 50% of simulation replicates.

3. Results

3.1 Estimating autocorrelation

We first seek to determine whether an integrated assessment model can provide an accurate and precise estimate of ρ. We therefore evaluate estimates produced either when treating ρ as a fixed effect ("Internal") or when calculating the sample autocorrelation of estimated recruitment deviations ("External"). "Internal" estimation is biased towards extreme values in all scenarios (i.e., towards 1.0 when true autocorrelation is positive and towards -1.0 when true autocorrelation is negative; Fig. 2, top row). "Internal" estimation also has a high proportion of simulation replicates that does not converge when the true autocorrelation is 0.9. In these cases, the estimated autocorrelation approaches the bound at 1.0 and the Hessian matrix is generally not positive definite. By contrast, external estimates of ρ are approximately unbiased for all levels of autocorrelation (Fig. 2, bottom row). "External" estimation also leads to a larger proportion of converged replicates compared to "Internal" estimation. As a sensitivity analysis, we also show "External" estimates of ρ given different quantities of data for estimating recruitment (Fig. 3; i.e., with fishery compositional data and survey data starting in either year 41 or 56, compared with year 26 by default). This shows that ρ can be estimated reasonably well with as few as 25 years of informative data (Fig. 3, bottom row), although estimates become more precise with increasing years of data. Additionally, "External" estimation was on average less biased than "Internal" estimation for the "less-information" scenario ($\overline{RE}$ = -0.21 and 0.42, respectively).

3.2 Impact of autocorrelation on population forecasts

We next seek to determine the impact of autocorrelated recruitment on population forecasts, and whether estimating and accounting for ρ improves model performance. To do so, we first illustrate the effect of autocorrelated recruitment on estimated spawning biomass for all years (years 1-100) for a single replicate of the simulation experiment (Fig. 4). As expected, fixing autocorrelation at its true value results in a forecast interval that expands rapidly during the forecast period (years 81-100) whenever autocorrelation is substantially different from zero. Most notable, the lower confidence bound for forecasts of spawning biomass declines over time when recruitment autocorrelation is 0.9, despite the forecast model correctly assuming that fishing is absent during this period (Fig. 4, top right).

These patterns also hold for the average absolute relative error (AARE) in estimates of spawning biomass across replicates (Fig. 5). During the "fishing" period (years 26-80), the AARE in estimates of spawning biomass is generally less than 0.04 for all estimation methods and all levels of true autocorrelation. We therefore conclude that increased recruitment autocorrelation, or mis-specifying recruitment autocorrelation, has relatively little impact on the precision and accuracy of estimates of spawning biomass during the period with information to estimate recruitment deviations, given an otherwise correctly specified model. However, increased autocorrelation leads to a large increase in AARE during the forecast period (years 81-100), such that AARE is 0.20-0.26 when autocorrelation is 0.9. All estimation methods have an AARE of 0.1 during the forecast period when recruitment is not autocorrelated, but when ρ is high ($\rho =$ 0.75 or 0.9) the "True" and "External" methods have lower AARE (0.17-0.18 and 0.20-0.21) than the "Zero" method (0.19 and 0.26). All estimation methods have a small positive bias in spawning biomass during the forecast period when autocorrelation is 0.75 and even more so when autocorrelation is 0.9. Exploratory analysis indicates that this bias arises due to the nonlinear stock-recruit function, i.e., because calculating forecasts based on the mean of the stock-recruit function is not identical to the expectation of the forecast due to this nonlinearity.

Finally, we illustrate 50% forecast interval coverage for each estimation method, defined as the proportion of simulation replicates where true spawning biomass falls within a 50% forecast interval (Fig. 6). A well-performing estimation method will have nominal coverage probability, i.e., 50% of simulation replicates will fall within the 50% interval. When autocorrelation is absent (Fig. 6, column "0.00"), all estimation methods have approximately nominal coverage, although they exhibit less-than-50% coverage (indicating too narrow of forecast intervals) in years 84-87.

When ρ is fixed at its true value (Fig. 6, top row), coverage remains close to 50% for all levels of true autocorrelation. However, increasing true autocorrelation leads to a large decline in coverage for the "Zero" estimation method (Fig. 6, 2^{nd} row). Coverage is close to 20% in year 90 for this estimation method (only 10 years into the forecast period) when true autocorrelation is 0.75, and is approximately 10% in this year when true autocorrelation is 0.9. By contrast, coverage is slightly smaller than 50% for the external estimation method when true autocorrelation is 0.75 or 0.9. We therefore conclude that external estimation substantially improved forecast interval performance relative to a model that neglects autocorrelated recruitment. Coverage was similar for a 75% forecast interval, though more variable and less optimistic (Fig. 6, open circles). Coverage was less than expected for all estimation methods in the "less-information" scenario (Fig. 7).

4. Discussion

Fisheries management in the United States and worldwide increasingly uses integrated stock assessment models to evaluate the likely impact of alternative management measures on fish population abundance. The United States and Europe both seek to end overfishing and rebuild overfished stocks (see Magnuson-Stevens Fishery Conservation and Management Reauthorization Act of 2006, http://www.nmfs.noaa.gov, and European Union Common Fisheries Policy, http://ec.europa.eu/fisheries/cfp/index_en.htm). Rebuilding plans for overfished stocks in the United States are based upon forecasts of population abundance, and each United States Regional Fisheries Management Council is required to develop an approved Rebuilding Plan that will result in rebuilding within a pre-determined time frame. Rebuilding Plans are also required to be more likely than not to succeed in their stated timeframe, i.e., rebuilding plans are premised on a probabilistic interpretation of the forecasts generated from integrated stock assessment models. A probabilistic interpretation of catch advice arising from stock assessment models is also used in many United States regions to incorporate scientific uncertainty when defining catch limits (Shertzer et al., 2008) or when interpreting stock status relative to biological reference points (e.g., Stewart et al., 2013).

In this study, we demonstrate that autocorrelated recruitment has a substantial impact upon both the accuracy of forecasts (i.e., how close they are to the true value) as well as the width of forecast intervals (i.e., the magnitude of the estimated standard error for forecasts). In particular, high levels of autocorrelation (i.e., $\rho>0.5$) result in substantial increases in the relative error of population forecasts, regardless of whether the stock assessment accounts for recruitment autocorrelation or not. Also, a model where autocorrelation is fixed at its true value showed that forecast interval width is substantially increased when autocorrelation is high compared to when it is zero. These results confirm that the certainty of population forecasts is highly dependent upon the presence or absence of recruitment autocorrelation. Presumably, high recruitment autocorrelation could contribute to the lack of rebuilding for some fishes under rebuilding plans worldwide, particularly if forecasted biomass is overestimated, as in our results (Hutchings, 2001; Neubauer et al., 2013). Previous analysis of model output from stock assessment models suggests that recruitment may have intermediate, positive autocorrelation for marine fishes (Ianelli, 2002; Thorson et al., 2014). However, care should be taken when interpreting these previous results, as well as results from the "External" estimation method, which are based on model-output (Brooks and Deroba, 2015; Thorson et al., 2015a).

We have also shown improvements in forecast interval performance when fixing autocorrelation at the sample autocorrelation of estimated recruitment deviations (the "External" estimation method). Accuracy of forecast interval width is less important for forecasts that only utilize the median, but if fisheries managers use other quantities from the forecast (i.e., seek a management procedure that achieves a target biomass with 75% probability), or have Harvest Control Rules where the percentile for catch advice depends on the degree of depletion, then it is necessary to have accurate estimates of forecast interval width. Our simulation results show that the "External" estimate of autocorrelation provides less biased estimates of autocorrelation than estimating autocorrelation as a fixed effect, as currently implemented in SS.

The poor forecast interval performance when estimating autocorrelation as a fixed effect likely arises from the use of penalized-likelihood estimation methods. Penalized likelihood has previously been shown to result in negatively biased estimates of the variance of recruitment deviations (Thorson et al., 2015b), and a sample-based statistic has therefore been developed for estimating this variance (Methot and Taylor, 2011). We tried modifying the Methot and Taylor (2011) approach to account for the impact of ρ on the realized variance of recruitments, by replacing the negative log-likelihood computation (Eq. 5) with the following:

$$-\log(\mathcal{L}_t) = \begin{cases} \gamma_t \log(\sigma_r) + \frac{\varepsilon_t^2}{(1-\rho^2)\sigma_r^2} & \text{if } t = t_{first} \\ \gamma_t \log(\sigma_r) + \frac{(\varepsilon_t - \varepsilon_{t-1})^2}{(1-\rho^2)\sigma_r^2} & \text{if } t > t_{first} \end{cases} \tag{6}$$

This modification resulted in estimates of ρ that were biased towards zero (results not shown), and we chose to proceed with Eq. 5, given that it has a stronger statistical justification. We note that fixing ρ at an externally derived value does not propagate uncertainty about the magnitude of autocorrelation when estimating standard errors for other parameters or derived quantities for management (e.g., the CV of average unfished spawning biomass may be different when ρ is estimated compared to when ρ is fixed).

Results presented here are representative of the best case scenario. Estimation methods were fit to a relatively large amount of informative data (i.e., data was available from both the fishery and a survey on a yearly basis) and were correctly specified. Furthermore, steepness and the marginal standard deviation of recruitment deviations were fixed at their true values. Previous research documented an inability to estimate steepness when autocorrelated recruitment deviations were accounted for (i.e., fixed at an externally estimated value) within the stock assessment framework (Butterworth et al., 2003; Ianelli, 2002), but did not investigate the effect of estimating steepness and autocorrelation on forecasts. Estimating steepness proved to be difficult no matter which estimation method was used to account for autocorrelated recruitment deviations, reminding us that poor forecast coverage can arise from causes other than autocorrelated recruitment. Future research could explore sensitivity to many types of model mis-specification, including: estimating steepness with more-informative data (e.g., catches from a stock experiencing a large contrast in spawning biomass) or mis-specifying its value; mis-specifying selectivity or growth parameters, such that estimated recruitment deviations incorporate process errors from mis-specifying other model components; and alternative forms for recruitment. In particular, we hypothesize that periodic changes in average recruitment ("regime shifts") will appear as 2nd or higher-order autocorrelation, and that our specification of 1st-order autocorrelation might be a poor approximation in these causes.

Based on our results here, we identify several useful avenues for future research:

1. Most obviously, research could explore whether a mixed-effects estimate of autocorrelation could improve performance when estimating autocorrelation as a model parameter. Mixed-effects estimation is increasingly feasible using either the Laplace approximation (Kristensen

et al., 2016; Skaug and Fournier, 2006; Thorson et al., 2015b) or Markov-chain Monte Carlo sampling (Stewart et al., 2013).

2. Future research could also explore the impact of autocorrelated recruitment on harvest strategy performance when either estimating or ignoring autocorrelation. Autocorrelated errors during forecast intervals are likely to impact the performance of harvest strategies (Wiedenmann et al., 2015), but it remains unclear whether the magnitude of improvements from estimating the extent of autocorrelation outweigh the additional complexity when developing and explaining the model.
3. Bias adjustment methods (Methot and Taylor, 2011) were developed without accounting for ρ, and future research should investigate how to account for this bias as well as autocorrelated recruitment deviations. In particular, we recommend further investigation of mixed-effects estimation and associated bias-correction methods (Thorson and Kristensen, 2016; Thorson and Minto, 2015) as a generic solution to bias-correction for autocorrelated errors.
4. Finally, many parameters are likely to vary over time in stock assessment models, including growth, maturity, selectivity, and productivity (Martell and Stewart, 2014; Thorson et al., In press). These processes (e.g., time-varying selectivity) could affect the interpretation of length composition samples, so neglecting time-varying selectivity could in some cases appear as autocorrelated recruitment (Butterworth et al., 2003). We did not explore the impact of multiple time-varying parameters on estimates of recruitment autocorrelation, and its potential impact remains difficult to predict. We therefore recommend ongoing research to develop tools to identify and account for time-varying parameters in stock assessment models.

5. Conclusions

We conclude that "External" estimation will likely result in better estimates of the magnitude of autocorrelated recruitment when estimation is based on penalized likelihood. The estimation of ρ appears to be most important for the forecast period as bias and precision were similar among mis-specified and correctly specified models for the estimation period. Consequently, future research should prioritize including ρ in all forecasts regardless of its magnitude and obtaining the best external estimate of ρ possible, especially if forecasts are performed outside of the stock assessment model. Unfortunately, even when ρ is fixed at its true value forecast coverage is poor for the first ten years when autocorrelation is high. Therefore, rebuilding within 10 years for stocks likely to have autocorrelated recruitment may necessitate updating the assessment more than once during the 10 year period, and, potentially, even more frequently depending on the quality of available data.

6. Acknowledgements

We gratefully acknowledge help from Ian Taylor, Chantel Wetzel, and Allan Hicks (NMFS, NOAA) on implementation issues regarding *ss3sim* and forecasting within Stock Synthesis. We thank James Ianelli for his comments on an earlier draft, which helped improve the content and clarity of the manuscript. This publication is partially funded by the Joint Institute for the Study of the Atmosphere and Ocean (JISAO) under NOAA Cooperative Agreement No. NA10OAR4320148, Contribution No. ?. Partial funding for KFJ was provided by a grant from Washington Sea Grant, University of Washington, pursuant to National Oceanic and Atmospheric Administration Award No. NA140AR4170078. This investigation contributes towards the work plan of the ICES Working Group on Recruitment Forecasting in a variable Environment

(WGRFE). The views expressed herein are those of the authors and do not necessarily reflect the views of NOAA or any of its sub-agencies.

References

Anderson, S.C., Monnahan, C.C., Johnson, K.F., Ono, K., Valero, J.L., 2014a. ss3sim: An R package for stock assessment simulation with Stock Synthesis. PLoS ONE 9, e92725.

Anderson, S.C., Monnahan, C.C., Johnson, K.F., Ono, K., Valero, J.L., Cunningham, C.J., Hurtado-Ferro, F., Licandeo, R., McGilliard, C.R., Szuwalski, C.S., Vert-pre, K.A., Whitten, A.R., 2014b. ss3sim: Fisheries stock assessment simulation testing with Stock Synthesis. R package version 0.9.0.

Bailey, K.M. 2000. Shifting control of recruitment of walleye pollock *Theragra chalcogramma* after a major climatic and ecosystem change. Mar. Ecol. Prog. Ser. 198, 215-224.

Breslow, N.E., Clayton, D.G., 1993. Approximate inference in generalized linear mixed models. J. Am. Stat. Assn. 88, 9-25.

Brooks, E.N., Deroba, J.J., 2015. When “data” are not data: the pitfalls of post hoc analyses that use stock assessment model output. Can. J. Fish. Aquat. Sci. 72, 634-641. doi:10.1139/cjfas-2014-0231

Butterworth, D.S., Ianelli, J.N., and Hilborn, R., 2003. A statistical model for stock assessment of Southern Bluefin Tuna with temporal changes in selectivity. Afr. J. Mar. Sci. 25, 331-361.

Clark, W.G., 1993. The effect of recruitment variability on the choice of a target level of spawning biomass per recruit, in: Kruse, G., Engers, D.M., Marasco, R.J., Pautzke, C., Quinn, T.J.I. (Eds.), Proceedings of the International Symposium on Management Strategies for Exploited Fish Populations. University of Alaska, Alaska Sea Grant Report 93-02, Fairbanks, AK, pp. 233-246.

Deroba, J.J., Butterworth, D.S., Methot, R.D., Jr., De Oliveira, J.A.A., Fernandez, C., Nielsen, A., Cadrin, S.X., Dickey-Collas, M., Legault, C.M., Ianelli, J., Valero, J.L., Needle, C.L., O’Malley, J.M., Chang, Y-J., Thompson, G.G., Canales, C., Swain, D.P., Miller, D.C.M., Hintzen, N.T., Bertignac, M., Ibaibarriaga, L., Silva, A., Murta, A., Kell, L.T., de Moor, C.L., Parma, A.M., Dichmont, C.M., Restrepo, V.R., Ye, Y., Jardim, E., Spencer, P.D., Hanselman, D.H., Blaylock, J., Mood, M., Hulson, P-J.F., 2015. Simulation testing the robustness of stock assessment models to error: some results from the ICES strategic initiative on stock assessment methods. ICES J. Mar. Sci. 72, 19-30.

Fournier, D.A., Skaug, H.J., Ancheta, J., Ianelli, J., Magnusson, A., Maunder, M.N., Nielsen, A., and Sibert, J., 2012. AD Model Builder: using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. Optim. Methods Softw. 27, 233-249.

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.

Hollowed, A.B., Hare, S.R., Wooster, W.S., 2001. Pacific Basin climate variability and patterns of Northeast Pacific marine fish production. Prog. Oceanogr. 1-4, 257-282.

Holt, C.A., Punt, A.E., 2009. Incorporating climate information into rebuilding plans for overfished groundfish species of the U.S. west coast. Fish. Res. 100, 57-67.

Hutchings, J.A., 2001. Influence of population decline, fishing, and spawner variability on the recovery of marine fishes. J. Fish Biol. 59, 306-322.

Ianelli, J.N., 2002. Simulation analyses testing the robustness of productivity determinations from West Coast Pacific Ocean Perch stock assessment data. N. Am. J. Fish. Manage. 22, 301-310.

Jørgensen, C., Ernande, B., Fiksen, Ø., Dieckmann, U., 2006. The logic of skipped spawning in fish. Can. J. Fish. Aquat. Sci. 63, 200-211.

Kristensen, K., Nielsen, A., Berg, C.W., Skaug, H., Bell, B.M. 2016. TMB: Automatic Differentiation and Laplace Approximation. J. Stat. Softw. 70, 1-21.

Lindegren, M., Checkley, D.M., 2013. Temperature dependence of Pacific sardine (*Sardine sagax*) recruitment in the California Current Ecosystem revisited and revised. Can. J. Fish. Aquat. Sci. 70, 245-252.

Martell, S., Stewart, I., 2014. Towards defining good practices for modeling time-varying selectivity. Fish. Res. 158, 84-95.

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.

Methot, R.D., Wetzel, C.R., 2013. Stock synthesis: a biological and statistical framework for fish stock assessment and fishery management. Fish. Res. 142, 86-99.

Methot, R.D., Jr., Tromble, G.R., Lambert, D.M., Greene, K.E., 2013. Implementing a science-based system for preventing overfishing and guiding sustainable fisheries in the U.S. ICES J. Mar. Sci. 71, 183-194. 10.1093/icesjms/fst119

Mueter, F.J., Boldt, J.L., Megrey, B.A., Peterman, R.M., 2007. Recruitment and survival of Northeast Pacific Ocean fish stocks: temporal trends, covariation, and regime shifts. Can. J. Fish. Aquat. Sci. 64, 911-927.

Neubauer, P., Jensen, O.P., Hutchings, J.A., Baum, J.K., 2013. Resilience and recovery of overexploited marine populations. Sci. 340, 347-349.

NRC, 2013. Evaluating the Effectiveness of Fish Stock Rebuilding Plans in the United States. The National Academies Press, Washington, D.C. http://www.nap.edu/catalog.php?record_id=18488

Okamoto, D.K., Schmitt, R.J., Holbrook, S.J., Reed, D.C., 2012. Fluctuations in food supply drive recruitment variation in marine fish. Proc. R. Soc. B. 23, 365-373.

Peterman, R.M., Anderson, J.L., 1999. Decision analysis: a method for taking uncertainties into account in risk-based decision making. Hum. Ecol. Risk Assess. 5, 231-244.

Punt, A.E., 2011. The impact of climate change on the performance of rebuilding strategies for overfished groundfish species of the U.S. west coast. Fish. Res. 109, 320-329.

R Core Development Team, 2014, R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.

Shertzer, K.W., Prager, M.H., Williams, E.H., 2008. A probability-based approach to setting annual catch levels. Fish. Bull. 106, 225-232.

Skaug, H., Fournier, D., 2006. Automatic approximation of the marginal likelihood in non-Gaussian hierarchical models. Comput. Stat. Data Anal. 51, 699-709.

Stewart, I.J., Hicks, A.C., Taylor, I.G., Thorson, J.T., Wetzel, C., Kupschus, S., 2013. A comparison of stock assessment uncertainty estimates using maximum likelihood and Bayesian methods implemented with the same model framework. Fish. Res. 142, 37-46. doi:10.1016/j.fishres.2012.07.003

Szuwalski, C.S., Vert-Pre, K.A., Punt, A.E., Branch, T.A., Hilborn, R., 2014. Examining common assumptions about recruitment: a meta-analysis of recruitment dynamics for worldwide marine fisheries. Fish Fish. 16, 633-648. doi:10.1111/faf.12083

Thorson, J.T., Jensen, O.P., Zipkin, E.F., 2014. 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., Cope, J.M., Kleisner, K.M., Samhouri, J.F., Shelton, A.O., Ward, E.J., 2015a. Giants' shoulders 15 years later: lessons, challenges and guidelines in fisheries meta-analysis. Fish Fish. 16, 342-361. doi:10.1111/faf.12061

Thorson, J.T., Hicks, A.C., Methot, R.D., 2015b. Random effect estimation of time-varying factors in Stock Synthesis. ICES J. Mar. Sci. J. Cons. 72, 178-185. doi:10.1093/icesjms/fst211

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., Kristensen, K., 2016. Implementing a generic method for bias correction in statistical models using random effects, with spatial and population dynamics examples. Fish. Res. 175, 66-74.

Thorson, J.T., Monnahan, C., Cope, J.M., In press. The effect of nonstationary biological processes on fisheries management targets. Fish. Res.

Wiedenmann, J., Wilberg, M.J., Sylvia, A., Miller, T.J., 2015. Autocorrelated error in stock assessment estimates: Implications for management strategy evaluation. Fish. Res. 172, 325-334. doi:10.1016/j.fishres.2015.07.037

Wilderbuer, T.K., Hollowed, A.B., Ingraham, W.J., Spencer, P.D., Conners, M.E., Bond, N.A., Walters, G.E., 2002. Flatfish recruitment response to decadal climatic variability and ocean conditions in the eastern Bering Sea. Prog. Oceanog. 55, 235-247.

Wooster, W.S., Bailey, K.M., 1989. Recruitment of marine fishes revisited. In: Beamish, R.J., McFarlane, G.A. (Eds) Effects of ocean variability on recruitment and evaluation of parameters used in stock assessment models. Can. Spec. Publ. Fish. Aquat. Sci. 108, 153-159.

Table 1. Parameter specifications used in the operating models (OMs) and estimation methods (EMs). Parameter specifications that vary among scenarios (combinations of OMs and EMs) are denoted in the table.

Parameter Name	Symbol	**OM** True value	**EM** Fixed (F) or Estimated (Est)
Natural mortality rate	M	0.2 yr^{-1}	F
Length at age 1	$L_{a=1}$	20 cm	F
Asymptotic maximum length	L_{∞}	132 cm	F
Von Bertalanffy growth coefficient	k	0.2 yr^{-1}	F
Coefficient of variation for length at age 1	$CV_{a=1}$	0.1	F
Coefficient of variation for asymptotic maximum length	CV_{∞}	0.1	F
Length at 50% maturity	θ_1^{mat}	38.2 cm	F
Length at 95% maturity	θ_2^{mat}	48.9 cm	F
Average recruits for the unfished population (natural log)	$ln(r_0)$	18.7	Est
Steepness of the Beverton-Holt stock recruit function	h	0.65	F[1]
Marginal log-standard deviation of recruitment	σ_R	0.4[2]	F
Magnitude of autocorrelated recruitment	ρ	Varies	varies
Random coefficients for recruitment variability (years 1-100)	δ_t	Varies	Est
Catchability coefficient for survey index of abundance (natural log)	$ln(q)$	0	Est
Length at 50% selection in the fishery	$\theta_1^{fishery}$	38.2 cm	Est
Length at 95% selection in the fishery	$\theta_2^{fishery}$	48.9 cm	Est
Length at 50% selection in the survey	θ_1^{survey}	30.6 cm	Est
Length at 95% selection in the survey	θ_2^{survey}	39.1 cm	Est

[1]Steepness is estimated in the "less-information" scenario using a beta prior (mean = 0.65, sd = 0.147).
[2]Marginal log-standard deviation of recruitment is 0.6 in the "less-information" scenario.

Fig. 1. Examples of fifty years of autocorrelated recruitment deviations for three levels of ρ: (i) -0.25 (dashed line), (ii) 0.00 (solid line), and 0.75 (dotted line), where each example used the same set of process error deviations ($\boldsymbol{\delta_t}$).

Fig. 2. Estimates of recruitment autocorrelation (ρ) from two estimation methods: (i) estimated as a fixed effect within Stock Synthesis simultaneously with other parameter estimation (“Internal”; top row) and (ii) calculated as the sample autocorrelation of recruitment deviations estimated in Stock Synthesis when ρ is fixed at zero (“External”; bottom row), for six (true) levels of recruitment autocorrelation (columns). The dashed red line illustrates the true level of autocorrelation, while the black shaded area is a histogram representing the simulation distribution for each scenario and estimation method. The number in the top left of each plot indicates the number of converged runs (out of 100).

Fig. 3. Estimates of recruitment autocorrelation (ρ) from the "External" estimation method, where it is calculated as the sample autocorrelation of recruitment deviations estimated in Stock Synthesis, for six (true) levels of recruitment autocorrelation (columns) and three different starting years for fishery length- and age-composition samples. The dashed red line illustrates the true level of autocorrelation, while the black shaded area is a histogram representing the simulation distribution for each scenario and estimation method. The number in the top left of each plot indicates the number of converged runs (out of 100).

Fig. 4. Illustration of estimated spawning biomass during 100 simulated years for different scenarios (columns, where recruitment autocorrelation is ρ={-0.25, 0.0, 0.25, 0.5, 0.75, 0.9}), and four estimation method (rows: “True”, “Zero”, “Internal”, and “External”), where each panel shows the true spawning biomass (black line) and the red shaded area shows the 95% confidence and forecasting intervals for the estimated spawning biomass.

Fig. 5. Relative error in spawning biomass during years for which the estimation method was provided data (years 26 through 80) and the forecast period (years 81 through 100, to the right of vertical red dashed lines) for six levels of autocorrelation in the simulated data (columns) and four estimation methods (rows). Horizontal dashed red lines indicate a relative error of zero. Upper and lower edges of the boxes correspond to the first and third quartiles (the 25th and 75th percentiles) and the whiskers correspond to 1.5 times the distance between the first and third quartiles. In each plot, the number in the top left indicates the number of converged runs (out of 100), the bottom left number is AARE for the years with data, while the bottom right number is AARE in the forecast.

Fig. 6. Performance of forecast interval estimates for different estimation methods (rows) and levels of autocorrelation (columns), where each panel shows the proportion of 50% (closed circles) and 75% (open circles) forecast intervals for spawning biomass that contain the true value. A well calibrated 50% forecast interval will contain the true value 50% of the time. Calibration lines for both 75% and 50% forecast intervals are indicated by the red dashed lines in each panel, respectively. Points above or below the line indicate forecast intervals were too conservative (wide) or permissive (not wide enough), respectively. In each plot, the number in the top left indicates the number of converged runs (out of 100).

Fig. 7. Relative error in spawning biomass (left column) and forecast coverage of spawning biomass (right column) for the "less-information" scenario across four estimation methods (rows) when estimating steepness. Relative error in spawning biomass is shown for years for which the estimation method was provided data (years 41 through 80) and the forecast period (years 81 through 100, to the right of vertical red dashed lines), where the horizontal dashed red lines indicate a relative error of zero. Upper and lower edges of the boxes correspond to the first and third quartiles (the 25th and 75th percentiles) and the whiskers correspond to 1.5 times the distance between the first and third quartiles. Performance of forecast interval estimates shows the proportion of 50% (closed circles) and 75% (open circles) forecast intervals for spawning biomass that contain the true value. A well calibrated 50% forecast interval will contain the true value 50% of the time. Calibration lines for both 75% and 50% forecast intervals are indicated by the red dashed lines in each panel, respectively. Points above or below the line indicate forecast intervals were too conservative (wide) or permissive (not wide enough), respectively. In each plot, the number in the top left indicates the number of converged runs (out of 100) and the number in the top right indicates the relative error in steepness. In each plot, the number in the top left indicates the number of converged runs (out of 100), the top right is the relative error in steepness, the bottom left number is AARE for the years with data, while the bottom right number is AARE in the forecast.

Fig. 1.
Click here to download high resolution image

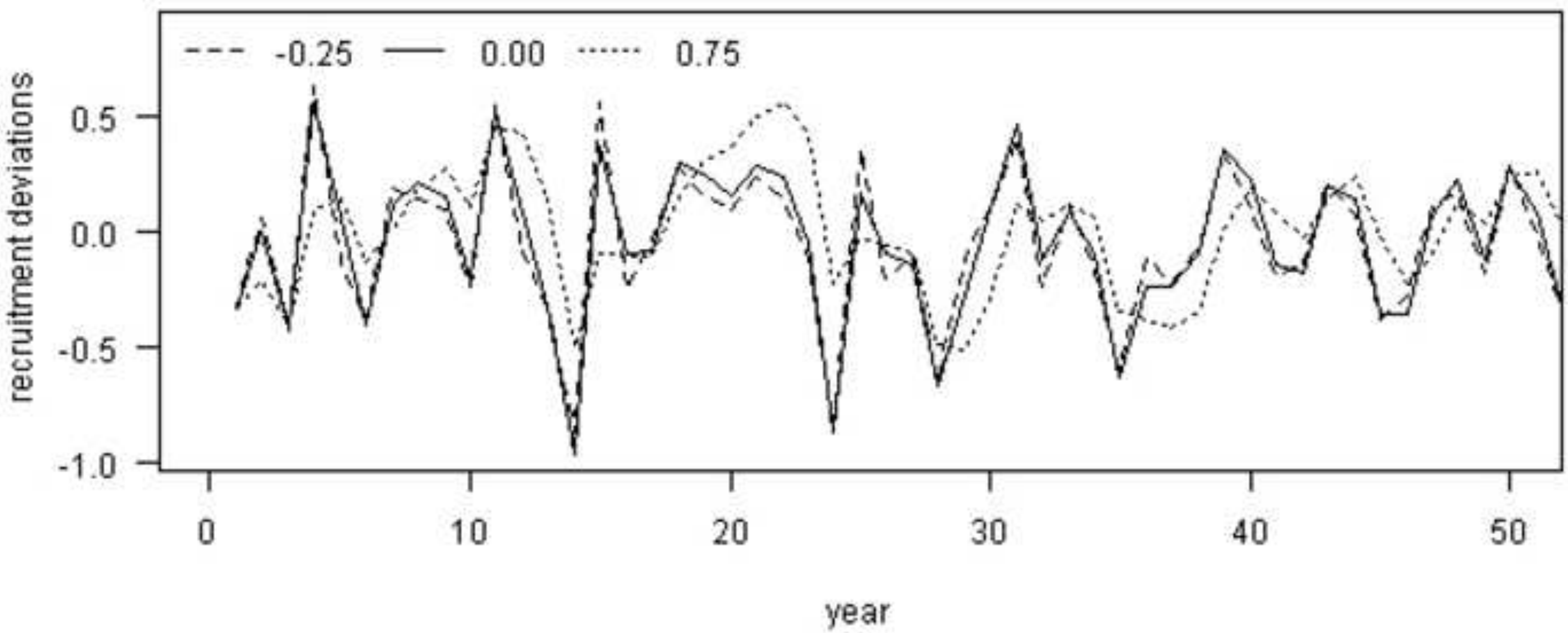

Fig. 2.
Click here to download high resolution image

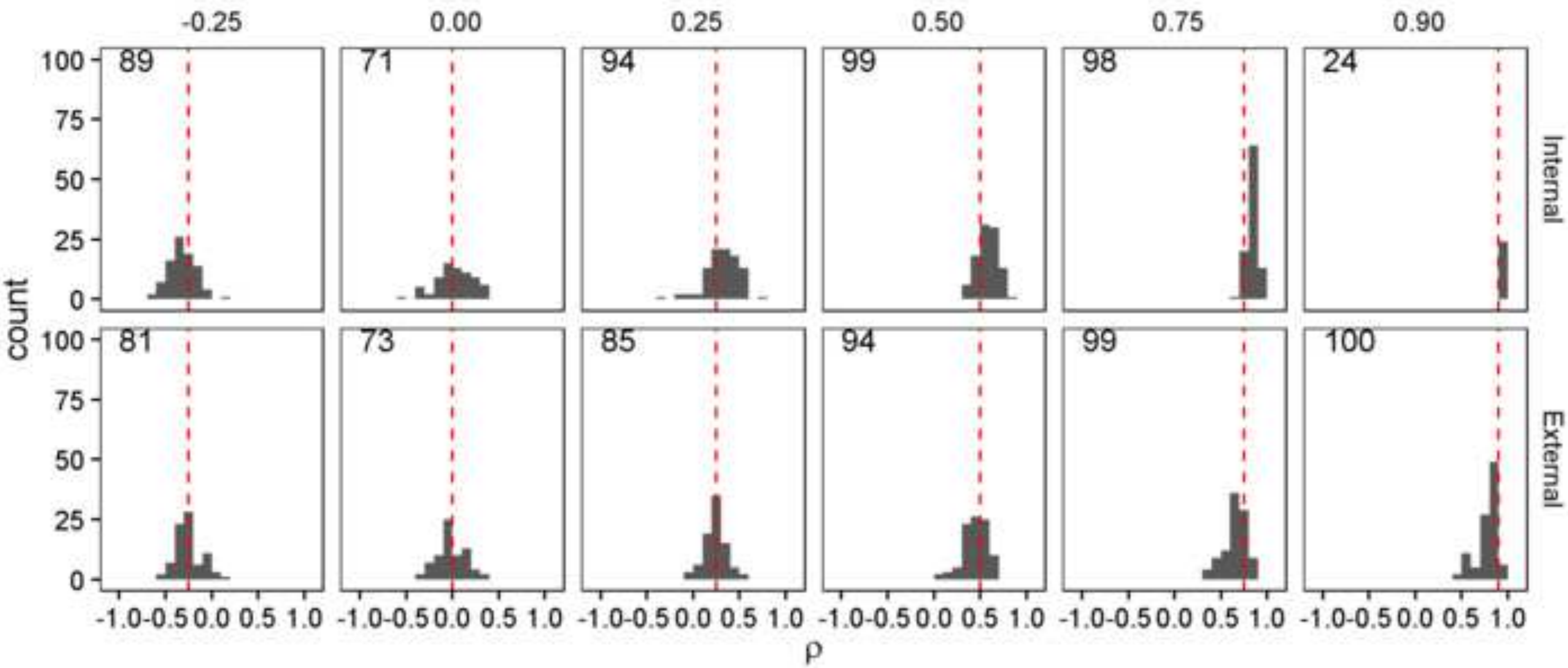

Fig. 3.
Click here to download high resolution image

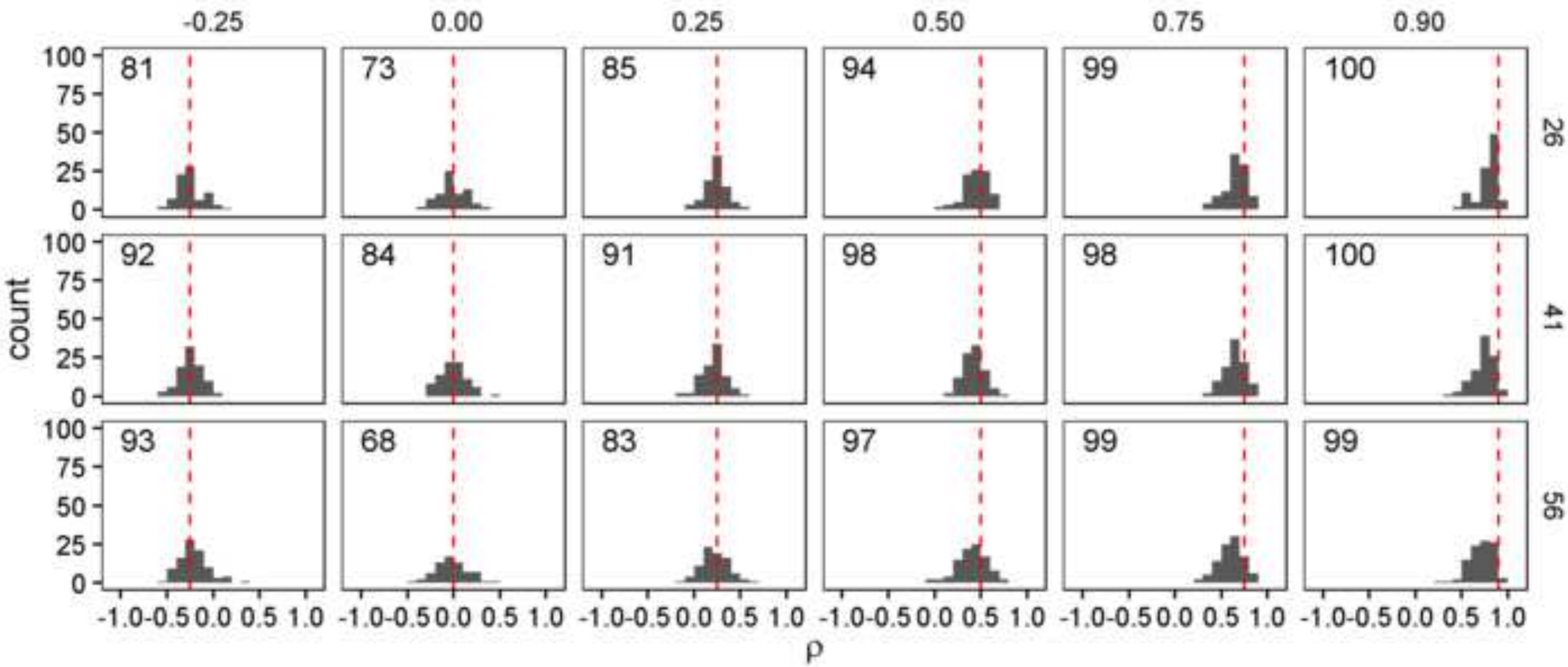

Fig. 4.
Click here to download high resolution image

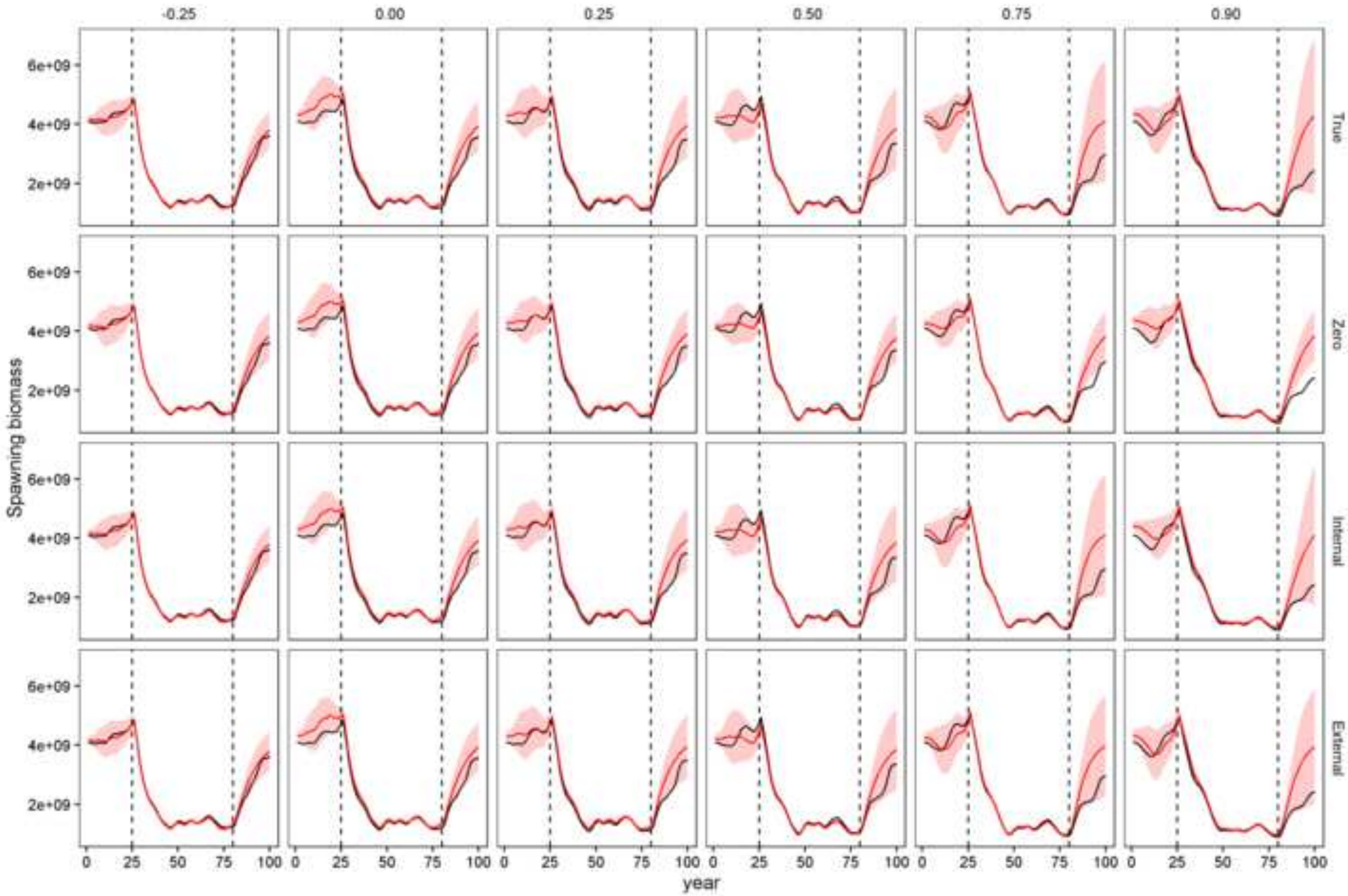

Fig. 5.
Click here to download high resolution image

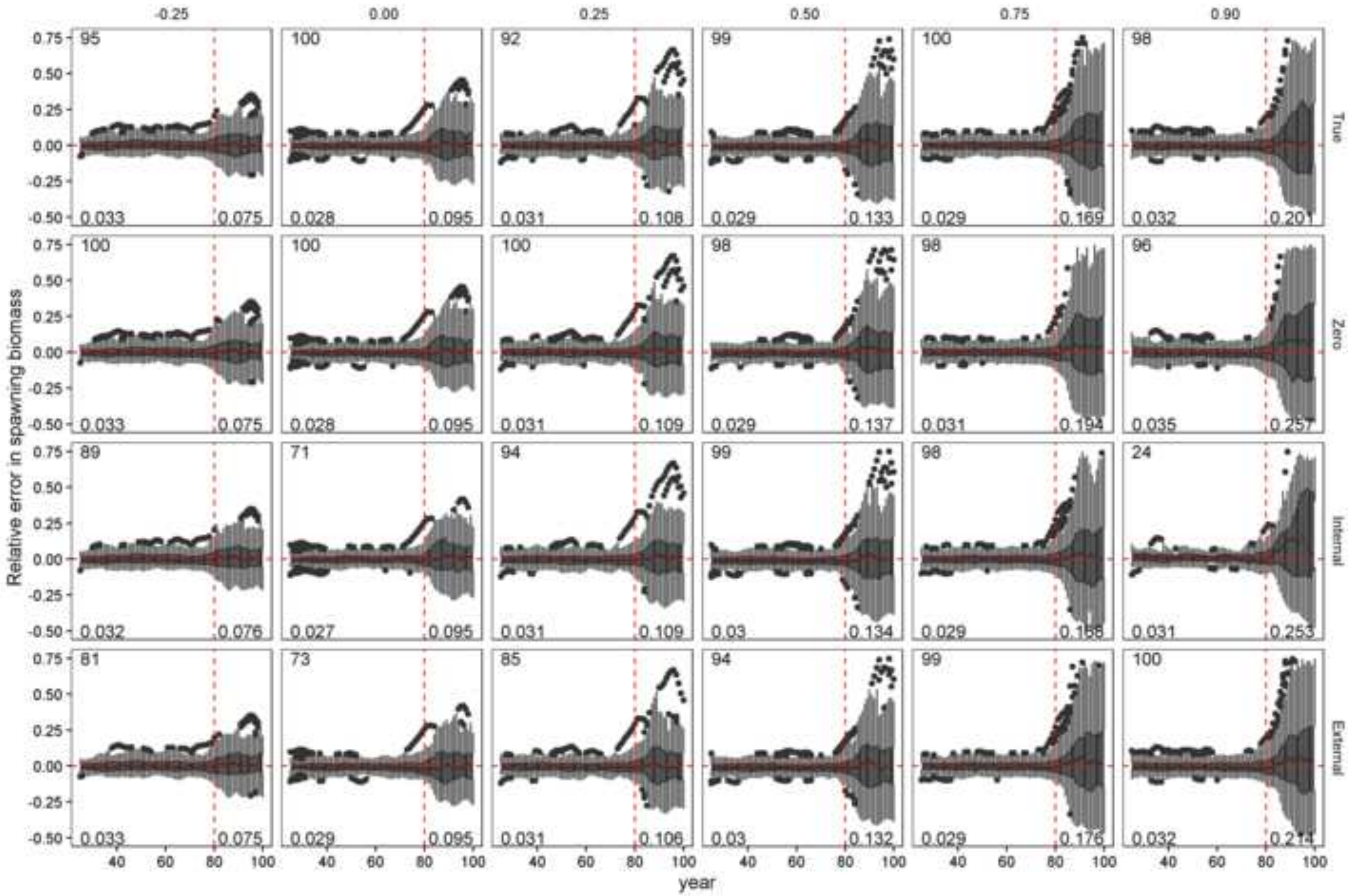

Fig. 6.
Click here to download high resolution image

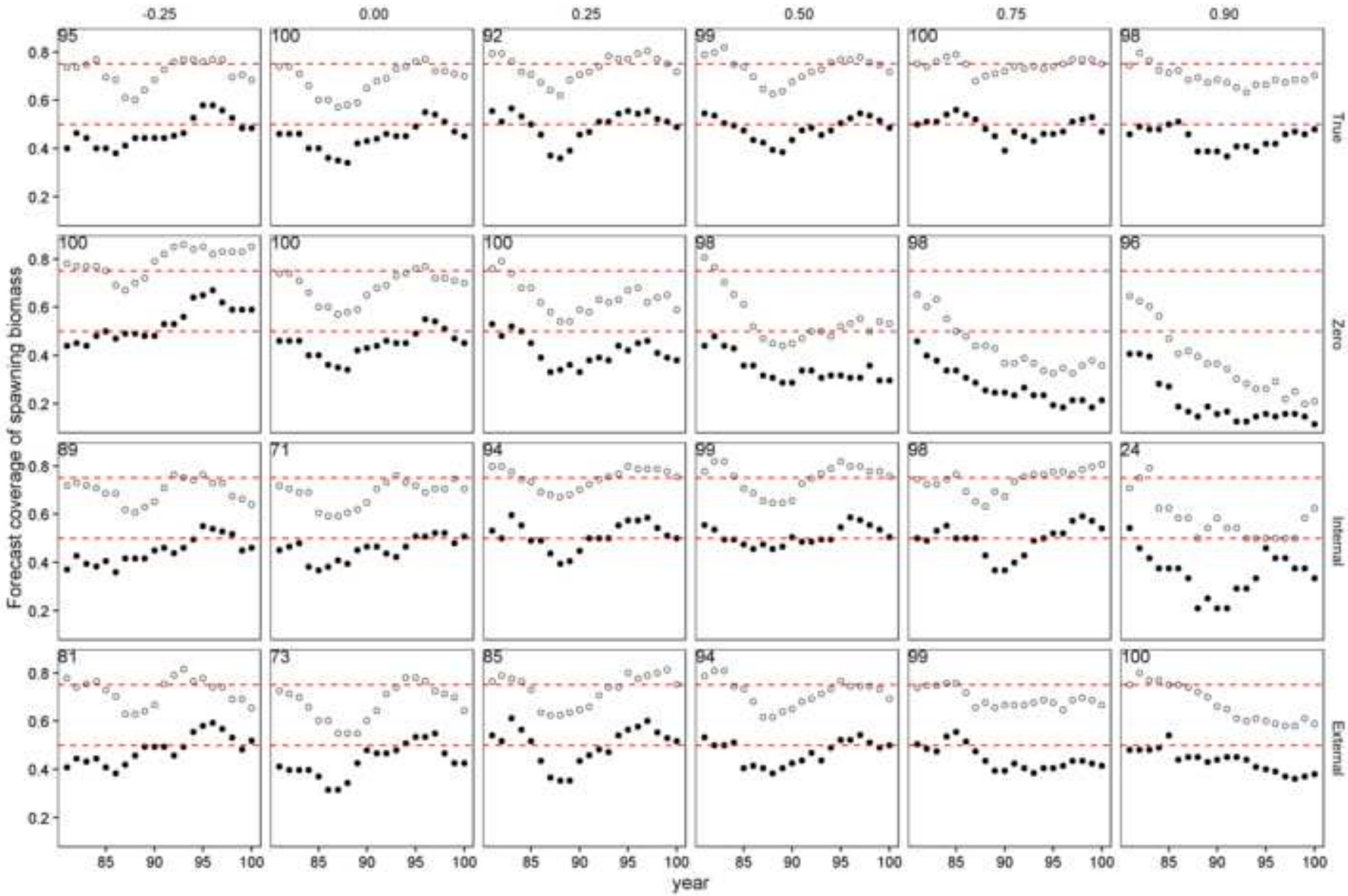

Fig. 7.
Click here to download high resolution image