**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. Methot4, 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

4NOAA 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

**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 , has not generally been analyzed. Monte Carlo simulations were used to evaluate how well Stock Synthesis, 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 autocorrelation 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 autocorrelation produced by Stock Synthesis were biased toward zero when autocorrelation was larger than zero, but were unbiased when the true level of autocorrelation was less than or equal to zero. However, a less biased estimate of autocorrelation was obtained by estimating it from the recruitment deviations estimated within Stock Synthesis. Additionally, the forecast interval for estimates of spawning stock biomass during the forecast period was more uncertain when the true level of autocorrelation was high. Our results suggest that autocorrelation 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 autocorrelation that have small bias, and results in accurate forecast interval coverage (i.e., a 50% forecast interval that includes the true value of spawning stock biomass 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 the sum of one generation time plus the median time for rebuilding given no fishing is used). 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 treats 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). Where parameter estimation is increasingly conducted using an “integrated” population dynamics model that typically incorporates many data types, 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 includes 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 helps 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 correlated measurable environmental factors remain unidentified, regime shifts can instead be treated as autocorrelation in recruitment deviations (i.e., where 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 forecasting intervals compared with assuming recruitment follows 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). 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 autocorrelation in recruitment deviations, 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 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**

The simulation was conducted 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 that govern stock productivity and status, therefore allowing uncertainty about past recruitment to be propagated into rebuilding forecasts. Simulations and analyses were accomplished using the *ss3sim* software package (Anderson et al. 2014a, 2014b), and a public repository houses our simulation code online (github.com/kellijohnson/AR-perf-testing) to ensure our 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., 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 estimation method (i.e, fixing the autocorrelation parameter at its true value, ignoring autocorrelation, or estimating an autocorrelation parameter internally or externally to the assessment). One hundred simulation replicates were generated 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” – Years 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 simulated without any 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 (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 and one survey, 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 output and spawning biomass, 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 (1) 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 biomass (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 *ρ* include -0.25, 0, 0.5, 0.75 and 0.9, where these values are centered approximately around estimates from recent meta-analyses (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 26-80, and were drawn from a Dirichlet distribution with twice the standard deviation of a multinomial distribution given an annual sample size of 100. 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 every other year providing an index of relative abundance for years 41-79, was drawn from a lognormal distribution with log-standard deviation of 0.1.

**2.3 Estimation model**

An age-structured stock assessment model was generated to each simulated data set, using data generated during the “fishing period (years 25-80, see Table 1 for a list of estimated parameters), and we refer to this as the “estimation model”. The estimation model also estimated recruitment deviations for years 1-25 (to simulate initial age-structure given plausible deviations away from the unfished age-distribution), and for years 81-100 (to simulate recruitment variability when forecasting population dynamics forward for 20 years after the simulated assessment in year 80). The correct input sample size for both the multinomial and Dirichlet composition samples (*Ninput* = 100 and 100/22, respectively) were specified in each estimation method (i.e., the estimation model had correct weighting for age- and length-composition sampling data). The bias correction factor was also estimated using five simulation replicates for each level of autocorrelation used to generate the data (Methot and Taylor, 2011).

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

1. “True” – For each simulation replicate, we include an estimation model where autocorrelation was fixed autocorrelation at the level used to generate the recruitment deviations input into the operating model. This estimation model is not plausible for any real-world assessment (given that the true level of autocorrelation will be unknown), but is included as a reference case, to demonstrate model performance if autocorrelation were known exactly.
2. “Zero” – We also include an estimation model where autocorrelation is fixed at zero. This estimation model represents the most common assumption in stock assessment models to date.
3. “Internal” – We next include an estimation model where the magnitude of autocorrelation (ρ) is estimated as a fixed effect. This scenario will likely result in biased estimates of autocorrelation, given that Stock Synthesis 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 magnitude of recruitment deviation (*σr*, Thorson *et al*., 2014). However, its performance when estimating the magnitude of recruitment autocorrelation has not been previously explored.
4. “External” – Finally, we include an estimation model where recruitment autocorrelation is estimated externally to Stock Synthesis. Specifically, this involves extracting estimates of recruitment deviations from the “Internal” estimation model, and then estimating the sampling autocorrelation of these estimates using the *acf* function in R (R Core Development Team 2015). This level of autocorrelation is then input as a fixed value in Stock Synthesis, and Stock Sythesis is run a second time to estimate other parameters for a given data set. This estimation model will likely have different estimation performance than the “internal” scenario, given that sample and population-level estimates are often different in mixed-effects models.

Each estimation method generates forecasts of population abundance during years 81 to 100 following the simulated stock assessment in year 80. All estimation models have no additional data during this forecast period, so recruitment deviations for years 81-100 are estimates at their expected value (i.e., zero when autocorrelation is absent, or decaying towards zero from the value of the estimated recruitment deviation in year 80 when autocorrelation is nonzero). Stock Synthesis uses the delta-method when calculating uncertainty in population abundance during the forecast period. Therefore, forecast period abundance has a standard error that includes uncertainty about recruitment deviations, and this uncertainty is affected by the level of recruitment autocorrelation.

For each estimation model, 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 forecaast period, any bias or imprecision in population abudance during the forecast period arises either from (1) bias and imprecision for estimating parameters during the fishing period, or (2) the impact of mis-specifying autocorrelation on forecasts of recruitment during the forecast period.

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 defifinite Hessian matrix were assumed to have converged. Models that failed to converge were removed from the analysis, and exploratory analysis confirms that results are qualitatively similar when changing the gradient threshold used to identify model convergence. Estimation performance was evaluated using relative error, , where and are estimated and true parameter values, respectively, and the forecast interval coverage, defined as the proportion of 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., where a 50% forecast interval will contain the true value in 50% of simulation replicates.

**3. Results**

**3.1 Estimates of ρ**

Estimates of recruitment autocorrelation (ρ) produced internally by SS while SS simultaneously estimated all other parameters (internal) were more precise but biased compared to estimates of ρ derived from the estimated recruitment deviations produced by SS (external) (Fig. 3). When ρ was positive, internal estimates were negatively biased, as was the external estimates of ρ when the true level of autocorrelation used to generate the recruitment deviations was 0.9, while the opposite was true for internal estimates when ρ was negative (Fig. 3).

**3.1 Estimated population dynamics**

For all scenarios, estimates of spawning stock biomass, recruitment (not shown), and fishing mortality (not shown) were generally unbiased during years the assessment method was provided data (years 26 through 80), though uncertainty was larger at the beginning and end of the time series, as expected (Fig. 5).

**3.2 Forecast coverage**

As the true value of ρ increased, SS forecast coverage decreased, with coverage being the worst for the highest level of true ρ when ρ was fixed at zero in the estimation method (Fig. 6).

As the true value of ρ increased, the forecast coverage decreased, with coverage being the worst when ρ was wrongly fixed at zero, for all estimation methods (Fig. 6). 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.

**3.2**

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 method 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).

**4. Discussion**

* 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 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., 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 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 |  | OM | EM |
| Name | Symbol | True value | Fixed (F) or estimated |
| Natural mortality rate | *M* | 0.2 yr-1 | F |
| Length at age 1 | *La=1* | 20 cm | Est |
| Asymptotic maximum length | *L∞* | 132 cm | Est |
| Von Bertalanffy growth coefficient | *k* | 0.2 yr-1 | Est |
| Coefficient of variation for length at age 1 | *CVa=1* | 0.1 | Est |
| Coefficient of variation for asmptotic maximum length | *CV∞* | 0.1 | Est |
| Length of 50% maturity |  | 38.2 cm | F |
| Length at 95% maturity |  | 48.9 cm | F |
| Average recruits for the unfished population (natural log) | *ln(r0)* | 18.7 | Est |
| Steepness of the Beverton-Holt stock recruit function | *h* | 0.65 | F |
| Marginal log-standard deviation of recruitment | *σR* | 0.4 | F |
| 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 |

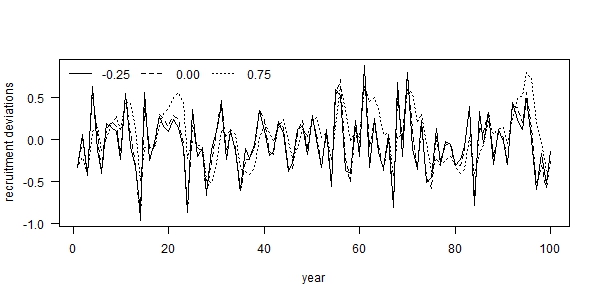


Fig. 1. Time series examples of autocorrelated recruitment deviations for three levels of autocorrelation: (i) -0.25 (solid line), (ii) 0.00 (dashed line), and 0.75 (dotted line), where each example used the same set of process error deviations ().

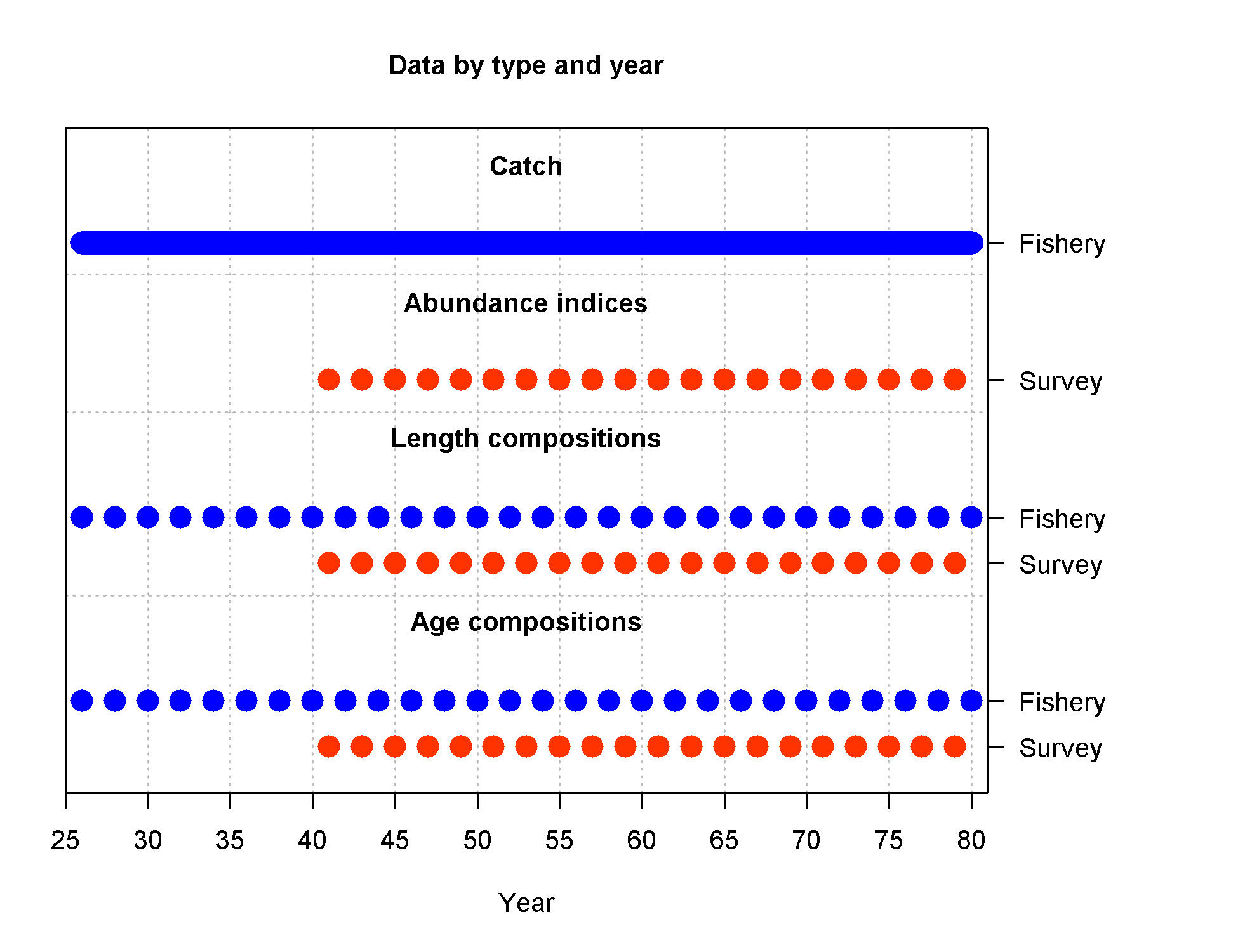
****

Fig. 2. Summary of simulated data available to the estimation method.

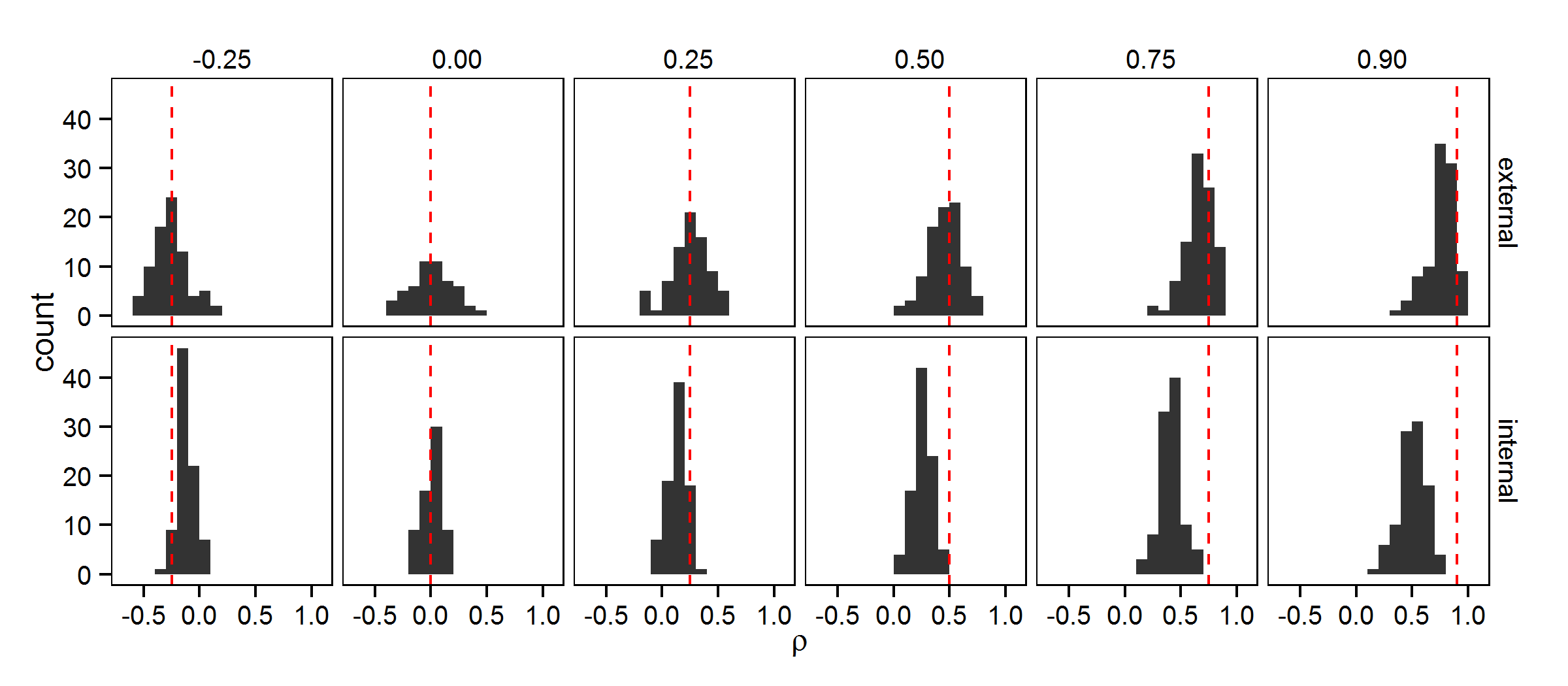


Fig. 3. Estimates of recruitment autocorrelation (*ρ*) from two estimation methods: (i) external to Stock Synthesis (external; top row) and (ii) internally within Stock Synthesis (internal; bottom row), for six levels of autocorrelation in the simulated recruitment deviations (columns). The dashed red line indicates the level of autocorrelation used to generate the simulated recruitment deviations used in the operating model.

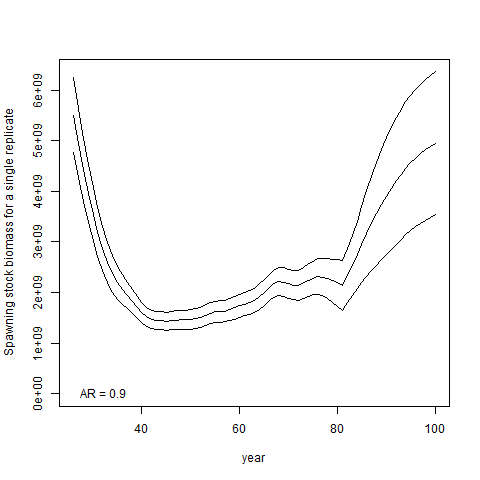


Fig. 3. Illustration of estimated spawning stock biomass (SSB) during 100 simulated years, where the stock assessment in year 80 is indicated with a vertical red line (i.e., the model is forecasting without data for years 81-100). Rows show results for different scenarios (where recruitment autocorrelation 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.

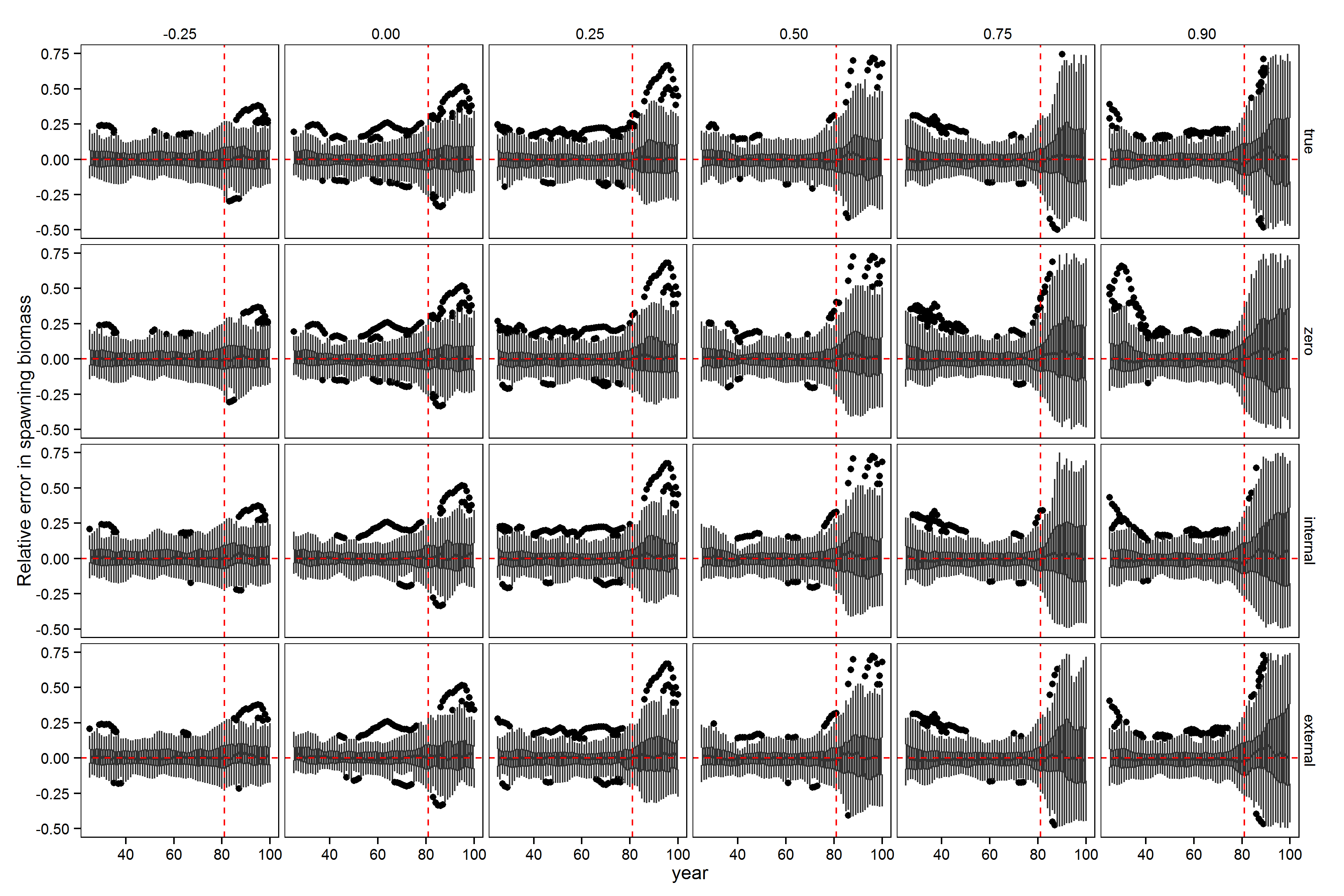


Fig. 5. Relative error in spawning stock biomass during years for which the assessment method was provided data (years 26 through 80) and the forecast period (years 81 through 100) for six levels of autocorrelation in the simulated data (columns) and four estimation methods (rows). Horizontal dashed 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.

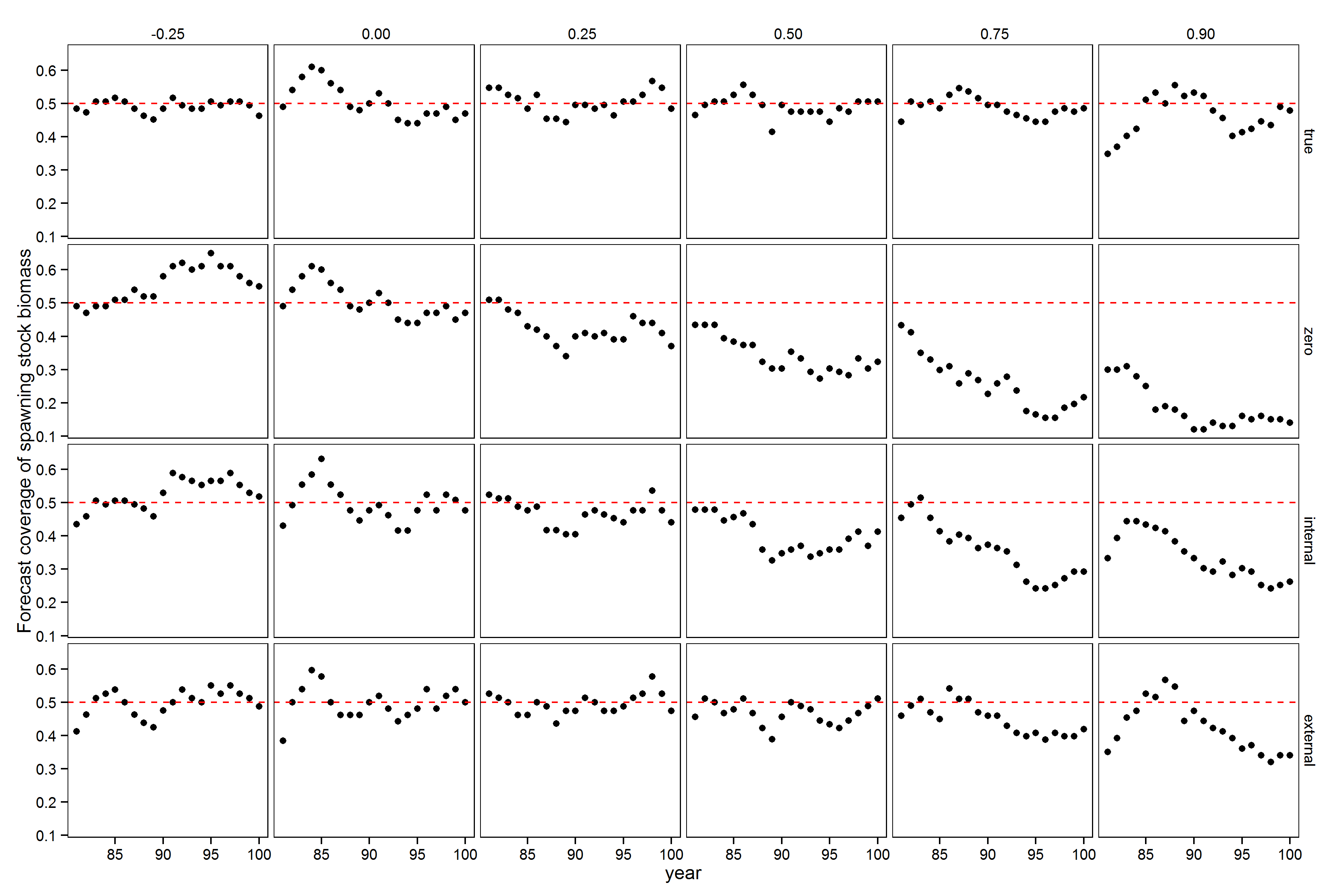


Fig. 6. Proportion of forecast intervals for spawning stock biomass that contain the true value across levels of true autocorrelation used to generate recruitment deviations (columns) and estimation methods (rows). The red dashed line indicates the level at which 50% of the forecast intervals contain the true.