CSIRO
MARINE RESEARCH

SCALIA Simulation-Estimation Study Results Relevant to CCSBT Management Procedure Development

Dale Kolody
Paavo Jumppanen

Prepared for the CCSBT 2nd Management Procedure Workshop
7-9 and 13-14 April 2003, Queenstown, New Zealand, CCSBT-SC/0304/10

Table of Contents

List of Tables

List of Figures

Abstract

CSIRO is currently using simulation-estimation methods to evaluate a range of stock assessment models using complex multi-fishery operating models. The study has a particular emphasis on system dynamics and data that resemble SBT, and the Statistical Catch-at-Age/Length Integrated Analysis (SCALIA) assessment modelling framework closely resembles the operating model (and historical conditioning procedures) developed under the guidance of the CCSBT Scientific Committee for the evaluation of Management Procedures. We present a few preliminary results here that are relevant to the CCSBT MP operating model.

We are generally encouraged by the inferential performance of SCALIA under the conditions described here, but there are some estimation biases evident (even though assumptions about the structural and data characteristics of the operating model are very good). Recent stock biomass levels relative to historical (or unfished) levels appear to be a bit low. Steepness estimates for the Beverton-Holt stock recruitment relationship also seem to be a bit low, but only when the operating model steepness is actually high. However, we included one assessment model with substantial differences (and presumably less credibility) from the others explored, and it demonstrated a somewhat different set of biases that are arguably no worse on average. Therefore, it is currently unclear if these conclusions are robust, or highly dependent on the exact specifications that we happened to explore. Estimates of historical dynamics and steepness were not sensitive to the specification of recruitment deviation CV or auto-correlation.

Introduction

Project SESAME (Simulation-Estimation Stock Assessment Model Evaluation) was initiated to evaluate various stock assessment models for large pelagic fish species. Southern Bluefin Tuna provided most of the impetus for the project, in part because of the diverse assortment of assessment models tabled at the 2001 Commission for the Conservation of Southern Bluefin Tuna (CCSBT) Stock Assessment Group (SAG) meeting. The models demonstrated a range of complexity, from simple two parameter surplus production models (that used only total catch and aggregate CPUE data) to rather complicated models with several hundred parameters that attempted to integrate many diverse data (including age and length frequency distributions dis-aggregated by fishery, and tag releases/recaptures). While there were some consistent features among the assessment model inferences, there were also important differences, with different implications for managers. It was difficult to form a general synthesis of the state of the SBT stock because there are no commonly accepted procedures for assigning degrees of credibility to the results of such different models. We hope that project SESAME will help to address some of these issues. While the CCSBT Scientific Committee has recently shifted emphasis from stock assessment modelling, to the evaluation of Management Procedures (MPs), this type of study remains highly relevant, because the operating model (used for evaluating candidate MPs) is essentially conditioned to the historical data in the same manner that a traditional asssessment model would be.

It is generally recognized as good practice to test stock assessment models by fitting them to simulated data with known characteristics (eg Hilborn and Walters 1992). In reality, the nature of these simulation exercises involves many difficult decisions about what sort of process and observation errors should be included, and how much prior information about the simulated system should be available to the analyst. As one component of a much more ambitious project, we decided to focus SESAME on key issues related to the productivity of the SBT stock. The motivation for this focus is driven by the importance that future recruitment will have for the recovery prospects of the SBT stock. It is important to define the plausible uncertainty in future recruitment to define robust management procedures, and we would like to have some confidence that this can be done. We attempted to address the following:

1) Are these models generally providing sensible inferences about the current state and historical dynamics of the stock?

2) Can these models be expected to estimate a stock recruitment relationship if it is known to exist, and does relatively high auto-correlation in recruitment deviations affect this ability?

3) Is the quality of the assessment model stock recruitment relationship estimation dependent upon correct specification of the input recruitment deviation variance and auto-correlation?

We made a first attempt to address these questions using 5 operating models with different stock and recruitment characteristics (different levels of stock productivity; with and without auto-correlation in recruitment deviations) and used the simulated data to challenge a range of assessment models. The Statistical Catch-at-Age/Length Integrated Analysis (SCALIA) assessment models were very similar except for different assumptions about recruitment dynamics (productivity estimated or held at a constant and sometimes incorrect level; recruitment auto-correlation and variance correct or mis-specified). Only one SCALIA model was substantially different from the rest, in that tagging data was ignored and natural mortality was estimated.

This study represents a small extraction of the simulation results that we have undertaken to date. We expect that additional results will be presented along similar lines at future meetings, and we welcome suggestions for specific simulation scenarios with respect to the SBT assessment modelling (the simulated data sets are also available for other interested parties to test models with).

Methods

In the real world, we never know the actual state of a fish population (unless perhaps it is extinct), so we can never know exactly how well our stock assessment model has described the fishery. Instead, we use an operating model (eg Linhart and Zuchini 1986) to simulate fish and fishery dynamics, and data characteristics that are known exactly. The performance of stock assessment models are then evaluated by fitting the model to the simulated data and comparing estimated values (eg fishing mortality, stock biomass, etc), with the known values. This is the general approach that has been used in many assessment model evaluation studies (eg Anon. 1998, 2002).

Performance Evaluation Criteria

We have gone through a process of identifying many different assessment model performance criteria, and it seems inescapable that several need to be compared simultaneously to make a convincing comment about performance. However, the indicators tend to be highly correlated, and, in the interest of brevity, we present only the following descriptors:

1) B(T)/B(t=1), the ratio of terminal exploitable biomass to initial exploitable biomass.

2) B(T)/B_NF(T), the ratio of terminal exploitable biomass to the terminal exploitable biomass that would have been observed in the absence of any fishery historically. This descriptor would be identical to B(T)/B(t=1) if the stock was initially unfished and there was no recruitment variation. However, it is a more informative criteria because it provides a measure of fishery impact on the stock that includes not only direct fishery removals, but also indirect fishery effects via recruitment over-fishing.

3) F(T-2:T), the mean fishing mortality in the last three years, where F(t) is (total catch mass)/(total exploitable biomass); this index of fishing mortality avoids discussion of appropriate reference ages and incorrect selectivities.

4) Steepness, a measure of the productivity of a stock with a Beverton-Holt stock-recruitment relationship (specifically, the ratio of mean recruitment at 20% of virgin spawning stock to mean recruitment at virgin spawning stock).

5) SR-rho, the auto-correlation in the annual recruitment deviations from the stock recruitment curve (calculated empirically from the SCALIA ML estimates after model fitting).

6) SR RMSE, the approximate CV of the annual recruitment deviations from the stock recruitment curve (calculated empirically from the SCALIA ML estimates after model fitting).

VSM-SBT Operating Model Definitions

Our Virtual Stock Model (VSM) simulator, was defined to qualitatively resemble the SBT fishery, but there was no explicit conditioning to actual data. It included the following features common to all scenarios, (distinctions among scenarios are defined in Table 1):

- spatially-aggregated

- growth, mortality, etc iterated in monthly time-steps

- 4 fisheries: historical longline on spawning grounds, longline on feeding grounds, recent longline on spawning grounds and juvenile fishery; selectivities, catch and effort histories roughly correspond to the Japanese longline fleets on spawning grounds and feeding grounds, Indonesian spawning ground, and Australian surface fisheries for SBT. Since it is spatially-aggregated, the actual location is irrelevant, but does reflect a distinction in perceived selectivity.

- 50 year exploitation history resembling SBT, with largest catches on the spawning grounds in the first 15 years, followed by increasing catch in the feeding grounds and juvenile fishery, drastic cuts to these fisheries after about 40 years, and an increasing spawning ground fishery in the last 10 years

- Age-specific natural mortality resembling common SBT assumptions

- knife-edged maturity at age 12; spawning potential directly proportional to mass-at-age

- mean length-at-age is constant over time and follows a simple von Bertalanffy growth curve estimated for SBT

- variance on length-at-age resembles SBT estimates

- mass-at-length relationship taken from SBT

- selectivity was constant over time for all fisheries

- mean recruitment followed a Beverton-Holt stock recruitment relationship with a CV ~ 0.4 (irrespective of auto-correlation). VSM scenarios varied in SR Steepness (0.3 – 0.9) and recruitment deviation auto-correlation (0 – 0.8) (Table 1)

- initial population was unfished, and in a random state determined by the stock-recruitment relationship and natural mortality

Simulated data included:

- The relationship between effort and fishing mortality was only reliable for the longline feeding grounds fishery (CV = 0.2, no temporal trends in catchability)

- catch-at-length data available for all fisheries (truly random sample of 1000)

- catch-at-age data only available for the recent spawning ground fishery (truly random sample of 1000)

- 6000-12000 perfectly mixed juvenile tags were released in years 41-45

- tag reporting rates were 100%

The actual terminal stock depletion varied within and among VSM scenarios from about 50-90%; catch ratios were intended to be similar across realizations and scenarios, but there was probably considerable variability. VSM models are split into two parts : a *system dynamics model*, which creates a state realization containing the complete time history of the simulation, and an observation model, which creates a data realization (simulated data with observational error). For each of the 5 VSM scenarios defined in Table 1, 10 stochastic state realizations were simulated, and one stochastic data realization per state realization.

SCALIA Definitions

SCALIA is an evolving assessment framework with many different options. Most of the technical details have been described elsewhere (Kolody and Polacheck 2001; Kolody, 2002). The SCALIA versions that we have included here have the following differences from the SBT Operating Model of Haist et al (2002):

- SCALIA is using only 4 fisheries (as is the VSM simulator)
- None of the input variances are estimated: actual values from the VSM model are specified (or intentionally mis-specified). In some cases, these variances are only approximate due to incompatabilities between SCALIA and VSM (eg variance in length-at-age does not translate easily across different time-scales due to within-timestep growth and mortality).
- Longline fishery catch is predicted as a function of observed effort (and transient effective effort deviations)
- The juvenile fishery uses catch-at-length prediction (instead of catch-at-age)
- selectivity is parameterized with a length-based constraint (an alternative to age-based curvature penalties)
- fishing mortality is parameterized as a continuous process, with all fisheries using the same fishing season
- total catch in numbers are predicted with error (but CV = 0.01 means that essentially perfect removals result)

The likelihood-based objective function consisted of the following components:

- catch-at-age
- catch-at-length
- tag recoveries
- total catch in numbers
- effective effort deviations (CPUE), longline feeding grounds only
- recruitment deviations from the SR
- curvature smoothing penalties (selectivity and natural mortality)
- random walk time series constraints on selectivity

Six versions of SCALIA are described in Table 1. Five of these versions differ only in assumptions about the stock recruitment relationship. The SCALIA models were given (unrealistically) good prior information including:

- natural mortality known (in 5 of 6 versions),
- CV of effort deviations (CPUE) known,
- tag reporting rates known (1 version did not use tags),
- length-at-age, mass-at-age and maturity schedule known

Out of concern that these scenarios were far too ideal, SCALIA 115 was added (tagging data ignored, natural mortality estimated).

Results and Discussion

Under the conditions of this study, we are encouraged to see that the SCALIA models are generally providing reasonable inferences. Qualitatively, the predictions and observations are in general agreement, and estimated biomass time trends closely resemble the operating model. However, there also appear to be definite biases in some of the ML estimates. The comments below are mostly qualitative (because we do not trust the small sample sizes for quantifying these biases) and are roughly organized according to the original 3 questions posed in the introduction.

Are these models generally providing sensible inferences about the current state and historical dynamics of the stock?

SCALIA tends to estimate the stock to be slightly more depleted than it really is (Fig. 1). The assessment model with the smallest bias in depletion generally seems to be SCALIA 115 (which uses the least prior information: M is estimated and tags ignored), however, this model does the poorest job of estimating current fishing mortality (which is not surprising). All of the other SCALIA models tend to estimate similar levels of depletion from initial conditions, B(T)/B(1), despite having different specifications for the input recruitment auto-correlation, CV, and/or steepness. Estimates of B(T)/B_NF(T) are generally similar to B(T)/B(1), except in the cases where steepness is constant (SCALIA 3003, and 3009). When steepness is held at the wrong value, the assessment does a poor job of estimating the biomass that would have occurred without fishing, because the indirect effects of the stock recruitment relationship are badly estimated. This effect is discussed further below.

Can these models be expected to estimate a stock recruitment relationship if it exists, and does relatively high auto-correlation in recruitment deviations affect this ability?

SCALIA was generally able to distinguish different levels of SR steepness, irrespective of the recruitment deviation auto-correlation (Fig. 2). ML steepness estimates seemed to be very good for unproductive scenarios (VSM 1,4; steepness of 0.3 and 0.4), but somewhat under-estimated for productive scenarios (VSM 2,3,15; steepness 0.6,0.9,0.8). The empirical recruitment deviation auto-correlation estimated from SCALIA tended to be slightly higher than the actual value in most scenarios. The empirical CV (RMSE) of the recruitment deviations was clearly low in the operating model scenarios with high auto-correlation (VSM 4, 15; rho = 0.8), while the other scenarios were reasonable. Of the models that estimated steepness, SCALIA

115 was the most dissimilar from the rest (and probably somewhat better in terms of steepness bias and auto-correlation).

SCALIA 3003 and 3009 (fixed steepness of 0.3 and 0.9 respectively) demonstrate that the empirical auto-correlation and CV of the recruitment deviations are inflated relative to the operating model (and other SCALIA models), when steepness is held constant at the incorrect value (Fig. 2). This effect was already recognized without the simulations, but the illustration may be helpful. In the cases presented here, the effect of the steepness mis-specification would not even be noticeable, if one focused only on the quality of the model fit to the simulated data and the biomass history, which can be very similar across models irrespective of the imposed steepness. This would cease to be the case if the input SCALIA recruitment CV was substantially reduced, or, presumably, if the variances in the other likelihood terms were increased.

Is the quality of the SCALIA stock recruitment relationship estimation dependent upon correct specification of the stock recruitment variance and auto-correlation?

SCALIA models 101, 106 and 111 differed only in the input assumptions about recruitment deviation auto-correlation (0, 0.8, and 0 respectively) and CVs (0.4, 0.4, and 0.8 respectively), however, the general estimation performance of these models was very similar, as illustrated by the resultant steepness estimates, empirical CV and auto-correlation (Fig. 2). This is encouraging, but presumably the results would have been more sensitive if the exploitation history was vastly different, or the other data more uncertain.

Future SESAME Plans

The estimation biases and input specification insensitivities that we have described here are interesting, but it is currently unclear if these are general results or related to the specific details of these simulations. The biases that we did observe remain unexplained, and may be due to several sources:

- SCALIA or VSM coding errors
- Structural inconsistencies between VSM and SCALIA. e.g. Different temporal resolution causes aggregation errors in length-at-age distributions and the calculation of summary statistics.
- Statistical biases. SCALIA uses the errors-in-variables estimation paradigm (eg Schnute 1994), that confuses the distinction between unknown states and estimable parameters. Also, many of the performance criteria are ratios and subject to misleading biases.
- 10 data realizations per VSM scenario is not very many.

We are interested in prioritizing future SESAME simulations to make the most useful inferences for both SBT stock assessment and operating model conditioning. In particular, we plan to systematically decrease the quality of the simulated data, add more complicated dynamics (catchability and selectivity temporal variability), and evaluate procedures for uncertainty quantification. This has already been undertaken to a greater extent than described here, but we welcome suggestions for the project, and encourage other analysts to test their models with the simulated data sets.

Conclusions and Implications for CCSBT Management Procedure Evaluations

We hesitate to make broad generalizations from the results presented above, because we do not know how robust they are. For example, SCALIA 115 (no tags, M estimated) would be expected to perform more poorly than the other models, and we would be tempted to downweight the importance of this scenario in drawing general conclusions. However, it seems that this model might have performed better than the others in many respects, and slightly different SCALIA models demonstrate similar variations (not shown). Thus we caveat the following list of conclusions as a synthesis of the very specific results above, biased by our a priori assumptions about which results are likely to be most meaningful, and with full expectation that some of the generalizations may be retracted at a later date (particularly 3-5):

1. SCALIA models were reasonably successful at estimating the stock-recruitment relationship parameters and historical stock dynamics; this provides encouragement for the general approach that has been taken by the CCSBT-SC with respect to formulating and conditioning the operating model for Management Procedures evaluations.

2. The imposition of an incorrect stock-recruitment relationship steepness inflated the empirical estimates of the recruitment deviation CV and auto-correlation. This had a relatively minor effect on the fit to the simulated data and estimated historical biomass dynamics, but the overall effect of fishing on the stock was poorly estimated (because indirect effects of fishing via the stock recruitment relationship were wrong). This was anticipated even without the simulations, but the results provide a convenient illustration of how the auto-correlation is exaggerated by incorrect SR relationships. In short-term projections, the dangerous implications of a bad steepness assumption would be mitigated to some extent by (an also incorrect) strong auto-correlation, and we would suggest that the empirical auto-correlation should be used in the projections (particularly if steepness values that are not supported by the data are imposed). There is also a related question that we have not addressed: what are the implications to the other relevant parameter estimates (eg effort deviations) when steepness is forced?

3. The terminal level of relative depletion seems to be somewhat over-estimated in SCALIA results. This is worth further exploration, but it a rather small bias, in a pre-cautionary direction, and possibly dependent on SCALIA assumptions.

4. SCALIA Stock recruitment steepness estimates were generally biased somewhat low for productive stocks. It would be worth trying to identify the cause of this. It does provide justification for some sort of steepness bias correction and exploration of alternative scenarios that are not entirely consistent with the data.

5. Stock recruitment parameter estimation was not very sensitive to the mis-specification of recruitment deviation CV or auto-correlation. This conclusion may seem counter-intuitive given the results of SBT operating model conditioning explorations (eg Polacheck and Kolody 2003). There are several things in the real SBT situation that may cause sensitivities to the stock recruitment relationship estimation, including incorrect assumptions about longline catchability trends, length-at-age prior to 1970, or recruitment regime shifts. Different data and structural assumptions support contradictory inferences, and we would suggest that these factors should be given a higher priority for exploration than the CV and auto-correlation, but unfortunately, these things are probably all inter-related.

References

Anonymous. 1998. Improving fish stock assessments / Committee on Fish Stock Assessment Methods, Ocean Studies Board, Commission on Geosciences, Environment, and Resources, National Research Council. National Academy Press. Washington, D.C.

Anonymous. 2002. Report of the Methods Working Group. 15th Meeting of the Standing Committee on Tuna and Billfish. 18-19 July 2002, Honolulu, USA

Haist, V., A. Parma, and J. Ianelli. 2002. Initial specifications of operating models for southern bluefin tuna management procedure evaluation. CCSBT-SC/0209/7

Hilborn, R. and C.J. Walters. 1992. Quantitative fisheries stock assessment: choice, dynamics, and uncertainty. Chapman and Hall, New York.

Kolody, D. 2002. SCALIA: application of an integrated analysis stock assessment model to the SCTB methods working group simulated tuna fishery data. Standing Committee on Tuna and Billfish 15 Working Paper MWG-5

Kolody, D. and T. Polacheck. 2001. Application of a statistical catch-at-age and – length integrated analysis model for the assessment of southern bluefin tuna stock dynamics 1951-2000. CCSBT-SC/0108/13.

Linhart, H. and W. Zuchini. 1986. Model selection. Wiley, NewYork.

Polacheck, T. and D. Kolody. 2003. The Behaviour and Fit of Alternative Operating Model Specification for Testing the Performance of Southern Bluefin Tuna Candidate Management Procedures. T. Polacheck and D. Kolody

Schnute, J.T. 1994. A general framework for developing sequential fisheries models. Can. J. Fish. Aquat. Sci. 51:1676-1688.

Table 1 - Summary of key operating (VSM) and assessment (SCALIA) model features.

	Beverton-Holt SR Steepness	**SR auto-correlation**	**SR CV**	**juvenile and longline selectivity**	**tag recapture data**	**Longline catchability**
VSM 1	0.3	0	0.4	constant	5 releases	constant
VSM 2	0.6	0	0.4	constant	5 releases	constant
VSM 3	0.9	0	0.4	constant	5 releases	constant
VSM 4	0.4	0.8	0.4	constant	5 releases	constant
VSM 15	0.8	0.8	0.4	constant	5 releases	constant
SCALIA 101	estimated	0	0.4	estimated (5 y periods)	yes	estimated (constant)
SCALIA 106	estimated	0.8	0.4	estimated (5 y periods)	yes	estimated (constant)
SCALIA 111	estimated	0	0.8	estimated (5 y periods)	yes	estimated (constant)
SCALIA 115**	estimated	0	0.4	estimated (5 y periods)	not used	estimated (constant)
SCALIA 3003	0.3	0.8	0.4	estimated (5 y periods)	yes	estimated (constant)
SCALIA 3009	0.9	0.8	0.4	estimated (5 y periods)	yes	estimated (constant)

* For SCALIA 115, age-specific natural mortality was also estimated

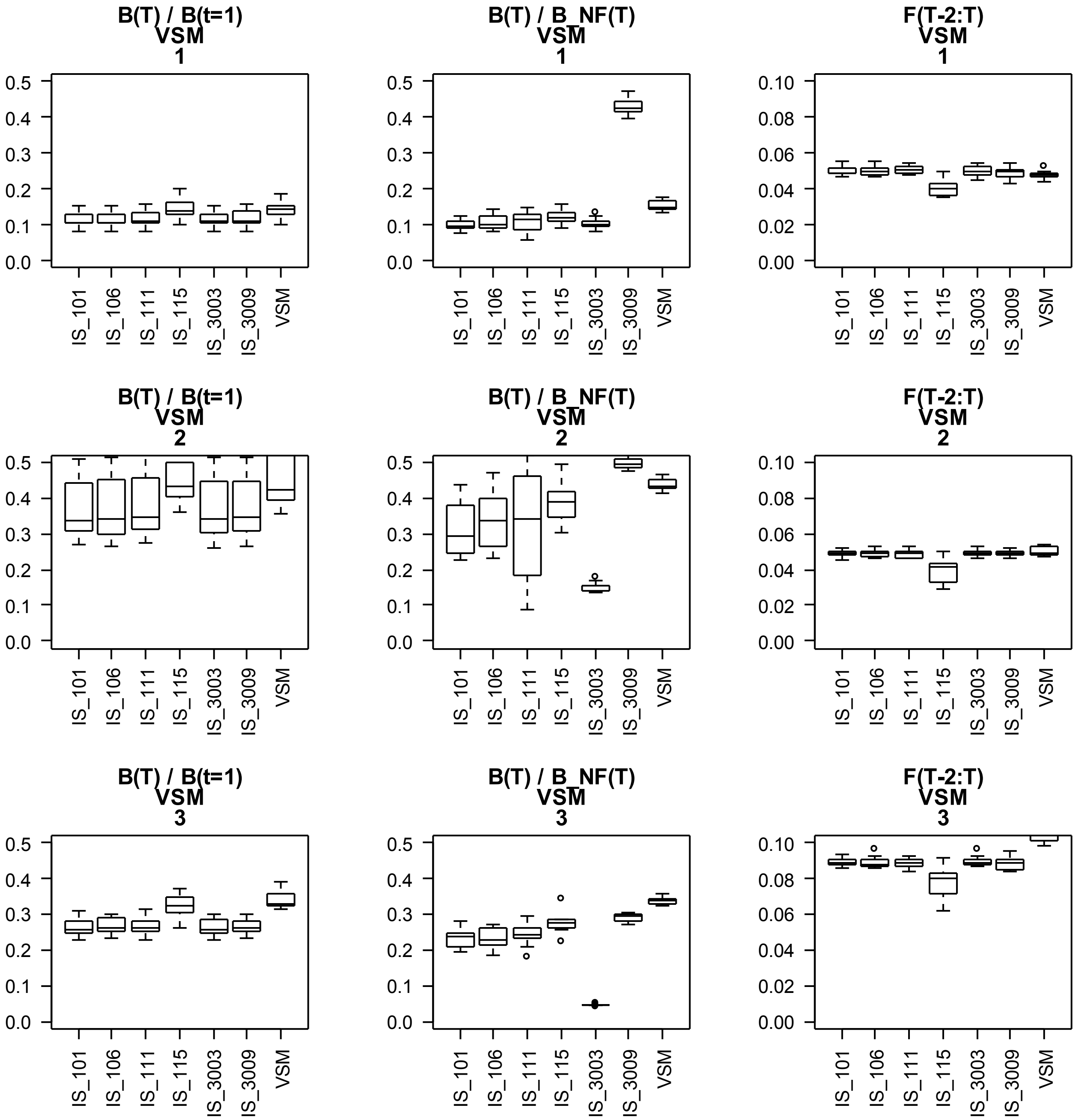

Figure 1. Box-and-whisker plots illustrating the distribution of ML biomass depletion and fishing mortality estimates (see text for explanations) from SCALIA models fit to each of 10 data realizations for each VSM operating model scenario). See Table 1 for VSM and SCALIA definitions.

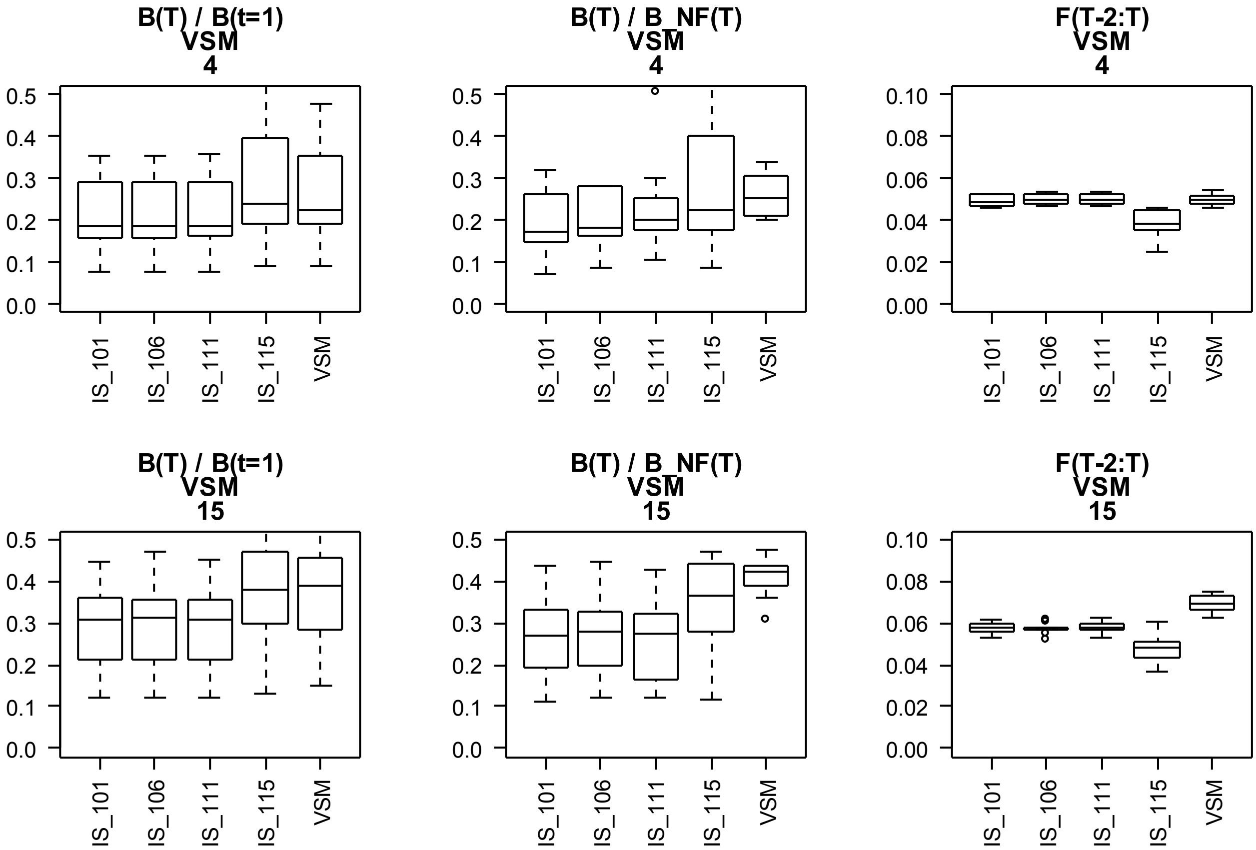
B(T) / B(t=1)
VSM
4
B(T) / B_NF(T)
VSM
4
F(T-2:T)
VSM
4
B(T) / B(t=1)
VSM
15
B(T) / B_NF(T)
VSM
15
F(T-2:T)
VSM
15
IS_101
IS_106
IS_111
IS_115
VSM

Fig. 1. (cont.)

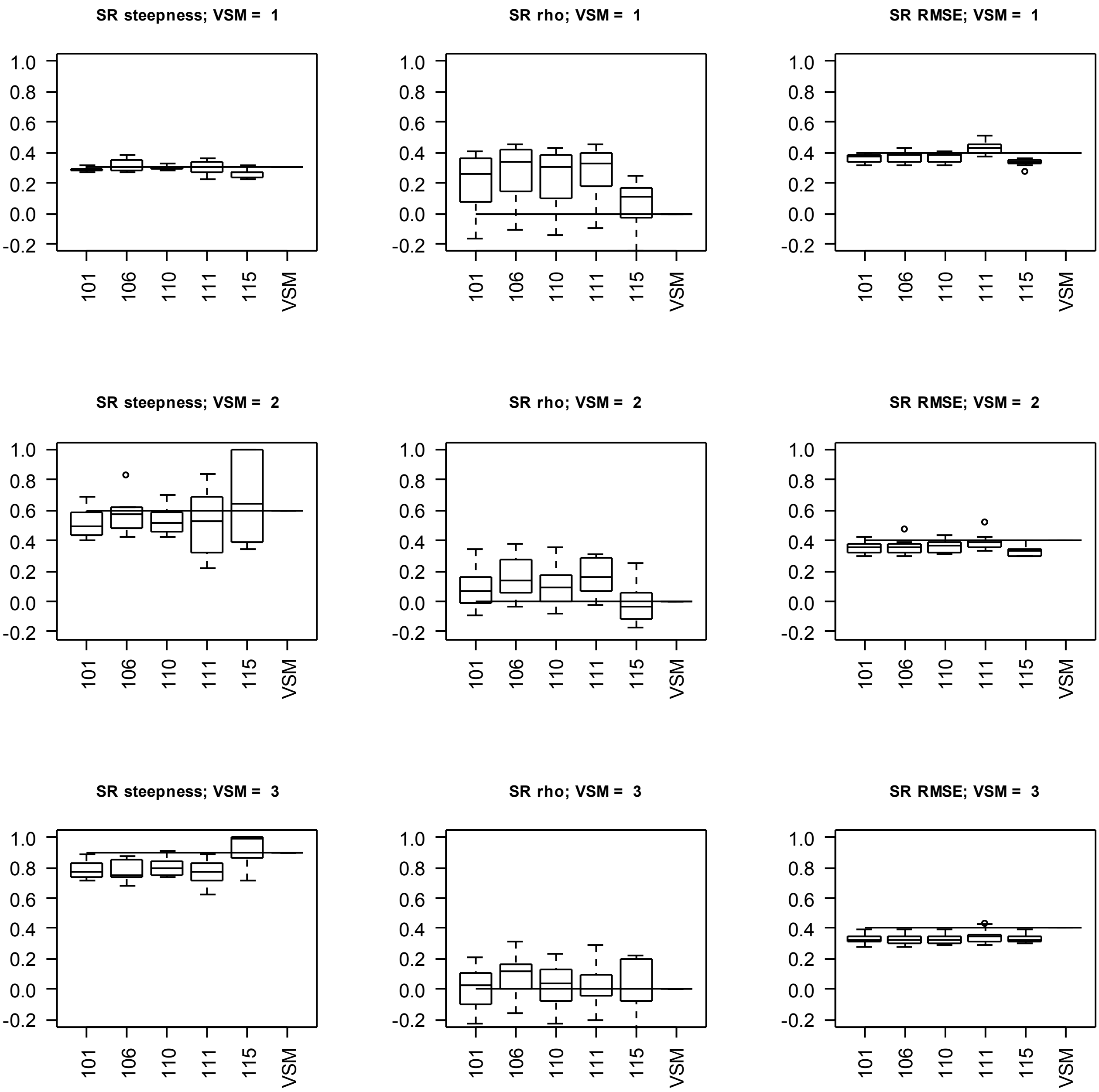

Figure 2. Box-and-whisker plots illustrating the distribution of stock recruitment relationship characteristics estimated by SCALIA models (10 data realizations for each VSM operating model scenario). The horizontal reference line indicates the actual value in each VSM scenario. The recruitment deviation CV (SR RMSE) and auto-correlation (SR rho) are calculated from the MLE for the recruitment deviations (i.e. not the assumed value used to fit the SCALIA model). See Table 1 for VSM and SCALIA definitions.

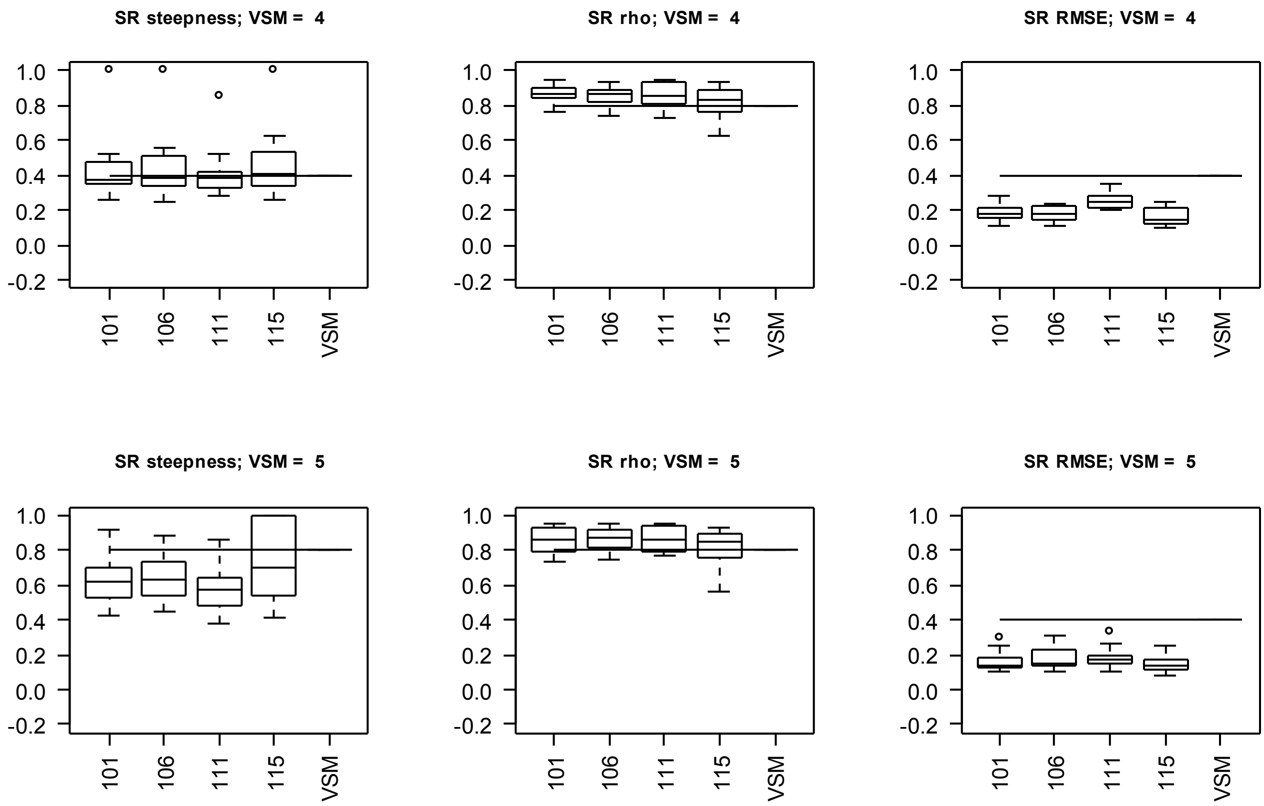
SR steepness; VSM = 4
SR rho; VSM = 4
SR RMSE; VSM = 4
SR steepness; VSM = 5
SR rho; VSM = 5
SR RMSE; VSM = 5
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
101
106
111
115
VSM

Fig. 2 (cont.)