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

Kelli Johnson1, Elizabeth Councill1,2\*, James T. Thorson1, Elizabeth Brooks3, Richard D. Methot1, André E. Punt2

1School of Aquatic and Fishery Sciences, University of Washington, Box 355020, Seattle, WA 98195-5020, USA

2Fisheries Resource Assessment and Monitoring Division, Northwest Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, 2725 Montlake Blvd. East, Seattle, WA 98112, USA

3Northeast Fisheries Science Center, 166 Water Street, Woods Hole, MA 02543, USA

\*Corresponding author:

telephone: +1 206 860 2425;

fax: +1 206 ??? ????;

email: Elizabeth.councill@noaa.gov

**Abstract**

Recruitment is often autocorrelated for marine fishes, and autocorrelation can arise due to numerous factors including regime shifts and periodicity in environmental drivers affecting survival rates during larval and juvenile stages. Patterns of first-order temporal autocorrelation in recruitment deviations are identified by periods of time where recruitment deviations are positive or negative for several years in a row. The ability of stock assessments to accurately estimate the magnitude of recruitment autocorrelation, and its effect on the quality of forecasts of spawning stock biomass (SSB), has not generally been analyzed. Monte Carlo simulations were used to evaluate how well Stock Synthesis (SS), an integrated age-structured stock assessment method used extensively in the management of fish stocks, estimates autocorrelation in the presence of a range of autocorrelated recruitment deviations. The precision and accuracy of estimated autocorrelation, and the ability of the stock assessment framework to forecast the true dynamics of the system, were compared for scenarios where AR was fixed at zero, fixed at its true value, internally estimated, and input as a fixed value determined using an external estimation procedure. Penalized-likelihood estimates of AR produced by SS internally were biased toward zero when AR was larger than zero, but were unbiased when the simulated value of AR was less than or equal to zero. However, a less biased estimate of AR was obtained by estimating it directly from the recruitment deviations estimated within SS. Additionally, the forecast interval for estimates of SSB during the forecast period were more uncertain when AR was high. Our results suggest that AR should first be estimated internally within a stock assessment, then externally calculated from the estimated recruitment deviations, especially in cases when the internally-estimated value is positive and nonzero. Results using this approach lead to estimates of AR that have small bias, and results in accurate forecast interval coverage (i.e., a 50% forecast interval that includes the true value of SSB for 50% of simulation replicates).

**1. Introduction**

Under the U.S. Magnuson-Stevens Fishery Conservation and Management Act (MSA; U.S. Public Law 104-297), overfished stocks in the U.S. must have a rebuilding plan. This plan involves specifying management measures to rebuild the stock to a biomass associated with maximum sustainable yield () within 10 years (or, if rebuilding within 10 years is impossible, then 1 generation plus the median time for rebuilding given no fishing). Legally, this rebuilding plan must be more likely than not to succeed, i.e., it is based upon a probabilistic forecast of future population dynamics given different potential levels of fishing. Additionally, the National Marine Fisheries Service (NMFS) must identify target and limit reference points for all stocks included in fishery management plans. As 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 population dynamics model that treat fluctuations in recruitment as a random process around a prediction derived from a presumed relationship between spawning output and recruits (Clark 1993; Methot and Wetzel 2013). Stock assessments are increasingly conducted using an “integrated” population dynamics model that typically incorporates many types of information, including samples of compositional data from fishery and surveys, indices of abundance, and information regarding total fishery harvest (Maunder and Punt 2013). These data are combined to estimate values for population productivity (parameters in the stock-recruit relationship) and status (spawning biomass in each year). Probabilistic forecasts are then calculated by simulating values for future recruitment and then projecting population dynamics forward into future years.

Recent studies illustrate that recruitment for many fishes has periods that are anomalously high or low (Szuwalski et al. 2014). 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 predicts future recruitment can be identified, it can then be used to inform rebuilding forecasts (Holt and Punt 2009, Punt 2011) and reference point calculations (Schirripa et al. 2009). If an environmental factor cannot be identified, population forecasts can calculated for different “states-of-nature”, where each state-of-nature depends upon a hypothetical scenario for future recruitment (e.g., high, average, and low productivity scenarios).

When measurable environmental factors remain unidentified, regime shifts can instead be treated as autocorrelation (i.e., where recruitment deviations are greater or less than zero for many years in a sequence). Modeling regime shifts and environmental influences as ‘autocorrelated recruitment’ may result in wider forecasting intervals compared with assuming that recruitment follows a white-noise process. This wider forecast interval may, in some cases, have better statistical coverage (i.e., a 75% forecast interval that contains the true value 75% of the time). 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 factorial design involving five plausible levels of recruitment autocorrelation, and several alternative configurations for the assessment model (e.g., ignoring autocorrelation, estimating an autocorrelation parameter internally or externally to the assessment, or fixing the autocorrelation parameter at its true value). We explore model 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 recruitment autocorrelation?

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

**2. Methods**

We conduct this simulation experiment using the Stock Synthesis (SS) assessment software (Methot and Wetzel 2013), which is widely used in the U.S. and provides a generic implementation of an integrated assessment model. SS estimates recruitment at the same time as other parameters which govern stock productivity and status, therefore allowing uncertainty about past recruitment to be propagated into rebuilding forecasts. Simulations and analyses are accomplished using the *ss3sim* software package (Anderson et al. 2014a, 2014b), and publicly reposit our simulation code online (XXXX) to ensure that 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 model that is applied to the simulated data, where the parameter estimates and derived quantities (i.e., population abundance during a forecast period) from the estimation method can be compared with their true values from the operating model. We use a factorial design involving five scenarios, each with a different level of recruitment autocorrelation, and four alternative configurations for the assessment model (i..e, ignoring autocorrelation, estimating an autocorrelation parameter internally or externally to the assessment, or fixing the autocorrelation parameter at its true value). We use 100 simulation replicates for each scenario, where each replicate has a different realization of process (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” – Tears 26-80 include a simulated fishery and survey, which generate data for an assessment model conducted in year 80; and
3. “Forecast period” – Years 81-100 are a forecast period, such that a forecast of population abundance from the assessment model conducted in year 80 can be compared with simulated abundance.

**2.1 Operating model**

The operating model (OM) represents a cod-like (i.e., slow-growing and long-lived) life history based on North Sea cod (*Gadus morhua*; R. Methot, NMFS, NOAA, pers. comm.). The operating model used biological parameters estimated from its respective stock assessment with some simplifications facilitating interpretation of the results (e.g., one fishery fleet and one survey fleet, combined sexes, and fishery selectivity that mirrors the maturity ogive; Table 1).

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

(1)

where *rt* and *bt* are the estimate of recruitment and spawning output, respectively, in year *t*, *h* and *r0* are estimated parameters representing steepness (the strength of recruitment compensation) and average recruitment at unfished spawning biomass *b0*, and recruitment deviation *εt* is calculated as:

(2)

where *δt* is a normally distributed coefficient representing recruitment variability:

(3)

where is the marginal variance of recruitment deviations and *ρ* is the magnitude of autocorrelation in recruitment. Equation (2) includes the term , where bias-correction term is included to ensure that *r0* is equal to the mean (not the median) recruitment given unfished spawning output (Methot and Taylor 2011).

Each replicate of the operating model involved simulating true dynamics over 100 years, where recruitment variation is simulated for all 100 years. 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 subsequent years, fully-selected fishing mortality, *F*, was fixed at the value that produced maximum sustainable yield. 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 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.

Tested values of autocorrelation, *ρ*, were -0.25, 0, 0.5, 0.75 and 0.9, where these values are centered approximately around estimates from recent meta-analysis (Mueter et al. 2007, Thorson et al. 2014). Random draws for process errors were the same across scenarios given an iteration (i.e., the values of for the first replicate of the *ρ* = 0 scenario were the same as for the first replicate of the *ρ* = 0.9 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; see Fig. 2). Fishery length- and age-composition data were simulated every other year for years 40-80, and were drawn from a multinomial distribution with an annual sample size of 25. Survey length- and age-composition data were simulated every other year for years 41-79, and were drawn from a multinomial distribution with an annual sample size of 100. The survey was simulated as providing an index of relative abundance for every other years for years 41-79. and was drawn from a lognormal distribution with log-standard deviation of 0.1.

**2.3 Estimation model**

The structure of the estimation method was based on that of the operating model, except in terms of autocorrelation (see Table 1 for a list of estimated parameters). Each operating model generated 100 years of simulated data though the estimation method was fit to data only from years 25 to 80. Estimation methids produced forecasts for years 81 to 100, while operating models generated data for years 1 to 100. During the forecasts, *F* was set to the estimated *F­MSY* from the estimation methods, thus forecasts are subject to errors from estimates of derived quantities and parameters governing the population dynamics of the system. For years during which composition sampling took place, the correct effective sample sizes were specified in all estimation models. Bias correction for the deviations about the stock-recruitment relationship was based on Methot and Taylor (2013). Estimation method convergence was evaluated using the maximum gradient of the objective function, where models with a maximum gradient of greater than 0.01 were removed from the analysis. Additionally, replicates with parameters estimated at either of their boundary conditions (Table 1) were also removed.

*Model Validation*

To verify that the EM can reproduce OM values, we performed #X simulations with highly informative data (multinomial sample size of 2000 for age and length composition data, and survey CV of 0.1). For these preliminary runs, the estimation model was able to estimate biomass without bias.

The process was repeated for each scenario with 100 age and length multinomial composition samples and an index of abundance with a CV of 0.1.Estimates of spawning biomass, SPB, and F from the estimation model were compared to the operating model’s estimates of SPB and F (Figures 3 and 4).

**Results**

*Single Replicate Results*

Estimates of spawning stock biomass, SSB, were extracted together with their asymptotic (delta method) 25%, 50%, and 95% confidence intervals from the estimation method for a single replicate (Figures 5 and 6). When AR is negative or zero, internal estimates of spawning stock biomass have comparatively low spread of uncertainty during the forecast period (Figures 5 and 6). As ρ increases, uncertainty around the time-series of estimates of SSB increases during the forecast period (years 80-100). This is especially true when ρ is fixed at its true value. While fixing ρ at zero gives a smaller interval of uncertainty in the estimate of SSB, the confidence interval generated by estimation method is unlikely to contain the true SSB as given by the operating model.

*Multi-Replicate Results*

Autocorrelation, AR, was estimated internally in SS with lower and upper bound specified in Table 1. The distribution of estimates of ρ produced by SS when ρ is estimated internally versus performance of the *acf* function (Venables and Ripley 2002) applied to the time-series of recruitment deviations produced by the estimation model when internal estimation of ρ was done by SS. Internal estimates of ρ are biased toward zero, but estimates of ρ calculated externally based on internally-estimated recruitment deviations produced by Stock Synthesis (Figure 7). Additionally, forecast coverage is better when ρ is fixed at its true value, while fixing ρ at 0 leads to strong bias in estimates of SSB during the forecast period when ρ is high (ρ > 0.5 in particular) (Figure 8).

**Discussion**

Without correcting our issue with bias in F during the forecast period, I’m hesitant to draw many strong conclusions at this time, but here are a few thoughts on how I foresee the discussion panning out, nonetheless.

* Perhaps estimating AR internally is not particularly helpful as SS seems to not produce an estimate of AR that is reliable. However, our external calculation of AR from rec devs estimated by SS are significantly less biased for relatively small values of AR. Perhaps it would be better to have AR as a derived quantity in SS instead of an estimable parameter.
* The seemingly default practice of fixing AR at zero may be leading to erroneous forecasts of SPB as the forecast coverage in this case was shown to be increasingly poor as forecast length increases despite forecast uncertainty remaining relatively low. Perhaps a better solution is to advocate for a double-pass estimation process in which AR is allowed to be estimated internally on the first pass. The second pass would estimate AR externally (or extract a derived quantity estimate of AR from a revised report file), fix AR at this value, and rerun SS to obtain the final estimates of biomass and other relevant estimated metrics. Our simulations show that this two-pass method increases the quality of the predictive coverage of SS at the expense of some tightness of the uncertainty interval.
* The effects of erroneous estimates of SPB in the forecast period on management proxies and MSY-based decision table hasn’t been explored by us just yet, but this task is on my list to complete before we submit the paper for peer-review and publication. Ideally, I’d like to generate a set of Kobe plots (or similar) showing how well our two-pass system compares to the default practice of fixing AR at zero when it comes to actual management advice.
* Future work directions include repeating our analysis utilizing Jim’s Laplace method to test how well it performs in comparison to the methods developed in this paper. Additional suggested follow-up includes repeating the analysis for non-cod-like species (specifically, sardines and other fast-growing short-lived species) to verify that our results hold across a range of life-history types.
* So far, a lot of this seems like an SS (or ss3sim) debugging exercise. I recognize that some things still need to be resolved, but we should keep in mind how to sell this message as ‘general’.

**Acknowledgements**

We gratefully acknowledge help from Ian Taylor, Chantel Wetzel, and Allan Hicks on implementation issues regarding ss3sim and forecasting within Stock Synthesis. This publication is partially funded by the Joint Institute for the Study of the Atmosphere and Ocean (JISAO) under NOAA Cooperative Agreement No. ?, Contribution No. ?. Partial funding for KFJ was provided by Washington Sea Grant.

**References**

Aanes, S. and Pennington, M. 2003. On estimating the age composition of the commercial catch of Northeast Arctic cod from a sample of clusters. ICES Journal of Marine Science, 60: 297-303.

Anderson, S. C., Monnahan, C. C., Johnson, K. F., Ono, K., and 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., and Whitten, A. R. 2014b. ss3sim: Fisheries stock assessment simulation testing with Stock Synthesis. R package version 0.8.9.9.

Clark, W. G. 1991. Groundfish exploitation rates based on life history parameters. Canadian Journal of Fisheries and Aquatic Sciences, 48: 734-750.

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

Hulson, P-J. F., Hanselman, D. H., and Quinn, T. J., II. 2011. Effects of process and observation errors on effective sample size of fishery and survey age and length composition using variance ration and likelihood methods. ICES Journal of Marine Science, 68: 1548-1557.

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

Maunder, M. N. 2011. Review and evaluation of likelihood functions for composition data in stock-assessment models: estimating the effective sample size. Fisheries Research, 109: 92-99.

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. Canadian Journal of Fisheries and Aquatic Sciences, 64: 911-927.

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

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

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. Fisheries Research, 109: 320-329.

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

Schirripa, M. J., Goodyear, C. P., and Methot, R. M. 2009. Testing different methods of incorporating climate data into the assessment of US West Coast sablefish. ICES Journal of Marine Science, 66: 1605-1613.

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

Thorson, J.T., Jensen, O.P., and 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**(7): 973–983. doi: 10.1139/cjfas-2013-0645.

Venables, W. N. and Ripley, B. D. (2002) Modern Applied Statistics with S. Fourth Edition. Springer-Verlag.

Vert-pre, K. A, Amoroso, R. O., Jensen, O. P., and Hilborn, R. 2013. Frequency and intensity of productivity regime shifts in marine fish stocks. Proceedings of the National Academy of Sciences, 110: 1779-1784.

Table : Parameters used in the operating model (OM) or estimation model (EM), listing the true value in the operating model and whether the model is fixed at the true value or freely estimated in the estimation model

|  |  |  |  |
| --- | --- | --- | --- |
| Parameter |  | OM | EM |
| Name | Symbol | True value | Fix or estimate? |
| Natural mortality rate | *M* | 0.2 yr-1 | Fix |
| Length at age 1 | *La=1* | 20 cm | Fix |
| Asymptotic maximum length | *L∞* | 132 cm | Fix |
| Von Bertalanffy growth coefficient | *k* | 0.2 yr-1 | Fix |
| Coefficient of variation for length at age | *CVL* | 0.1 | Fix |
| Length of 50% maturity |  | 38.2 cm | Fix |
| Length at 95% maturity |  | 48.9 cm | Fix |
| Average recruits for the unfished population (natural log) | *ln(R0)* | 18.7 | Est |
| Steepness of the Beverton-Holt stock recruit function | *h* | 0.65 | Fix |
| Marginal log-standard deviation of recruitment | *σR* | 0.4 | Fix |
| Magnitude of autocorrelated recruitment | *ρ* | varies | varies |
| Random coefficients for recruitment variability (years 1-100) | *δt* | varies | Est |
| Catchability coefficient for survey index of abundance (natural log) | *ln(q)* | 0 | Est |
| Length of 50% selection in the fishery |  | 38.2 cm | Est |
| Length of 95% selection in the fishery |  | 48.9 cm | Est |
| Length of 50% selection in the survey |  | 30.6 cm | Est |
| Length of 95% selection in the survey |  | 39.1 cm | Est |

|  |  |
| --- | --- |
| Run | Tested setting for AR |
| Internal | ρ estimated within estimation model internally in SS |
| External | ρ is fixed at the calculated autocorrelation estimated using recruitment deviations produced by the estimation model |
| True | ρ is fixed at the true value of AR used to generate recruitment deviations in the operating model |
| Zero | ρ is fixed at zero |

Table 2: Summary of four estimation models, where each model estimates the recruitment autocorrelation (AR) parameter differently.

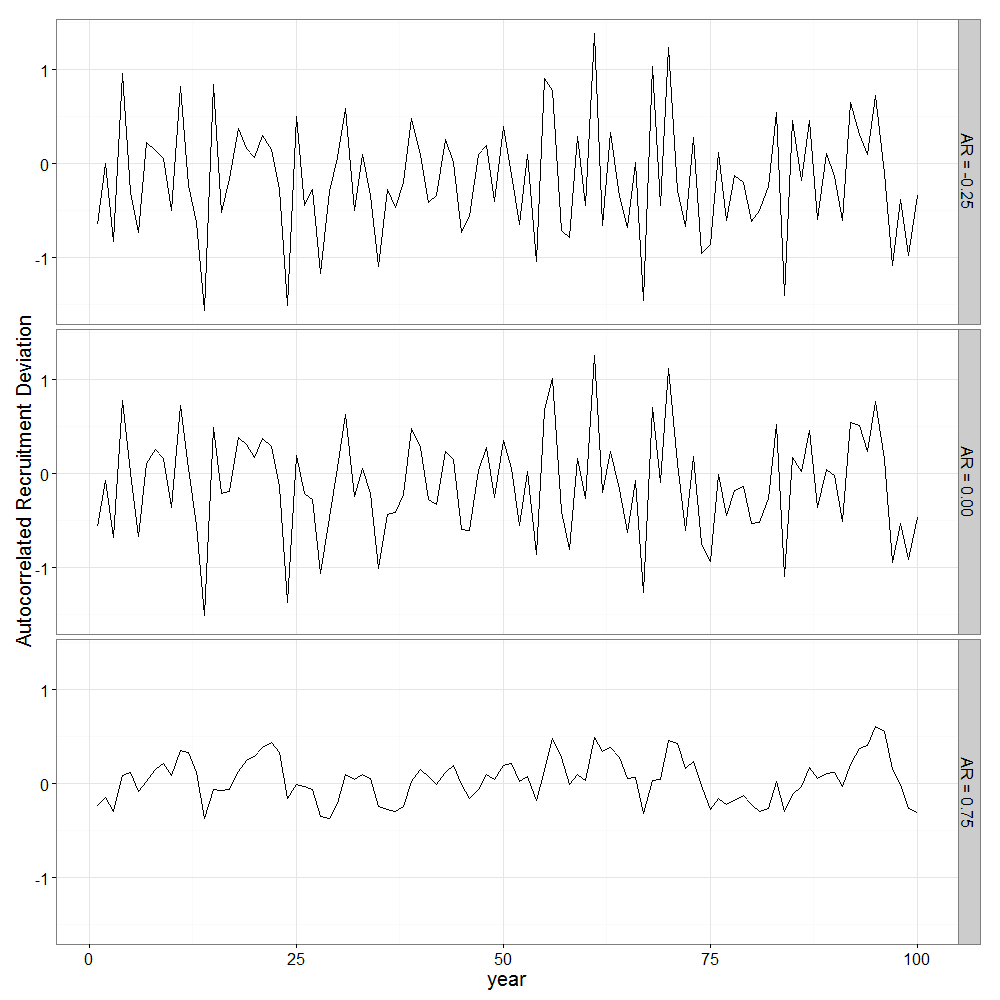


Figure 1: Examples of autocorrelated recruitment deviations for three levels of autocorrelation (-0.25, 0.00, and 0.75; top to bottom respectively), where each example used the same set of process error deviations ().

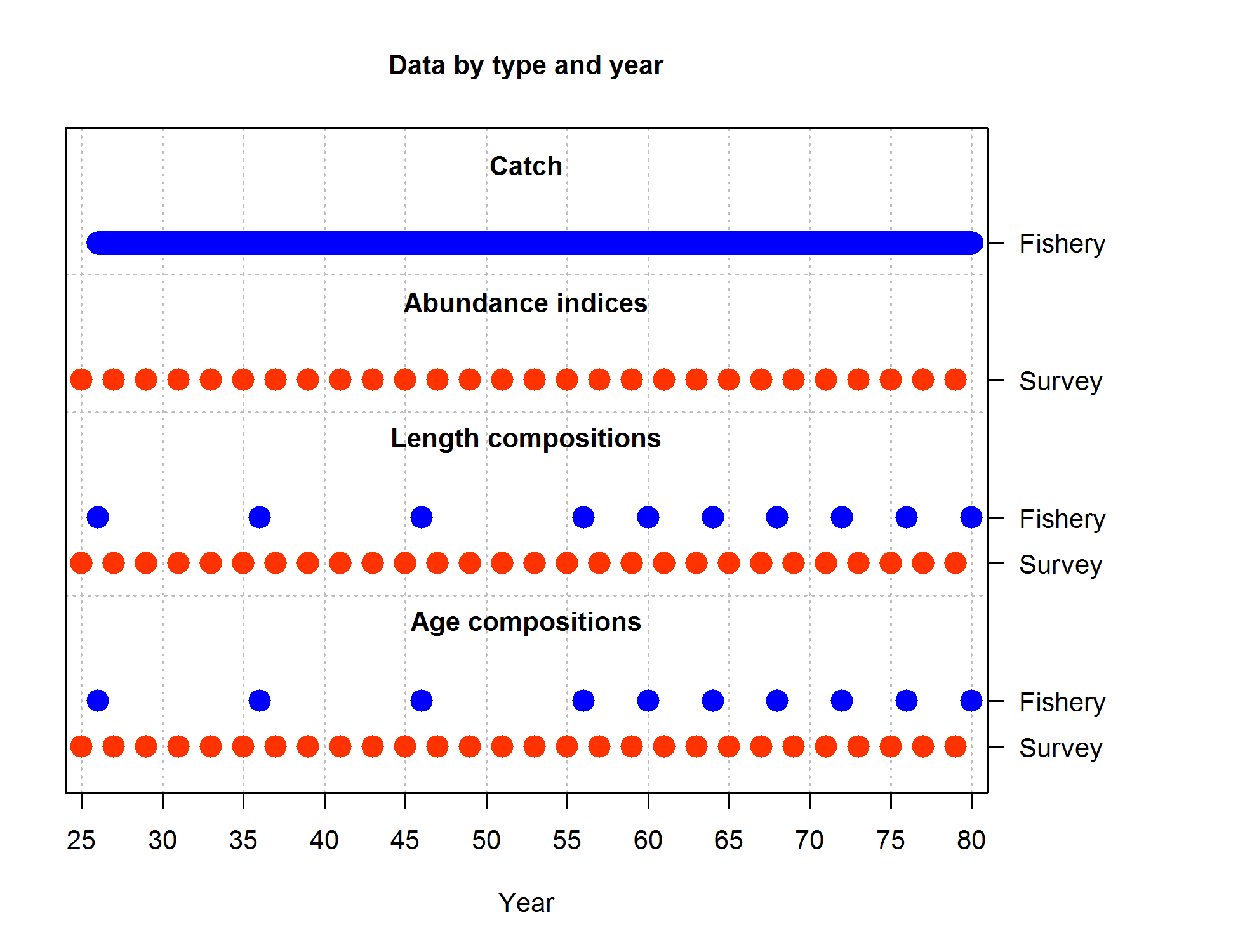
****

Figure 2: Summary of simulated data available to the estimation method. Data availability was consistent across all scenarios and 100 multinomial composition samples were generated each year compositions were available.

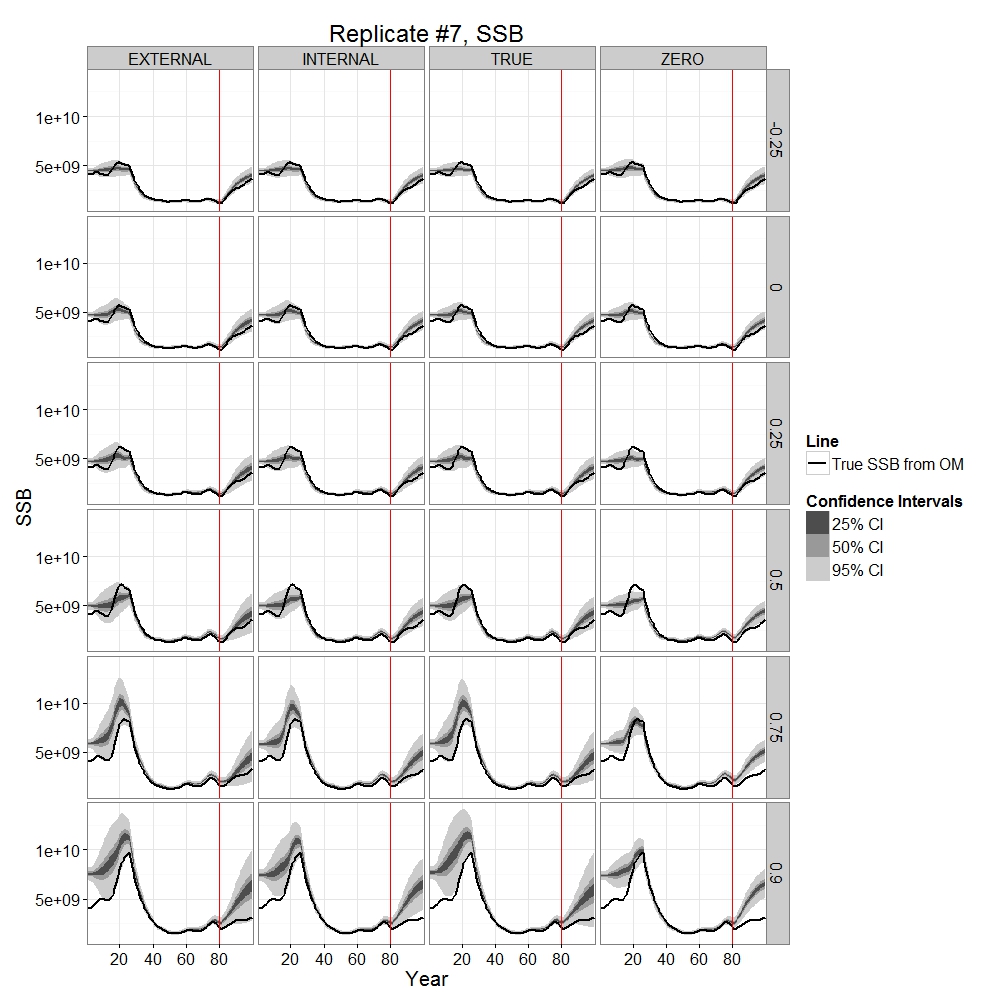


Figure 3: Illustration of estimated spawning stock biomass (SSB) during 100 simulated years, where the stock assessment in year 80 is indicated with a vertical red line (i.e., the model is forecasting without data for years 81-100). Rows show results for different scenarios (where recruitment autocorrelation is ρ={-0.25, 0.0, 0.25, 0.5, 0.75, 0.9}), while columns show four different estimation models (external estimation, international estimation, fixing at the true value, or fixing at zero), and each panel shows the true spawning biomass (black line) and grey shading shows the confidence and forecasting interval for the estimating spawning biomass.

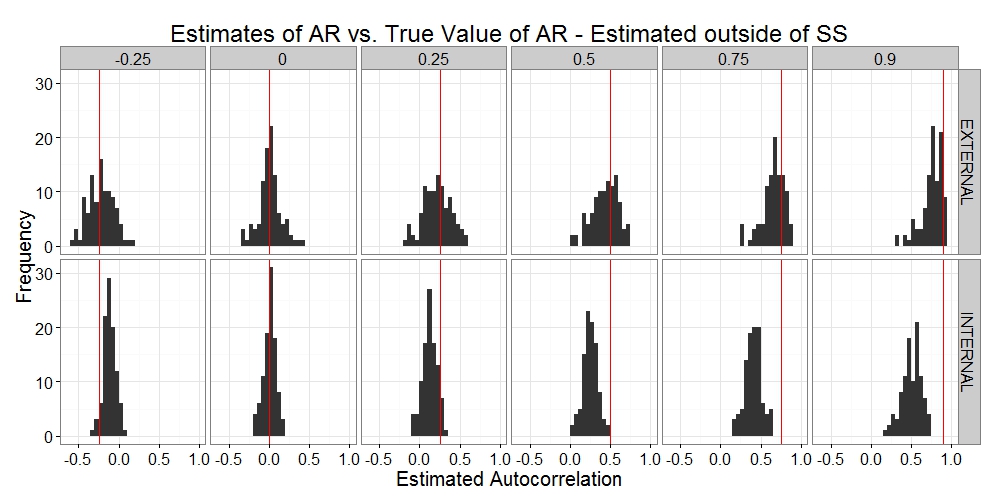


Figure 4: Estimates of recruitment autocorrelation (*ρ*) produced via external estimation of autocorrelation from estimated recruitment deviations (top row) or via estimates internally as a parameter in Stock Synthesis (bottom row), for different true levels of autocorrelation (different columns), where the red line in each panel shows true value of autocorrelation for that scenario.

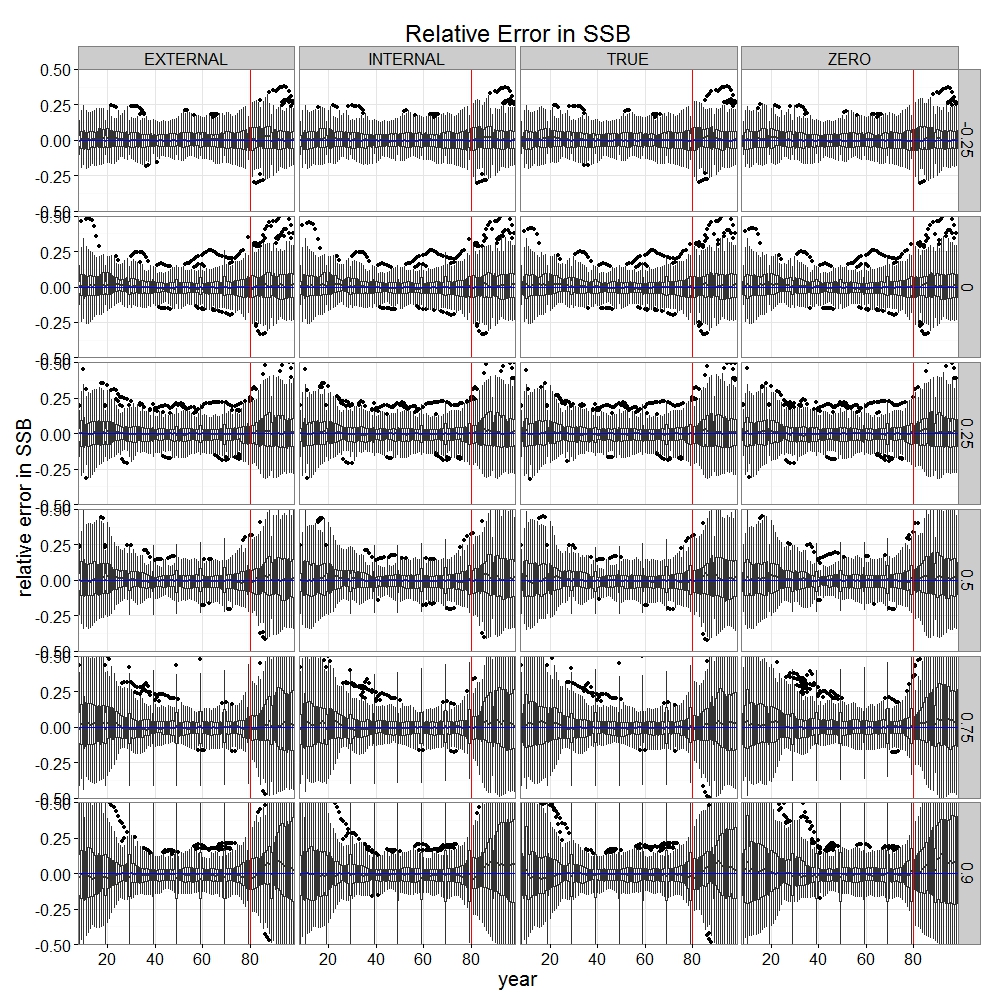


Figure 5: Timeseries of boxplots showing relative error in spawning stock biomass (SSB) for variety of scenarios by year. Red lines show the year at which forecasts begin. Horizontal index on panels indicate scenario (see Table 2) and vertical index on panels indicate true value of AR.

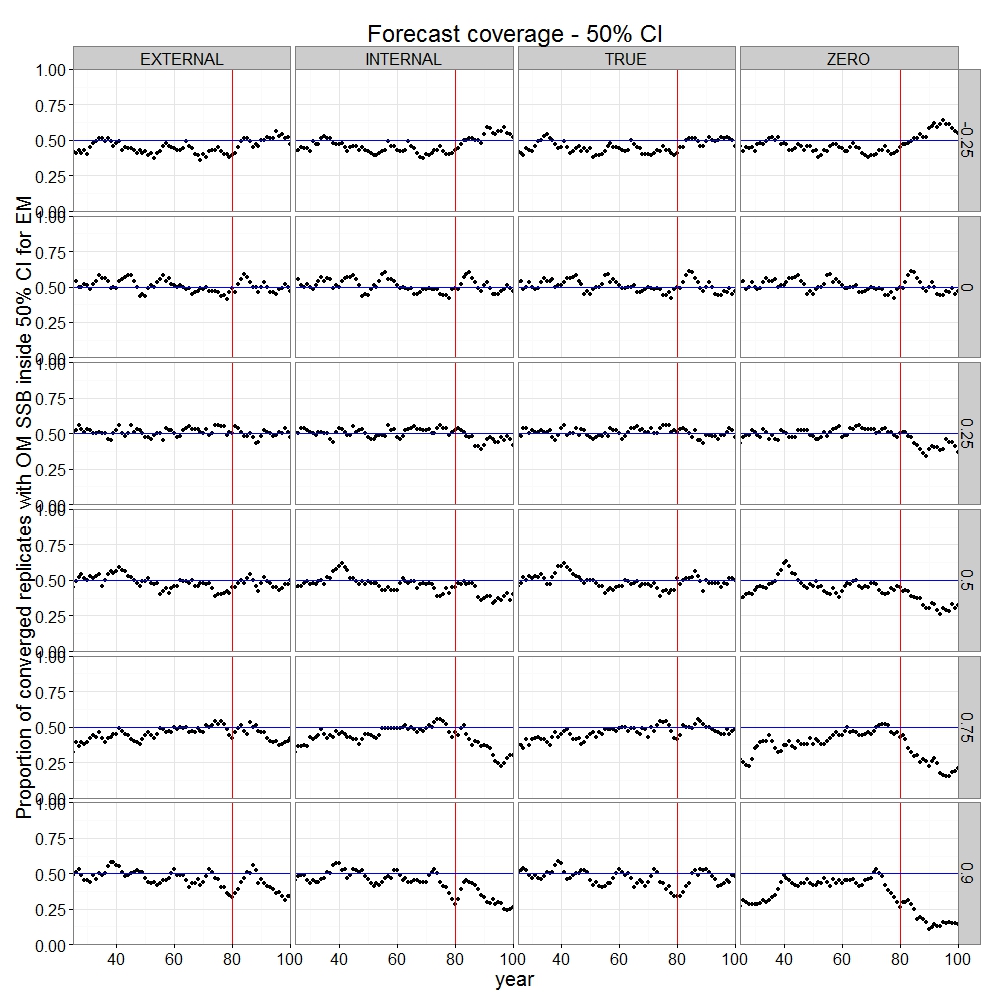


Figure 6: Proportion of converged replicates whose true value of SSB lies outside the 50% confidence interval estimated by the estimation method.