SCRS/2016/014 Collect. Vol. Sci. Pap. ICCAT, 73(9): 3000-3008 (2017)

CONDITIONING OPERATING MODELS ON DATA AND KNOWLEDGE AND REJECTING AND WEIGHTING OF HYPOTHESES

Laurence T. Kell[1] and Iago Mosqueira[2]

SUMMARY

Important steps when conducting Management Strategy Evaluation are the selection of that represent the simulated versions of reality, conditioning the Operating Model based on data and knowledge, and then to weight and reject those hypotheses depending on their plausibility. There are many alternative ways to do this, one way is to use the currently-used stock assessment model as the Operating Model. Although use of the assessment model as the operating model seems to imply that assessment models describe nature almost perfectly, if a Management Procedure cannot perform well when reality is as simple as implied by an assessment model, it is unlikely to perform adequately for more realistic representations of uncertainty. Basing an operating model on the current assessment model also has arguably the lowest demands for knowledge and data. In a stock assessment, however, due to limitations in time often only a limited number of hypotheses are considered for developing assessment scenarios. Given the need to evaluate robustness and the longer time scale required for conducting an MSE, a broader range of hypotheses for conditioning an Operating Model is both desirable and possible. As an example we present diagnostics from an Operating Model developed for Indian Ocean albacore conditioned using Stock Synthesis.

RÉSUMÉ

Des étapes importantes lors de la réalisation de l'évaluation de la stratégie de gestion sont la sélection des hypothèses aux fins de l'examen dans le modèle opérationnel qui représentent les versions simulées de la réalité, le conditionnement du modèle opérationnel fondé sur les données et les connaissances, et ensuite la pondération et le rejet de ces hypothèses en fonction de leur plausibilité. Il y a beaucoup d'autres façons de le faire, par exemple en se servant du modèle d'évaluation des stocks actuellement utilisé comme modèle opérationnel. Bien que l'utilisation du modèle d'évaluation comme modèle opérationnel semble impliquer que les modèles d'évaluation décrivent la nature presque à la perfection, si une procédure de gestion ne fonctionne pas correctement quand la réalité est aussi simple que celle issue d'un modèle d'évaluation, il est peu probable qu'elle fonctionne correctement avec des représentations de l'incertitude plus réalistes. Faire reposer un modèle opérationnel sur le modèle d'évaluation actuel a aussi sans doute les exigences les plus basses en termes de connaissances et de données. Ceci dit, dans une évaluation des stocks, en raison de limitations de temps, souvent seul un nombre limité d'hypothèses sont envisagées pour élaborer des scénarios d'évaluation. Compte tenu de la nécessité d'évaluer la solidité et l'échelle temporelle plus longue requise pour réaliser une MSE, un plus large éventail d'hypothèses aux fins du conditionnement d'un modèle opérationnel est à la fois souhaitable et possible. À titre d'exemple, nous présentons les diagnostics obtenus à partir d'un modèle opérationnel élaboré pour le germon de l'océan Indien conditionné au moyen de Stock Synthesis.

RESUMEN

Pasos importantes al realizar la evaluación de estrategias de ordenación son la selección de la hipótesis a considerar en el modelo operativo que representa las versiones simuladas de la realidad, el condicionamiento del modelo operativo basado en los datos y el conocimiento y después la ponderación o descarte de hipótesis en función de su plausibilidad. Existen varias

[1] ICCAT Secretariat, C/Corazón de María, 8. 28002 Madrid, Spain.
[2] European Commission, Joint Research Center (EC JRC), Institute for the Protection and Security of the Citizen (IPSC), Maritime Affairs Unit G03, Via E. Fermi 2749, 21027 Ispra VA, Italy.

alternativas para los modos de hacerlo, una forma es utilizar el modelo de evaluación utilizado actualmente como modelo de operativo. Aunque el uso del modelo de evaluación como modelo operativo parece implicar que los modelos de evaluación describen la naturaleza casi a la perfección, si un procedimiento de ordenación no funciona bien cuando la realidad es tan simple como se supone en un modelo de evaluación, es poco probable que funcione adecuadamente con representaciones más realistas de la incertidumbre. También puede decirse que basar un modelo operativo en el modelo de evaluación actual implica el nivel más bajo de requisitos en cuanto a conocimientos y datos. Sin embargo, en una evaluación de stock, debido a las limitaciones de tiempo, se considera a menudo solo un número limitado de hipótesis para el desarrollo de escenarios de evaluación. Dada la necesidad de evaluar la robustez y la escala de tiempo más larga requerida para la realización de una MSE, sería deseable y es posible considerar una gama más amplia de hipótesis para el condicionamiento de un modelo operativo. Como ejemplo presentamos el diagnóstico de un modelo operativo desarrollado para el atún blanco del océano Índico condicionado utilizando stock shynthesis.

KEYWORDS

Life History; FLR; Density Dependence; Stochasticity;
Reference Points; Growth; Fecundity; Maturity; Natural Mortality

Introduction

When conducting Management Strategy Evaluation (MSE) there are six steps namely i) Identification of management objectives; ii) Selection of hypotheses for the Operating Model (OM); iii) Conditioning of the OM based on data and knowledge, and weighting and rejection of model hypotheses; iv) Identifying candidate management strategies; v) Running the Management Procedure (MP) as a feedback control in order to simulate the long-term impact of management; and vi) Identifying the MP that robustly meet management objectives. In this paper we describe a case study to develop an OM (steps ii and iii) for Indian Ocean albacore using Stock Synthesis.

There are many ways to condition OMs (Kell *et al.*, 2006), in the case of the Indian Ocean albacore MSE the currently-used stock assessment model was used. Although the use of the assessment model as the OM seems to imply that assessment models describe nature almost perfectly, if a MP cannot perform well when reality is as simple as implied by an assessment model, it is unlikely to perform adequately for more realistic representations of uncertainty. Basing an OM on the current assessment model also has arguably the lowest demands for knowledge and data.

In a stock assessment, due to limitations in time often only a limited number of hypotheses are considered for developing assessment scenarios. Given the need to evaluate robustness and the longer time scale involved when conducting an MSE, a broader range of hypotheses for conditioning an OM is both desirable and possible.

To evaluate the impact of uncertainty in stock assessment often a base case scenario is agreed and then the different assumptions modelled as factors with levels. Scenarios can then be run for all interactions, as a factorial design, or for only the main effects. Ways of rejecting or weighting scenarios should be agreed in advance, e.g. are model assumptions violated, model mis-specified or are there systematic failure in fits?

When fitting an assessment model to data there is often insufficient information and contrast in the data to estimate parameters for important population processes (e.g. Lee *et al.*, 2012, 2011; Pepin and Marshall, 2015). The data may also appear equally likely given alternative parameter values, and different data sets (CPUE, catch and length distributions) may have conflicting signals and often scenarios are developed by down weighting datasets. As a result of these problem some parameters may need to be fixed, the functional form of processes assumed and assumptions made about density dependence. A variety of scenarios should be run therefore to reflect scepticism about the capacity of the model to estimate key parameters.

Material and Methods

The Indian Ocean Tuna Commission (IOTC) uses Stock Synthesis (SS Methot and Wetzel, 2013) to assess Indian Ocean albacore and has recently started to conduct MSE for this stock to evaluate simpler management procedures. As pointed out by Kolody *et al.* (2009) the stock assessment process often appears to involve a haphazard search for a few model specifications which appear to be plausibly, consistent with the data, and a priori expectations. Conditioning an OM in contrast is an attempt to evaluate uncertainty and so rather than attempting to find a 'best' model it is an attempt to characterise what we don't know about resource dynamics. This requires many scenarios to be run, and agreeing procedures to weight and possibly reject scenarios.

In the assessment only a limited number of assessment scenarios were considered, while in the MSE seven factors with various levels were used to condition the OM (**Table 1**) i.e. 720 scenarios as all interactions were run. Factors considered were for different values of natural mortality (M), steepness, variability of recruitment, selection pattern, and assumptions about effort creep, variance of CPUE and weighting of data.

Likelihoods

Integrated models are so called as they integrate multiple diverse datasets to try to extract as much information as possible about modelled processes (Fournier *et al.*, 1998). An implicit assumption is that integrated models can compensate for lack of good data. Indices may be conflicting between themselves and with the catch and length frequency data. Fitting therefore involves weighting averages of contradictory trends (Schnute and Hilborn, 1993). Models are, by definition, simplifications of reality and model misspecification can cause degradation of results when including additional potentially conflicting data sets (Maunder and Piner, 2016). Payne *et al.* (2009) showed that including all available data in stock assessments may cause high noise levels and lead to poor-quality assessments, and recommended that the choice of data should be based on rational and justifiable selection criteria. It is therefore critical to determine what drives the assessment (Francis and Hilborn, 2011).

Likelihood can be used to compare stock assessment scenarios or to weight multiple runs (Hobbs and Hilborn, 2006). **Figure 1** summarises the log likelihoods from the 720 scenarios. The data, however, are not the same for all scenarios as changing the sample size between models invalidates a direct comparison of the likelihoods and the use of criteria such as AIC or BIC. Furthermore to weight the OMs by their likelihoods, it is assumed the data were generated according to well-defined probabilistic processes and that the model is correct, which is unlikely to be true (Kolody *et al.*, 2009).

Likelihood profiles can be used to check the information content of the data, however. Profiling allows the impact of different data sets on key parameters to be evaluated; likelihood profiles are shown for R_0 for three scenarios in **Figure 2**. Three features are seen, i) that some datasets, i.e catch, suggest that R_0 is large and not within the range explored, ii) the different datasets conflict and imply different estimates of R_0 and the profiles by data component are not smooth.

Figure 3 summarise the condition number of the Hessian matrix H. H is a square matrix of second-order partial derivatives of a scalar-valued function, or scalar field and describes the local curvature of a function of many variables. In the case of non-linear minimisation this is the likelihood with respect to the estimated parameters. The convergence rate for non-linear minimisation is linked to the eigenvalues of the H, in particular, to the ratio of the smallest eigenvalue to the largest one on the log scale, i.e. the condition number. A large condition number indicate that updates when fitting (i.e. convergence) is likely to be poor. The structure of the minimum is essentially determined by H and plays a major role in the optimization problem and its solution processes. Ideally the problem should be reformulated (i.e. rescaled) so that the condition number is small. Another problem potentially identified by a large condition number is collinearity due to the non-independence of predictor variables.

Post-hoc model selection

The OMs are all complicated highly parameterised models with many assumptions that are poorly justified. Simulation studies have shown that some important parameters cannot be estimated from the available data (e.g. Lee *et al.*, 2011, 2012; Pepin and Marshall, 2015). Given the problems in the likelihoods the feasibility of the scenarios was evaluated based on knowledge of the available albacore habitat which provides a limit on the carrying capacity (K). Estimates of K for albacore stocks across all oceans were obtained from the relevant tRFMO-approved stock assessments and estimates of suitable habitat by ocean for albacore from Arrizabalaga (2014) **Table 2**.

A linear model of the form K ~ 0 + h, where h is the potential habitat size, and 0 indicates a zero intercept, was fitted using the lm function in the R statistical language and estimates of the coefficient and standard error (**Table 3** were used to generate an upper plausible limit for K, and by assuming equilibrium conditions prior to the start of industrial fishing for B_0. This was computed as the upper 99.9% confidence interval around the estimate of the slope coefficient, by using the calculated ratio for the Indian Ocean. The upper limit obtained, B_0 = 878,127, could then be used to select scenarios deemed plausible. This reduced the total number of scenarios from 720 to 258. **Figure 8** summarises R_0 ($B_0 = R_0$ times a constant given by the spawner per recruit in the absence of fish) by scenario. The scenarios that resulted in the largest values of R 0 were when a dome shaped selection pattern and high M at older age-classes were assumed.

Confounding

Collinearity, data conflicts and lack of information in the data may mean that the estimated values, e.g. stock status relative to reference points (e.g. K, SSB_{MSY}, SSB_{2012}, F_{2012}, F_{MSY}, MSY, cov(SSB_{2012}, F_{2012}), var(SSB_{2012}), var(F_{2012}), SSB:SSB_{MSY}, F:F_{MSY}, H_{MSY} and H:H_{MSY}) may be correlated. Therefore the relationship between them were explored using principle components analysis (PCA) in **Figure 4.** The first principal axis maximizes the variance, as reflected by its eigenvalue. The second component is orthogonal to the first and maximizes the remaining variance. The first two component account for over 80% of the variance and therefore yield a good approximation of the relationship between the original variables. They therefore correspond to the interesting dynamics and lower ones to noise. The main features of the data as given by the first component is a contrast between F_{MSY} and current F (i.e. F_{2012}); B_{MSY} and F_{MSY} are close to each other and so are correlated. The second component summarises absolute abundance (e.g. SSB 2012).

Time series

When evaluating feedback control systems the nature of the dynamics, i.e. the properties and relationships between time series are very important. Therefore cross-correlation was used to separate the influence of recruitment on SSB from the influence of SSB on recruitment (Szuwalski *et al.*, 2014), evaluated in **Figure 5**. If recruitment estimates are lagged to the year of fertilization, the correlation at zero lag represents the influence of SRP on recruitment. Negative lags represent the influence of recruitment 1,2,3, ... years in the past on the current years SRP. If the influence of recruitment on SRP is much larger than the influence of SRP on recruitment, it is possible that recruitment is environmentally driven, even if there is an apparent stock-recruit relationship (Gilbert, 1997). Therefore, only if SSB has a larger and significant influence on recruitment than recruitment does on SSB, then the existence of a stock recruitment relationship is unequivocal. The cross-correlations are calculated using Spearmans correlation (Spearman, 1904) to identify monotonic relationships between recruitment and SSB. The fact that correlations are mainly positive is problematic.

Next the recruitment time series was evaluated using the STARS algorithm (Rodionov, 2004) to identify potential regime shifts (**Figure 6**).

Conclusions

Possible procedures for conditioning OMs using integrated stock assessment models were discussed and include agreeing a base case and scenarios based on a grid of factors with levels for processes and fixed values. It will be impossible to run all possible diagnostic procedures on a full grid. Remaining issues are how to propagate of uncertainty into projections, generate historic data for the Observation Error Model (OEM), check the nature of time series, and weighting of OMs.

References

Arrizabalaga, H., Dufour, F., Kell, L., Merino, G., Ibaibarriaga, L., Chust, G., Irigoien, X., Santiago, J., Murua, H., Fraile, I. and Chifflet, M., 2015. Global habitat preferences of commercially valuable tuna. *Deep Sea Research Part II: Topical Studies in Oceanography*, *113*, pp.102-112.

D. A. Fournier, J. Hampton, and J. R. Sibert. MULTIFAN-CL: a length-based, age-structured model for fisheries stock assessment, with application to south pacific albacore, *Thunnus alalunga*. Can. J. Fish. Aquat. Sci., 55(9):2105–2116, 1998.

R. C. Francis and R. Hilborn. Data weighting in statistical fisheries stock assessment models. Can. J. Fish. Aquat. Sci., 68(6):1124–1138, 2011.

D. Gilbert. Towards a new recruitment paradigm for fish stocks. Can. J. Fish. Aquat. Sci., 54(4): 969–977, 1997.

N. T. Hobbs and R. Hilborn. Alternatives to statistical hypothesis testing in ecology: a guide to self-teaching. Ecological Applications, 16(1):5–19, 2006.

L. Kell, J. A. De Oliveira, A. E. Punt, M. K. McAllister, and S. Kuikka. Operational management procedures: an introduction to the use of evaluation frameworks. Developments in Aquaculture and Fisheries Science, 36:379-407, 2006.

D. Kolody, M. Herrera, and J. Million. 2009 Indian Ocean skipjack tuna stock assessment (stock synthesis). Technical report, IOTC-2011-WPTT-14 (Rev1), 2009.

H.-H. Lee, M. N. Maunder, K. R. Piner, and R. D. Methot. Estimating natural mortality within a fisheries stock assessment model: an evaluation using simulation analysis based on twelve stock assessments. Fish. Res., 109(1):89–94, 2011.

H.-H. Lee, M. N. Maunder, K. R. Piner, and R. D. Methot. Can steepness of the stock-recruitment relationship be estimated in fishery stock assessment models? Fish. Res., 125:254–261, 2012.

M. Maunder and K. Piner. Dealing with data conflicts in statistical inference of population assessment models that integrate information from multiple diverse data sets dealing with data conflicts in statistical inference of population assessment models that integrate information from multiple diverse data sets. Submitted to Fish. Res., 2016.

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

M. R. Payne, L. W. Clausen, and H. Mosegaard. Finding the signal in the noise: objective data-selection criteria improve the assessment of western baltic spring-spawning herring. ICES Journal of Marine Science: Journal du Conseil, 2009. doi: 10.1093/icesjms/fsp185. URL http://icesjms.oxfordjournals.org/content/early/2009/07/09/icesjms.fsp185.abstract.

P. Pepin and C. T. Marshall. Reconsidering the impossible??? Linking environmental drivers to growth, mortality, and recruitment of fish 1. Canadian Journal of Fisheries and Aquatic Sciences, 72(999): 1–11, 2015.

K. R. Piner, H.-H. Lee, M. N. Maunder, and R. D. Methot. A simulation-based method to determine model misspecification: examples using natural mortality and population dynamics models. Marine and Coastal Fisheries, 3(1):336–343, 2011

S. N. Rodionov. A sequential algorithm for testing climate regime shifts. Geophysical Research Letters, 31(9), 2004.

J. T. Schnute and R. Hilborn. Analysis of contradictory data sources in fish stock assessment. Can. J. Fish. Aquat. Sci., 50(9):1916–1923, 1993.

C. Spearman. ”general intelligence”, objectively determined and measured. The American Journal of Psychology, 15(2):201–292, 1904.

C. S. Szuwalski, K. A. Vert-Pre, A. E. Punt, T. A. Branch, and R. Hilborn. Examining common assumptions about recruitment: a meta-analysis of recruitment dynamics for worldwide marine fisheries. Fish and Fisheries, 2014.

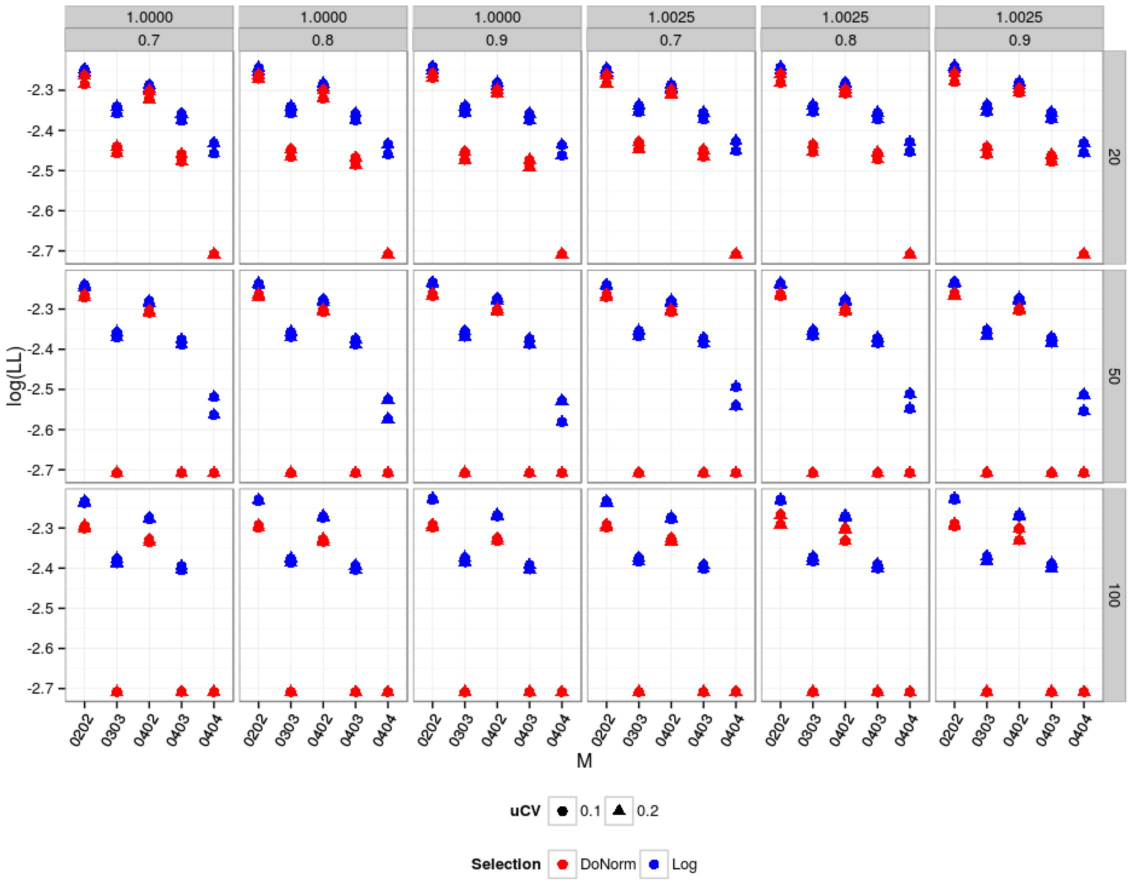

Figure 1. Log-ikelihoods for all 720 scenarios.

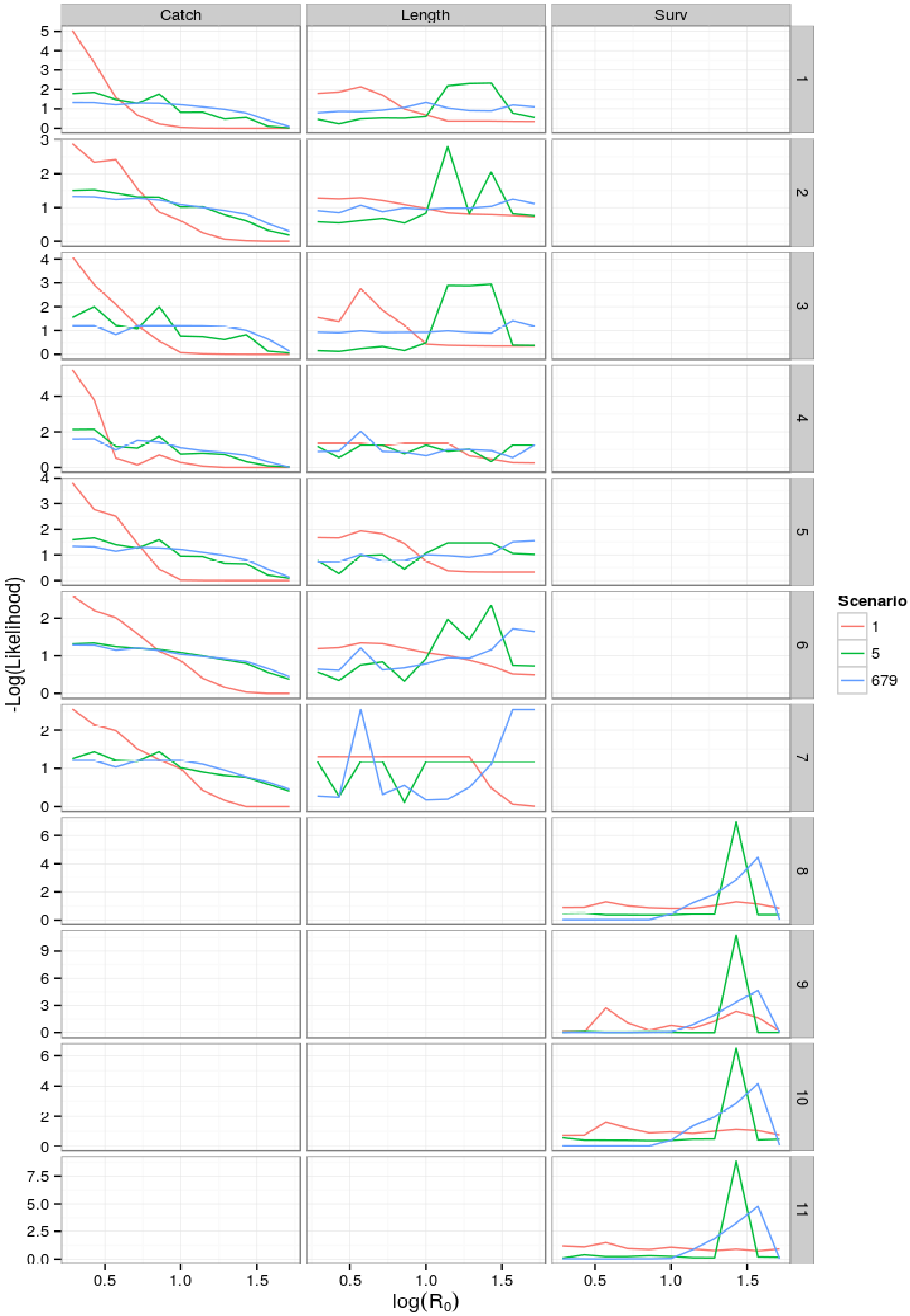

Figure 2. Profiles of negative log likelihoods by data component (column) and fleet.

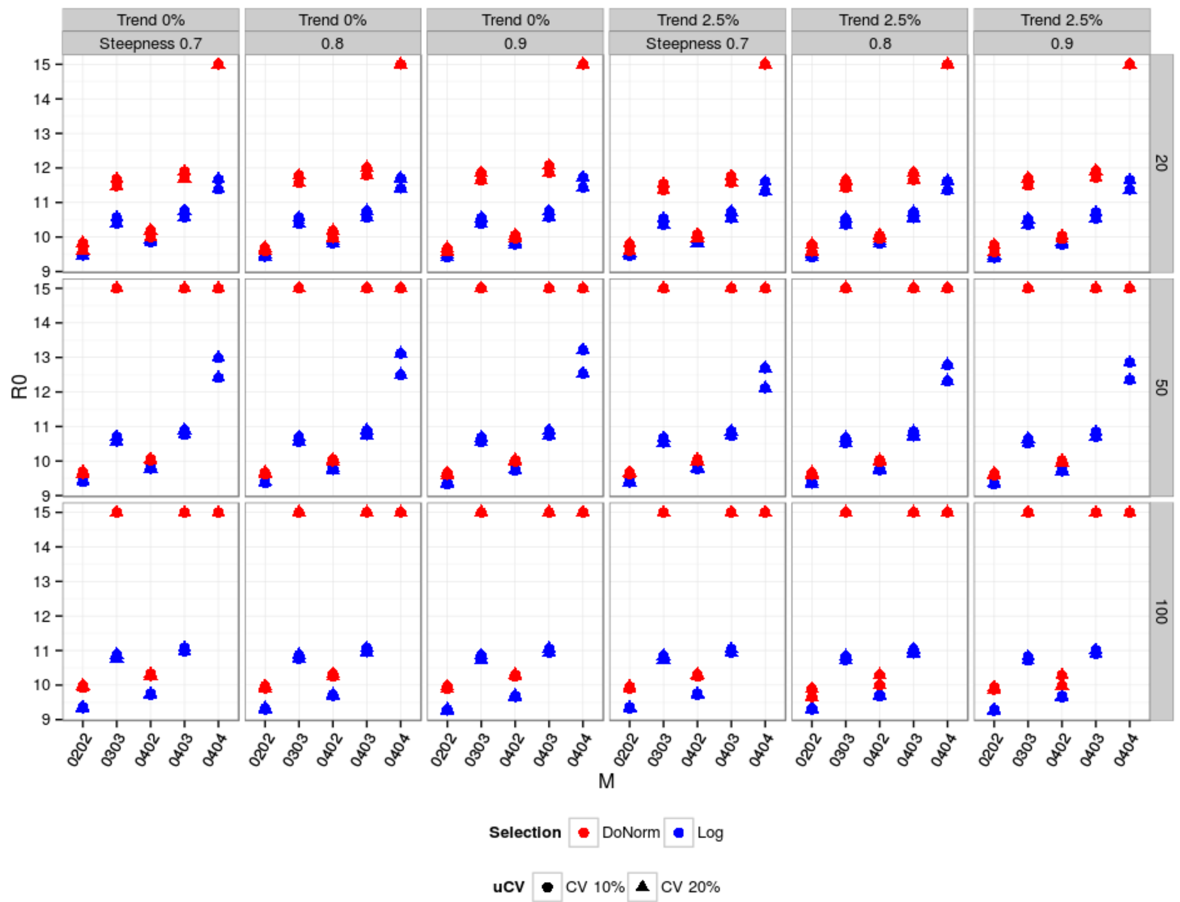

Figure 3. Condition Numbers for the grid.

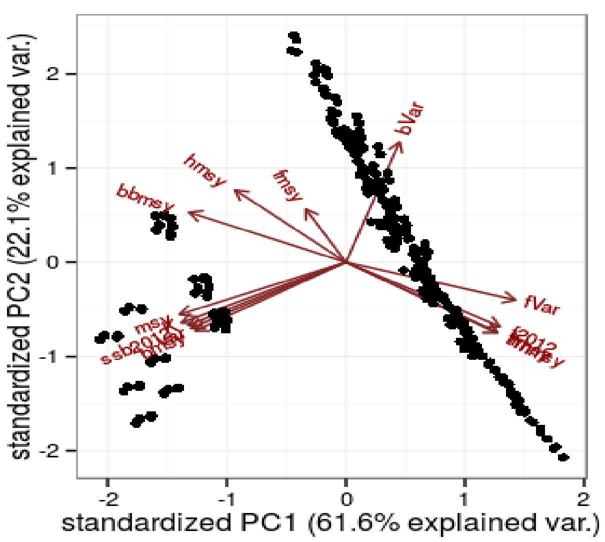

Figure 4. Bi-plot showing the scenarios by the 1st 2 principle components.

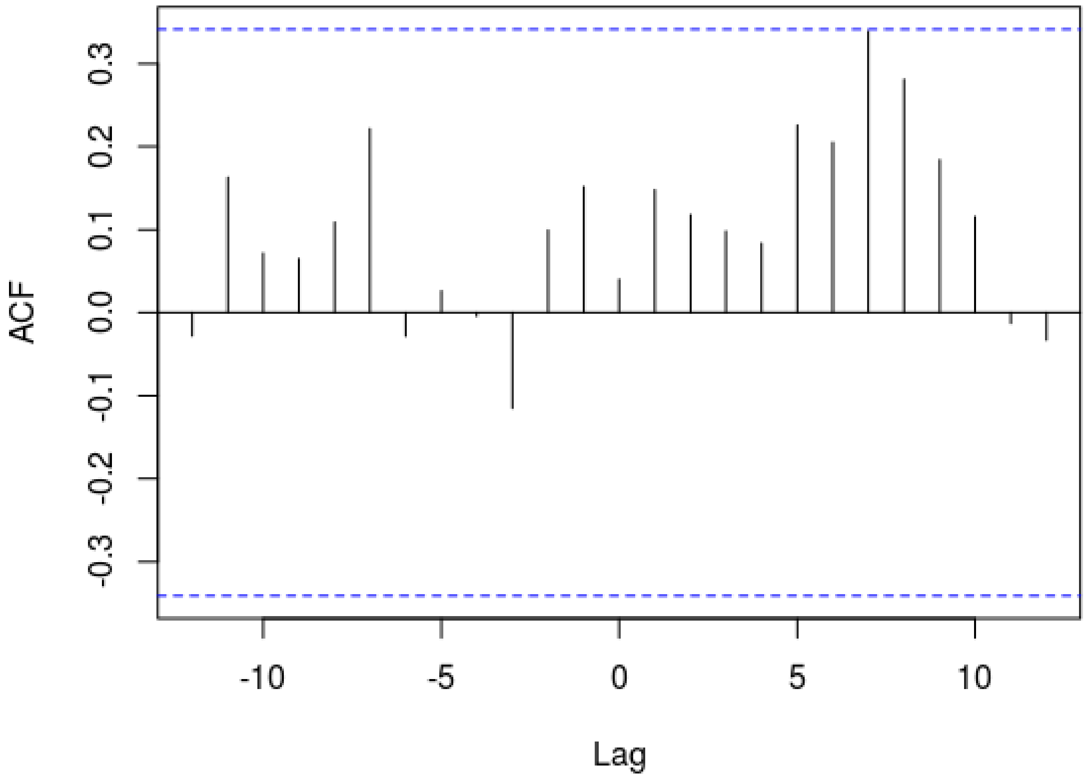

Figure 5. Cross correlations between recruitment and SSB.

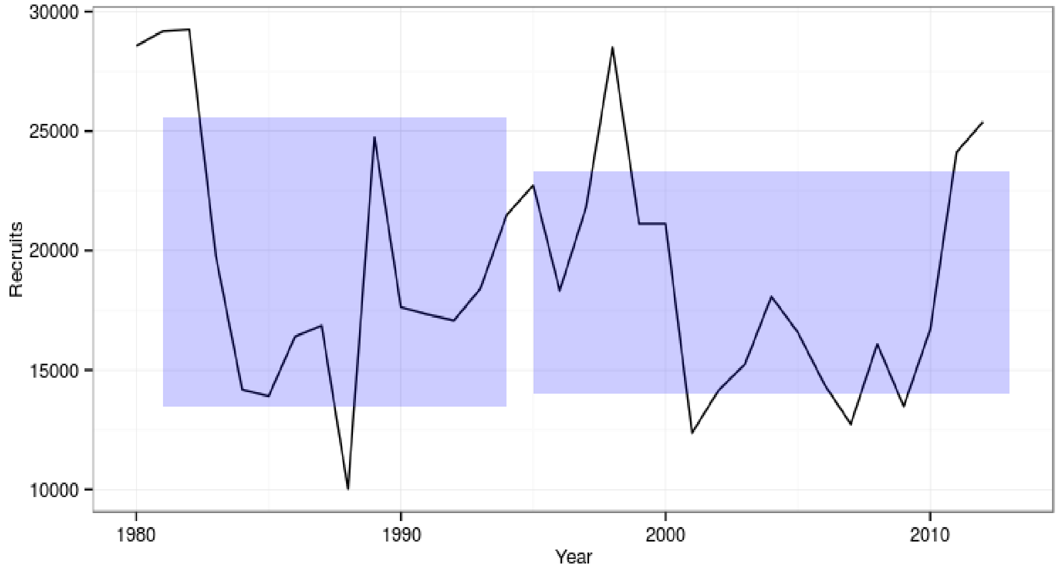

Figure 6. Recruitment regimes.