**How well can autocorrelated recruitment be estimated for individual stocks, and how does it affect stock forecasting?**

E Councill1,2\*, K Johnson2, J Thorson1, L Brooks3, R Methot1, A Punt2

1Fisheries 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

2School of Aquatic and Fishery Sciences, University of Washington, Box 355020, Seattle, WA 98195-5020, USA

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

\*Corresponding author: tel: +1 206 860 2425; fax: +1 206 ??? ????; email: Elizabeth.councill@noaa.gov

**Abstract**

Patterns of autocorrelation, AR, in recruitment deviations, often identified by periods of time where recruitment deviations are positive or negative for several years in a row, can appear due to numerous factors including regime shifts and periodicity in environmental drivers affecting recruit survivorship. The ability of stock assessments to accurately characterize autocorrelation and its effect on the quality of forecasts of spawning biomass generally remains unknown. Monte Carlo simulations were used to test how well Stock Synthesis, SS, an integrated age-structured stock assessment software package used extensively in the management of fish stocks, estimates AR in the presence of a range of autocorrelated recruitment deviations. The precision and accuracy of quantities of interest to management 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 their true value, internally estimated, and input as a fixed value determined from an external estimation procedure. Estimates of AR produced by SS internally were biased toward zero when AR was larger than zero but were unbiased when the AR was less than or equal to zero. However, a reasonably unbiased estimate of AR was obtained by estimating it directly from the recruitment deviations produced by SS’s internal estimation procedure. Additionally, estimates of spawning biomass during the forecast period were more uncertain when AR was high. Our results suggest that users first estimate AR internally within the stock assessment framework, then externally calculate AR from the estimated recruitment deviations, especially in cases when the internally-estimated value is positive and nonzero. Results using this method produce estimates of AR that are unbiased except in cases in which the true value of AR is large (e.g. AR > 0. 5).

**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 plan that is more likely than not to succeed in rebuilding the stock to biomasses associated with maximum sustainable yield () within 10 years or 1 generation plus the time for rebuilding given no fishing. Additionally, the National Marine Fisheries Service (NMFS) must identify target and limit reference points for all stocks included in fishery management plans, where reference points are defined probabilistically based on stock dynamics. Reference points and rebuilding forecasts are often estimated using an age-, size-, or stage-structured population dynamics model that treats fluctuations in recruitment as a random process around a mean or pre-specified spawner-recruitment relationship. Often, reference points are not estimable and thus proxies must be identified. Unfortunately, proxies make a similar assumption that yearly estimates of recruitment are statistically-independent, and fluctuate around a time-invariant process.

However, recent studies illustrate the prevalence of regime shifts in the underlying stock-recruit relationship (Vert-pre et al. 2013). From a management perspective, issues surrounding regimes are more related to forecasting than their existence. Ideally, researchers can identify measureable environmental factors that are correlated with recruitment variability or regime shifts, and which can be forecast into the future. Consequently, the environmental factor can facilitate the appropriate characterization of future recruitment variability, thus influencing rebuilding forecasts (Holt and Punt 2009, Punt 2011) and reference point determination (Schirripa et al. 2009). Alternatively, population forecasts can be ‘state-of-nature-dependent’ such that multiple scenarios are presented, each of which is dependent upon a hypothetical scenario for future recruitment (e.g., high, average, and low productivity scenarios).

Regime shifts can appear as autocorrelated recruitment (i.e., where recruitment deviations are greater or less than zero for many years in a sequence) when interpreted using a model that assumes constant average recruitment or a constant spawner-recruit relationship. Where measurable environmental factors remain unidentified, modeling regime shifts and environmental influences as ‘autocorrelated recruitment’ (and given a fixed level of recruitment variance) will likely result in wider forecasting intervals compared with assuming recruitment follows a white-noise process. This wider forecast interval may, in some cases, have better statistical coverage (i.e., a 75% forecast interval may include 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), and recruitment autocorrelation will also influence commonly-used proxies for reference points (Clark 1991).

As NMFS works to reduce the number of overfished stocks, projection success is being examined more critically, and the accuracy of those projections in meeting rebuilding targets could certainly be improved (Neubauer et al. 2013, NRC 2013). Accounting for autocorrelated recruitment and resulting streaks of poor or strong recruitment could help improve the accuracy of forecasting, as well as improve the success of rebuilding stocks within forecasted timelines.

We explore and test the estimation performance of an integrated age-structured assessment model when recruitment is autocorrelated. We use a factorial design involving five plausible levels of recruitment autocorrelation, and several alternative configurations for the assessment model (e.g., ignoring autocorrelation, estimating the autocorrelation parameter internally or externally, or fixing the autocorrelation parameter at its true value). We explore model performance by answering two questions: how well can the magnitude of autocorrelation be estimated; and does accounting for autocorrelation improve the accuracy and predictive coverage of forecasts compared with ignoring recruitment autocorrelation? We conduct this exploration 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 integrated assessment modeling. 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.

**Methods**

Simulations and analyses were accomplished via the *ss3sim* software package (Anderson et al. 2014a, 2014b), an R package (R Core Team 2014) that facilitates rapid end-to-end simulation testing with SS (Methot and Wetzel 2013). *ss3sim* allows for reproducible simulation testing of SS, and we used a modified version of SS (v3.24f), where these modifications are incorporated in all versions later than v3.24o.

The simulation framework consists of three components: an operating model that specifies the “true” system dynamics, simulated data generated by the operating model, and an estimation model fit 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. Each combination of operating model and estimation model, hereinafter referred to as scenario, was replicated 50 times, each with different process (recruitment deviations) and observation errors.

**Operating model**

The operating model 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., single fleet, combined sexes, and selectivity which mirrors the maturity ogive; Table 1).

Five configurations of the operating model were used to produce autocorrelated recruitment deviations:

(1)

where the magnitude of first order autocorrelation (ρ) varied among configurations (Table 1); and are observed log-transformed residuals around the stock recruitment curve, assumed to be of Beverton-Holt form, corresponding to years and respectively; and is a process error coefficient drawn from a normal distribution with mean zero and standard deviation representing uncorrelated errors in year (Figure 1). Equation (1) is modified by subtracting so has an expected value of one after exponentiation.

For all scenarios, years 1 through 25 had zero fishing and acted as a burn-in period, which ensured that the population age-structured started with plausible deviations away from its expectation in an unfished state. In subsequent years, fully-selected fishing mortality, *F*, was constant and equal to the value that produced maximum sustainable yield. Fishery selectivity was length-based, mirrored the maturity ogive (asymptotic), and was time-invariant. 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, AR, 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). Recruitment deviations (process errors) were the same across scenarios given an iteration (e.g., recruitment deviations produced for AR = 0 and replicate 1 are based on the same random number seed as for the AR = 0.9 case for replicate 1).

**Data generation**

Catch was reported yearly without error from the start of the fishery to year 80, where year 80 represents the last year where data was available to the estimation model. Fishery length- and age-composition data were generated every 10 years for the first 30 years and then every 4 years until the terminal year to emulate an increase emphasis on obtaining composition data. A multinomial sample size of 100 fishery length- and age-compositions were sampled each year sampling took place. An index of abundance was generated using the lognormal distribution with a log-standard deviation of 0.2. Surveys occurred every other year beginning in year 25. Survey length- and age-compositions were generated every two years starting in year 25 with the same multinomial sample size of 100 per year (Figure 2). Survey and fishery composition data were generated using the multinomial distribution, which assumes homogenous capture probabilities across bins and perfect mixing.

**Estimation model**

The structure of the estimation model was based on that of the corresponding 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 each estimation was fit to data only from years 25 to 80. Estimation models 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 EM, thus forecasts are subject to errors from estimates of derived quantities and parameters governing the population dynamics of the system. For years in which composition sampling took place, the correct effective sample sizes were specified in all estimation models. Bias correction was done via Methot and Taylor (2013). Estimation model convergence was evaluated using the maximum gradient of the objective function, where models with a 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 ensure convergence properties of the model were realistic, provided an appropriate amount of data was available, preliminary runs included 2000 age and length multinomial compositions samples sampled biannually from years 25-80 and an index of abundance for those same years with a 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 bias-checked against the operating model’s estimates of SPB and F (Figures 3 and 4). While estimates of SPB were relatively unbiased (Figure 3), estimates of F during the forecast period were biased low (Figure 4).

**Results**

*Single Replicate Results*

Estimates of spawning biomass, SPB, were extracted together with their 25%, 50%, and 95% confidence intervals from the estimation model for a single replicate (Figures 5 and 6). When AR is negative or zero, internal estimates of spawning biomass have comparatively low spread of uncertainty during the forecast period (Figures 5 and 6). As AR increases, uncertainty around the timeseries estimates of SPB increases during the forecast period (years 80-100). This is especially true when AR is fixed at its true value. While fixing AR at zero gives a smaller interval of uncertainty in the estimate of SPB, the confidence interval generated by estimation model is unlikely to contain the true SPB 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 AR produced by SS when AR is estimated internally versus performance of the *acf* function (Venables and Ripley 2002) applied to the timeseries of recruitment deviations produced by the estimation model when internal estimation of AR was done by SS. Internal estimates of AR are biased toward zero, but estimates of AR calculated externally based on internally-estimated recruitment deviations produced by Stock Synthesis (Figure 7). Additionally, forecast coverage is better when AR is fixed at its true value, while fixing AR at 0 leased to strong bias in estimates of SPB during the forecast period when AR is high (AR > 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.

**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.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Parameter | Operating Model | Estimation Model | | |
|  | Initial | Phase | Min bound | Max bound |
| L\_at\_Amin\_Fem\_GP\_1 | 20 | 2 | 1 | 100 |
| L\_at\_Amax\_Fem\_GP\_1 | 132 | 1 | 6.6 | 660 |
| VonBert\_K\_Fem\_GP\_1 | 0.2 | 2 | 0.01 | 1 |
| CV\_young\_Fem\_GP\_1 | 0.1 | 1 | 0.01 | 0.5 |
| CV\_old\_Fem\_GP\_1 | 0.1 | 1 | 0.01 | 0.5 |
| SR\_LN(R0) | 18.7 | 3 | 4 | 20 |
| AR | -0.25, 0.00, 0.25, 0.50, 0.90 | 5 | -1 | 1 |
| Q\_base\_2\_Survey | 0 | 1 | -20 | 20 |
| SizeSel\_1P\_1\_Fishery | 50.8 | 5 | 5.08 | 101.6 |
| SizeSel\_1P\_3\_Fishery | 5.1 | 4 | 0 | 25.5 |
| SizeSel\_2P\_1\_Survey | 41.8 | 2 | 4.18 | 83.6 |
| SizeSel\_2P\_3\_Survey | 5.2 | 3 |  |  |

Table 1: Life history, fishery, and modelling parameters used for the operating model and estimation model, where parameters were initiated at their true value from the operating model in each estimation model. Estimation of AR varies by scenario (see Table 2).

|  |  |
| --- | --- |
| Run | Tested setting for AR |
| Internal | AR estimated within estimation model internally in SS as in Table 1 |
| External | AR is fixed at the calculated autocorrelation estimated using recruitment deviations produced by the estimation model |
| True | AR is fixed at the true value of AR used to generate recruitment deviations in the operating model |
| Zero | AR is fixed at zero |

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

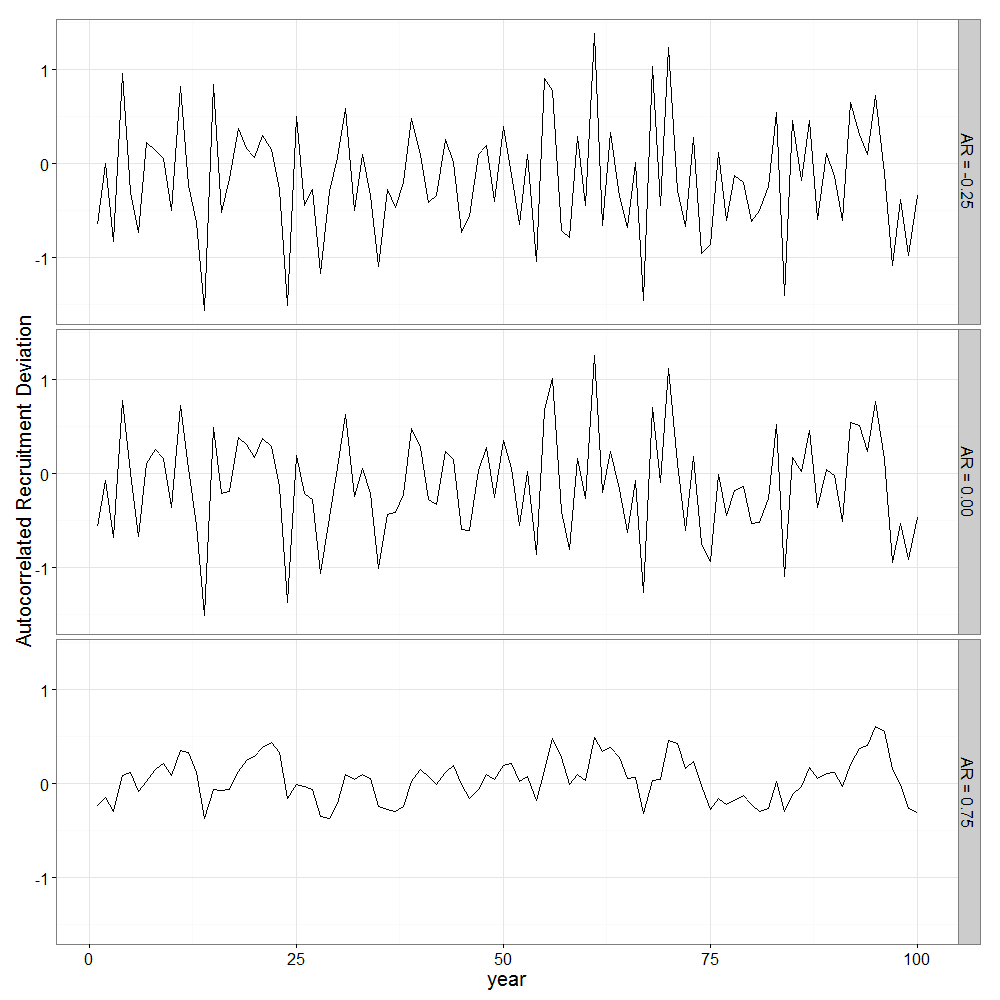


Figure 1: Examples of autocorrelated recruitment deviations for three levels of autocorrelation (-0.25, 0.00, and 0.75; top to bottom respectively), where each exampled used the same set of process error deviations ().

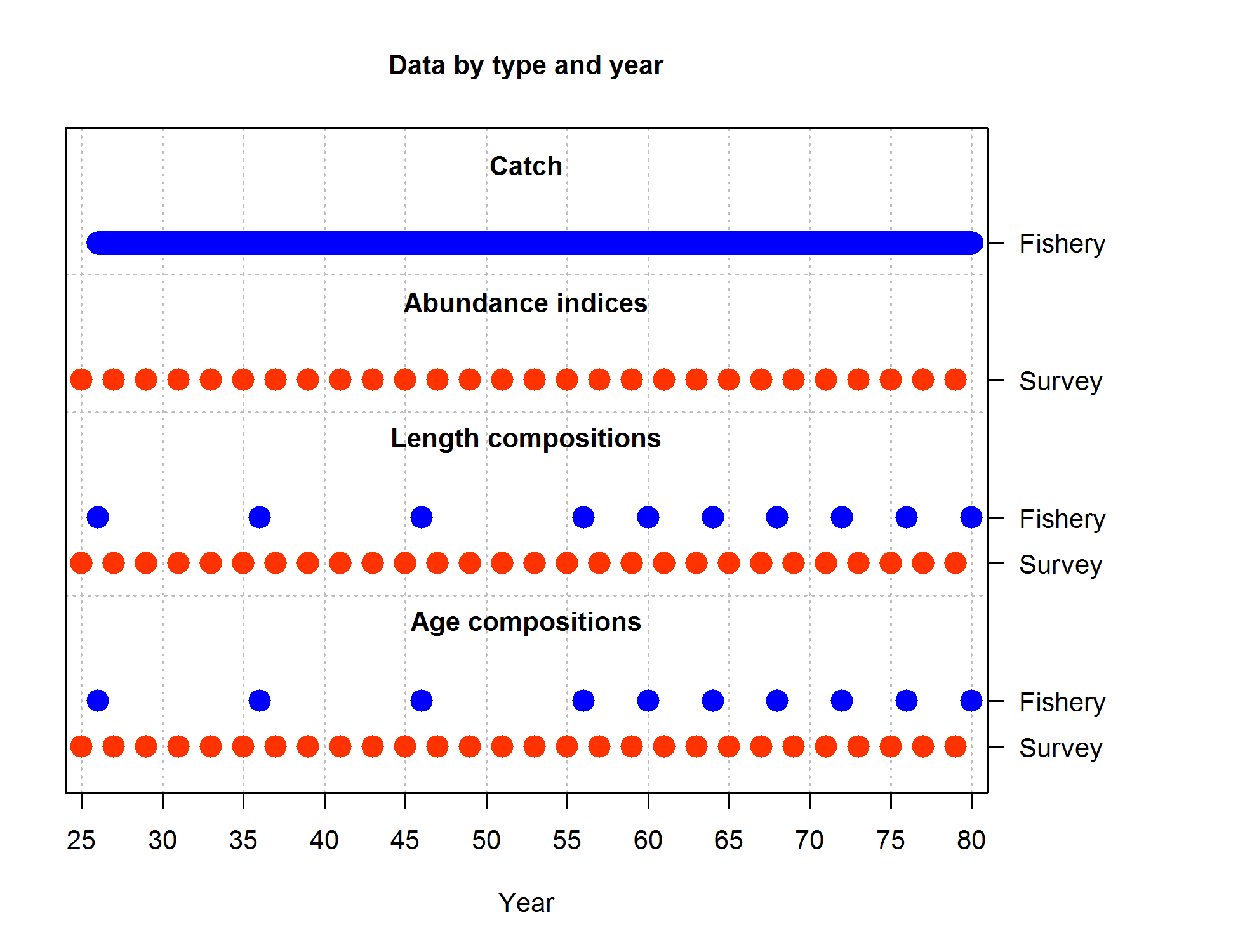
****

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.

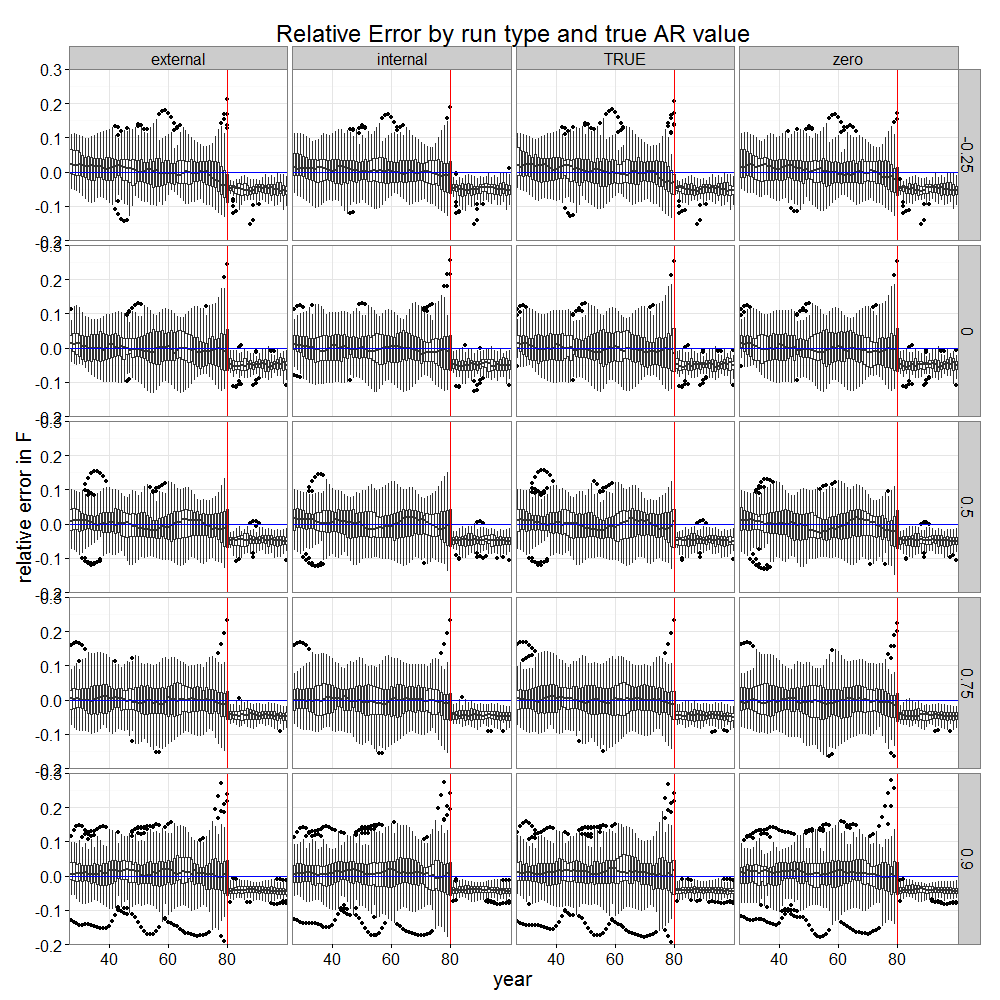


Figure 3: Boxplots showing relative error in F between operating and estimation models. Red lines show the year at which forecasts begin. Horizontal index on panels indicate scenario (see Table 2) and vertical index on panels indicate true value of AR.

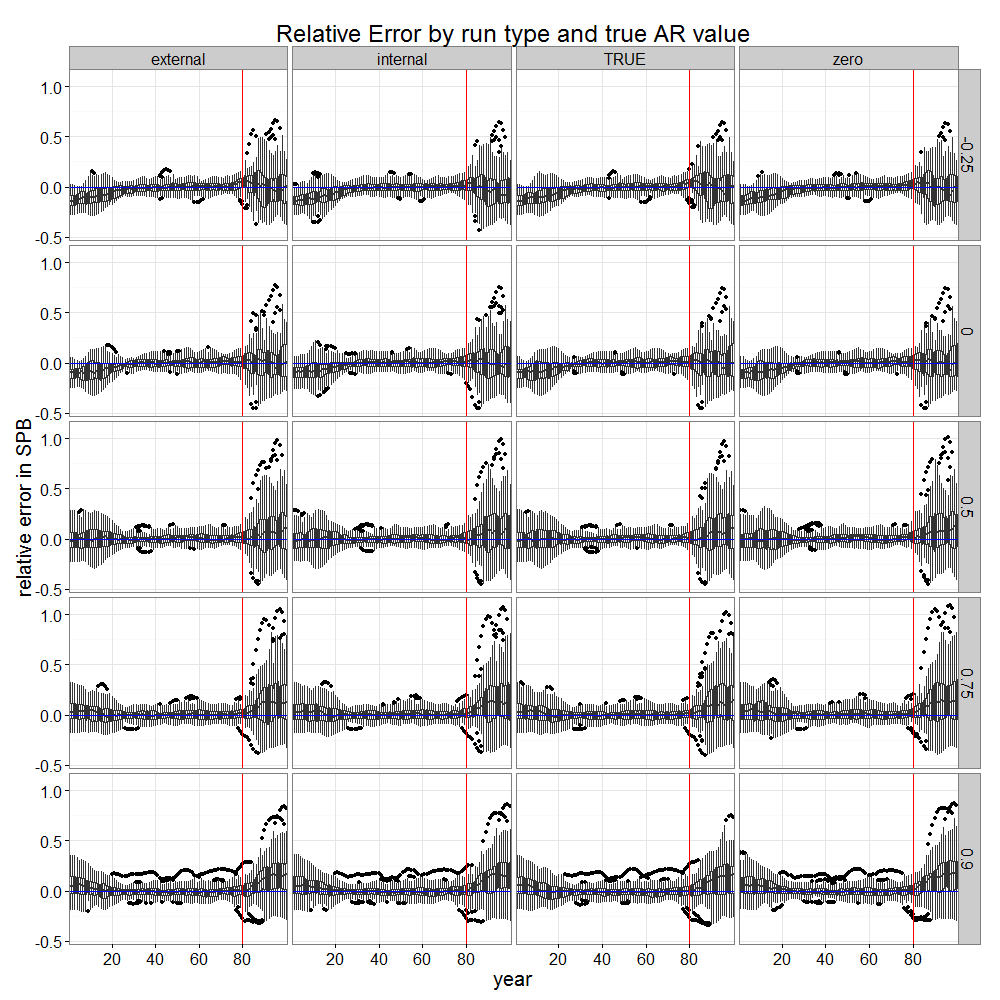


Figure 4: Timeseries of boxplots showing relative error in spawning biomass (SPB) for variety of scenarios by year. Red lines show the year at which forecasts begin. Horizontal index on panels indicate scenario (see Table 2) and vertical index on panels indicate true value of AR.

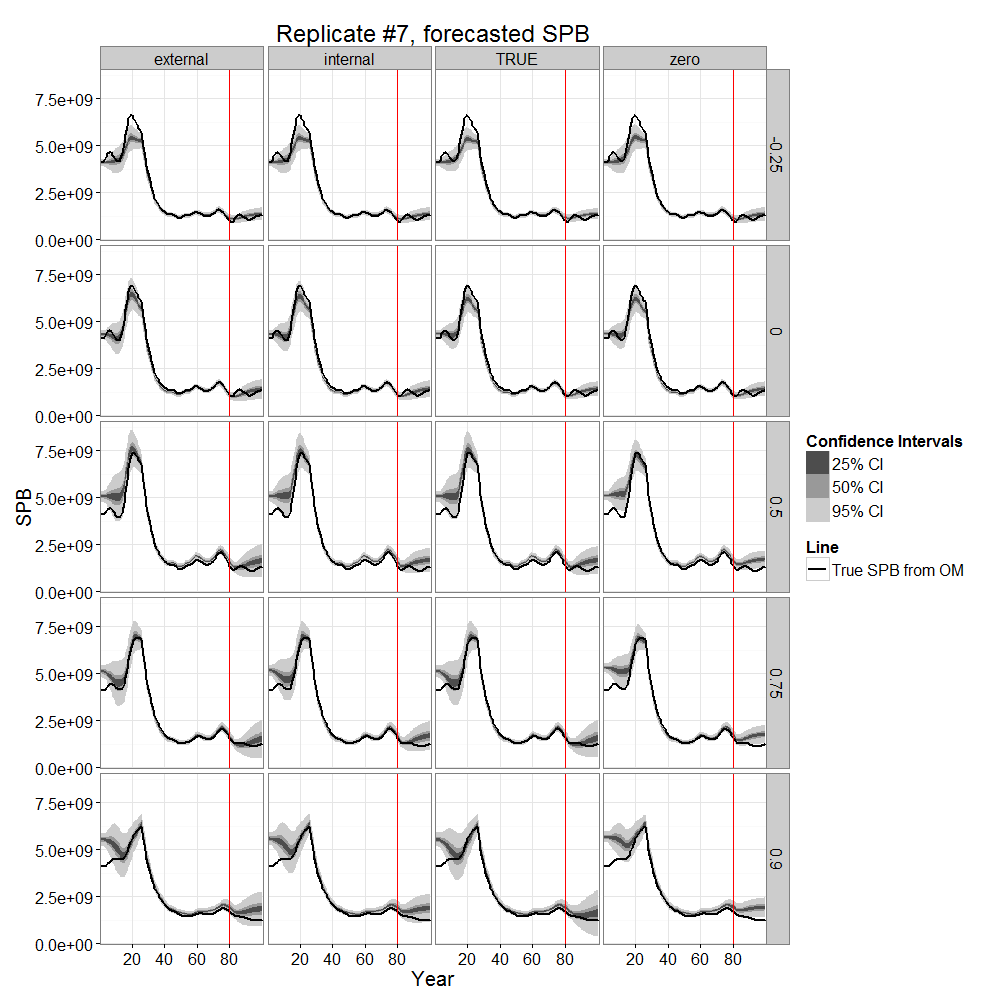


Figure 5: Single-replicate estimate of SPB over time. Red lines show the year at which forecasts begin. Horizontal index on panels indicate scenario (see Table 2) and vertical index on panels indicate true value of AR.

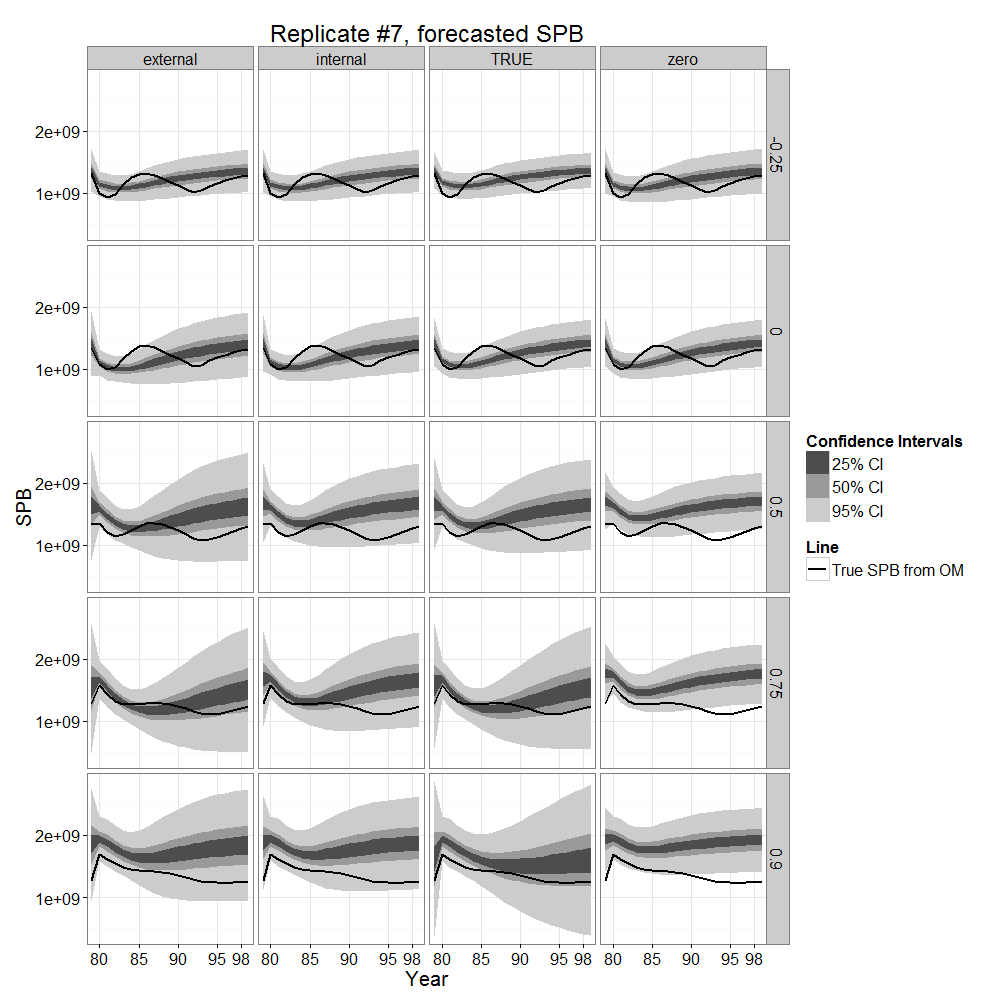


Figure 6: Single-replicate estimate of SPB over time from years 80 to 100. Horizontal index on panels indicate scenario (see Table 2) and vertical index on panels indicate true value of AR.

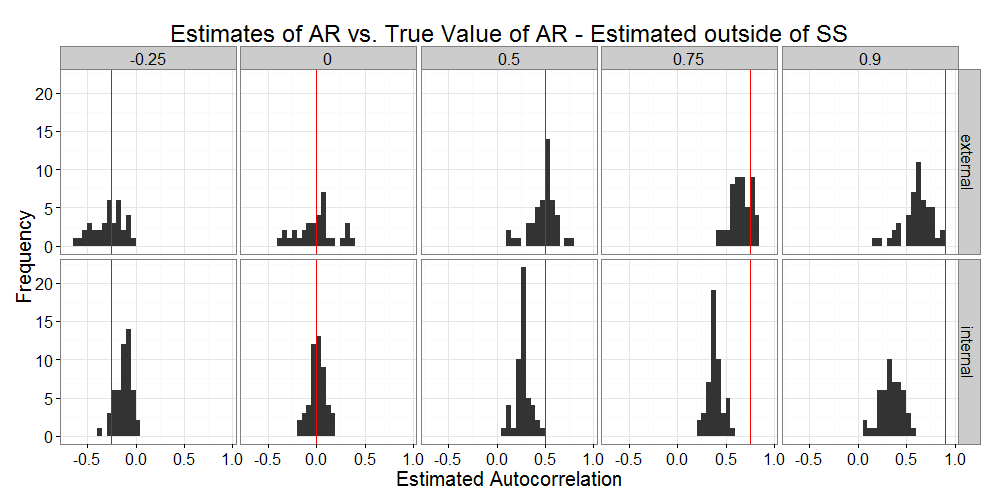


Figure 7: Estimates of AR produced internally by SS (second row) as compared to estimates of AR from external estimation of autocorrelation from estimated recruitment deviations (top row). Red line shows “true” value of AR.

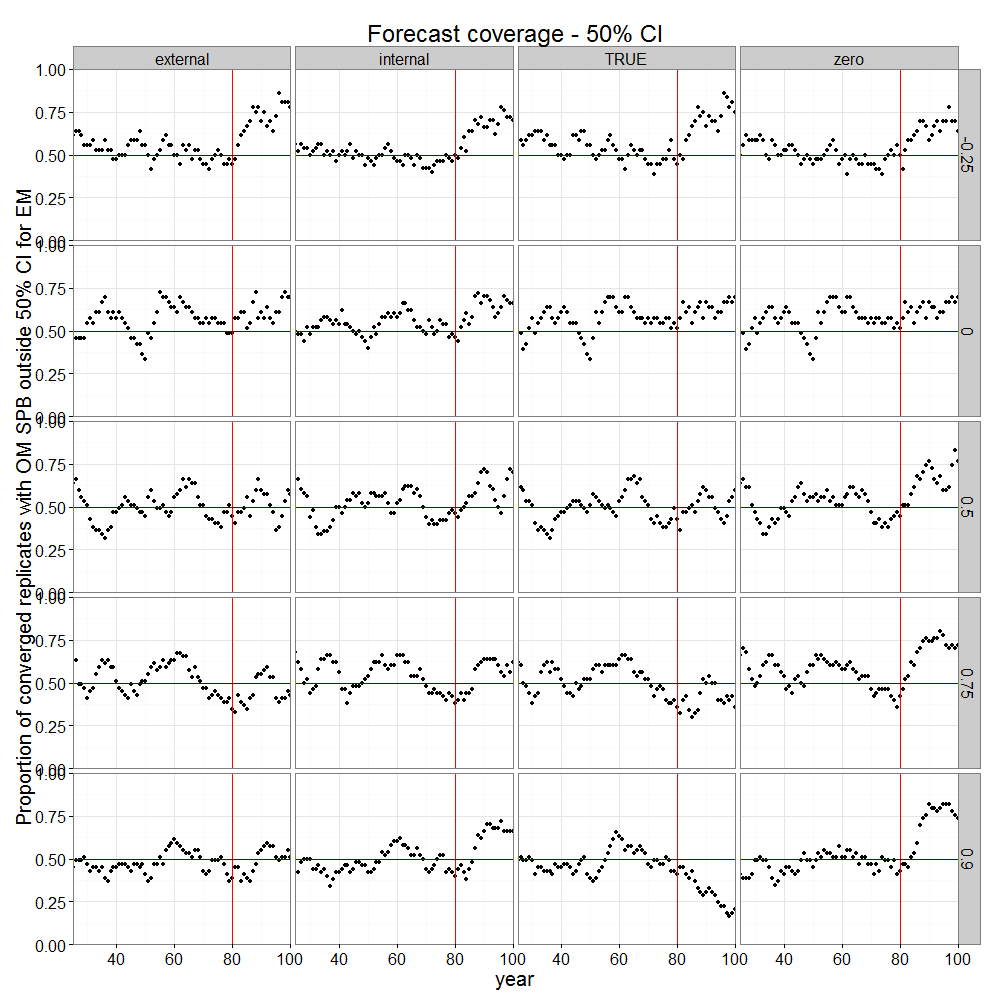


Figure 8: Proportion of converged replicates whose true value of SPB lies outside the 50% confidence interval estimated by the EM.