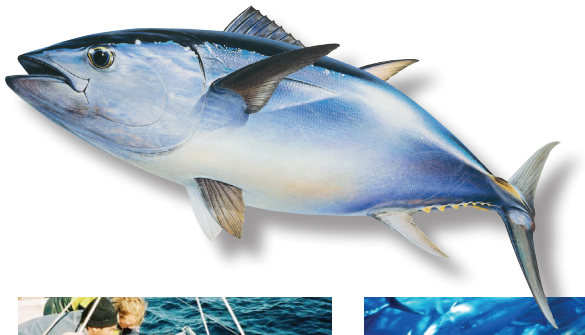
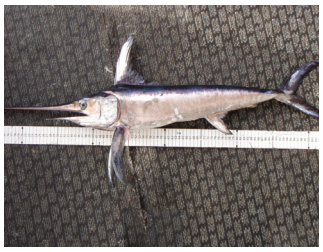
National Research
FLAGSHIPS

CSIRO





www.csiro.au

Development of a robust suite of stock status indicators for the Southern and Western and the Eastern Tuna and Billfish fisheries

Wealth from Oceans National Research Flagship

FRDC Project No. 2003/042

June 2008

Marinelle Basson
and Natalie A. Dowling

Australian Government
Fisheries Research and Development Corporation

Author: Basson, Marinelle.

Title: Development of a robust suite of stock status indicators for the Southern and Western and the Eastern Tuna and Billfish Fisheries / Marinelle Basson ; Natalie A. Dowling.

ISBN: 9781921424359 (pbk.)

Subjects: Tuna fisheries--Australia--Evaluation.
Billfish fisheries--Australia--Evaluation.
Fish stock assessment--Australia--Research.
Fishery management--Australia.
Fishery resources--Australia--Management.

Other Authors/Contributors:
Dowling, Natalie Anne, 1975-
CSIRO. Marine and Atmospheric Research.

Dewey Number: 639.277830994

Table of Contents

1. Non Technical Summary 1
2. Acknowledgements 4
3. Background and Need 4
4. Objectives 6
5. Outline of Research 6
6. Candidate Stock Status Indicators 11
Summary 11
6.1. Introduction 11
6.2. Candidate indicators of stock status or harvest rate 13
6.3. Candidate indicators of change in spatial distribution 15
6.4. Quantities which could reflect changes in targeting 15
6.5. Indicators used in further analyses 16
7. Standardisation of size-based indicators: three case studies 19
Summary 19
7.1. Introduction 19
7.2. Postscript regarding Swordfish standardisation 20
7.3. Attachment 1: Bigeye and yellowfin tuna example 22
7.4. Attachment 2: Broadbill swordfish example 48
8. An investigation of the responsiveness of CPUE- and size-based indicators 71
Summary 71
8.1. Introduction 71
8.2. Methods 73
8.3. Results 80
8.4. Discussion and Conclusions 98
8.5. Annex of additional figures to Section 8 101
9. The performance of single indicators and groups of indicators as predictors of relative spawning biomass 115
Summary 115
9.1. Introduction 116
9.2. Methods 117
9.3. Results: Prediction errors of single indicators 129
9.4. Results: Classification Trees for multiple indicators 146
9.5. Discussion and Conclusions 158
9.6. Annex of additional figures to Section 9 161
10. An evaluation of the performance of single, and groups of, indicators in feedback decision rules 163
Summary 163
10.1. Introduction 164
10.2. Methods 165
10.3. Results: Decision rule performance for a long-lived life-history and high steepness 176
10.4. Sensitivities: steepness, selectivity and effort creep 192
10.5. Discussion and Conclusions 207
11. Feedback dynamics of simple catch and harvest rate decision rules: a theoretical investigation 209
Summary 209
11.1. Introduction 211
11.2. General methodology 213
11.3. Catch decision rules based on change in abundance 215
11.4. Harvest rate Decision rules based on change in abundance 220
11.5. Sigmoid Catch and Harvest rate rules 225

11.6. Discussion and Conclusions 238
12. Implications of sexual dimorphism in swordfish growth for size-based indicators 241
Summary 241
12.1. Introduction 241
12.2. Methods 242
12.3. Results 248
12.4. Discussion and conclusions 262
13. Swordfish: The performance of decision rules in the presence of fish and fleet movement 265
Summary 265
13.1. Introduction 266
13.2. Methods 267
13.3. Results: No additional uncontrolled harvest 276
13.4. Results: Additional uncontrolled harvest 291
13.5. Results: Sensitivity trials 292
13.6. Discussion and conclusions 302
13.7. Annex of additional Figures to Section 13. 306
14. Benefits and adoption 313
15. Further Development 314
16. Planned Outcomes 314
17. Conclusion 315
18. References 317
19. Intellectual Property 320
20. Staff 320
21. Appendix 1: Technical description of the operating model 321
1. General population dynamics 321
2. General fishery dynamics 323
3. Stock Status Indicators 326
4. Inputs to the simulation model 327
5. Starting a simulation run 328
6. Decision rules 330
7. Translating catch quota fishing mortality 334
8. Spatial triggers and external fishing 335
9. Performance indicators 335
22. Appendix 2. Reference points in fisheries management: a background paper 337
Summary 337
1. Introduction 338
2. What are reference points and why do we need them? 338
3. Types of reference points: targets and limits. 339
4. Multiple objectives - multiple reference points 340
5. What 'currency' should we use to define reference points? 341
6. Candidates for reference points 342
7. Implementation 345
8. Management procedures, Decision Rules and Reference points 347

List of Tables

Table 1. Candidate Indicators of Stock status. 14
Table 2. Candidate Indicators of change in spatial distribution 15
Table 3. Quantities which could reflect changes in targeting or fishery characteristics (excluding measures in Tables 1 and 2 which may also reflect such changes). 16
Table 4. Grid showing the three values of the input parameters evaluated: steepness (h), von-Bertalanffy growth rate (k), natural mortality, age at maturity (tm), age at first capture (tc= tm/2, tm, tm+1).....76
Table 5. Input parameters for the two life-history scenarios: 'low m' or long-lived, and 'high m' or short-lived....78
Table 6. Return times (for an indicator to get within 5% of its long term mean at F_{msy}) following a drop harvest rate from $3*F_{msy}$ to F_{msy}. Steepness is 0.999, and recruitment variability is relatively high: sigR=σ_R=0.4. 98
Table 7. Summary of parameter values for components of each scenario. The naming conventions used to describe a scenario are given in italics (e.g. High m, high h, flat, HHH, etc.) 126
Table 8. CVs of proportion at age implied by the α-parameter for the Derichlet multinomial distribution for flat and dome-shaped selectivity. CVs are given for the first age class, and an upper age class which approximately covers the main part of the catch-at-age distribution. At older age classes the CVs are even higher than those shown against 'upper class'. The minimum CV (for the age class with highest proportion in the catch) is also shown. 127
Table 9 Summary of harvest rate values input to the sine function which generates the time-series of harvest rates in scenarios, and implications in terms of the mean, minimum and maximum harvest rate for each scenario. 128
Table 10. Correlation matrix for the 13 indicators for the low m, HIGH steepness, high noise runs. Correlations are for the raw simulated data, not the converted binary values. The correlation of 1 between mdW and mdL and between W90 and L90 is due to the fact that there is no error or noise in the length-to-weight conversion. 136
Table 11. Correlation matrix for the 13 indicators for the low m, LOW steepness, high noise runs. Correlations are for the raw simulated data, not the converted binary values. The correlation of 1 between mdW and mdL and between W90 and L90 is due to the fact that there is no error or noise in the length-to-weight conversion.... 137
Table 12 Prediction error for scenarios with low m, high measurement errors (HHH) and for two levels of steepness and two types of selectivity. Datasets are indicated along the top, trees are indicated down the side. Where a prediction is based on a 'wrong' assumption, both the tree and the trigger/reference points for indicators are based on the 'wrong' assumptions. 148
Table 13. Prediction error for scenarios with high m, high measurement errors (HHH) and for two levels of steepness and two types of selectivity. Datasets are indicated along the top, trees are indicated down the side. Where a prediction is based on a 'wrong' assumption, both the tree and the trigger/reference points for indicators are based on those 'wrong' assumptions. 151
Table 14. Summary table of 'correct' tree-data matches prediction errors compared with best single indicator under SPR triggers 154
Table 15. Summary of split level, indicator(s) at that split level and the primary surrogate indicator. These results are for trees based on data from the low m life history, high measurement error (HHH) and SPR triggers. 155
Table 16. Summary of split level, indicator(s) at that split level and the primary surrogate indicator. These results are for trees based on data from the high m life history, high measurement error (HHH) and SPR triggers. 155
Table 17. Prediction errors for trees based on data with and without effort creep and then applied to data with or without effort creep. The scenario is for low m, low steepness (h=0.4), flat selectivity and SPR triggers. 156
Table 18. Prediction errors for trees and datasets derived from different assumptions about steepness, and for a tree based on the combined data for all steepness levels. Other quantities are: low m, flat selectivity and SPR triggers. As before, h=0.9 implies h~U(0.85, 0.95) and similarly for other h-values. 157
Table 19. List of indicators with associated abbreviated names (more detail in Section 8 and Appendix 1) 175

Table 20. List of performance measures generated by the simulation model for each scenario (of 'n' realisations). 'over hs period' means over the period that the decision rule was implemented – years 11 to 60. 176
Table 21. Prediction error for two trees and two datasets..... 201
Table 22. Summary of equations for Decision rule examples used with the Schaefer and Fox stock production models, together with equations for the non-zero eigenvalue and the stability conditions at the extremes of possible biomass values, B*=0 and B*=K. For the Fox model the shorthand notation lnB*=ln(B*) and lnK=ln(K) are used. 225
Table 23. Parameters for von Bertalanffy growth curves from literature 243
Table 2: Von Bertalanffy growth parameter values for the stylised sexually dimorphic growth curves, and the parameter values reported by Young and Drake (2004) for swordfish from the west coast of Australia 245
Table 24. The 8 base case scenarios, run for each of the 3 decision rules, both with no external fishing mortality, and with external fishing mortality equating to 3.F_{msy}..... 275
Table 25. List of scenarios considered in sensitivity trials; a blank cell means no runs were done for that scenario. The default value for steepness is h=0.65 unless otherwise stated, e.g. in scenarios 6-9. . 276
Table A1.1 Listing of the parameters used in the decision tree..... 333
Table A1.2. Performance measures for the Spatial model simulations. Time-series are generated for all quantities and in some cases summary statistics are derived (2nd column) 336

Please note: Tables in Attachment 1 (p22-47) and Attachment 2 (p48-70) are not included in this list.

List of Figures

Figure 1. Correlated life-history parameters used in the evaluation of the expected relative change in indicators when harvesting at Fmsy. Parameters are natural mortality (m), von Bertalanffy growth rate (k), steepness (h), age at maturity (tm) and age at first capture (tc). See main text (methods) for more detail. 78
Figure 2. B_{msy}/B_0 for the subset with h=0.6, and by m*k pairs as indicated on the x-axis. For example, 0.4*0.2 means m=0.4, k=0.2. Each 'box' shows the range of t_m and t_c values for the given m*k combination. 80
Figure 3. B_{msy}/B_0 (relative spawning biomass) for different values of k, m and steepness (h). Numbers on the x-axis refer to m*k pairs with m=(1=0.1, 2=0.2, 3=0.4) and k=(1=0.2, 2=0.4, 3=0.6). Factors reflected in the 'boxes' are age at maturity and first capture, t_m and t_c. 80
Figure 4. Relative mean length (meanL(F_{msy})/meanL(F=0)) for harvest at F_{msy} given 3 steepness (h) levels. Numbers on the x-axis refer to m*k pairs with m=(0.1,0.2,0.4) and k=(0.2,0.4,0.6). Factors reflected in the 'boxes' are age at maturity (t_m) and age at first capture (t_c). 81
Figure 5 Relative 90th percentile of weight (W90(F_{msy})/W90(F=0)) for harvest at F_{msy} given 3 steepness (h) levels. Numbers on the x-axis refer to m*k pairs with m=(0.1,0.2,0.4) and k=(0.2,0.4,0.6). Factors reflected in the 'boxes' are age at maturity (t_m) and age at first capture (t_c) 82
Figure 6. Expected relative change in mean length when harvesting at Fmsy for flat selectivity (top panel; identical to Figure 4), dome-shaped selectivity with 4 fully selected age classes (middle panel) and with 2 fully selected age classes (lower panel). Each panel shows results by steepness level (h) and m*k pairs, as indicated on the x-axis. 83
Figure 7. Expected relative change in the 90th percentile of weight when harvesting at F_{msy} for flat selectivity (top panel), dome-shaped selectivity with 4 fully selected age classes (middle panel) and with 2 fully selected age classes (lower panel). Each panel shows results by steepness level (h) and m*k pairs, as indicated on the x-axis. 84
Figure 8. 'Change ratios', [1 – indicator(F_{msy})/indicator(F=0)]/[1-B_{msy}/B_0], for 7 size-based indicators and 5 CPUE-based indicators (see sub-section 2.3 above for descriptions of indicator labels). Scenarios are the grid set of life-history parameters. The top panel is for flat selectivity, and the lower panel is for dome-shaped selectivity with 2 fully selected age classes. 85
Figure 9. Flat selectivity, Fmsy. Relative change in SSB (circles) and other indicators as a function of steepness (h). Indicators are identified in the title of each panel: triangles relate to the first listed indicator and '+'s to the 2nd indicator in the list. 86
Figure 10. Dome-shaped selectivity, width 4, Fmsy. Relative change in SSB (circles) and other indicators as a function of steepness (h). Indicators are identified in the title of each panel: triangles relate to the first listed indicator and '+'s to the 2nd indicator in the list. 87
Figure 11. Dome-shaped selectivity, width 2, Fmsy. Relative change in SSB (circles) and other indicators as a function of steepness (h). Indicators are identified in the title of each panel: triangles relate to the first listed indicator and '+'s to the 2nd indicator in the list. 87
Figure 13. Change ratio, [1 – indicator(F_{msy})/indicator(F=0)]/[1-B_{msy}/B0], for 7 size-based indicators and 5 cpue-based indicators – see sub-section 2.3 for indicator names. Scenarios are for the correlated set of life-history parameters. The top panel is for the flat selectivity curve; the lower panel for dome-shaped selectivity with 2 fully selected age classes 89
Figure 14. Ratios, F_{spr40}/F_{msy} plotted against steepness (h) for flat and dome-shaped selectivity (of width 2 age classes) and for the correlated set of life-history parameters. 91
Figure 15. Relative values for indicators under the harvest rate F_{SPR40} (triangles) and F_{MSY} (open circles) plotted against steepness (h). The panels are for median length (mdL, top left), median weight (mdW, top right), cpue in weight (uw, bottom left) and cpue of big fish (ub, bottom right). Results are for the correlated life-history dataset with flat selectivity. 91
Figure 16. Relative differences between expected (relative) indicator values under the harvest rate F_{SPR40} and F_{MSY}, plotted against steepness (h). The panels are for median length (mdL, top left), median weight (mdW, top right), cpue in weight (uw, bottom left) and cpue of big fish (ub, bottom right). Results are for the correlated life-history dataset with flat selectivity. 92
Figure 17. Relative differences between expected (relative) indicator values under the harvest rate F_{SPR40} and F_{MSY}, plotted against steepness (h). The panels are for median length (mdL, top left), median

weight (mdW, top right), cpue in weight (uw, bottom left) and cpue of big fish (ub, bottom right). Results are for the correlated life-history dataset with dome-shaped selectivity.....92
Figure 18. Time series of recruitment (relative to the mean) showing high variability (σ_R=0.7, solid line) and low variability (σ_R =0.1, dashed line). 93
Figure 19. Time series of median weight (relative to the mean) for the high recruitment variability (σ_R=0.7, solid line) and low recruitment variability (σ_R =0.1, dashed line). 93
Figure 20. Time series of cpue of spawners (relative to the mean) for the high recruitment variability (σ_R =0.7, solid line) and low recruitment variability (σ_R =0.1, dashed line). 94
Figure 21. CVs of spawning biomass (B), fishable biomass (Fb) and 13 size- and cpue-based indicators as a function of (i) different harvest rates (F, left panel) at a recruitment CV of 40% and (ii) different recruitment variabilities (sigR=σ_R, right panel) at F=2*F_{msy}=0.5. 94
Figure 22. CVs of spawning biomass (B), fishable biomass (Fb) and 13 size- and cpue-based indicators as a function of different recruitment variabilities (sigR=σ_R) and (i) flat selectivity (left panel) or (ii) dome-shaped selectivity (right panel). In all cases the harvest rate was F=F_{msy}, which is 0.23 for the flat and 0.36 for dome-shaped selectivity. 95
Figure 23. CVs of spawning biomass (B), fishable biomass (Fb) and 13 size- and cpue-based indicators as a function of different recruitment variabilities (sigR=σ_R) for the high natural mortality case and (i) flat selectivity (left panel) or (ii) dome-shaped selectivity (right panel). In all cases the harvest rate was F=F_{msy} which is 1.2 for both flat and dome-shaped selectivity. 95
Figure 24. Autocorrelation in mean length (mL) and mean weight (mW) for flat and dome-shaped selectivity. The solid line is for a population with high m (=0.4); the dashed line is for a population with low m (=0.1). Both examples were harvested at F_{msy} and had a high recruitment variability, (sigR=σ_R=0.7). 96
Figure 25. Autocorrelation in the proportion big fish (pb) and cpue in numbers (u) for flat and dome-shaped selectivity. The solid line is for a population with high m (=0.4); the dashed line is for a population with low m (=0.1). Both examples were harvested at F_{msy} and had a high recruitment variability, (sigR=σ_R=0.7). 97
Figure A 1. Relative SSB at F_{msy} for a range of steepness (h) levels and combinations of m and k (shown on the x-axis as m*k). The panels are: flat selectivity (top), dome-shaped with 4 fully selected age classes (middle) and dome-shaped with 2 fully selected age classes (bottom). 101
Figure A 2. Relative median length at F_{msy} for a range of steepness (h) levels and combinations of m and k (shown on the x-axis as m*k). The panels are: flat selectivity (top), dome-shaped with 4 fully selected age classes (middle) and dome-shaped with 2 fully selected age classes (bottom). 102
Figure A 3. Relative 90th percentile of length at F_{msy} for a range of steepness (h) levels and combinations of m and k (shown on the x-axis as m*k). The panels are: flat selectivity (top), dome-shaped with 4 fully selected age classes (middle) and dome-shaped with 2 fully selected age classes (bottom). 103
Figure A 4. Relative mean length at F_{msy} for a range of steepness (h) levels and combinations of m and k (shown on the x-axis as m*k). The panels are: flat selectivity (top), dome-shaped with 4 fully selected age classes (middle) and dome-shaped with 2 fully selected age classes (bottom). 104
Figure A 5. Relative median weight at F_{msy} for a range of steepness (h) levels and combinations of m and k (shown on the x-axis as m*k). The panels are: flat selectivity (top), dome-shaped with 4 fully selected age classes (middle) and dome-shaped with 2 fully selected age classes (bottom). 105
Figure A 6. Relative 90th percentile of weight at F_{msy} for a range of steepness (h) levels and combinations of m and k (shown on the x-axis as m*k). The panels are: flat selectivity (top), dome-shaped with 4 fully selected age classes (middle) and dome-shaped with 2 fully selected age classes (bottom). 106
Figure A 7. Relative proportion big fish at F_{msy} for a range of steepness (h) levels and combinations of m and k (shown on the x-axis as m*k). The panels are: flat selectivity (top), dome-shaped with 4 fully selected age classes (middle) and dome-shaped with 2 fully selected age classes (bottom). 107
Figure A 8. Relative proportion small fish at F_{msy} for a range of steepness (h) levels and combinations of m and k (shown on the x-axis as m*k). The panels are: flat selectivity (top), dome-shaped with 4 fully selected age classes (middle) and dome-shaped with 2 fully selected age classes (bottom). 108
Figure A 9. Relative cpue in numbers at F_{msy} for a range of steepness (h) levels and combinations of m and k (shown on the x-axis as m*k). The panels are: flat selectivity (top), dome-shaped with 4 fully selected age classes (middle) and dome-shaped with 2 fully selected age classes (bottom). 109
Figure A 10. Relative cpue of big fish at F_{msy} for a range of steepness (h) levels and combinations of m and k (shown on the x-axis as m*k). The panels are: flat selectivity (top), dome-shaped with 4 fully selected age classes (middle) and dome-shaped with 2 fully selected age classes (bottom). 110

Figure A 11. Relative cpue of small fish at F_{msy} for a range of steepness (h) levels and combinations of m and k (shown on the x-axis as m*k). The panels are: flat selectivity (top), dome-shaped with 4 fully selected age classes (middle) and dome-shaped with 2 fully selected age classes (bottom). 111
Figure A 12. Relative cpue in weight at F_{msy} for a range of steepness (h) levels and combinations of m and k (shown on the x-axis as m*k). The panels are: flat selectivity (top), dome-shaped with 4 fully selected age classes (middle) and dome-shaped with 2 fully selected age classes (bottom). 112
Figure A 13. Relative cpue of spawners (in weight) at F_{msy} for a range of steepness (h) levels and combinations of m and k (shown on the x-axis as m*k). The panels are: flat selectivity (top), dome-shaped with 4 fully selected age classes (middle) and dome-shaped with 2 fully selected age classes (bottom). 113
Figure 26. Life-history with low m. Combinations of m, k, t_m and t_c for 100 realisations. Values were generated according to the description in the text (above). 124
Figure 27. Life-history with high m. Combinations of m, k, t_m and t_c for 100 realisations. Values were generated according to the description in the text (above). 124
Figure 28. Examples of selectivity curves for the long-lived (low m) life history in the top two panels and for the short-lived (high m) life-history in the bottom panels. The flat-topped selectivity curves are on the left, and the dome-shaped selectivity curves are on the right. 125
Figure 29. Two examples of harvest rate time-series (over 90 time-steps) used in simulations. The solid line is for an input F=0.3; the dashed line is for an input F=0.1. 128
Figure 30. Prediction errors for a range of relative trigger point values for mean length (left panel) and cpue (right panel). The biomass target is 0.23 (B_{msy}/B_0). The dotted lines show the type I and type II errors respectively and the solid line shows the overall prediction error. The dashed vertical line (on the left) shows where the MSY-based trigger is and the vertical dotted line (on the right) shows where the SPR40% trigger lies. Both panels are based on a simulation with input F=0.3 ($1.2F_{msy}$). 129
Figure 31. Prediction errors for a range of relative trigger point values for mean length. The biomass reference is 0.23 (B_{msy}/B_0). The dotted lines show the type I and type II errors respectively and the solid line shows the overall prediction error. The vertical dashed line (on the left) shows where the MSY-based trigger is and the vertical dotted line (on the right) shows where the SPR40% trigger lies. The left-hand panel is based on a simulation with input F=0.1 and the right-hand panel on input F = 0.6 to generate sine-wave time-series of F. F_{msy} is 0.25. 130
Figure 32. Prediction errors of indicators for 4 different harvest rate scenarios. The first three are based on the sine wave with input values of 0.1, 0.3 and 0.6 respectively (f01, f03, f06). 'f03.const' has a lognormal distribution around 0.3, and only varies between 0.19 and 0.44 (CV=0.2). The sine wave version (f03) varies between 0.03 and 0.62 (see Methods sub-section for more detail; descriptions of indicator labels, horizontal axis, are given in sub-section 2.1 above) 131
Figure 33. Indicators ranked by prediction error (low error implies a low rank; low rank is good) for each of the harvest rate series and sorted according to the scenario with input F=0.3. Harvest rate scenarios are as in Figure 32. Descriptions of indicator labels are given in sub-section 2.1 above. . 131
Figure 34. Prediction error of size based indicators using MSY-based triggers, either as absolute quantities ('abs') or relative to unexploited values ('rel'). This example is for low m, low noise, high h and widely varying harvest rate. 132
Figure 35. Prediction errors for the 13 indicators and at 4 different levels of noise in the 3 sources of error on indicators. The 3-letter code reflects the levels (L=low, H=high) for the 3 sources of error: catch-at-age, cpue and size-frequency sample size. See text for more detail. The left panel is for high steepness, the right panel for very low steepness. All scenarios have low m and a flat selectivity curve. 132
Figure 36. Prediction errors of indicators for 4 harvest rate scenarios with high noise for all three error sources (see text). The first three are based on the sine wave with mean values of 0.1, 0.3 and 0.6 respectively. 'f03nosin' also has a mean F of 0.3, but only varies between 0.19 and 0.44 (CV=0.2). The sine wave version (f03) varies between 0.03 and 0.62. 133
Figure 37. Indicators ranked by prediction error (low error implies a low rank; low rank is good) for each of the harvest rate series and sorted according to the scenario with mean F=1.2*Fmsy. Harvest rate scenarios are as in Figure 36. 133
Figure 38. Prediction errors for different steepness ranges (the means are shown in the legend) and for low m, high noise, flat selectivity. The harvest rate is matched to the steepness with a sine wave constructed around 1.2*Fmsy (given h) 134

Figure 39. Prediction errors for two ranges of recruitment variability: high σ_R = (0.6,0.8) and low σ_R = (0.2,0.4). In all cases the measurement error is low (LLL), selectivity is flat, and steepness is high (left panel), or very low (right panel). 135
Figure 40. Prediction errors for two ranges of recruitment variability: high σ_R = (0.6,0.8) and low σ_R = (0.2,0.4). In all cases the measurement error is high (HHH), selectivity is flat, and steepness is high (left panel), or very low (right panel). 135
Figure 41. Range of prediction errors for each indicator (i.e. when calculating this by realisation) for high steepness (left panel) and low steepness (right panel). In both cases the noise level is HHH. 136
Figure 42. Prediction errors for the 13 indicators under the high (HHH) and low (LLL) noise assumptions and for flat selectivity (left panel) and dome-shaped selectivity (right panel). In all cases the scenario is low m, high steepness. 138
Figure 43. Correlation between each indicator and the spawning biomass ('raw' simulated data, not binary values) for flat and dome-shaped selectivity. The scenarios are low m, high steepness and high noise (HHH). Note that the ps (proportion small fish) indicator has not been plotted because it has negative correlation which confuses the graph. The correlations for ps are: flat= -0.61, dome= -0.46. 138
Figure 44. Prediction errors for a range of steepness levels and for four noise levels: low noise (LLL), high noise (HHH), and the combinations LHL and HLH (see text for interpretation). All cases are for low m, FLAT selectivity and MSY triggers. In each panel, indicators are sorted by the prediction errors for h=0.5 (triangles). 139
Figure 45. Prediction errors for a range of steepness levels and for two noise levels: low noise (LLL) and high noise (HHH). All cases are for low m, DOME-shaped selectivity and MSY triggers. In each panel, indicators are sorted by the prediction errors for h=0.5 (triangles). 139
Figure 46. Prediction errors for a range of steepness levels and for low noise (LLL; left panel) and high noise (HHH; right panel). All cases are for high m, FLAT selectivity and MSY triggers. 140
Figure 47. Prediction errors for the very low and high steepness levels (means of 0.4 and 0.9) and for the LHL and HLH noise levels with either flat or dome-shaped selectivity. All cases are for high m and MSY triggers. Indicators are in the same order in each panel and not sorted according to prediction error. 140
Figure 48. Proportion of survivors at age (under natural mortality only) and dome-shaped selectivity at age for the long-lived (m=0.1) and short-lived (m=0.4) scenarios. 141
Figure 49. Prediction errors for two levels of recruitment variability: high σ_R = (0.6,0.8) and low σ_R = (0.2,0.4). In all cases the measurement error is high on CPUE (LHL) and steepness is high (0.9). . 141
Figure 50. Prediction errors for a range of steepness levels and for four noise levels: low noise (LLL), high noise (HHH), and the combinations LHL and HLH (see text for interpretation). All cases are for low m, FLAT selectivity and SPR triggers. In each panel, indicators are sorted by the prediction errors for h=0.5 (triangles). 142
Figure 51. Prediction errors for a range of steepness levels and for four noise levels: low noise (LLL), high noise (HHH), and the combinations LHL and HLH (see text for interpretation). All cases are for low m, DOME-shaped selectivity and SPR triggers. In each panel, indicators are sorted by the prediction errors for h=0.5 (triangles). 142
Figure 52. Prediction errors for a range of steepness levels and for four noise levels: low noise (LLL), high noise (HHH), and the combinations LHL and HLH (see text for interpretation). All cases are for high m, FLAT selectivity and SPR triggers. In each panel, indicators are sorted by the prediction errors for h=0.5 (triangles). 143
Figure 53. Prediction errors for a range of steepness levels and for four noise levels: low noise (LLL), high noise (HHH), and the combinations LHL and HLH (see text for interpretation). All cases are for high m, DOME-shaped selectivity and SPR triggers. In each panel, indicators are sorted by the prediction errors for h=0.5 (triangles). 143
Figure 54. Prediction errors for widely varying harvest rate (F is a sine function, 'sine') or randomly varying F around a constant ('const') and for low (LLL) and high (HHH) noise levels. All cases are for low m, dome-shaped selectivity and using SPR triggers. 144
Figure 55. The effect of effort creep on prediction errors for high (h=0.9) and very low steepness (h=0.4) with low mortality, high noise and flat selectivity. 145
Figure 56. The effect of effort creep on prediction errors for high (h=0.9) and very low steepness (h=0.4) with low mortality, high noise and dome-shaped selectivity. 146

Figure 57. Relative error as a function of the number of splits in a tree for low mortality, high noise and using MSY triggers. Four cases are shown: high and very low steepness (h=0.9 or h=0.4) combined with flat or dome-shaped selectivity (see legend). 147
Figure 58. Structures of pruned trees for the low mortality scenario with high noise and using MSY trigger points. The four trees are for high and very low steepness (h=0.9 or h=0.4) combined with either flat or dome-shaped selectivity. See main text (above) for interpretation of notation in trees. 148
Figure 59. Low mortality, high noise and MSY triggers. The relative change in prediction error between the best single indicator, the correct tree and several 'wrong' trees. See text for more detail. 149
Figure 60. Relative error as a function of the number of splits in a tree for low mortality, high noise and using MSY triggers. Four cases are shown: high and very low steepness (h=0.9 or h=0.4) combined with flat or dome-shaped selectivity (see legend). 149
Figure 61. Structures of pruned trees for the high mortality scenario with high noise and using MSY trigger points. The four trees are for high and very low steepness (h=0.9 or h=0.4) combined with either flat or dome-shaped selectivity. See main text (above) for interpretation of notation in trees. 150
Figure 62. High mortality, high noise and MSY triggers. The relative change in prediction error between the best single indicator, the correct tree and several 'wrong' trees. See text for more detail. 151
Figure 63. Relative error as a function of the number of splits in a tree for low mortality, high noise and using SPR triggers. Four cases are shown: high and very low steepness (h=0.9 or h=0.4) combined with flat or dome-shaped selectivity (see legend). 152
Figure 64. Relative error as a function of the number of splits in a tree for high mortality, high noise and using SPR triggers. Four cases are shown: high and very low steepness (h=0.9 or h=0.4) combined with flat or dome-shaped selectivity (see legend). 152
Figure 65. Structures of pruned trees for the low mortality scenario with high noise and using SPR trigger points. The four trees are for high and very low steepness (h=0.9 or h=0.4) combined with either flat or dome-shaped selectivity. 153
Figure 66. Structures of pruned trees for the high mortality scenario with high noise and using SPR trigger points. The four trees are for high and very low steepness (h=0.9 or h=0.4) combined with either flat or dome-shaped selectivity. 154
Figure 67. The relative change in prediction error between the best single indicator, the correct tree and several 'wrong' trees (used with 'wrong' triggers). The left panel is for low m, the right panel for high m. In both cases the HHH noise level was used. The x-axis identifies the combination of steepness (high, h9; very low, h4) and selectivity curve (flat; dome). See text for more detail. 155
Figure 68. Comparison of pruned trees with no lagged indicators included in the fit (left panel) and with lagged versions of the indicators mL, ub and uSB (namely mL2, ub2 and uSB2; lagged by 1 time step; right panel) included in the fit. The scenario is for low m, high h, flat selectivity and SPR triggers. 157
Figure 69. Tree based on data for 4 levels of h combined. The life-history is low m, selectivity flat, and SPR triggers were used. 158
Figure A 14. Prediction errors for a range of steepness levels and for low noise (LLL; left panel) and high noise (HHH; right panel). All cases are for low m, FLAT selectivity and MSY triggers. 161
Figure A 15. Prediction errors for a range of steepness levels and for low noise (LLL; left panel) and high noise (HHH; right panel). All cases are for low m, DOME-shaped selectivity and MSY triggers. . 161
Figure A 16. Prediction errors for the long-lived life history at a range of steepness levels. The top row shows flat selectivity with high noise (HHH, left panel) and high CPUE CV (LHL, right panel). The bottom row shows dome-shaped selectivity with high noise (HHH, left panel) and high CPUE CV (LHL, right panel). SPR triggers based on 40% SPR and a relative biomass reference point 40% were used. 162
Figure A 17. Prediction errors for the short-lived life history at a range of steepness levels. The top row shows flat selectivity with high noise (HHH, left panel) and high CPUE CV (LHL, right panel). The bottom row shows dome-shaped selectivity with high noise (HHH, left panel) and high CPUE CV (LHL, right panel). SPR triggers based on 40% SPR and a relative biomass reference point 40% were used. 162
Figure 70. Illustration of a single realisation for two traffic light decision rules based on the same set of indicators (mL, uSB, ub, ps, us), but with either 3 'red lights' (out of 5) triggering a reduction in F

(solid line), or 1 'red light' (out of 5) triggering a reduction in F (dashed line). The left panel is for spawning biomass relative to unexploited; the right panel for harvest rate, F. 170
Figure 71. Example of values of m, k, tm and tc used in a set of 200 replicates. The (red) lines show the true values of the parameters..... 174
Figure 72. Results of a single realisation of the 'Up-down' decision rule based on 2 years' indicator values to calculate the slope, and an adjustment factor d=0.9; mean length as indicator (dashed line) and CPUE as indicator (grey dotted line). A constant harvest rate at Fspr40 from year 11 onward is shown for comparison (solid line). The left panel shows spawning biomass relative to unexploited (relB); middle panel shows harvest rate (F); right panel shows catch relative to that in the first year (relC). 178
Figure 73. Results of a single realisation of the 'Up-down' decision rule based on 5 years' indicator values to calculate the slope, and an adjustment factor d=0.9; mean length as indicator (dashed line) and CPUE as indicator (grey dotted line). A constant harvest rate at Fspr40 from year 11 onward is shown for comparison (solid line). The left panel shows spawning biomass relative to unexploited (relB); middle panel shows harvest rate (F); right panel shows catch relative to that in the first year (relC). 178
Figure 74. Results of two sets of realisation for the 'Up-down' decision rule without upper limit on F (solid line) and with an upper limit Fmax=3.Fspr40 on F (dashed line). In both cases the indicator is CPUE, the slope is based on 2 years and d=0.9. The top set of panels is for one pair of realisations, the lower set of panels for a second pair of realisations. The left panel shows spawning biomass relative to unexploited (relB); middle panel shows harvest rate (F); right panel shows catch relative to that in the first year (relC). 179
Figure 75. Results of one set of realisation for the 'Up-down' decision rule without an upper limit on F (solid line) and with Fmax=Fspr40 (dashed line). In both cases the indicator is CPUE, the slope is based on 2 years and d=0.9. The left panel shows spawning biomass relative to unexploited (relB); middle panel shows harvest rate (F); right panel shows catch relative to that in the first year (relC). 180
Figure 76. Boxplots of summary results of 50 realisations for 7 scenarios using the up-down decision rule without, or with a maximum on the harvest rate. The scenario is identified on the horizontal axis: cf=constant F=0.1; 'ml' refers to mean length as indicator, 'u' to CPUE as indicator. Further extensions to the names are: 'x3' for Fmax=3.Fspr40, and 'x1' for Fmax=Fspr40. The panels show the average spawning biomass relative to unexploited, the average F and the average catch relative to Cmsy; in all cases averaged over the period that the decision rule (or hs=harvest strategy) is applied (i.e. the first 10 years are omitted from calculations)..... 181
Figure 77. Boxplots of summary results for the same 7 scenarios shown in Figure 76. The panels show the minimum relative spawning biomass, the percentage time that F was reduced and the average relative change in catch; in all cases over the period that the decision rule is applied. 181
Figure 78. Classification trees fitted to simulated data based on high steepness (h9,flat) and on low steepness (h4,flat), for the long-lived life history with flat selectivity, high measurement error levels and SPR40 reference points. Repeated from Section 9; see text for more detail. 182
Figure 79. Time series plots of a single realisation for four decision rules: binary based on mean length (dashed line), binary based on CPUE in weight (dotted line), classification tree 'h9.flat' (dash-dot-dash line) and classification tree 'h4.flat' (solid line). See text for detail. Panels are for relative spawning biomass (relB), harvest rate (F) and relative catch (relC). 182
Figure 80. Summary results of 100 realisations for four decision rules: binary based on mean length (ml), binary based on cpue in weight (uw), classification tree 'h9.flat' (h9) and classification tree 'h4.flat (h4). The extension '1' or '3' indicates whether $F_{max}=1.F_{spr40}$, or $F_{max}=3.F_{spr40}$. Results for a constant harvest rate $F=F_{spr40}$ (cf) is plotted first for comparison. Panels are as in figures above, with averages taken over the period that the decision rule was operational..... 183
Figure 81. The percentage time that relative biomass is below 30% of unexploited under the constant harvest rate F=Fspr40 (cf) and for the binary rule based on mean length with Fmax=Fspr40 (ml1) or Fmax=3.Fspr40 (ml3) in the left panel. The right panel is again for constant F=Fspr40, and for the 'up-down' rule based on mean length with Fmax=Fspr40 (mlx1) or Fmax=3.Fspr40 (mlx3). Note the difference in scale on the y-axis for the two panels. 184
Figure 82. Summary results for traffic light rules triggered by 3 'red lights'. The first two boxes are for the traffic light rule with the h9 set of indicators and low F_{max} (L91) or high F_{max} (L93). The last two

boxes are for the traffic light rule with the h4 set of indicators and low F_{max} (L41) or high F_{max} (L43). 185
Figure 83. Results for the traffic light decision rule with h9 set of indicators, low Fmax, and different settings for the number of red lights that trigger action: 1 red light (L91.1); 3 red lights (L91.3) or 4 red lights (L91.4). The classification tree (h9.flat) is shown for comparison..... 186
Figure 84. Results for the 20-40 decision rule with mean length as indicator (ml) or CPUE (u) as indicator. The constant harvest rate (Fspr40=0.1) is shown first, followed by the pair (ml24 and u24) with Fmax=Fspr40. The last pair (ml24f3, u24f3) is for Fmax=3.Fspr40. The performance statistics are as in the figures above. 187
Figure 85. The 20-40 rule with mean length (solid line) and 'up-down with maximum' rule with mean length (dashed line); both rules have Fmax=Fspr40. 188
Figure 86. The 20-40 rule with mean length (solid line) and the 10-40 rule with mean length (dashed line); both rules have Fmax=Fspr40. 188
Figure 87. Results for the 20-40 ('24') and 10-40 ('14') decision rules with mean length as indicator (ml) or CPUE (u) as indicator. The constant harvest rate (Fspr40=0.1) is shown first, followed by the pair based on mean length (ml24 and ml14) and then the pair based on CPUE (u24 and u14). All four decision rules have Fmax=Fspr40. The performance statistics are as in the figures above. 189
Figure 88. Results for 10-40 ('14') decision rules with mean length as indicator (ml), CPUE (u) as indicator or a set of indicators (A and B; see main text) combined using a mean statistic (A14 and B14) or a minimum statistic (A14min). The constant harvest rate (Fspr40=0.1) is shown for comparison. All five decision rules have Fmax=Fspr40. The performance statistics are as in the figures above..... 190
Figure 89. Prediction errors for a range of steepness levels with high noise (HHH), low m, FLAT selectivity and SPR triggers. Indicators are sorted by the prediction errors for h=0.5 (triangles). .. 191
Figure 90. Results for 10-40 ('14') decision rules with mean length (ml), CPUE (u), set A and CPUE of small fish (us) as indicators. The constant harvest rate (Fspr40=0.1) is shown for comparison. All decision rules have Fmax=Fspr40. The performance statistics are as in the figures above. 191
Figure 91. Comparisons of 'up-down with Fmax' and 10-40 rules under high steepness (h=0.9) and low steepness (h=0.4). Results are plotted in pairs for high, then low steepness, indicated by '.4' extension to the name. Rules are: 'up-down with Fmax=0.1 using mean length (mx1 and mx1.4); 10-40 rule with mean length (m14 and m14.4); 'up-down' with Fmax=0.1 using CPUE (ux1 and ux1.4); 10-40 rule with CPUE (u14 and u14.4). 192
Figure 92. Comparisons of binary and classification tree rules under high steepness (h=0.9) and low steepness (h=0.4). Results are plotted in pairs for high, then low steepness, indicated by '.4' extension to the name. Rules are: classification tree h9 (h9 and h9.4); binary rule using mean length (ml and ml.4); binary rule using CPUE (in weight) (uw and uw.4); classification tree h4 (h4 and h4.4). 193
Figure 93. Comparisons of traffic light and classification tree rules under high steepness (h=0.9) and low steepness (h=0.4). Results are plotted in pairs for high, then low steepness, indicated by '.4' extension to the name. The left section of each panel is for rules based on the h9 set of indicators, i.e. traffic light rule (4 trigger lights) (L9 and L9.4) and classification tree rule (h9 and h9.4). The right section of each panel is for rules based on the h4 set of indicators, i.e. traffic light rule (2 trigger lights) (L4 and L4.4) and classification tree rule (h4 and h4.4). 194
Figure 94. Comparisons of binary and classification tree rules under high steepness (h=0.9) and low steepness (h=0.4) when selectivity is dome-shaped. Results are plotted in pairs for high, then low steepness, indicated by '.4' extension to the name. Rules are: classification tree d9 (d9 and d9.4); binary rule using mean length (ml and ml.4); binary rule using CPUE (in weight) (uw and uw.4); classification tree d4 (d4 and d4.4). 194
Figure 95. True selectivity is flat, steepness is high and Fmax=0.1 in all cases. Binary rules and classification trees either assume the correct (flat) selectivity or the incorrect (dome-shaped, prefix 'd.' to name) selectivity and results are plotted in these 'correct – incorrect' pairs for each rule. The rules are: classification tree h9, binary rule based on mean length (ml), binary rule based on cpue (in weight, uw) and classification tree h4. Measurement error was high on all components (HHH scenario). 195
Figure 96. True selectivity is flat and true steepness is low; Fmax=0.1 in all cases. Binary rules and classification trees either assume the correct (flat) selectivity or the incorrect (dome-shaped, prefix 'd.' to name) selectivity and results are plotted in these 'correct – incorrect' pairs for each rule. The rules are: classification tree h9, binary rule based on mean length (ml), binary rule based on cpue (in

weight, uw) and classification tree h4. Measurement error was high on all components (HHH scenario). 196
Figure 97. True selectivity is flat. The first 3 boxes are for true h=0.9, the second three are for true h=0.4 (prefix '4' to scenario name). In all cases it is the binary decision rule using CPUE. Assumed selectivity and Fmax for decision rules are: flat, Fmax=0.10 (uw, 4uw); dome-shaped, Fmax=0.10 (duw, 4duw); dome-shaped with Fmax=0.18 (duwF, 4duwF)..... 196
Figure 98. Classification trees fitted to simulated data based on high steepness (h9,dome) and on low steepness (h4,dome), for the long-lived life history with DOME-shaped selectivity, high measurement error levels and SPR40 reference points. Repeated from Section 9; see text for more detail. 197
Figure 99. True selectivity is dome-shaped. Binary rules and classification trees either assume the correct (dome-shaped) selectivity or the incorrect (flat, prefix 'f.' to name) selectivity and results are plotted in these 'correct – incorrect' pairs for each rule. The rules are: classification tree dh9, binary rule based on mean length (ml), binary rule based on cpue (in weight, uw) and classification tree dh4. Measurement error was high on all components (HHH scenario). 198
Figure 100. True selectivity is dome-shaped and steepness is very low (h=0.4). Binary rules and classification trees either assume the correct (dome-shaped) selectivity or the incorrect (flat, prefix 'f.' to name) selectivity and results are plotted in these 'correct – incorrect' pairs for each rule. The rules are: classification tree dh9, binary rule based on mean length (ml), binary rule based on cpue (in weight, uw) and classification tree dh4. Measurement error was high on all components (HHH scenario). 198
Figure 101. True selectivity is dome-shaped. The first 3 boxes are for true h=0.9, the second three are for true h=0.4 (prefix '4' to scenario name). In all cases it is the binary decision rule using CPUE. Assumed selectivity and Fmax for decision rules are: dome-shaped, Fmax=0.18 (uw, 4uw); flat, Fmax=0.18 (fuw, 4fuw); flat with Fmax=0.10 (fuwF, 4fuwF). 199
Figure 102. Performance of the dh9 classification tree rule compared to the binary rule with mean length (dml), binary rule with CPUE in weight (duw) and the dh4 classification tree. All the decision rules assume dome-shaped selectivity and true selectivity is dome-shaped. 199
Figure 103. Classification trees fitted to simulated data based on F varying only slightly around a constant value (dh9,F=constant), and F varying widely according to a sine wave (h9,dome=dh9, F=sine). Both are based on the long-lived life history, dome-shaped selectivity, high measurement error levels and SPR40 reference points..... 200
Figure 104. As Figure 102, but including the decision rule based on the new classification tree, dh9.C, fitted to data simulated under approximately constant F. The rules are: the dh9 classification tree rule (based on F=sine), the dh9.C classification tree rule (based on F=constant), the binary rule with mean length (dml), binary rule with CPUE in weight (duw) and the dh4 classification tree. 201
Figure 105. Time series plots for simulations with and without effort creep. The top row of panels are for the binary decision rule based on CPUE. Panels are relative SSB, the F that is generated by the decision rule, or assumed to be operating, and the F (F.imp) that is actually implemented; the solid line is for no 202
Figure 106. Summary results of six performance measures and for decision rules without and with effort creep, plotted in pairs; the version with effort creep has a 'cr' extension to the name. Decision rules are: binary rule based on CPUE in weight (uw and uwcr); binary rule based on mean length (ml and mlcr); classification tree h9 (h9 and h9cr) and traffic light rule based on indicator set h9 with 3 red lights triggering action. The harvest rates are those from the decision rule rather than the implemented Fs. 203
Figure 107. Summary results of six performance measures and for decision rules without and with effort creep, plotted in pairs; the version with effort creep has a 'cr' extension to the name. Decision rules are: classification tree h4 (h4 and h4cr) and traffic light rule based on indicator set h4 with 2 red lights triggering action. The harvest rates are those from the decision rule rather than the implemented Fs. 204
Figure 108. Summary results of the decision-rule harvest rates and implemented harvest rates under 6 decision rules and in the presence of effort creep (hence the 'cr' extension in the name). The rules are: binary rule based on CPUE (in weight, uwcr); binary rule based on mean length (mlcr); the classification tree h9 (h9cr); the traffic light rule based on the h9 indicator set and 3 lights to trigger action; the classification tree h4 (h4cr); and the traffic light rule based on the h4 indicator set and 2 lights to trigger action. The left panel is the mean harvest rate generated by the decision rule; the

right panel is the mean implemented, or realised, harvest rate over the last 10 years of simulations. 204
Figure 109. Low steepness in true dynamics. Summary results of six performance measures and for decision rules without and with effort creep, plotted in pairs; the version with effort creep has a 'cr' extension to the name. Decision rules are: binary rule based on mean length (ml and mlcr); classification tree h9 (h9 and h9cr) and traffic light rule based on indicator set h9 with 4 red lights triggering action. The harvest rates are those from the decision rule rather than the implemented Fs. 205
Figure 110. Summary results of six performance measures and for decision rules without and with effort creep, plotted in pairs; the version with effort creep has a 'cr' extension to the name. Decision rules are: binary rule based on CPUE in weight (uw and uwcr); classification tree h4 (h4 and h4cr) and traffic light rule based on indicator set h4 with 2 red lights triggering action. The harvest rates are those from the decision rule rather than the implemented Fs. 206
Figure 111. The percentage time that spawning biomass is below 20% of the unexploited level for the low steepness scenario, and for decision rules without, or with effort creep (plotted in pairs, with 'cr' indicating with effort creep). The left panel is for: binary mean length rule (ml, mlcr), classification tree h9 (h9, h9cr), and traffic light rule based on h9 set of indicators (L9,L9cr) and triggered by 4 red lights. The right panel is for: binary cpue rule (uw, uwcr), classification tree h4 (h4, h4cr) and traffic light rule based on the h4 set of indicators (L4, L4cr) triggered by 2 red lights. 206
Figure 112. Stability criterion, $\mu = \lambda+1$ for: top row=example 1 with lines (upper to lower) coinciding with α= (0.05,0.5, 1, 2) respectively and for 3 values or rate of increase, r; bottom row is the same but for example 2, and α= (0.5,5,20,50). In all cases K=100. Horizontal lines indicate the stability bounds at -1 and +1. 218
Figure 113. Stability criterion, $\mu = \lambda+1$ for example 3 with lines (upper to lower) coinciding with α= (0.05, 1, 2,4) respectively and for 4 values or rate of increase, r. In all cases K=100. Horizontal lines indicate the stability bounds at -1 and +1. 218
Figure 114. Time series of biomass following a small perturbation at an equilibrium value, for example 1 decision rule. The title line shows the parameter values for r, K, and α. Initial biomass (before the perturbation) and the sign of the perturbation is: top row, left to right, 20+, 20-, 80- ; middle row left to right, 20-, 80-, 50-; bottom row left to right 20+, 20-, 80-. The magnitude of the perturbation was 0.01 in all cases. 220
Figure 115. Stability criterion, $\mu = \lambda+1$ for: top row=example 4 with lines (upper to lower) coinciding with α= (0.001,0.01, 0.02) respectively and for 3 values or rate of increase, r; middle row is the same but for example 5, and α= (0.1,1,2) and the bottom row is for example 6 and α= (0.05,1,2,4). In all cases K=100. Horizontal lines indicate the stability bounds at -1 and +1; (in the bottom right panel the line for α= 1 coincides with the horizontal line at -1. 223
Figure 116. Stability criterion, $\mu = \lambda+1$ for the Fox model with catch-based DR at different values of r and with lines (upper to lower) coinciding with α= (0.5, 1,1.5, 2, 2.5) respectively. In all cases K=100. Horizontal lines indicate the stability bounds at -1 and +1. 224
Figure 117. Change in Biomass as a function of the level of biomass and Catch as a function of biomass for a r=0.2, K=100, $\widetilde{B}$ =50 and three values of α: 1=thin black line; 10=bold black line; 50=grey line. C_{max}=5 in all cases (i.e. MSY=rK/4). Equilibrium points lie where ΔB=delB=0 on the left-hand panel. 227
Figure 118. Change in biomass as a function of biomass level for r=0.2, K=100 and a range of values of Cmax and $\widetilde{B}$ as shown in the heading. In each panel, the lines are for α 1,10,50 (thin black, bold black, grey). Recall from Figure 6 that α=1 is not really sigmoid over the range of B-values. 227
Figure 119. Stability criterion as a function of any biomass level with symbols indicating where the equilibria are (i.e. where delB=0). The horizontal line as at the stability boundary 1; equilibrium points above that are unstable. In each panel, the lines/symbols are for α 1,10,50 (thin black/squares, bold black/dots, grey/triangles). 228
Figure 120. Bifurcation plots of biomass (B) as a function of $\widetilde{B}$ in the sigmoid decision rule based on B_t for different values of the maximum catch, Cmax. Other parameters are: r=1.0, K=100, α =50. . 229
Figure 121. Catch versus biomass for the last 200 (out of 300) iterations. Values of Cmax and $\widetilde{B}$ are given above each panel. Other parameters are: r=0.2,K=100 and α =50 as in Figure 123. 230

Figure 122. Catch versus biomass for the last 200 (out of 300) iterations. Values of Cmax and $\widetilde{B}$ are given above each panel. Other parameters are: r=1.0,K=100 and α =50 as in Figure 120. 230
Figure 123. Bifurcation plots of biomass (B) as a function of $\widetilde{B}$ in the sigmoid decision rule based on B_{t-1} instead of B_t for different values of the maximum catch, C_{max}. Other parameters are: r=0.2, K=100, α =50. 231
Figure 124. Bifurcation plots of biomass (B) as a function of $\widetilde{B}$ in the sigmoid decision rule based on B_{t-1} instead of B_t for different values of the maximum catch, C_{max}. Other parameters are: r=1.0, K=100, α =50. 231
Figure 125. Bifurcation plots as a function of $\widetilde{B}$ with a sigmoidal catch-based decision rule (left-hand panels) or a harvest rate based decision rule (right-hand panels), in both cases based on B_t. The values of α and r are shown above each panel. Catch-based rules have C_{max} = rK/4; harvest rate-based rules have F_{max}=r/2. 232
Figure 126. Bifurcation plots as a function of $\widetilde{B}$ with a sigmoidal catch-based decision rule (left-hand panels) or a harvest rate based decision rule (right-hand panels), in both cases based on B_{t-1} instead of B_t. The values of α and r are shown above each panel. Catch-based rules have C_{max} = rK/4; harvest rate-based rules have F_{max}=r/2. 233
Figure 127. Sigmoid decision rule curves which could represent (and are similar to) 10-40 or 20-40 decision rules. The versions with $\widetilde{B}$ =30 are closer to the 10-40 or 20-40 rules, but the versions with $\widetilde{B}$ =40 are meant to reflect the notion of precaution which is sometimes implemented via a rule which reduces catches at higher levels of B/B_0. 234
Figure 128. Bifurcation plots showing approximate maximum r-values (given in headers) which avoids limit cycles and complex behaviour for $\widetilde{B}$ <=40% of B_0 with α =25 in both the Schaefer and Fox stock-production models and for Catch-based rules (top row) and F-based rules (bottom row). ... 235
Figure 129. Schaefer model with sigmoid DR and 30% CV on process error. The top row is for a stable equilibrium and the bottom row for a limit cycle. The left-hand panels are time-series, the middle panels are phase plots of B_{t+1} versus B_t and the right hand panels are autocorrelation plots for a range of time lags. 237
Figure 130. The pairs of panels (top and bottom) illustrate the time-series and autocorrelation plots for: left-hand pair= 30%CV process error on a stable equilibrium; middle pair = limit cycle with no process error; right-hand pair the same limit cycle with 30%CV process error. The titles above the top panels give the parameters: r, α, $\widetilde{B}$, C_{max}, σ. In all examples K=100. 237
Figure 131: Male and female length-at-age for the 4 sets of swordfish growth curves from the literature 243
Figure 132. Stylised growth curves for sexually dimorphic male (solid line) and female (dashed line) growth, together with the growth curves reported by Young and Drake (2004) for swordfish from the west coast of Australia. 245
Figure 133. Schematic illustration of the simple decision rule based on mean length (relative to virgin mean length) 247
Figure 134. Expected values for indicators relative to their values at F=0 by growth curve for pooled genders. Absolute equilibrium spawner biomass is also plotted. Indicators are: proportion mature 249
Figure 135. Expected values for indicators relative to their values at F=0 by gender (male, female, pooled), for the Young and Drake (2004) growth curves. Absolute equilibrium spawner biomass is also plotted. Indicators are as in Figure 134. 250
Figure 136. Expected values for indicators relative to their values at F=0 by gender (male, female, pooled), for the Berkeley and Houde (1983) growth curves. Absolute equilibrium spawner biomass is also plotted. Indicators are as in Figure 134. 251
Figure 137 Expected values for indicators relative to their values at F=0 by gender (male, female, pooled), for the Mantiel (1996) growth curves. Absolute equilibrium spawner biomass is also plotted. Indicators are as in Figure 134. 252
Figure 138. Expected values for relative mean length: meanL(Fmsy)/meanL(F=0) by growth curve and gender. Other factors are steepness, m, and implicitly age at first capture noting that size at first capture is identical in all cases. Gender is indicated by: 1=Male, 2= Female, 3=Pooled. Y&D refers to the Young and Drake growth curves with age-at-maturity of 5 years (Y&D 5), or 9 years (Y&D 9). 253

Figure 139 Change in indicator relative to change in biomass, [1 – indicator(Fmsy)/indicator(F=0)]/[1-Bmsy/B0], by gender, for 11 size-based indicators: proportion mature (pm), proportion big (pb), mean length (ml), median length (dl), 95th percentile of length (95l), mean weight (mw), median weight (dw) and 95th percentile of weight (95w), mean length of big fish (lb), mean weight of big fish (wb); and two CPUE indicators: cpue in numbers (cn) and cpue in weight (cw)..... 254
Figure 140 Change in the 'proportion big' indicator relative to change in biomass, [1 – indicator(Fmsy)/indicator(F=0)]/[1-Bmsy/B0], by gender and growth curve: 1=divergent, 2=parallel, 3=intersecting, 4=Young and Drake with age-at-maturity=5, and 5=Young and Drake with age-at-maturity=9. 255
Figure 141. Percentage 'no-change errors' (from 100 runs) for ml=mean length, ul=95th percentile of length, mw=mean weight and uw=95th percentile of weight for 3 levels of sampling (1=good, 2=poor, 3= very poor) and 4 lengths of the time-series, all starting in year 1. The indicators are based on pooled data, the parallel growth curve was used with a recruitment CV of 40% and harvesting at Fmsy..... 256
Figure 142. Overall proportion of 'no change' errors for pooled data, over periods of 5,10,15 and 20 years, for harvesting at F_{msy} and for the 3 growth curves with good sampling (1, indicated by 'div','int','par') and very poor sampling (3, indicated by 'div3','int3','par3'). See text for more detail on sampling level. Indicators are: proportion big (pbig), mean length (meanl), median length (medl), upper 95th percentile of length (up95l), mean weight (meanw), median weight (medw), upper 95th percentile of weight (up95w), mean length of big fish (meanlbig) and mean weight of big fish (meanwbig)..... 257
Figure 143. As Figure 142, but separately for males and females..... 258
Figure 144. Mean squared error from GAM models for indicators by gender and pooled data. Other assumptions are: recruitment variability CV of 40%, F=Fmsy and poor sampling (level 2). Indicators are as in Figure 142. 259
Figure 145. Summary results for 100 replicates of the 7 scenarios (numbered on the x-axis; see text in sub-section 2.4 above for definitions). All quantities are relative to their initial values (unexploited stock), except for final catch which is relative to the MSY of the underlying growth curve, and last panel which is the number of years out of 30 that SSB is below 40% of unexploited SSB. 261
Figure 146. Trade-off plot in terms of spawning biomass (B/B0) and catch (Catch/MSY), averaged over the 100 replicates, for the 7 scenarios (see text). The right-hand points in each pair are for the decision rule with higher threshold (0.93), those on the left, for the decision rule with lower threshold (0.88). 261
Figure 147. The proportion of the stock by area (A1, A2, A3) for 2 levels of harvest rate: F=0.2 i.e. relatively low (closed symbols), and F=0.6 which is very high (open symbols). The horizontal axis indicates the proportion of the population assumed to be resident. (Legend: A1 F6 implies Area 1, F=0.6 etc.) 271
Figure 148. The proportion of the stock by area (A1, A2, A3) for 2 levels of harvest rate: F=0.2 i.e. relatively low (closed symbols), and F=0.6 which is very high (open symbols). The horizontal axis indicates the proportion of the population assumed to be resident. (Legend: A1 F6 implies Area 1, F=0.6 etc.). 272
Figure 149. Single, long time-series, realisations for 3 decision rules and a constant harvest rate (without a spatial indicator). All scenarios start with a harvest rate of $2.F_{msy}$ over the first 5 years. The decision rule, or constant harvest rate (F_{msy}) starts in year 6. The top panel is for the total (all areas) relative spawning biomass; the lower panel is the total catch (in weight; all areas) relative to C_{msy}..... 278
Figure 150. Four realisations of the decision tree rule with a lower target reference for CPUE. The scenario is for 50% residency, 'gradual' move-on criteria, and no spatial indicator was used. Spawning biomass relative to unexploited is plotted over time..... 278
Figure 151. Four different tunings, or parameter settings, for the decision tree, run without a spatial indicator and assumptions of 50% residency and 'gradual move-on criteria. The top panel shows spawning biomass relative to unexploited; the lower panel shows total catch in weight relative to MSY. 280
Figure 152. Single, long time-series, realisations of catch in weight (relative to C_{msy}) for the two 40/20 decision rules used with a spatial indicator which reduces the catch by 30% when it is triggered. All scenarios start with a harvest rate of $2.F_{msy}$ over the first 5 years. The decision rule starts in year 6. 281
Figure 153. Average relative biomass (over 10 realisations) for the 40/20 rule based on mean weight (left panel) and the decision tree (right panel), and for two residency levels and move-on criteria..... 282

Figure 154. Average catch times series for two 40/20 decision rules, one using mean weight, the other using CPUE, and the decision tree. The left panel is for 50% residency, the right panel for 90% residency. The move-on criteria are 'gradual' and there is no spatial indicator. 284
Figure 155. Average mean weight used in the 40/20 decision rule under the 'gradual' move-on criteria with no spatial indicator, and for 50% and 90% residency. 284
Figure 156. Time series of Mean weight from the 10 realisations in two scenarios: 50% residency (left panels) and 90% residency (right panels). In each case, the first 5 realisations are plotted in the top panel and the second 5 in the lower panel. The mean (over all 10) is plotted in bold, and the horizontal line is the target reference level. 285
Figure 157. Scenario: 50% residency, gradual move-on criteria. Boxplots of catch rates (minimum and average over years 20-30) in the inshore area, and catches (variability and total, averaged over the decision rule period) in all areas, over 10 realisations. The 'factor' number identifies the decision rule used, as indicated in the legend. Time series of the average relative biomass (all areas) over the 10 realisations is also shown. 286
Figure 158. Time series of relative biomass from individual realisations for the decision tree without (left panels) and with a spatial indicator (right panels). The scenario is 50% residency and gradual move-on criteria. In each case, the first 5 realisations are plotted in the top panel and the second 5 in the lower panel. The mean (over all 10) is plotted in bold. 287
Figure 159. Scenario: 50% residency, stronger move-on criteria. Boxplots of catch rates (minimum and average over years 20-30) in the inshore area, and catches (variability and total, averaged over the decision rule period) in all areas, over 10 realisations. The 'factor' number identifies the decision rule used, as indicated in the legend. Time series of the average relative biomass (all areas) over the 10 realisations is also shown. 288
Figure 160. Scenario: 90% residency, gradual move-on criteria. Boxplots of catch rates (minimum and average over years 20-30) in the inshore area, and catches (variability and total, averaged over the decision rule period) in all areas, over 10 realisations. The 'factor' number identifies the decision rule used, as indicated in the legend. Time series of the average relative biomass (all areas) over the 10 realisations is also shown. 289
Figure 161. Scenario: 90% residency, stronger move-on criteria. Boxplots of catch rates (minimum and average over years 20-30) in the inshore area, and catches (variability and total, averaged over the decision rule period) in all areas, over 10 realisations. The 'factor' number identifies the decision rule used, as indicated in the legend. Time series of the average relative biomass (all areas) over the 10 realisations is also shown. 290
Figure 162. Average CPUE in area 1 over the last 11 years of the decision rule, for scenarios with uncontrolled harvest outside. Panels are: 50% residency, gradual move-on (top left); 50% residency, stronger move-on (top right); 90% residency, gradual move-on (bottom left); and 90% residency, stronger move-on. Within each panel, the decision rules are as in Figure 161, i.e. in pairs for without and with the spatial indicator, for constant harvest rate (1,2), 40/20 mean weight (3,4), 40/20 CPUE (5,6) and the decision tree (7,8). 292
Figure 163. Scenario: 90% residency, gradual move-on criteria. Boxplots of catch rates (minimum and average over years 20-30) in the inshore area, and catches (variability and total, averaged over the decision rule period) in all areas, over 10 realisations. Results are plotted in pairs of scenarios without and with effort creep for the following decision rules: 40/20 CPUE rule without spatial indicator (1,2) and with spatial indicator (3,4); to the right of the vertical line are the decision tree rule without spatial indicator (5,6) and with spatial indicator (7,8). 293
Figure 164. Time series of average (10 realisations) relative biomass for the 40/20 CPUE decision rule without and with effort creep, and without and with a spatial indicator. Residency is 90% and the move-on criteria are gradual. 294
Figure 165. Time series of average (10 realisations) relative biomass for the decision tree rule without and with effort creep, and without and with a spatial indicator. Residency is 90% and the move-on criteria are gradual. Average trajectories should be interpreted with caution for this decision rule. 294
Figure 166. Time series of relative biomass from individual realisations for the 40/20 CPUE rule without effort creep (left panels) and with effort creep (right panels). The scenario is 90% residency and gradual move-on criteria, and no spatial indicator was used. In each case, the first 5 realisations are plotted in the top panel and the second 5 in the lower panel. The mean (over all 10) is plotted in bold. 295

Figure 167. Time series of relative biomass from individual realisations for the decision tree rule without effort creep (left panels) and with effort creep (right panels). The scenario is 90% residency and gradual move-on criteria, and no spatial indicator was used. In each case, the first 5 realisations are plotted in the top panel and the second 5 in the lower panel. The mean (over all 10) is plotted in bold. 296
Figure 168. Scenario: 90% residency, gradual move-on criteria. Boxplots of catch rates (minimum and average over years 20-30) in the inshore area, and catches (variability and total, averaged over the decision rule period) in all areas, over 10 realisations. Results are plotted in pairs of scenarios with Cmax set correctly and with Cmax set 50% too high for the following decision rules: 40/20 mean weight rule without spatial indicator (1,2) and with spatial indicator (3,4); to the right of the vertical line is the 40/20 CPUE rule without spatial indicator (5,6) and with spatial indicator (7,8). 297
Figure 169. Time series of average (10 realisations) relative biomass for the 40/20 mean weight decision rule with Cmax set correctly or incorrectly, and without and with a spatial indicator. Residency is 90% and the move-on criteria are gradual..... 298
Figure 170. Time series of average (10 realisations) relative biomass for the 40/20 CPUE decision rule with Cmax set correctly or incorrectly, and without and with a spatial indicator. Residency is 90% and the move-on criteria are gradual..... 298
Figure 171. Time series of average (10 realisations) relative biomass for the 40/20 mean weight decision rule with Cmax set too high, but its calculation based on the correct steepness (h=0.3), and without and with a spatial indicator (1,2); and Cmax set too high, and based on the incorrect steepness (h=0.65), without and with a spatial indicator (3,4). In all cases true steepness is 0.3. Residency is 90% and the move-on criteria are gradual..... 299
Figure 172. Time series of average (10 realisations) relative biomass for the 40/20 CPUE decision rule with Cmax set too high, but its calculation based on the correct steepness (h=0.3), and without and with a spatial indicator (1,2); and Cmax set too high, and based on the incorrect steepness (h=0.65), without and with a spatial indicator (3,4). In all cases true steepness is 0.3. Residency is 90% and the move-on criteria are gradual. 300
Figure 173. Time series of average (10 realisations) relative biomass for the decision tree rule with target CPUE based on the correct steepness (h=0.3), and without and with a spatial indicator (1,2); and target CPUE based on the incorrect steepness (h=0.65), and hence set too low, without and with a spatial indicator (3,4). In all cases 'true' steepness is 0.3. Residency is 90% and the move-on criteria are gradual..... 301
Figure 174. Time series of relative biomass from individual realisations for a scenario with low steepness (h=0.3), 90% residency and gradual move-on criteria. The decision tree rule has a CPUE target based on the correct (low) steepness and is used without a spatial indicator (left panels) and with a spatial indicator (right panels). In each case, the first 5 realisations are plotted in the top panel and the second 5 in the lower panel. The mean (over all 10) is plotted in bold. 301
Figure 175. Time series of relative biomass from individual realisations for a scenario with low steepness (h=0.3), 90% residency and gradual move-on criteria. The decision tree rule has a CPUE target based on the incorrect (high) steepness, h=0.65, and is used without a spatial indicator (left panels) and with a spatial indicator (right panels). In each case, the first 5 realisations are plotted in the top panel and the second 5 in the lower panel. The mean (over all 10) is plotted in bold..... 302
Figure A 18. Scenario: 50% residency, gradual move-on criteria, additional uncontrolled harvest. Boxplots of catch rates (minimum and average over years 20-30) in the inshore area, and catches (variability and total, averaged over the decision rule period) in all areas, and time series of the average relative biomass in all 'domestic' areas, and all areas (including the 'outside' area) over 10 realisations. The 'factor' number identifies the decision rule used, as indicated in the legend (factor numbers are 1:8 from top to bottom). 306
Figure A 19. Scenario: 50% residency, stronger move-on criteria, additional uncontrolled harvest. Boxplots of catch rates (minimum and average over years 20-30) in the inshore area, and catches (variability and total, averaged over the decision rule period) in all areas, and time series of the average relative biomass in all 'domestic' areas, and all areas (including the 'outside' area) over 10 realisations. The 'factor' number identifies the decision rule used, as indicated in the legend (factor numbers are 1:8 from top to bottom)..... 307
Figure A 20. Scenario: 90% residency, gradual move-on criteria, additional uncontrolled harvest. Boxplots of catch rates (minimum and average over years 20-30) in the inshore area, and catches (variability and total, averaged over the decision rule period) in all areas, and time series of the average relative

biomass in all 'domestic' areas, and all areas (including the 'outside' area) over 10 realisations. The 'factor' number identifies the decision rule used, as indicated in the legend (factor numbers are 1:8 from top to bottom). 308
Figure A 21. Scenario: 90% residency, stronger move-on criteria, additional uncontrolled harvest. Boxplots of catch rates (minimum and average over years 20-30) in the inshore area, and catches (variability and total, averaged over the decision rule period) in all areas, and time series of the average relative biomass in all 'domestic' areas, and all areas (including the 'outside' area) over 10 realisations. The 'factor' number identifies the decision rule used, as indicated in the legend (factor numbers are 1:8 from top to bottom)..... 309
Figure A 22. Effort Creep. Scenario: 90% residency, gradual move-on criteria, with spatial indicator and additional uncontrolled harvest. Boxplots of catch rates (minimum and average over years 20-30) in the inshore area, and catches (variability and total, averaged over the decision rule period) in all domestic areas, over 10 realisations. Results are plotted in pairs of scenarios without and with effort creep for the following decision rules: 40/20 CPUE rule with spatial indicator (1,2); the decision tree rule with spatial indicator (3,4). The bottom panel shows relative spawning biomass in all areas over 10 realisations..... 310
Figure A 23. C_{max} too high. Scenario: 90% residency, gradual move-on criteria, with spatial indicator and additional uncontrolled harvest. Boxplots of catch rates (minimum and average over years 20-30) in the inshore area, and catches (variability and total, averaged over the decision rule period) in domestic areas, over 10 realisations. Results are plotted in pairs of scenarios with [C_{max}=C_{msy}]and [C_{max}=1.5C_{msy}] for the following decision rules: 40/20 mean weight rule with spatial indicator (1,2); 40/20 CPUE rule with spatial indicator (3,4). The bottom panel shows relative spawning biomass in all areas over 10 realisations..... 311
Figure A1.1 Representation of the "40/10" style decision rule used for adjusting quota or fishing mortality..... 330
Figure A1.2 Flowchart diagram of the four levels of the decision tree..... 334

Please note: Figures in Attachment 1 (p22-47) and Attachment 2 (p48-70) are not included in this list.

1. Non Technical Summary

Project: 2003/042
Development of a robust suite of stock status indicators for the Southern and Western, and the Eastern Tuna and Billfish fisheries.

Principal Investigator: Dr Marinelle Basson
Address: CSIRO, Marine and Atmospheric Research
GPO Box 1538
Hobart, Tas 7000

Objectives

1. Design a candidate set of potential stock status indicators (SSI) which reflect a wide range of aspects of stock and fishery status, and develop appropriate standardisation procedures for the SSIs
2. Modify existing operating models to reflect fish population and fishery dynamics of each relevant case study with particular focus on stock structure uncertainty
3. Develop candidate frameworks and methodology for a management system based on a suite of indicators
4. Test the robustness of SSIs (individually and jointly in a framework) relative to arbitrary trigger or reference points without built-in decision rules
5. Develop meaningful reference points, in consultation with FAGs, and test best-performing SSIs with built-in decision rules
6. Provide Southern and Western Tuna MAC[1], Eastern Tuna MAC and AFMA (the Australian Fisheries and Management Authority) with an evaluation of the robustness of alternative SSIs in the context of management strategies.

Outcomes Achieved

The most important outcome from this project, jointly with the work conducted within the Harvest Strategy Working Group (HSWG) for the Eastern Tuna and Billfish Fishery (ETBF) and the Western Tuna and Billfish Fishery (WTBF)), is a harvest strategy framework for tunas and billfish. The HSWG developed a decision rule with associated reference and trigger points which has been adopted, thus meeting the fisheries' first set of requirements under the Ministerial Directive. The design of the decision rule and the use of several indicators, including size-based indicators, was substantially influenced by the work conducted and results obtained under this project. In particular, the design of the decision rule benefitted from:
(i) a much improved understanding of the responsiveness and performance of a wide range of size-based and CPUE-based stock status indicators under different assumptions about the population and harvesting dynamics of the stock;
(ii) a much improved understanding of the general dynamics and behaviour of different types of decision rule.
In addition, the simulation models developed under his project allowed for rapid preliminary

[1] At the start of the project, the fishery off the west coast was defined (within AFMA) as the 'Southern and Western Tuna and Billfish fishery' (SWTBF); this was later changed to the 'Western Tuna and Billfish fishery' (WTBF). The report uses both acronyms.

testing of the decision rules developed by the HSWG, and therefore evaluation of the choice of particular reference and trigger points. This would not have been feasible without the outputs from this project.

The general nature of the results and outputs from the study also implies broader applicability and relevance to a wider range of species and fisheries than the Eastern and Western tuna and billfish fisheries. This could lead to additional outcomes in future e.g. potential application in the Indian Ocean Tuna Commission (IOTC), following broader dissemination of project results.

Non Technical Summary

The aim of the project was to explore whether a fishery could be managed by using indicators to adjust total allowable catch or effort. The idea is to monitor indicators, such as catch per unit effort (CPUE), or mean length in the catch, and use them in a decision rule. For example, depending on whether an indicator was above or below its reference point, or whether it was increasing or decreasing, the total allowable catch or effort could be adjusted accordingly, using a pre-agreed rule. This approach would be of particular value for data poor fisheries, or in cases where there is no stock assessment.

A simulation approach was used to conduct the explorations. In addition to CPUE, eight size-based indicators were considered: mean, median and 90th percentile of length and weight, and the proportion of 'big' and of 'small' fish (Section 6). Simulated data area easily generated to be consistent over time, but in reality indicators are usually based of fisheries data and therefore susceptible to changes in fishing operation. Standardisation, to take account of such changes, is routine for CPUE data, but not for size frequency data. We therefore first illustrate the modelling approaches that should be used to standardise an indicator time-series, or explore whether standardisation is required, before using it as a management tool (Section 7).

How responsive are indicators to changes in population size? Under ideal circumstances, changes in CPUE are proportional to changes in fishable abundance. Size-based indicators change by far less, and are therefore much more sensitive to measurement error or random noise, e.g. due to recruitment variability (Section 8). Results strongly suggest that size-based indicators, in terms of length or weight, are only likely to be informative for populations where individual growth is slow. Of all the size-based indicators considered, the means (of length or weight) generally perform best. In a few cases, the upper 90th percentiles perform well, but they should be used with caution because they are more susceptible to sampling error, and they perform poorly when the harvest rate is approximately constant.

The performance of size-based indicators also depends on the relationship between the number of recruits (offspring) and the parent biomass (called the stock-recruit relationship). This relationship is notoriously difficult to quantify, and it is therefore risky to rely only on size-based indicators. Decision rules which use (unbiased) CPUE perform better because they show the least difference in performance measures (e.g. average and minimum biomass) under different assumptions about the stock-recruit relationship. This strongly underlines the importance of obtaining reliable catch-effort data, and additional data on fishing operations for standardisation purposes. Obtaining fishery independent indices of relative abundance are even more important to strive for.

In order to evaluate the performance of groups of indicators we fitted classification trees to simulated data, and used the trees as decision rules (Sections 9, 10). Results suggest that there is

very little to be gained by using more than about 4 to 5 indicators together; often as few as 2 or 3 were sufficient. The choice of indicators depends on whether the fish population is long- or short-lived, slow- or fast-growing. Decision rules based on groups of indicators did not generally outperform those based on the best single indicator in the group. At face value this suggests that there is little to gain from using multiple indicators, though we caution against over-interpreting this result. The benefit of multiple indicators may only really come into play when indicator time series are biased or prone to serious unpredictable 'failure' (e.g. changes in fishery operations which are not detected when an indicator is standardised). Therefore, we consider it prudent to not (yet) abandon the notion that multiple indicators may perform better than a single indicator in a real situation.

There are different ways in which indicators can be combined in a decision rule. There were no strong differences in overall performance between the classification tree and traffic light rules using the same set of indicators (Section 10). Some simulation results suggest that the firm structure of the classification tree can sometimes be a disadvantage, whereas the very free structure of the traffic light rule could be an advantage.

We note that, when data are available to construct several different indicators, those data may be sufficient to use in a statistical framework (e.g. of a stock assessment model). Even a relatively simple model can be a powerful tool for integrating the data, and this should not be overlooked.

Even good indicators can perform poorly when used in a decision rule that is badly designed. A mathematical analysis (called 'local stability analysis') was used to explore the characteristics of a decision rule in relatively simple cases (Section 11). We suggest that this is a very useful approach for eliminating rules with poor characteristics at an early stage. Results for ten example decision rules highlight the importance of building a 'target' into a rule rather than just having a rule which responds to an increase or decrease in an indicator. Results also show that decision rules which adjust catches behave quite differently from those that are defined in a similar way but adjust effort. They cannot simply be used interchangeably. Care should be taken when choosing the limit and target reference points in a '20-40' type rule[2]. If these points are too close together, the decision rule may lead to highly variable catches which are likely to induce large cycles in the population abundance.

Swordfish displays particularly strong differences in growth patterns between males and females. A brief comparison of gender-specific and pooled size-based indicators did not suggest a high priority for looking at gender-specific size frequency data when constructing indicators (Section 12). Exactly how the male and female curves differ, does, however, seem important. Although results suggest that it is not essential to collect gender-specific size data, there are other important reasons why size data should be collected by gender if at all possible. One important reason is that size-based indicators are likely to be sensitive to changes in the sex ratio of the catch when there are differences between growth of males and females. This information would be needed to standardise size-based indicators.

We used a general spatial model to explore the effects of uncertainties about stock structure and fish and fleet movement on the performance of decision rules (Section 13). It is not surprising that the advantages of management are compromised by additional uncontrolled harvesting (e.g. by international fleets) outside the management area. This issue will be very important when a

[2] Such a rule has a minimum catch (often zero) when the indicator is below its limit reference point, a maximum catch when the indicator is at or above its target reference point and a linear increase in catch between the minimum and maximum for values of the indicator inbetween its limit and target reference points. See Appendix 1, Figure A1.1.

decision rule is being tuned to achieve a specific objective, and should not be ignored even when a non-spatial model is being used.

If spatial management is envisaged, the potential for 'missing data' (due to no fishing in a given area) should be considered, because it can affect the calculation/standardisation of an indicator and hence the performance of a decision rule. The inclusion of a spatial indicator (e.g. how far offshore the fleet is fishing) in a decision rule, can be useful, though for most scenarios in this report, a spatial indicator was arguably not needed. A spatial indicator could be used to 'trigger' a re-distribution of effort, rather than a cut in catches as modelled here. This does not have to imply the closure of any areas, so avoiding issues of 'missing data'.

The performance of decision rules may also be sensitive to the way in which the fleet dynamics are modelled. Experience from this study suggests it may be better to use a case-specific, data driven model, particularly for the fleet dynamics, to explore the behaviour of decision rules in a spatial context. There is still a need to explore the uncertainties in the model and inputs, but this approach should help focus on the most realistic or relevant subset of scenarios.

KEYWORDS: stock status indicator, decision rule, classification tree, harvest strategy evaluation, fisheries management

2. Acknowledgements

We are grateful to the Secretariat of the IOTC for assistance with obtaining and interpreting size frequency data, for the tropical tunas and broadbill swordfish, from the IOTC database. We acknowledge the valuable interactions with members of the Harvest Strategy Working Group, as well as members of the WTBF RAG and MAC, and the ETBF RAG and MAC. Thanks too for useful discussions with many colleagues at CSIRO, particularly with Dale Kolody, Robert Campbell, Jock Young, Chris Wilcox and Tom Polacheck at various stages of the project. Paavo Jumppanen provided valuable software assistance at crucial times. The project was funded by the FRDC grant 2003/042 and by the CSIRO.

3. Background and Need

This project was proposed in 2002, and at that stage the substantial expansion in the longline sector of both the Eastern and the Southern and Western Tuna and Billfish fisheries (ETBF and SWTBF) in the preceding years, as well as AFMA's requirement for formal management plans, brought about a need for well-designed assessment and management systems for these two fisheries. The ETBF had opted for input control via total allowable effort (TAE), whereas the SWTBF opted for output control via total allowable catch (TAC). When the project was proposed, a start had been made in both fisheries to set initial TACs or TAEs for each of the key target species (bigeye and yellowfin tuna, and broadbill swordfish), and to consider definitions of decision rules for changing those TAC/TAEs.

The SWTBF fishery forms a small part of the fisheries on yellowfin and bigeye tuna, and broadbill swordfish in the broader Indian Ocean. The stock structure for all three species is, however, still broadly unknown, or at best highly uncertain. All three species fall under the remit of the Indian Ocean Tuna Commission (IOTC). In theory, the IOTC (through member scientists) should conduct stock assessments and provide advice to the Commission on stock status and on sustainable catch or effort levels. In practice though, most stock assessments (e.g.

for bigeye, yellowfin tuna) are highly uncertain, and difficult because of data issues. In other cases there is no stock assessment (e.g. broadbill swordfish[3]). In addition, and possibly to some extent because of the uncertainties in the stock assessment, the IOTC Commission has not set any catch quotas yet. At the time of completion of this project (2007), some resolutions to limit effort have at least been agreed by the Commission (IOTC Commission, 2007). The IOTC Scientific Committee (SC) has been recommending catch reductions and/or limits on effort for all three species over several years. The SC has also recommended that current catches for bigeye are likely to be above the maximum sustainable yield (Reports of the IOTC Working Party on Tropical Tunas and Working Party on Billfish, and of the Scientific Committee, 2001, 2002 and subsequent reports). The research proposed in this project was identified as a high priority task at IOTC Working party meetings particularly given the difficulties associated with (or absence of) stock assessments.

The objective of this project was to develop a robust suite of stock status indicators for tuna and billfish-like stocks, particularly for potential use as management tools in the Western Tuna and billfish fishery. The idea was, however, to look at the issues in a more general way so that results would be relevant to the Eastern tuna and billfish fishery, as well as (potentially) for other species and in the international context of regional fisheries management organisations (RFMO's) such as the IOTC.

The proposal was also seen as a first step towards developing harvest strategies, not just for the WTBF, but also within the IOTC. At the time, informal discussions between members and the IOTC secretariat confirmed interest (at least within the scientific community) to start developing the tools for doing harvest strategy evaluation with an aim to develop harvest strategies (decision rules). The idea was that, even where there are no stock assessments, indicators could be used to define a decision rule which would adjust the catch (or effort) levels in the fishery. Recall that this was in 2002, several years before the Australian Ministerial Directive (see Australian Government, 2007) which called for the development of harvest strategies for all Commonwealth managed fisheries.

At the time this project was proposed, significant progress was being made with existing projects focusing on the Eastern Tuna and Billfish fisheries (ETBF), but the need to develop and test suites of stock indicators with associated reference points and decision rules for use by the Resource Assessment Groups (RAGs) for each fishery remained, particularly for the Southern and Western region (SWTBF). The need was also urgent, given the imminent introduction of TAE/TAC-based management plans and the need to formally evaluate them on an annual basis.

Within the SWTBF, the project 'Review and analysis of information required for the determination of TACs and decision rules relevant to the Southern and Western Tuna and Billfish Fishery' was providing scientific advice to AFMA on these issues. However, because of the short time frame of this project it could only identify performance measures and indicators and suggest preliminary decision rules for the adjustment of TACs, based on outcomes from previous studies conducted on the Eastern Tuna and Billfish fishery (e.g. Punt *et al.* 1999). Also, for practical reasons, the outcomes from that project could only focus on performance measures and indicators that were already available, or indicators that could relatively quickly and easily be implemented. The key aspect of testing whether the identified indicators would be robust to underlying uncertainties, and whether particular decision rules were likely to be effective, were not covered.

[3] At the time the proposal was being developed there was no swordfish assessment in the Indian Ocean. There has since been initial attempts to fit stock production models to the data, with some success.

Within the ETBF, the project 'Development of an operating model and evaluation of harvest strategies for the ETBF' (Campbell and Dowling, 2003) focused on the issue of decision rules with regard to swordfish in the SW Pacific. The project was, however, primarily focusing on the use of catch per unit effort (CPUE), either as an empirical indicator, or in a production model assessment. The testing of a wider range of additional indicators (e.g. spatial indices) was not included, and neither were the other target species (bigeye and yellowfin tuna). The project to develop robust stock status indicators was therefore proposed as a much more extensive project than the SWTBF project aimed at swordfish.

An 'indicator' could be any quantity that is expected to reflect something of the population (or fishery) status, for example, catch per unit effort (CPUE) or the mean length of fish in the catch. Given the potential difficulties and limitations of using only CPUE as an indicator of stock status there was a need to also consider other indicators, e.g. ones which might reflect local depletion, an issue which was of real concern for swordfish. Standardisation of such indicators also needed to be considered. Different indicators (e.g. CPUE or mean length in the catch) reflect different aspects of the population dynamics, and there was increasing recognition of the value of considering a suite of indicators rather than relying on a single one. When the project was proposed there were only limited and relatively unsophisticated frameworks for combining information from several indicators. There was a need to develop this further, and to design defensible and robust frameworks to use in management decision-making. There was also a need to test the robustness of suites of indicators within the context of a feed-back management loop rather than simply in a non-feedback sense.

4. Objectives

1. Design a candidate set of potential stock status indicators (SSI) which reflect a wide range of aspects of stock and fishery status, and develop appropriate standardisation procedures for the SSIs
2. Modify existing operating models to reflect fish population and fishery dynamics of each relevant case study with particular focus on stock structure uncertainty
3. Develop candidate frameworks and methodology for a management system based on a suite of indicators
4. Test the robustness of SSIs (individually and jointly in a framework) relative to arbitrary trigger or reference points without built-in decision rules
5. Develop meaningful reference points, in consultation with FAGs, and test best-performing SSIs with built-in decision rules
6. Provide Southern and Western Tuna MAC, Eastern Tuna MAC and AFMA with an evaluation of the robustness of alternative SSIs in the context of management strategies.

5. Outline of Research

Evolution of the project and external drivers

The Ministerial Directive (Australian Government, 2007), which called for the development of harvest strategies for all Commonwealth managed fisheries within a very tight timeline, had a direct impact on the project. First, the need for urgent and unplanned additional work meant that direct work on this project was temporarily reduced while staff time was redirected to the more direct work on the development of harvest strategies for the ETBF and WTBF. This was done through a Harvest Strategy Working Group (HSWG), set up through AFMA with members from the ETBF and WTBF RAGs. At the same time, however, the work to date on this project

was directly fed into the HSWG workshops, both formally through presentations, and informally through the participation of the project team (Marinelle Basson and Natalie Dowling) in the HSWG. Second, many of the concepts expressed in the Ministerial Directive and associated draft policy documents reinforced the relevance of the approaches being taken in this project.

Third, the decision rule developed by the HSWG was evaluated using the software developed under this project, and was used in some of the simulation trials in this project (see Section 13). Finally, part of the fifth objective (development of reference points) was essentially taken on by the HSWG, driven and guided by the HS policy associated with the ministerial directive. This project therefore shifted the focus of that objective to the evaluation of decision rules, though we had already done some work towards the definition of reference points[4]. Overall, we consider that the impact of the Ministerial Directive was positive, and strengthened the outcomes from this project. The uptake of results and relevance of this project should be seen in the context of the additional work done through the HSWG.

It is worth briefly commenting on the relative timing of the work of the HSWG and the chapters in this project. For example, at the time that the HSWG started its work we had not completed the evaluation of groups of indicators. The 'ad-hoc' (hierarchical) decision tree suggested by Jeremy Prince, and developed in the HSWG inspired us to fit classification trees to simulated data as a tool for evaluating groups of indicators. Also, much of the work on indicators and decision rules in feedback context was underway, but had not been completed. This work also therefore proceeded in parallel with that of the HSWG. This is important to bear in mind when reading this project report as well as the HSWG reports.

With respect to interaction in the IOTC forum, the rapid decline in the number of active vessels the WTBF fishery, combined with other commitments meant that members of this project team did not attend as many of the IOTC Working Party meetings as originally intended. Nonetheless, several working papers were submitted to the IOTC WPs as working papers (including progress reports which have not been included here to avoid duplication). The firm intention is also to publish much of the work in this report as peer reviewed papers, and draw the attention of the IOTC Working Parties to this work, by submitting publications as information papers to their meetings.

Organisation of project methods and results

The research presented in this report does not lend itself to being organised into a single 'methods' section and single 'results' section. Instead, the various components of the work have been organised in Sections (which can be thought of as Chapters), each with its own sub-section on methods, results, and discussion. Throughout we refer to the main sections as "Section" (i.e. Sections 6 through 13), and numbered sections within are referred to as "sub-sections". Also, because of the size of the report, each of the main Sections has a short Summary at the start to assist the reader.

In this project we define and evaluate the suitability of a wide range of potential indicators, mostly based on the size distribution of the catch, but also including the familiar CPUE-based indicators. **Section 6** (Candidate stock status indicators) lists a wide range of potential indicators and associated literature. Note that here we only consider single-species indicators.

[4] This work is reported in Appendix 2 rather than as main chapter, because it has essentially been superseded by the work of the HSWG and the HS Policy document (Australian Government, 2007).

The need to standardise CPUE before interpreting it as an indicator of relative abundance, or using it in a stock assessment, is well known in fisheries science. The process of standardisation is meant to remove 'false' signals due to, for example, a change in the timing and/or location of fishing, or a change in the setting or nature of the gear. Other indicators of stock status, such as mean length of the catch, could also be subject to the same – or other – factors, and should ideally also first be standardised. **Section 7** (Standardisation of size-based indicators: three case studies) is a collation of two papers submitted to IOTC Working party meetings, considering the standardisation of size-based indicators for bigeye and yellowfin tuna, and for swordfish. The aim of this exercise was primarily to illustrate how one might approach the standardisation of non-CPUE indicators, and to explore whether the standardisation would strongly affect the time-trends of the indicators for the particularly international longline fisheries in the IOTC context. (We did not attempt to do this exercise for the WTBF, or ETBF because of short time-series and the lack of a direct link between the location of fishing and each of the size frequency samples).

The third, fourth and fifth chapters are closely related, and deal with the issues of 'robustness' in a relatively general way. **Section 8** (An investigation of the responsiveness of CPUE- and size-based indicators) considers how responsive the different indicators are likely to be, depending on the life-history strategy (i.e. growth rate, natural mortality and stock-recruit dynamics) of the harvested species, and the nature of the fishery, particularly its selectivity pattern. Here we also look at the effect of recruitment variability on time-series of indicators.

Section 9 (The performance of single indicators and groups of indicators as predictors of relative spawning biomass) looks at the performance of single indicators by considering them as 'predictors' of spawning stock biomass status, i.e. whether SSB is above or below a given reference point, depending on whether the indicator itself is above or below a given 'trigger' point. Performance is quantified in terms of a prediction error and these quantities are compared between indicators and across a range of scenarios. The scenarios reflect differences in fishery selectivity, life-history of the stock, and process and measurement error. An important part of this project is to evaluate the performance of groups of indicators used together. Although the approach used for single indicators can be expanded to groups of indicators, this quickly becomes unwieldy because of the large numbers of possible combinations. We therefore use classification trees to form specific combinations of indicators. The trees can be pruned to include a smaller number of indicators (or decision nodes), and prediction errors are easily obtained for the trees.

The final step in this general investigation is to use single or groups of indicators in feedback decision rules, and this is discussed in **Section 10** (An evaluation of the performance of single, and groups of, indicators in feedback decision rules). There are many potential ways of constructing a decision rule for adjusting the catch (or effort), even from a single indicator. The aim is not to find an optimal decision rule, but rather to compare the behaviour of different decision rules, based on different indicators (or groups of indicators) under a range of scenarios. In particular, we focus on those quantities which are likely to be unknown in reality, e.g. the stock-recruit relationship, and investigate whether some indicators, or types of decision rule, perform better than others.

In the course of exploring the behaviour of different decision rules, we noticed that the simulated population dynamics very often displayed a cyclic behaviour after implementation of the decision rule. This was not simply due to recruitment variability. As far as we are aware, the development of decision rules has been relatively ad-hoc (albeit based on common sense), and primarily within a simulation context, so that the theoretical characteristics of such systems have not been explored. **Section 11** (Feedback dynamics of simple catch and harvest rate decision rules: a theoretical investigation) considers the likely behaviour of different types of decision rules from a

theoretical point of view, with the aim to raise general awareness of the strange behaviour that could be associated with some forms of decision rule, and to guide the design of decision rules.

The analyses referred to so far was done either in a very general context of different life histories, or in the context of a swordfish-like (long-lived) and a yellowfin-like (short-lived) species, but not necessarily with closely matching parameter values for each species. The reason is that it is more the general behaviour of indicators that we are concerned about than the case-specific behaviour. However, given the importance of swordfish in the WTBF (and the ETBF), particularly at the time the proposals was being developed, we considered two issues that are very specific to, and potentially very relevant to, swordfish.

The first issue is sexual dimorphism in growth. Swordfish does appear to display particularly strong sexual dimorphism in growth, with females generally growing larger than males. It is, however, currently very difficult to collect information about gender from the catch, and hence to get separate male and female size frequency distributions. It is obvious to ask whether size-based indicators need to be gender-specific, or whether a pooled indicator could perform as well, or at least adequately. This is discussed in **Section 12** (Implications of sexual dimorphism in swordfish growth for size-based indicators).

The second issue is local depletion. In the ETBF there has been some experience of local depletion of swordfish (Campbell and Hobday, 2003). The fishery started relatively close inshore, but following a drop in CPUE in this region, the fleet moved further offshore. This was again followed by a drop in CPUE and a further offshore movement of the fleet. Such an effect may, or may not, imply that the overall stock is overfished, or being overfished, but even if the stock as a whole is at a sustainable level, the local depletion can have serious economic implications for the fishery. The question then arises whether relatively simple 'spatial' indicators could be used to avoid local depletion, and how this interacts with different assumptions about the level of residency of fish and harvest rates outside the Australian fishery. **Section 13** (Swordfish: The performance of decision rules in the presence of fish and fleet movement) considers these spatial issues in the context of swordfish.

6. Candidate Stock Status Indicators

Summary

This Section addresses the first objective of the project, namely to define a large range of potential stock status indicators for tuna and billfish fisheries. A very broad set of indicators, some very familiar in fisheries science, and other less familiar ones taken from the literature, is identified. A subset of size-based indicators, in terms of length, weight, or proportions, as well as a set of CPUE-based indicators are chosen for further analyses. The choices are based both on the likely availability of the necessary data, and on the feasibility of generating such data in a simulation context.

In addition to indicators which reflect an aspect of the population (or catch), we also consider a group of indicators meant to reflect aspects of the spatial distribution of the fleet. The relevance of this is for use in a spatially explicit model aimed at exploring the issue of local depletion.

6.1. Introduction

What are stock status indicators?

The most familiar indicator of stock status is catch per unit effort (CPUE). There are, however, many potential problems with CPUE which could make it unreliable or, at worst, useless. The most obvious issues include:

effort creep: increased efficiency due to learning, improved technology and/or changes in operation. This is usually very difficult to separate from the 'signal' which reflects relative abundance. If there is effort creep, the CPUE could be increasing when the stock is in fact declining.

targeting: changes in targeting practices, which could be subtle and difficult to record or quantify, can affect CPUE and provide a false signal

gear type: purse seine CPUE, particularly fishing based on FADS (fish aggregation devices), is particularly difficult to interpret and standardise

It is well-known that there are other quantities associated with the catch (and population) that change as a stock is harvested. For example, at the start of a fishery, there may be a large number of very big fish in the catch. As the fishery continues there is likely to be fewer and fewer very big fish. This will be reflected in quantities such as the mean length of the catch, or the 90^{th} percentile (say) of the catch size frequency distribution. Unfortunately, most size-based indicators can be sensitive to recruitment. For example a decrease in mean size in the catch could either be due to a reduction in the number of large fish in the population, or due to an increase in the number of small fish (or recruits). Indicators based on proportions, for example the proportion big fish in the catch, also suffer from this problem.

Where a stock assessment is available, the key quantities estimated in the stock assessment, and considered when evaluating the status of a harvested stock are:

Spawning biomass
Recruitment
Age or size structure of the population (in some assessments)

Often these quantities are expressed relative to their expected (or estimated) average levels for a stock that is not harvested. The notions behind these quantities are relatively obvious. First, there needs to be 'enough' adult, reproducing individuals. Second, the adults need to be producing 'enough' recruits to replace the removals due to natural deaths and fishing, and third, there needs to (ideally) be a spread of numbers over several age classes. This is clearly more relevant for longer-lived than short-lived stocks.

The idea of using stock status indicators is based around the above concepts of a 'healthy' stock. Where a stock assessment is not available, simplified versions of the three quantities mentioned above, could be:

the proportion mature fish in the catch compared to the proportion in an unexploited stock;
the proportion recruits in the catch compared to the proportion in an unexploited stock;
the average size of fish in the catch compared to the average size in an unexploited stock.

Two points should, however, be noted. First, ideally, one would want the above quantities for the population, not just the catch, but it is only the latter that is likely to be easily obtainable. This is particularly true if only a few age classes are taken in the catch, or if the catch only consists of immature fish. Second, these indicators are based on the size-frequency (or weight frequency) of the catch. This is deliberate; the concept of 'stock status indicators' is specifically meant to be broader than just catch per unit effort (CPUE).

Identifying potential indicators for further investigation

One of the objectives of this study (see Section 4. Objectives) is to identify a wide, but realistic, range of potential indicators of stock status: realistic from the point of view of calculating the indicator and realistic from the point of view of testing its likely performance in a simulation context. The list of candidate indicators, identified and selected during the first phase of this study, exclude ones that cannot easily be tested in simulation, irrespective of whether the data are currently available for the relevant fisheries or not. It is beyond the scope of this study to model the economics or decisions driving fleet dynamics, for example, so the choice of indicators to evaluate, need to be compatible with this constraint. We also emphasise that this study focuses on single species, and indicators that reflect 'ecosystem' status have therefore not been considered either.

Recall that indicators are being considered for situations where there are no direct measures of relative (or absolute) abundance, such as surveys for example, and where assessments which could provide estimates of abundance are not (yet) feasible or reliable. In general, many indicators are primarily indirect measures of total mortality on the stock. For example, simple population dynamics theory shows that harvesting is likely to lead to a decrease in mean size in the population over time due to the increased mortality. It is also possible to infer the likely stock, or spawning, biomass relative to the unexploited stock, or spawning, biomass for a given level of total mortality. The first group of indicators (Table 1) are fairly standard indicators of this kind.

In cases where a fishery is concentrated in particular areas, the response of the local fish population could be different from the overall average. Also, if a fishery then moves to a previously unexploited area, and this is not recognised, changes in an indicator could be misinterpreted. The second group of indicators (Table 2) are aimed at detecting and monitoring spatial changes in the fishery and/or stock distribution, both to identify situations of local (or serial) depletion and to help with the interpretation of other indicators.

At this stage we merely summarise quantities that could be used as indicators; we do not fully discuss potential problems of interpretation or implications of wrong assumptions. Neither do we discuss here exactly how an indicator might be used; this is explored in Sections 8 and 9. In brief, the idea is to calculate an indicator annually (or more, or less, frequently) and monitor the time-series for rapid changes or trends. In Section 8 the indices are used as 'predictors' of whether spawning biomass is above or below a specified reference point, depending on whether the indicator is above or below a specified 'trigger' point. In Section 9 indicators are used to construct decision rules which, depending on the value of the indicator, trigger management action, such as an increase or reduction in total allowable catch.

6.2. Candidate indicators of stock status or harvest rate

It is impossible to harvest a stock and yet keep it at its unharvested level, or keep its age structure or size structure the same as it is when unharvested. By looking at the value of a 'stock status indicator', the hope is that this quantity would indirectly reflect the changes in the population due to harvesting. The most familiar, and relatively direct, indicators are those based on catch per unit effort (CPUE). These indicators are usually assumed to be proportional to stock abundance so that, for example, a 20% drop in the indicator reflects a 20% drop in the abundance of the exploited part of the population. They can be thought of as indicators of the density of fish over the area where the long-line is set, or the volume of water trawled by a trawl net.

Another set of indicators are based on size (length or weight) frequency data from the catch. As noted above, here the 'signal' is more related to the harvest rate through the changes it causes to the age and size structure of the population. Ideally, one would want to sample the whole population, i.e. all age classes, but with pelagic species it is usually only size frequency data from the catch that are available. This may mean that only a subset of age classes are represented in the data. Nonetheless, through inferences about the total mortality rate, one can to some extent infer the relative population size.

Size-based indicators that reflect certain aspects of the population may have particular potential as indicators. For example, a substantial drop, or strong downward trend, in the average size of mature fish in the catch, could be a warning signal that the spawning biomass is declining. Unfortunately, size-based indicators are likely to be sensitive to big changes or trends in the fishery (or gear) selectivity pattern. It is also likely to be difficult to distinguish between changes in indicators that are due to selectivity, versus changes that are due to changes in the underlying population. These indicators can, for example, be particularly sensitive to large variability in recruitment.

In addition to quantities such as mean, median or some percentile of size or weight, one can construct indicators that reflect the proportion of, say, small fish in the catch, or of mature fish in the catch. Proportion-based indicators also potentially suffer from misinterpretation. For example, the proportion small fish in the catch can increase either because the number of small fish has increased, or because the number of large fish has decreased.

This leads us to another set of indicators, which we have generally grouped together with CPUE-based indicators, and which avoid some of the problems associated with proportions. These are composite indicators, based on both catch per unit effort information and size frequency information, for example, the CPUE of small fish, or the CPUE of mature fish.

There is an important group of indicators, which we list in Table 1 but which we do not consider in the simulation evaluations, namely indicators based on age. This omission is intentional. For the fisheries we are considering here, the catch is not yet routinely aged, and in some cases ageing is difficult and expensive. There are many reasons why age information would be (should be) preferred over size-based information as indicators of total mortality in the population, but from a practical point of view, we considered it more important to evaluate the performance of size-based indicators in this project.

Table 1. Candidate Indicators of Stock status.

Data	**Measures**	**Available?**	**Easy to Simulate?**
Catch per unit Effort	Average CPUE CPUE by 'area' etc.	Yes	Yes
Size frequency	Mean length Median Upper 5th percentile Proportion > size at maturity (or proxy) Proportion < size at recruitment to fishery	Yes (limited data for domestic fleet)	Yes
Weight frequency	Mean weight Median Upper 5th percentile Proportion > weight at maturity (or proxy)	Yes (direct or via size freq. and L-W relationship)	Yes
Age data	Catch curve estimate of total mortality (also e.g. Rochet and Trenkel, 2003)	No (could be available for some species)	Yes
Maturity information[5]	Maturity at age or size (in conjunction with other indicators)	No	Yes (with assumption about density dependent response)
Size-specific CPUE	CPUE of 'small' fish CPUE of 'big' fish CPUE of spawners	Yes	Yes
Catch data	Relative rate of catch increase RRCI (Fonteneau and Gaertner, 2002)	Yes	Yes

The last measure in Table 1, relative rate of catch increase (RRCI; Fonteneau and Gaertner, 2002), can give crude estimates of MSY as determined from average catch when RRCI is zero or negative. The assumption underlying the approach is that the yearly rate of increase of catches tends to decrease as MSY is approached, and be null or negative when MSY is reached or exceeded. One problem is that catches can be driven by many external factors other than stock size. Another problem is that MSY is an equilibrium concept whereas fisheries are usually in transition rather than at equilibrium. Gaertner *et al.* (2001) obtained a crude estimate of MSY via

[5] A change in maturity at age or size is usually assumed to take place as a response to reduced density of the population. The relationship between the change in maturity at size and change in population size is likely to be unknown and we do not propose to use this measure as a direct indicator. It is, however, something one would like to know if it is occurring.

the RRCI. For yellowfin tuna, a strong relationship was found between MSY estimates from standard equilibrium production models and those from RRCI. This leads to the notion that RRCI can be used to obtain proxies of MSY for fisheries, particularly where it is difficult to estimate effective fishing effort, e.g. skipjack. Catch data can be notoriously misleading as an indicator of stock status, but it tends to be used because it is sometimes the only available information.

6.3. Candidate indicators of change in spatial distribution

There is a growing body of evidence that suggests swordfish is less mobile than previously thought, and may be particularly vulnerable to local depletion. There is evidence of this having occurred in the (Australian) East Coast fishery (Campbell and Hobday, 2003). It is therefore important to monitor spatial aspects of the fishery. The extent to which some of these type of indicators can be tested in simulations depends on the extent to which fleet dynamics can be modelled realistically. It is beyond the scope of this study to model the economics or decisions driving fleet dynamics. It is, however, possible to model stock distribution and fleet dynamics under the assumption that fishing will occur in areas of highest catch rates. The list below reflect indicators which are likely to be meaningful in this particular context.

Table 2. Candidate Indicators of change in spatial distribution

Data	Measures	Available?	Easy to Simulate?
Effort at location	Mean distance offshore of effort (e.g. by vessel size or other appropriate categories) Mean longitude of effort Number of squares fished for 'n' years (index of contraction or expansion)	Yes (fine scale data available for domestic fleet)	Yes
Catch per unit effort at location	Catch rate by the above measures, e.g. Catch rate by longitude Catch rate by number of years a square has been fished	Yes (fine scale data available for domestic fleet)	Yes
Size frequency data	Size-based indices (see Table 1) on spatial scale, e.g. by distance offshore	No (possibly very limited data)	Yes

6.4. Quantities which could reflect changes in targeting

Both the WTB and ETB fisheries are essentially mixed species fisheries. The catch data from both fisheries show that one species can, and often does, dominate the catch. There are also, however, mixed sets, and the catch composition changes through the year (and from year to year). These changes almost certainly depend on a complex combination of stock distribution, which affects catch rates, and external factors, such as operating costs and market price of products. There is little doubt that the fisheries can, to quite some extent, target the different species (swordfish, yellowfin, bigeye, and more recently albacore). Targeting is something which affects CPUE, and which is considered very difficult to characterise and standardise for. If

targeting is directed at specific sizes of fish and this changes from year to year, then the size-based indicators are also likely to be affected.

The list of quantities given in Table 3 are not indicators of stock status on their own but rather indicators of possible changes in targeting which could in turn affect the interpretation of other indicators. This project was set up to investigate the robustness of stock status indicators in a single-species context. The intention was not to construct a complex multi-species model during this project. It is nonetheless important to bear the targeting issue in mind even if we cannot easily simulate and test indicators fully in that context. Instead we propose that the list of quantities (and possibly others) in Table 3 should be considered when looking at and interpreting changes in indicators based on real data. This table therefore lacks the column related to simulations.

Table 3. Quantities which could reflect changes in targeting or fishery characteristics (excluding measures in Tables 1 and 2 which may also reflect such changes).

Data	Measures	Available?
Effort data	Locations fished Time of year Time of day or night Hooks per basket (or direct measures of fishing depth) Other (e.g. hook type, environmental or ocean conditions)	Yes (locations, fishing depth not always available)
Catch data	Proportions by species (whole fleet) Proportions by species by 'area' and time	Yes
Other Fishing/fleet characteristics	e.g. Direct evidence (from fishermen) Are light sticks being used? - tends to favour swordfish catches Setting by night and/or day Vessel type, vessel characteristics	Yes

6.5. Indicators used in further analyses

When the project was set up, the main focus was on the performance of size-based, and possibly spatial, indicators. Even during the early stages of the project, it became clear that it was essential to also consider CPUE and composite CPUE indicators, such as the CPUE of large fish. These indicators are so familiar and so directly related to stock abundance when assumptions are met, that they can serve as 'best performance' benchmarks for the other indicators. The general investigations of indicator responsiveness and of indicator performance when used as predictors of spawning biomass have therefore been done with CPUE, size-based, weight-based and size-specific CPUE indicators in Table 1. For the swordfish-specific investigations associated with local depletion, we have used the 'mean distance offshore' of effort in a spatially explicit model.

We have already indicated that we have not considered age data because these data are not yet available in the fisheries we are considering, and these data are often difficult and expensive to collect. We have also used size as a proxy for maturity rather than assuming that direct maturity information are available.

In the simulation studies, we generate the indicators in such a way that they can be considered to be standardised, similar to the way in which CPUE data are standardised before being used as an indicator, or in an assessment. In reality, however, there may have been systematic changes in operation (e.g. the location and timing of fishing) which should be standardised. This is a very familiar procedure for CPUE data, but not for other indicators. In the next Section (7), we look at actual size-frequency data and explore ways of standardising these data.

7. Standardisation of size-based indicators: three case studies

Summary

This Section addresses the second part of the first objective, namely to develop approaches for standardisation of size-based indicators. Aggregated size frequency data from the Indian Ocean Tuna Commission database were used to calculate indicators and general linear models were used to standardise the series. In most cases examined here, there were only small differences between the nominal and standardised series. The largest difference was observed for the swordfish indicators, where area fished (within the whole of the Indian Ocean) was an important factor.

Although standardisation did not make a big difference to most series, this work illustrates the kinds of approaches that should be considered to standardise an indicator time-series, or at least to explore whether standardisation is required, before using it as a management tool.

7.1. Introduction

Here the objective is to explore and illustrate how size-based indicators can be standardised (see Section 4. Objectives). This is analogous to the familiar standardisation of catch per unit effort (CPUE) data before use as an indicator or in a stock assessment. Size frequency data for three case studies are considered: bigeye tuna, yellowfin tuna and broadbill swordfish. The data are taken from the IOTC (Indian Ocean Tuna Commission) database and therefore represents a much larger geographic area than, for example, data from the Western Tuna and Billfish fishery (WTBF). There are two main reasons why the IOTC data rather than the WTBF data were used to address this objective of the project.

First, the WTBF time series of size frequency data is much shorter than those from the IOTC database, particularly the size frequency data from the Japanese longline fleet. Second, the WTBF size frequency data are not directly related to a location, since data are collected during processing. The IOTC data are also relatively coarsely aggregated, but do provide direct information on location in terms of a 5x5 degree grid square.

In addition to the above reasons, discussions within the IOTC Working Party for Tropical Tunas (WPTT), suggested that if mean size were to be used as an indicator of stock size, then these data should ideally first be standardised. The work presented in the two working papers, which form this Section, showed how this could be achieved.

The time-series of mean length for both tuna stocks, yellowfin and bigeye tuna, were not in fact much affected by the standardisation, suggesting that standardisation may not always be necessary. It is nonetheless important to note that this cannot simply be taken for granted; the standardisation exercise needs to be done to confirm whether standardisation makes a difference or not. There is also an informative and useful by-product of the standardisation exercise, namely estimates of uncertainty (CV) in each year for the standardised indicator.

In the case of swordfish, it would have been ideal to analyse the Taiwanese size frequency data, because this fleet targets swordfish, whereas the Japanese fleet takes swordfish primarily as a by-catch. It was not, however, possible to obtain the Taiwanese size frequency data to undertake the standardisation. The time-series of mean length of swordfish in the Japanese catch was affected by the standardisation, particularly in some years.

We note here that standardisation cannot (easily) remove all factors which might affect size-based indicators. For example, a change in the selectivity pattern of the fleet which is not related to obvious (measurable, or recorded) changes in operation, or which is due to a substantial change in the stock's spatial distribution by size, would not be removed by standardisation. Indicators – both size- and CPUE-based - should therefore always be considered carefully for potential impacts of changes in the fishery which may not be removed by standardisation, and which may affect the interpretation of the indicator.

The analyses and results for the three case studies are presented in the two working papers that follow:
1. Standardisation of size-based indicators for bigeye and yellowfin tuna
in the Indian Ocean
2. Standardisation of size-based indicators for broadbill swordfish in the Indian Ocean
In these two working papers, references to scientific papers are given at the end of each working paper, not in Section 18.

Please note that tables and figures in these attachments are not included in the List of Tables or List of Figures.

7.2. Postscript regarding Swordfish standardisation

At the end of the paper on standardisation of swordfish size-based indicators, we note that results of standardized CPUE and size-based indices for the Indian Ocean broadly suggest declines in CPUE but no apparent trend in mean size. This is in contrast with observations from the Mediterranean Sea, where swordfish CPUE has remained stable but mean size has decreased (e.g. ICCAT, 2007), though it should be added that it is not clear whether changes in fishing practices have lead to this. It is also in contrast to swordfish in the Atlantic (see e.g. IOTC, 2004). We also note that this combination of indicator trends could perhaps have more serious biological implications than what is observed in the Indian Ocean, since it hints towards hyper depletion, whereby an aggregating stock or highly efficient fleet allows catch rates to be maintained despite declines in overall abundance. Additionally, a decrease in mean size despite constant catch rates has potential implications for reproduction and recruitment, since fecundity is positively related to size. It is worth considering this contrast in a little more detail, since further investigations and comparisons may well be informative in terms of the spatial structure and habitat use of different swordfish populations.

The first obvious comment is that the Mediterranean is a substantially smaller area than the Indian Ocean and this is highly likely to have some effect on the spatial structure and movement dynamics of the respective populations. In the stock assessment report of ICCAT (ICCAT, 2007) it is commented that the potential reproductive area (for swordfish) in the Mediterranean is probably relatively larger than that in the Atlantic, and the productivity of the Mediterranean Sea is thought to be very high. The International Commission for the Conservation of Atlantic Tunas (ICCAT) held a stock structure workshop (ICCAT, 2006a) and the stock assessment report (ICCAT, 2006b) notes that "while the delineation between stock boundaries remains imprecise, the results of the research presented gave general support to the stock structure currently assumed for Atlantic Swordfish (Mediterranean and North and South Atlantic stocks)". In other words, the Mediterranean stock is treated as a separate stock for assessment and management purposes. There are also indications that age at maturity is lower for Mediterranean than Atlantic swordfish (Macías *et al.* 2005); growth rates may be higher, but there are concerns about comparability of data and methods (ICCAT, 2007). It is also not clear whether such

differences are inherent in the population or whether they are a result of fishing (possibly heavy fishing, over many years).

Regarding the stock structure of swordfish more broadly, Ward et. al. (2001, and references therein) comment on work which showed that swordfish are comprised of two main phylogenetic clades. The authors explain that: "Clade I is found throughout the distribution of the species range while Clade II is found in highest frequency in the Mediterranean Sea. The most clearly, consistently and strikingly differentiated stock was that in the Mediterranean and early studies concluded that swordfish are geographically subdivided on ocean basin scales." The authors also note that, within the Atlantic Ocean, some studies support heterogeneity whereas others do not, but they conclude that the evidence favours genetic stock heterogeneity within the Atlantic. Furthermore, they note that Pacific Ocean samples of swordfish show small but significant genetic differences from Atlantic fish. Within the Pacific Ocean, a complicated pattern of heterogeneity and homogeneity seems apparent. The authors (Ward *et al.*, 2001) conclude that both allozyme and mtDNA data point towards population heterogeneity in the Pacific.

Indian Ocean swordfish appear to be genetically very similar to Pacific fish. Unfortunately, no single study seems to have compared sufficient samples from a large number of geographic sites. We comment on two studies here. First, a study by Ward and colleagues (Ward *et al.* 2001) examined the genetic stock structure of swordfish caught in the Australian fisheries on the west and east coasts, and to determine how these populations might relate to other populations in the Indian and Pacific Oceans. The study used a combination of mitochondrial DNA and microsatellite DNA analyses. Samples from four areas were considered: Australia's east coast fishery, Australia's west coast fishery, from off Reunion Island, and from the Pacific Ocean. The study concludes that combing the mtDNA and nuclear DNA evidence from all the samples suggests several stocks: (1) a northern Pacific stock (differentiated by mtDNA), (2) a southern Pacific stock (differentiated by mtDNA) which is, surprisingly, genetically similar to Reunion Island, and (3) a western Australia stock (differentiated by microsatellite DNA). The study concludes that, if there are distinct stocks, then the degree of separation of them is small.

According to Lu et. al. (2006), their genetic analyses of samples from the Indian and Pacific Ocean suggest that the stock structure of swordfish in the Indo-Pacific region can be summarized into the following groups: an area off northern Madagascar, the Bay of Bengal, and the rest of the Indian Ocean and western Pacific.

In summary, with regard to the Indian Ocean, there is still much uncertainty about population structure. We note, however, that the (apparently) heaviest fished areas in the Western Indian Ocean (off Madagascar) are those where declines in CPUE have been observed for some fleets and where there have not been clear trends in mean size (see IOTC, 2004). The basic question remains in the light of the genetic and stock structure information: why might some populations exhibit declines in mean size of the catch but no trend in CPUE, whereas others exhibit declines in CPUE but no clear trend in mean size of the catch? It is beyond the scope of this project to answer this question and the answer can lie either in the fisheries themselves (e.g. changes in fishing practices or targeting etc.), in the population characteristics, or a combination of both.

7.3. Attachment 1: Bigeye and yellowfin tuna example

IOTC-2004-WPTT-15

Standardisation of size-based indicators for bigeye and yellowfin tuna in the Indian Ocean

Marinelle Basson & Natalie Dowling
CSIRO, Marine Research
Castray Esplanade, Hobart
Australia

Abstract
Information on the size distribution, for example mean length, in the catch can potentially be used as an indicator of stock status. This study looked at the standardisation of such indices particularly for spatial and seasonal effects. Results for bigeye and yellowfin in the longline and purse seine fisheries in the Indian Ocean show that the standardised series are very similar to the 'nominal' indices. This may suggest that, for the current datasets, there is not a strong need to standardise. Standardisation does, however, have other benefits, including the availability of standard errors and insight into the spatial and seasonal patterns from the estimated coefficients.

1. Introduction

Simple indicators from fishery data, e.g. mean length in the catch, are increasingly used as potential indicators of stock status. Such indicators are either used on their own where no assessment is available (e.g. swordfish, skipjack; IOTC WPB and WPTT reports), used instead of an extensive assessment (e.g. southern bluefin tuna during development of management procedure), or used in conjunction with an assessment (bigeye, yellowfin tuna; IOTC WPTT reports). Changes or trends in these indicators could be caused by many different factors and it has been argued that indicators should ideally be standardised, in the same way the commercial longline CPUE is standardised. It is, for example, easy to see that if a stock is spatially disaggregated by size, changes in the timing and location of a fishery could lead to a change in the mean length in the catch which may have nothing to do with a change in the underlying size distribution of the overall exploitable stock.

In this study we consider only the size frequency data and attempt to standardise different measures from these data, including the mean, median and upper 80th percentile, with a view to developing standardised 'stock status indicators'. The idea is simple. At high harvest rates one may see fewer large, old fish. The longline fishery size frequency data could, for example, reflect an absence or decrease in the proportion of old, big fish. The purse seine fishery, on the other hand, could reflect an absence of small fish, which may indicate low recruitment. Given that this gear takes the young faster growing age classes, it could also reveal changes in mean size that may reflect changes in growth rate. Monitoring these indices could therefore provide information on changes in the underlying population structure. It is obvious that changes in the fishery itself could lead to changes in the size frequency, and this should always be borne in mind. This study does not address reasons why there may, or may not, have been changes.

It is important to note that this is only one step in a much wider study that started in September 2003 (SSI-study). The SSI-study aims to explore, via simulation, the robustness and sensitivity of a wide range of indicators, and to develop ways of combining several such indicators into a

coherent framework for management decision-making. The study is addressing these issues primarily in the context of the domestic (Australian) fishery, but also in the wider context of the IOTC. The simulation part of the SSI-study may find that mean length is insensitive to stock changes (see e.g. Punt et. al. 2001), and that some other measure, possibly even based on data not currently available for the IOTC fisheries, is more reliable. The work presented here should be seen in this wider context, as an exploration of the effects of, and need for, standardisation of indices based on existing data.

2. Data and Methods

Basic data were taken from the raw size frequency data held by the IOTC. Instead of converting from different measurement types, only data in terms of fork length (FL) were used. The raw data were manipulated to obtain the lower *n*th percentile, median, mean and upper *n*th percentile by gear type, fleet, area, month and quarter (month=quarter for longline).

We considered only two gear types: longline and purse seine. We chose fleets with most samples and longest time series: Japanese longline (LL) data and Spanish, French and NEI-EUR purse seine (PS) data. For the purse seine fleets, temporal trends were almost identical, so all analyses were conducted on data combined across fleets. For the longline data we also looked at the proportion of fish greater than a given length, chosen to reflect approximate size at maturity. For both species this was taken to be 100cm FL, based on the results of Shung (1973) for yellowfin tuna and Whitelaw and Unnithan (1997) for bigeye tuna. For bigeye, this length corresponded to most of the size frequency from longlines (LL), so a larger length threshold of 140cm FL was chosen based on examination of length-frequency data. The indicator is referred to as the 'proportion big'.

For purse seine data we also constructed an indicator based on the proportion of small fish (size corresponding approximately to 1 y.o. fish based on published growth curves: 60cm FL for bigeye and 55cm FL for yellowfin), with the notion that it might reflect recruitment dynamics. Changes in targeting (size) or in fishing techniques could also lead to increases in the proportion small fish in the catch, and the modelling we undertake here does not address the reasons for changes in indicators.

Prior to 1991, sampling of purse seine catch was poorer both in terms of sample size and sampling coverage. For example, there is a lack of data in some months and areas. Sample sizes are also not available in the database. The mean lengths plotted against year showed markedly more variability prior to 1991, and initial modelling attempts showed that it was difficult to deal with the early data. We therefore excluded purse seine data prior to 1991 from the GLM analysis.

Preliminary explorations showed so much similarity between annual values of mean, median, lower 20th and upper 80th percentile that we chose only to apply standardisations to one of these, namely the mean. The mean length and proportion of fish greater than a certain length were also similar. This is not surprising, but both were considered given that the next phase of the study will explore the relative robustness of these different candidate indicators.

In all cases simple general linear models (in Splus software) were used for standardisation. We identified plausible covariates to include, explored those and interactions between them where sensible, and extracted the appropriate 'standardised' series. Although it is possible to fit a model to size-based indicators from several fleets and gears combined, we decided to separate the gears. We considered that the size frequencies from the two gears are different enough that they may reflect different aspects of the population. Combining the data and 'correcting' for gear type could potentially lead to a loss of signal or information. Given that gear types were treated

separately, the main covariates to consider were time and area effects.

There are two reasons for considering spatial factors: 1) to take account of changes in the locations of fishing over time and/or 2) to take account of changes in the size frequency in different areas over time. The second reason could be very important for a species like swordfish which is now thought to have a relatively high expected residence time in an area (Campbell and Hobday, 2003). Different levels of fishing intensity in different areas could therefore have different effects on the mean size in the catch over time. This type of change is more likely to be observed at a relatively coarse spatial scale than a very fine spatial scale. Given our interest in this type of effect for swordfish, we also explored this for bigeye and yellowfin (see below).

We consider that it is inappropriate to incorporate environmental factors into this type of standardisation unless there is clear evidence of a plausible mechanism for catchability changing differentially by size under different environmental conditions. Where environmental conditions in different areas or at different times affect the size distribution, the area or time factors should take care of that aspect. Although it is plausible, and indeed likely, that growth may be affected by environmental conditions, it is unlikely that the size frequency in any given year should be affected by environmental conditions in that year. This is because a size frequency distribution does not reflect the growth in a single year or the growth of a given cohort, but rather the cumulative growth of several cohorts over several years.

Models for longline mean length and proportion big

Multiplicative models were chosen for mean length, with area and time (year and quarter) as the main factors. The log of mean length is therefore an additive model of the relevant factors and we assume that errors are normally distributed. For both species, we chose several different spatial scales. At the fine scale level, we used 10-degree latitude bands and the lat-long grid positions for each quarter. The next level of detail was the "CPUE areas" that had been assigned when standardising catch rates (for bigeye, Okamoto & Miyabe (2003) (Figure 1), for yellowfin, Nishida *et al.* (2003) (Figure 2)), with the notion that the areas have some coherence in terms of fishery and other characteristics. For interest, we also looked at the Longhurst areas (Longhurst 2001), which are meant to represent "habitats" and which may therefore contain fish of certain age or size classes, if distribution is driven by habitat preference, and if the Longhurst area is at a spatial scale capable of distinguishing between those habitats. At the coarsest spatial scale, we considered the eastern and western Indian Ocean with a model that estimates a year effect in each area (covariate 'WestEast').

During the exploratory phase, more complex models with main effects and interactions, for example, between year, quarter and area were fitted and compared to more parsimonious models. Recall that a model with main effects implies that the mean length in any year, area and quarter is the sum of a year effect, an area effect and a quarter effect. A model with an interaction between, say, area and quarter implies that the quarterly effects are different in each of the spatial areas. It is important to bear the meaning of each model in mind, particularly when including interaction terms. Splus software was used and the model definitions for some of the simpler examples we considered are given in terms of S+ notation below:

Model type 1:glm(log(mean length) ~ year + area + quarter)

Model type 2 : glm(log(mean length) ~ year + area*quarter)

Model type 3: glm(log(mean length) ~ year*WestEast + quarter)

The 'proportion big' was fitted with a binomial model and factors were again time and the three different area definitions. The data were weighted by sample size, and the models were fitted to the subset of the data where sample size was greater than 50 (corresponding roughly to the lower quartile of a summary of sample sizes across the whole data set).

Models for Purse Seine mean length and 'proportion small'

Data based on purse seine fishing contains a further important factor: fishing on free schools (FS data) or fishing on FADs (log school or LS data). The importance of this for the size frequency in the catch has been identified previously in IOTC reports and working papers. Essentially, the FS mean length had a bimodal distribution, whereas the LS mean length had a unimodal distribution (see e.g. Figure 12). Due to these differences, the free school and FAD data were considered using separate GLMs rather than trying to model the data with School Type as a factor. However, no systematic factor could be identified that could remove the bimodality in the free school data. Since this causes serious problems for the analysis and invalidates some of the model assumptions[6], analysis was only undertaken for the associated school data.

The area designations used with the longline data, particularly those based on CPUE standardisation, were not considered appropriate for the purse seine data. This fishery is concentrated in a much smaller area compared to the longline fleets. The lat-long grid points were used as one option, and another was constructed based on the 10-degree latitude bands containing the majority of the data (9N-19S). Data outside the 9N and 19S latitude bands were excluded from the analyses.

Modelling of the proportion small yielded problematic residuals under a binomial assumption. Since the nominal temporal patterns for the proportion small were largely inversely related to those for the mean length, the standardized indices yielded no new information and, given that they were statistically unreliable, the results are not presented here.

Model selection

The key aim of fitting GLMs to the two indicators is to remove changes due to different timing (within the year) or different locations of fishing, and to extract the so-called year-effects from the models. The aim is not to determine statistically which factors are significant or not. Nonetheless, it is good practice to compare the goodness of fit of different models to avoid under- or over-fitting. The deviance residuals and q-q plots associated with each model were used to evaluate goodness of fit, homogeneity of variance and extent of conformance to a normal distribution. Where models were nested, appropriate significance tests were used to compare them, but in the case of the binomial models, these are unlikely to be reliable because of over-dispersion. Where models were not nested, particularly where the different area-designations were used, Akaike's information criterion (AIC) was used. Our main concern was, however, with the year-effects and whether they were sensitive to the different model formulations or variables included.

[6] The problems are similar to a situation where one is trying to fit a regression line through two clusters of points. The clusters will have strong leverage on the resulting regression line, but the model itself may be meaningless.

3. Results

3.1 Longline: Bigeye Tuna

Nominal trends and exploratory analyses

The mean lengths by area, quarter and year were highly variable. The annual averages of mean length were also very variable, ranging between about 117 and 134cm. Although there appear to be increases and decreases in the annual average series, these should be seen within the context of the range of the size frequency. Figure 3a shows the annual averages (over location and quarter) of the mean and of the 20^{th} and 80^{th} percentiles. Changes in the mean (and median) time series were small relative to the inter-percentile range. Similarly, the decrease in bigeye mean size over the last 5 years appears relatively insignificant relative to the long-term pattern, and the size range. The inter-quantile range has not changed substantially over the whole period, varying between 22 and 35cm with no systematic temporal pattern.

The pattern of the nominal time series plot for the proportion of fish >140cm FL was similar to that of mean length (Figure 3b), as expected. There was an increase in the proportion of large fish in the early 1990s to historically high levels (approx. 40%), followed by a decrease to the pre-1990 levels.

Exploratory analysis was undertaken to investigate the effect of alternative temporal scales and area designations on the inter-annual trends. This included comparing nominal time series broken down by quarter and area to find any obvious patterns, and comparing standardised indices across alternative models. The following points summarise the minor findings from the exploratory analyses:

- Plots of mean length (and the other candidate indicators) over time and by the different area designations did not show any clear patterns to inform model choice. Plots were generally noisy, with missing data and much inter-annual variability.

- For both mean length and "proportion big", the relative pattern of the year coefficients was very similar between model types and choice of area designation. Although models with interaction terms generally gave a better fit to the data, the standardised indices were essentially unchanged. Figure 4 shows the similarity in mean length GLM year coefficients for 5 alternative model types/area designations.

- Models using the CPUE standardisation areas generally gave lower AIC values compared to those using other area designations. Therefore, only the main effects models using the CPUE standardisation areas are subsequently considered in detail.

GLM standardisation of mean length and proportion >140cm FL

Results are presented for mean length modelled as:

glm(log(mean length) ~ year + CPUEarea +quarter).

This model has a common spatial pattern for each year and a common quarterly pattern in each year and area. The standardized indices showed a similar temporal trend to the nominal pattern (Figure 5a). However, the standardised indices for the proportion of bigeye > 140cm FL, using the same factors as the mean length model, showed a clear decline from the mid 1970s to 1990,

which was not as apparent in the nominal trend (Figure 5b).

Figure 6 shows diagnostics and results for the mean length model. Residuals did not show any systematic patterns, either overall or when plotted against each set of predictor variables. The q-q plot showed that the data were almost normally distributed, with deviation mainly at the lower tail. These large negative residuals persisted in models with interactions. It is interesting to note that areas 1, 2, 3 and 6 had higher coefficients. These areas corresponded to the western Indian Ocean (Figure 1). Areas 4, 5 and 7, corresponding to the eastern Indian Ocean, had lower coefficients. The lowest-value quarter coefficient was in quarter 3. The results imply a 5cm difference between mean length in quarter 1 versus quarter 3, and a 14cm difference between mean length in area 1 versus area 7.

Residuals for the model of proportion of bigeye >140cm FL were also reasonably homoscedastic, close to normal and showed no pattern when plotted against each set of predictor variables. The area and quarter coefficients followed a similar pattern to those from the mean length model.

The standard errors of the standardised indices tended to be inversely related to sample size (Figure 7). This effect was more marked when weighting by sample size was used in the models. There is therefore less relative certainty associated with the standardised trends from more recent years, and this will have an effect on how such a series is implemented as a "stock status indicator". This comment also applies to the other standardised indices presented below.

3.2 Longline: Yellowfin Tuna

Nominal trends and exploratory analyses

Mean fork length of Japanese yellowfin catch showed the most dramatic change in the first 5 or 6 years of the fishery with a drop from ~145cm to ~115cm (Figure 8a). As with bigeye, plots of the annual average 20th percentile, mean and upper 80th percentile showed that, apart from the first 5-6 years in the time series, changes in the mean were small relative to the size range of the length frequency. The inter-percentile range varied between 12 and 29cm, with the narrowest range at the start of the time series, followed by an increase and then a decrease in the final few years. We did not have time to enquire whether there are any problems with, or less reliability in the data, prior to, say, 1965 (the first year for which bigeye data are available). Since catches in these early years were low relative to catches in later years, it seems unlikely that this would have been the effect of fishing, but this cannot be ruled out.

It is interesting that time series of minimum, median and maximum length values generally also showed parallel trends, with the notable exception of a decrease in minimum size in the late 1960's and early 1970's. This could possibly indicate one or more strong recruitments. One should also remember that the minimum and maximum would be far more sensitive to sample size and sampling noise than, say, the mean or median.

There was much inter-annual variation but no consistent temporal trend in the proportion of fish >100cm (Figure 8b). As expected, this series showed the same features as the mean length series.

The following points summarise the minor findings from the exploratory analyses:

- Plots of candidate indicators over time and by the different area designations did not show any clear patterns to inform model choice. Plots were generally noisy, with missing data and much inter-annual variability.

- A decrease in mean length since 1952 (see below) was seen in both the western and eastern Indian Ocean, but the increase in the mid 1980s until 1994 was mainly in the western Indian Ocean, and most prominent in the northern hemisphere.
- For mean length, Model 3, which incorporated latitude, longitude and quarter in a single factor, gave the lowest AIC and residual deviance of all the models. For the proportion of fish >100cm FL, models including the CPUE standardisation areas gave the lowest AIC among all model types. However, for both indicator variables, there was very little difference between model types and choice of area designation, in terms of the relative pattern of the time series of year coefficients, with the extent of similarity consistent with that shown for bigeye (Figure 4). As such, only the main effects models using the CPUE standardisation areas are subsequently considered in detail.

GLM standardisation of mean length and proportion >100cm FL
Results are presented for mean length modelled as:

glm(log(mean length) ~ year + CPUEarea +quarter).

Standardised indices showed a similar trend relative to the nominal plot, with the most recent value similar to the long-term mean (Figure 9a). Figure 10 shows diagnostics and the other estimated coefficients. Residuals showed some heteroscedasticity and many large negative values. Apart from the tails, the distribution was reasonably normal. The area coefficients were highest for area 4 (south-east Indian Ocean, Figure 2), and lowest for area 5 (north-east Indian Ocean south of the Bay of Bengal). The quarter coefficients varied depending on whether an interaction term was included in the model, but were consistently low for quarter 2 and high for quarters 1 and 4.

Fitted and observed values of the 'proportion big' (>100cm FL) were all clustered close to 1.0, indicating that for the majority of samples, all or almost all fish are "mature". The standardized indices showed a consistent pattern with the nominal trend (Figure 9b). Figure 11 shows that residuals were reasonably homoscedastic but there is again a tail of large negative residuals. This feature was present irrespective of the model choice. Area coefficients showed somewhat different relative patterns than in the mean length model. Quarter coefficients were similar, except for quarter 1, relative to those from the mean length model. As for bigeye, standard errors were, as expected, larger in years where sampling was low.

3.3. Purse seine: Bigeye Tuna

Nominal trends and exploratory analyses

We have already commented on the bimodal nature of the mean length for free school data. This bimodality is illustrated in a histogram of mean length (with means being calculated for each month/year/5-degree square/fleet combination, Figure 12b). The annual average (over month, location, fleet) of mean length and of 20^{th} and 80^{th} percentiles generally resulted in parallel trends but, unlike the longline data, changes in the mean were large relative to the inter-percentile range (Figure 12a). This suggests the possibility that more than one age class is being exploited in very different proportions from year to year. It is interesting to note that results in Stequert and Conand (2003) imply that the mode at 60cm coincides with 1-year old fish and the mode at 110cm (or above) coincides with 2-and-a-half year or older fish. As stated in the methods, a systematic factor responsible for the bimodality could not be identified and thus no GLM standardisations were undertaken on the free school data.

Mean lengths from associated schools (LS data) were generally smaller than those from free schools and the distribution of mean length was unimodal (Figure 12). Changes in the mean length were small relative to the inter-percentile range, which varied between 12 and 22cm.

The proportion of 'small' fish (<60cm FL) for free and associated schools showed an inverse temporal patterns to that for mean length. Since these temporal patterns gave no new information to that provided by the mean length, and due to persistent problems with the residuals under a binomial assumption, no GLM analyses are presented for the proportion of small fish.

The exploratory analyses revealed the following minor points:

- Temporal plots of free and associated mean size and by 10-degree latitude bands and Longhurst areas did not show any clear patterns to inform model choice. Data were generally concentrated within one area, with missing data and much inter-annual variability characterizing the other areas.

- The substitution of month for quarter in the GLMs, and/or the inclusion of fleet or area (Longhurst, latitude band) factors in the GLMs made no discernable difference to the standardised index. The similarity between models was even greater than for the longline analyses shown in Figure 4.

GLM standardisation of mean length

There appears to be a strongly significant interaction between year and quarter, and including this interaction reduced the deviance considerably[2]. This implies that the quarterly pattern in mean length varied according to year. Results for the following simple models are presented:

Model 1 (main effects): glm(log(mean length) ~ year + quarter)
Model 2 (interaction effects): glm(log(mean length) ~ year + year:quarter)

The standardized indices for the main effects model were almost identical to the nominal series (Figure 13), but those for the interaction effects model were slightly different. Note that the indices in Figure 13 are relative to quarter 1. For the main effects model, the pattern over time would be identical for each quarter, but the level would be different. For the interaction model, however, the pattern over time would differ slightly depending on which quarter one is looking at.

Diagnostics for the two models are shown in Figures 14 and 15. There are some positive outliers but no systematic patterns or heteroscedasticity. Outliers were not removed as doing so did not change the overall time series pattern of the standardized indices. The q-q plots for both models showed some deviation from normality, particularly at the right tail. However, as the emphasis of the GLMs is not on optimal model selection, some deviance is acceptable. The comparison shows that, although the residuals are slightly smaller for the more complicated model, the large positive values, and the departure from normality persist.

The quarterly coefficients in the main effects model were highest for quarters 1 and 2, suggesting that the mean size of associated school fish is larger over the summer months (Figure 14). In the interaction model, the quarterly effects are estimated relative to quarter 1 in each year. Figure 15

[2] Model 1 reduced the null deviance (19.83) by 43% (residual deviance 11.36), while model 2 reduced the null deviance by 59% (residual deviance 8.07) (ANOVA of model 1 vs. model 2: df=33, F=51.88, P < 0.000001).

shows that the relative quarterly patterns were similar for years i) 1991-1994, and 2000, ii) 1996 and 2001, and c) 1999 and 2002. There is no pattern apparent in this, and it is probably simply due to 'natural' inter-annual variability. It does, however, show that quarter 2 has the highest coefficients of the three in most of the years, as suggested by model 1.

3.4 Purse seine: Yellowfin Tuna

Nominal trends and exploratory analyses

Mean lengths from associated schools were again generally smaller than those from free schools. The mean and 20th percentile trends were mostly parallel, but the 80th percentile was very variable and appears to have dropped from ~100cm in 1992 to ~70cm in the most recent years (Figure 16). This could be worrying, particularly when seen together with the slight decrease in the longline mean length since 1994.

Free school data and the proportion small yellowfin (<55cm FL) in the associated school data were not used in GLM analyses for the same reasons as given for bigeye. Exploratory analyses for the yellowfin data lead to similar conclusions as those summarised for bigeye above.

GLM standardisation of mean length

Residual and q-q plots for the two models were very similar and are only shown for model 2. There was a cluster of very large positive residuals for the lower fitted values (Figure 18), due to positive outliers in 2001 and 2002, from all quarters. Apart from this cluster of residuals, there are no other systematic patterns or heteroscedasticity. Outliers were not removed as doing so did not change the overall time series pattern of the standardized indices. More complex models did not 'eradicate' the large positive residuals.

As with bigeye tuna, the quarterly coefficients of the main effects model were highest for quarters 1 and 2, suggesting that the mean size of associated school fish is larger over the summer months. The temporal pattern for the interaction effects was reasonably similar for quarters 3 and 4, particularly prior to 1998 (Figure 18). Relative quarterly patterns are similar for years i) 1993-1994, 1998 and 2000 and ii) 1996-1997, 2001 and 2002.

4. Discussion

The results presented here show that it is possible to standardise the mean length and other similar indices reasonably successfully. Admittedly, there are cases where the model fit is rather poor with undesirable features in the residuals. In our experience, there is very little that can be done (given the set of available covariates) to improve this. The standardisation does not, however, seem to have much of an effect on the indicators considered here. The pattern of an annual 'nominal' index was found to be very similar to its standardised version. This suggests that, apart from occasional checks to see that this still holds, it may not be necessary to standardise these particular indices for yellowfin and bigeye.

Although the standardised and 'nominal' series have almost identical patterns over time, there are a few useful by-products of the standardisation process. Most importantly, there are estimates of standard error that are likely to be more robust than direct estimates from the raw data. When implementing such an index in a monitoring framework the standard errors would play a strong role in assessing whether there has been enough of a change in the index to warrant management action. The increase in standard errors when sample sizes decrease was noted, and this would have to be taken into account when implementing such indices into a monitoring or management

framework.

The standardisation also allows for the integration of data in a statistical framework and for 'like with like' comparisons in different areas or in the same area in different quarters. The standardisation can reduce the effects of 'noise' in the data. This does not mean that careful scrutiny of the detailed data are not necessary. Particularly when changes are observed, there would be a need to try to understand why and where those changes are occurring.

Although one would hope that standardisation could remove the effects of spatial or temporal changes in the fishery, we have already noted that there may be other potential factors that affect catch at size but which cannot be incorporated due to an absence of data or the aggregated nature of the data. Therefore, this type of analysis does not, and cannot, take into account changes due to factors that could not be included in the model.

We have also noted that mean or median length may not be a sensitive or reliable indicator. It is easy to show, using simple yield per recruit analysis, that the largest change in mean length as fishing mortality increases occurs for slow growing species. This suggests that mean length may be particularly poor for yellowfin which is relatively fast growing, but it could be reasonable for a species like swordfish. This will be investigated under the SSI-study, and results will be presented to the IOTC WPB in the future.

Acknowledgements

This work was supported by funding from the Australian Fisheries Research Development Corporation (project 2003/042) and CSIRO. We thank the IOTC Secretariat, Miguel Herrera in particular, for assistance with the data and for prompt responses to our questions. Any misinterpretations of the data are entirely our fault.

References

Campbell, R.A. and A. Hobday 2003. Swordfish-environment-seamount-fishery interactions off eastern Australia. Report to the Australian Fisheries Management Authority, Canberra, Australia.

Longhurst, A.R. 2001. Ecological geography of the sea. Academic Press, San Diego, California.

Nishida, T, Bigelow, K., Mohri, M. and F. Marsac 2003. Comparative study on Japanese tuna longline CPUE standardisation of yellowfin tuna (*Thunnus albacares*) in the Indian Ocean based on two methods: general linear model (GLM) and habitat-based model (HBM)/GLM combined – (1958-2001). IOTC WPTT-03-05.

Okamoto, H. and N. Miyabe 2003. Standardised Japanese longline CPUE for bigeye tuna in the Indian Ocean up to 2001. IOTC WPTT-03-11.

Punt, A.E.; Campbell, R.A.; Smith, A.D.M. 2001. Evaluating empirical indicators and reference points for fisheries management: application to the broadbill swordfish fishery off eastern Australia. Marine & Freshwater Research Vol. 52, no. 6, pp. 819-832.

Shung, S.H. 1973. The sexual activity of yellowfin tuna caught by the longline fishery in the Indian Ocean, based on the examination of ovaries. Bulletin of the Far Seas Fisheries Laboratory Vol. 9, pp. 123-142.

Stequert, B. and F. Conand. 2003. Age and Growth of bigeye tuna (Thunnus obesus) in the western Indian Ocean. IOTC WPTT-03-Inf2.

Whitelaw, A.W. and Unnithan, V.K. (1997). Synopsis of the distribution, biology and fisheries of the bigeye tuna (*Thunnus obesus*, Lowe) with a bibliography. Australia, CSIRO Marine Laboratories Report 228, 62pp.

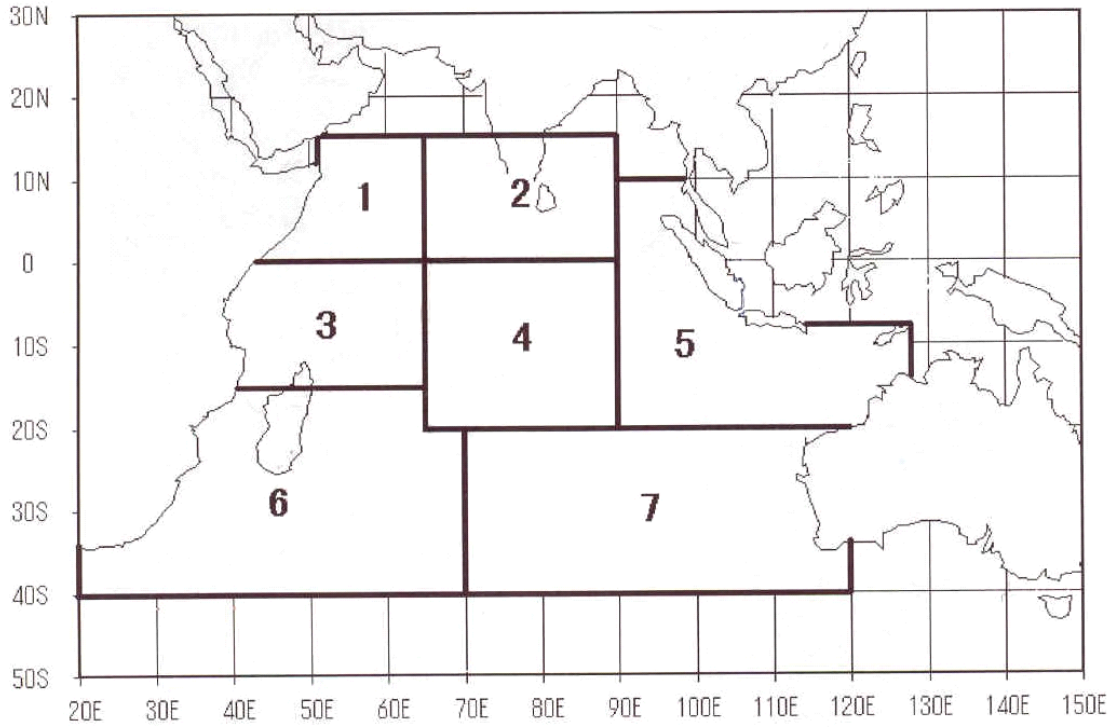


Figure 1: The areas used for standardisation of bigeye tuna longline CPUE (Okamoto and Miyabe, 2003).

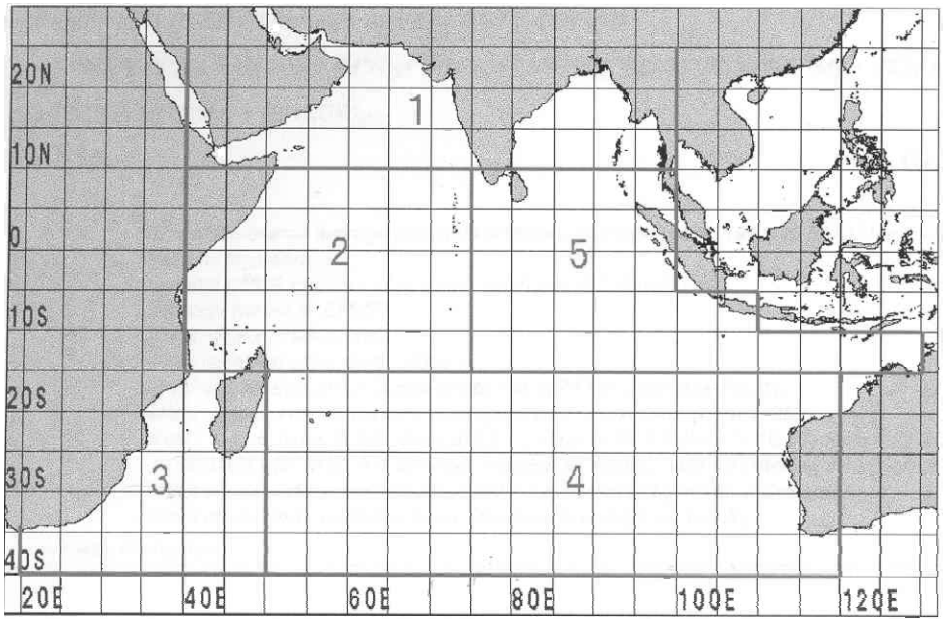


Figure 2: The areas used for standardisation of yellowfin tuna longline CPUE (Nishida *et al.*, 2003).

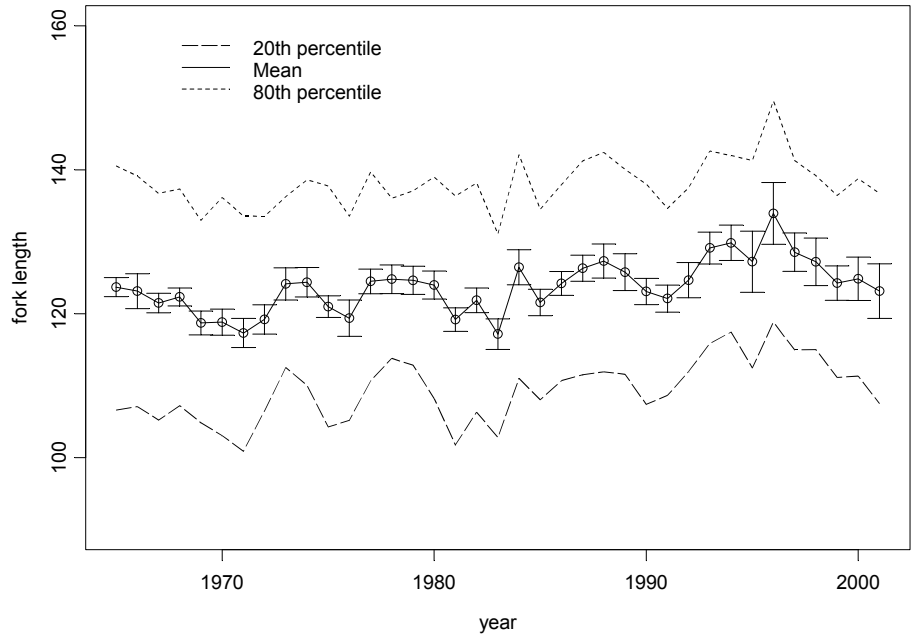


Figure 3a: Bigeye, Japanese Longline. Annual averages of mean length and average 20th & 80th percentiles (FL in cm) for the whole of the IO. Error bars are ±1 standard error (s.e.).

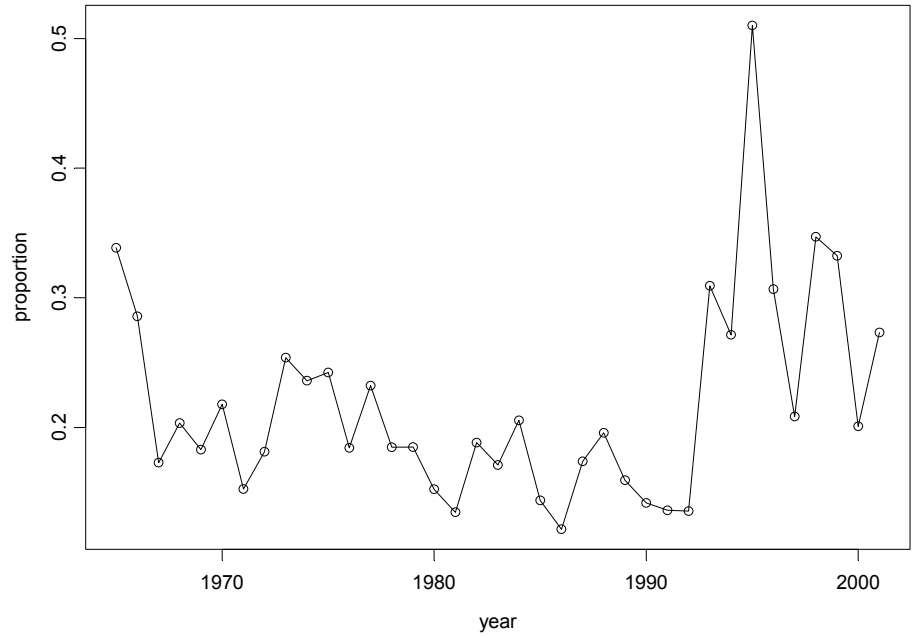


Figure 3b: Bigeye, Japanese Longline. Annual proportion >140cm FL (FL in cm) for the whole of the IO.

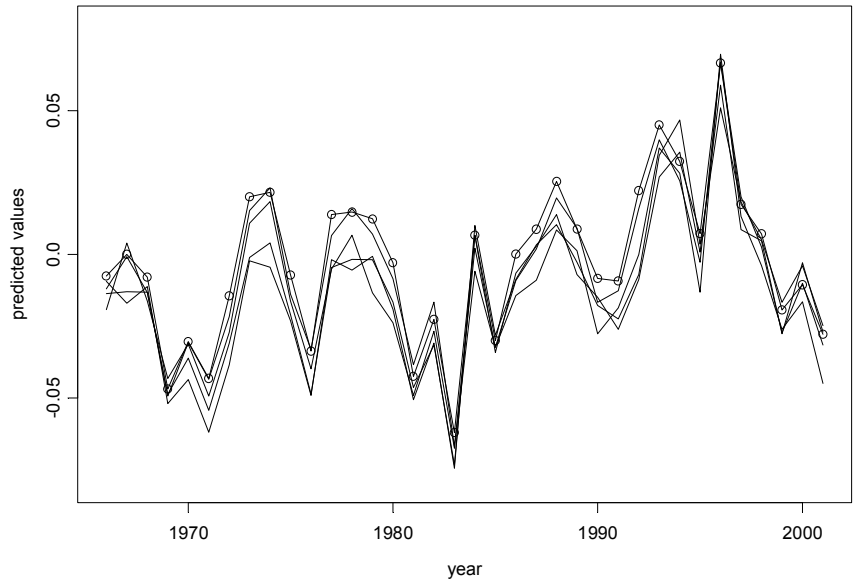


Figure 4: Bigeye, Japanese Longline. Year coefficients overlaid for GLMs of log(mean length) as a function of i) Year + CPUEarea + Quarter, ii) Year + Longhurst + Quarter, iii) Year + Latitude band + Quarter, iv) Year + LongLatQtr, v) Year + CPUEarea*Quarter. (Individual models have not been identified; the figure is meant to illustrate the similarity between models.)

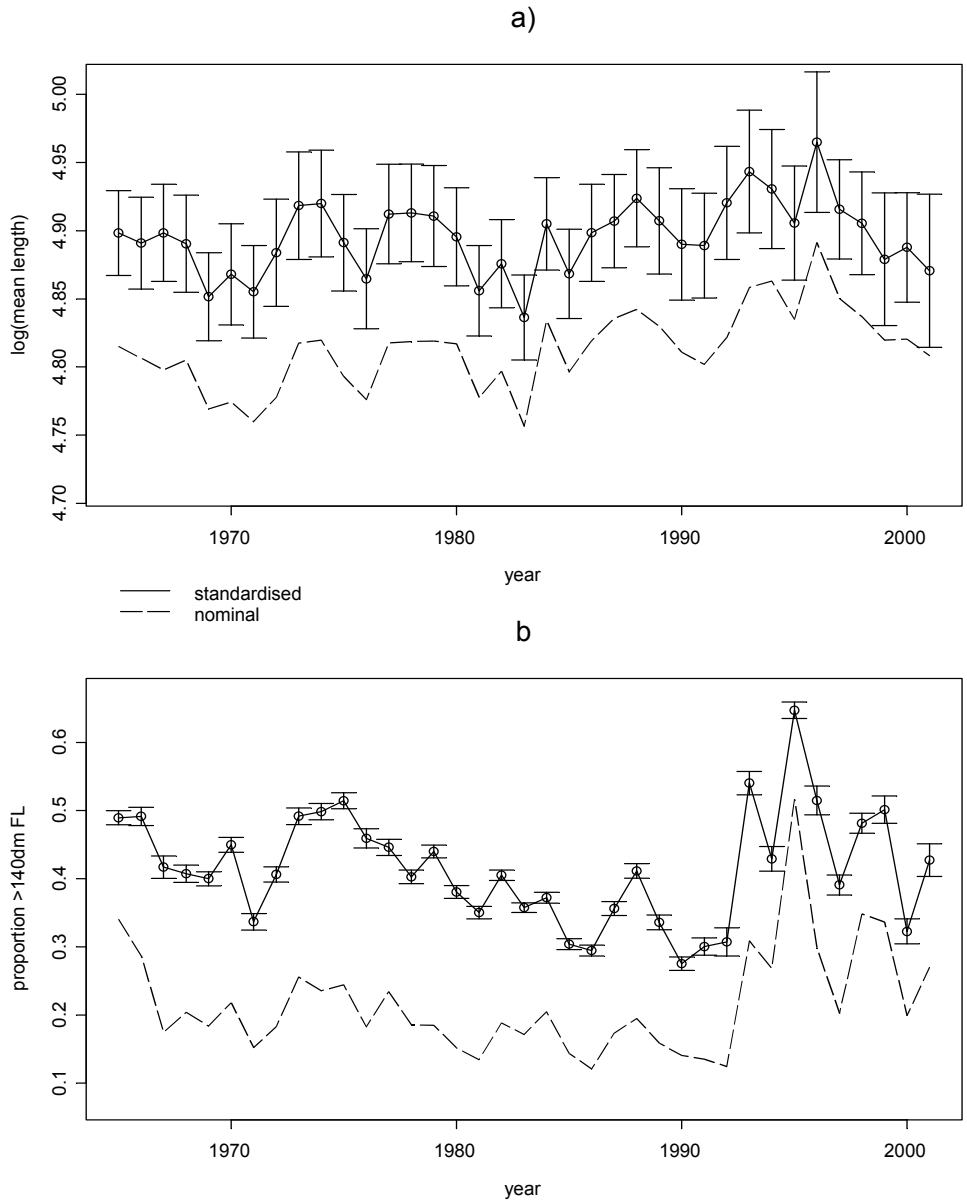


Figure 5: Bigeye, Japanese Longline. Time series of a) standardized and nominal bigeye mean length indices on the log scale, and b) standardized and nominal indices for proportion of bigeye >140cm FL. Error bars are ±1 s.e. The differences in magnitude are because the standardized indices are the predicted values for CPUE area 1 in quarter 1, while the nominal indices are the annual averages across all areas and quarters.

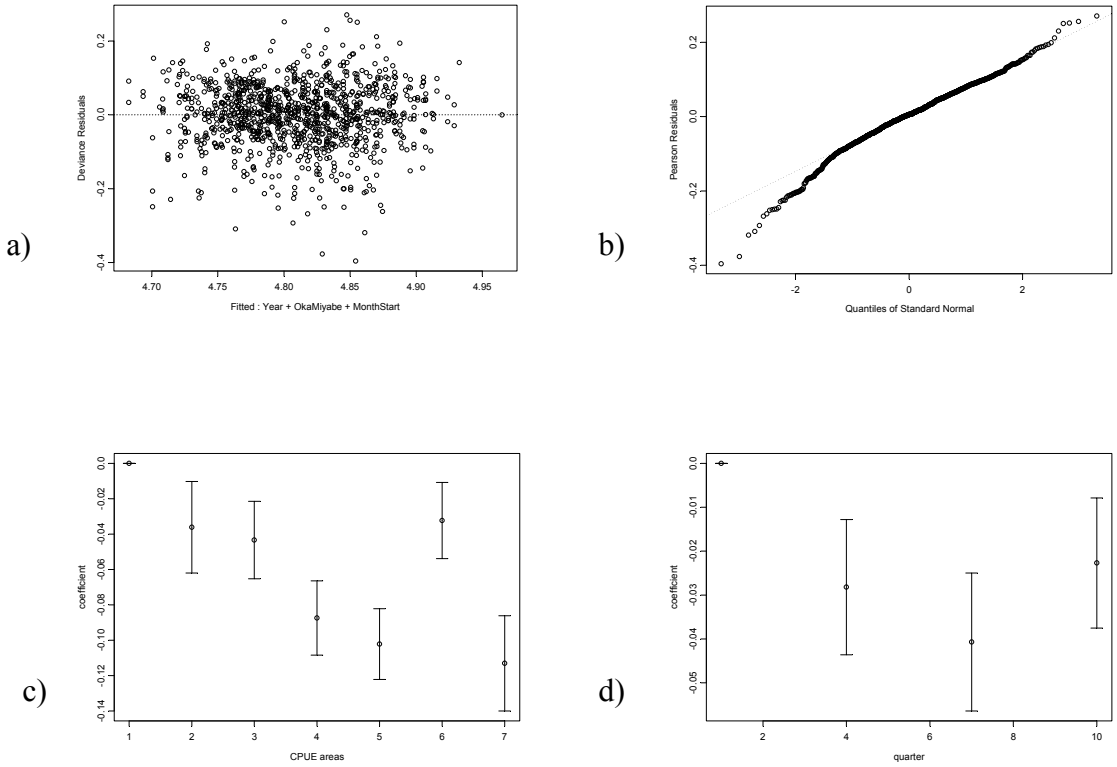


Figure 6: Bigeye, Japanese Longline. Plots of a) deviance residuals vs. fitted values, b) Pearson residuals vs. quantiles of standard normal, c) CPUE area coefficients and d) quarter coefficients, for the bigeye mean length main effects GLM. Error bars are ±1 s.e.

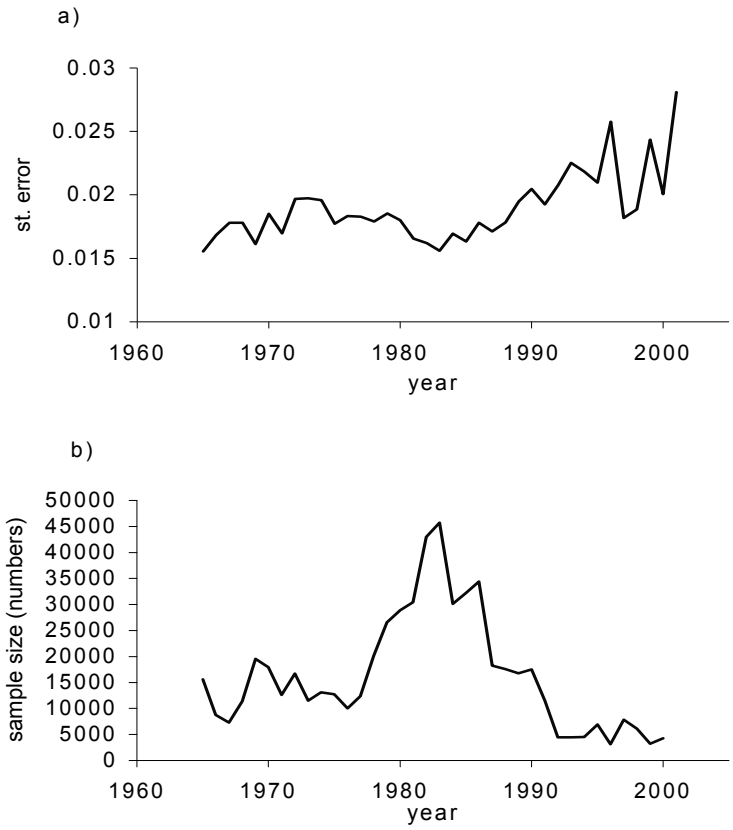


Figure 7: Bigeye, Japanese Longline. Standard errors associated with standardized indices for a) bigeye mean length and b) annual total sample sizes for bigeye mean length.

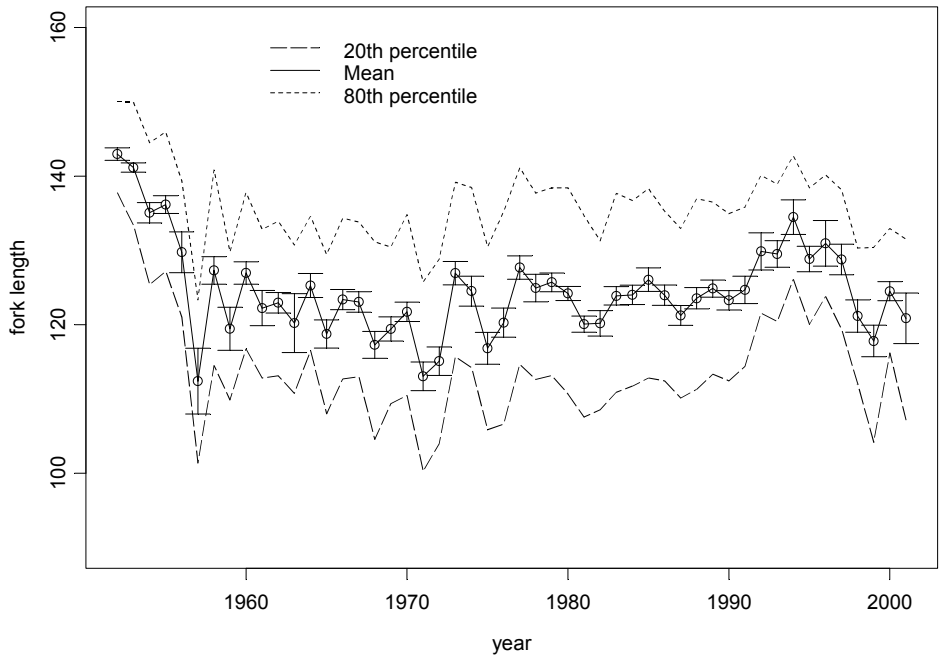


Figure 8a: Yellowfin, Japanese Longline. Annual averages of mean length and 20th & 80th percentiles (FL in cm) for the whole of the IO. Error bars are ±1 s.e.

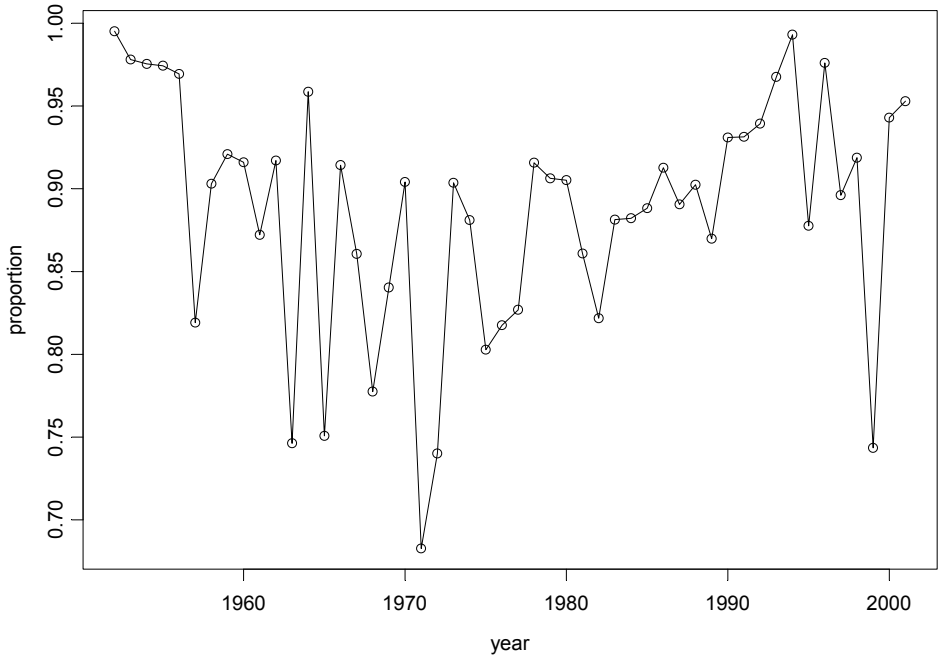


Figure 8b: Yellowfin, Japanese Longline. Annual proportion >100cm FL (FL in cm) for the whole of the IO.

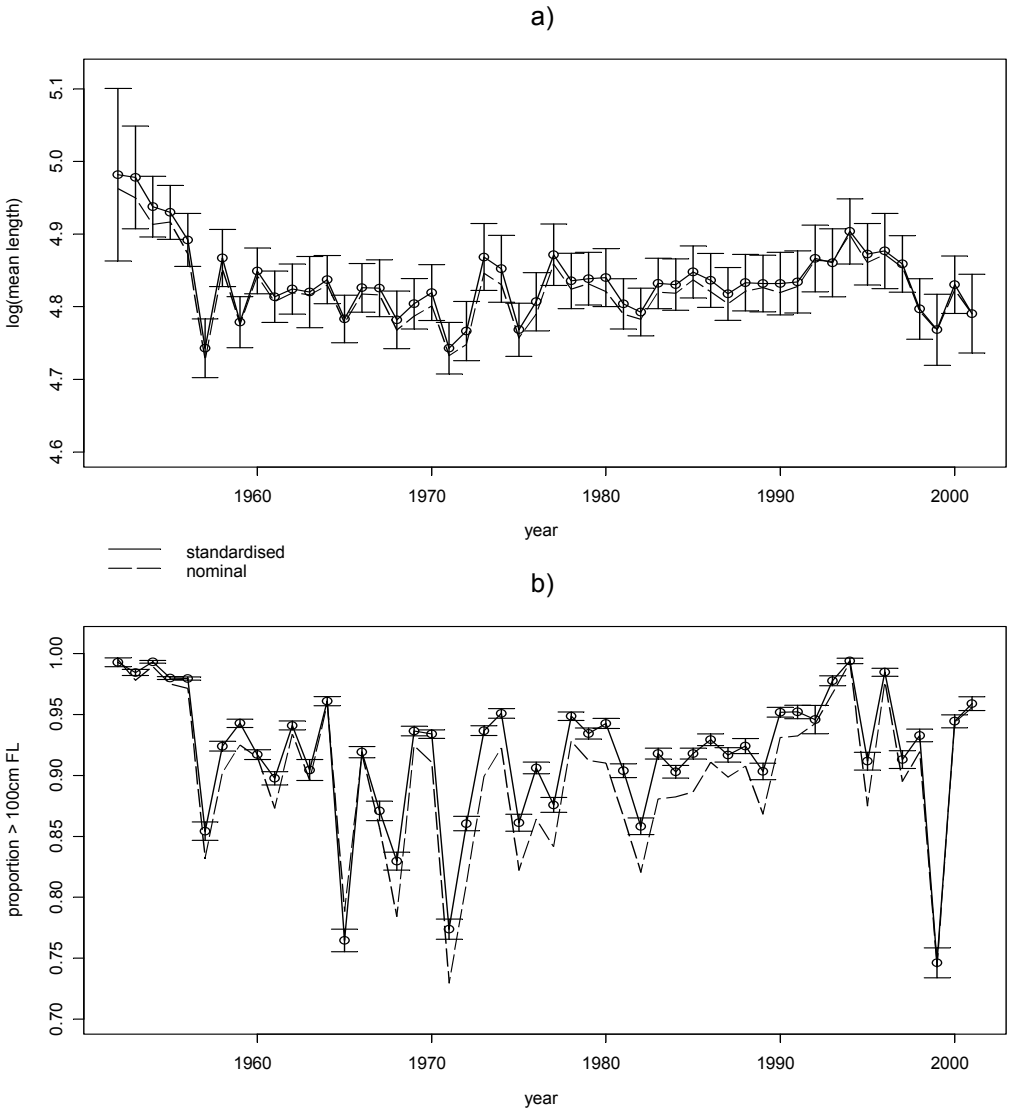


Figure 9: Yellowfin, Japanese Longline. Time series of a) standardized and nominal yellowfin mean length indices on the log scale, and b) standardized and nominal indices for proportion of yellowfin >100cm FL. Error bars are ±1 s.e. Note that the standardized indices are the predicted values in CPUE area 1 in quarter 1.

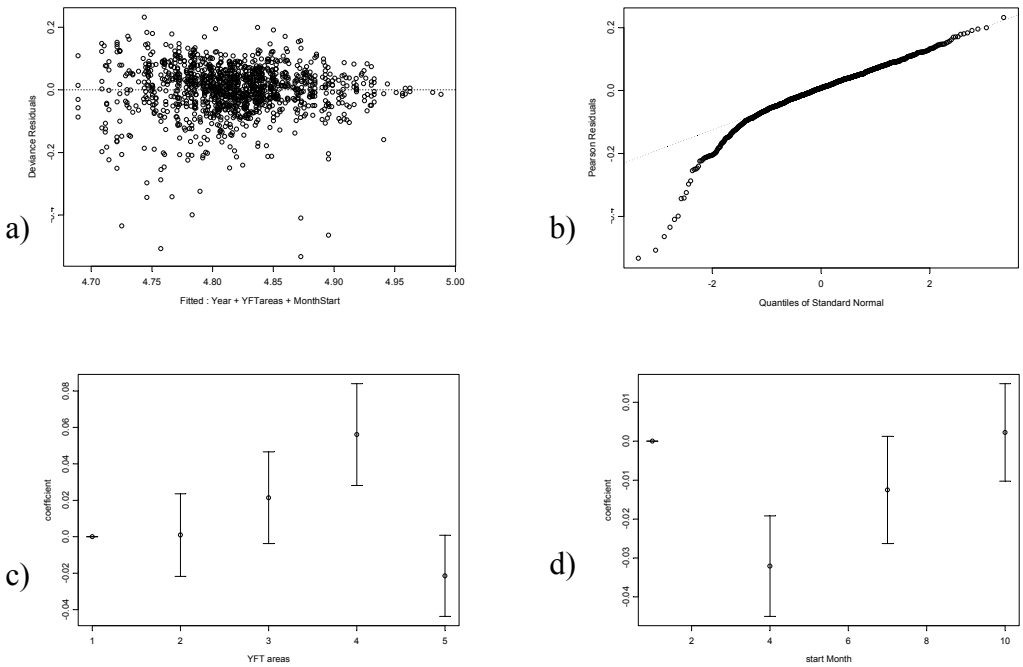


Figure 10: Yellowfin, Japanese Longline. Plots of a) deviance residuals vs. fitted values, b) Pearson residuals vs. quantiles of standard normal, c) CPUE area coefficients and d) quarter coefficients, for the yellowfin mean length main effects GLM. Error bars are ±1 s.e.

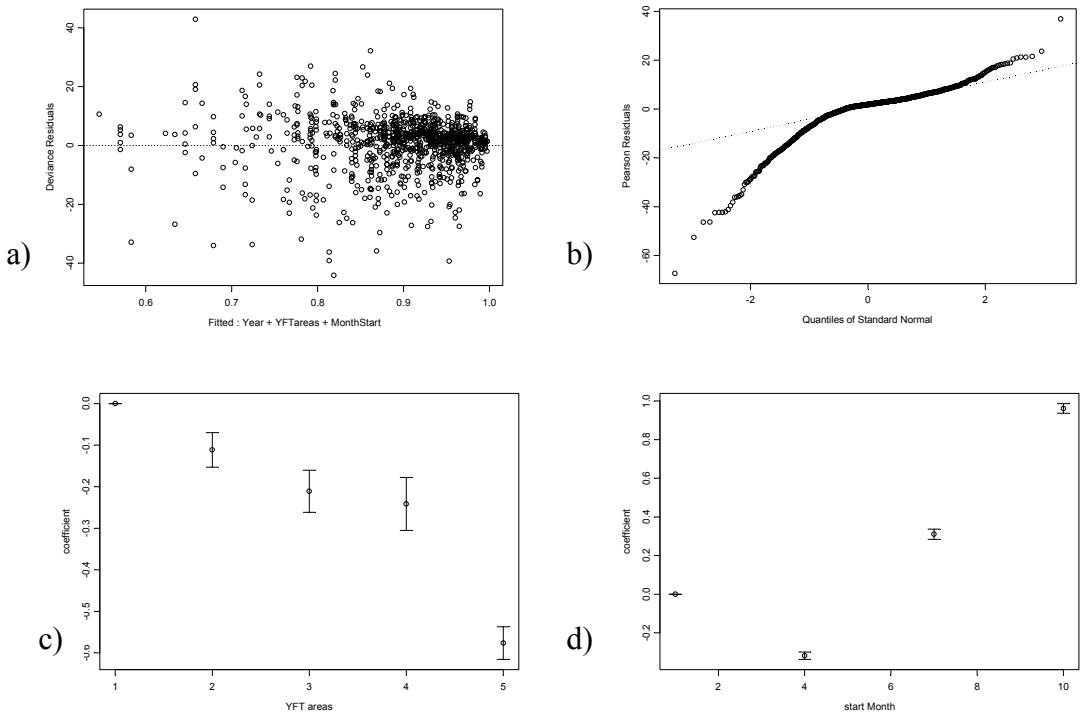


Figure 11: Yellowfin, Japanese Longline. Plots of a) deviance residuals vs. fitted values, b) Pearson residuals vs. quantiles of standard normal, c) CPUE area coefficients and d) quarter coefficients, for the GLM of proportion of yellowfin >100cm FL. Error bars are ±1 s.e.

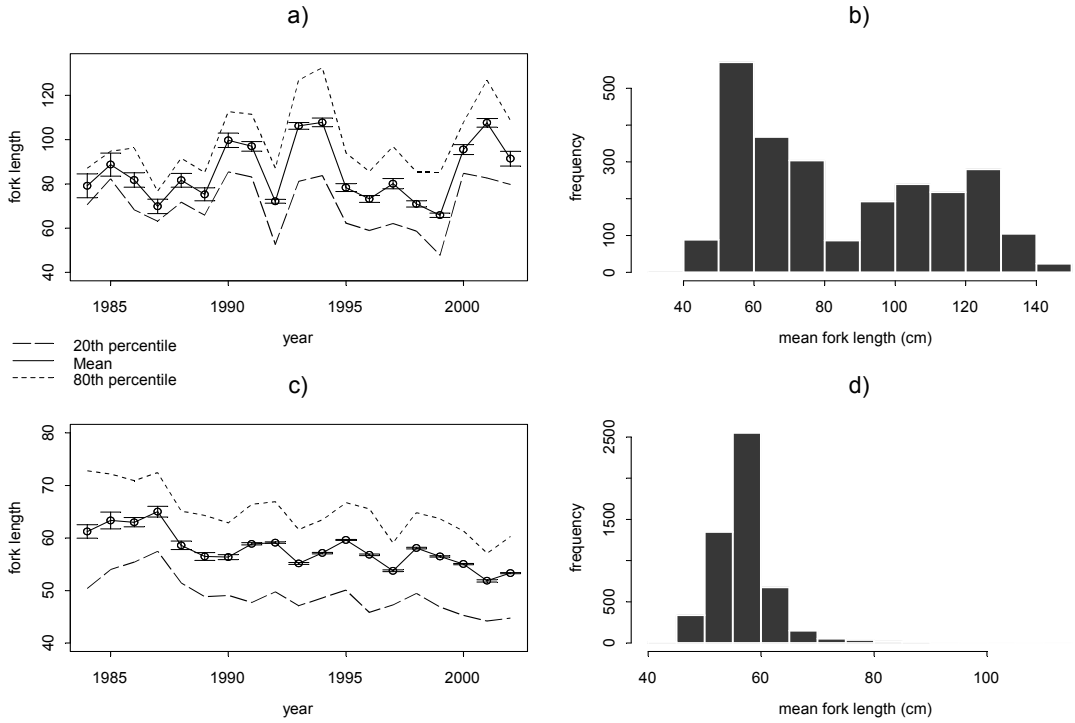


Figure 12: Bigeye, Purse Seine. Free school a) average annual mean, 20th and 80th percentiles of length, and b) histogram of mean lengths. Associated school c) average annual mean, 20th and 80th percentiles of length, and d) histogram of mean lengths. Mean lengths are calculated for each available month/year/5-degree square/fleet. combination for Spanish, French and NEI-EUR fleets, across all areas fished. Error bars are ±1 s.e.

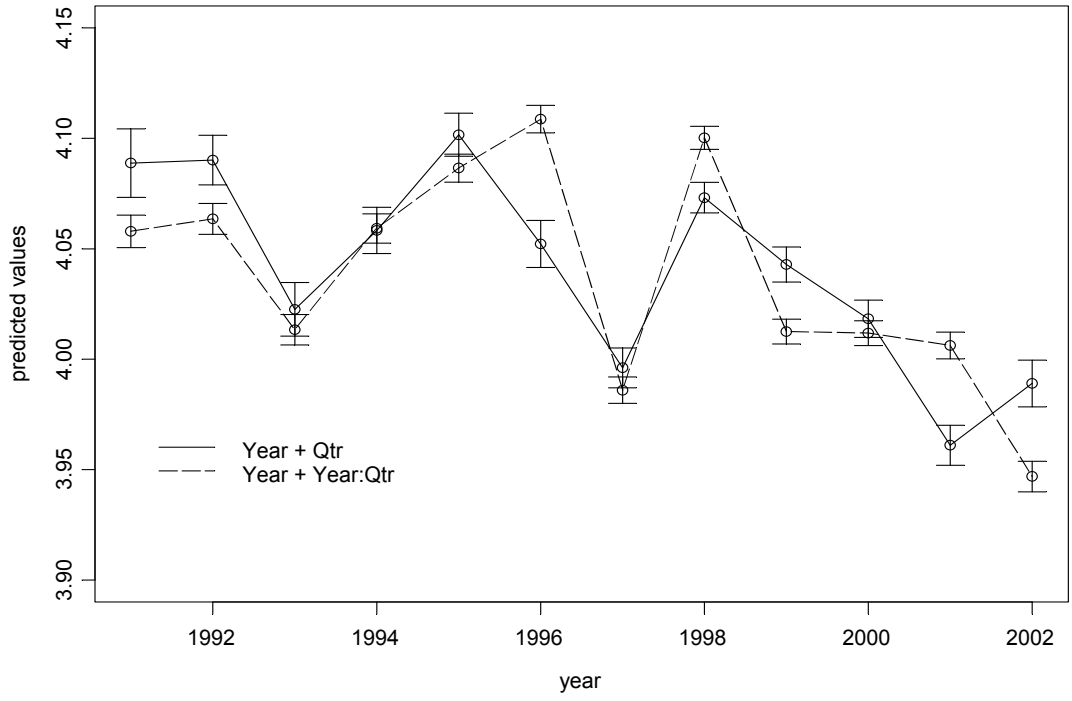


Figure 13: Bigeye, Purse Seine. Time series of standardized mean length indices, on the log scale, for each of the two models (see text). Error bars are ±1 s.e.

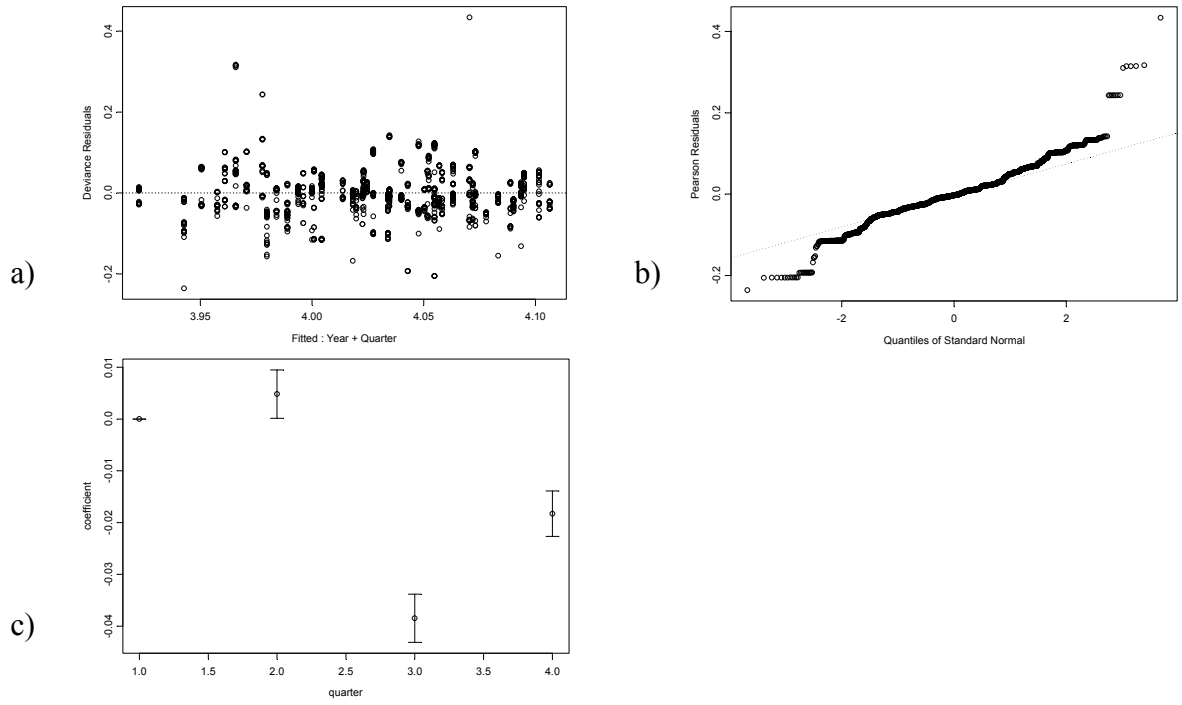


Figure 14: Bigeye, Purse Seine. Plots of a) deviance residuals vs. fitted values, b) Pearson residuals vs. quantiles of standard normal, c) quarter coefficients, for the bigeye GLM model 1. Error bars are ±1 s.e.

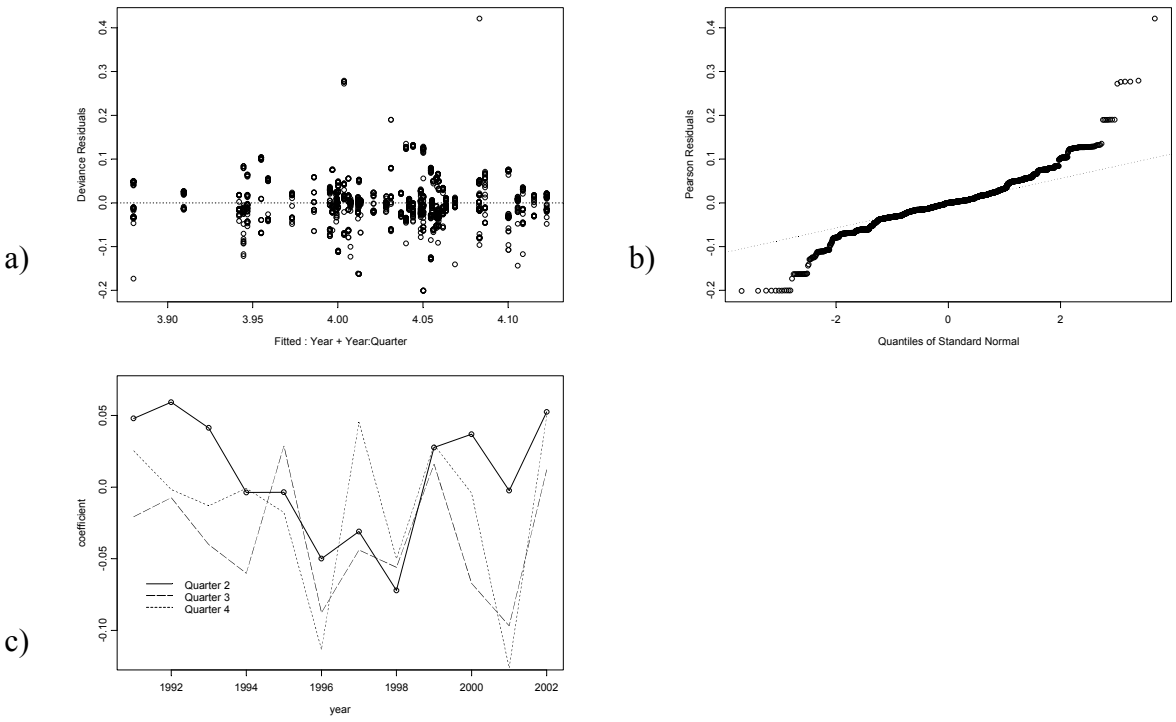


Figure 15: Bigeye, Purse Seine. Plots of a) deviance residuals vs. fitted values, b) Pearson residuals vs. quantiles of standard normal, c) quarter coefficients, for the bigeye GLM model 2. Error bars are ±1 s.e.

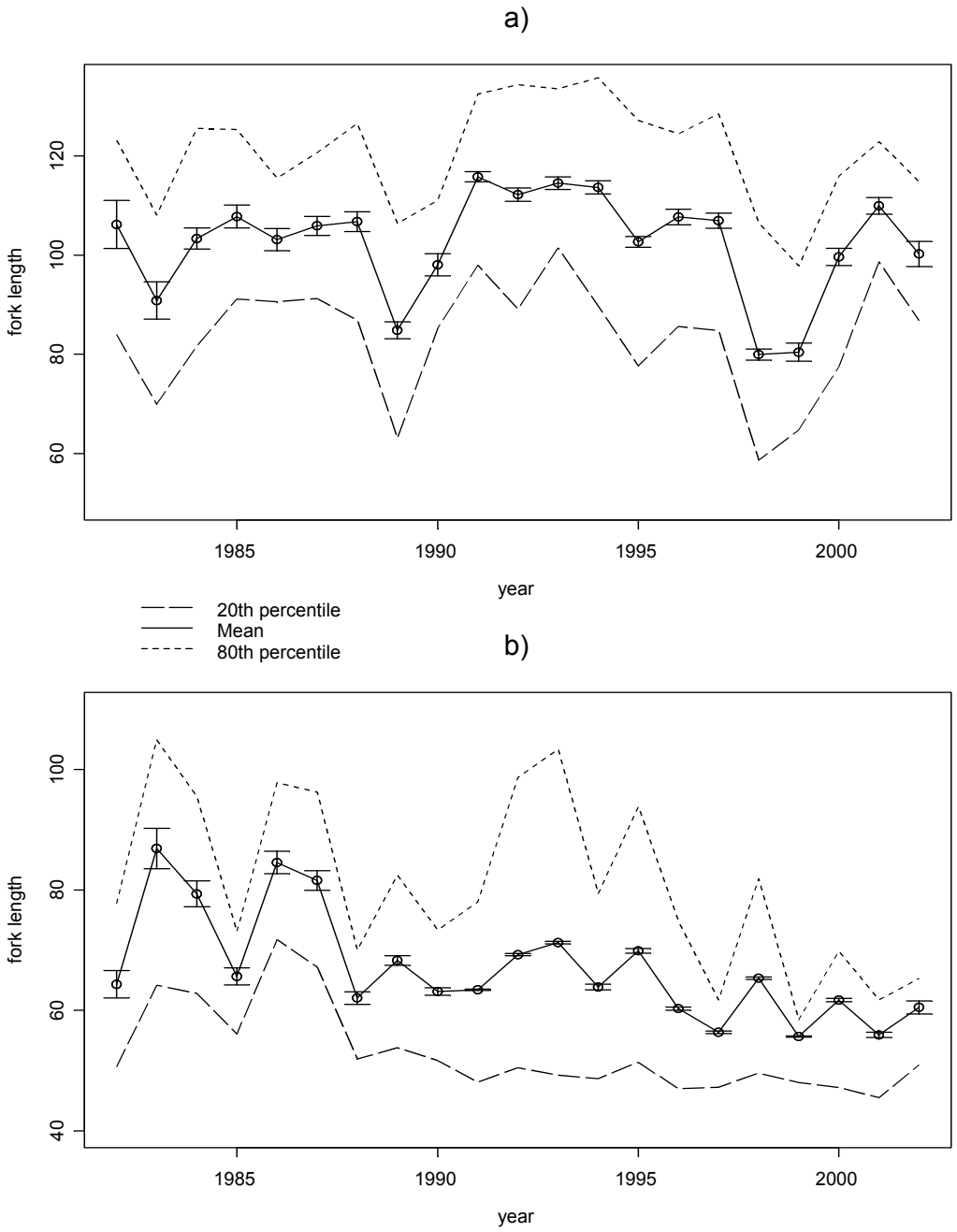


Figure 16: Yellowfin, Purse Seine. Annual averages of mean length and 20th & 80th percentiles (FL in cm) combined Spanish, French and NEI-EUR fleets for a) free schools and b) log schools. Error bars are ±1 s.e.

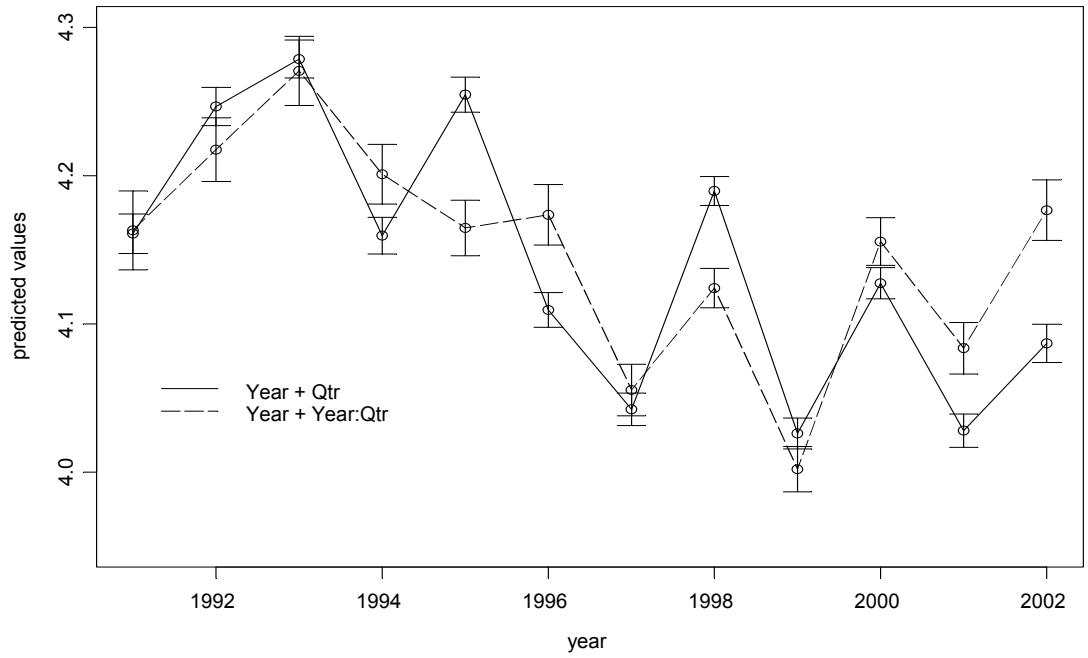


Figure 17: Yellowfin, Purse Seine. Time series of standardized LS mean length indices (on the log scale) for each of the two models (see text). Error bars are ±1 s.e.

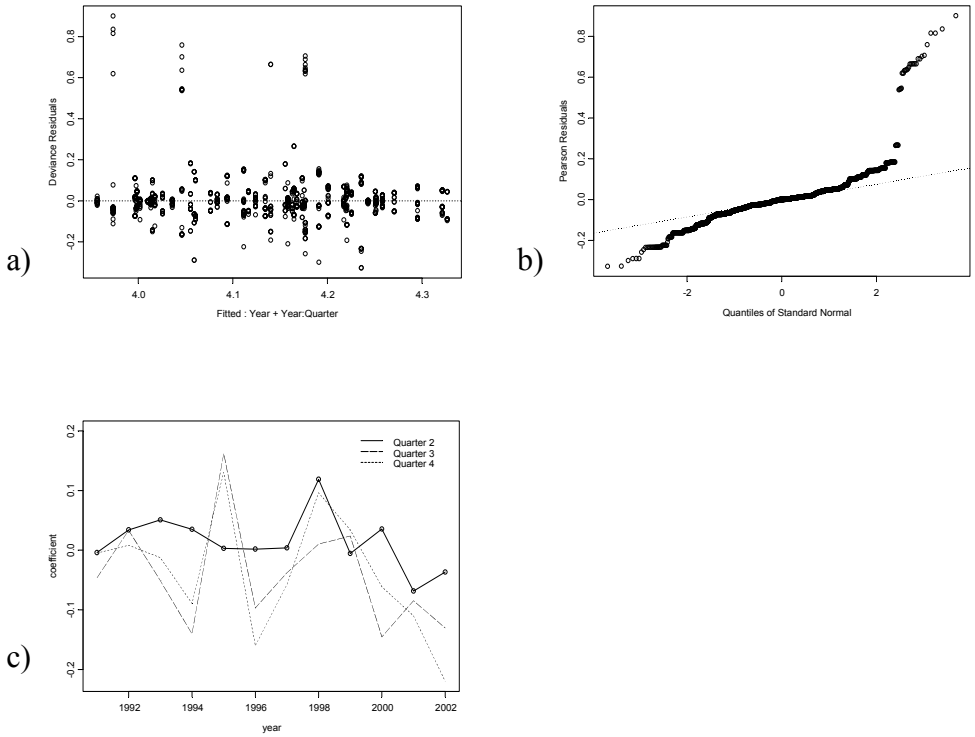


Figure 18: Yellowfin, Purse Seine. LS data. Plots of a) deviance residuals vs. fitted values, b) Pearson residuals vs. quantiles of standard normal, c) quarter coefficients, for the yellowfin GLM model 2. Error bars are ±1 s.e.

7.4. Attachment 2: Broadbill swordfish example

IOTC – 2004 - WPB - 07

STANDARDISATION OF SIZE-BASED INDICATORS FOR BROADBILL SWORDFISH IN THE INDIAN OCEAN

Natalie Dowling and Marinelle Basson
CSIRO Marine Research, Castray Esplanade
Hobart 7000, Tasmania
Australia

ABSTRACT

Information on the size distribution, for example mean length, in the catch can potentially be used as an indicator of swordfish stock status. This study looked at the standardisation of such indices particularly for spatial and seasonal effects. Standardization of size-based indices smoothed temporal trends that were evident in the nominal data, largely by accounting for the significant effect of area. Nominal peaks in the early and late 1990s were dampened, such that the standardized indices showed no overall temporal trend. GLMs indicated a strong effect of area on the size-based indices that predominated over the quarterly effects. This is to be expected from a species with size-based spatial segregation. The lack of a significant quarter effect was consistent with the lack of evidence of seasonal swordfish spawning or migrations. An increase in the inter-annual variability in the proportion of large fish since 1990 was a feature of the combined data. The increased variability was not smoothed by standardisation, and may be a feature warranting closer attention, or may simply be an artefact of lower sampling intensity. It was noted that the length-based swordfish indices did not show the pattern of decrease observed for the standardized CPUE trends in various sub-regions. These results contrasted with those from the Mediterranean Sea, where CPUE has remained stable but mean size has decreased.

INTRODUCTION

Large pelagic fisheries, such as the Indian Ocean longline fisheries for tuna and billfish, are typically data-poor in terms of fisheries-independent sampling and/or abundance surveys, with the consequence that stock status has to be indirectly inferred, whether from empirical indicators or formal stock assessment models using only fisheries dependent data. Of the main species caught on longlines in the Indian Ocean, broadbill swordfish (*Xiphias gladius*) are considered the most vulnerable to overexploitation, due to their relatively lower productivity, longer life span and hence time to mature, and their apparent higher viscosity (Campbell *et al.* 2003; Fonteneau and Richard 2003; Ward and Elscot 2000). Moreover, CPUE-based analyses have suggested that broadbill swordfish are almost fully exploited in longline fisheries across most oceans (IOTC 2003).

Attempts have been made to fit age-structured production models to Japanese and Taiwanese swordfish longline data in the Indian Ocean, but these have been relatively unsuccessful, largely due to a lack of contrast in, and/or contradictory catch-per-unit-effort (CPUE) indices of abundance, and a rapid increase in catch in the 1990s that could not be easily resolved by the

models (Yokawa 2001; IOTC 2001). Additional problems were that swordfish were generally not a target species of the Japanese and Taiwanese longline fleets, targeting practices shifted over time, the higher degree of residency suggests multiple stocks and requires appropriate area stratification (Yokawa, 2001; Campbell and Hobday 2003), and that there is a lack of biological knowledge about the species in the Indian Ocean.

Given the lack of feasibility of formal stock assessments for swordfish, simple indicators from fishery data are increasingly used to evaluate stock status. With the imminent introduction of Total Allowable Catch (TAC)-based management plans for the Australian domestic fishery, and the suspected vulnerability of swordfish to overfishing, there is a need to develop and test suites of stock indicators. As many different factors could be responsible for changes or trends in these indicators, it has been argued that they should ideally be standardised, in the same manner that commercial longline CPUE is standardised. If, for example, a stock is spatially disaggregated by size, changes in the timing and location of a fishery could lead to a change in the mean length in the catch that may be unrelated to a change in the underlying size distribution of the overall fishable stock.

In the face of data availability, obvious simple indicators additional to CPUE are those based on size-frequency data. Indeed, the Report of the 3rd Session of Indian Ocean Tuna Commission (IOTC) WP Billfish (2003) recommended that size-based indices for the whole fishery and by area should be considered. At high harvest rates large, old fish are generally the first to be depleted (ITOC 2003), and an absence or decrease of these fish could be reflected in the size frequency data. Monitoring size-based indices could therefore provide information on changes in the underlying population structure, as illustrated by Froese (2004) for cod stocks. We note that Punt *et al.* (2001) found that upper percentile weight for swordfish catch in the western Pacific was a more sensitive indicator that mean weight or length; however, only length data are available from the IOTC.

In this study we develop standardised length-based indices for swordfish in the Indian Ocean, and contrast these with results from CPUE standardisations. The mean, median and upper 80th percentile length from the length-frequency data are considered. Due to the suspected higher residency of swordfish relative to the tunas, we also considered trends within sub-regions that showed patterns of decline in standardised CPUE.

This work is part of a wider study on stock status indicators, that will use simulation to investigate the robustness and sensitivity of a range of indicators, and to develop ways of combining several such indicators into a coherent framework for management decision-making. The study is addressing these issues primarily in the context of the domestic (Australian) fishery, but also in the wider context of the IOTC. Simulation work may indicate that mean length is insensitive to stock changes (see e.g. Punt et. al. 2001), and that some other measure, possibly even based on data not currently available for the IOTC fisheries, may be more reliable. Hence the current results should be viewed in this wider context as an exploration of the effects of, and need for, standardisation of indices based on existing data.

METHODS

Data

Quarterly, area-specific (on the scale of five-degree squares) lower nth percentiles, medians, means and upper nth percentiles were derived from the raw size frequency data held by the

IOTC. Eye-fork length (EFL) data from the Japanese fleet was used since this contained the most records over the most number of years. We also considered the proportion of fish greater than 150cm EFL, chosen to reflect approximate size at maturity (Yabe *et al.* 1959; Nakano and Bayliff 1992; Di Martini 1999, Young and Drake 2002)

As with yellowfin and bigeye tuna (Basson and Dowling 2004), the similarity between annual values of mean, median, lower 20th and upper 80th percentile was such that standardisations were applied only to the mean length, and to the proportion of large fish. The proportion of small fish could also have been used as an aggregate index of recruitment, but was not considered here.

Standardisation technique

The methodology applied was analogous to that for bigeye and yellowfin tuna (Basson and Dowling 2004). General linear models were used for standardization, with time and area effects as the main covariates.

Spatial factors were considered for two reasons: 1) to account for changes in the fishing locations over time and/or 2) to account for changes in the size frequency in different areas over time. The second reason could be important for swordfish, which is now thought to have a potentially high residence time (Campbell and Hobday 2003). Different levels of fishing intensity in different areas could therefore have different effects on the mean size in the catch over time. This type of change is more likely to be observed at a relatively coarse spatial scale than a very fine spatial scale.

Incorporating environmental factors into this type of standardisation is considered inappropriate unless there is evidence of a plausible mechanism for catchability changing differentially by size under different environmental conditions. Otherwise, area or time factors should account for changes in size distribution associated with temporally or spatially varying environmental conditions (see Basson and Dowling (2004) for further detail).

Models for mean length and "proportion large"

Multiplicative models were used for mean length, with year and quarter as the main factors. The log of mean length is therefore an additive model of the relevant factors and we assume that errors are normally distributed. Several different spatial scales were considered. At the fine scale level, we used 10-degree latitude bands and the (5-degree resolution) lat-long grid positions for each quarter. On a broader scale, we considered the catch rate standardization areas (IOTC 2003) (Figure 1), with the notion that the areas were assigned on the basis of common fishery and/or other characteristics.

We also considered the Longhurst areas (Longhurst 2001), which supposedly represent "habitats" and may therefore contain fish of certain age or size classes, if distribution is driven by habitat preference.

During the exploratory phase, more complex models with main effects and interactions, for example, between year, quarter and area were fitted and compared to more parsimonious models. S-PLUS software was used and the model definitions were of the following forms:

Model type 1:glm(log(mean length) ~ year + area + quarter)

Model type 2 : glm(log(mean length) ~ year + area*quarter)

We also generated area-specific standardized size indices to enable comparison with the area-specific CPUE indices (IOTC 2003). Three areas in which Japanese CPUE had declined, and a "reference area" in which CPUE showed no temporal trend, were examined. Factors for the within-area models were year and quarter.

The 'proportion large' was fitted with a binomial model with logit link and factors were again time and the four different area definitions. The data were weighted by sample size, and the models were fitted to the subset of the data where sample size was greater than 10 (approximately the lower quartile of a summary of sample sizes across the whole data set).

Model selection

In fitting GLMs to the two indicators, the main aim is to eliminate changes due to seasonal effects or different locations of fishing, and to extract the year effects as standardized indices, rather than to determine statistically whether factors are significant. However, the goodness of fit of different models should be compared to avoid under- or over-fitting. Deviance residuals and q-q plots were used to evaluate goodness of fit, homogeneity of variance and extent of conformance to a normal distribution. Where models were nested, appropriate significance tests were used to compare them, but in the case of the binomial models, these are unlikely to be reliable because of over-dispersion. Where models were not nested, particularly where the different area-designations were used, Akaike's information criterion (AIC) was used. The main priority, however, was to obtain the year-effects and determine whether they were sensitive to the different model formulations or variables included.

RESULTS

Nominal trends and exploratory analyses

The mean lengths by area, quarter and year were highly variable. The annual averages of mean length were also very variable, ranging between about 132 and 162cm EFL. Although there appear to be increases and decreases in the annual average series, and in particular, an overall increase from 1978 to 1999, with decreases in 1995 and since 1999, these should be seen within the context of the range of the size frequency. Figure 2a shows the annual averages (over location and quarter) of the mean and of the 20^{th} and 80^{th} percentiles. Changes in the mean (and median) time series were small relative to the inter-percentile range. Similarly, the decrease in swordfish mean size over the last 2 years appears relatively insignificant relative to the long-term pattern, and the size range. The inter-quantile range has not changed substantially over the whole period, varying between 31 and 56cm with no systematic temporal pattern.

The pattern of the nominal time series plot for the proportion of fish >150cm EFL was similar to that of mean length (Figure 2b), as would be expected. There were increases in the proportion of large fish in the early and late 1990s to historically high levels (to greater than 70% "large" fish), followed by a decreases to the pre-1990 levels.

Exploratory analysis was undertaken to investigate the effect of alternative temporal scales and area designations on the inter-annual trends. This included comparing nominal time series broken down by quarter and area to find any obvious patterns, and comparing standardised indices across alternative models. The following points summarise the minor findings from the exploratory analyses:

- Plots of mean length (and the other candidate indicators) over time and by the different area designations did not show any clear patterns to inform model choice. Plots were generally noisy, with missing data and much inter-annual variability.

- Figure 3 shows the mean length GLM year coefficients for 5 alternative model types/area designations. For both mean length and "proportion big", the relative pattern of the year coefficients was very similar between model types and choice of area designation. The exception was the model using the Longhurst area, which did not appear to smooth/reduce the nominal increases in mean length observed in the early and late 1990's. The Longhurst area of the Indian South Subtropical Gyre province (Longhurst 2001) incorporates the CPUE standardization areas 5 through to 9 (Figure 1). For GLMs using the CPUE areas, area coefficients were high for area 8, intermediate for areas 7 and 9 and low for areas 5 and 6 (Figure 5c). The larger Longhurst area perhaps could not reconcile these different effects within one area and standardize for them appropriately. Finally, although models with interaction terms sometimes gave a better fit to the data, the standardised indices were essentially unchanged.

- Models using the CPUE standardisation areas generally gave lower AIC values compared to those using other area designations. Therefore, only the main effects models using the CPUE standardisation areas are subsequently considered in detail. These models were also convenient in allowing for direct comparison with standardised CPUE trends.

GLM standardisation of mean length and proportion >150cm EFL

Results are presented for mean length modelled as:

glm(log(mean length) ~ year + CPUEarea +quarter).

This model has a common spatial pattern for each year and a common quarterly pattern in each year and area. The standardized indices generally showed a similar temporal trend to the nominal pattern, except that the relative increases in nominal mean length in the early and late 1990s were dampened (Figure 4a). The standardised indices for the proportion of swordfish > 150cm EFL, using the same factors as the mean length model, showed no overall increase from 1980 to the early 1990's, as was apparent in the nominal trend (Figure 4b). Standardised values varied about consistently higher proportions than the nominal values, particularly prior to the mid 1990s. The standardisation did not smooth the increase in inter-annual variability in the proportion of large fish that was observed since 1990.

Figure 5 shows diagnostics and results for the mean length model. Residuals did not show any systematic patterns, either overall or when plotted against each set of predictor variables. The q-q plot showed that the data were almost normally distributed, with deviation mainly at the lower tail. These large negative residuals persisted in models with interactions. The area coefficients were estimated relative to CPUE area 4 because this area had high levels of sampling over the whole time period (see Table 1). Note that area 8, in the middle of the southern Indian Ocean (Figure 1), had the highest-value coefficient, but this was based on a single, 1982 sample; hence the high standard error. Area 4, encompassing Indonesia and the northeast Indian Ocean south of the Bay of Bengal (Figure 1), has the lowest-value coefficient (Figure 5c). The quarter coefficients were small in value relative to the area coefficients, indicating a lesser relative effect on mean length (by approximately an order of magnitude). The highest-value quarter coefficient was in quarter 3, and the lowest was in quarter 2 (Figure 5d). The results imply a 7cm difference

between mean length in quarter 2 versus quarter 3, and a substantial 60cm difference between mean length in area 4 versus area 8.

The trends in the area coefficients reflect the size segregation by latitude that has been observed for swordfish. Spawning occurs in tropical/sub-tropical regions where surface temperatures exceed 20°C, that is, in latitudes that rarely extend north of 35N or south of 35S (Ward and Elscot 2000). Juveniles are confined to these regions for at least their first year, and migrate to higher latitudes as they grow (Campbell and Miller 1998). Females attain larger sizes than males and occupy the highest latitudes of the range (Campbell and Miller 1998). In the northern Pacific, the average size of swordfish in longline catches increases with latitude for both sexes (DeMartini 1999). This is reflected in the values of the area coefficients of the mean length model, with these being largest for the higher latitude areas (Figure 1; Figure 5c). In the Indian Ocean, the greatest concentrations of larvae have been found in the eastern Indian Ocean, off of north-western Australia and south-west of Java (Nishikawa and Ueyanagi 1974), which is consistent with the adjacent area 4 having the lowest area coefficient. Larvae have also been reported in the Mozambique Channel, and east of Madagascar (Palko *et al.* 1981), and in equatorial waters (Nishikawa et al. 1985). Correspondingly, area coefficients were low in areas 3, 5 and 6, which encompass the equator and Madagascar.

Residuals for the model of proportion of swordfish >150cm EFL (again fitted relative to CPUE area 4) were also reasonably homoscedastic and close to normal (Figure 6), and showed no pattern when plotted against each set of predictor variables. Area 8 had no data (due to the sample size restriction on the model), so the value of the coefficient was automatically zero, with no standard error. Of the areas for which there was data input to the model, areas 7, 9, 10 (all south of 20°S) and 1 (north of 10°N) had the highest value coefficients, and hence higher proportions of large fish, while areas 4 and 5 (the latter in the western Indian Ocean encompassing northern Madagascar) had the lowest (Figure 6c). The lowest-value quarter coefficient was in quarter 2, while quarters 3 and 4 had the highest-value coefficients. Hence large fish appear to be caught in greater proportions in the latter half of the year (Figure 6d). As with mean length, the quarterly coefficients were of an order of magnitude lower value than the area coefficients, indicating a lesser effect of quarter on the proportion of large fish in the catch, relative to area.

As with the mean length, the area coefficients for the proportion large model reflect the size-differentiated spatial distribution of swordfish. The lowest proportion of "large" fish by area was lowest in the areas for which larval concentrations are highest, while the greatest proportions of "large" fish were associated with higher latitude regions, consistent with the observation that swordfish progressively migrate towards higher latitudes as they grow.

The standard errors of the standardised indices tended to be inversely related to sample size, in numbers of fish (Figure 7). This effect was more marked when weighting by sample size was used in the models. There is therefore less relative certainty associated with the standardised trends from more recent years, and this will have an effect on how such a series is implemented as a "stock status indicator". This comment also applies to the other standardised indices presented below.

Within-area analyses

At the 3rd Session of the IOTC Working Party on Billfish (2003) catch rate standardizations were undertaken on longline data. Relatively large decreases in Japanese CPUE were noticed in areas 3, 5&6 and 7 since 1990, which coincided with the timing of an increase in catch (Figure 10). It

should be noted also that areas 3 and 7 are those from which the most catches are taken (IOTC 2003). In contrast, the standardized CPUE series from area 4 showed no temporal trend with the possible exception of the last couple of years. In terms of the length-frequency data, Table 1 shows that the highest length-frequency sample sizes over the most continuous time series were also taken from areas 3, 4, 6 and 7.

To see whether similar patterns of decrease were observed using size-based indices, we constructed separate standardized indices of mean length and proportion large for areas 3, 6 and 7. Area 5 was not considered because of a lack of data. We also constructed indices for area 4, chosen as a "reference" area in which CPUE had not declined. The GLMs considered above suggested that the mean length and proportion of large fish was lowest from area 4, moderately low for areas 3 and 6, and high for area 7. It is re-iterated that these area-specific GLMs are standardizing only for the effect of quarter.

The standardized mean length indices for each area showed very similar temporal trends to the nominal patterns (Figure 8). Only the standardised trend for area 4 showed any appreciable differences in some years to nominal mean length pattern (Figure 8), as this was the only area-specific model for which quarter was a significant factor (ANOVA, $P < 0.005$), with lower-value coefficients in quarters 2 and 4 (Figure 11). Figure 8 shows that there were essentially no temporal trends in standardised (or nominal) mean length in any of these areas, except for a sharp drop in mean size in area 6 in 2000 and 2001.

Figure 9 shows the standardized and nominal proportions of swordfish greater than 150cm EFL. Again, the standardized patterns are very similar to the nominal trends, indicating that the quarterly pattern in each year is not strong. The greatest differences between the nominal and standardized values were for some years in areas 6 and 7.

None of the areas showed even vaguely similar patterns to the CPUE trends. The only consistent feature across CPUE and the proportion large indices was the peaks in the mid to late 1980s in areas 3 and 7 (Figure 9).

The proportion of "large" fish caught in area 3 showed the highest inter-annual variability of the four areas, with the "proportion large" varying between about 90% and less than 5%(Figure 9a) In area 4, the "proportion large" ranged within 15% and 70%. (Figure 9b). Overall, the proportion of large fish was generally lower (in area 4) relative to the other areas, consistent with the low-value coefficient for this area from the main model (Figure 6c). In area 6, between 1970 and 1992, the proportion of large fish was relatively constant (between 30% and 60%), but has varied much more since 1990. The proportion dropped to less than 20% in 1992 and subsequently peaked above 70% in 1993 and 1994 (Figure 9c). Finally, the proportion of large fish in area 7 showed high inter-annual variability. With the exception of 1983 and 1984, the proportion was greater than 50% (Figure 9d). This is consistent with the result that area 7 had one of the highest-value coefficients in the main model (Figure 6c).

Trends in the quarterly coefficients for the area-specific mean length and proportion large models were not particularly consistent across areas, with area 4 showing the most difference (Figure 11). This could be interpreted that there is an interaction effect between area and season (i.e. that the effect of season is different for each area). It should be recalled that areas 3 and 4 straddle both hemispheres, while areas 6 and 10 are in the southern hemisphere, such that a quarter of the year corresponds to opposite seasons for each hemisphere, so perhaps the inconsistencies should be expected. However, the inclusion of an area-quarter interaction in the main GLMs for mean length and "proportion large" was not statistically significant at the 5% level although it was at the 10% level for the mean length model (ANOVA; $P = 0.068$) (note also that quarter as a main

effect was marginally insignificant in the mean length model; P=0.057). Moreover, the inclusion of an interaction term had a negligible effect on the values of the year coefficients (see Figure 3 as an example for the mean length model).

DISCUSSION

As with bigeye and yellowfin tuna (Basson and Dowling 2004), the results presented here for swordfish show that it is possible to standardise the mean length and other similar indices reasonably successfully. In particular, and as opposed to bigeye and yellowfin standardisations, which showed little difference to the nominal patterns, the standardization of size-based indices for swordfish smoothed temporal trends that were evident in the nominal data, largely by accounting for the significant effect of area. Standardization of swordfish mean length and proportion "large" tended to dampen nominal peaks in the early and late 1990s, such that the standardized trends showed no overall temporal trend. While mean length did not appear to be a sensitive or reliable indicator for bigeye or yellowfin tuna (Basson and Dowling 2004), simple yield per recruit analysis indicates that the largest change in mean length as fishing mortality increases is likely to occur for slow growing species. This suggests that mean length may be a poor stock status indicator for yellowfin, which is relatively fast growing, but it could be reasonable for a species like swordfish. However, selectivity patterns can also influence mean length and render it misleading, particularly if selectivity patterns change over time.

The GLMs indicated a strong effect of area on the size-based indices that predominated over the quarterly effects. This is to be expected from a species with higher residency, and, more particularly, size-based spatial segregation, where fish migrate to progressively higher latitudes as they grow, and larval concentrations are highest about the equator and in specific regions. The lack of a significant quarter effect is possibly also unsurprising, since there is no evidence of seasonal spawning or migrations for swordfish.

An increase in the inter-annual variability in the proportion of large fish since 1990 was a feature of the combined data, and of the temporal trends from CPUE standardisation areas 4 and 6. Simultaneously, it was noted that the greatest increases in catch did not occur in areas 4 and 6, but rather in areas 3 and 7 (Figure 10). The increased variability was not smoothed by standardisation, and may be a feature warranting closer attention (e.g. sample size issues or differences in sex ratio, given that females grow to larger sizes than males (Ward and Elscot 2000)). Typically, a continuous decline in the value of a stock status indicator over a time series alerts decision makers to potential problems, but the onset of, or an increase in instability could be of equal concern. That the increase in variability occurred in areas with lower increases in catch may suggest that closer attention needs to be paid to more marginal areas of the fishery, though it could simply be an artefact of lower sampling intensity.

The standardization process yields useful by-products: most importantly, estimates of standard error that are likely to be more robust than direct estimates from the raw data. When implementing such an index in a monitoring framework the standard errors would play a strong role in assessing whether change in indices are substantial enough to warrant management action. The increase in standard errors when sample sizes decreases was noted, and would have to be taken into account when implementing such indices into a monitoring or management framework.

The standardisation also allows for the integration of data in a statistical framework and for 'like with like' comparisons in different areas or in the same area in different quarters. The standardisation can reduce the effects of 'noise' in the data. This does not mean that careful

scrutiny of the detailed data is unnecessary. Particularly when changes are observed, there would be a need to try to understand why and where those changes are occurring.

Since swordfish are often caught together with bigeye tuna, there is value in comparing the standardized size-based indices for the two species. Referring to Basson and Dowling (2004), the trends in standardized mean length for bigeye tuna showed no similarity to those for swordfish, and the only similarity in the "proportion large" indices was a peak occurring in the mid 1990s. Quarterly coefficients for mean length showed opposing patterns, with bigeye having the lowest quarterly coefficient in quarter 3, which yielded the highest value coefficient for swordfish.

The length-based swordfish indices did not show the pattern of decrease observed for the standardized CPUE trends in areas 3, 6 and 7. There are many possible explanations for this observation, and it is not currently feasible to say which mechanism(s) may be operating. It may, for example, suggest that standardised CPUE is a more sensitive stock status indicator, or that the abundance of all sizes reflected in the CPUE is changing in a similar way, so that means size remains relatively constant. Areas with CPUE declines could be areas with slow replenishment times. With few new recruits coming in to these areas, CPUE could decline while the size of the catch could remain relatively constant. Indeed, area 4, the "reference" area showing no apparent decline in CPUE, had the lowest mean length coefficient value and one of the lowest proportion "large" coefficients, and lies close to an identified area of high larval concentration. Area 4 could thus be seen as a recruitment area that is thereby less likely to show declines in catch rates relative to other heavily fished areas.

It is interesting to note that the findings for standardized CPUE and size-based indices for the Indian Ocean contrast with those from the Mediterranean Sea, where CPUE has remained stable but mean size has decreased. This combination of indicator trends could perhaps have more serious biological implications than what is observed in the Indian Ocean, since it hints towards hyperdepletion, whereby an aggregating stock or highly efficient fleet allows catch rates to be maintained despite declines in overall abundance. Additionally, a decrease in mean size despite constant catch rates has potential implications for reproduction and recruitment, since fecundity is positively related to size.

ACKNOWLEDGEMENTS

This work was supported by funding from the Australian Fisheries Research Development Corporation (project 2003/042) and CSIRO. We thank the IOTC Secretariat, Miguel Herrera in particular, for assistance with the data and for prompt responses to our questions. Any misinterpretations of the data are entirely our fault.

REFERENCES

Basson, M. and Dowling, N. (2004). Standardisation of size-based indicators for bigeye and yellowfin tuna in the Indian Ocean. IOTC-2004-WPTT-15

Campbell, R.A. and Hobday, A. (2003). Swordfish-Environment-Seamount-Fishery Interactions off eastern Australia. Report to the Australian Fisheries Management Authority, Canberra, Australia.

Campbell, R.A. and Miller, R.J. (eds.) (1998). Fishery Assessment Report 1997: Eastern Tuna and Billfish Fishery. Australian Fisheries Management Authority, Canberra, Australia.

Campbell, R.A., Basson, M. and Dowling, N. (2003). Review and analysis of information for determination of Total Allowable Catches and decision rules for the Southern and Western Tuna and Billfish Fishery. Report to the Australian Fisheries Management Authority, Canberra.

DeMartini, E. (1999) Size at maturity and related reproductive biology. Pp. 161-169 in DeNardo, G.T. (ed.) Proceedings of the Second International Symposium on Swordfish in the Pacific Ocean. Southwest Fisheries Science Centre Administrative Report NOAA-TM-NMFS-SWFSC-263. National Marine Fisheries Service, Honolulu.

Fonteneau, A. and Richard, N. (2003). Relationship between catch, effort, CPUE and local abundance for non-target species, such as billfishes, caught by Indian Ocean longline fisheries. Marine and Freshwater Research 54: 383-392.

Froese, R. (2004). Keep it simple: three indicators to deal with overfishing. Fish and Fisheries 5(1): 86-91.

Indian Ocean Tuna Commission (2003). Report of the 3rd Session of the IOTC Working Party on Billfish. Perth, Australia, November 10-12, 2003

Indian Ocean Tuna Commission (2001). Report of the 2nd Session of the IOTC Working Party on Billfish, St. Gilles, La Reunion, November 5-8, 2001

Internation Commission for the Conservation of Atlantic Tunas (2003). Report of the 2003 ICCAT Mediterranean Swordfish Stock Assessment Session. Col. Vol. Sci. Pap. ICCAT 56(3):

Longhurst, A.R. (2001). Ecological geography of the sea. Academic Press, San Francisco, California.

Nakano, H. and Bayliff, W.H. (1992). A review of the Japanese longline fishery for tunas and billfishes in the eastern Pacific Ocean, 1981-87. Inter-American Tropical Tuna Commission Bulletin 20(5): 183-355.

Nishikawa, Y. and Ueyanagi, S. (1974). The distribution of the larvae of swordfish, *Xiphias gladius*, in the Indian and Pacific Oceans. pp. 261-264 in Shomura, R.S. and Williams, F. (eds.) Proceedings of the International Billfish Symposium held Kailua-Kona, Hawaii, 9-12 August 1972. Part 2. Review and contributed papers. NOAA Technical Report, NMFS SSRF-675.

Nishikawa, Y., Honma, M., Ueyanagi, S. and Kikawa, S. (1985). Average distribution of larvae of oceanic species of scombrid fishes, 1956-1981. Far Seas Fisheries Research Laboratory, Shimizu. S. Series 12.

Palko, B.J., Beardsley, G.L. and Richards, W.J. (1981). Synopsis of the biology of the swordfish, *Xiphias gladius* Linnaeus. United States Department of Commerce, NOAA Technical Report NMFS Circular 441 (FAO Fisheries Synopsis No. 127).

Punt, A.E., Campbell, R. and A.D.M. Smith. 2001. Evaluating empirical indicators

and reference points for fisheries management: Application to the broadbill swordfish fishery off Eastern Australia. *Mar. Freshwater Res.* 52: 819-832.

Ward, P. and Elscot, S. (2000). Broadbill swordfish: status of world fisheries. Bureau of Rural Sciences, Canberra, 208pp.

Yokawa, K. (2001). Problems in swordfish stock assessment in the Indian Ocean. WPM01-09 IOTC Proceedings no. 4. pp 232-233.

Yabe, H., Ueyanagi, S., Kikawam S, and Watanabe, H. (1959). Study of the life history of the sword-fish,, *Xiphias gladius* Linnaeus. Nankai Reg.Fish. Res. Lab. Rep. 10: 107-150.

Young, J. and Drake, A. (2002). Reproductive dynamics of broadbill swordfish (*Xiphias gladius*) in the domestic longline fishery off eastern Australia. FRDC Project 1999/108 Final Report.

Table 1: The total number of length-frequency sampled fish, by year and CPUE standardization area, normalized by the approximate number of 5-degree squares in that area.

YEAR	CPUE STANDARDISATION AREA									
	1	2	3	4	5	6	7	8	9	10
	APPROXIMATE NUMBER OF 5-DEGREE SQUARES									
	12	10	26	28	14	14	23	6	14	20
1970	NA	59.80	3.62	12.14	NA	27.79	0.04	NA	NA	NA
1971	NA	3.20	6.08	8.68	NA	18.14	NA	NA	5.00	0.45
1972	NA	9.90	0.42	12.00	NA	18.14	NA	NA	1.50	NA
1973	NA	NA	1.15	8.79	1.21	14.00	NA	NA	NA	NA
1974	NA	0.50	NA	15.04	NA	21.64	NA	NA	1.43	NA
1975	NA	5.30	1.31	8.39	2.14	31.21	0.04	NA	0.07	NA
1976	7.08	2.80	0.04	6.57	0.36	48.07	0.04	NA	NA	NA
1977	1.83	NA	12.50	7.68	NA	41.07	0.04	NA	3.21	NA
1978	NA	2.40	2.35	7.61	NA	40.71	1.26	NA	10.07	NA
1979	2.92	8.30	1.42	9.36	0.07	61.57	0.17	NA	1.43	NA
1980	NA	NA	2.88	10.14	NA	72.86	NA	NA	NA	NA
1981	NA	0.50	2.73	12.14	0.29	94.64	0.48	NA	NA	NA
1982	NA	12.50	10.81	9.50	1.57	89.07	NA	0.17	0.71	NA
1983	NA	14.90	4.73	9.71	0.14	72.29	4.43	NA	0.64	0.25
1984	NA	4.80	9.00	6.21	1.21	62.21	38.48	NA	NA	NA
1985	NA	19.00	4.54	7.36	0.14	79.36	20.65	NA	NA	NA
1986	NA	19.90	8.08	4.61	0.50	60.79	10.70	NA	NA	NA
1987	0.50	13.90	11.96	5.46	0.71	38.50	2.65	NA	NA	NA
1988	0.42	17.60	15.81	6.25	NA	26.57	11.78	NA	0.57	0.25
1989	NA	13.70	1.42	5.36	NA	20.50	32.87	NA	0.07	2.20
1990	NA	13.30	4.00	4.21	NA	20.93	44.74	NA	NA	NA
1991	NA	NA	1.00	3.61	NA	14.93	34.04	NA	1.93	8.75
1992	NA	3.60	0.77	0.86	NA	6.14	14.83	NA	2.00	2.30
1993	NA	NA	NA	1.07	NA	3.29	39.39	NA	2.36	NA
1994	NA	NA	2.27	0.86	0.36	1.00	41.57	NA	NA	NA
1995	NA	2.70	4.23	1.61	NA	1.93	45.17	NA	NA	NA
1996	NA	0.30	1.65	2.04	NA	2.36	22.35	NA	0.21	0.20
1997	NA	NA	4.19	0.43	NA	3.93	37.13	NA	NA	NA
1998	NA	1.80	5.54	1.32	NA	1.29	33.39	NA	0.07	NA
1999	NA	0.20	NA	1.04	NA	NA	16.70	NA	NA	NA
2000	NA	1.80	2.08	1.75	2.71	0.43	13.65	NA	NA	NA
2001	NA	0.20	NA	0.14	NA	0.29	15.17	NA	NA	NA
2002	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA

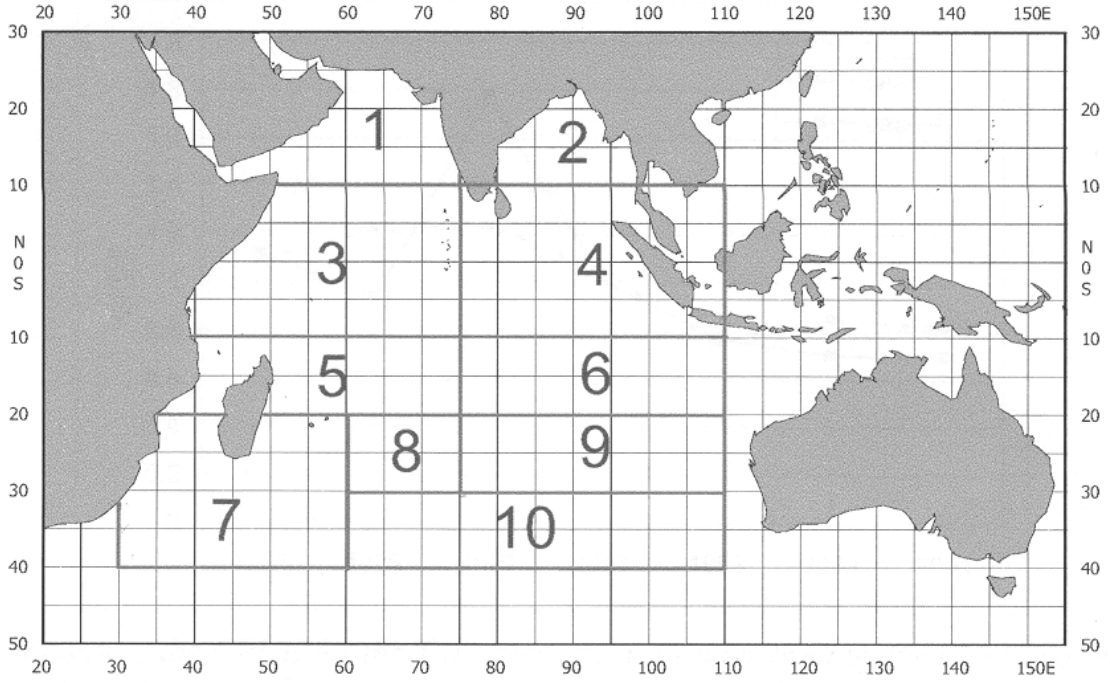


Figure 1: The areas used for standardization of broadbill swordfish longline CPUE (Report of 3rd Session IOTC WP Billfish, 2003)

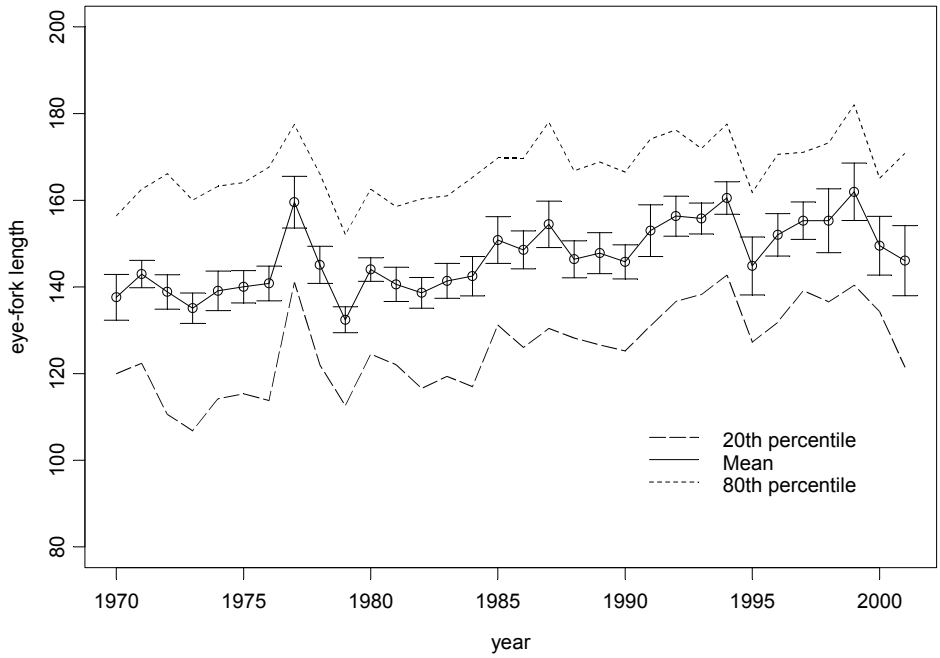
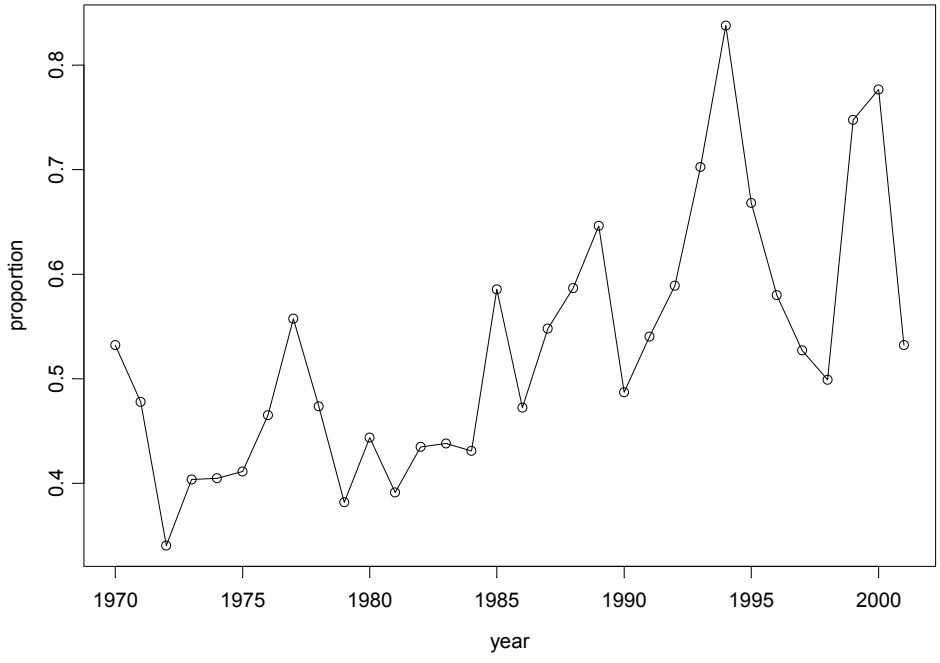


Figure 2a (above) : Japanese longline annual averages of mean length and average 20th & 80th percentiles (FL in cm) for the whole of the IO. Error bars are ±1 standard error (s.e.).

Figure 2b (below) : Japanese longline annual proportion >150cm EFL (EFL in cm) for the whole of the IO.



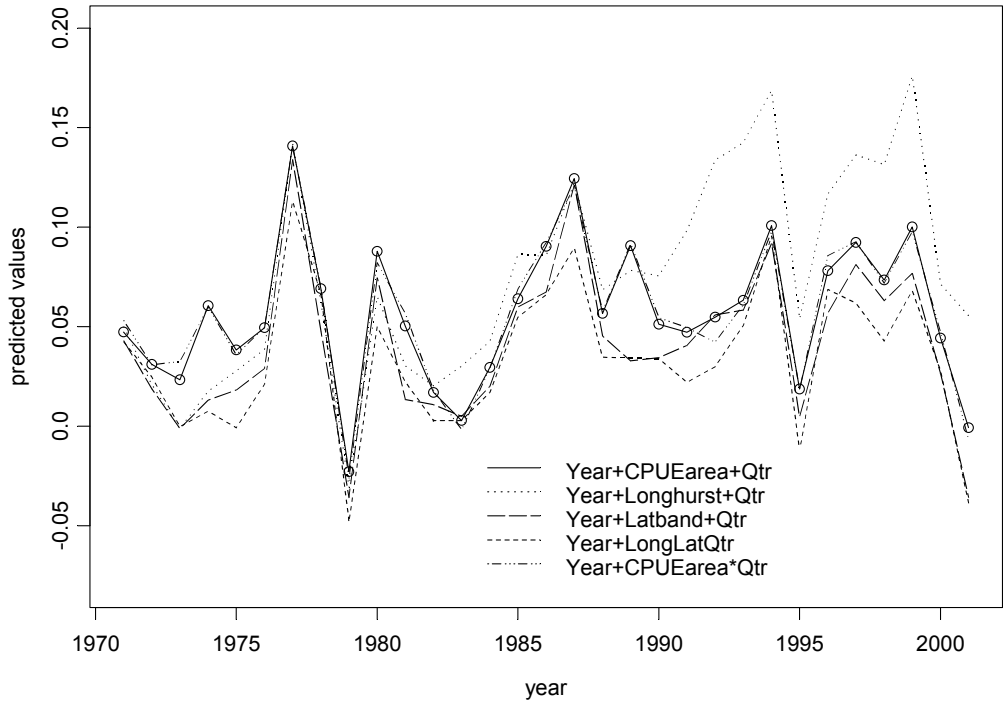


Figure 3: Japanese longline year coefficients overlaid for GLMs of log(mean length) as a function of i) Year + CPUEarea + Quarter, ii) Year + Longhurst + Quarter, iii) Year + Latitude band + Quarter, iv) Year + LongLatQtr, v) Year + CPUEarea*Quarter.

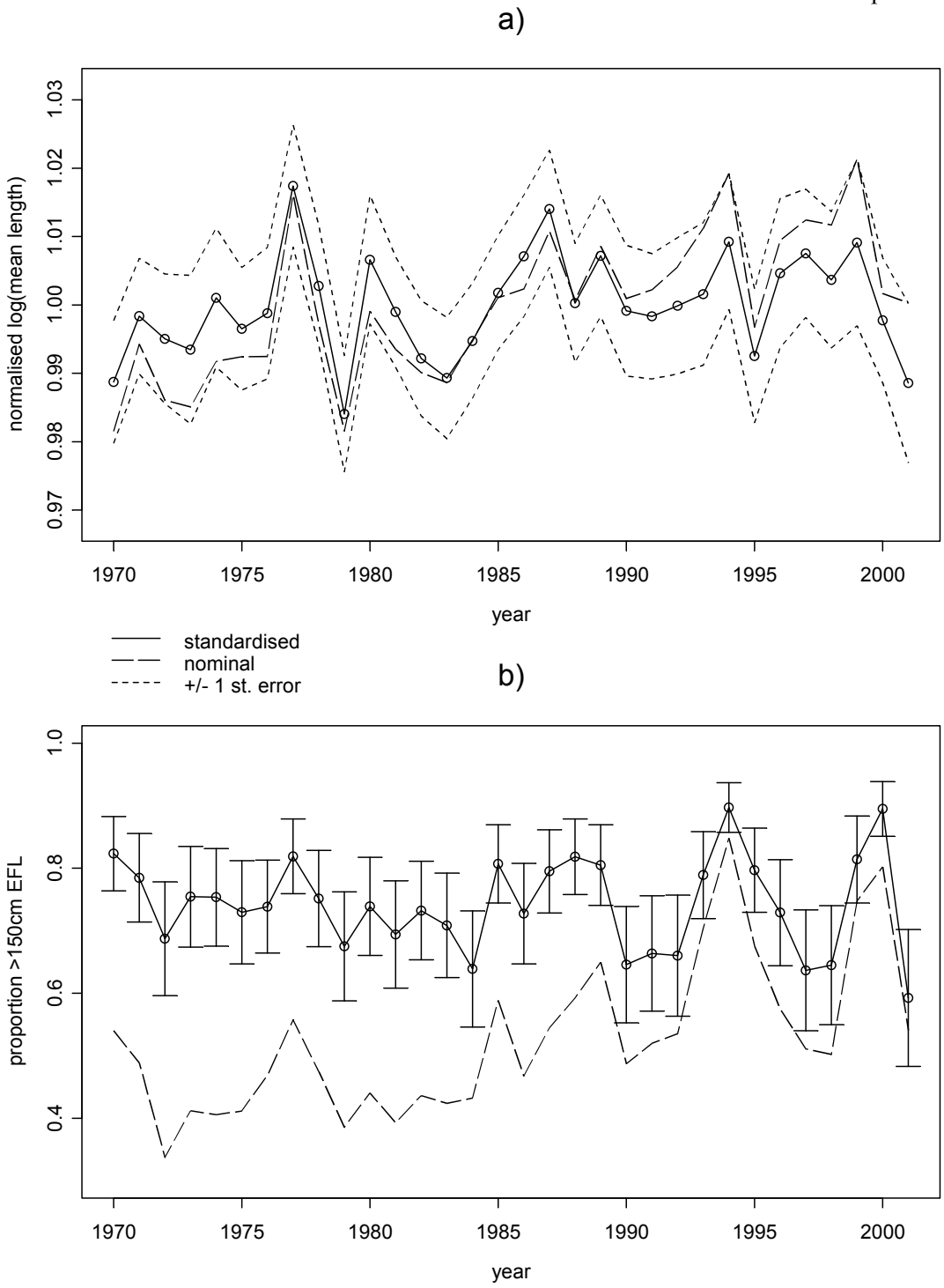


Figure 4: Japanese longline time series of a) standardized and nominal swordfish mean length indices on the log scale, normalized using the mean values over the time series, and b) standardized and nominal indices for proportion of swordfish >150cm EFL. Error bars are ±1 s.e.

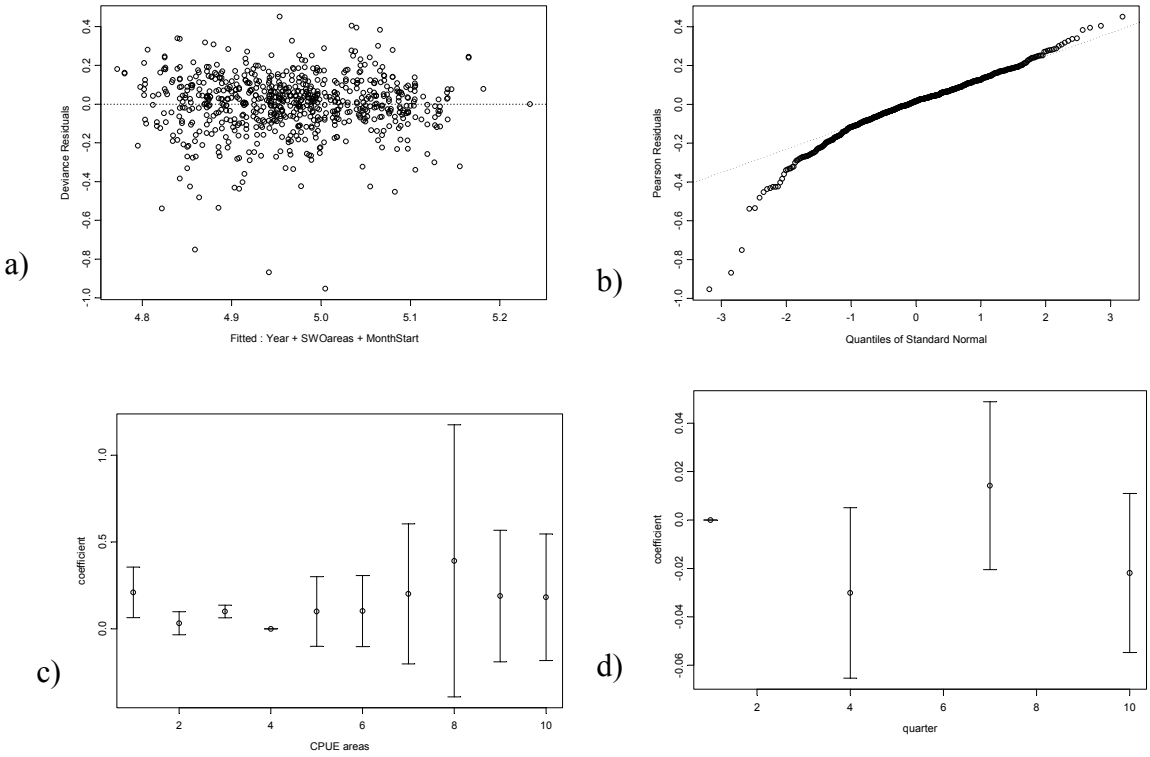


Figure 5: Japanese Longline. Plots of a) deviance residuals vs. fitted values, b) Pearson residuals vs. quantiles of standard normal, c) CPUE area coefficients, relative to area 4, and d) quarter coefficients relative to quarter 1, for the mean length main effects GLM (model coefficients relative to CPUE area 4). Error bars are ±1 s.e.

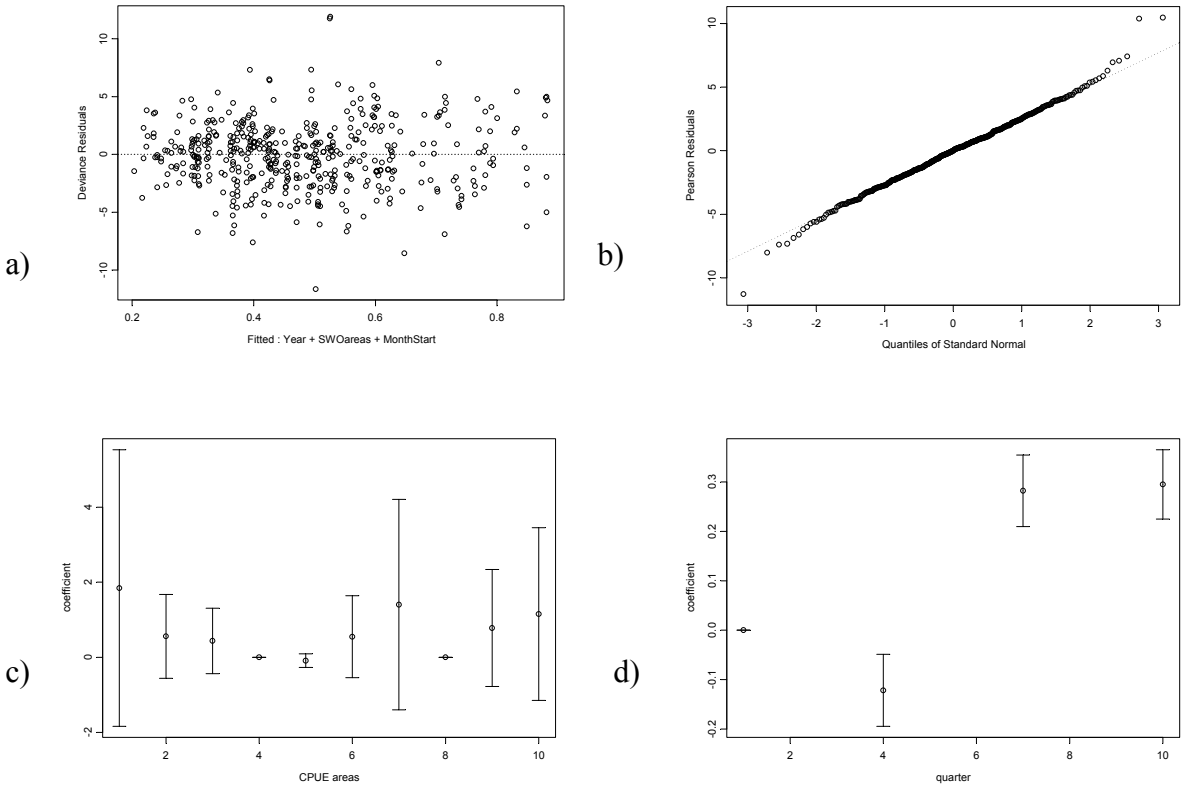


Figure 6: Japanese Longline. Plots of a) deviance residuals vs. fitted values, b) Pearson residuals vs. quantiles of standard normal, c) CPUE area coefficients relative to area 4, and d) quarter coefficients relative to quarter 1, for the GLM of proportion of swordfish >150cm EFL (model coefficients relative to CPUE area 4). Error bars are ±1 s.e.

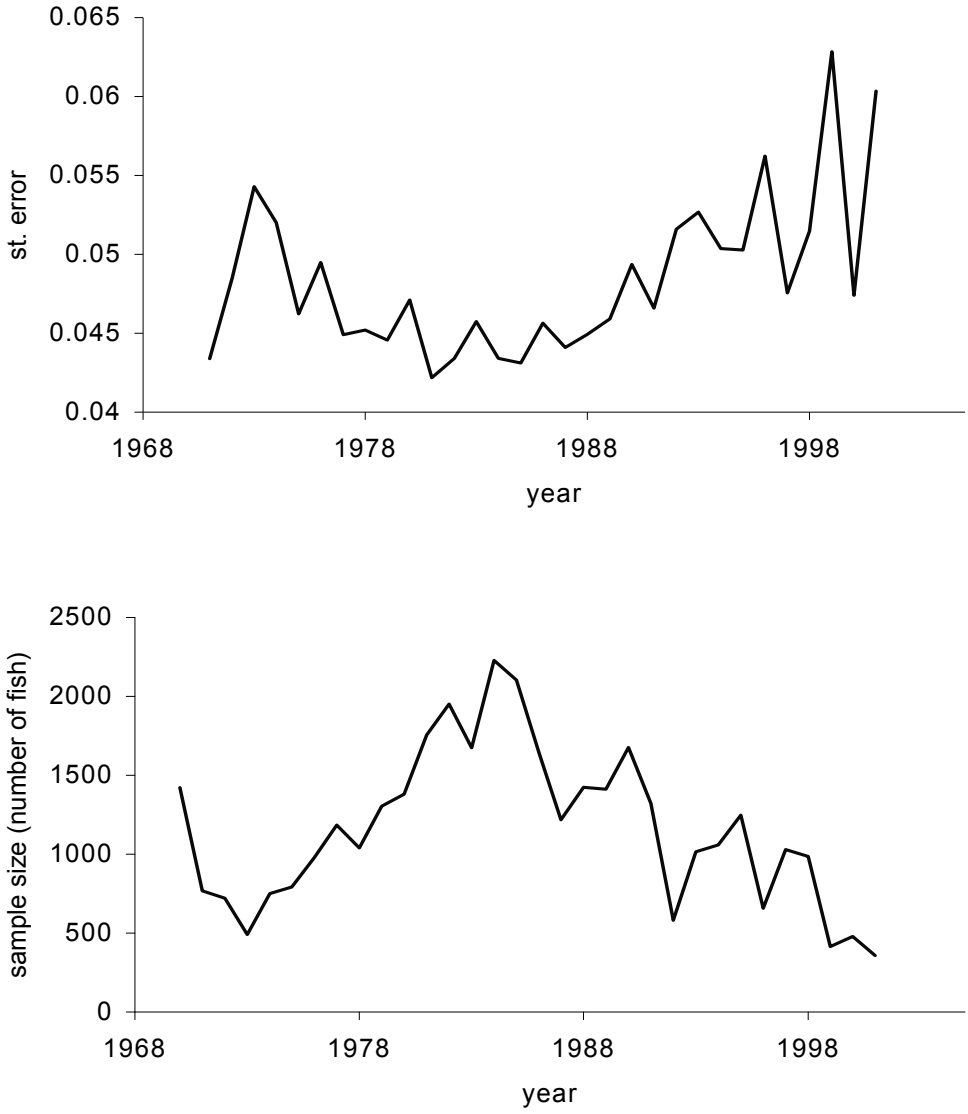


Figure 7: Japanese longline. a) standard errors associated with standardized indices for mean length and b) annual total sample sizes (numbers of fish) used in calculating mean length.

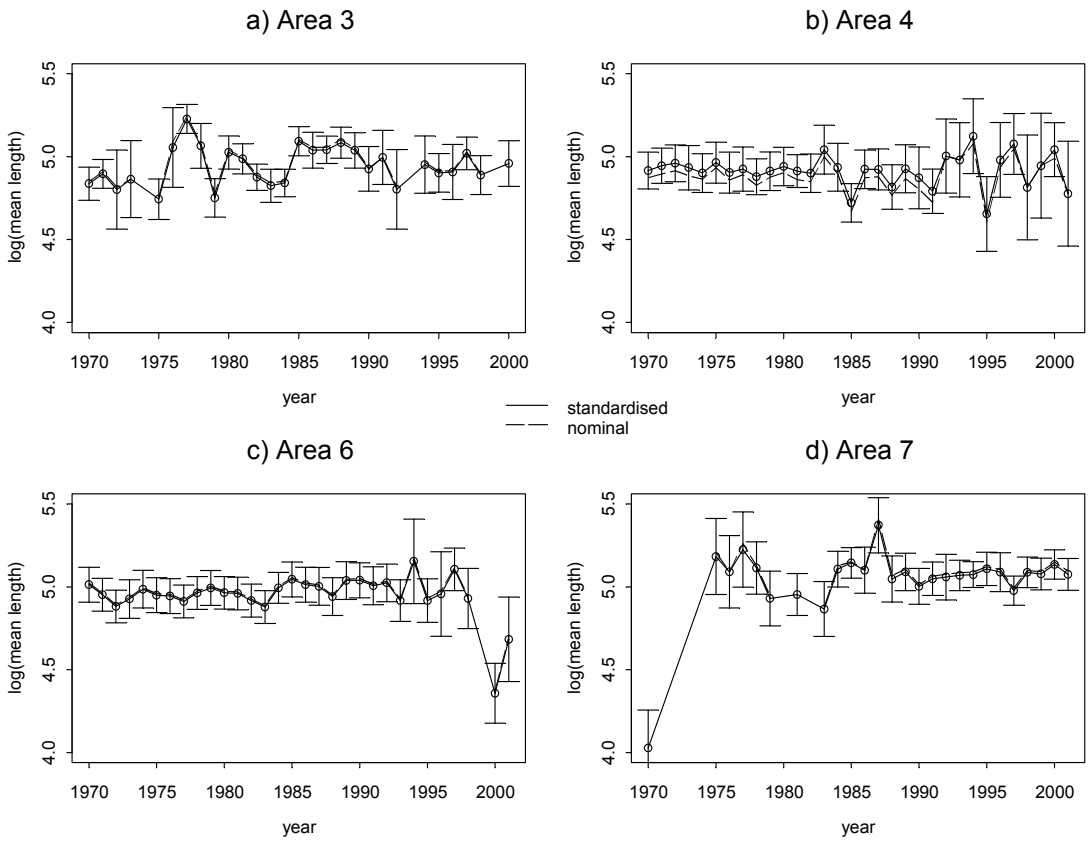


Figure 8: Japanese longline time series of standardized and nominal swordfish mean length indices on the log scale, for a) CPUE standardization area 3, b) CPUE standardization area 4, c) CPUE standardization area 6, d) CPUE standardization area 7. Error bars are ±1 s.e.

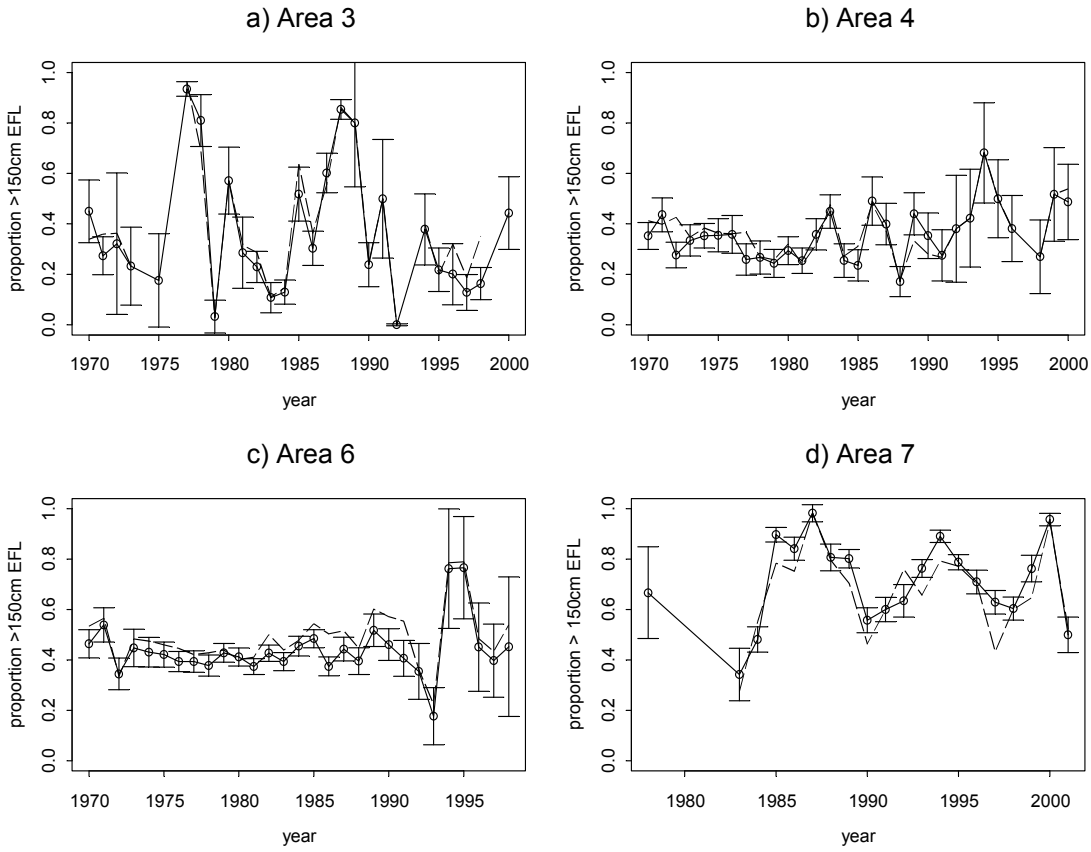


Figure 9: Japanese longline time series of standardized and nominal indices for proportion of swordfish >150cm EFL, for a) CPUE standardization area 3, b) CPUE standardization area 4, c) CPUE standardization area 6, d) CPUE standardization area 7. Error bars are ±1 s.e.

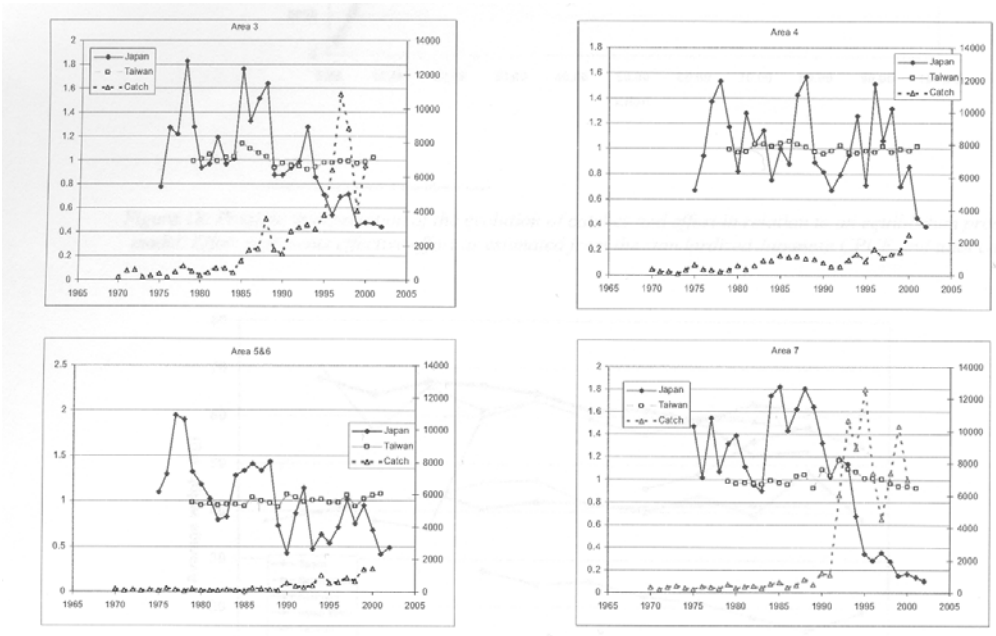


Figure 10: Standardised CPUE indices for areas 3, 4, 5&6 and 7, from the Report of the 3rd Session of the IOTC Working Party on Billfish, 2003.

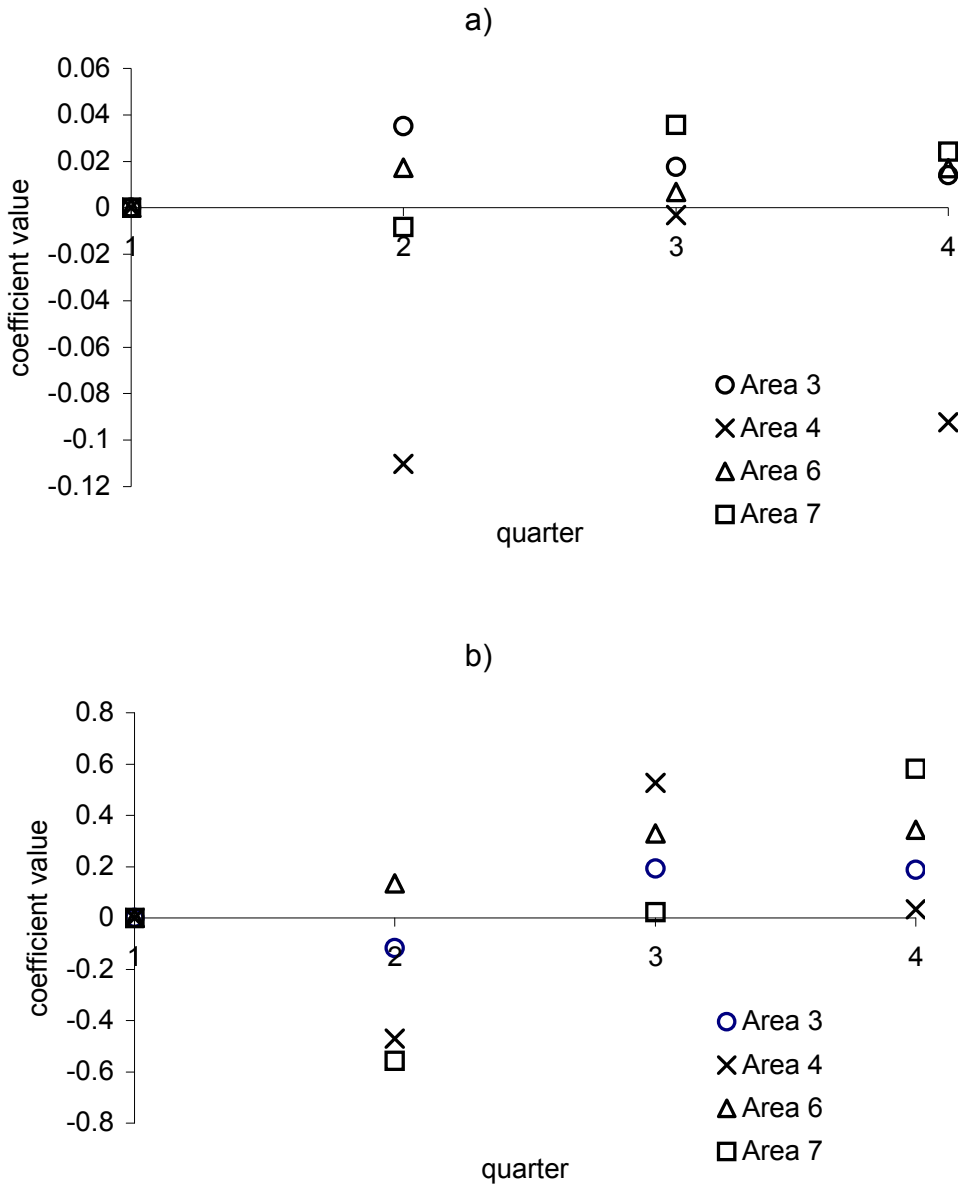


Figure 11: Quarterly coefficients for the area-specific GLMS undertaken for a) mean length and b) proportion of fish > 150cm EFL

8. An investigation of the responsiveness of CPUE- and size-based indicators

Summary

The work in this Section is the first step towards addressing the third and fourth objectives of the project, namely to develop and test the robustness of single indicators and of groups of indicators, for potential use in decision rules.

It is unsurprising that CPUE-based indicators and CPUE-time series have been used so extensively in assessing stock status. Compared to size-based indicators, CPUE-based indicators are substantially more responsive. Under ideal circumstances, the change in CPUE is proportional to the change in the fishable abundance. The size-based indicators change by far less, and are therefore much more sensitive to measurement error or random noise, for example, due to recruitment variability.

Results strongly suggest that size-based indicators, in terms of either length or weight, are only likely to be informative for stocks which have slow growth rates. Size-based indicators are in fact also unlikely to be informative if the steepness of the stock-recruit curve is low, irrespective of the mortality and growth rates. This is because the maximum sustainable harvest rate is low when steepness is low. Unfortunately, it is very likely that we would not know what steepness is, or if there is an estimate, it is likely to be very uncertain.

The expected change in an indicator also depends on the selectivity pattern of the fishery. A flat selectivity pattern leads to more responsiveness in size-based indicators than dome-shaped selectivity, particularly if only a narrow range of ages (or sizes) are being harvested.

Harvest rates based on 40% relative spawner-per-recruit, F_{spr40} , and associated trigger points based on these harvest rates, are likely to be conservative compared to F_{msy} when steepness is above about 0.5. If steepness is lower than 0.5, however, F_{spr40} harvest rates will tend to be higher than F_{msy} and therefore not conservative. If there is evidence or a serious concern (e.g. based on the reproductive strategy of a particular stock) that steepness is low, then a harvest rate coinciding with a higher spawner per recruit ratio (e.g. F_{spr50}) should be considered.

Although CPUE-based indicators are the most responsive, they are more strongly affected by recruitment variability than the size-based indicators. They also have positive autocorrelations for longer time lags than the size-based indicators. The time it takes for an indicator to reach a new equilibrium level (or to be within, say, 5% of it's new long term average) following a step change in the harvest rate, depends on the type of indicator, the life history parameters, and the extent of change in the harvest rate. Such 'return times' are longer when natural mortality is low, when there is a broad range of age classes in the catch, and when steepness is low.

8.1. Introduction

Stock status indicators, such as mean length or catch per unit effort (CPUE), are considered as management tools, because it is assumed that they reflect the status of the stock, or the rate of exploitation, in some way. CPUE is particularly attractive because it is potentially so directly and strongly related to stock size. The most common assumption is that CPUE is proportional to stock size, implying that a 20% drop in CPUE, say, signals a 20% drop in the fishable

population[7]. There are, however, two potential problems with CPUE-based indicators: (i) the relationship between CPUE and population size may not be linear and (ii) the constant of proportionality could change over time. The second issue is usually referred to as 'effort creep'. It is because of these potential problems, as well as the possibility that CPUE information may not be available or interpretable, that other indicators are being explored as alternative or additional indicators.

This raises the question: how does mean length or median weight relate to population size? This is not straightforward because, unlike CPUE, size-based indicators are not direct measures of population density. Instead, they reflect the size and age structure in the catch (or fishable population) which reflects the total mortality, and therefore implicitly the fishing mortality. The size structure is also affected by the growth rate, and both size- and age-structures are affected by the fishery selectivity curve (i.e. the extent to which different sizes are caught).

Here we explore the relationship between stock size and size-based indicators for a wide range of life-history strategies and selectivity patterns. We do this by first considering a deterministic population model with no recruitment variability or measurement error. There are several reasons why we consider this exploration to be important and generally relevant. First, the expected change in an indicator under a given harvest rate identifies the signal that one would be trying to detect and respond to in a real situation. The signal would, however, be attenuated (weakened) by random noise and measurement errors. If the signal is likely to be very weak to start with, i.e. if the change in an indicator is expected to be small when the stock is harvested at a sustainable rate, then that indicator is unlikely to perform well in practice.

The second reason is that by looking at a wide range of life-history strategies, we should be able to provide general guidance about when a particular size-based indicator is likely to be worth considering and when it is likely to behave poorly, based on the expected relative change at sustainable harvest rates.

The performance of range of CPUE-based indicators under ideal conditions and perfect assumptions (i.e. no effort creep, and a linear relationship between CPUE in numbers and population abundance) is also explored. We note that it is not possible to consider spatial indicators[8] in the same way, because they relate to the dynamics of the fishery rather than the biology of the stock.

The deterministic considerations play another important role. Calculations of the expected value (absolute, or relative value) of an indicator under a given harvest rate, can potentially be used as a reference point or a trigger point for that indicator (also see Sections 9 and 10). Here we compare expected values for the range of indicators and life-history parameters based on the associated maximum sustainable yield harvest rate (F_{msy}) and the harvest rate which implies 40% of unexploited spawner per recruit (F_{spr40}).

Having looked at the deterministic expected response of indicators, we consider the potential effect of recruitment variability. Although we know that all wild marine populations are subject to recruitment variability, the magnitude of this variability is often very hard to measure, even if there is a quantitative assessment which provides an estimate of this quantity. We use a simulation model to generate time-series of indicators under a constant harvest rate and under low, moderate or high recruitment variability. We then explore the variability, as well as the

[7] The 'fishable population' refers to the ages and size classes present in the catch, and therefore reflected in the CPUE.

[8] For example, the fishing fleet's mean distance between the shore and fishing positions or 'mean distance offshore'.

autocorrelation, of the time-series of indicators. The variability is relevant because it provides an indication of how much an indicator could vary even when measured without any error.

The final sub-section looks at the responsiveness of indicators to a large change in harvest rate. We ask how many years it takes for the each indicators to get within 5% of its new long term average level associated with the lower harvest rate. This, together with the variability and autocorrelation properties of an indicator, has potential implications for how one might use such an indicator in a decision rule. It also has implications for whether a decision rule, and hence catch or harvest rate adjustment, should be applied annually, or at longer time-intervals such as every three or every five years.

The work in this Section is the first step towards addressing the third and fourth objectives of the project (see Section 4):

3. Develop candidate frameworks and methodology for a management system based on a suite of indicators
4. Test the robustness of SSIs (individually and jointly in a framework) relative to arbitrary trigger or reference points without built-in decision rules

Results in this Section are taken further in Section 9.

8.2. Methods

8.2.1 The structure of the simulation model

The simulation model for this part of the study is a very simple annual, age-structured population model. The model has no spatial structure, and represents a single stock with no movement dynamics. It is harvested by a single fleet, which is meant to represent longlining, but we explore both flat-topped and dome-shaped selectivity curves for the fishery. Purse seine gear usually has a dome-shaped selectivity curve, so this type of harvesting is implicitly represented here. Model equations are given in Appendix 1, but a verbal description is given here for completeness.

The main population parameters, that are input to the model, are:

steepness[9] (h) which defines the Beverton-Holt stock-recruitment curve,
growth rate, k, in the von Bertalanffy growth curve,
natural mortality (m), and
age at maturity (t_m).

All these quantities are assumed to be constant over time. Although the model is age-based, there is an implicit size-distribution associated with each age class. We assume a normally distributed length distribution for each age class, with the mean length obtained from the von Bertalanffy growth curve and a variance specified to decrease with age (Appendix 1). This implies a mixture normal distribution for the size frequency of the catch or population. A weight-frequency-distribution can be derived from any length frequency distribution using a power relationship ($W = a.L^b$, where the value of b is usually around 3).

It is important to note that scaling quantities such as the level of recruitment R, the maximum length, L_∞ and t_0 in the growth curve, and the constant in the weight-length relationship, are of limited importance because we are primarily interested in the value of an indicator at some harvest rate, F, relative to the value of the indicator when there is no harvest, i.e. when F=0.

[9] Steepness is defined as the ratio between recruitment (R) when spawning biomass is at 20% of unexploited spawning biomass (B_0), and recruitment at the unexploited spawning biomass i.e. $h = R(0.2B_0) / R(B_0)$.

Setting these scaling quantities at arbitrary values should therefore not affect results of relative change in indicators.

Interannual variability in recruitment can either be 'turned off' (i.e. no variability in recruitment) or specified through a non-zero standard deviation term. Variable recruitment is modelled as a lognormal distribution with standard deviation σ_R and mean taken from the stock-recruit curve.

Parameter values are given in the description of scenarios below.

8.2.2 The Fishery

The fishery, or harvesting, is described by a harvest rate, F, and an age-specific selectivity vector , S, which specifies the proportion of F which applies to each age class. The selectivity of longline fishing gear is commonly assumed to be flat-topped, increasing from 0 to 1 over some range of sizes, but then remaining at 1 for all older/larger fish. Many assessments of pelagic fish stocks which estimate selectivity parameters, suggest that the selectivity of longline gear may be dome-shaped, and we therefore consider three types of selectivity curve. The simplest is knife-edge selectivity which is zero for all ages younger than the age at first capture (t_c), and one for all ages at and above t_c. The other two selectivity curves are: flat (or flat-topped), which increases from 0 to 1 over several, rather than just 1, age classes; and dome-shaped (i.e. where selectivity drops back to zero for older, larger fish).

Note that these curves can easily be shifted to younger or older animals by specifying, for example, the youngest age at which selectivity is 50%. Also, dome-shaped selectivity pattern can represent either a wide or a narrow range of fully selected age classes. At the one extreme, with a wide range of fully selected age classes, the dome-shaped and flat selectivity curves are very similar. They are obviously most different when the dome-shaped selectivity curve has only a narrow range of fully selected age classes. We consider two dome-shaped curves, one with 4 and one with 2 fully selected age classes. Note however, that the total mortality rate also plays a role in whether results from flat and dome-shaped selectivity curves are different or similar (see sub-section 3. Results).

The indicators are assumed to come from 'fisheries data' and are modelled as such. The full set of indicators are listed in sub-section 2.3 below and also described in Appendix 1. In this Section's explorations we have 'turned off' the measurement and sampling error, implying that the indicators can be perfectly 'measured' from the catch-at-age vector (which is without error). For example, CPUE in numbers, is the actual catch in numbers from the standard catch equation divided by the implicit effort[10]. There is no error in the catch-at-age distribution and there is no additional 'measurement' error in CPUE. Measurement errors on indicators are introduced in Section 9.

The size distribution of the catch is implicitly for the whole catch, though this still has a finite sample size, albeit it several thousand (~ 4000 in most cases). The size frequency distribution is obtained from the catch at age distribution and the associated normal distributions of length at each age. Given the finite total catch sample, and the fact that some age classes could have relatively few individuals in the catch, some randomness enters this step so that even for an identical catch-at-age, different realisations of length frequency distribution would not be identical (unless the same random seed is used for each sampling). This level of randomness in the indicators is, however, small enough to not be a problem for this investigation.

[10] Effort is not explicitly modelled, but can be inferred from the assumption that the harvest rate F = qE (catchability * Effort) so that the implicit effort is F/q. CPUE then becomes C/(F/q).

As an aside we note that the simple indicators such as mean length and mean weight can be obtained very easily without going through any 'sampling' from the catch-at-age as described above. The medians and upper percentiles of length and weight do, however, require sampling of this kind, particularly if the more realistic case of mixture distributions of length-at-age (rather than single lengths-at-age) are used.

For most of the scenarios we consider fishing at F_{msy}, which is the harvest rate that produces the maximum sustainable yield (also see sub-section 2.4 below). The main reason for using F_{msy} is that it is considered a limit reference point for the harvest rate, and should therefore indicate the largest relative change that one would want to observe in an indicator. This means that if there is only a very small expected change in an indicator when harvesting at F_{msy} (relative to the indicator value for the unexploited stock), then a lower harvest rate would lead to an even smaller expected change which may be hard to detect when variability and measurement errors are taken into account.

We also briefly consider fishing at $F_{spr40,}$ which is the harvest rate where spawner per recruit is 40% of the spawner per recruit value at no harvest. Note that spawner per recruit calculations are analogous to yield per recruit calculation, with no assumption about a stock-recruit relationship. F_{msy} on the other hand, requires an assumption about the stock-recruit relationship and is, in fact, strongly affected by the steepness. The major advantage of using spawner-per-recruit considerations rather than maximum sustainable yield considerations is that the steepness is one of the main life-history quantities which is likely to be unknown or, at best, very poorly estimated.

8.2.3 Indicators

The calculation of most of the indicators used here are self-evident. For example, the mean length is simply the mean of all the sampled lengths in the size-frequency distribution (also see Appendix 1).
The size-based indicators are:

Name	*Description*
mL, mdL, L90	- the mean, median, 90th percentile of length
mW, mdW, W90	- the mean, median, 90th percentile of weight
ps	- the proportion of "small" fish in the catch (defined as the proportion of fish $\leq t_c+1$)
pb	- the proportion of "large" fish in the catch (defined as the proportion of fish $\geq t_c+3$)

where t_c is the age-at-first capture when selectivity is knife-edged, or the youngest age for which selectivity is ≤ 0.5 when selectivity is flat-topped (but not knife-edge) or dome-shaped.

The CPUE-based indicators are:

Name	*Description*	
u	- CPUE in numbers	
us	- CPUE of small fish (numbers)	= u.ps
ub	- CPUE of big fish (numbers)	= u.pb
uw	- CPUE in weight	= u.mW
uSB	- CPUE of spawning fish (weight)	= $u.W_s.p_m$

(where W_s is the mean weight of fish larger than size at maturity in the size-frequency sample, and p_m is the proportion mature).

The indicator 'proportion small fish' increases rather than decreases when biomass decreases and is more closely related to recruitment than spawning biomass, particularly when the catch contains many juveniles. This indicator therefore needs to be interpreted with some care.

8.2.4 Scenarios

8.2.4.1 Theoretical and equilibrium response

i) Grid of life-history parameters

Since the way in which a size-based indicator changes when a stock is harvested, will depend on the life-history parameters of the stock we first consider a grid of these values and explore the response of indicators for all the combinations of these parameters, as summarised in Table 4. These values span a range of likely extremes for most tropical tuna and billfish-like species.

Table 4. Grid showing the three values of the input parameters evaluated: steepness (h), von-Bertalanffy growth rate (k), natural mortality, age at maturity (tm), age at first capture (tc= tm/2, tm, tm+1).

h	k	M	t_m	t_c
0.3	0.2	0.1	2	$t_m/2$
0.6	0.4	0.2	4	t_m
0.9	0.6	0.4	6	t_m+1

The first set of scenarios therefore consist of 243 combinations of life-history parameter in Table 4, combined with a single set of the other (scaling) parameters. The first step is to determine the harvest rate associated with the maximum yield, $F_{msy.}$ This is done by running the simulation model to equilibrium for a given harvest rate and determining the catch. An optimisation routine is used to find the harvest rate (F_{msy}) which produces the largest catch, to within a specified tolerance level.

The deterministic investigations of the expected change in indicators when harvesting at F_{msy} compared to no harvest, are done with no variability in recruitment. The population model is started in an unexploited state, and the values of all indicators are calculated. The simulated population is then harvested at F_{msy} until the stock reaches its new equilibrium. The values of all indicators at the new, harvested, equilibrium are calculated, as is the expected relative change in each indicator.

ii) correlated life-history parameters

In reality, some combinations of life-history parameters in Table 4 are much less likely than others. For example, it is most unusual to find a short-lived species with a high age at maturity (it would be a very risky and probably non-viable life-history!). It is also unusual to find very high ratios of m to k, and these parameters are usually correlated. Such unlikely combinations are included in the full set of (243) combinations in Table 4. We therefore also generate a somewhat wider range of life-history dynamics, but we do this with correlation between m and k, and between m and t_m (age at maturity), as well as m and t_c (age at first capture). The reason for also doing this exercise is to check whether the patterns seen above are much changed, or not, when the life-history parameters are a more realistic sub-set.

For each realisation within a life-history scenario, the four parameters (m, k, t_m, t_c) are generated in the following way.

- Natural mortality and growth rate. Values of m and k are generated so that log(m) and log(k) are normally distributed with variances of 0.3 and a correlation of 0.8. The log-means were the logs of 0.2 for both parameters. This choice generates a wide range of parameter values, but values are mostly between about 0.1 and 0.5, which is of main relevance for the kinds of species we are interested in. The resulting set of parameters is shown in Figure 1 (below).
- Age at maturity. Values of t_m are determined in such a way that the expected proportion alive at age t_m (given the value of m from step 1 above) is between 20% and 40%. For each m-value, a minimum (T_{min}, based on 40%) and maximum (T_{max}, based on 20%) potential age at maturity is calculated. The t_m value associated with that m-k pair is drawn from a uniform distribution [T_{min}, T_{max}], and rounded to obtain an integer age. Maturity-at-age is knife-edge at age t_m
- Age at first capture. Values of t_c are also drawn from uniform distribution [T_{min}, T_{max}], and rounded to obtain an integer age. This does mean that extreme differences between age-at-maturity and age-at-first capture are not represented in this simulated set. The most likely omissions are ones where t_c is much smaller than t_m. However, we consider that the ranges of t_c values for each t_m value shown in Figure 1 are sufficiently wide for the purposes of this exercise.

The level of correlation and the related assumptions used to generate this dataset, are not based on data, but are based on general observations and qualitative conclusions that, for example, natural mortality rate and growth rate are usually close to a ratio of 1. Also, the age at maturity is inversely related to growth rate and natural mortality rate. In practice, age at first capture is also related to the growth and mortality rate. This is most easily understood by thinking about yield per recruit and the fact that a fishery would tend to trade off the average size of animals they capture with the numbers. There is no real reason for age at first capture to be related to age at maturity, but this occurs inevitably through the links with m and k.

For the correlated set of life-history parameters, we also calculate the harvest rates which leads to a spawner per recruit value of 40% of the unexploited spawner per recruit value (i.e. SPR ($F=F_{spr40}$) / SPR ($F=0$) =0.4), and the coinciding expected changes in indicators when harvesting at this harvest rate, F_{spr40}.

8.2.4.2 Effects of recruitment variability

Deterministic, equilibrium results for the expected change in an indicator under a given harvest rate indicates the magnitude of the 'pure' signal when recruitment is constant. If recruitment varies from year to year, as it does in real populations, the indicators will also vary from year to year, even if they are measured without any error. This in itself is not a problem. The problem arises when this 'true' variability is combined with 'noise' or measurement error, because it is usually very difficult to separate the 'variability' from the 'noise'. It is therefore informative to look at the effect of only recruitment variability on the different indicators. In particular, we look at the magnitude of variability in each of the indicators by calculating the coefficient of variation (CV) of each indicator for a fixed harvest rate, but varying recruitment.

Even if recruitment variability is random and without autocorrelation, the time-series of spawning biomass, and many of the indicators, will tend to be autocorrelated because the spawning biomass consists of several age classes. The lower the total mortality, the higher the autocorrelation is likely to be. In addition to the CVs, we also calculate the autocorrelation of each indicator at different time-lags.

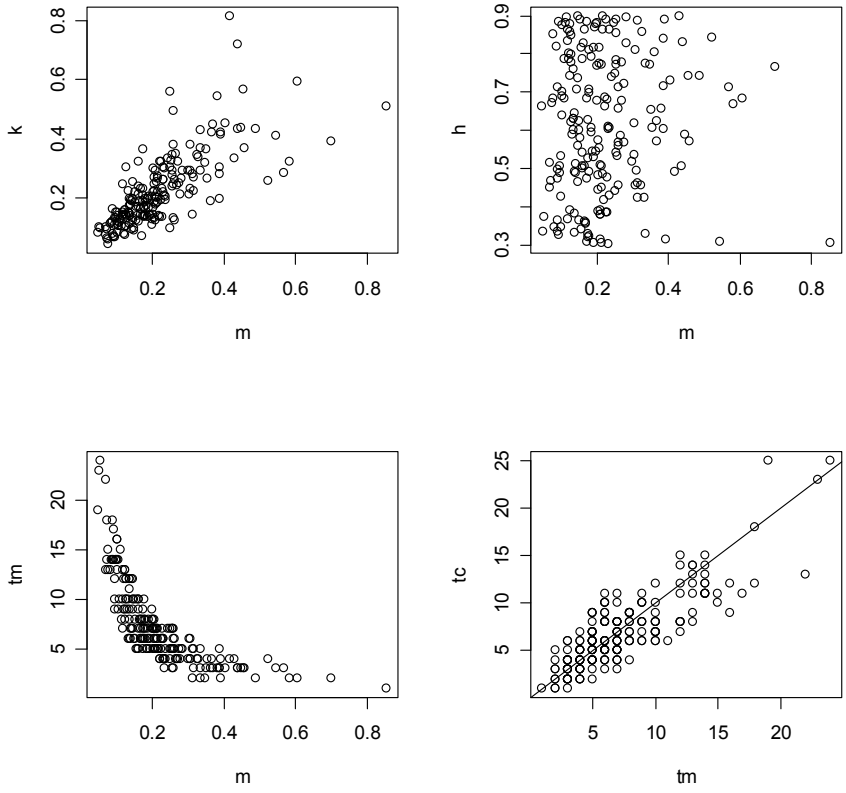


Figure 1. Correlated life-history parameters used in the evaluation of the expected relative change in indicators when harvesting at Fmsy. Parameters are natural mortality (m), von Bertalanffy growth rate (k), steepness (h), age at maturity (tm) and age at first capture (tc). See main text (methods) for more detail.

For this investigation, we again start the population at an unexploited level, then run the simulation forward under a fixed harvest rate until it stabilises around its new (harvested) equilibrium. Here 'stabilises' means that it is varying around the deterministic equilibrium. The CVs are calculated from the last 250 years of a 300-year time series. Very long time-series are required to obtain reasonable estimates of autocorrelation, particularly for long time lags. Again a time-series of 250 years was used (ignoring the first 50), and autocorrelations of lags up to 20 years was calculated, though in most cases it is only the first 10 or so terms that are of interest.

Two life-history scenarios were selected for this investigation: 'low m', or long-lived life history and 'high m', or short-lived life history. Parameter for these scenarios are summarised in Table 5. (The high m is not necessarily the highest value for actual tuna or billfish stocks, but it is high enough to show the differences). These scenarios had very high steepness to avoid confounding by a reduction in recruitment for the harvested stock.

Table 5. Input parameters for the two life-history scenarios: 'low m' or long-lived, and 'high m' or short-lived.

	h	k	m	t_m
low m scenario	0.99 (see text)	0.2	0.1	6
high m scenario	0.99 (see text)	0.6	0.4	3

Each scenario was run with either flat or dome-shaped selectivity, positioned at an appropriate (youngest) age at 50% selectivity.

8.2.4.3 Pulse change in harvest rate

In the context of using an indicator in a decision rule to adjust the catch or harvest rate, another quantity of interest is the time it takes for a pulse (or change) in harvest rate to 'die out'. This is relevant because a slowly changing indicator may need to be given time to respond rather than making adjustments to the control (i.e. harvest rate) each year. From a fishery's point of view, the desirability of stability in the fishery and the catch also suggests that it may be preferable to set catch levels for multiple rather than single years[11]. The question is then how often should a decision rule should be used: every year? every third or every fifth year? The ultimate test will be in simulations that mimic the feedback rule at different frequencies, but this investigation gives some insight into which time scales would be worth exploring further.

In this set of explorations we have used the same low and high natural mortality scenarios (Table 5). Both the flat and dome-shaped selectivity curves were used as described in the previous sub-section. The simulation is started with an unexploited stock harvested at 3 times F_{msy}. When the stock is varying around its new (overexploited) level, the harvest rate is cut back to F_{msy} in one year, and then kept at that level thereafter. The time-series of biomass and indicators, following the pulse-change in harvest rate, is then analysed to determine the number of years it takes to get to within 5% of the long term average at F_{msy}. The reason for using a criterion such as this, rather than within 1 or 2 standard deviations of the long term average, is because the standard deviation depends heavily on recruitment variability, σ_R. At low σ_R there would be a false impression of much longer return times than at high σ_R. The criterion based on being within some percentage of the long term average suffers far less from this problem.

These scenarios were run with a very high steepness so that there is no change in the average level of recruitment as a function of a change in spawning biomass. This gives the shortest 'return times'. If steepness is low, recruitment would increase slowly as spawning biomass increases, and this would lead to longer return times. This exercise was done for a single realisation of each scenario. We consider this to be sufficient for illustrative purposes and to highlight the main differences between indicators and scenarios. If this type of information were to be used more formally in subsequent analyses or more directly as input to decision rules, it would be preferable to run many realisations and consider the medians and standard deviations of 'return times' for each scenario.

8.2.5 Interpreting box plots

Many of the results are presented as box (or box-whisker) plots (e.g. Figure 2, Figure 3). The 'box' shows the median as a horizontal line inside the 'box', and the inter-quartile range as the upper and lower 'edges' of the box. The whiskers (vertical lines) extend outside the box to the most extreme data point which is no more than 1.5 times the inter-quartile range from the box. Data points beyond this range are shown as points outside the box.

[11] There is an important distinction between catch and effort regulated fisheries. In theory, effort regulated fisheries should require changes in effort less often since the catch ought to vary according to population size.

8.3. Results

8.3.1 Theoretical and Equilibrium considerations

8.3.1.1 Results for grid of life-history parameters

First consider results from the grid of values (Table 4) with a flat selectivity curve. Results are presented as boxplots of relative change in an indicator, or biomass, by three main life-history quantities: m, k and h. Figure 2 shows results of B_{msy}/B_0 for just the middle steepness value h=0.6, but for all the combinations of natural mortality and growth rate (m and k). Plots like this one are then put together for all three levels of steepness, as shown in Figure 3. (The vertical lines now separate the steepness levels rather than the k-levels as in Figure 2.)

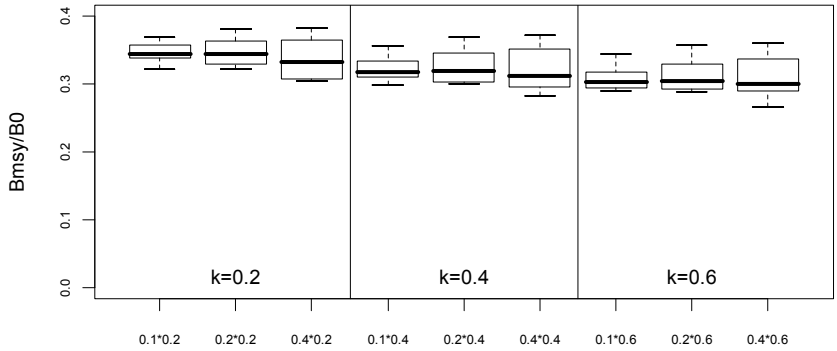


Figure 2. B_{msy}/B_0 for the subset with h=0.6, and by m*k pairs as indicated on the x-axis. For example, 0.4*0.2 means m=0.4, k=0.2. Each 'box' shows the range of t_m and t_c values for the given m*k combination.

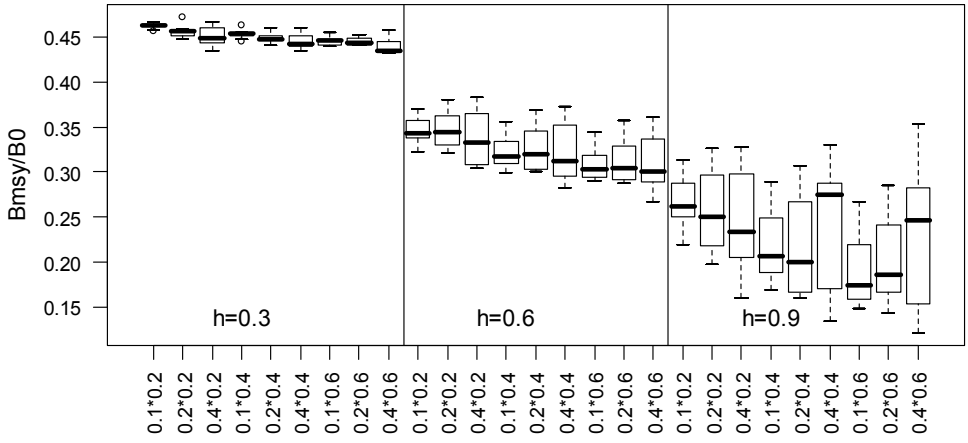


Figure 3. B_{msy}/B_0 (relative spawning biomass) for different values of k, m and steepness (h). Numbers on the x-axis refer to m*k pairs with m=(1=0.1, 2=0.2, 3=0.4) and k=(1=0.2, 2=0.4, 3=0.6). Factors reflected in the 'boxes' are age at maturity and first capture, t_m and t_c.

Steepness, which is likely to be the least well known quantity for a stock, has a strong effect on F_{msy} (higher steepness can support a higher F_{msy}) and therefore also on B_{msy}/B_0 (Figure 3). B_{msy}/B_0 is also generally lower for high values of m and k, but the effect of steepness dominates. The pattern seems to break down somewhat at high steepness, particularly when combined with high k. This is not real, but rather due to the fact that the yield curves for some of these cases are so flat that it is difficult to find the maximum. In the automated simulations which are stopped on the basis of a specified tolerance level, the harvest rate estimated as F_{msy}, may be on the low side because the yield increases more slowly (with further increasing F) than the tolerance requires. This would lead to higher values of B_{msy}/B_0 than seem to 'fit in' with the patterns shown in the lower steepness examples.

A similar plot for relative mean length, i.e. $mL(F_{msy})/mL(F=0)$ (Figure 4) also shows the strong effect of steepness. The expected change in mean length is largest at high steepness and when k is small (i.e. slow growth), particularly when m is also high (e.g. k=0.2, m=0.4). The clearest difference between the relative change in mean length and in spawning biomass (Figure 4 and Figure 3) is the magnitude. B_{msy}/B_0 is just above 0.45 at its highest, whereas relative mean length is very close to 1 at its highest, suggesting almost no change when steepness is low. B_{msy}/B_0 is around 0.15-0.20 at its lowest. Relative mean length is 0.80-0.85 at its lowest, which occurs only at high steepness combined with low growth rate. In other words, a drop of between 50 to 80% in spawning biomass, leads to hardly any change, or at most a 20% drop in the mean length.

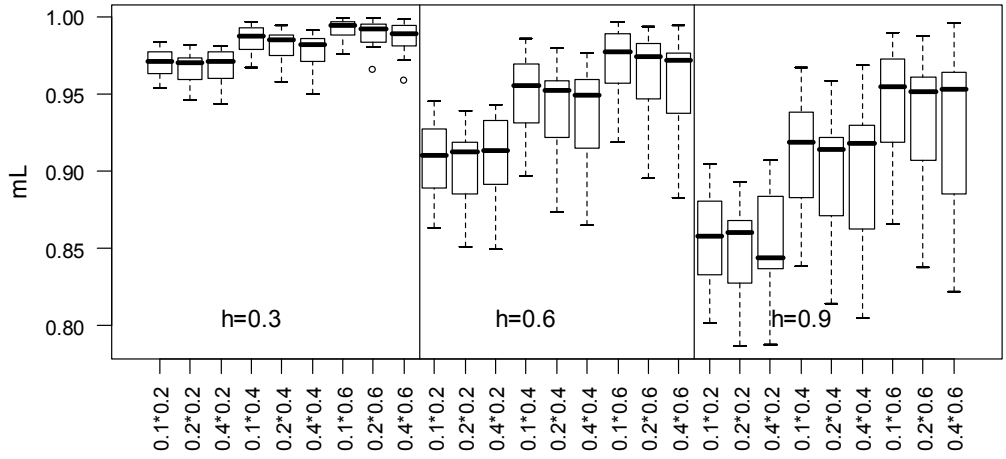


Figure 4. Relative mean length (meanL(F_{msy})/meanL(F=0)) for harvest at F_{msy} given 3 steepness (h) levels. Numbers on the x-axis refer to m*k pairs with m=(0.1,0.2,0.4) and k=(0.2,0.4,0.6). Factors reflected in the 'boxes' are age at maturity (t_m) and age at first capture (t_c).

Figure 4 also implies that, if a large change in mean length is observed for a stock with high growth rate, particularly in combination with low mortality rate, then this should be cause for concern, because it is likely that the harvest rate is already too high.

The 90th percentile of weight (Figure 5) is also sensitive to steepness, but note how much more sensitive this indicator is to natural mortality than the mean length, for example. The 90th percentile of length shows a similar pattern (see sub-section 5, Figure A3). Overall, the relative change in 90th percentiles is larger than for the means or medians of both length and weight.

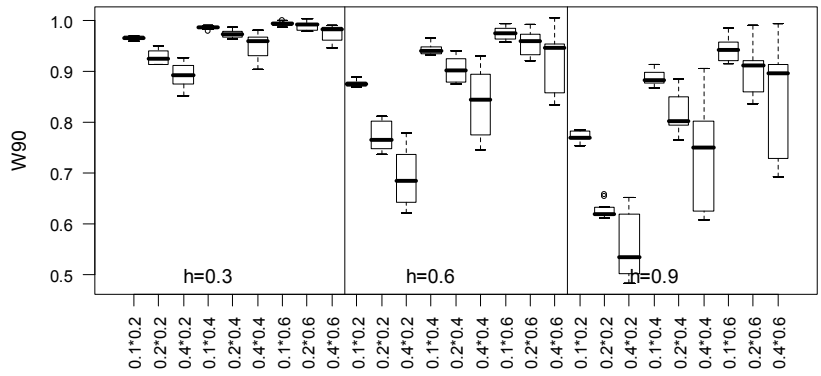


Figure 5 Relative 90th percentile of weight ($W90(F_{msy})/W90(F=0)$) for harvest at F_{msy} given 3 steepness (h) levels. Numbers on the x-axis refer to m*k pairs with m=(0.1,0.2,0.4) and k=(0.2,0.4,0.6). Factors reflected in the 'boxes' are age at maturity (t_m) and age at first capture (t_c).

When selectivity is dome-shaped, the oldest, largest fish are not caught and therefore not sampled for the size-based indicators. Their absence would also affect the CPUE-based indicators, but one would expect the effect to be greatest on the size-based indicators. A comparison of the relative change in mean length for different selectivity curves shows that the relative change is smaller when selectivity is dome-shaped, particularly when the dome is narrow (i.e. contains very few fully selected age classes).

It is also unsurprising that the 90th percentiles of weight (Figure 7) and length are more affected by the different selectivity curves than the means and medians. Similar graphs of results for other indicators and B_{msy}/B_0, and for each of the three types of selectivity curve, are shown in Annex 1.

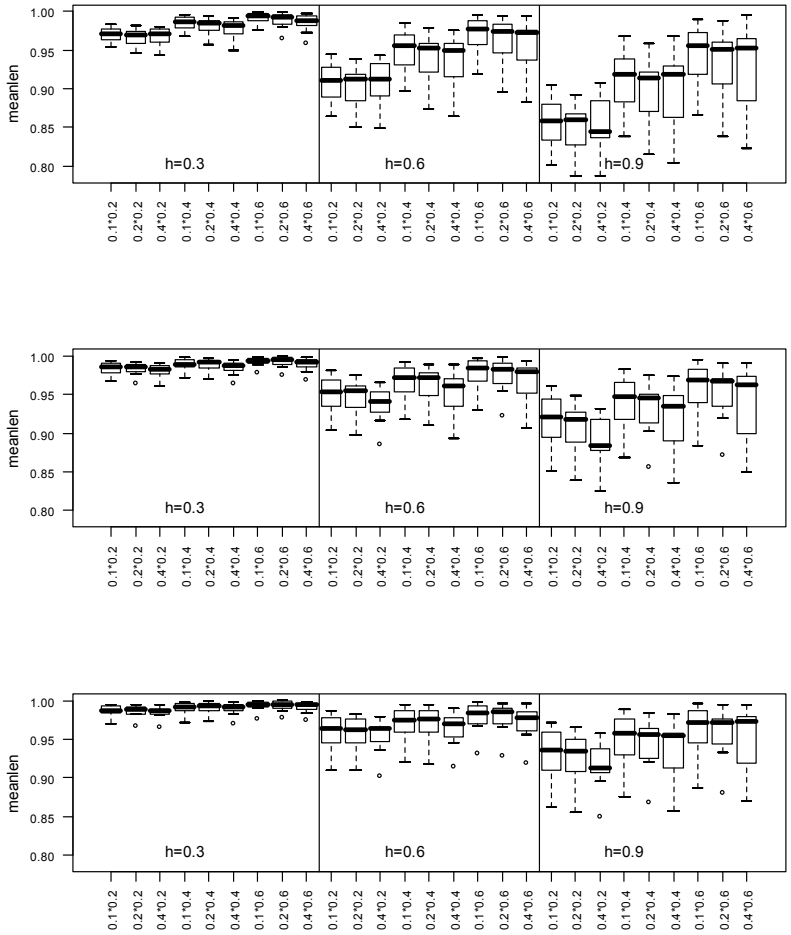


Figure 6. Expected relative change in mean length when harvesting at Fmsy for flat selectivity (top panel; identical to Figure 4), dome-shaped selectivity with 4 fully selected age classes (middle panel) and with 2 fully selected age classes (lower panel). Each panel shows results by steepness level (h) and m*k pairs, as indicated on the x-axis.

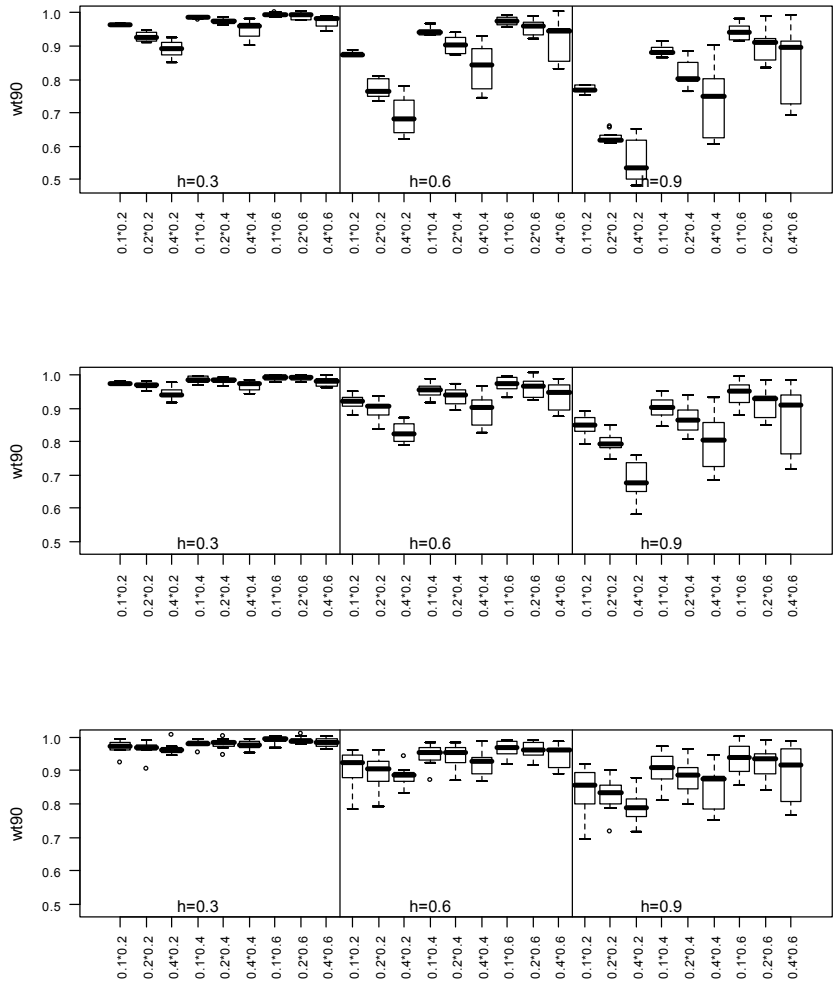


Figure 7. Expected relative change in the 90th percentile of weight when harvesting at F_{msy} for flat selectivity (top panel), dome-shaped selectivity with 4 fully selected age classes (middle panel) and with 2 fully selected age classes (lower panel). Each panel shows results by steepness level (h) and m*k pairs, as indicated on the x-axis.

As noted above, the mean length is not a particularly responsive indicator of spawning biomass, because it changes much less than the spawning biomass. This is also illustrated in Figure 8 where the change in indicator is expressed relative to the change in spawning biomass:

$$\{1 - \text{Indicator}(F_{msy})/\text{Indicator}(F=0)\} / \{1 - B_{msy}/B_0\}$$

An ideal indicator would have large values, close to or above 1 for this performance measure, because the indicator would be changing almost as much or more than spawning biomass. An indicator with low values is likely to be poor in terms of detecting the signal when stochasticity in underlying dynamics and sampling error come into play.

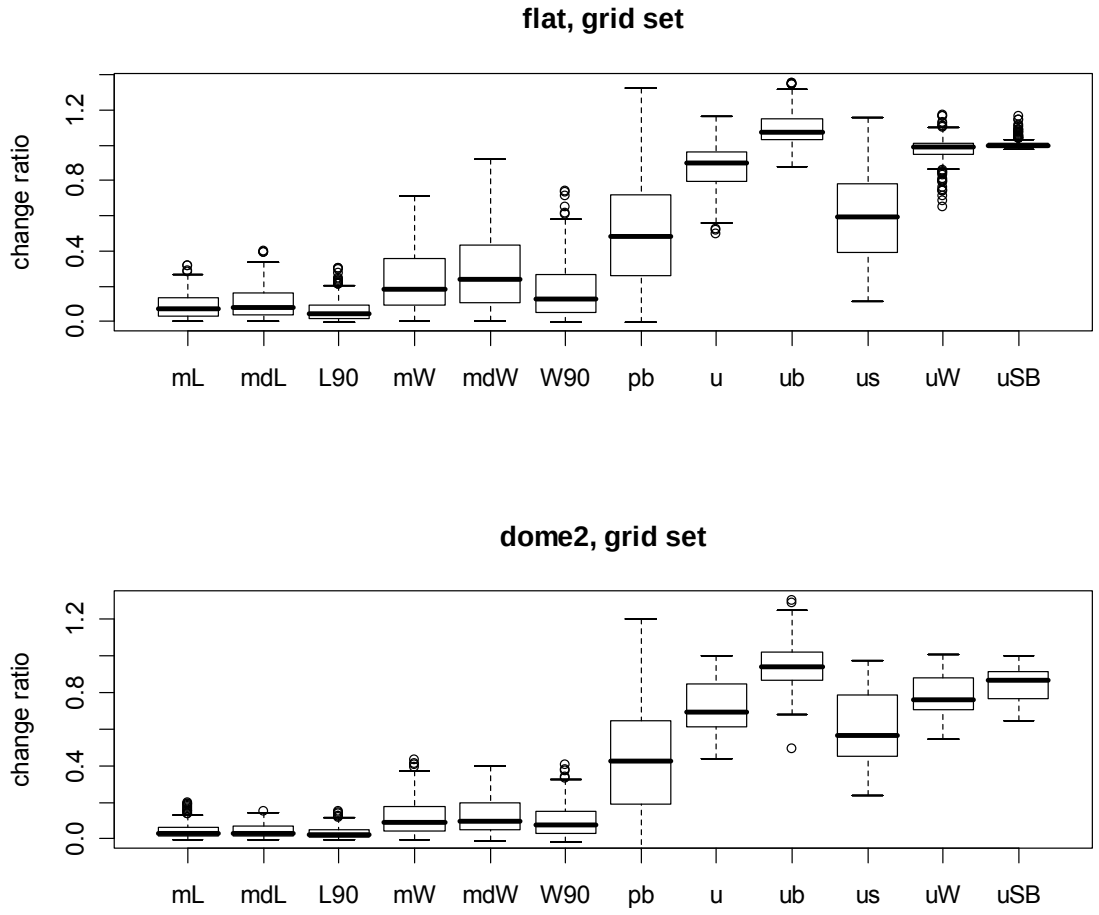


Figure 8. 'Change ratios', [1 – indicator(F_{msy})/indicator(F=0)]/[1-B_{msy}/B_0], for 7 size-based indicators and 5 CPUE-based indicators (see sub-section 2.3 above for descriptions of indicator labels). Scenarios are the grid set of life-history parameters. The top panel is for flat selectivity, and the lower panel is for dome-shaped selectivity with 2 fully selected age classes.

Figure 8 shows that indicators based on weight are generally more responsive than those based on length. This result is not new; see for example Punt *et. al.* (2001). The proportion big indicator is variable in its performance, depending on the size/age at first capture relative to the size/age at maturity. Obviously, if age at first capture is greater than age at maturity, then the proportion mature is not informative, and those cases are included in Figure 8. Unsurprisingly, the CPUE-based indicators are the most responsive, changing as much as, or more than spawning biomass.

8.3.1.2 Results for correlated life-history parameters.

Results in this sub-section are for the wider range of life-history dynamics than the grid given in Table 4, with correlation between m and k, and between m and t_m, as well as between m and t_c (see Figure 1). Results are presented differently here because the life-history parameters are no longer a small number of distinct values, but cover a range of values.

Figure 9, which is for flat selectivity, again shows that both mean length and mean weight are expected to change only slightly compared to SSB when harvested at F_{msy}. The change is particularly small at low steepness, and somewhat greater at high steepness. The range for each

indicator (around a given value of h) now reflects different values for m, k, t_m and t_c. There is not much difference between the mean, median or 90th percentile of length or weight-based indicators – there is more difference between coinciding length and weight based indicators (i.e. mean length vs. mean weight, or median length vs. median weight). The proportion big shows more change than the means, medians and 90th percentiles of either length or weight, particularly at high steepness. (Recall that 'big fish' were defined as the age at first capture plus 3 years. In practice this definition can be fine-tuned to give the best performance for a specific case).

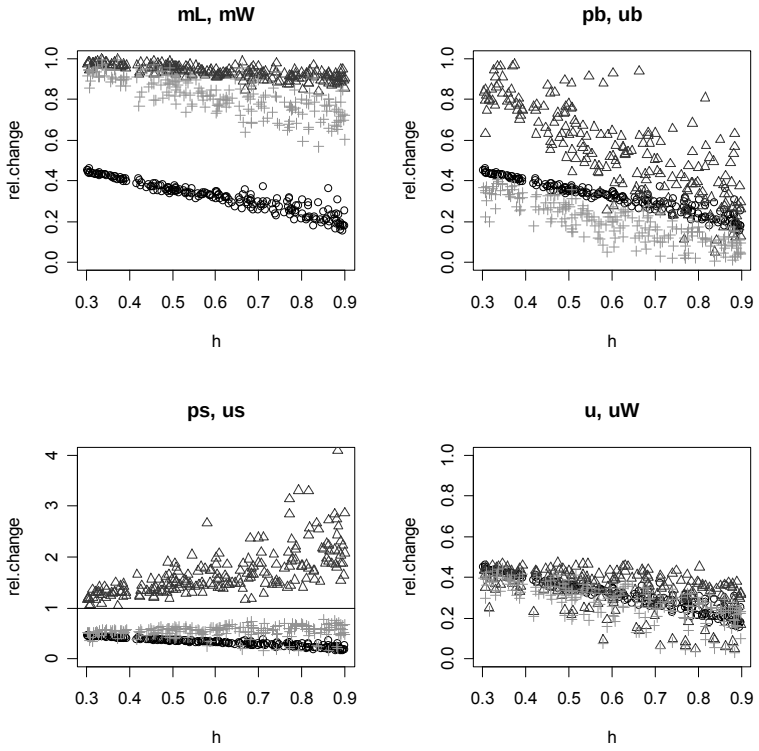


Figure 9. Flat selectivity, Fmsy. Relative change in SSB (circles) and other indicators as a function of steepness (h). Indicators are identified in the title of each panel: triangles relate to the first listed indicator and '+'s to the 2nd indicator in the list.

The proportion small fish increases when the stock is harvested, so it is inversely related to the change in SSB. This indicator also shows most change when steepness is high; CPUE of small fish does change (relative to unexploited) but does not differ much over the range of steepness values. The final panel which shows CPUE in numbers and CPUE in weight both track the relative change in SSB when harvested at F_{msy} well. It is very clear why cpue-based indicators, if unbiased and reliable, are so desirable.

Patterns are very similar for dome-shaped selectivity, but comparisons with results for flat selectivity confirms that relative changes are generally smaller with dome-shaped than flat selectivity.

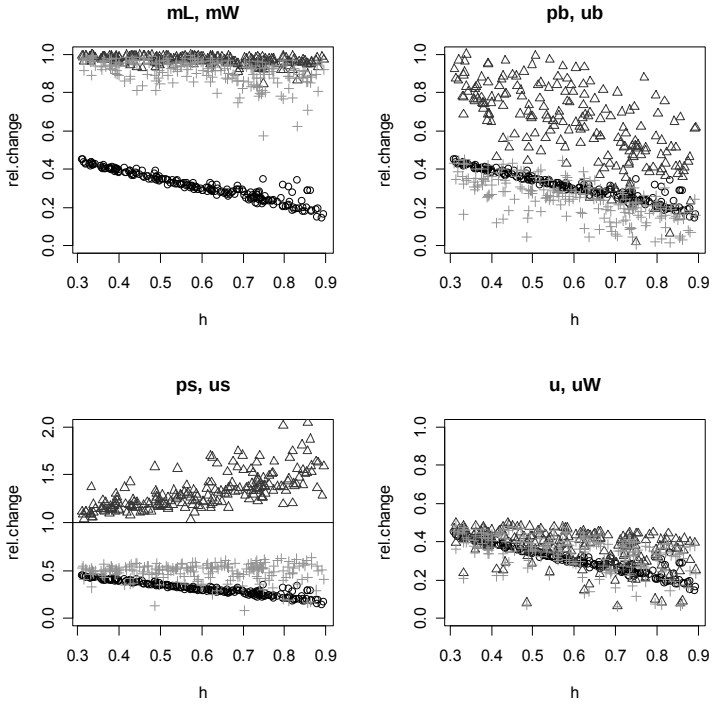


Figure 10. Dome-shaped selectivity, width 4, Fmsy. Relative change in SSB (circles) and other indicators as a function of steepness (h). Indicators are identified in the title of each panel: triangles relate to the first listed indicator and '+'s to the 2nd indicator in the list.

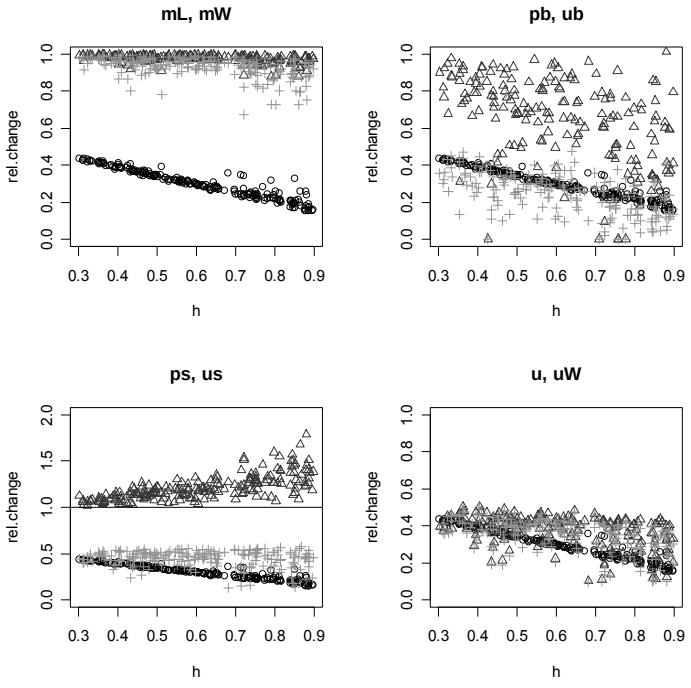


Figure 11. Dome-shaped selectivity, width 2, Fmsy. Relative change in SSB (circles) and other indicators as a function of steepness (h). Indicators are identified in the title of each panel: triangles relate to the first listed indicator and '+'s to the 2nd indicator in the list.

All the above results show a very strong effect of steepness on the expected relative change in indicators when harvesting at F_{msy}. This is unsurprising, because steepness has a strong influence on the magnitude of F_{msy}. Any harvest rate less than F_{msy} would lead to even smaller expected

changes in indicators and therefore even less chance of detecting such a change when natural variability in the system (e.g. via recruitment variability) and sampling/measurement errors are superimposed on the theoretical, deterministic results.

Results from the grid of life-history parameters suggested that the indicators based on 90th percentiles have a larger expected change than those based on means (or medians) when mortality is high relative to growth rate, i.e. m/k>1, and when selectivity is flat. We look at this for the correlated dataset by plotting the difference between the relative change in the 90th percentile of length (or weight) and the relative change in the mean of length (or weight) against the ratio m/k. A negative value for the difference indicates that the 90th percentile indicator is more responsive than the mean-based indicator. Figure 12 confirms the result for flat selectivity since most of the negative points lie in the region where m/k>1. As suggested in the boxplots for the grid dataset (e.g. Figure 8), this pattern does not persist when selectivity is dome-shaped.

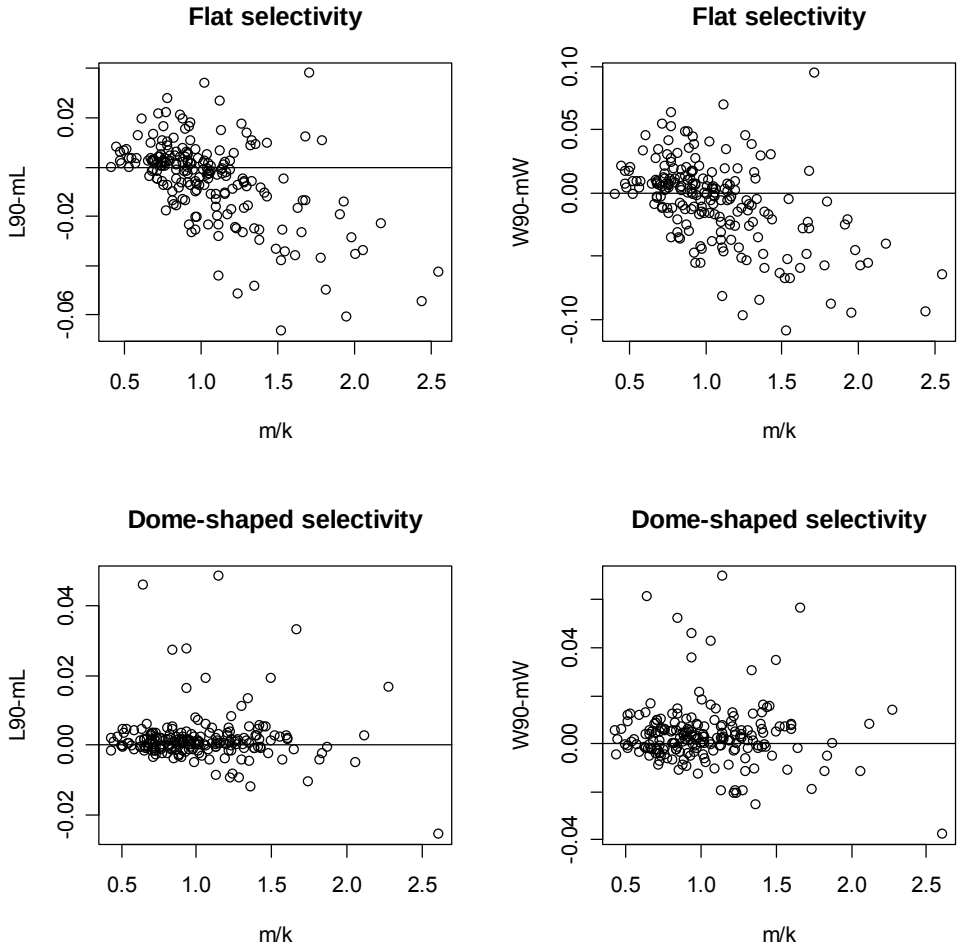


Figure 12. Differences between relative changes in mean and 90th percentile indicators for flat and dome-shaped selectivity (with 2 age classes fully selected), plotted against the ratio of natural mortality to growth rate (m/k). The left panels are for length, the right panels for weight. Points show: [relative change in the 90th percentile-based indicator] – [relative change in mean-based indicator]

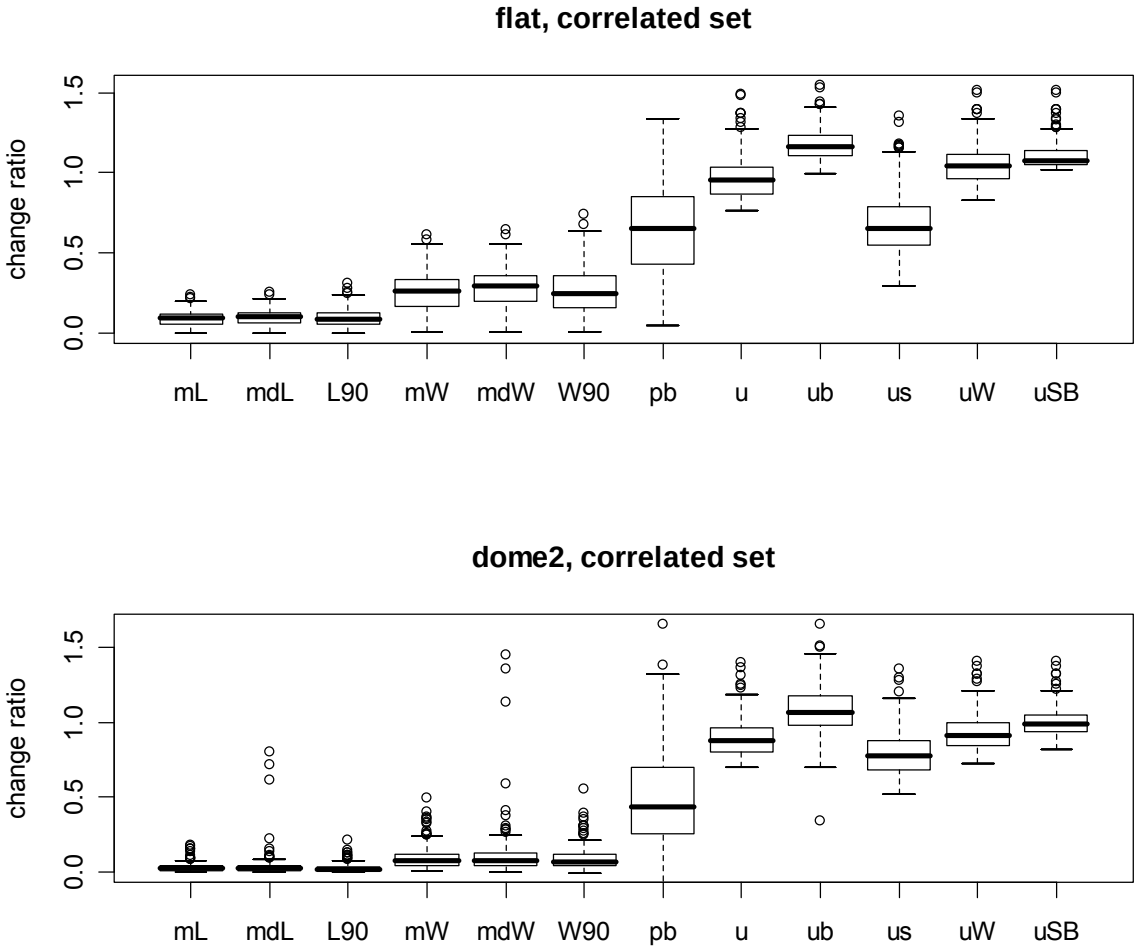


Figure 13. Change ratio, [1 – indicator(F_{msy})/indicator(F=0)]/[1-B_{msy}/B0], for 7 size-based indicators and 5 cpue-based indicators – see sub-section 2.3 for indicator names. Scenarios are for the correlated set of life-history parameters. The top panel is for the flat selectivity curve; the lower panel for dome-shaped selectivity with 2 fully selected age classes.

Figure 13 shows very similar patterns to those seen in Figure 8 (for the grid of life-history parameters), confirming that the unrealistic combinations of parameters did not distort the outcomes.

8.3.2 Potential reference or trigger points

The fourth objective of the project, as originally stated, is to test the robustness (performance) of indicators relative to arbitrary trigger or reference points. Although the detail of this is discussed in Section 9, it is probably useful to briefly explain what is meant by this. The notion is to define a trigger point for each indicator, to run a simulation which generates a time-series of indicator values. At each time step we then determine whether the indicator is above or below its trigger point. We also determine whether the spawning biomass (B) is above or below a specified reference point, and then calculate how often the indicator was 'wrong'. In other words, how

often the indicator was above and B below, or the indicator was below and B above the relevant reference or trigger points[12].

Use of the word 'arbitrary' requires some explanation. When putting the proposal together, we anticipated that, at the time that this part of the work was being done, reference points or management objectives would not yet have been agreed. The intention of the word arbitrary, was therefore to distinguish between 'agreed' reference points (or reference points that would meet a certain objective) and reference points which could be used for comparative purposes, i.e. to evaluate the performance of the indicators. The word arbitrary is also not meant to suggest that one can arbitrarily choose any value to use as a trigger point for each indicator, because this approach would not provide a fair comparison of the performance between indicators. Instead we use a common concept, such as F_{msy} or the harvest rate coinciding with some relative spawner per recruit (SPR) level, to calculate the expected values of each of the indicators at that harvest rate. This set of expected values can then be used as a set of trigger points which should lead to a 'fair' comparison of the performance of the indicators.

Note that the expected (deterministic) value of an indicator, given a certain harvest rate, can be used both as (a) an indication of the responsiveness of an indicator and (b) as a potential reference point. In addition, given the notion of F_{msy} as a limit reference point (rather than a target reference point; see Appendix 2 on Reference points), it is clear why indicator values coinciding with F_{msy} are treated as measures of 'maximum' response. Here, however, we focus more on the use as potential reference points.

Results presented so far were based on F_{msy}. We have already noted that steepness is unlikely to be known at all. Spawner-per-recruit (SPR)-based reference points do not require an assumption about steepness. Here we briefly consider the expected changes in indicators at a harvest rate of F_{spr40} and compare these values to those at F_{msy}. Recall that F_{spr40} will be identical for life-histories with the same {m,k,t_m,t_c (or selectivity)}, even if steepness differs. Figure 14 shows that F_{spr40} will tend to be conservative (lower than F_{msy}) for high steepness and values down to about h=0.6, but will be too high for steepness below 0.5. The pattern is similar for flat and dome-shaped selectivity. To achieve more robustness at low steepness, a higher SPR ratio could be chosen (e.g. 50%), but this will of course mean more conservatism if steepness is in fact high.

The implications of the two different types of harvest rates on the expected values of some indicators are shown in Figure 16 for flat selectivity. The figures for mean length and the 90th percentile of length are very similar to that for median length. Similarly, those for mean weight and the 90th percentile of weight are similar to that for median weight. Figures for the CPUE-based indicators are also very similar to one another. The same information is shown in a slightly different way in Figure 17 for flat selectivity and Figure 17 for dome-shaped selectivity. This again shows that the SPR-based reference points would be conservative if steepness is high, but risky if steepness is low. The figures also, however, show that the differences are small for the length-based indicators, but can be quite substantial for CPUE-based indicators.

[12] We deliberately distinguish between 'trigger points' for indicators and 'reference points' for spawning biomass and harvest rate. We do this because we consider that when used in a decision rule, the trigger points for indicators may have to be adjusted to achieve the required objectives for spawning biomass and harvest rate, and may not be the values that coincide directly with, say, the harvest rate associated with the SSB reference point.

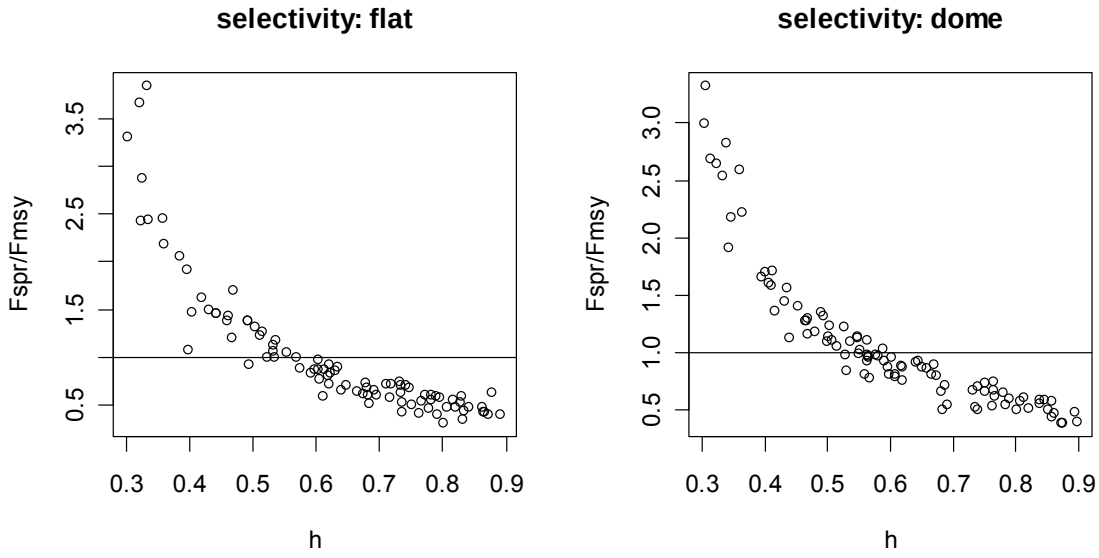


Figure 14. Ratios, F_{spr40}/F_{msy} plotted against steepness (h) for flat and dome-shaped selectivity (of width 2 age classes) and for the correlated set of life-history parameters.

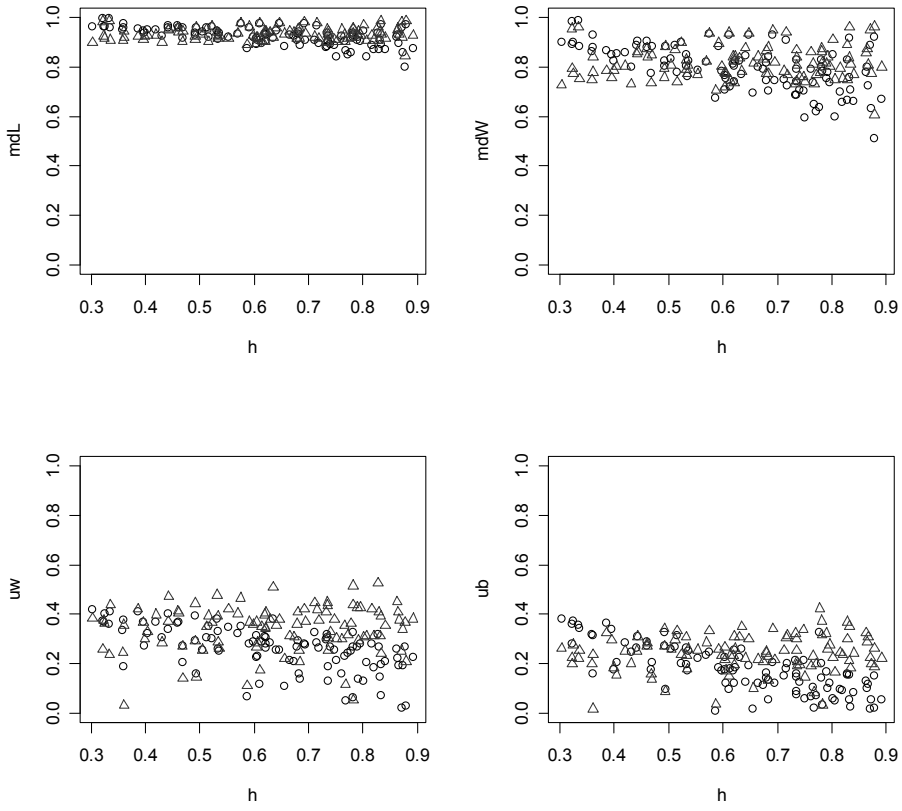


Figure 15. Relative values for indicators under the harvest rate F_{SPR40} (triangles) and F_{MSY} (open circles) plotted against steepness (h). The panels are for median length (mdL, top left), median weight (mdW, top right), cpue in weight (uw, bottom left) and cpue of big fish (ub, bottom right). Results are for the correlated life-history dataset with flat selectivity.

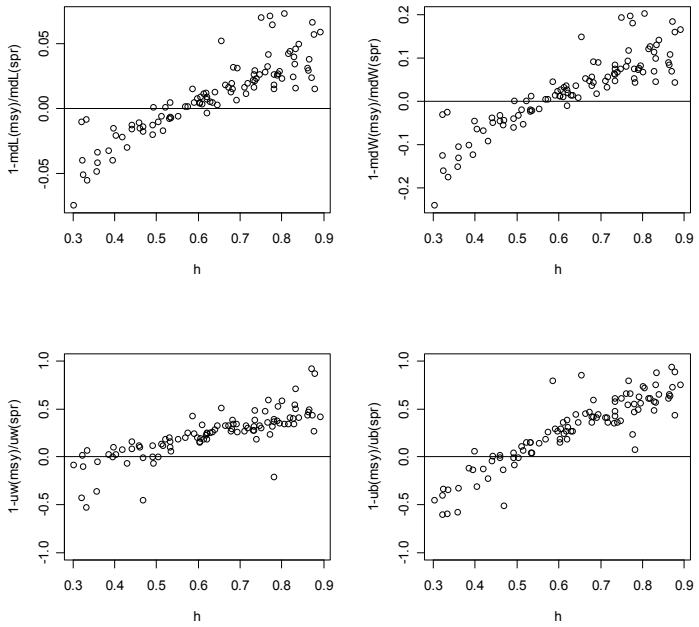


Figure 16. Relative differences between expected (relative) indicator values under the harvest rate F_{SPR40} and F_{MSY}, plotted against steepness (h). The panels are for median length (mdL, top left), median weight (mdW, top right), cpue in weight (uw, bottom left) and cpue of big fish (ub, bottom right). Results are for the correlated life-history dataset with flat selectivity.

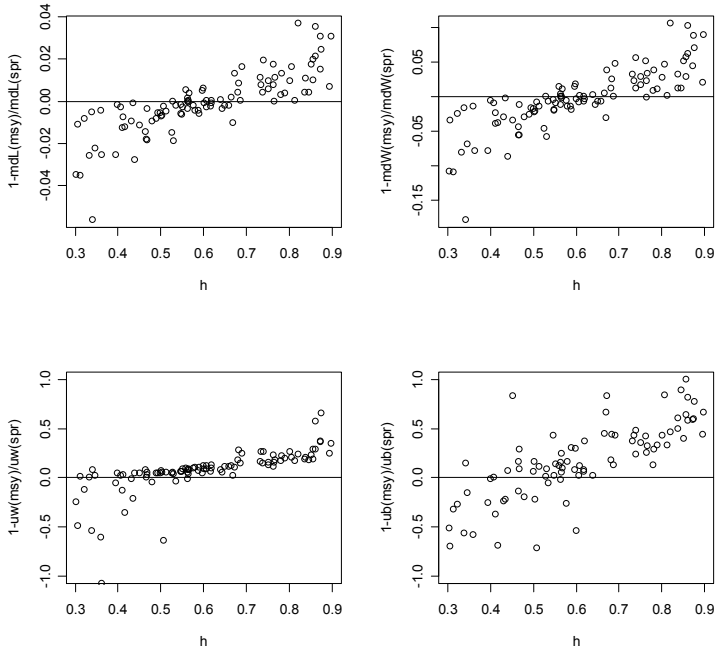


Figure 17. Relative differences between expected (relative) indicator values under the harvest rate F_{SPR40} and F_{MSY}, plotted against steepness (h). The panels are for median length (mdL, top left), median weight (mdW, top right), cpue in weight (uw, bottom left) and cpue of big fish (ub, bottom right). Results are for the correlated life-history dataset with dome-shaped selectivity.

8.3.3 The effect of recruitment variability on indicators.

Common sense suggests that quantities based on several age classes, for example the spawning biomass, should have less variability than numbers in the first age class, the recruits. If the total mortality is low and there are many age classes in the spawning stock or catch then this 'damping' effect is likely to be strong. For a short-lived stock one would expect the 'damping' to be weaker, and the variability of the spawning stock biomass and catch therefore more similar to the variability in recruitment. Time series of quantities based on several age classes, such as spawning biomass or mean length of the catch will also be autocorrelated. In this sub-section we look at the CV and, in the next sub-section, at the autocorrelation of indicators resulting from different levels of variability in recruitment.

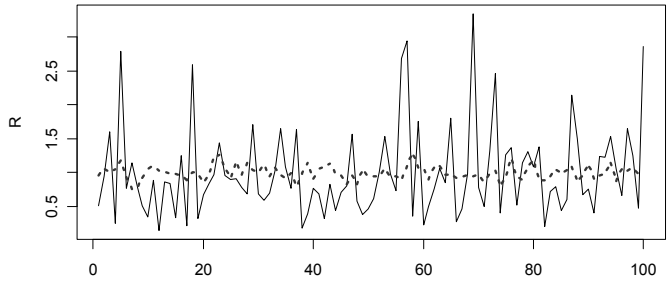


Figure 18. Time series of recruitment (relative to the mean) showing high variability (σ_R=0.7, solid line) and low variability (σ_R =0.1, dashed line).

Recruitment variability is generated as a log-normal distribution about the mean recruitment from the stock-recruit relationship. Figure 18 shows two examples of variability in recruitment (and no explicit autocorrelation). The plot is an extract of 100 values from the middle of a series of 250. At this point in the time-series, the effect of starting with an unexploited stock, has disappeared. Indicators, such as median weight (mdW) for example (Figure 19, showing the same 100 time-steps as the recruitment series above), however, exhibit some autocorrelation and much less variability than recruitment. This example is based on low natural mortality (m=0.1) and a relatively high harvest rate of F=0.5.

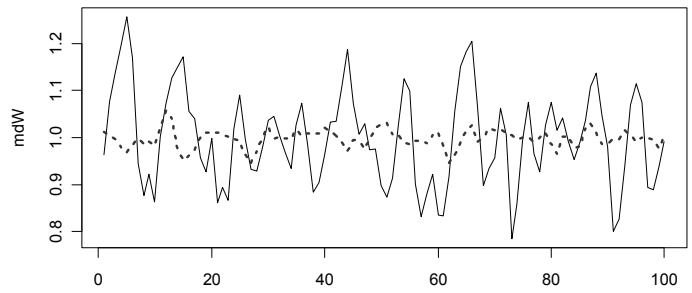


Figure 19. Time series of median weight (relative to the mean) for the high recruitment variability (σ_R=0.7, solid line) and low recruitment variability (σ_R =0.1, dashed line).

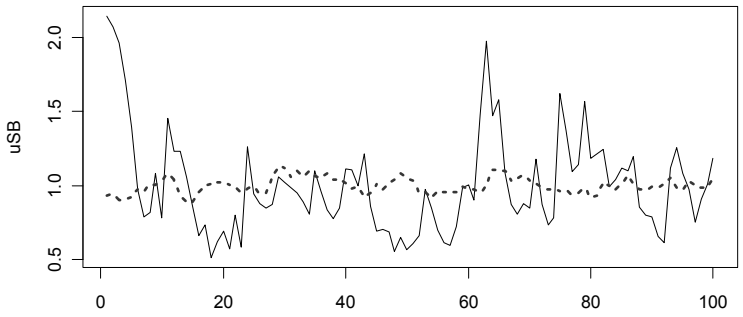


Figure 20. Time series of cpue of spawners (relative to the mean) for the high recruitment variability (σ_R =0.7, solid line) and low recruitment variability (σ_R =0.1, dashed line).

The CPUE-based indicators show much higher variability than the size-based indicators, particularly when recruitment variability is high. This example is from the same simulation as the median weight shown above. Figure 20 illustrates another interesting feature, namely the time-lag between a very high pulse in recruitment and the effect of that becoming evident in the CPUE. The high first point on the solid line in Figure 20 is due to a high peak in recruitment 6 years earlier (and therefore not evident from the recruitment plot, Figure 18).

The harvest rate also has an effect on the variability of indicators since it modifies the age structure and the number of age classes in the population. A comparison between the extent to which a range of harvest rates affect the CVs of indicators and of population abundance, and the extent to which different recruitment variabilities (at the same F-value) affect these quantities, show that recruitment variability is likely to be the more important factor (Figure 21).

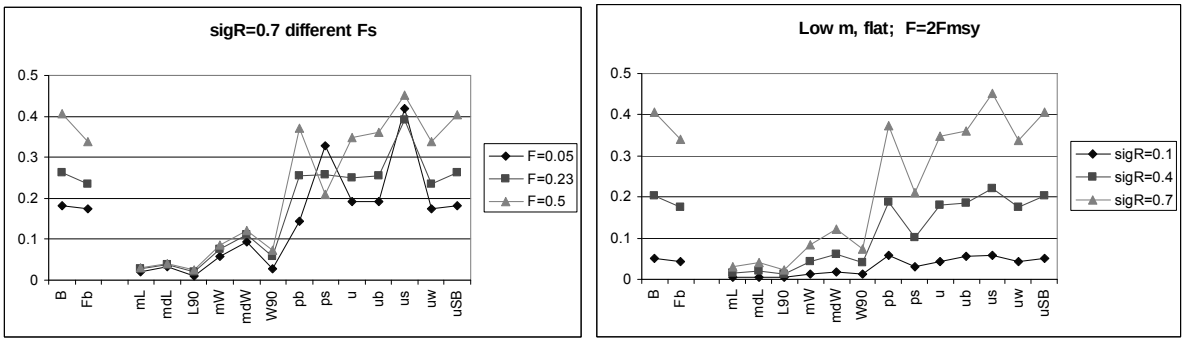


Figure 21. CVs of spawning biomass (B), fishable biomass (Fb) and 13 size- and cpue-based indicators as a function of (i) different harvest rates (F, left panel) at a recruitment CV of 40% and (ii) different recruitment variabilities (sigR=σ_R, right panel) at F=2*F_{msy}=0.5.

Figure 21 is again based on low m (m=0.1), and shows that the CPUE-based indicators have similar CVs to the population abundance (spawning and fishable biomass). The size-based indicators generally have much lower CVs, with the length-based indicators having lower CVs than the weight-based indicators. Recall that these quantities are simulated with no measurement error and with large numbers in the population and therefore in the catch samples. The only other random component in the simulated data comes from the fact that the size frequency sample is based on a mixed normal distribution (see Methods sub-section).

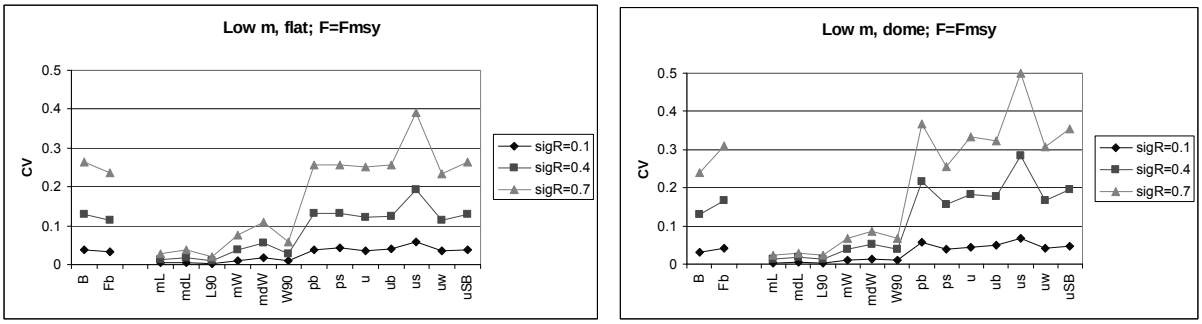


Figure 22. CVs of spawning biomass (B), fishable biomass (Fb) and 13 size- and cpue-based indicators as a function of different recruitment variabilities (sigR=σ_R) and (i) flat selectivity (left panel) or (ii) dome-shaped selectivity (right panel). In all cases the harvest rate was F=F_{msy}, which is 0.23 for the flat and 0.36 for dome-shaped selectivity.

The catch resulting from dome-shaped selectivity consists of a larger proportion of small fish compared to flat selectivity, and this implies somewhat higher CVs than for the flat selectivity scenario. Figure 22 shows that this is the case, particularly for the higher levels of recruitment variability. It also shows that the 'proportion big' (pb) indicator has a higher CV relative to the other indicators when selectivity is dome-shaped than when it is flat. Recall, however, that F_{msy} is higher for dome-shaped than flat selectivity.

When natural mortality is relatively high (m=0.4) and the harvest rate is again, the difference between dome-shaped and flat selectivity is much less. This is because there is less accumulated biomass in the older age classes so the proportions in the catch are relatively small even if selectivity is flat.

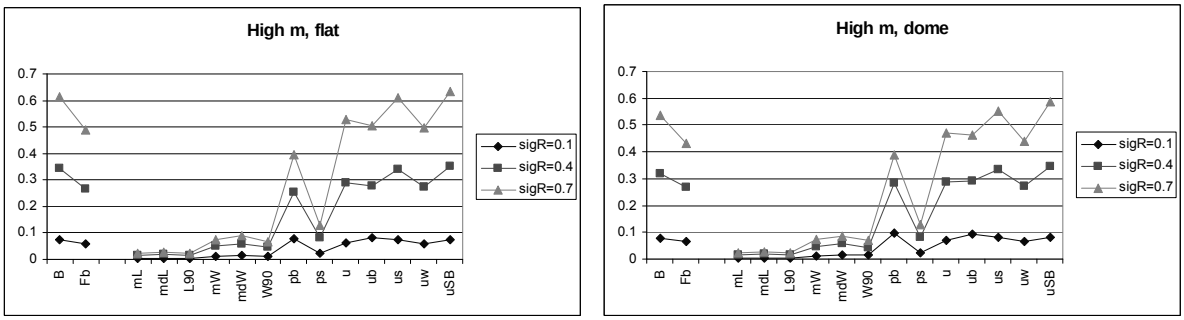


Figure 23. CVs of spawning biomass (B), fishable biomass (Fb) and 13 size- and cpue-based indicators as a function of different recruitment variabilities (sigR=σ_R) for the high natural mortality case and (i) flat selectivity (left panel) or (ii) dome-shaped selectivity (right panel). In all cases the harvest rate was F=F_{msy} which is 1.2 for both flat and dome-shaped selectivity.

Now the CVs for the flat selectivity curve are slightly higher than for the dome-shaped selectivity when recruitment variability is 0.7, but this could simply be because these CVs are from single runs rather than mean CVs from many runs. Limited explorations with multiple runs confirmed that the overall obvious patterns shown above persist, although there are slight differences between different realisations.

8.3.4 Autocorrelation of indicators

We again consider the two examples of low natural mortality (m=0.1) and relatively high natural mortality (m=0.4), and for each we look at flat selectivity and dome-shaped selectivity. Figure 24 shows that, as expected, the autocorrelation dies out much faster when mortality is high than when it is low, particularly when the selectivity curve is flat rather than dome-shaped. It is also interesting to note how the autocorrelation becomes negative before dying out. As a pulse of high recruitment enters the catch, it would initially lead to a drop in mean length (or weight) because of a high number of small fish in the catch. Later, as individuals in the cohort grow, there will be a large number of big fish (recall that the mortality rate is constant and therefore the same for each cohort) leading to a higher mean length and negative autocorrelation for a time-lag of some years. In these examples, the negative autocorrelation starts at around a lag of 2 years for the high mortality case and around 4 years for the low mortality case when selectivity is flat.

Also note that the autocorrelation becomes negligible after different time-periods for the different scenarios. The longest time-period is for the low mortality with flat selectivity (about 10 years or more). The shortest is for the high mortality with either flat or dome-shaped selectivity (about 4 years).

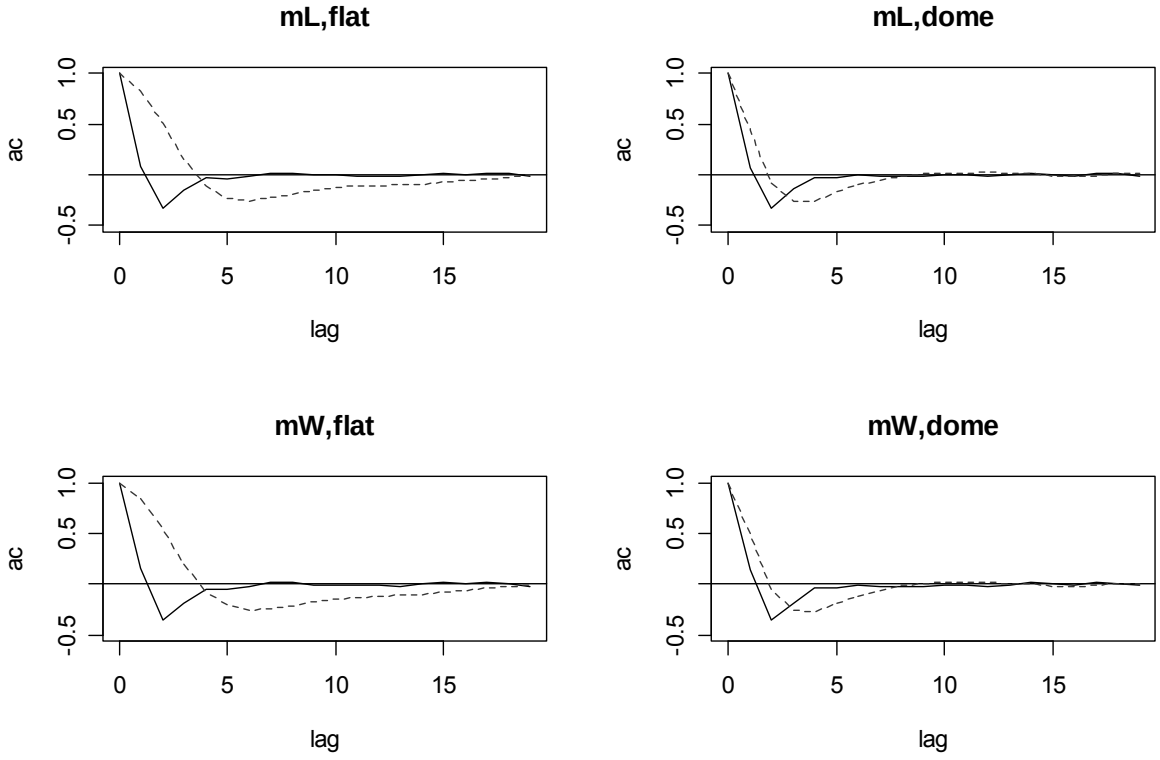


Figure 24. Autocorrelation in mean length (mL) and mean weight (mW) for flat and dome-shaped selectivity. The solid line is for a population with high m (=0.4); the dashed line is for a population with low m (=0.1). Both examples were harvested at F_{msy} and had a high recruitment variability, (sigR=σ_R=0.7).

The patterns of autocorrelation is very similar for mean, median and 90th percentile of length, and for mean, median and 90th percentile of weight. The 'proportion big' indicator (pb) also behaves similarly to the size-based indicators, and the autocorrelation patterns are very similar for all the CPUE-based indicators (Figure 25). What is particularly noticeable in the autocorrelation pattern

for CPUE, is that it does not become negative, before decaying, but remains positive. This is also the case for the CPUE indicators in weight (e.g. uW and uSB).

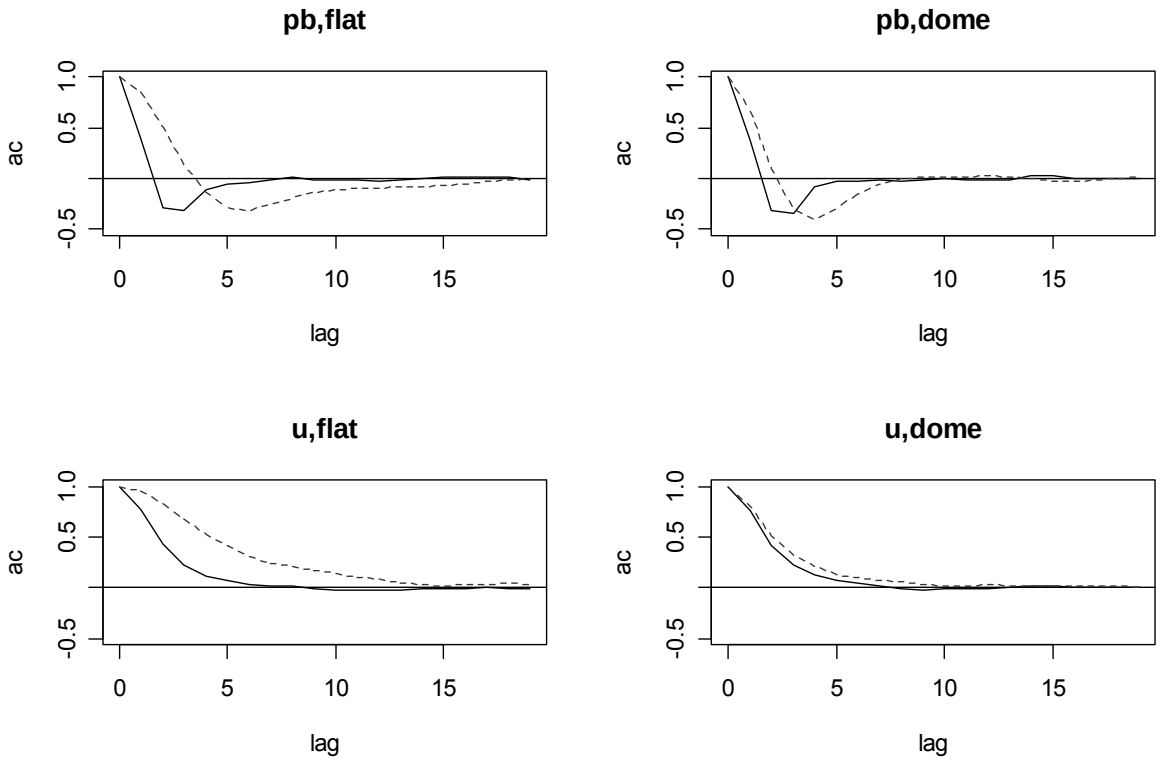


Figure 25. Autocorrelation in the proportion big fish (pb) and cpue in numbers (u) for flat and dome-shaped selectivity. The solid line is for a population with high m (=0.4); the dashed line is for a population with low m (=0.1). Both examples were harvested at F_{msy} and had a high recruitment variability, (sigR=σ_R=0.7).

8.3.5 Return times following a rapid decrease in F

Return times for all the indicators under the four scenarios (low m, high m, combined with flat or dome-shaped selectivity) are shown in Table 6 for a relatively high recruitment variability (σ_R=0.40).

It is clear, and unsurprising, that return times are much shorter when natural mortality is high than when it is low. Recall that the higher natural mortality scenario has an associated higher F_{msy} than the low mortality scenario, so that total mortality is also much higher. There is again very little difference between results for flat- and dome-shaped selectivity when mortality is high. When mortality is low, however, the flat selectivity scenario has longer associated return times than the dome-shaped selectivity. This is partly because F_{msy} is higher when selectivity is dome-shaped. It is also because the size-based indicators have a more limited range of values they can take on (the catch contains a smallish subset of ages compared to flat selectivity). In general, indicators based on large animals (e.g. L90, W90, pb, ub) have longer return times. The weight-based indicators also have longer return times than the length-based indicators. The CPUE indicator in numbers (u) has shorter return times than the weight-based CPUE indicators (uW,uSB).

Table 6. Return times (for an indicator to get within 5% of its long term mean at F_{msy}) following a drop harvest rate from $3*F_{msy}$ to F_{msy}. Steepness is 0.999, and recruitment variability is relatively high: sigR=σ_R=0.4.

	Low m, flat	Low m, dome	High m, flat	High m, dome
mL, mdL, L90	5, 5, 6	1, 1, 4	1, 1, 1	1, 1, 1
mW, mdW, W90	7, 6, 9	4, 4, 5	2, 2, 2	2, 2, 2
pb, ps	9, 6	6, 4	2, 2	2, 2
u, ub, us	5, 11, 3	3, 6, 2	1, 2, 1	5, 6, 5
uW, uSB	8, 8	3, 3	1, 2	5, 6

Results above are for high steepness. When steepness is low, the return times are all much longer, because recruitment changes more between the over-exploited case and the constant harvest at F_{msy}.

Finally, as noted in the methods sub-section, estimates given here are from a single realisation within each scenario. If this information were to be used directly, it would be preferable to do multiple realisations and calculate the means and variances of return times for each indicator. Here the main purpose is to compare between scenarios and indicators, and in our view this is sufficiently well illustrated by the single realisation.

8.4. Discussion and Conclusions

We started this investigation by asking how a change in the population abundance is likely to manifest itself in an indicator such as mean length of the catch, or catch per unit effort (CPUE), and how strong such a signal would be when there is no variability in recruitment and no measurement error. We defined the signal as the ratio of, the value of the indicator when the stock is harvested at F_{msy}, to the value of the indicator when the stock is not harvested. F_{msy} was used because, as a limit reference point, this should indicate the largest change one would want to observe and hence the strongest signal that one would need to detect in the indicator.

It is not surprising that the extent of the change is strongly driven by the assumed steepness in the stock-recruit relationship. This is because the value of F_{msy} is strongly affected by steepness, with higher harvest rates possible for a stock with high steepness. Given that steepness is unlikely to be known, the broad message from these results is that size-based indicators are only likely to provide a strong signal, and hence perform well, when the growth rate (of body size) is low and when a broad range of age classes are being harvested.

Results also, however, suggest that, if there is evidence or other reasons to believe that a stock might have low steepness (e.g. specifics of reproductive biology), then a size-based indicator may not be responsive enough to detect changes in biomass in time. On the other hand, if a large change in a size-based indicator has been observed in such a case, it would suggest that the harvest rate has already been too high.

Of the indicators considered here, the length-based indicators are least responsive, i.e. have the smallest expected change, for a wide range of life-history parameters. The weight-based indicators are more responsive in most cases, or at worst, they are similar to their length-based counterparts. The strongest responses are expected when steepness is high, growth rate is low, and there is a broad range of fully-selected age-classes in the catch. The 90th percentiles of length

or weight are only likely to be more responsive than the means or medians if natural mortality is high relative to growth rate and, again, when a broad range of age classes are being harvested.

The proportion big indicator (pb) shows a wider range of responses than the 90th percentiles of length or weight. Note, however, that the behaviour of the 'proportion big' indicator will also be a function of how it is specified. In practice, one would explore the definition of this indicator for a given stock in order to optimise its performance. Here we have only used one definition (animals older than the (youngest) age at 50% selectivity plus 3 years) which is unlikely to be optimal in all cases.

CPUE-based indicators show by far the strongest response, or signal. This underlines why they are so common, and why it is so important to consider ways in which reliable, unbiased CPUE information can be obtained or derived from raw data. Although CPUE standardisation is common practice, it is always a concern that not all causes of bias have been (or can) be 'standardised out'. As noted before, two common examples are 'effort creep' and targeting.

Spawner per recruit (SPR) considerations do not require an assumption about steepness. If we compare the harvest rate coinciding with SPR at 40% of the unexploited SPR (F_{spr40}) with F_{msy}, we note that F_{spr40} values are mostly similar to or lower than F_{msy} when steepness is above about 0.5. When steepness is lower, F_{spr40} values are higher than F_{msy}. This has clear implications for the expected changes in indicators when we consider the expected relative changes as potential reference points for indicators. For example, if we assume that steepness is unknown and we consider the expected relative change in an indicator, when the stock is harvested at F_{spr40}, as a limit reference points for that indicator, it implies the following. If steepness is above 0.5, then the F_{spr40} reference point will be conservative. It will be very conservative if steepness is very high. If, however, steepness is really low (<0.5) then the F_{spr40} reference point will be risky. It is possible to be even more robust to low steepness but increasing the SPR ratio to, say 50% and using F_{spr50}, or to be less conservative by decreasing the SPR ratio to, say 30% and using F_{spr30}.

Although the CPUE-based indicators are more responsive than the size-based indicators, they are more strongly affected by recruitment variability. CPUE-based indicators have CVs similar to the CV of the population abundance. The length-based indicators have the lowest CVs (<5%) and the weight based indicators have only slightly higher CVs (< or around 10%) even when recruitment variability is high (CV of 70%).

All indicators are autocorrelated, but when total mortality is high, the autocorrelation drops to zero at much shorter lags than when total mortality is lower. It is interesting that the size-based indicators switch from positive autocorrelation at short lags to a negative autocorrelation, before dying out. This switch in the sign can be understood when we consider a strong pulse of recruitment and a constant harvest rate. When the cohort enters the catch, the mean size will drop below average. The mean is likely to remain 'below average' for some time (depending on the total mortality rate). When the cohort has grown to a large enough size, there will then be more large animals than 'on average', and this will lead to a larger mean size which, in turn, will persist for some years, depending on the total mortality.

CPUE on the other hand, shows declining, but always positive, autocorrelation with lag. The slowest decline occurs when total mortality is low (low m with flat selectivity). For this scenario, the autocorrelation only drops to very low levels after about 10 years. The other scenarios have very low autocorrelation after about 5 years.

The time it takes for an indicator to reach a new equilibrium level (or to be within, say, 5% of it's new long term average) following a step change in the harvest rate, depends on the type of

indicator, the life history parameters, and the extent of change in the harvest rate. When mortality is high, the return times are short and similar for CPUE and size-based indicators When mortality is low, the 90th percentiles take longest to respond, and the length-based indicators respond faster than the weight-based indicators. This makes sense when we note that the larger, older fish contribute most to the 'tails' of the size distributions (and hence to the 90th percentile), and also that the weight-at-age curve is much steeper than the length-at-age curve (weight is approximately length cubed).

It is important to recall that all the simulated indicators considered here, including the size-based indicators, are unbiased. In practice, size frequency data may first need to be standardised for factors such as time (month or quarter), fishing area and possibly some details of gear (e.g. setting practices for longlines). It is obviously important to keep size data from very different gear types, for example longline and purse seine, separate.

The fact that the CPUE-based indicators show by far the strongest response, or signal, underlines why they are so common, and why it is so important to consider ways in which reliable, unbiased CPUE information can be obtained or derived from raw data. Although CPUE standardisation is common practice, it is always a concern that not all causes of bias have been (or can) be 'standardised out'. As noted before, two common examples are 'effort creep' and targeting.

The 'return times' and autocorrelation characteristics of indicators may be important for the design of decision rules, particularly the frequency of application of such a rule.

8.5. Annex of additional figures to Section 8.

The next set of figures are all for $F=F_{msy}$ and top to bottom the 3 panels are for flat, wide dome-shaped (4 age classes wide) and narrow dome-shaped (2 age classes) selectivity curves. Within each steepness level (h), m*k combinations (natural mortality * von Bertalanffy growth rate) are shown on the x-axis.

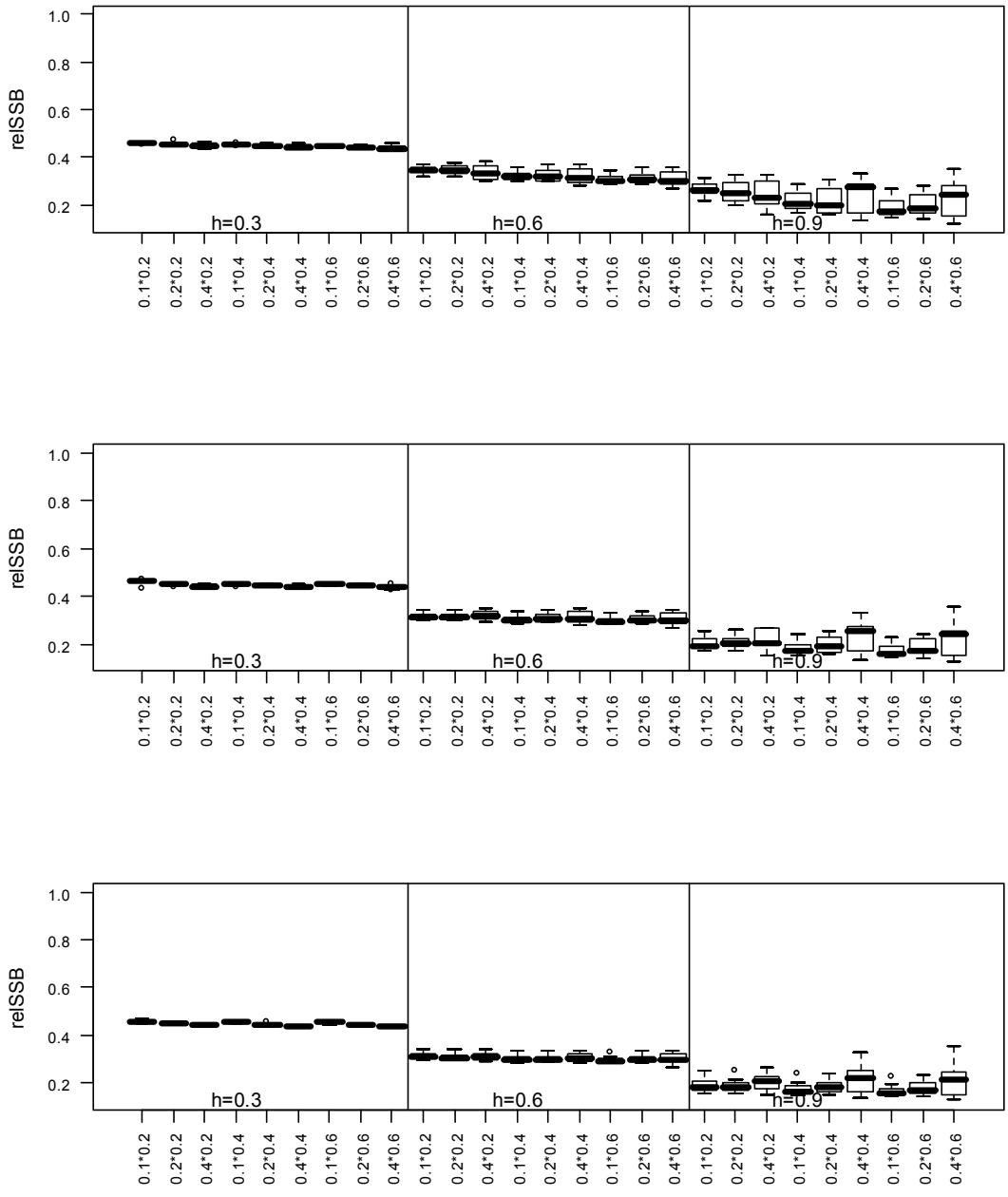


Figure A 1. Relative SSB at F_{msy} for a range of steepness (h) levels and combinations of m and k (shown on the x-axis as m*k). The panels are: flat selectivity (top), dome-shaped with 4 fully selected age classes (middle) and dome-shaped with 2 fully selected age classes (bottom).

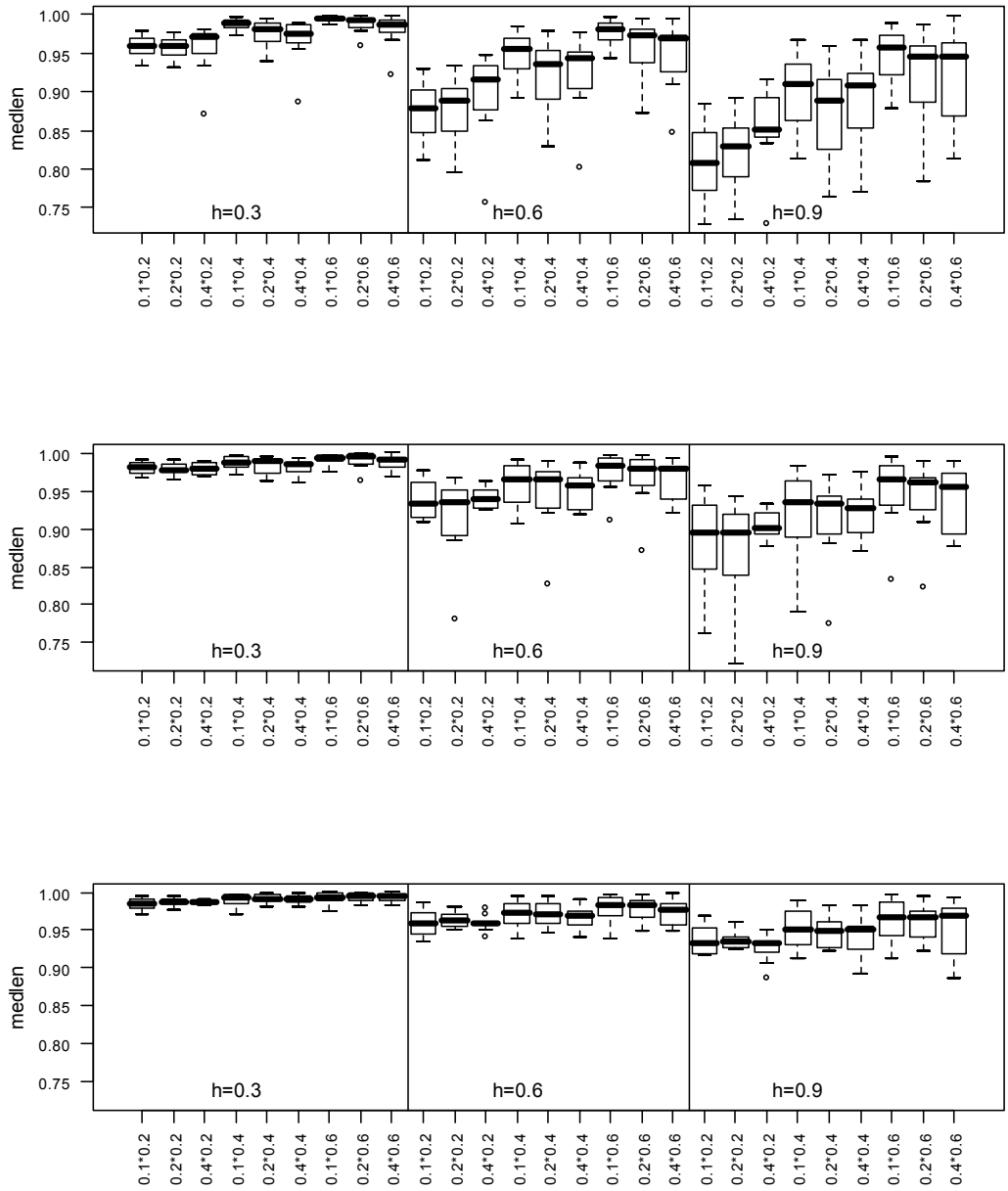


Figure A 2. Relative median length at F_{msy} for a range of steepness (h) levels and combinations of m and k (shown on the x-axis as m*k). The panels are: flat selectivity (top), dome-shaped with 4 fully selected age classes (middle) and dome-shaped with 2 fully selected age classes (bottom).

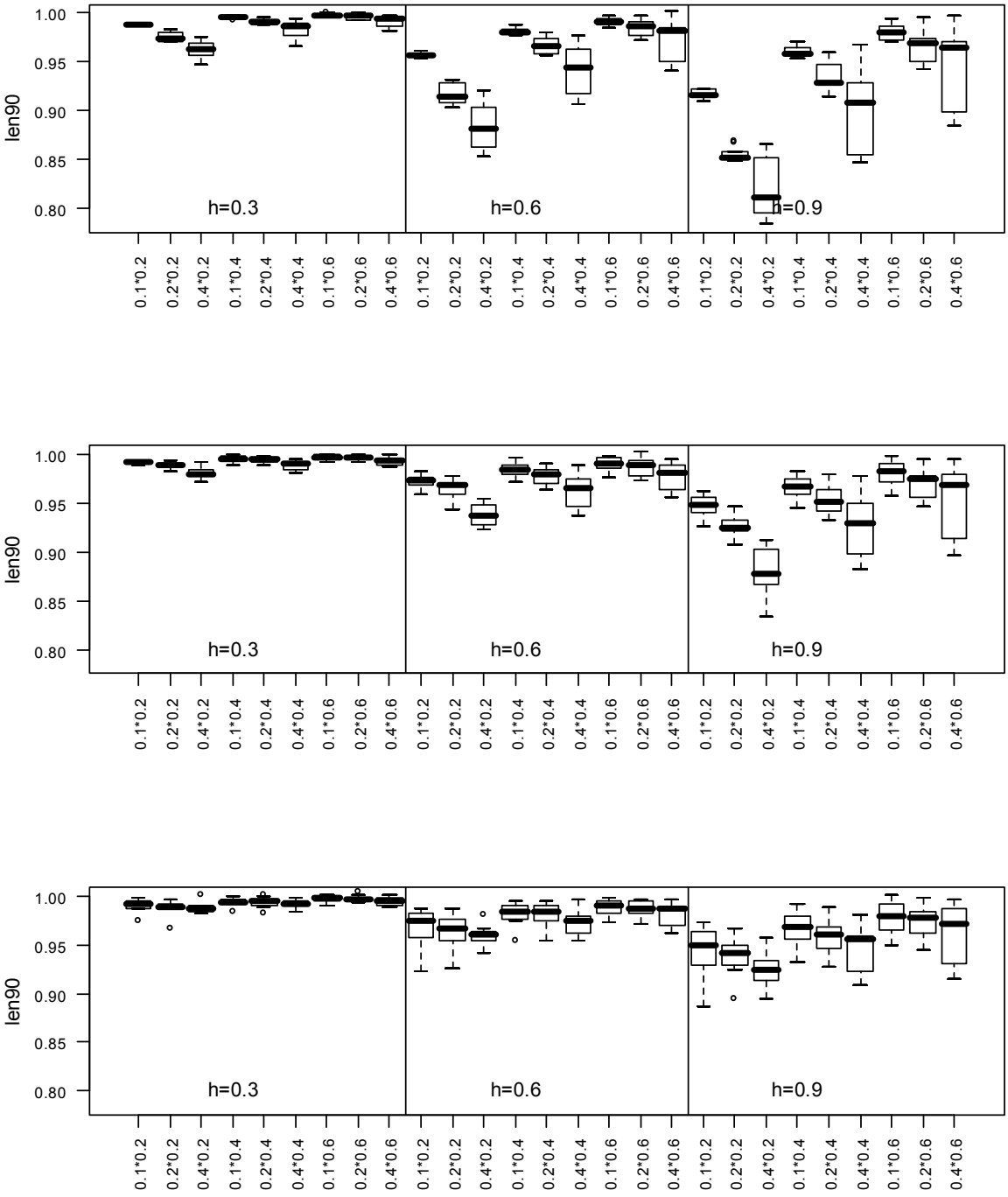


Figure A 3. Relative 90th percentile of length at F_{msy} for a range of steepness (h) levels and combinations of m and k (shown on the x-axis as m*k). The panels are: flat selectivity (top), dome-shaped with 4 fully selected age classes (middle) and dome-shaped with 2 fully selected age classes (bottom).

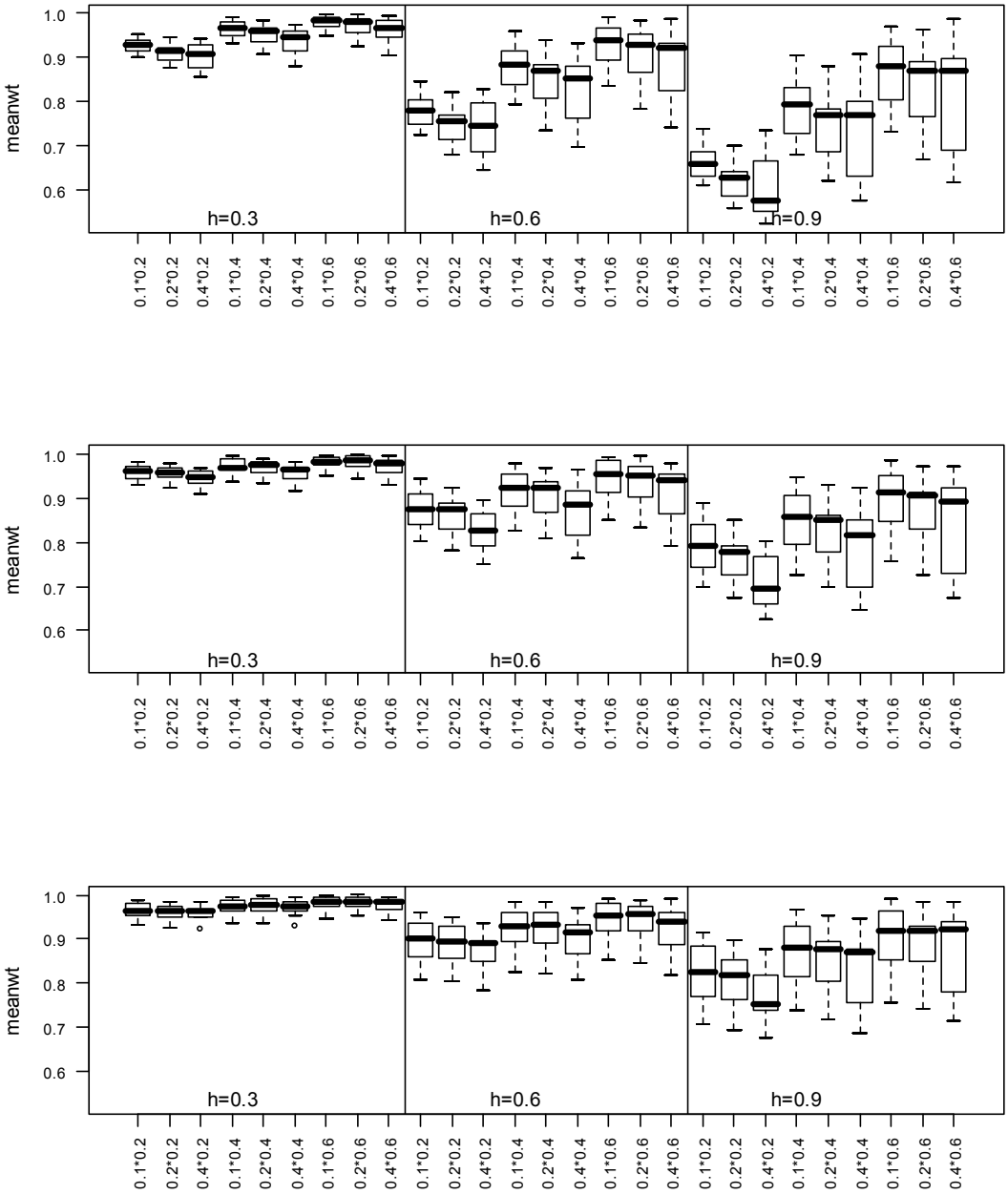


Figure A 4. Relative mean length at F_{msy} for a range of steepness (h) levels and combinations of m and k (shown on the x-axis as m*k). The panels are: flat selectivity (top), dome-shaped with 4 fully selected age classes (middle) and dome-shaped with 2 fully selected age classes (bottom).

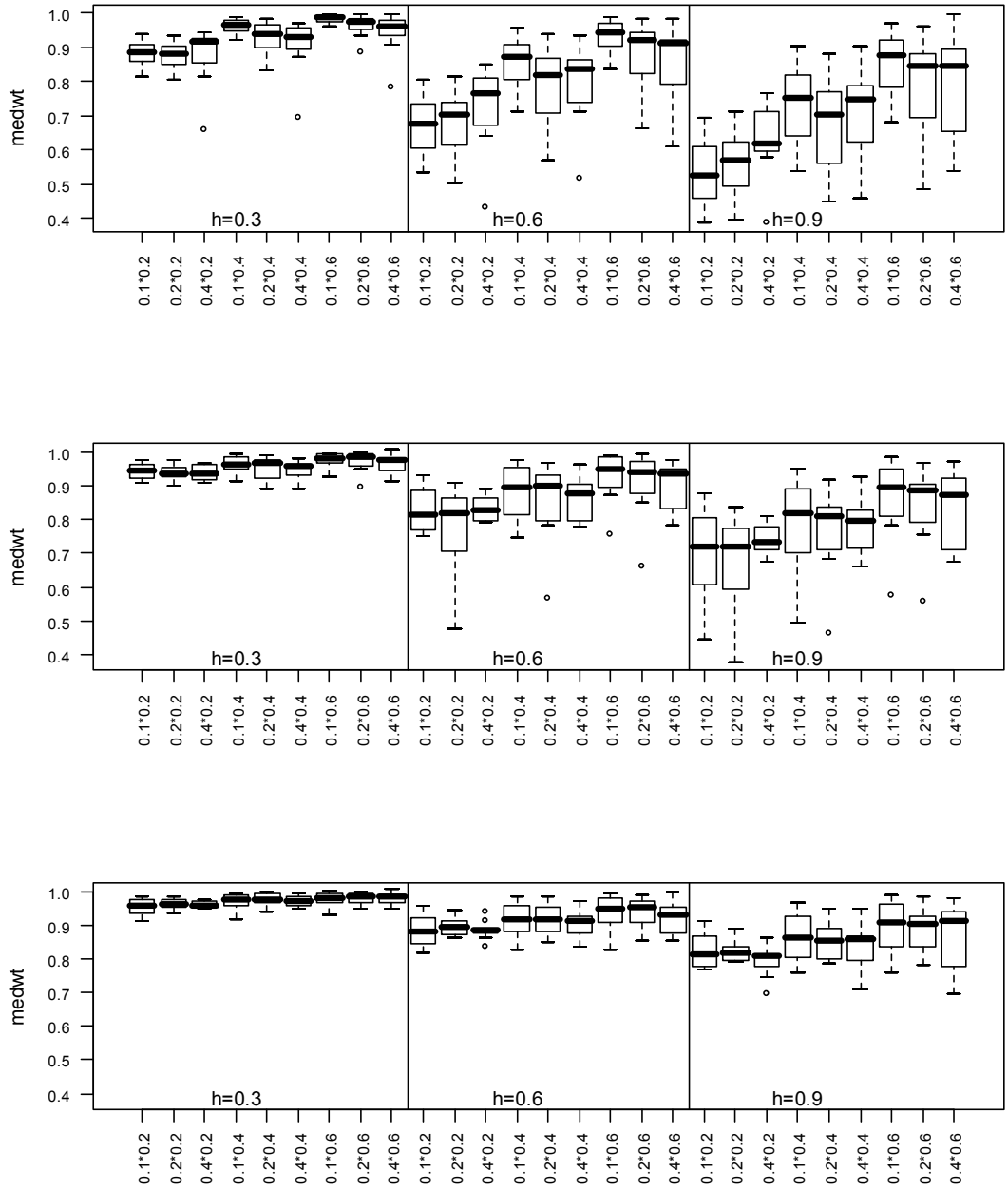


Figure A 5. Relative median weight at F_{msy} for a range of steepness (h) levels and combinations of m and k (shown on the x-axis as m*k). The panels are: flat selectivity (top), dome-shaped with 4 fully selected age classes (middle) and dome-shaped with 2 fully selected age classes (bottom).

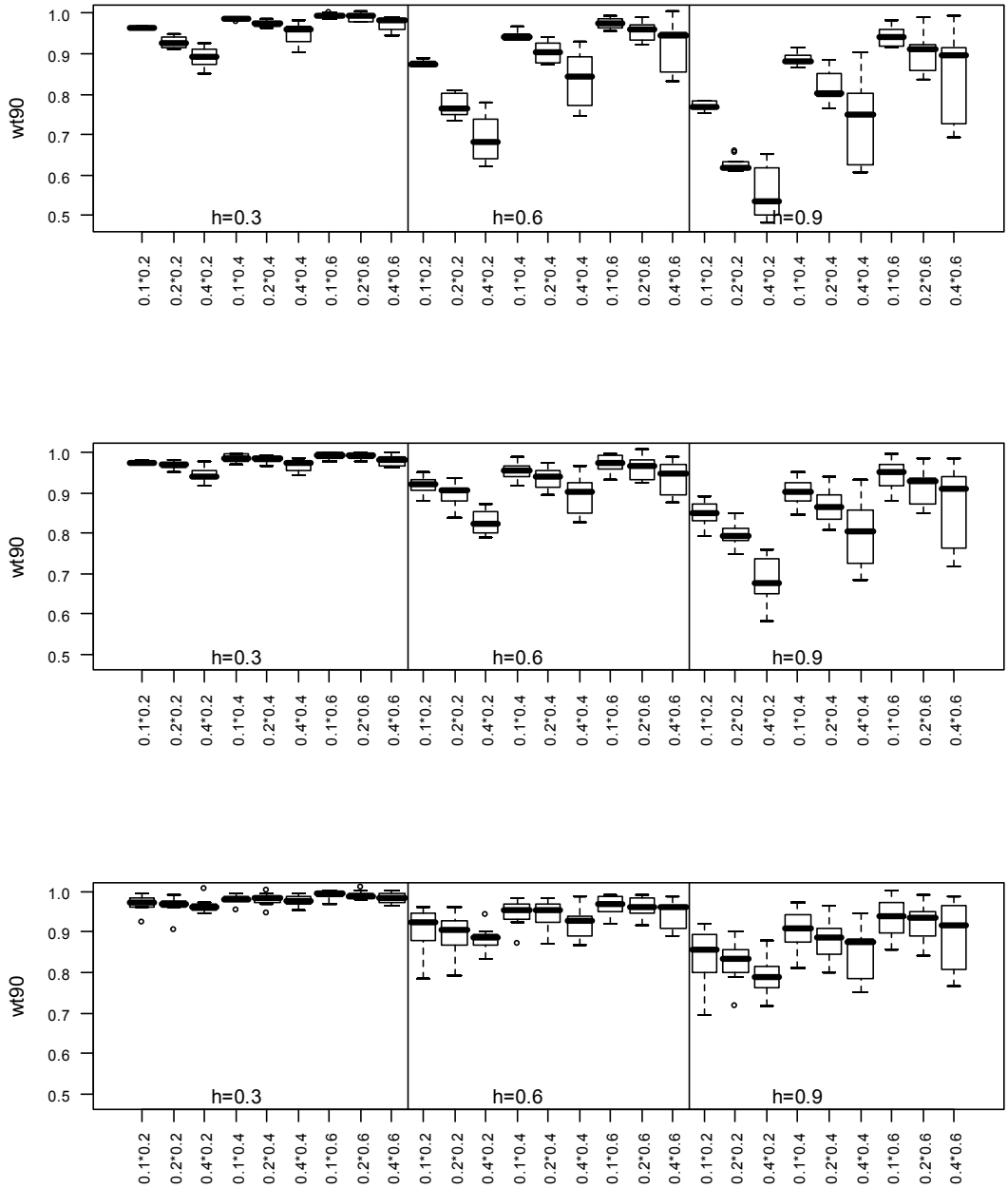


Figure A 6. Relative 90th percentile of weight at F_{msy} for a range of steepness (h) levels and combinations of m and k (shown on the x-axis as m*k). The panels are: flat selectivity (top), dome-shaped with 4 fully selected age classes (middle) and dome-shaped with 2 fully selected age classes (bottom).

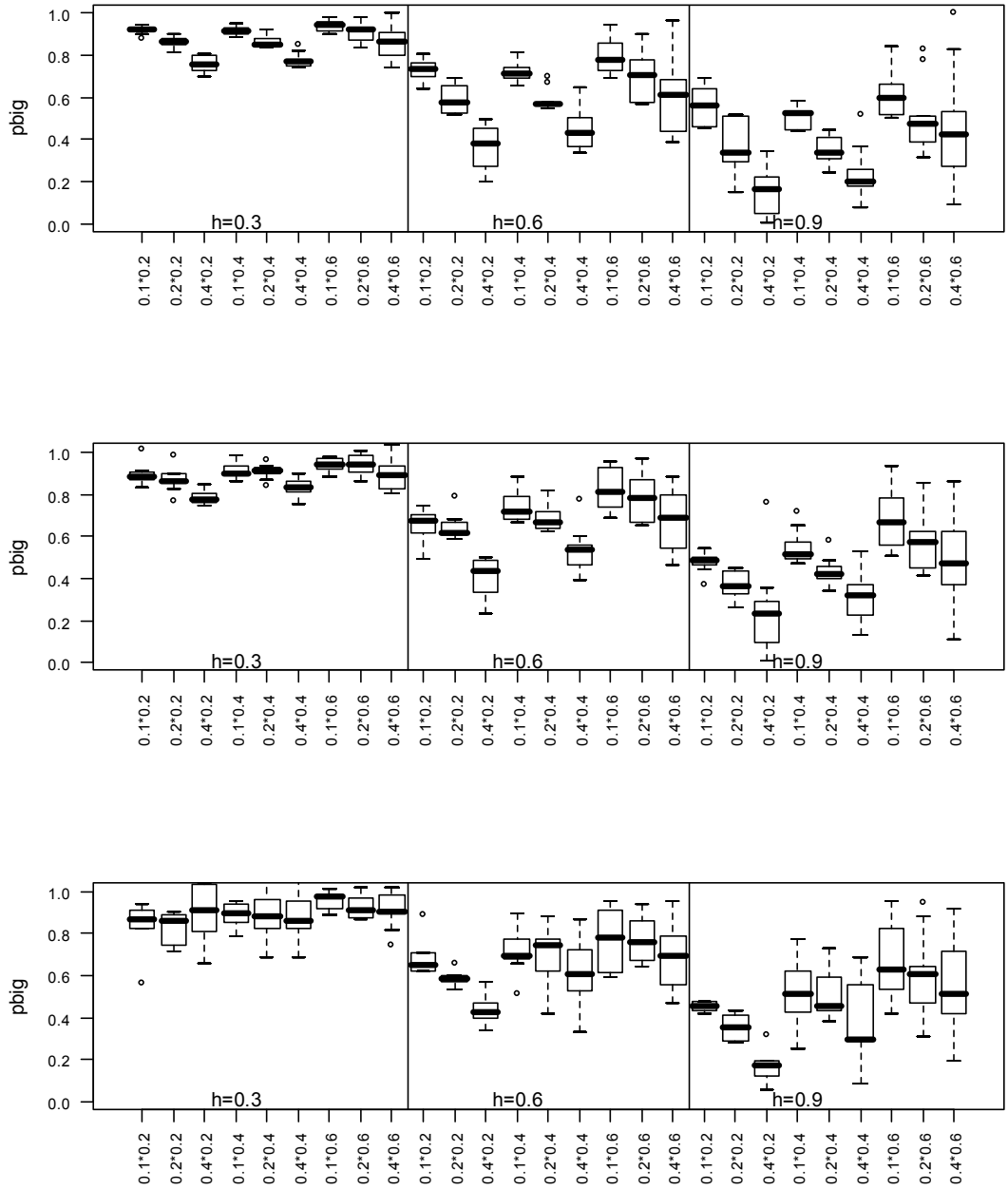


Figure A 7. Relative proportion big fish at F_{msy} for a range of steepness (h) levels and combinations of m and k (shown on the x-axis as m*k). The panels are: flat selectivity (top), dome-shaped with 4 fully selected age classes (middle) and dome-shaped with 2 fully selected age classes (bottom).

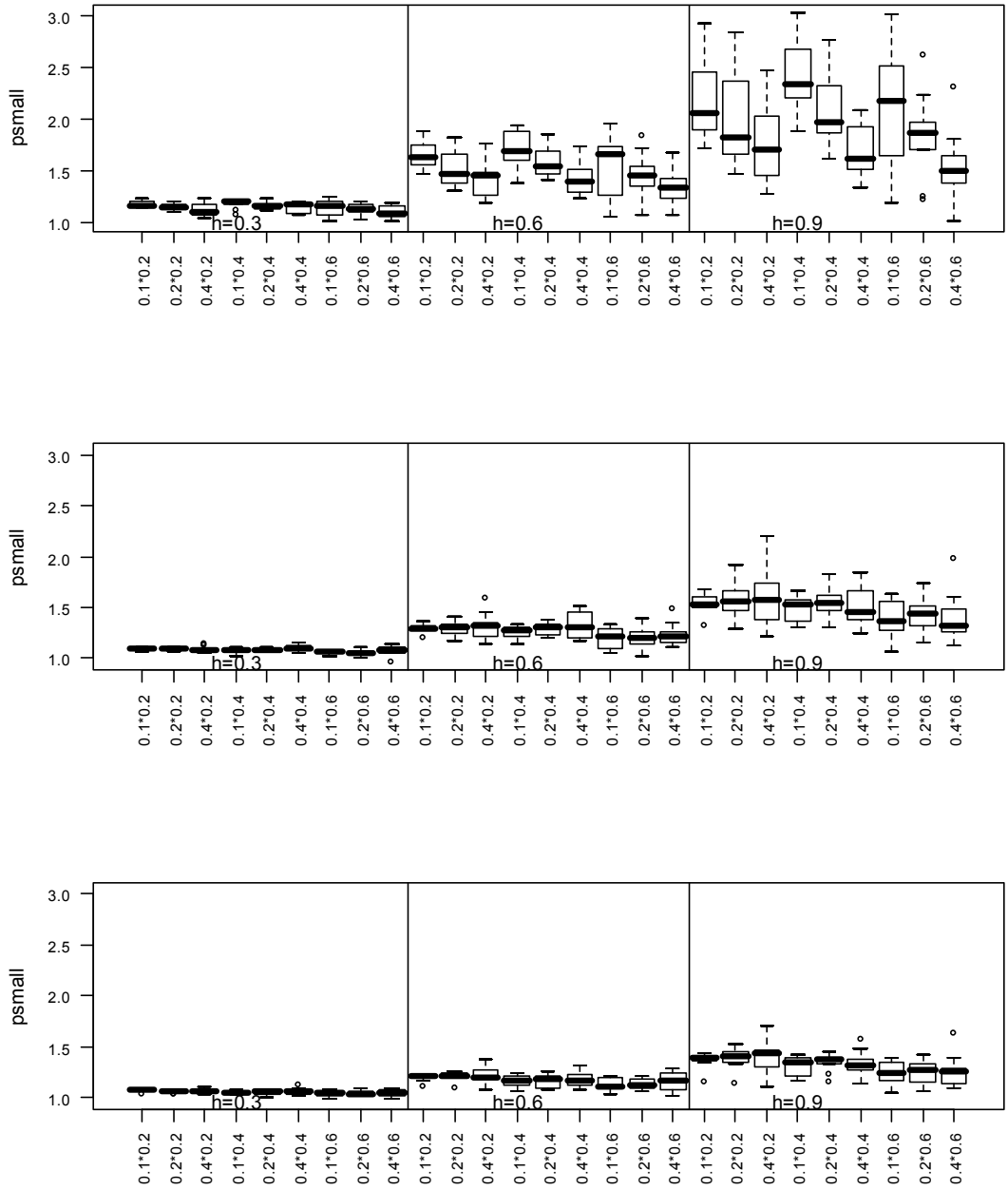


Figure A 8. Relative proportion small fish at F_{msy} for a range of steepness (h) levels and combinations of m and k (shown on the x-axis as m*k). The panels are: flat selectivity (top), dome-shaped with 4 fully selected age classes (middle) and dome-shaped with 2 fully selected age classes (bottom).

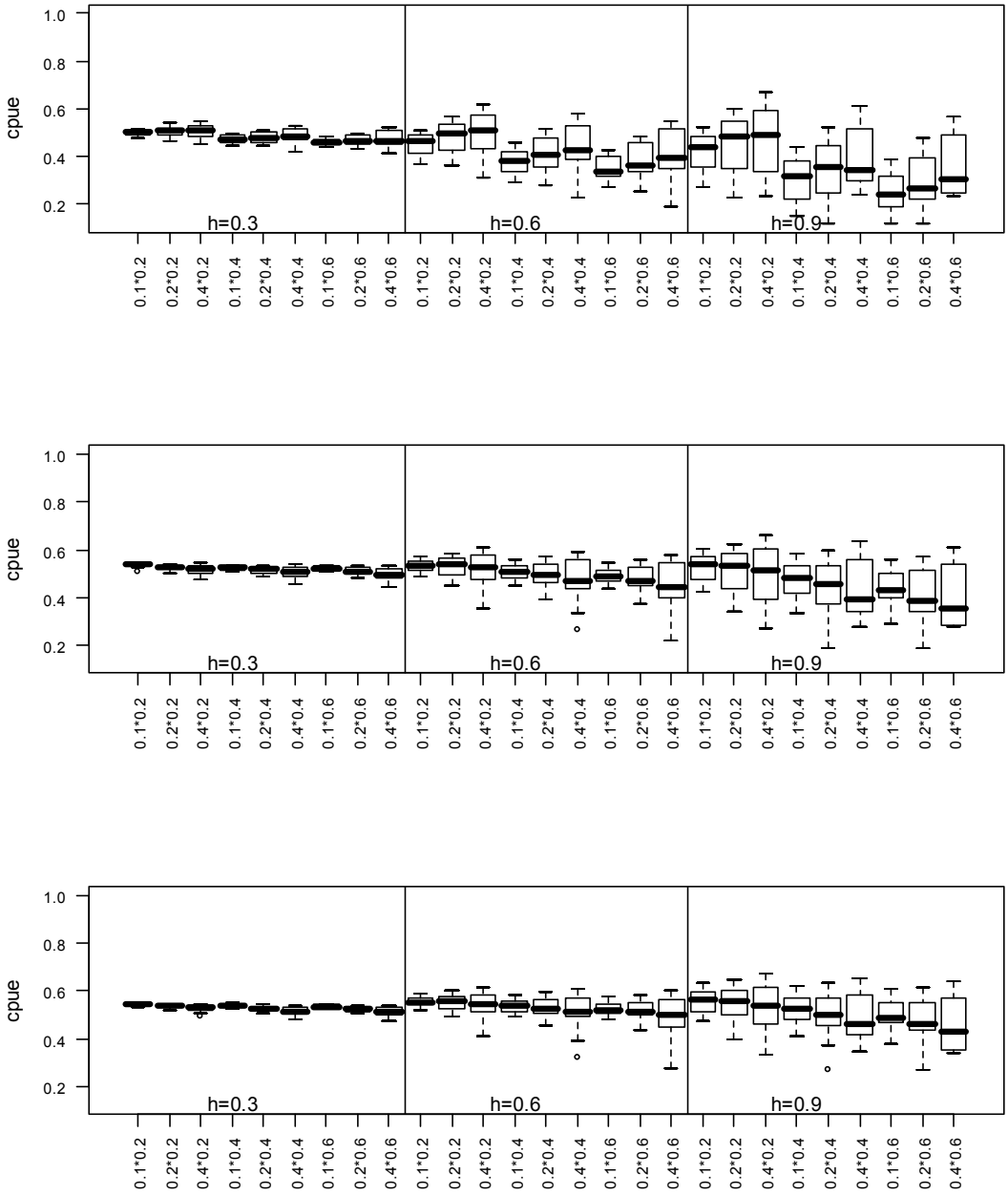


Figure A 9. Relative cpue in numbers at F_{msy} for a range of steepness (h) levels and combinations of m and k (shown on the x-axis as m*k). The panels are: flat selectivity (top), dome-shaped with 4 fully selected age classes (middle) and dome-shaped with 2 fully selected age classes (bottom).

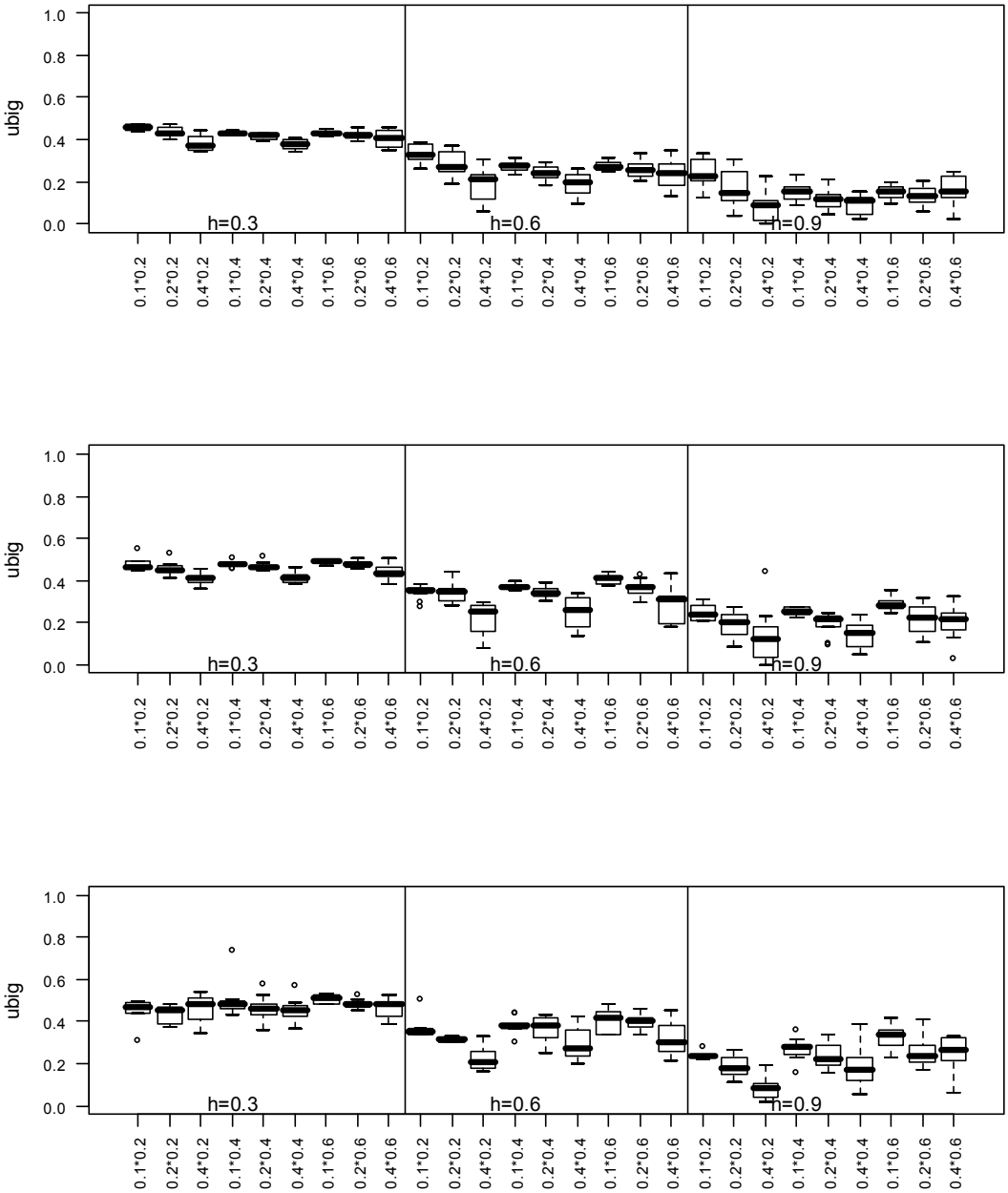


Figure A 10. Relative cpue of big fish at F_{msy} for a range of steepness (h) levels and combinations of m and k (shown on the x-axis as m*k). The panels are: flat selectivity (top), dome-shaped with 4 fully selected age classes (middle) and dome-shaped with 2 fully selected age classes (bottom).

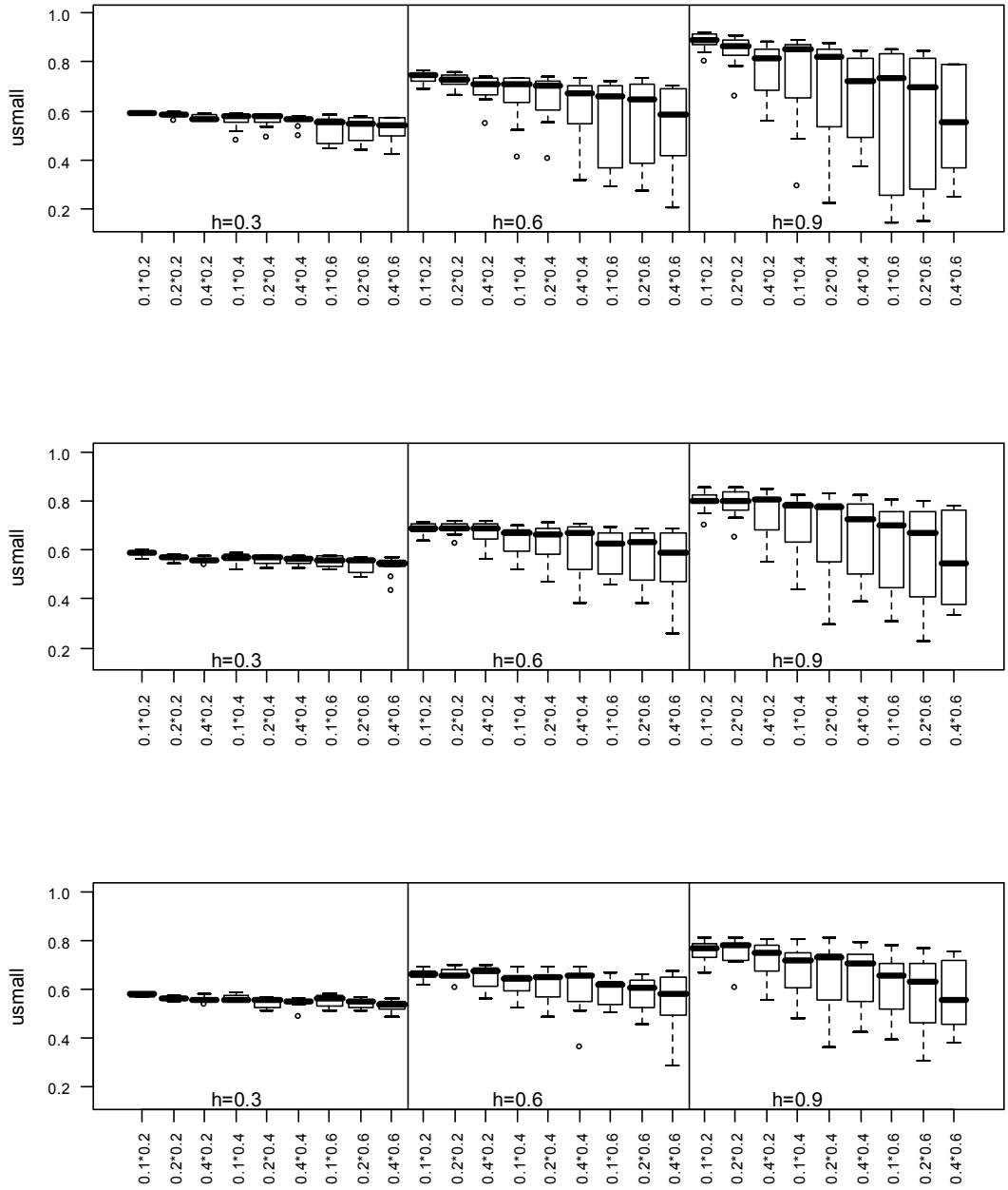


Figure A 11. Relative cpue of small fish at F_{msy} for a range of steepness (h) levels and combinations of m and k (shown on the x-axis as m*k). The panels are: flat selectivity (top), dome-shaped with 4 fully selected age classes (middle) and dome-shaped with 2 fully selected age classes (bottom).

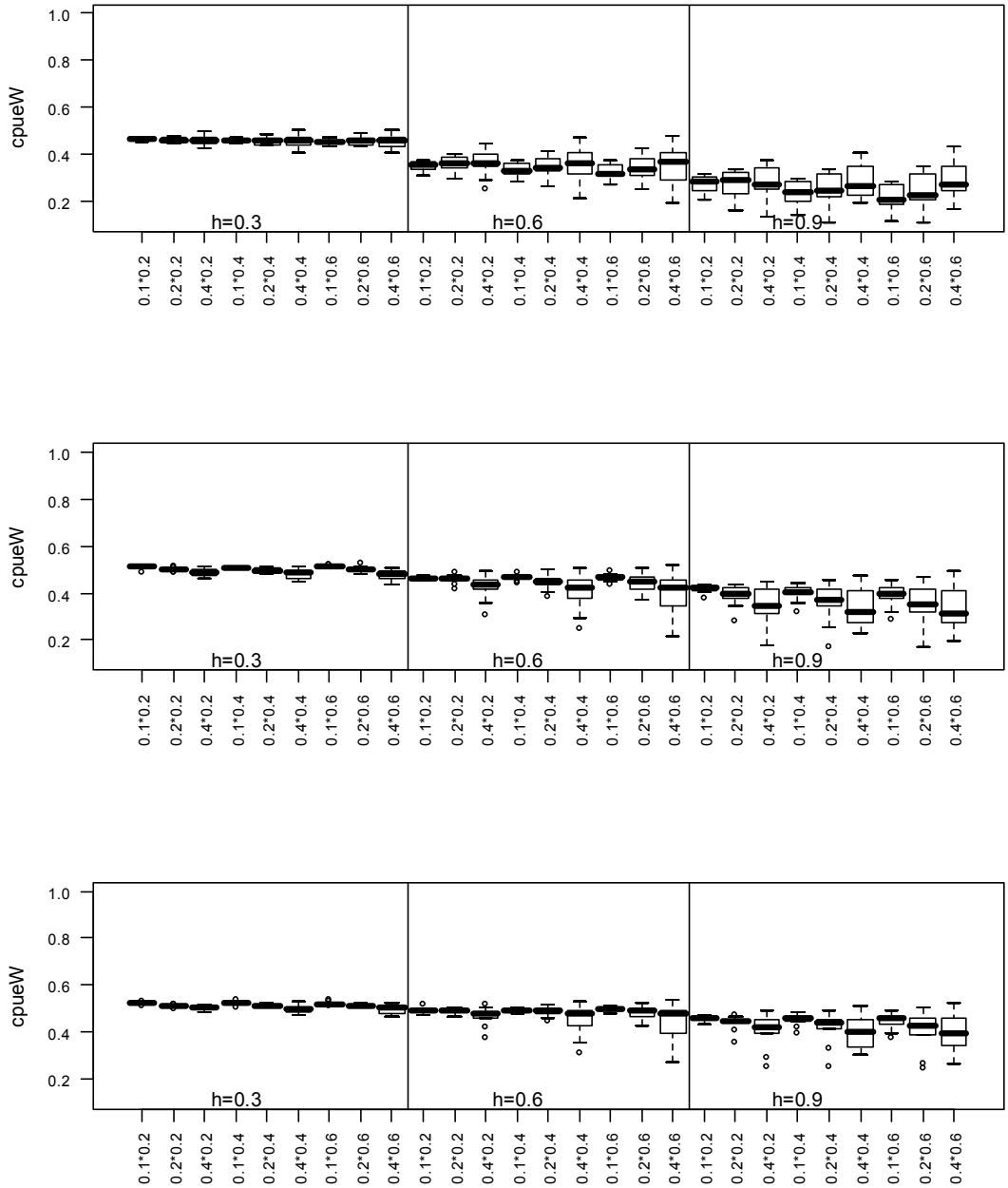


Figure A 12. Relative cpue in weight at F_{msy} for a range of steepness (h) levels and combinations of m and k (shown on the x-axis as m*k). The panels are: flat selectivity (top), dome-shaped with 4 fully selected age classes (middle) and dome-shaped with 2 fully selected age classes (bottom).

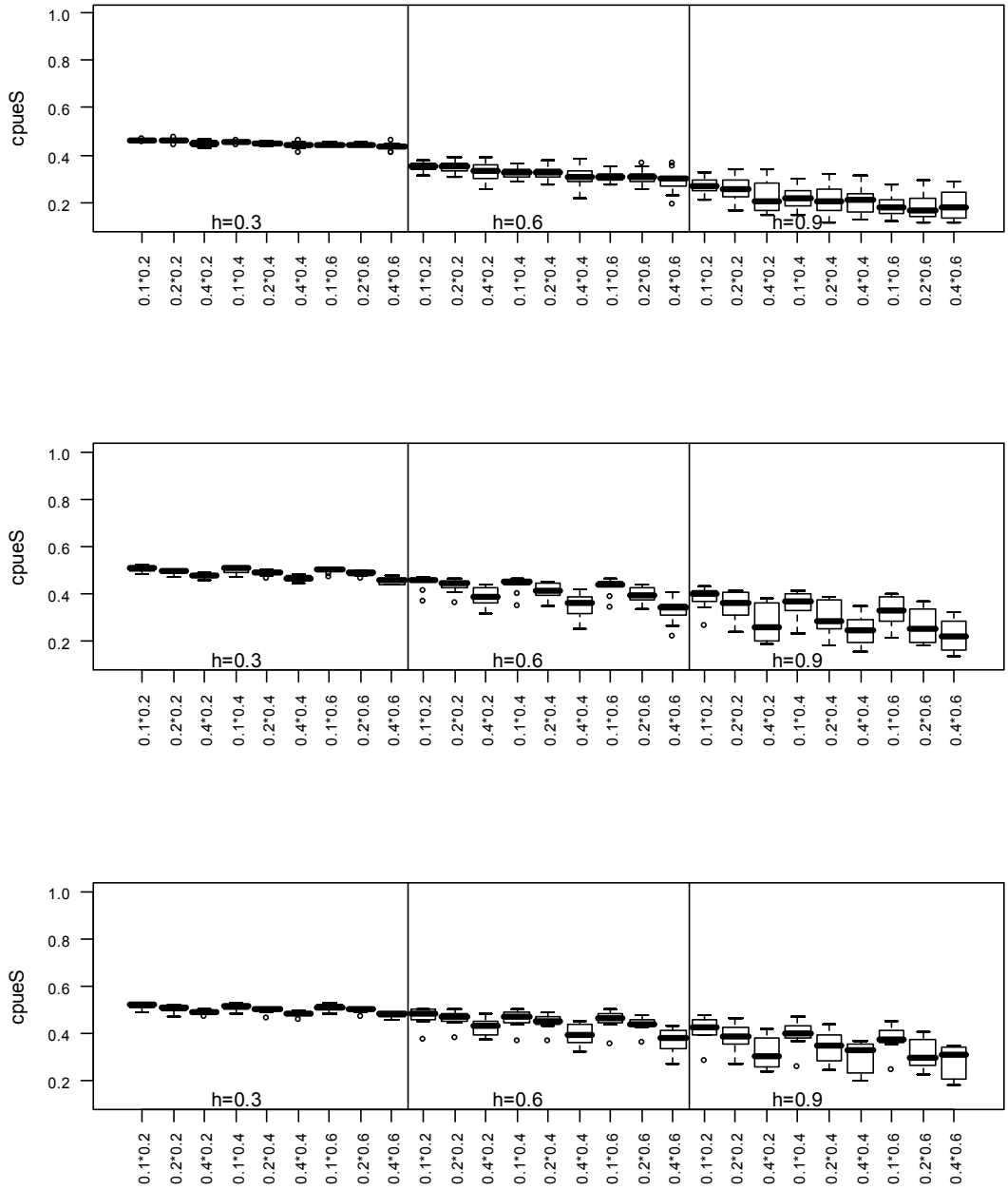


Figure A 13. Relative cpue of spawners (in weight) at F_{msy} for a range of steepness (h) levels and combinations of m and k (shown on the x-axis as m*k). The panels are: flat selectivity (top), dome-shaped with 4 fully selected age classes (middle) and dome-shaped with 2 fully selected age classes (bottom).

9. The performance of single indicators and groups of indicators as predictors of relative spawning biomass

Summary

The work in this Section addresses the third and fourth objectives of the project, namely to develop and test the robustness of single indicators and of groups of indicators, for potential use in decision rules.

Size-based indicators only perform well as predictors of whether spawning biomass is above or below its reference point, when steepness is high and when the von Bertalanffy growth rate is low. Although steepness is strongly related to the performance of these indicators via the level of harvest rate that is sustainable, the trials in this Section use a widely varying harvest rate over time, so it is not simply the extent of change expected at a constant harvest rate of, say, F_{msy} that is relevant to performance. It is likely that the level of recruitment and the associated population age/size structure at low compared to at high steepness also plays a role.

Of all the size-based indicators considered, the means (mean length or mean weight) generally perform best. In a few cases, the upper 90^{th} percentiles perform well. These indicators should, however, be used with caution because they perform poorly when the harvest rate is approximately constant.

Using indicators and triggers relative to their unexploited values, lead to lower prediction errors than using absolute values. Although there may be practical difficulties in defining the unexploited values, relative indicators should be considered in preference to absolute values.

Prediction errors from classification trees, which use multiple indicators, are only marginally lower than predictions from the best performing single indicator. At face value this suggests that there is little to gain from using multiple indicators, though we caution against over-interpreting this result. It is likely that the benefit of multiple indicators only really comes into play when series are biased or prone to serious an unpredictable 'failure' (e.g. changes in the fishery operation which is not detected when an indicator is interpreted or standardised).

Steepness in the stock-recruit relationship is an important factor in the evaluation of the performance of indicators. In reality, however, it is unlikely that steepness would be known. Results, particularly the predictions from classification trees, show that it is more risky to assume that steepness is high when it is in fact low, than to assume it is low when it is in fact high. The very low steepness values considered here (e.g. h=0.3 or h=0.4) are hopefully not common in tuna and billfish populations.

Results also suggest that, in the absence of knowledge about selectivity, it is safer to assume that selectivity is dome-shaped than flat, particularly when natural mortality is thought to be low. When natural mortality is high, there is much less difference between results for the two types of selectivity.

9.1. Introduction

In this Section we evaluate the performance of a wide range of size-based and CPUE-based indicators as predictors of the status of spawning biomass. In Section 8 we commented on the likely performance of indicators based on the expected level of change when a fish stock is harvested at a sustainable level (i.e. at a maximum of F_{msy}). Here we take this a step further by: (i) testing the performance of indicators under different scenarios when there is also measurement error affecting the indicators and when recruitment varies from year to year and (ii) testing the performance of groups of indicators for the same set of scenarios. This investigation does not yet involve testing of feedback decision rules, but it serves to inform how one might construct good decision rules. The work in this Section builds on that in Section 8 to address the third and fourth objectives (see Section 4. Objectives) of the project.

The performance of indicators, or groups of indicators, is evaluated by using them as 'predictors' of the status of spawning biomass and by calculating an associated 'prediction error'. We do this by turning a time-series of each indicator into a binary series of '0's and '1's depending on whether the indicator is below or above a specified trigger point[13]. The notion is that, when the trigger point is sensibly defined, the indicator should reflect whether spawning biomass (B) is above or below a given reference point. In Section 8 we showed that one can calculate the expected value of an indicator for any given harvest rate. We also showed that it is more useful to consider the relative expected value, i.e. the expected indicator value at a given harvest rate, divided by its expected value when there is no harvest. So, if we are interested in whether spawning biomass is above or below B_{msy}, we consider whether an indicator, say mean length, is above or below its expected value when the harvest rate is F_{msy}.

It is obvious that there are several reasons why an indicator may be wrong in its prediction. The indicator is likely to be measured with some error and some of the assumptions about parameter values that underlie our calculations of the trigger value may be wrong. Also, the trigger and reference point calculations are based on the implicit assumption of equilibrium, and this is unlikely to be met in reality, even simply because recruitment will vary from year to year.

Nonetheless, in the absence of a full stock assessment, this approach is still likely to be useful and the simulations should inform us just how useful the indicators can be when used in this way. We should add that we are not proposing that indicators should always be used in this binary way. When the time-series of an indicator is turned into a binary series, quite a lot of information is lost. A series of '0's and '1's do not reveal whether the indicator was just below (or above) the trigger point or well below (or above) it. This level of information is, however, likely to be more important in the context of a feedback decision rule where it is used to determine a response, such as a reduction in the catch. Here we are less interested in this subtlety and more interested in the overall patterns of performance between different indicators and groups of indicators.

There is another aspect of indicators that is not given much attention in this analysis (though we briefly look at one example), namely the autocorrelation in the time-series of indicator values. It is intuitive that if, say CPUE, is below its trigger point in a given year, and had been for the previous 3 years, one would be more concerned than if it is below its trigger point but had been fluctuating apparently without pattern prior to that. The strong autocorrelation in most of the

[13] We deliberately distinguish between 'trigger points' for indicators and 'reference points' for spawning biomass and harvest rate. We do this because we consider that when used in a decision rule, the trigger points for indicators may have to be adjusted to achieve the required objectives for spawning biomass and harvest rate, and may not be the values that coincide directly with the harvest rate implied by the SSB reference point.

indicators suggests that looking at trends or at multiple years, rather than a single year at a time, may be more informative.

As mentioned above, we compare the binary series of each indicator with the binary series of spawning biomass from the simulation to estimate how often the indicator is right and how often it is wrong in its prediction. This exercise can also be done by comparing the binary indicator with the binary harvest rate series (e.g. whether F is above or below F_{msy}). Note however, that this approach will only be meaningful if the harvest rate is really varying. In the context of a decision rule specified in terms of harvest rate, it will only be meaningful if there is implementation error in the simulations (i.e. the simulated 'actual' harvest rate is not the same as the harvest rate specified by the decision rule). We therefore prefer to focus on the prediction of spawning biomass.

The performance, or prediction error, of groups of indicators can in theory be evaluated in a similar way to that for single indicators. This exercise does, however, quickly become unwieldy because of the large number of potential combinations of different numbers of indicators. Instead, we use classification trees as a tool for evaluating whether (and if so, by how much) groups of indicators outperform single indicators, an also as a basis for constructing decision rules which are subsequently evaluated in feedback simulations (Section 10).

The set of scenarios considered here are not intended as exhaustive robustness testing of all combinations of indicators under all possible 'realities'. Instead, we focus on several scenarios, including two life-history strategies, namely long-lived (like swordfish) and short- lived (like yellowfin tuna or skipjack tuna). For each life-history strategy we consider two types of gear selectivity pattern: flat selectivity and dome-shaped selectivity. The details of the scenarios are discussed further below.

9.2. Methods

9.2.1 The simulation model

The performance of indicators is evaluated using a simulation model of the population dynamics of a harvested stock. Here we essentially use the same model as in Section 8. It is a discrete time, age-structured, single gender model with annual time-steps. A Beverton-Holt stock-recruit curve is used, defined in terms of the steepness parameter (h). Harvesting is modelled using a harvest rate, F, and a selectivity curve for a single fleet. There is no spatial structure in this version of the model and the assumption is a closed population in a single area. The equations are given in Appendix 1, but a verbal description is given here.

The life-history parameters (natural mortality, stock-recruit relationship, growth curve, length-weight relationship and maturity at age) are assumed constant over time for a particular scenario or realisation. Recruitment in each year is, however, a lognormal random variable with mean given by the stock-recruit curve and a specified variance, σ_R. The 'true' (simulated) quantity of main interest is the spawning biomass (B) in each year. (Note it is explicitly stated if we are referring to total biomass or fishable biomass).

The selectivity of longline fishing gear is commonly assumed to be flat-topped, increasing from 0 to 1 over some range of sizes, but then remaining at 1 for all larger sizes. Many assessments of pelagic fish stocks which estimate selectivity parameters, suggest that the selectivity of longline gear may be dome-shaped, with selectivity dropping down to 0 for very old/large fish (e.g. CCSBT 2003 and 2004; Davies *et al.* 2006). It is difficult to know whether there really are

mechanisms that would lead to dome-shaped selectivity of longline gear[14], or whether it is an artefact of the datasets and model structure. In our experience, both selectivity curves are often compatible with the data (depending on what other assumptions are made). We therefore consider this an important factor to investigate, and the simulated scenarios consider both types of selectivity with each of the life-history strategies. Although the main focus of this work is on longline fishing gear, a dome-shaped selectivity curve is very common for purse seine fishing. This reinforces the value and relevance of looking at both types of selectivity curve. What we have not attempted here is to consider a single stock being harvested by both a purse seine fishery and a longline fishery at the same time[15].

The indicators are generated from the simulated catch. The standard catch equation is used to first generate 'error free' catch numbers at age (see Appendix 1). This vector is then sampled with a Derichlet-multinomial distribution. The α parameter governs the variability of the 'with error' catch numbers at age, and this is further discussed below (sub-section 2.5). The total catch number is the same as the original, derived from the catch equation, but the proportions at age are different. It is these 'with error' catch numbers at age which are used to update the population dynamics, to calculate total catch weight and to generate size-based indicators.

Catch per unit effort (CPUE) is generated from the total catch numbers by first getting a nominal effort level from the harvest rate to catchability ratio[16] (F/q). The catchability coefficient, q, is mostly assumed to be constant over time, but we also consider scenarios with 'effort creep', i.e. where catchability increases over time. The observed CPUE is a lognormal random number with mean the total catch numbers over the nominal effort, and with a pre-specified variance which can differ between realisations. Note that the implicit assumption is that CPUE is proportional to fishable numbers in the population, although it is sampled or 'measured' with random error. We have not considered cases where CPUE is non-linearly, or only weakly related to abundance, or seriously biased. It is quite clear that such cases would lead to poor performance of CPUE indicators, if they are assumed to be linearly related to abundance. In practice it is usually very difficult, if not impossible, to determine whether a CPUE series is a reasonable indicator of abundance unless an additional independent measure of abundance, such as a survey, is available. In such a case one would almost always use the designed survey in preference to the CPUE.

The catch numbers at age (with error) are then sampled using a specified sampling proportion (e.g. 50% of the catch numbers), to give a total number, S, for sampling from the length-frequency distribution. We assume that length-at-age is a mixed normal distribution with means (i.e. mean length at each age) obtained from the von Bertalanffy growth curve, and variances specified. In this particular set of explorations, we have used a decreasing variance over age, based on limited size-at-age data from swordfish (see Appendix 1). Other assumptions about variance at age should not have a strong effect on the relative performance of indicators. The proportion for each component is simply the proportion of the sampled catch at that age.

The sampled length frequency consists of N_s lengths from which the size-based indicators (mean, median and 90th percentile of length) can be obtained. By converting each length to a weight via the length-weight relationship, the weight-based indicators can be obtained (mean, median and

[14] It could be that natural mortality increases for older fish when it is actually assumed to be constant, or there could be difference in spatial dynamics for animals of different ages.

[15] Note that the project was very strongly driven by the need for advice on swordfish which is not fished by purse seine fleets and only by longline fleets in the Indian (and Pacific) Oceans. Yellowfin and bigeye tuna are also primarily fished by longline fisheries in the EASTERN Indian Ocean, the area closest to the Australian Western Tuna and billfish fishery.

[16] This is derived from the approximate Catch equations: $C \simeq qEN$ (catchability x Effort x Population numbers) which is approximately equal to FN (harvest rate x Population numbers), implying $E \simeq F/q$

90th percentile of weight). Although this step could also be performed with error in the assumed length-weight relationship, this has not been done here. This primarily implies that the correlation between length and weight based indicators is stronger in these simulations that they might be in reality.

Finally, the proportion of the sample above or below a specified size can be calculated from the size frequency distribution to give the 'proportion big' and 'proportion small' indicators. Three additional CPUE indicators are calculated by taking the product of CPUE in numbers and another indicator, such as the proportion big fish, or the mean weight in the catch.

The list of indicators is as in Section 8, but repeated here for completeness (also see Appendix 1). The size-based indicators are:

Name	*Description*
mL, mdL, L90	- the mean, median, 90th percentile of length
mW, mdW, W90	- the mean, median, 90th percentile of weight
ps	- the proportion of "small" fish in the catch (the proportion of fish $\leq t_c+1$)
pb	- the proportion of "large" fish in the catch (the proportion of fish $\geq t_c+3$)

where t_c is the age-at-first capture when selectivity is knife-edged, or the youngest age for which selectivity is ≤ 0.5 when selectivity is flat-topped (but not knife-edge) or dome-shaped.

The CPUE-based indicators are:

Name	*Description*	
u	- CPUE in numbers	
us	- CPUE of small fish (numbers)	= u.ps
ub	- CPUE of big fish (numbers)	= u.pb
uw	- CPUE in weight	= u.mW
uSB	- CPUE of spawning fish (weight)	= $u.W_s.p_m$

(where W_s is the mean weight of fish larger than size at maturity in the size-frequency sample, and p_m is the proportion mature).

Although one could use the catch numbers at age and the detailed weight-at-age sample to obtain CPUE in weight, it is more realistic, with respect to the way data are usually collected, to use an estimate of CPUE in numbers and to multiply that by a mean weight of fish in the catch in that year.

The outputs, at each time step, consists of the population and catch quantities (fishable numbers and biomass, spawning biomass, recruitment, harvest rate, catch) and the value of each indicator. Details of how the time-series are generated and the additional assumptions involved are given in sub-section 2.3.

9.2.2 Trigger points

Given the outputs, the idea is to evaluate how well each indicator can 'predict' whether the spawning biomass, B, is above or below a specified reference point. For this, however, each indicator requires a 'trigger' point, chosen so that the comparison between indicators is fair. We chose to use two sets of reference points as 'triggers', namely those coinciding with harvesting at F_{msy} and those coinciding with harvesting at an F which would imply a 40% spawner per recruit level (i.e. SPR at 40% of unexploited SPR; we refer to this as F_{spr40}). The coinciding reference points for B are: B_{msy} , or B_{msy}/B_0 in relative terms, and $B=0.4B_0$. Note that the constant harvest rate which would imply $B/B_0=0.4$ is not necessarily the same as F_{spr40} because of the absence of a

stock-recruit relationship in the SPR calculations. In practice is similar for high and moderately high steepness, but the harvest rate would be lower than F_{spr40} for relatively low steepness.

The choice of 40% is arbitrary and F_{msy} can also, obviously, be replaced by a lower value, for example $0.8F_{msy}$. The exact choices are not crucial to this analysis, because they are not proposed as management quantities or objectives. The approach serves to illustrate important points and to evaluate the performance of indicators relative to one another. In the text that follows, we refer to these quantities as 'MSY' and 'SPR' trigger (or reference) points respectively.

It is worth emphasising that the MSY reference and trigger points require an assumption about the steepness (h) of the stock-recruit relationship. The SPR reference and trigger points do not require any assumption about the stock-recruit relationship; SPR calculations are similar to 'yield per recruit' calculations. The inclusion of the SPR approach is particularly relevant since steepness will mostly not be known. Even in cases where there is a full stock assessment, steepness can often not be estimated, or is estimated with very wide confidence intervals.

A second point worth noting is that the size-based indicators are independent of the actual level of recruitment, so the SPR calculations can provide an absolute estimate of the expected mean length (or weight) in the catch when harvesting at F_{spr40}. We nonetheless mostly work with the size-based indicators relative to their unexploited levels. These are, in fact, less sensitive to wrong assumptions about the life-history parameter values.

CPUE-based indicators are dependent on the actual level of recruitment, so the SPR calculations are used to provide relative CPUE; i.e. CPUE at F_{spr40} relative to CPUE of the unexploited stock. Even for the MSY reference points, it is more sensible to consider relative CPUE because in many cases catchability and absolute population size may not be known. (It goes without saying that in a real situation it would then be necessary to infer, from historic data, what the CPUE or size-based indicators of the unexploited stock might have been.)

9.2.3 Calculating Prediction Errors

The first step in the processing of the simulated outputs ('true' spawning stock biomass, B, and the set of indicators, e.g. CPUE, mean length, mean weight etc. at each time step) is to calculate the prediction error of each (single) indicator for different trigger points and a spawning biomass reference point.

To explain this, consider MSY trigger and reference points. We determine the ratio B_{msy}/B_0 and use this as the spawning biomass reference point. We also determine the expected values of each of the indicators if the stock is harvested at F_{msy} (over a long period, until the stock stabilises). For example, say the B_{msy}/B_0 ratio is 0.2 and the mean length when harvesting at F_{msy} is expected to be 250cm. The spawning biomass outputs from the simulation are converted to relative values (i.e. $relB = B/B_0$) and a binary vector (of 0's and 1's) is constructed to reflect whether B is above or below its reference point:

relB > 0.2 --> b=1
relB ≤ 0.2 --> b=0

where 'b' is the binary value of relative spawning biomass. Similarly, the mean length vector is turned into a binary vector with i=1 if mean length > 250 and i=0 if the mean length is <=250. Here 'i' is the binary value of the indicator. The two binary vectors are then compared, and a simple table constructed to show how many times both were equal to 1 or both equal to 0 (i.e. the correctly predicted outcomes), and how many times they were different ('01' or '10' combinations). The number of '01' and '10' combinations over the total number of observations is the prediction error for mean length as an indicator.

Although this explanation was based on the absolute mean length relative to its absolute trigger (250cm in the example above), we mostly use the mean length relative to its value in the first time-step, and compare that with its relative trigger point. Recall that this is ratio of the expected mean length when F=0 to the expected mean length when the stock is harvested at F_{msy}. The indicator values for the so-called unexploited stock still relate to the fishable age classes with the associated selectivities and not to the stock as a whole. It is therefore a somewhat theoretical construct, though it can be thought of as the value likely to be obtained in the first year of a harvest on a previously unexploited stock.

It is informative to separately consider the two types of errors that make up the prediction error of an indicator:
Type I: Biomass is below its reference (b=0), but the indicator suggests it is above (i=1)
Type II: Biomass is above its reference (b=1), but the indicator suggests it is not (i=0)

The first error has potentially more serious implications for conservation, if no action is taken. The second error has potentially more serious implications for the catch, if action is taken and the catch or effort is reduced. Also note that the overall prediction error can be written as:

Pred.Err = P(i=1 | b=0). P(b=0) + P(i=0 | b=1).P(b=1)
where
P(i=1 | b=0) is the Type I error
P(i=0 | b=1) is the type II error

The overall prediction error is therefore a weighted sum of the type I and type II errors, where the weights are the probabilities that b=0 or b=1. These probabilities, of whether the spawning biomass is above or below its reference point, are obviously strongly linked to the harvest rate. If the harvest rate is high, it is likely that P(b=1) = 0, or close to 0, and the overall prediction error would be identical, or very similar, to the type I error. Alternatively, if the harvest rate is low, and P(b=0) = 0, the overall prediction error would be identical to the type II error. As will become clear from results, the overall prediction error of indicators depend on the harvest rate. This issue is discussed further below.

9.2.4 Classification trees

Although the above approach can be expanded to combinations of indicators, it very quickly becomes unmanageable. Even with pairs of indicators, there would be 78 (unique) pairs from 13 individual indicators. A pair of indicators (i_1 and i_2) taken with the relative B (b) imply three binary vectors which, for each data point (or simulated year), can take on 2^3=8 possible values, i.e. ($b.i_1.i_2$ = 000, 001, 010, ... , 111). Although it is obvious that the combinations '100' and '011' are 'wrong' predictions, an implicit interpretation is required when the two indicators are different. With two indicators it may seem obvious to count predictions where either i_1 or i_2 is different from b as 'wrong', but when the group of indicators is larger, it becomes less obvious because the question of how they will ultimately be used in a decision rule becomes relevant[17].

Instead of pursuing this approach for combinations of indicators, we have used classification trees to model the binary spawning biomass 'data', b, as a function of the full set of indicators (i_1,

[17] For example, if only 1 out of 5 indicators are predicting b incorrectly (as 0, say), should this count as 'wrong'? Arguably, it should count as 'wrong' if the set of indicators are used in a traffic-light type of decision rule based on 'if any indicator is 0, then predicted b=0'. It should not count if the decision rule is based on 'if any three indicators are 0 then predicted b=0'.

i_2, ... i_{13}; see sub-section 2.1 above). General information on classification trees can be found in, for example, Venables and Ripley (2002) or Chambers and Hastie (1993). In brief, however, a classification tree is a decision structure which takes inputs, and follows a logic to determine an outcome, or in this case a classification, of whether spawning biomass is above or below its reference point. The concept will become clear in the results; a glance at Figure 58 will show an example of a classification tree.

It is important to note that the idea is not to find the optimal tree, but rather to use classification trees as a tool to:
(a) identify the most relevant or important indicators for predicting whether B is above or below its reference point
(b) evaluate whether a set of indicators used together perform better than the best performing single indicator
(c) provide insight to and possible guidance for designing decision rules

Since the decision rules will ultimately be used to modify the catch or harvest rate in a feedback simulation model, and since the performance of the indicators/decision rule will be tested in that context, it is clear that fitting trees to the data is an intermediate step. Also recall that the performance of a decision rule is not only measured in terms of the performance of spawning biomass, but also in terms of catch and the variability of catch. The classification trees only consider performance in terms of biomass.

Trees were fitted using the 'rpart' library in the statistical software package R (www.R-project.org). The 'rpart' function, which fits the classification tree is used with the following call for cases where MSY triggers were used to generate the binary data:

```
rpart( Bmsy ~ mL+mdL+L90+mW+mdW+W90+pb+ps+u+ub+us+uW+uSB,cp=1e-4)
```

where 'Bmsy', the response variable, is the binary version of spawning biomass relative to B_{msy} and the explanatory variables (to the right of the ~) are the binary versions of the indicators (see sub-section 2.1 above for descriptions of indicator labels). A similar call is used when binary data are based on SPR triggers, with the response variable referred to as 'Bspr'. Since the data are binary, the method will fit a classification tree with the Anova method.

The parameter 'cp' is a control parameter, called the complexity parameter. With 'anova' splitting, for example, this means that the overall R^2 value must increase by 'cp' at each step. Any split that does not achieve this is not attempted. The main role of this parameter is to save computing time by pruning off splits that are obviously not worthwhile. The general approach is to first fit a relatively large tree by specifying a low value for the 'cp' parameter. These 'full' trees are used to evaluate the relative errors as the size of the tree increases. As will become obvious, some of these classification trees can be very large and therefore difficult to interpret or implement as a decision rule. The difficulties may lie in the availability of data if, say, most of the indicators in the set are included, but more importantly, the trees are unlikely to be understandable or intuitive to stakeholders.

We therefore also generate 'pruned' trees, i.e. smaller trees with fewer indicators and fewer branches. The approach to pruning trees is primarily based more on pragmatism than statistical rigour. We consider that a tree with more than about 5 "splits" (or decision points) would be hard to understand. As will be shown, however, this pragmatic choice of size is often very similar to the size that one might choose on the basis of the relative error as a function of the number of splits in the tree. This is illustrated and further discussed in sub-section 4 below.

A tree can then be used with the dataset on which it was based to obtain a prediction error (as before, the number of times the tree predicted b=1 when the data showed b=0 and *vice versa*). In R this is simply done with the associated 'predict' function for trees. The tree can also, however, be applied to a different dataset, based on the same or on different assumptions, and a prediction error obtained. This is used in sub-section 4 to explore the likely implications of wrong assumptions.

9.2.5 Scenarios and running simulations

A scenario is the combination of 4 components: a life-history, a steepness level (in the stock-recruit curve), a selectivity curve and measurement errors associated with the catch and hence the indicators. We first outline the options under each of these components, and then explain how each scenario is run. It is worth mentioning now that each scenario is run for 100 realisations. Some quantities are the same for each realisation, but others vary randomly, as will become clear under the discussions of the components.

(1) Life-history parameters

For the first component, there are two choices of a life-history, namely a 'Low m' (low natural mortality, slow growth rate and late maturity) and a 'High m' (higher natural mortality, faster growth rate and early maturity) strategy. These two life histories have been shown (Section 8) to be near the extremes of expected changes in the size-based indicators for tuna and billfish-like species. This range also includes other teleosts and elasmobranchs, but not the extremely short-lived life-histories like squid, or the very long-lived life histories like orange roughy. Nonetheless, the conclusions drawn from the two life-histories considered here can be extended to more extreme cases. Table 1 shows the mean input parameters for each life history. For each realisation within a life-history scenario, the four parameters (m, k, t_m, t_c) are varied in the following way.

Natural mortality and growth rate. Values of m and k are generated so that log(m) and log(k) are normally distributed with a variance of 0.02 and a correlation of 0.7. The log-means were simply the logs of the input values in Table 1 for that life-history. The implications of this are most easily seen in Figure 26 and Figure 27. (See sub-section 2.4 in Section 8 for more about correlated life-history parameters).

Age at maturity. Values of t_m are determined in such a way that the expected proportion alive at age t_m (given the value of m from step 1) is the same (or similar, given the need to round the result to an integer age) as for the specified input values for that life-history (Table 7). Maturity-at-age is knife-edge at age t_m

Age at first capture. Values of t_c are determined so that the length at first capture is the same in each realisation. In other words, the length at first capture (L_c) is determined from the specified input values (Table 1) for that life-history. Using the value of k from step 1, the age at first capture coinciding with L_c is then found (rounded to the nearest integer) and used as t_c in the specific realisation.

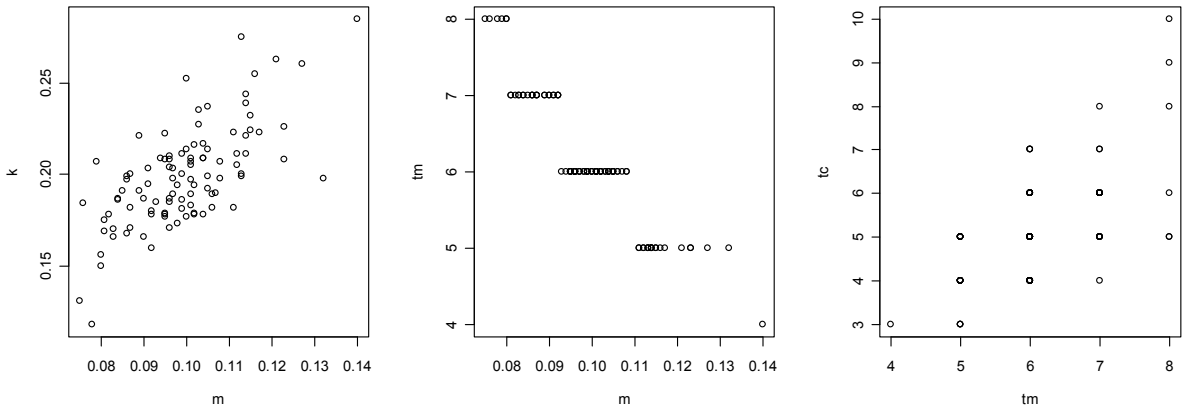


Figure 26. Life-history with low m. Combinations of m, k, t_m and t_c for 100 realisations. Values were generated according to the description in the text (above).

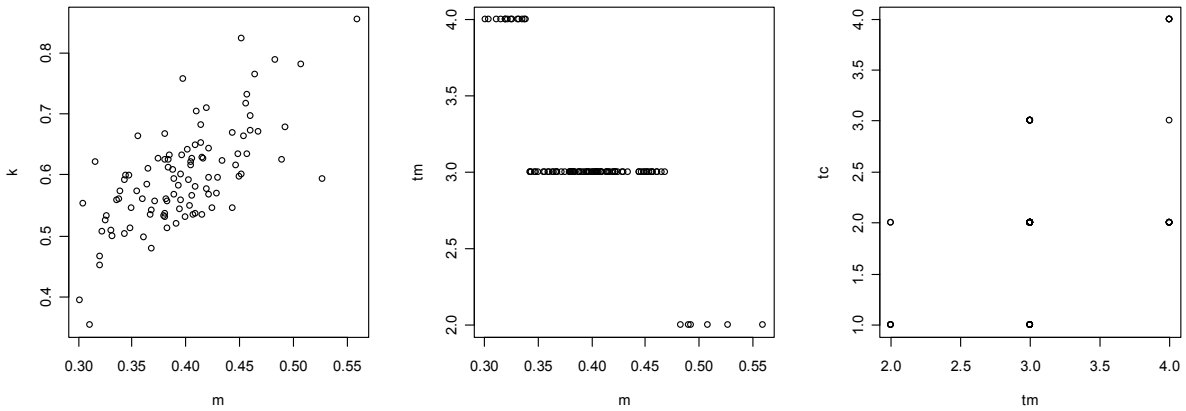


Figure 27. Life-history with high m. Combinations of m, k, t_m and t_c for 100 realisations. Values were generated according to the description in the text (above).

We note that this exercise of using slight variations in the input parameters, were simply done to introduce some variability into the simulated 'true' dynamics, but within relatively small ranges around the specified input values. The specified input values in Table 1 are subsequently used to calculate the trigger and reference points. This approach therefore implies that results will reflect, to some extent, the effect of slight mismatches between the 'true' life history, and that assumed for the purposes of calculating reference/trigger points.

(2) Steepness

Given the strong role that steepness plays in the dynamics of a stock and the level of harvest rate that is sustainable, as well as the fact that steepness is likely to be unknown in reality, we have separately considered four levels of steepness. The levels range from high (around 0.9) to very low (around 0.4). A steepness value for each realisation within a scenario is drawn from a uniform distribution as specified by the 'Range' in Table 1. For example, if the steepness component of a scenario is 'high' (h=0.9), each realisation would have a steepness value drawn from the uniform distribution U[0.85,0.95]. The 'very low' steepness is hopefully rare. However, if we assume that steepness is higher (or much higher) and it really is this low, the situation could be very risky indeed. From the point of view of exploring robustness, we therefore consider it

necessary to include this very low steepness scenario. For practical reasons, we have sometimes only considered the extremes (high and very low).

Recruitment variability is not considered as a separate component, but we have considered log-normal recruitment with a CV taken from the uniform distribution U[0.2, 0.4]. It is arguable that recruitment can have much higher CVs (e.g. even > 0.6). We consider that these higher recruitment variabilities are more relevant in the case of testing decision rules within a feedback system, and particularly when tuning decision rules to achieve certain objectives or minimise certain risks, than here where we are primarily interested in the relative performance of indicators.

(3) Selectivity

As noted above, there are two choices for the selectivity curve, flat or dome-shaped. The age at 50% selectivity (t_c; see above) specifies where the curve is placed, so when t_c varies between realisations, the same curve is simply shifted to the left or right. Examples are shown in Figure 28.

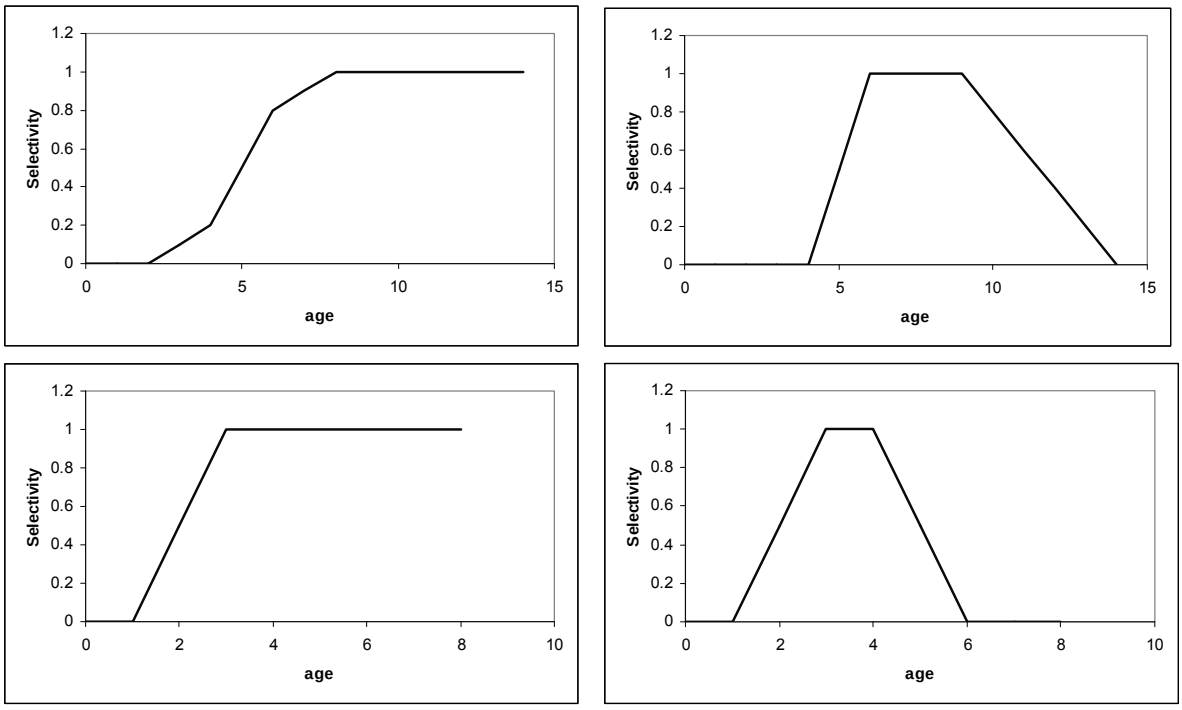


Figure 28. Examples of selectivity curves for the long-lived (low m) life history in the top two panels and for the short-lived (high m) life-history in the bottom panels. The flat-topped selectivity curves are on the left, and the dome-shaped selectivity curves are on the right.

(4) Measurement errors

The fourth component is the level of noise or measurement error in the indicators. The procedure for generating catch and indicators based on catch is described above in the 'Simulation model' sub-section. There are three sources of error:
1. error in the proportions of catch-at-age [multinomial Derichlet distribution governed by α]
2. the CV of CPUE [log-normal distribution governed by σ_{cpue}] and
3. the sample-size for length and weight frequency distributions [sample a proportion 'p' of the total catch numbers]

For each source of error we have defined a high (H) and a low (L) range of errors and denote a specific scenario by the three letter 'code' in the order given above (e.g. HHH implies all 3 sources have high noise levels; LHL implies low noise on catch-at-age and size frequencies, but high error on CPUE. It is important to note that the definitions and 'L' or 'H' codes refer to the resulting errors, not (necessarily) to the values of the input parameters[18]. For each realisation, a parameter values for a noise component is drawn uniformly from the specified range of parameter values given in part D of Table 1. For example, if the CV of CPUE is high (H), then a value for this component is drawn from the uniform distribution [0.4, 0.5].

The α-parameter of the Derichlet multinomial distribution is particularly difficult to interpret. The implications of different values of this parameter for the CV of catch-at-age are therefore summarised in

Table 8 to give some indication of the extent and pattern of variability. Note that the variability is always higher for age classes containing a small proportion of the population. Also, the total numbers that are being sampled, is used to scale up the α-parameter. The examples given below are for relatively large numbers, around 4000 individuals.

Table 7. Summary of parameter values for components of each scenario. The naming conventions used to describe a scenario are given in italics (e.g. High m, high h, flat, HHH, etc.)

A. Two life history strategies

	High m	*Low m*
natural mortality	0.4	0.1
vonBertalanffy k	0.6	0.2
maturity age, t_m	3	6
catch age 50%, t_c	2	6

B. Steepness, h, in the Beverton-Holt Stock recruit relationship

Steepness	*high, h=0.9*	*low, h=0.7*	*lower, h=0.5*	*very low, h=0.4*
Range	(0.85,0.95)	(0.65,0.75)	(0.45,0.55)	(0.35,0.45)
Mean	0.9	0.7	0.5	0.4

C. Selectivity (see Figure 28)

Flat

Dome-shaped

D. Noise in indicators

	L (low)	*H* (high)
1. error in catch-at-age (α parameter)	[1.0, 2.0]	[0.05, 0.15]
2. CV of CPUE in numbers	[0.1, 0.2]	[0.4, 0.5]
3. Size frequency sampling proportion	[0.5, 0.7]	[0.05, 0.15]

As indicated above, a scenario consists of a choice of life-history (2 options), level of steepness (4 options), selectivity curve (2 options) and measurement error (8 options). For each scenario, we ran 100 realisations of 90 time steps each. Recall that within a scenario, the quantities that vary between realisations are:

m, k, t_m, t_c (within the narrower ranges for each life-history option)
h (within the narrower range for the steepness option)
placement of selectivity curve to match t_c (shape is defined by the selectivity option)
noise parameters (within the narrower ranges of high or low noise level)

[18] For the first and third sources of errors, low values of the input parameters lead to HIGH errors and vice versa. It is only for the second source (CV of CPUE) that a high value of the input parameter implies HIGH errors.

Table 8. CVs of proportion at age implied by the α-parameter for the Derichlet multinomial distribution for flat and dome-shaped selectivity. CVs are given for the first age class, and an upper age class which approximately covers the main part of the catch-at-age distribution. At older age classes the CVs are even higher than those shown against 'upper class'. The minimum CV (for the age class with highest proportion in the catch) is also shown.

	Flat selectivity			Dome-shaped selectivity		
α parameter	first age class	minimum	upper age class	first age class	minimum	upper age class
2.00	0.13	0.06	0.16	0.06	0.05	0.19
1.00	0.20	0.09	0.24	0.09	0.07	0.27
0.15	0.53	0.23	0.58	0.22	0.19	0.66
0.05	0.93	0.40	1.04	0.37	0.30	1.23

The last and important quantity that needs to be specified is the harvest rate. There are two quite different arguments that can be put forward for specifying F. The first argument is as follows. Since we are testing whether indicators can correctly 'predict' the status of spawning biomass, it is important to test this in situations where the stock is well below, or well above its reference point. In these situations there should be a strong signal. If an indicator does not perform well under these circumstances then it cannot be relied upon; it will be a risky indicator. Taking this argument, it would be informative to run the simulations with a harvest rate that varies greatly over time and allows the stock to be very depleted, as well as lightly fished over a relatively long time-period.

The second argument is based on the idea that indicators are being tested in order to use them in decision rules. The decision rules would tend to be constructed and used in such a way that the harvest rate (or catch) is unlikely to vary greatly, and certainly not to become too large over an extended period of time. Also, apart from possibly at the start of implementation, the whole idea is to keep the stock at or close to a sustainably harvested level. Under this argument, it would make more sense to test indicators using a harvest rate that is close to some constant (or varying slightly around a constant) level.

Both arguments have merit, but we choose to concentrate primarily on the first argument to explore whether an indicator is likely to work at the extremes of very high and very low spawning biomass. We use a sine wave to generate a cycle of increasing and decreasing harvest rates over a period of 90 years:

$$F_t = F(1.1 + \sin(t/10))$$

where t is time (year) and F is an input harvest rate. The amplitude (value of 10 which scales time) is chosen so that depletion drops to low levels (e.g. B/B_0 <20%), but also rises to levels well above 50% of B_0. The first term (1.1) prevents F_t from dropping to exactly zero and the mean of F_t is also therefore slightly above the input value F. The input F values are specified as $1.2F_{msy}$ for each scenario[19], bringing some comparability to scenarios. Figure 29 illustrates the time-series of harvest rates for an input value of 0.3 and 0.1. Note that the 90-year time-period allows for two occurrences of high and one occurrence of low harvest rate, but also recall that

[19] F_{msy} is calculated from the average or input values for each scenario (Table 3), rather than the values in each realisation.

the simulated population starts in an unexploited state. Table 3 summarises the input F-values and their implications in terms of the mean, minimum and maximum values of F_t.

We also briefly compare a scenario where Ft is drawn from a log-normal distribution with mean 1.2Fmsy and a CV of 20%, for illustrative purposes. We return to this issue in Section 10 where the performance of indicators are evaluated within feedback decision rules.

Table 9 Summary of harvest rate values input to the sine function which generates the time-series of harvest rates in scenarios, and implications in terms of the mean, minimum and maximum harvest rate for each scenario.

		Flat selectivity			Dome-shaped selectivity		
	h	entered value	mean F	min,max F	entered value	mean F	min,max F
Low m	0.9	0.30	0.40	0.03, 0.63	0.52	0.68	0.05, 1.09
	0.7	0.17	0.23	0.02, 0.36	0.31	0.41	0.03, 0.65
	0.5	0.10	0.13	0.01, 0.20	0.18	0.24	0.02, 0.38
	0.4	0.065	0.09	0.006, 0.14	0.125	0.16	0.01, 0.26
High m	0.9	1.38	1.80	0.14, 2.90	1.50	1.97	0.15, 3.15
	0.7	0.70	0.92	0.07, 1.47	0.83	1.09	0.08, 1.74
	0.5	0.37	0.49	0.04, 0.78	0.48	0.63	0.05, 1.01
	0.4	0.25	0.33	0.03, 0.52	0.32	0.42	0.03, 0.67

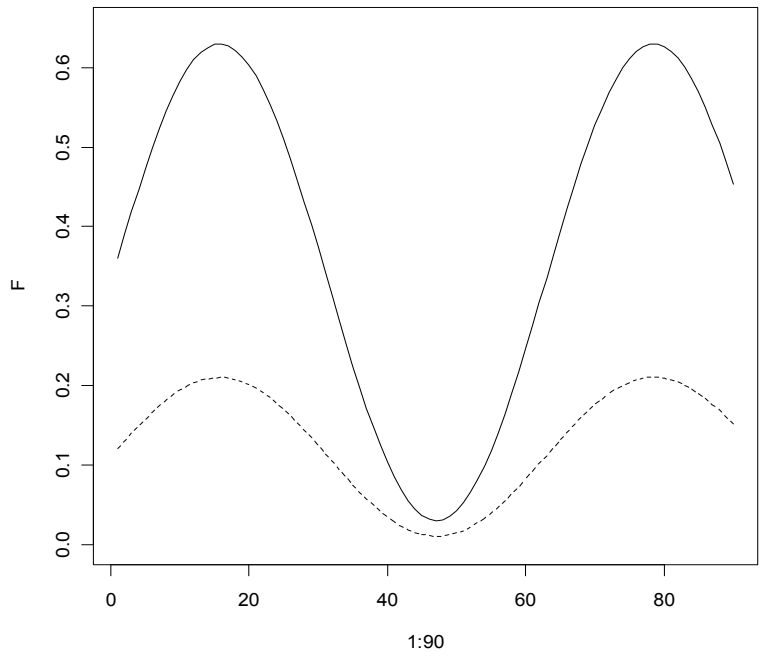


Figure 29. Two examples of harvest rate time-series (over 90 time-steps) used in simulations. The solid line is for an input F=0.3; the dashed line is for an input F=0.1.

The outputs from each scenario is the time series of spawning biomass, and various other population quantities such as fishable biomass, recruitment etc. (though these are not used much in these explorations), and the time series of the full set of indicators listed in sub-section 2.1 above. In addition to the values of the indicators (relative to their unexploited values), we also obtain their binary values relative to a set of trigger points. This is, however done in a post-

processing phase, to avoid having to rerun the whole scenario, if we want to explore an alternative set of trigger points.

9.3. Results: Prediction errors of single indicators

9.3.1. What affects Prediction error?

We explore the factors that affect prediction error to better understand (and illustrate) the performance of indicators under different scenarios. This sub-section mostly uses the low m life-history strategy with flat fishery selectivity and high steepness unless otherwise indicated.

9.3.1.1 Choice of Trigger

Two sets of trigger points, MSY-based and SPR-based, were chosen to evaluate the prediction error of each indicator. In an ideal situation, the trigger points would be based on parameter inputs that are identical to (or at least very close to) the true population parameters. Although this is the way most of the subsequent analyses have been done, it is informative to consider what happens when the trigger points are much lower or higher than the 'correct' values. Recall that many of the quantities we assume known here, will not be known in reality, so mis-specification of trigger points is not an unlikely. Furthermore, the value of a trigger point affects the prediction error for that indicator. In this sub-section we illustrate the effect of using a higher or lower value than the one coinciding with the 'true' parameters for the population.

If the trigger point is very high and the indicator is always below the trigger point (i=0), then the Type II prediction error will be equal to the probability that B is above its reference point (b=1). At the other extreme, if the trigger point is too low and the indicator always above the trigger point (i=1) then the Type I prediction error will be equal to the probability that B is below its trigger point (b=0). Recall that the overall prediction error is a weighted sum of the two types of error (see Methods above). It is also important to note that the probability of B being above or below its reference point similarly depends on the choice of its reference point (e.g. whether it is 20% or 40% of unexploited B, or Bmsy) and the harvest rate(s) used to simulate the data.

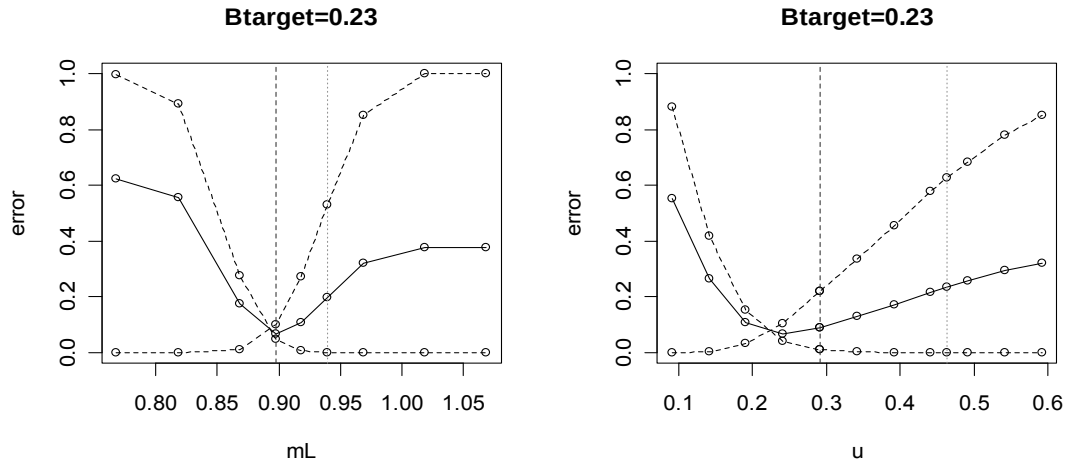


Figure 30. Prediction errors for a range of relative trigger point values for mean length (left panel) and cpue (right panel). The biomass target is 0.23 (B_{msy}/B_0). The dotted lines show the type I and type II errors respectively and the solid line shows the overall prediction error. The dashed vertical line (on the left) shows where the MSY-based trigger is and the vertical dotted line (on the right) shows where the SPR40% trigger lies. Both panels are based on a simulation with input F=0.3 ($1.2F_{msy}$).

Figure 30 illustrates the effect of the choice of trigger (x-axis) on the prediction errors of type I, type II and overall. Note that there is a point where the two lines cross, i.e. where the type I and type II errors are equal. This is often at, or very close to the minimum overall error. The biomass reference is B_{msy} which is why the minimum is close to the MSY triggers. The minimum would be closer to the SPR triggers (the green dotted line) if we use a higher biomass reference (e.g. $0.4B_0$)

9.3.1.2 Harvest rate

Figure 31 shows the effect of harvest rate, F, on the prediction error over a wide range of trigger points. It also shows how, at low F (left panel) the overall prediction error is almost identical to the '10' error (b=1, i=0) because p(b=1) is almost 1 and p(b=0) is almost 0. At higher F (right panel), the overall error tends more to the '01' error, because then p(b=0) is almost 1 and p(b=1) is almost 0.

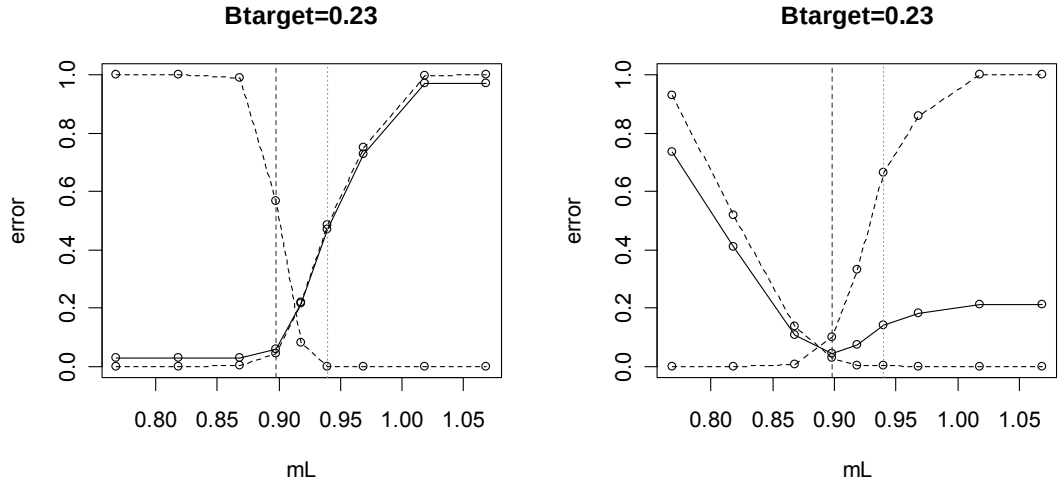


Figure 31. Prediction errors for a range of relative trigger point values for mean length. The biomass reference is 0.23 (B_{msy}/B_0). The dotted lines show the type I and type II errors respectively and the solid line shows the overall prediction error. The vertical dashed line (on the left) shows where the MSY-based trigger is and the vertical dotted line (on the right) shows where the SPR40% trigger lies. The left-hand panel is based on a simulation with input F=0.1 and the right-hand panel on input F = 0.6 to generate sine-wave time-series of F. F_{msy} is 0.25.

For the rest of this sub-section, we only use MSY-based triggers when evaluating prediction errors. Spawner-per-recruit based triggers are considered in sub-section 3.4.

The effect of different harvest rates on the overall prediction error for the 13 indicators is shown in Figure 32. Recall that this is the low m, high h, low noise scenario with MSY-based triggers. Three points are worth noting. First, the most sensitive indicator is the CPUE of small fish (us). As mentioned before, this indicator is not expected to be a good indicator of spawning biomass, but could potentially be considered as an indicator of recruitment. Second, the prediction error is (again, as expected) higher when the harvest rate varies by only a small amount around a constant level than when it varies very widely. Third, when F covers a wide range and the noise in the indicators is low, there is only relatively little difference between prediction errors for the other indicators under the different F-values.

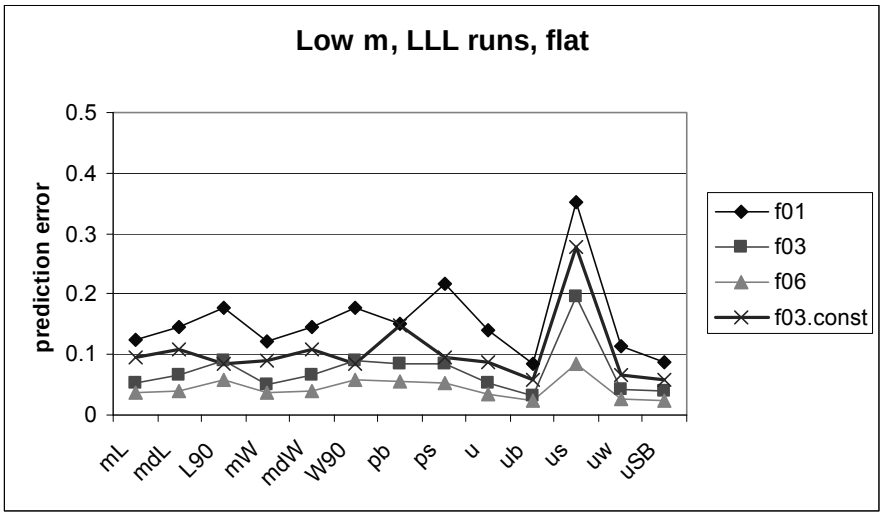


Figure 32. Prediction errors of indicators for 4 different harvest rate scenarios. The first three are based on the sine wave with input values of 0.1, 0.3 and 0.6 respectively (f01, f03, f06). 'f03.const' has a lognormal distribution around 0.3, and only varies between 0.19 and 0.44 (CV=0.2). The sine wave version (f03) varies between 0.03 and 0.62 (see Methods sub-section for more detail; descriptions of indicator labels, horizontal axis, are given in sub-section 2.1 above).

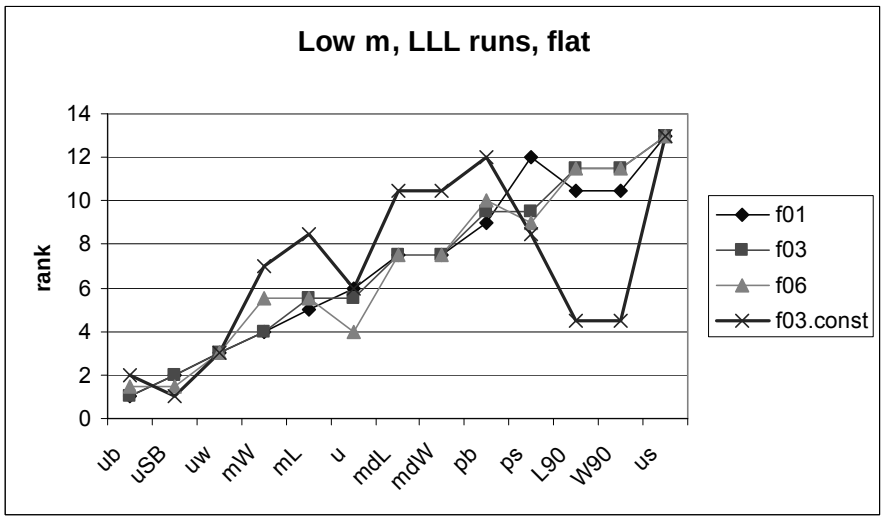


Figure 33. Indicators ranked by prediction error (low error implies a low rank; low rank is good) for each of the harvest rate series and sorted according to the scenario with input F=0.3. Harvest rate scenarios are as in Figure 32. Descriptions of indicator labels are given in sub-section 2.1 above.

When we turn the prediction errors into ranks, we again note that the harvest rate affects the ranking (though note that ranking exaggerates the differences between indicators). It is still, however, the case that the indicators ub, uSB, uw and mW perform best on the whole. Points of note are that L90 and W90 perform well when F varies slightly, but not when F varies widely.

9.3.1.3 Relative versus Absolute trigger points

In these simulations where the life-history parameters are allowed to vary between realisations, it also makes a difference whether absolute trigger points are used with absolute indicator values (e.g. mean length in cm) or whether the trigger is expressed relative to an expected unexploited level (e.g. $mL(F=F_{msy})/mL(F=0)$). Using relative trigger points lead to a lower prediction error than absolute trigger points.

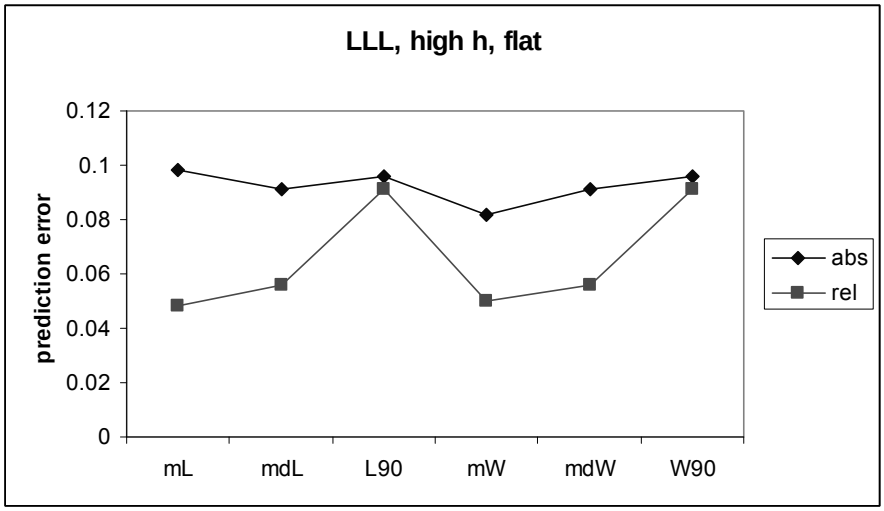


Figure 34. Prediction error of size based indicators using MSY-based triggers, either as absolute quantities ('abs') or relative to unexploited values ('rel'). This example is for low m, low noise, high h and widely varying harvest rate.

The proportion big and small (pb and ps) are treated as absolute because several comparisons showed very little difference between prediction errors for these two indicators when treated as absolute or as relative indicators.

The CPUE-based indicators are also treated as relative, but this is essentially because it is unlikely that the stock size and catchability would be known. It will also become clear that relative trigger points are more robust to errors in the input parameters used to derive them than absolute trigger points.

9.3.1.4 Increased measurement errors in indicators

The above simulations had relatively low errors for all three sources: 1. the catch-at-age error, 2. the CV of CPUE and 3. the sample-size for length / weight frequency distributions. Here we increase the level of error on the three components. Recall that a low sample size leads to higher error, so that 'high'=H coding for this component should be interpreted as 'high error'.

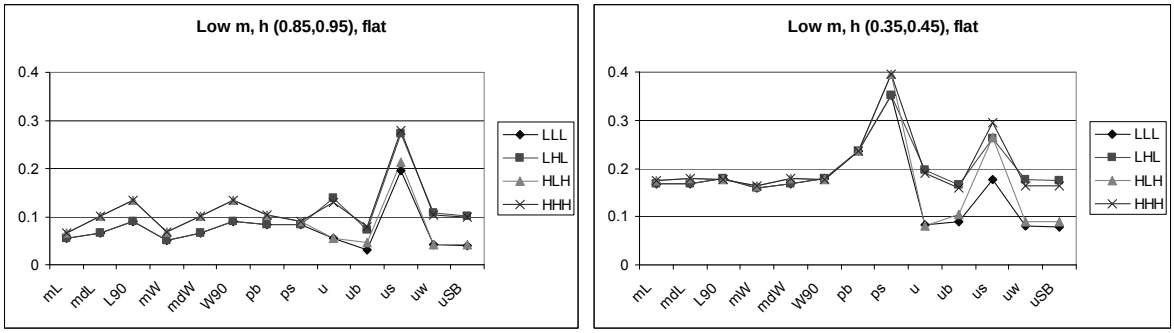


Figure 35. Prediction errors for the 13 indicators and at 4 different levels of noise in the 3 sources of error on indicators. The 3-letter code reflects the levels (L=low, H=high) for the 3 sources of error: catch-at-age, cpue and size-frequency sample size. See text for more detail. The left panel is for high steepness, the right panel for very low steepness. All scenarios have low m and a flat selectivity curve.

The patterns of prediction errors at different noise levels behave as one might expect. When the sampling error is high, the size-based indicators have higher prediction errors than when

sampling error is low. When the CV of CPUE is high, the prediction error of CPUE-based indicators is higher than when the CV of CPUE is low. Figure 35 shows 4 levels of noise and illustrates how the prediction errors differ under these scenarios. The higher catch-at-age error affects both types of indicator (e.g. LLL vs. HHH). The four noise levels shown here represent the biggest differences between patterns of prediction error (over the 13 indicators), and the rest of the explorations therefore focus on these cases: the extremes, case LLL and HHH, and the two combinations LHL (high CV on CPUE, others low) and HLH (low CV on CPUE, others high).

It is not easy to make direct comparisons between the level of noise on size-based indicators and on CPUE-based indicators, because they are very different in character. Also recall that some of the CPUE indicators are products of CPUE in numbers and a size-based indicator (e.g. mean weight), so they would be affected by all three sources of error.

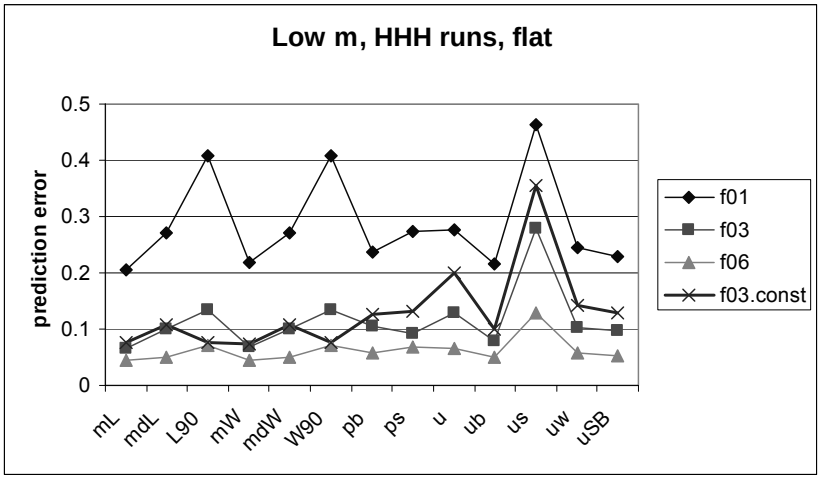


Figure 36. Prediction errors of indicators for 4 harvest rate scenarios with high noise for all three error sources (see text). The first three are based on the sine wave with mean values of 0.1, 0.3 and 0.6 respectively. 'f03nosin' also has a mean F of 0.3, but only varies between 0.19 and 0.44 (CV=0.2). The sine wave version (f03) varies between 0.03 and 0.62.

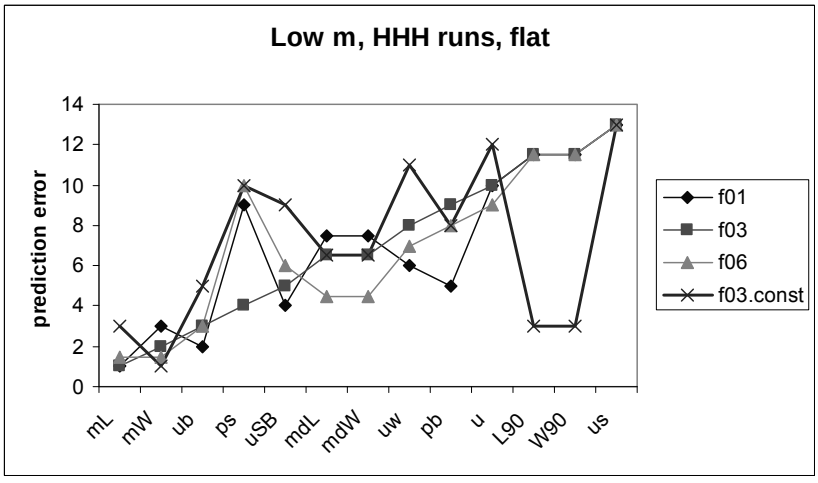


Figure 37. Indicators ranked by prediction error (low error implies a low rank; low rank is good) for each of the harvest rate series and sorted according to the scenario with mean F=1.2*Fmsy. Harvest rate scenarios are as in Figure 36.

For comparison, we also illustrate the effect of harvest rate on the prediction errors and ranks of the 13 indicators (Figure 36 and Figure 37). When noise is much higher, the difference between prediction errors at low F and high F is more extreme. This is understandable since the low F

implies less of a change in the indicator and with higher 'noise' the indicator is more likely to be mis-interpreted than with high F. For the widely varying F values, the top three indicators (noting again the fact that ranks exaggerate the differences) are mL, mW, and ub. The proportion small, ps, is only fourth for the F=0.3 (sine wave) scenario, and one could argue that uSB is a better candidate. Figure 36 shows that there is only a very small difference between prediction error for pb and uSB for the F=0.3 case. The L90 and W90 indicators perform well for the scenario where F varies only slightly, but poorly when F varies widely; this is the same as seen for the LLL example above.

9.3.1.5 The effect of Steepness

There are several reasons why one would expect prediction errors to be affected by the steepness in the stock-recruit relationship. First, F_{msy} is a function of steepness, with higher values when steepness is high than when it is low. This in turn means that the expected difference in an indicator for the unharvested stock and that harvested at F_{msy} is greater for high than low steepness. Trigger points (MSY-based ones) are obviously also affected. Even when spawner per recruit-based triggers are used, one would expect differences, but this is considered in the next sub-section.

For the comparisons below, the four steepness ranges (see Methods above) were used with a sine wave harvest rate around 1.2*Fmsy (Fmsy for that steepness range). The level of prediction error is higher at low steepness (Figure 38). We note that the proportion of B below or above its reference point is very similar for the different steepness ranges (0.66 for high h and around 0.60 to 0.61 for the others).

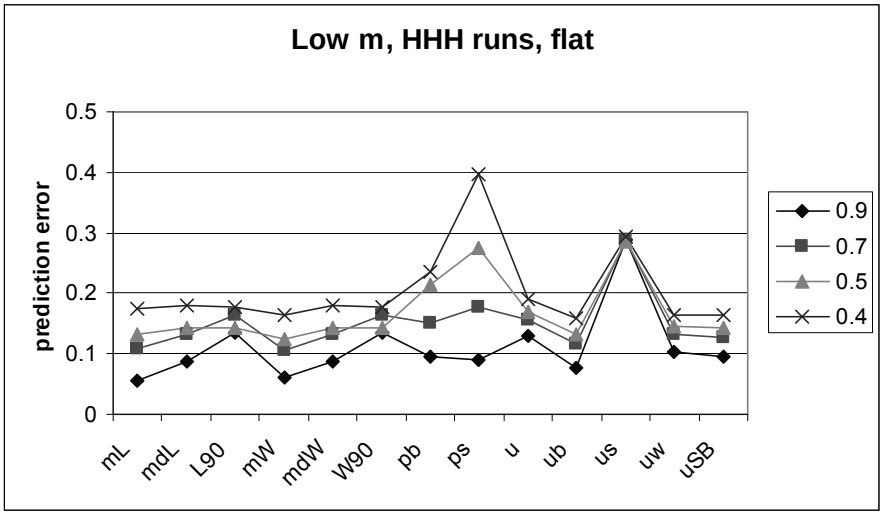


Figure 38. Prediction errors for different steepness ranges (the means are shown in the legend) and for low m, high noise, flat selectivity. The harvest rate is matched to the steepness with a sine wave constructed around 1.2*Fmsy (given h).

At low steepness, the performance of L90 and W90 are much more similar to that of the means and medians, whereas at high steepness, the means (mL and mW) perform better. The performance of the size-based indicators is more strongly affected by steepness than the CPUE-based indicators, and the proportion small fish is particularly strongly affected.

9.3.1.6 Recruitment variability

In most situations where indicators are considered as the basis for management, the magnitude of inter-annual variability in recruitment is likely to unknown or, at best, poorly estimated. Recruitment variability does, however, affect the prediction error of the different indicators. Patterns of prediction error are quite similar under low and high measurement errors (LLL and HHH scenarios). More marked is the difference between low and high recruitment variability when steepness is low. In this case (right-hand panels in Figure 39 and Figure 40) the size-based indicators perform substantially more poorly when recruitment variability is high. The effect is much weaker when steepness is high. It is also noticeable that the ranking of the three quantities mean, median and 90th percentile (of both length and weight) switches between high and low steepness when recruitment variability is high.

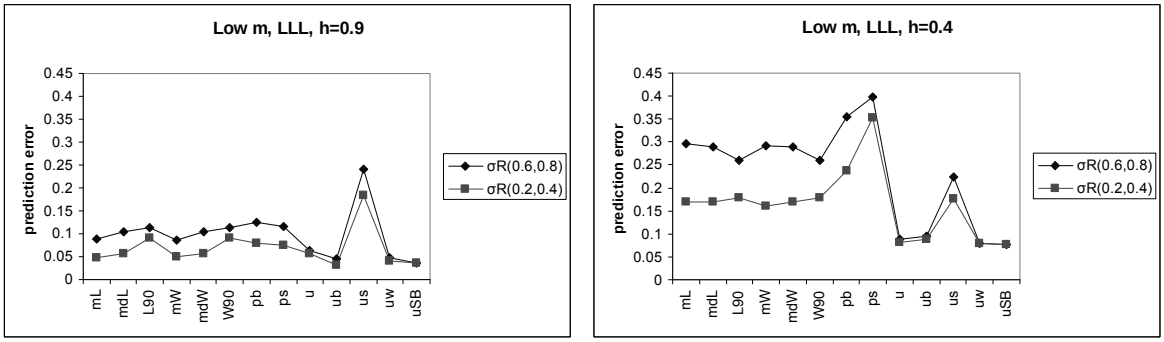


Figure 39. Prediction errors for two ranges of recruitment variability: high σ_R = (0.6,0.8) and low σ_R = (0.2,0.4). In all cases the measurement error is low (LLL), selectivity is flat, and steepness is high (left panel), or very low (right panel).

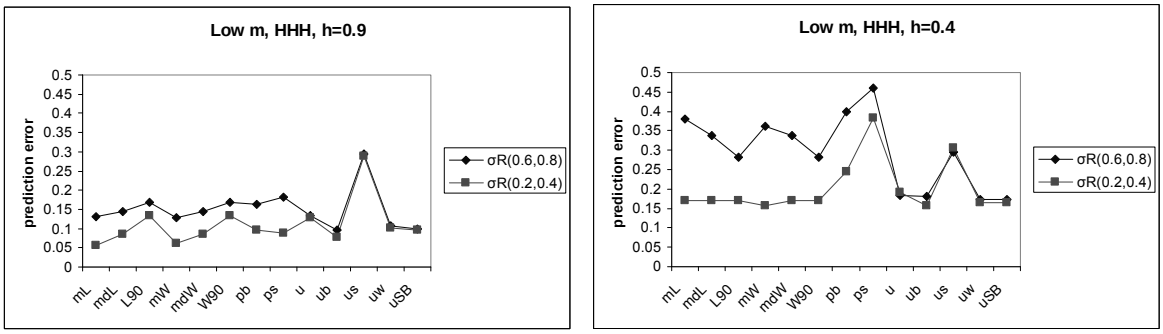


Figure 40. Prediction errors for two ranges of recruitment variability: high σ_R = (0.6,0.8) and low σ_R = (0.2,0.4). In all cases the measurement error is high (HHH), selectivity is flat, and steepness is high (left panel), or very low (right panel).

These results suggest that, when recruitment variability is high, the size-based indicators are unlikely to perform better than the CPUE-based indicators, and may actually perform much more poorly. We therefore do the rest of the explorations on the cases where the performance of the size- and CPUE-based indicators are more similar, i.e. when recruitment variability is low, σ_R = (0.2,0.4).

9.3.1.7 Two further observations

First, note that the above results have been summarised over all 100 replicates. Although this provides an indication of the difference between mean prediction error of each indicator, it does not reflect the range of prediction errors when viewed by realisation. Figure 41 shows that the lower quantiles and minima follow a very similar pattern to that of the mean and median

prediction errors, but that the upper quantiles and maxima can follow a somewhat different pattern. For example, although the median prediction error of L90 is higher than that of mL, its range is much narrower and its upper quartile and maximum are lower than those of mL. The range of prediction errors for W90 is similarly narrower than those of either mean or median weight. The CPUE-based indicators generally have similar inter-quartile ranges.

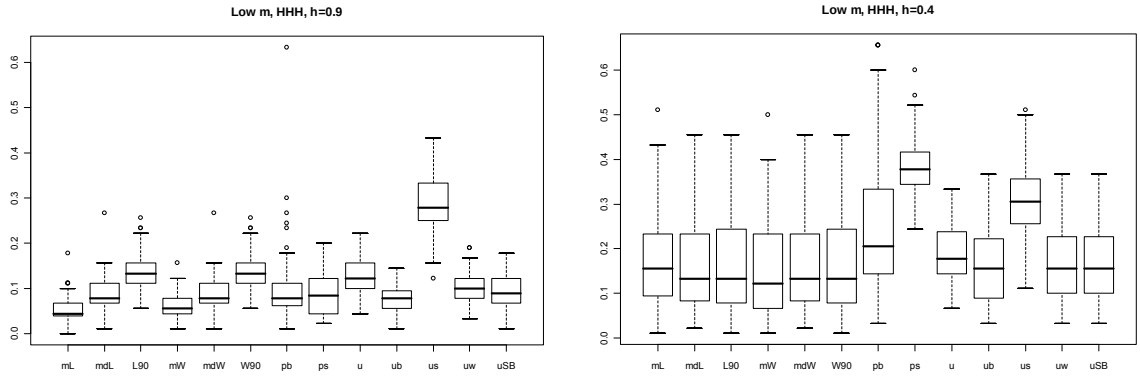


Figure 41. Range of prediction errors for each indicator (i.e. when calculating this by realisation) for high steepness (left panel) and low steepness (right panel). In both cases the noise level is HHH.

In the rest of the analyses, we only consider the overall prediction errors, rather than prediction errors by realisation. The 'by realisation' results are, however, relevant when doing this type of simulation study for a specific species with specific parameters and quantified uncertainties.

A second observation, based on results so far, is the similarity in performance between (i) all the size-based indicators, and (ii) all the CPUE-based indicators (except 'us'). This is not surprising when we look at the correlation between indicators. The size based indicators are highly correlated with one another, and the CPUE indicators are also highly correlated with one another. From the point of view of information content, it implies that we really only have two (possibly 3, if we count the two 'proportion' indicators separately) groups of indicators. The shaded area in Table 10 shows that the correlation between the size- and CPUE-based indicators is also quite high.

Table 10. Correlation matrix for the 13 indicators for the low m, HIGH steepness, high noise runs. Correlations are for the raw simulated data, not the converted binary values. The correlation of 1 between mdW and mdL and between W90 and L90 is due to the fact that there is no error or noise in the length-to-weight conversion.

	mL	mdL	L90	mW	mdW	W90	u	ub	uw	uSB
mL	1									
mdL	0.96	1								
L90	0.85	0.8	1							
mW	0.99	0.95	0.88	1						
mdW	0.95	1	0.8	0.96	1					
W90	0.85	0.8	1	0.88	0.8	1				
u	0.55	0.54	0.53	0.58	0.56	0.54	1			
ub	0.68	0.67	0.61	0.71	0.7	0.63	0.92	1		
uw	0.64	0.63	0.6	0.67	0.66	0.62	0.98	0.96	1	
uSB	0.65	0.64	0.61	0.69	0.67	0.62	0.97	0.97	1	1

The high correlation is only partly related to the way in which data are simulated, for example, the fact that weight is calculated as a function of length. In practice, the high correlations between size-based indicators are still likely to hold, particularly if a length-weight relationship is used instead of direct weight frequency data. The correlations may well be lower than in the table

above, but they are still likely to be higher than the correlations between the size and CPUE-based indicators.

Table 11. Correlation matrix for the 13 indicators for the low m, LOW steepness, high noise runs. Correlations are for the raw simulated data, not the converted binary values. The correlation of 1 between mdW and mdL and between W90 and L90 is due to the fact that there is no error or noise in the length-to-weight conversion

	mL	mdL	L90	mW	mdW	W90	u	ub	uw	uSB
mL	1									
mdL	0.9	1								
L90	0.7	0.62	1							
mW	0.99	0.9	0.75	1						
mdW	0.9	0.99	0.62	0.91	1					
W90	0.69	0.62	1	0.75	0.63	1				
u	0.25	0.26	0.26	0.27	0.28	0.27	1			
ub	0.48	0.48	0.39	0.5	0.5	0.39	0.9	1		
uw	0.4	0.4	0.37	0.43	0.42	0.38	0.98	0.95	1	
uSB	0.41	0.41	0.37	0.43	0.42	0.38	0.97	0.95	1	1

Although broad patterns are similar, it is noticeable how the correlation between CPUE in numbers and the size-based indicators is lower when steepness is low (Table 11) than when it is high (Table 10). Below, where several indicators are used together in classification trees, some of the implications of these correlations will become clear.

In Section 8 we also illustrated the potentially strong patterns of autocorrelation in each indicator time-series. The autocorrelation is essentially being ignored here. We note, however, that in principle, the autocorrelation could be taken into account by, for example, using indicator values for more than one year to 'predict' the status of spawning biomass. This is briefly considered in sub-section 4.2 below.

9.3.2. The effect of selectivity: long-lived life history, MSY-based triggers

When selectivity is dome-shaped, very large, old fish are not present in the catch. This is likely to have strong effects on the performance of indicators, particularly when natural mortality is low. The prediction errors are much higher for all indicators when selectivity is dome-shaped than when it is flat (Figure 42). It is important to remember that the harvest rate patterns are different because F_{msy} is higher (e.g. F_{msy}=0.25 for flat selectivity; F_{msy}=0.44 for the dome-shaped selectivity; with high steepness in both cases). It is noticeable that 'us' (cpue of small fish) has a much more similar prediction error to the other indicators when selectivity is dome-shaped than when it is flat.

It is also informative to look at the correlation between each indicator and spawning biomass (the 'raw' simulated data, not the binary values). This is shown in Figure 43. It makes intuitive sense that the proportion small fish (ps indicator) would be unaffected or only slightly affected, depending on where the selectivity curve lies relative to the definition of 'small'.

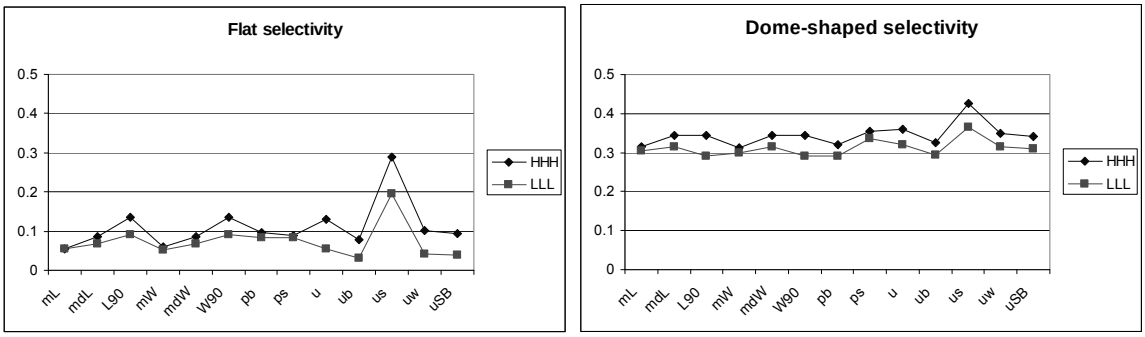


Figure 42. Prediction errors for the 13 indicators under the high (HHH) and low (LLL) noise assumptions and for flat selectivity (left panel) and dome-shaped selectivity (right panel). In all cases the scenario is low m, high steepness.

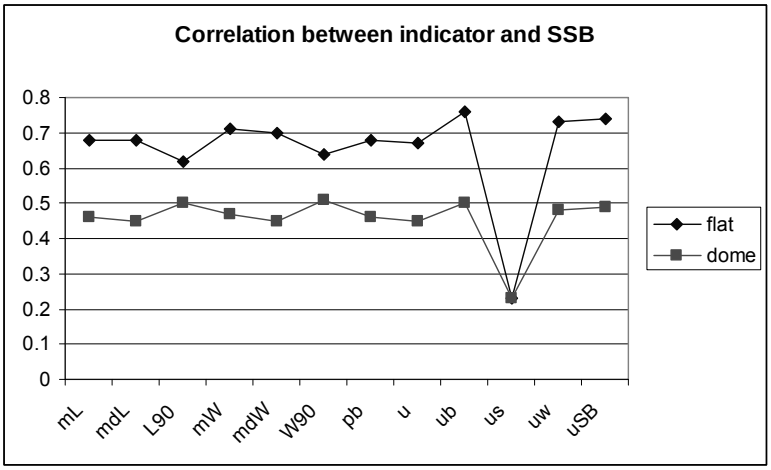


Figure 43. Correlation between each indicator and the spawning biomass ('raw' simulated data, not binary values) for flat and dome-shaped selectivity. The scenarios are low m, high steepness and high noise (HHH). Note that the ps (proportion small fish) indicator has not been plotted because it has negative correlation which confuses the graph. The correlations for ps are: flat= -0.61, dome= -0.46.

The set of figures below look at the prediction errors over the four levels of steepness and for both flat (Figure 44) and dome-shaped selectivity (Figure 45).

The general picture is that some size-based indicators perform as well as, or sometimes slightly better than, CPUE-based indicators when steepness is high, but more poorly (or much more poorly) when steepness is low. The difference in performance, between high and low steepness levels, is stronger for low noise scenarios than for high noise scenarios. When selectivity is dome-shaped and noise is high, there is relatively little difference in prediction errors between steepness levels (at least compared to the other scenarios).

The CPUE based indicators ub, uw, uSB perform consistently well over the steepness ranges, the two noise levels of low (LLL) and high (HHH), and the two selectivity patters (flat and dome-shaped). One could argue that if the size-based indicators had low noise compared to the CPUE-based indicators, they would/could perform better. In our experience, the trade-off between the level of noise and the extent of change in the reference point (i.e. signal) are such that it is difficult to get good performance from the size-based indicators over the relatively wide range of scenarios considered here. It is, of course, true that if CPUE is unreliable, highly biased or much more highly variable than considered here, the size-based indicators are likely to outperform the CPUE-based indicators.

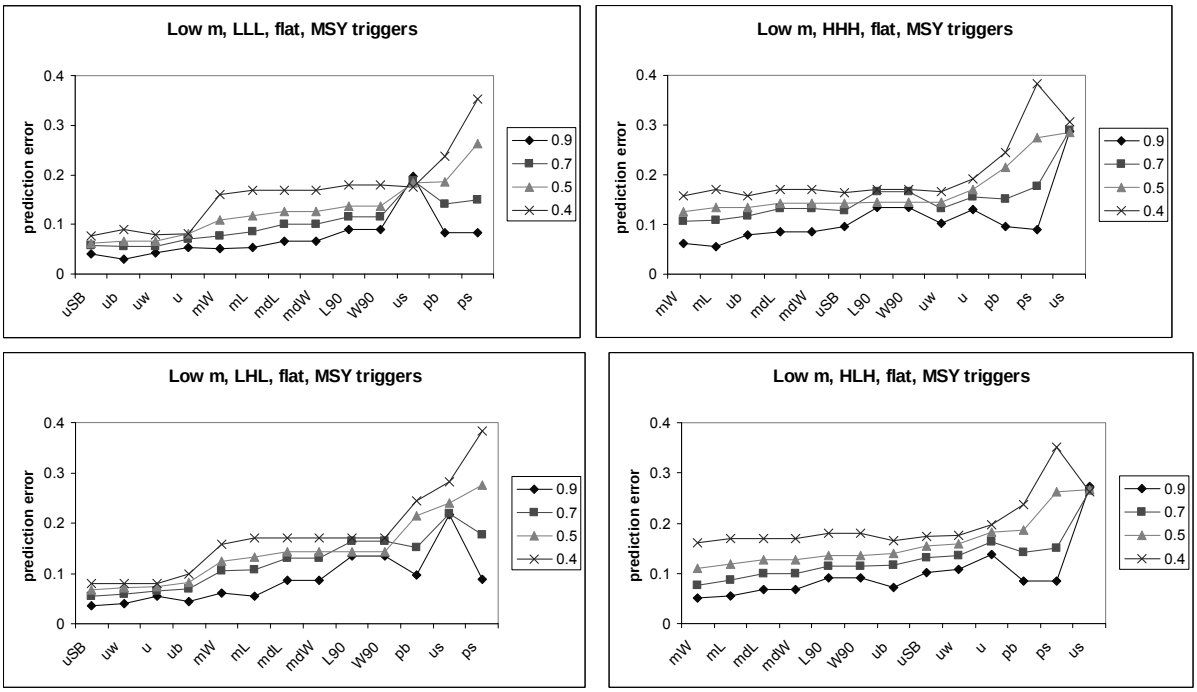


Figure 44. Prediction errors for a range of steepness levels and for four noise levels: low noise (LLL), high noise (HHH), and the combinations LHL and HLH (see text for interpretation). All cases are for low m, FLAT selectivity and MSY triggers. In each panel, indicators are sorted by the prediction errors for h=0.5 (triangles).

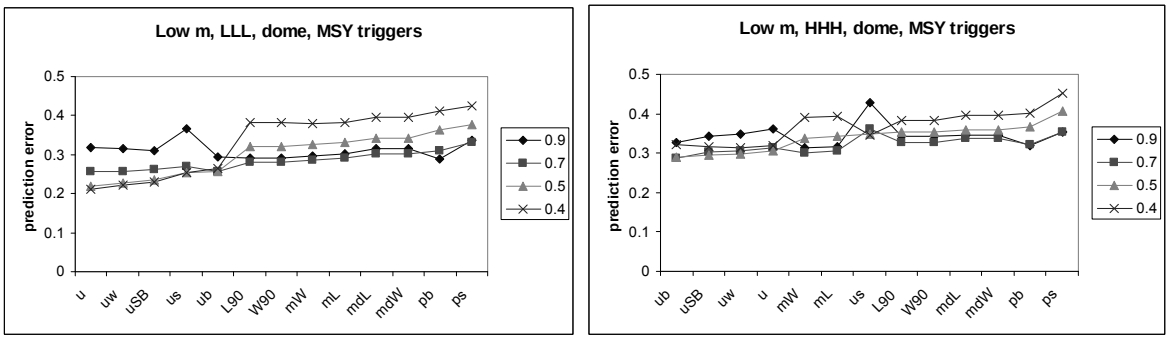


Figure 45. Prediction errors for a range of steepness levels and for two noise levels: low noise (LLL) and high noise (HHH). All cases are for low m, DOME-shaped selectivity and MSY triggers. In each panel, indicators are sorted by the prediction errors for h=0.5 (triangles).

9.3.3. The effect of selectivity: short-lived life history; MSY-based triggers

The short-lived life history scenario has much higher natural mortality and growth rate than the long-lived life history scenario (Table 1 above). We have also assumed a lower age-at-maturity and age at first capture as one would expect in reality. This implies that F_{msy} is higher and there are fewer age classes in the catch (and the population) than for the long-lived life history scenarios. The differences between indicators, for example between mean, median and upper 90th percentile of both length and weight, are now more pronounced than for the long-lived life history. Compare, for example, Figure 46 with Figure 44. The extent to which size-based indicators perform better than CPUE-based indicators at high steepness, and vice versa at low steepness, is also much stronger for the short-lived than long-lived scenario.

There is also now hardly any difference between the prediction error patterns for flat and dome-shaped selectivity (Figure 47). This is because of the high mortality which means that the catch is dominated by the first several age classes and it therefore matters less whether the older age classes have high or low selectivity. This is illustrated in Figure 48 which shows the proportion survivors at age when only natural mortality operates. The effect is even stronger when fishing mortality at age is also taken into account.

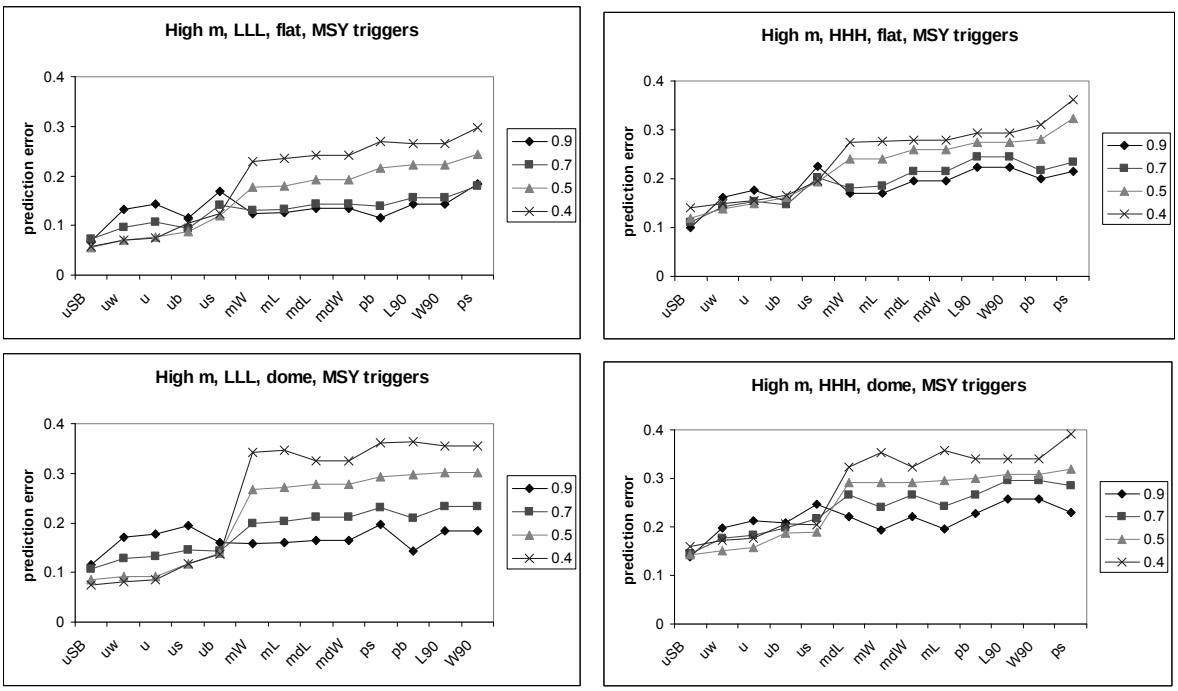


Figure 46. Prediction errors for a range of steepness levels and for low noise (LLL; left panel) and high noise (HHH; right panel). All cases are for high m, FLAT selectivity and MSY triggers.

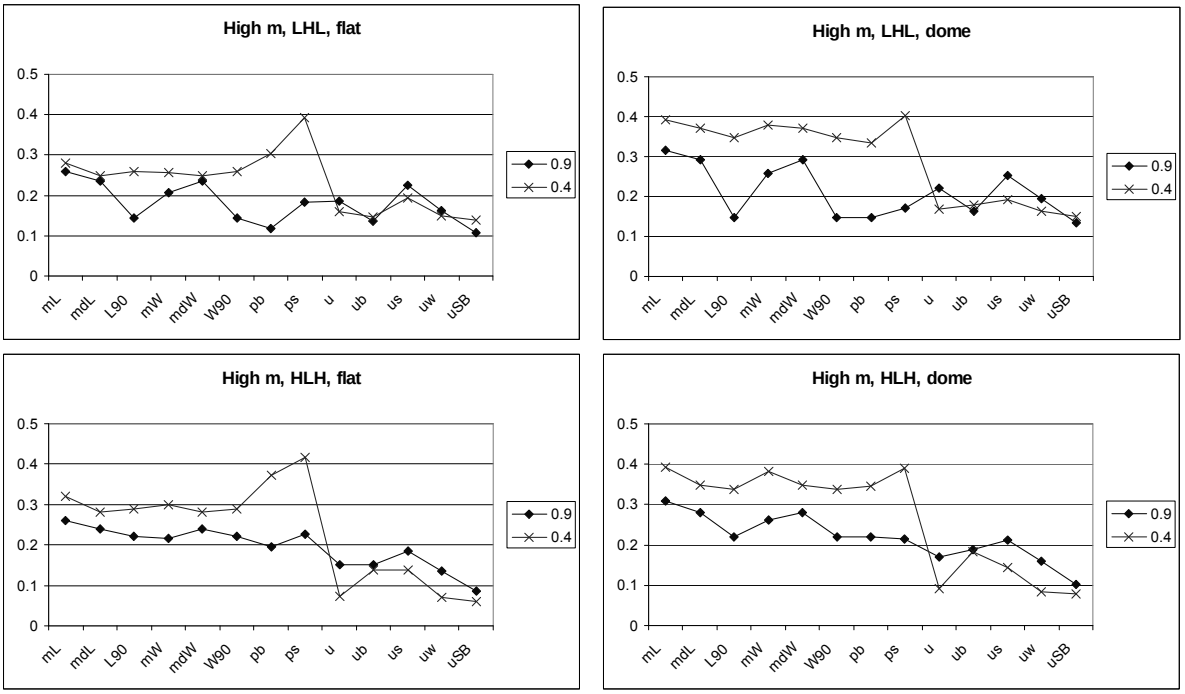


Figure 47. Prediction errors for the very low and high steepness levels (means of 0.4 and 0.9) and for the LHL and HLH noise levels with either flat or dome-shaped selectivity. All cases are for high m and MSY triggers. Indicators are in the same order in each panel and not sorted according to prediction error.

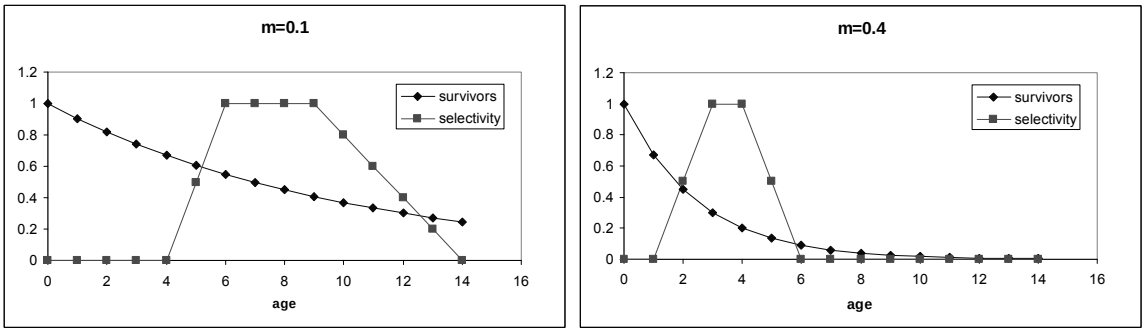


Figure 48. Proportion of survivors at age (under natural mortality only) and dome-shaped selectivity at age for the long-lived (m=0.1) and short-lived (m=0.4) scenarios.

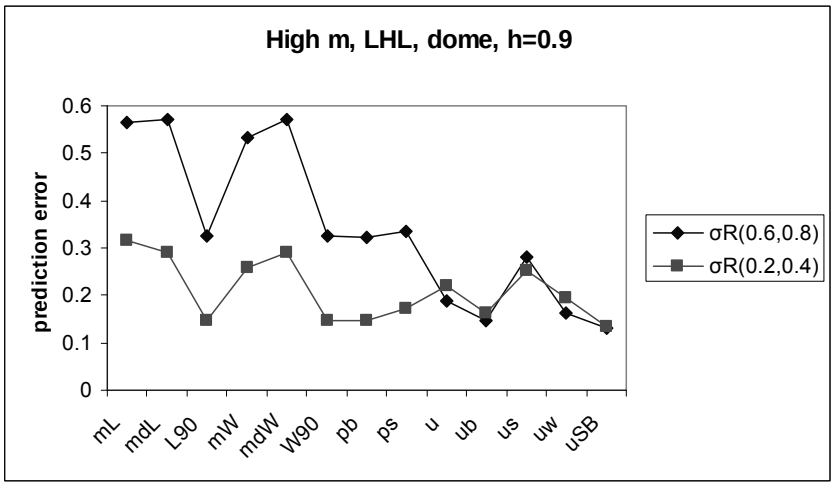


Figure 49. Prediction errors for two levels of recruitment variability: high σ_R = (0.6,0.8) and low σ_R = (0.2,0.4). In all cases the measurement error is high on CPUE (LHL) and steepness is high (0.9).

The effect of increased recruitment variability on the size-based indicators is also stronger than for the long-lived life history; even at high steepness, the prediction errors are almost double (compare Figure 49 with Figure 40).

9.3.4. Spawner-per-recruit based triggers for both life-histories and selectivities

So far, explorations were done with MSY-based triggers to turn the data into binary quantities. Given the difficulties with estimating steepness (even in cases where there is a quantitative or model-based stock assessment), we consider that SPR trigger points are likely to be much more practical in most applications where indicators are likely to be used in a decision rule. Recall that SPR triggers themselves are not dependent on steepness, but we have retained separate scenarios for the different ranges of steepness as before insofar as the steepness still relates to the 'true' simulated population dynamics. The harvest rate used here, F_{SPR40}, is the same for each steepness scenario (0.xx for low m and 0.xx for high m). (Note that a sine wave F-pattern is still used, but it is constructed around F_{SPR40}).

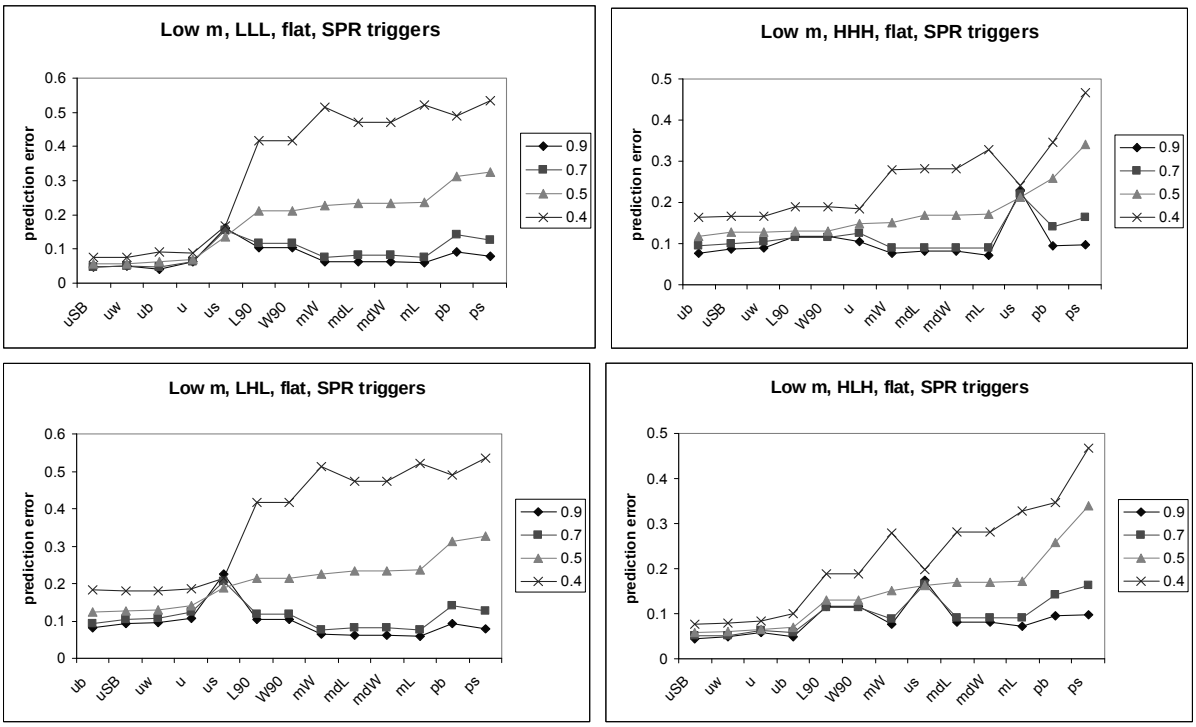


Figure 50. Prediction errors for a range of steepness levels and for four noise levels: low noise (LLL), high noise (HHH), and the combinations LHL and HLH (see text for interpretation). All cases are for low m, FLAT selectivity and SPR triggers. In each panel, indicators are sorted by the prediction errors for h=0.5 (triangles).

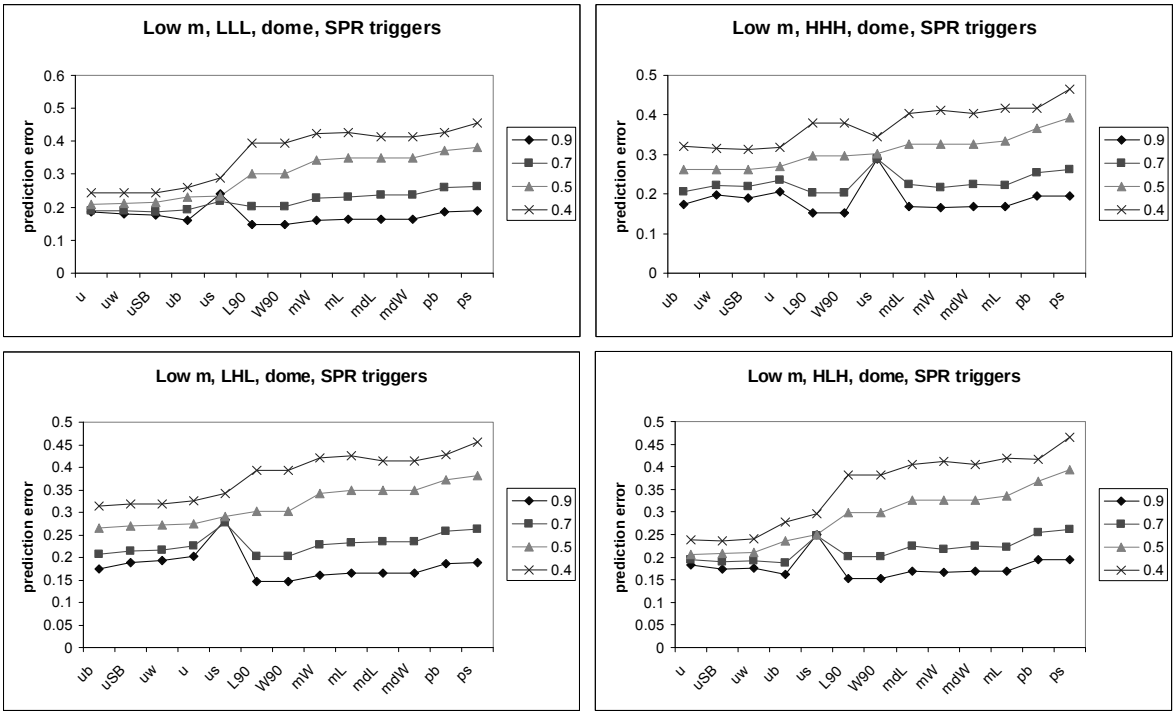


Figure 51. Prediction errors for a range of steepness levels and for four noise levels: low noise (LLL), high noise (HHH), and the combinations LHL and HLH (see text for interpretation). All cases are for low m, DOME-shaped selectivity and SPR triggers. In each panel, indicators are sorted by the prediction errors for h=0.5 (triangles).

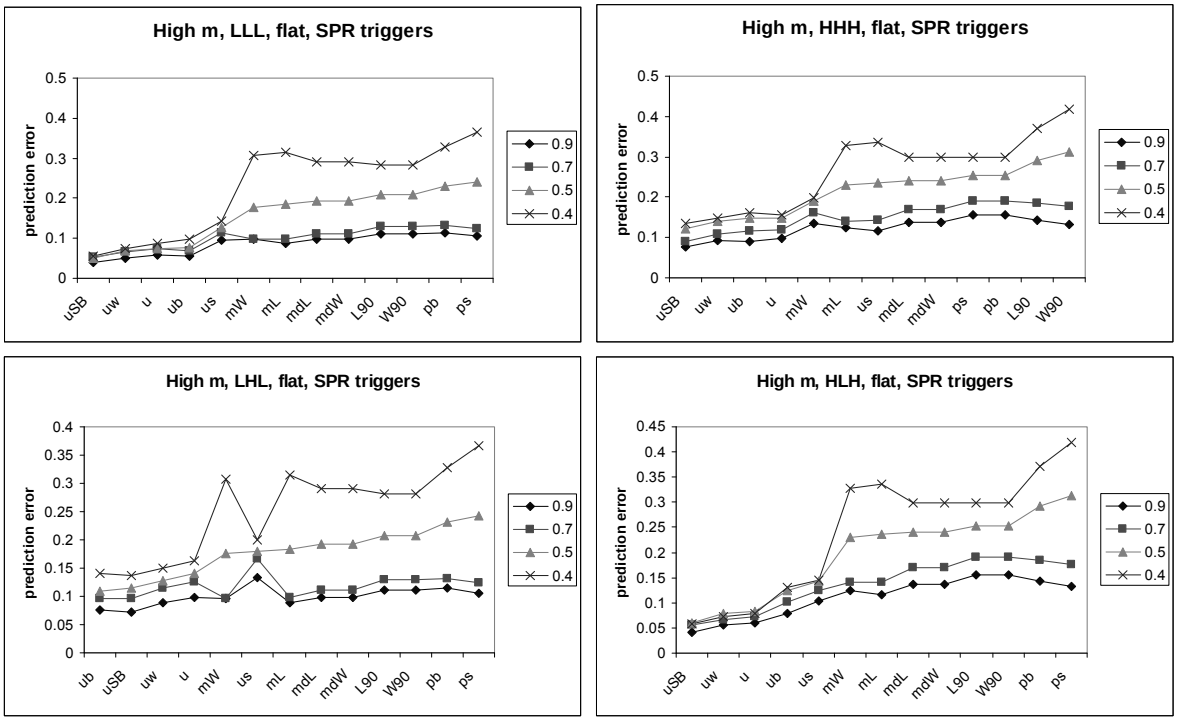


Figure 52. Prediction errors for a range of steepness levels and for four noise levels: low noise (LLL), high noise (HHH), and the combinations LHL and HLH (see text for interpretation). All cases are for high m, FLAT selectivity and SPR triggers. In each panel, indicators are sorted by the prediction errors for h=0.5 (triangles).

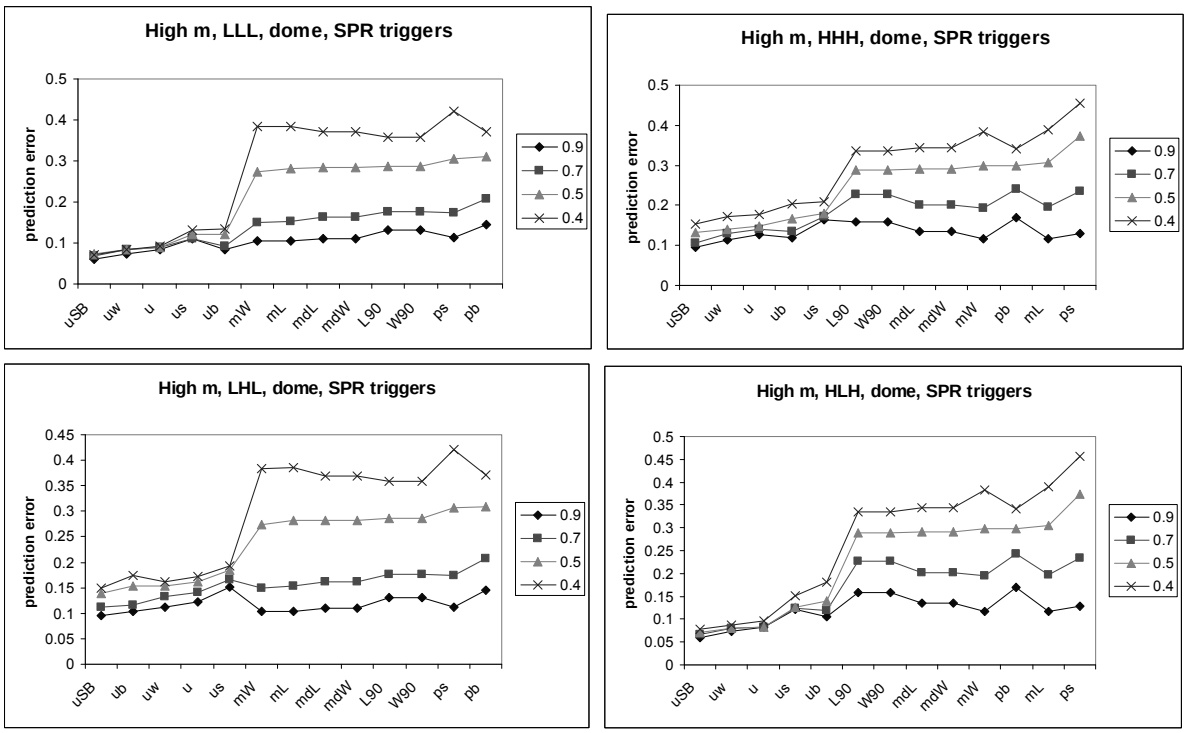


Figure 53. Prediction errors for a range of steepness levels and for four noise levels: low noise (LLL), high noise (HHH), and the combinations LHL and HLH (see text for interpretation). All cases are for high m, DOME-shaped selectivity and SPR triggers. In each panel, indicators are sorted by the prediction errors for h=0.5 (triangles).

First consider the low m scenarios (long lived). There is a big difference in prediction errors between h levels. CPUE indicators perform better at low h and size-based ones are as good (or slightly better) at high h. This is the case for flat and dome-shaped selectivity. Within the groups of size-based indicators, there is relatively little difference between the mean, median and 90th percentile performance. In most cases the mean performs best of the three, but when selectivity is flat and steepness is low the 90th percentiles tend to perform better.

Next, consider the high m scenario (short lived). Here the differences between the steepness levels is much less and, unlike the low m scenario, the relative performance of indicators remain more similar over the range of h values. As in the case of the MSY-based triggers, there is little difference between the two selectivity assumptions. There is a big difference between the means, medians and 90th percentiles of both length and weight, with the 90th percentiles performing much more poorly than the means or medians. Together with the proportion small and the CPUE of small fish, the L90 and W90 indicators are the 4 poorest performing. We repeat the caveat that it is not surprising that the proportion small and CPUE of small fish are usually not good indicators of spawning biomass, but we have not explored their use as indicators of recruitment.

Recall that the runs for each steepness level is based on a different range of harvest rates. For high steepness, the harvest rate is higher than F_{spr40}, but for the low steepness it is higher than F_{spr40}. Implications of evaluating indicators under different ranges of harvest rates were shown to be relatively minor under the scenario where selectivity was flat and MSY triggers were being used. In the process of evaluating the performance of trees as decision rules, we found that there are some cases where the implications are much stronger. One such example is for dome-shaped selectivity with SPR triggers. Figure 54 shows how the two different harvest rate patterns (sine wave or randomly varying (low sd) around a constant) affect the prediction errors and hence the relative performance of the different indicators.

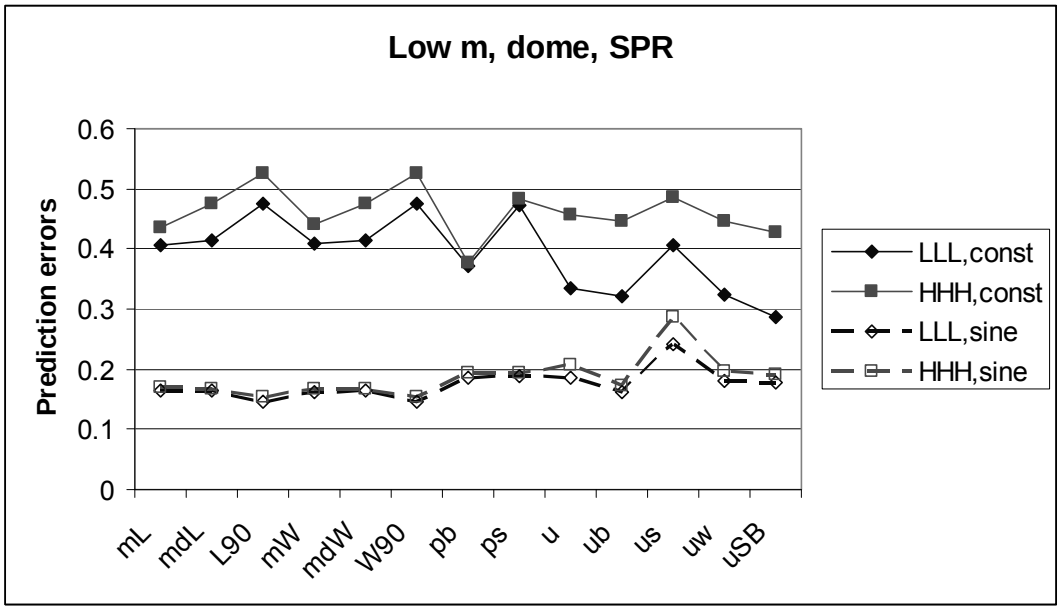


Figure 54. Prediction errors for widely varying harvest rate (F is a sine function, 'sine') or randomly varying F around a constant ('const') and for low (LLL) and high (HHH) noise levels. All cases are for low m, dome-shaped selectivity and using SPR triggers.

Note in particular how L90 and W90 are the best indicators when F varies widely as a sine wave, but how they are in fact the poorest when F is approximately constant and relatively low (near Fspr40). Also note that for constant F there is more difference in prediction error between the

low and high noise scenarios. This implies that prediction errors, particularly for the CPUE-based indicators, are more sensitive to noise levels when the harvest rate does not vary greatly. This is not surprising.

9.3.5. Effort Creep

Here we illustrate some of the possible effects of effort creep (i.e. an increase in catchability which affects all CPUE-based indicators), on the performance of single indicators.

Data generated with effort creep inevitably lead to poorer performance of CPUE-based indicators than when there is no effort creep. Whether they perform worse than the size-based indicators depend on the detail, particularly the extent of the effort creep and the difference in performance between the two sets under no effort creep (their relative noise levels, for example, and the life history parameters).

Figure 55 shows how the prediction errors increase when there is effort creep, for the high noise (HHH) scenario and high (h=0.9) and very low (h=0.4) steepness, and with flat selectivity. The difference is largest when steepness is low.

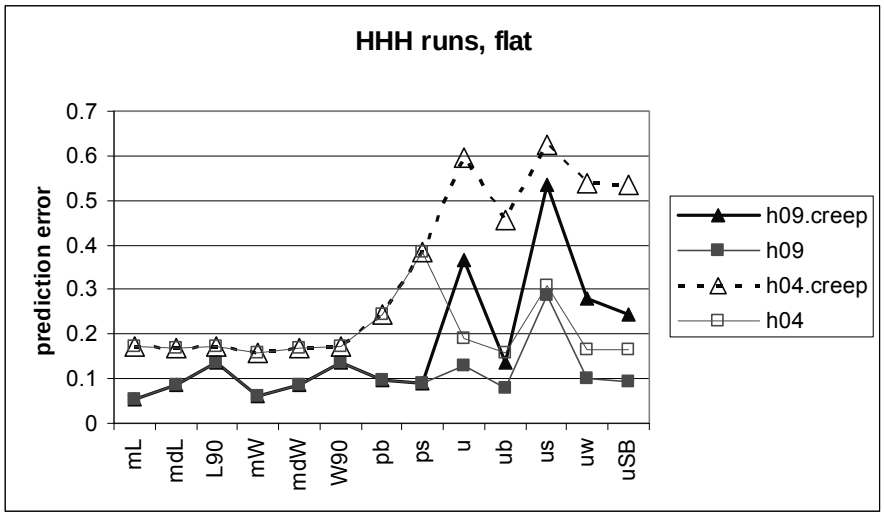


Figure 55. The effect of effort creep on prediction errors for high (h=0.9) and very low steepness (h=0.4) with low mortality, high noise and flat selectivity.

The relative increase in prediction error is less when selectivity is dome-shaped, but the overall patterns are the same (Figure 56). Differences between runs with or without effort creep are also very similar for other combinations of measurement error, e.g. HLH or LHL runs.

When selectivity is dome-shaped, the CPUE-big indicator (ub) is often below its trigger point when the biomass is above its reference point. The effort creep therefore sometimes inadvertently leads to a correct prediction of 'b=1' because it is higher than in the no-creep case. This is why this indicator has a lower overall prediction error for the effort creep scenario than that without effort creep (h=0.9; Figure 56).

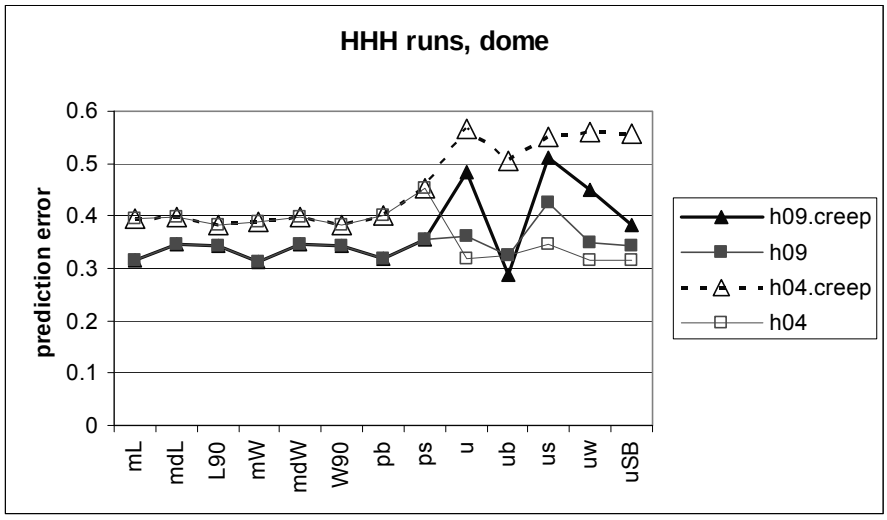


Figure 56. The effect of effort creep on prediction errors for high (h=0.9) and very low steepness (h=0.4) with low mortality, high noise and dome-shaped selectivity.

9.4. Results: Classification Trees for multiple indicators

Here we explore the use of multiple indicators through classification trees and their associated prediction errors. First we use the binary simulated data based on MSY triggers and Bmsy as a reference point. Then we use the binary simulated data based on SPR triggers and B/B0=0.4 as a reference point. All scenarios in this sub-section are for the high measurement errors (HHH), and we only consider the two extreme steepness levels, h=0.9 and h=0.4.

9.4.1 Classification trees based on MSY triggers

9.4.1.1 Low m, long-lived life history

We start with the four scenarios associated with the low m (long-lived) life history: high and very low steepness, crossed with flat and dome-shaped selectivity. It is informative to look at the relative error of the large trees as a function of the number of splits (Figure 57). Initially there is a reduction in the relative error as the number of splits increases. More 'splits' either implies that more than one indicator is being used, or that there are more 'decision points'[20]. Very quickly, however, after about 2 to 4 splits, there is either no further, or only very little further, reduction in the relative error.

The large fitted trees were then pruned to 5 or 6 splits (Figure 58). Although this choice of number of splits is arbitrary and not based on any strict criterion, it is based on two common sense principles. First, the fact that (as noted above) the relative error, as well as, cross validated errors, show that there is little to be gained after about 2 to 4 splits for these types of data. Second, in practice, there would be a lot of benefit for a parsimonious tree which stakeholders can relatively easily understand. Even the trees illustrated below are, in our view unnecessarily large, but they do show the indicators that come in at the top of the tree, the ones that appear lower down, and how the structure of each tree differs between the scenarios.

[20] We make this distinction because any indicator can be used multiple times in the same tree. The number of splits in a tree is therefore not the same as the number of indicators being used in the tree.

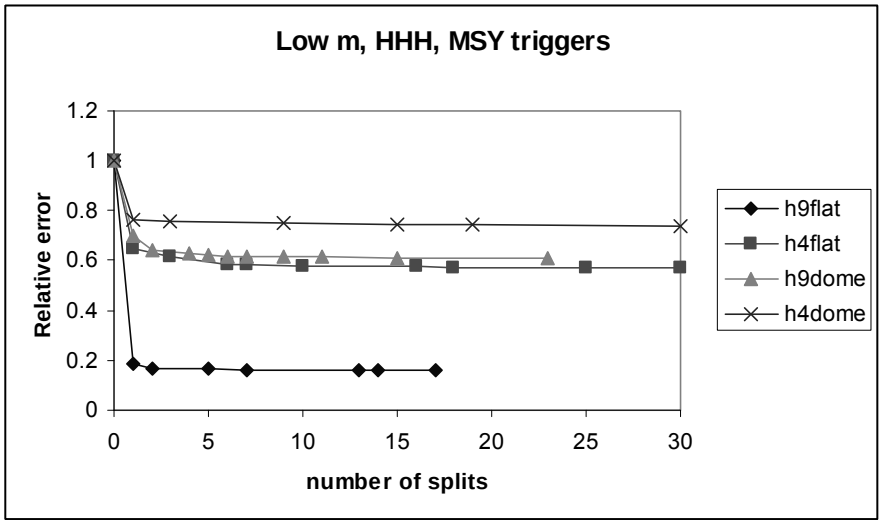


Figure 57. Relative error as a function of the number of splits in a tree for low mortality, high noise and using MSY triggers. Four cases are shown: high and very low steepness (h=0.9 or h=0.4) combined with flat or dome-shaped selectivity (see legend).

A brief comment on the interpretation of the 'tree' plots may be useful. The notation 'mL=a' indicates that if mL=0, the left-limb is followed and if mL=1 the right limb is followed. Since the binary indicators are used as factors in the model, the levels are, by default, referred to as 'a' and 'b' instead of '0' and '1'. (The notation 'L90=b' should be interpreted as: if L90=1, the left limb is followed, and if it is '0' the right limb is followed.)

It is interesting to note that mean length (mL) is at the top of the tree for h=0.9 and flat selectivity. Referring back to Figure 39, we note that mean length has the lowest prediction error for this scenario. It makes sense that the top node of the tree should be based in the indicator which performs best (has the lowest prediction error), and we always found this to be the case. Note, however, that the tree does not then necessarily follow the rest of the sequence of best performing indicators. In fact, sometimes one of the poorer performing indicators, but with potentially low correlation with the other indicators, appear quite high up in the tree. In general, however, the results in Figure 58 confirm the notion that the size-based indicators perform more poorly than CPUE-based indicators when steepness is low, and even more poorly when selectivity is dome-shaped rather than flat.

Table 12 summarises the prediction errors of the best single indicator (i.e. from sub-section 3), and the pruned fitted classification tree. A comparison of the first two rows in this Table shows that there is only a marginal decrease in prediction error between the single best indicator and a classification tree. The largest absolute improvement occurs for the scenario with high h and dome-shaped selectivity (a reduction from 0.31 to 0.27); the relative improvements are all less than 20%.

The rest of Table 12 gives the prediction errors when the 'wrong' (pruned) tree is used with a binary dataset. This is meant to illustrate how badly the prediction error might be affected if we make a wrong assumption about steepness, selectivity, or both. Note that this "wrong" assumption pertains to both the tree and the trigger points. For example, the cell in the row "wrong h" and column "low h, flat" which gives a prediction error of 0.68. This is the prediction error for a tree, and trigger points, based on the assumption that steepness is high (h=0.9) and selectivity is flat, but a simulated dataset which has low steepness (h=0.4) and flat selectivity.

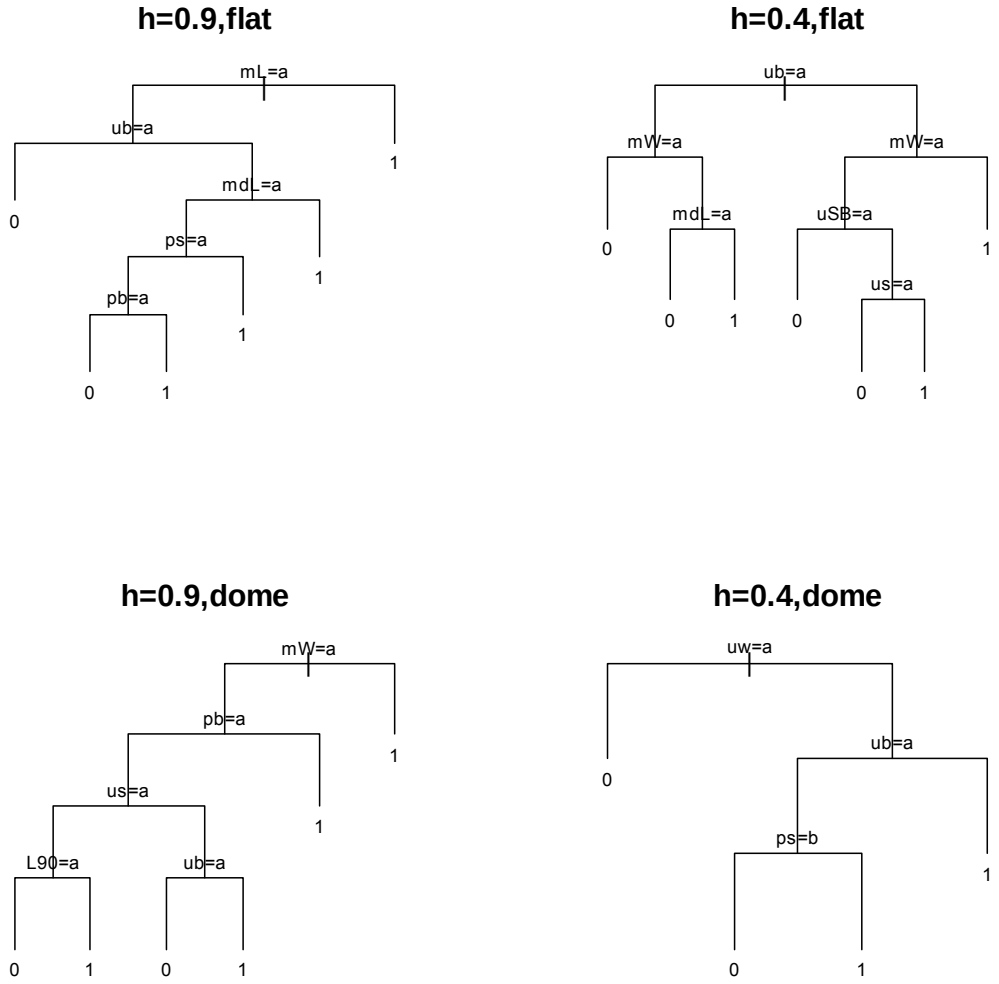


Figure 58. Structures of pruned trees for the low mortality scenario with high noise and using MSY trigger points. The four trees are for high and very low steepness (h=0.9 or h=0.4) combined with either flat or dome-shaped selectivity. See main text (above) for interpretation of notation in trees.

Results in Table 12 show that the underlying assumptions are important. The assumption about steepness is even more important than that about the shape of selectivity. Also, the wrong assumption increases the prediction error by much more when selectivity is (in reality) flat rather than dome-shaped. This is more obvious in Figure 59 where the relative change in prediction error from the single best indicator to the tree (correct or wrong) is plotted for each of the four scenarios.

Table 12 Prediction error for scenarios with low m, high measurement errors (HHH) and for two levels of steepness and two types of selectivity. Datasets are indicated along the top, trees are indicated down the side. Where a prediction is based on a 'wrong' assumption, both the tree and the trigger/reference points for indicators are based on the 'wrong' assumptions.

	high h, flat	low h, flat	high h, dome	low h, dome
best indicator	0.06	0.16	0.31	0.32
correct tree	0.05	0.14	0.27	0.30
wrong h	0.21	0.68	0.40	0.54
wrong selectivity	0.31	0.17	0.29	0.42
wrong h and selectivity	0.19	0.59	0.36	0.57

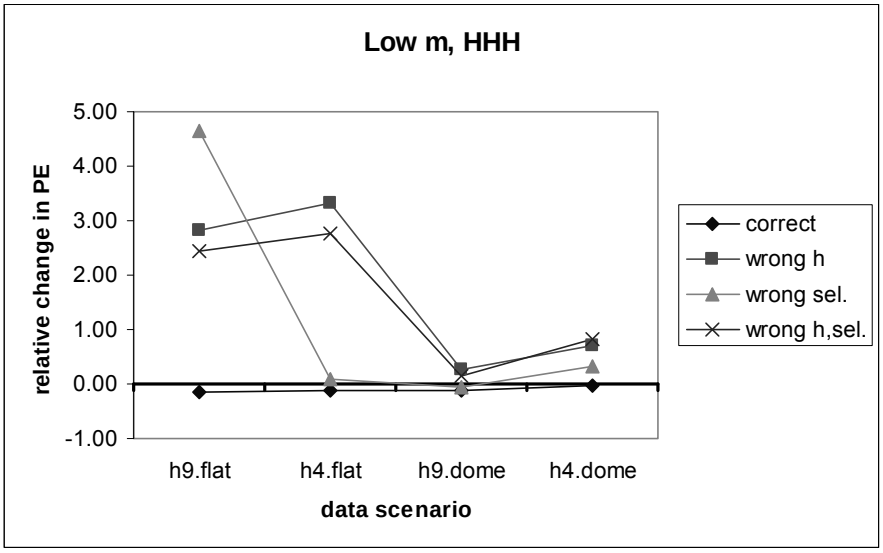


Figure 59. Low mortality, high noise and MSY triggers. The relative change in prediction error between the best single indicator, the correct tree and several 'wrong' trees. See text for more detail.

9.4.1.2. High m, short-lived life history

Very similar patterns emerge when we consider the high m (short-lived) life history. There is again little to gain from a tree with more than about 3 splits (Figure 60). It is worth noting that a tree does not necessarily contain all possible number of splits in its sequence, for example, the tree for h=0.4, flat selectivity goes from about 14 splits to 24 (see Figure 60). This implies that a tree which goes from, say, 1 to 3 to 8 splits cannot be pruned to a 4-split tree.

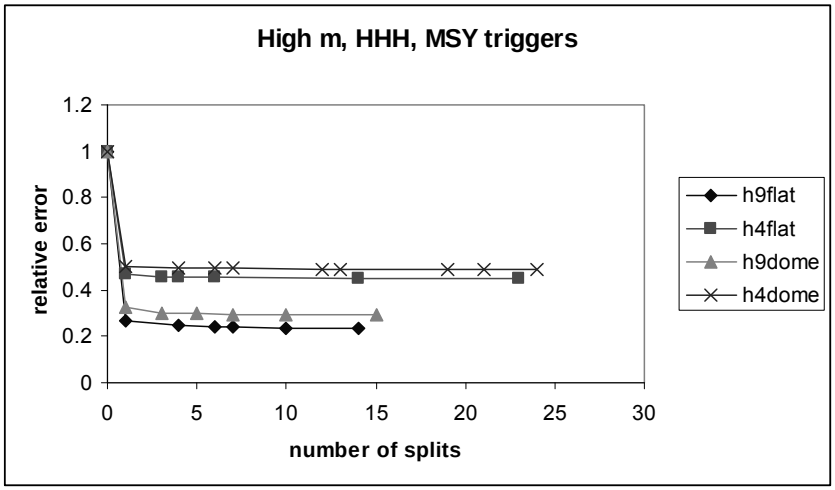


Figure 60. Relative error as a function of the number of splits in a tree for low mortality, high noise and using MSY triggers. Four cases are shown: high and very low steepness (h=0.9 or h=0.4) combined with flat or dome-shaped selectivity (see legend).

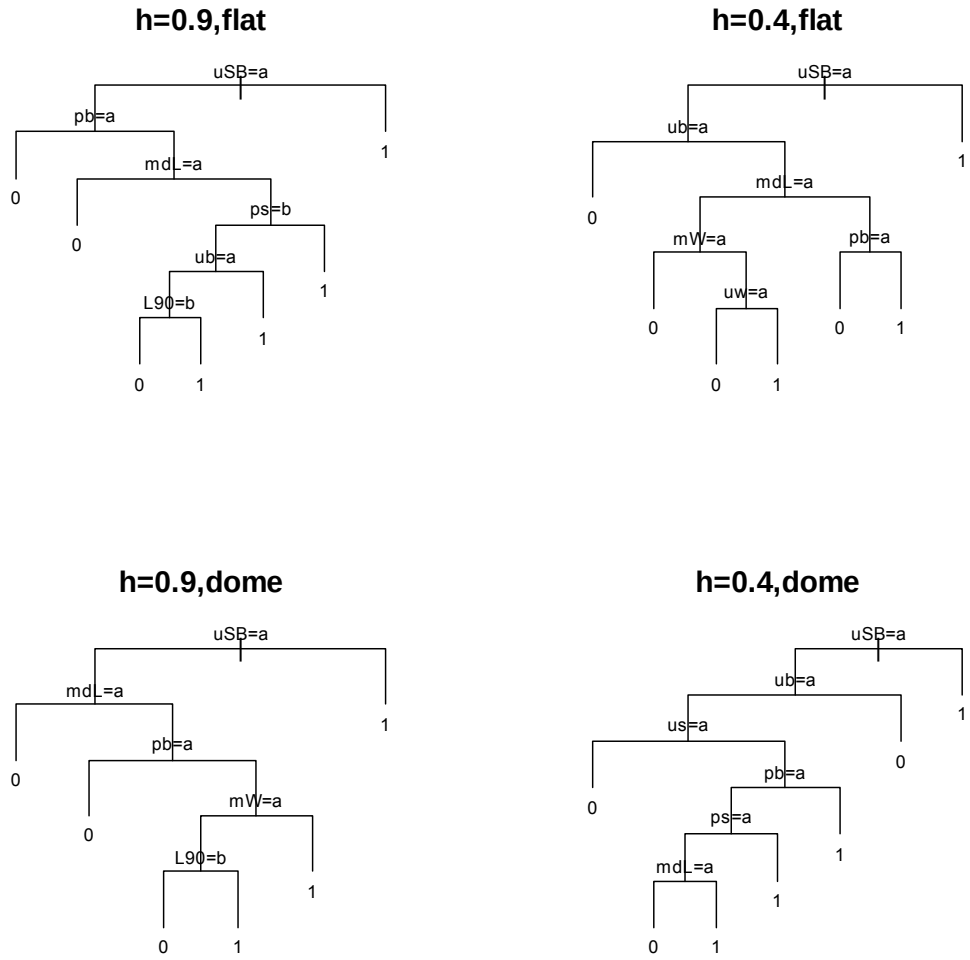


Figure 61. Structures of pruned trees for the high mortality scenario with high noise and using MSY trigger points. The four trees are for high and very low steepness (h=0.9 or h=0.4) combined with either flat or dome-shaped selectivity. See main text (above) for interpretation of notation in trees.

The same set of prediction errors have been calculated for the datasets and trees in this scenario (Table 13). Note we have not crossed the life-history scenarios with one another under the assumption that, in reality, one should in most cases at least be able to have a reasonable idea about the longevity of a harvested species. We consider that a badly wrong assumption about steepness or selectivity is more likely than a badly wrong assumption about life-history (e.g. whether the average lifespan is around 2.5 or around 10 years).

This scenario appears to be less strongly affected by wrong assumptions about steepness or selectivity. This is likely to be partly because the CPUE-based indicators almost always performed better than the size-based indicators (see sub-section 3), and Figure 61 shows that the top node is uSB for all four scenarios.

Table 13. Prediction error for scenarios with high m, high measurement errors (HHH) and for two levels of steepness and two types of selectivity. Datasets are indicated along the top, trees are indicated down the side. Where a prediction is based on a 'wrong' assumption, both the tree and the trigger/reference points for indicators are based on those 'wrong' assumptions.

	high h, flat	low h, flat	high h, dome	low h, dome
best indicator	0.10	0.14	0.14	0.16
correct tree	0.09	0.14	0.13	0.16
wrong h	0.25	0.48	0.27	0.43
wrong selectivity	0.17	0.16	0.11	0.17
wrong h and selectivity	0.24	0.34	0.25	0.49

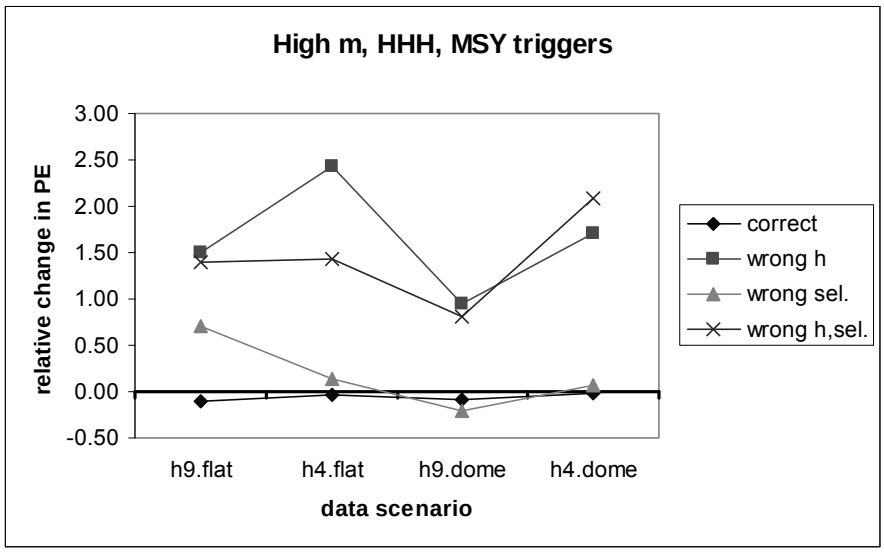


Figure 62. High mortality, high noise and MSY triggers. The relative change in prediction error between the best single indicator, the correct tree and several 'wrong' trees. See text for more detail.

9.4.2 Classification trees based on spawner-per-recruit triggers

In most cases where indicators are likely to be used there may not yet be a stock assessment and it is therefore most unlikely that there would be any knowledge of MSY-based quantities. Even if there is a stock assessment, uncertainty around MSY quantities may be large enough that they are considered risky to use or not meaningful. The next sub-section is based on SPR trigger points and we focus most of the subsequent analyses on SPR-based triggers and reference points. Where relevant or useful, comparisons with MSY-based cases are made.

The relative error as a function of the number of splits in a tree again show that there is little to be gained by having a tree larger than about 2 or 3 splits. This is the case for all four scenarios (high and low steepness crossed with flat and dome-shaped selectivity) and for both life histories (Figure 63 and Figure 64).

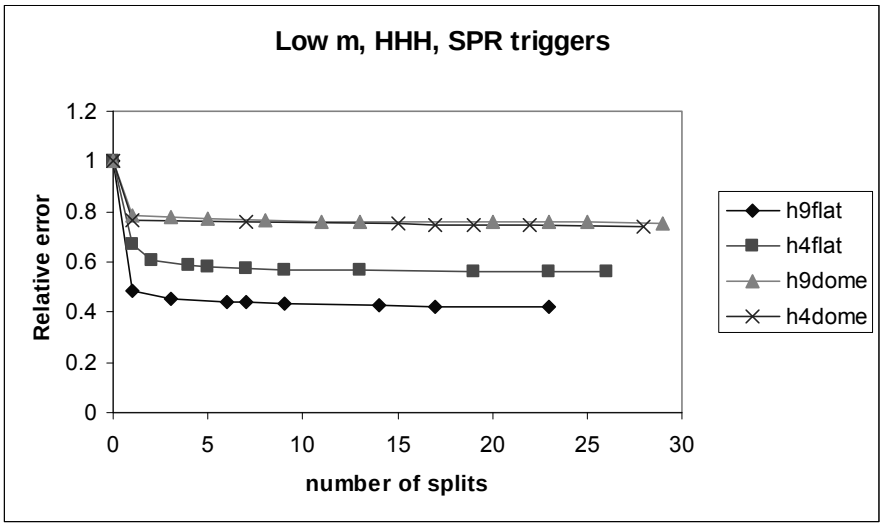


Figure 63. Relative error as a function of the number of splits in a tree for low mortality, high noise and using SPR triggers. Four cases are shown: high and very low steepness (h=0.9 or h=0.4) combined with flat or dome-shaped selectivity (see legend).

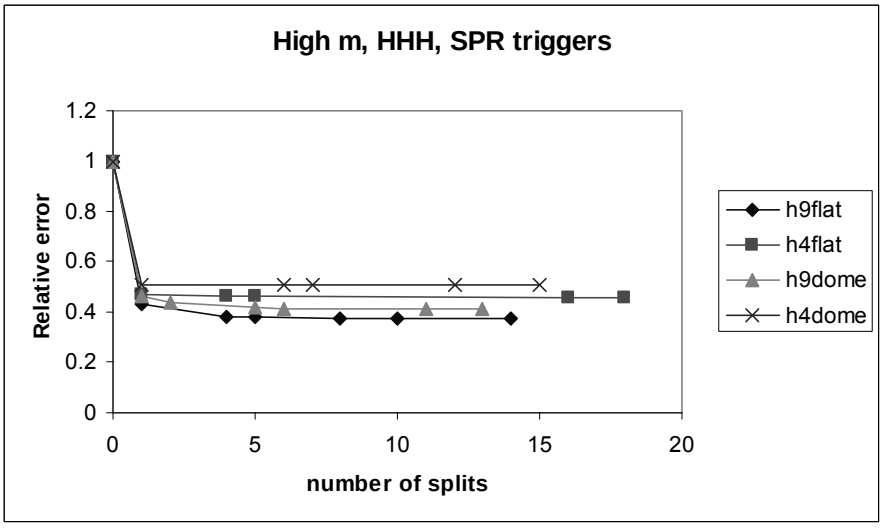


Figure 64. Relative error as a function of the number of splits in a tree for high mortality, high noise and using SPR triggers. Four cases are shown: high and very low steepness (h=0.9 or h=0.4) combined with flat or dome-shaped selectivity (see legend).

Recall that SPR trigger points are independent of steepness, so for the high and low h scenarios, the trigger points are the same. The biomass reference point was chosen to be B/B0=0.4. The pruned trees for the four scenarios of the low m life history are shown in Figure 65. The size-based indicators only form the top node of the trees for high steepness.

The four trees for the short-lived (high m) life history are shown in Figure 66. As for the trees based on MSY triggers, the top node of these four trees are again the indicator uSB i.e. the CPUE of spawners.

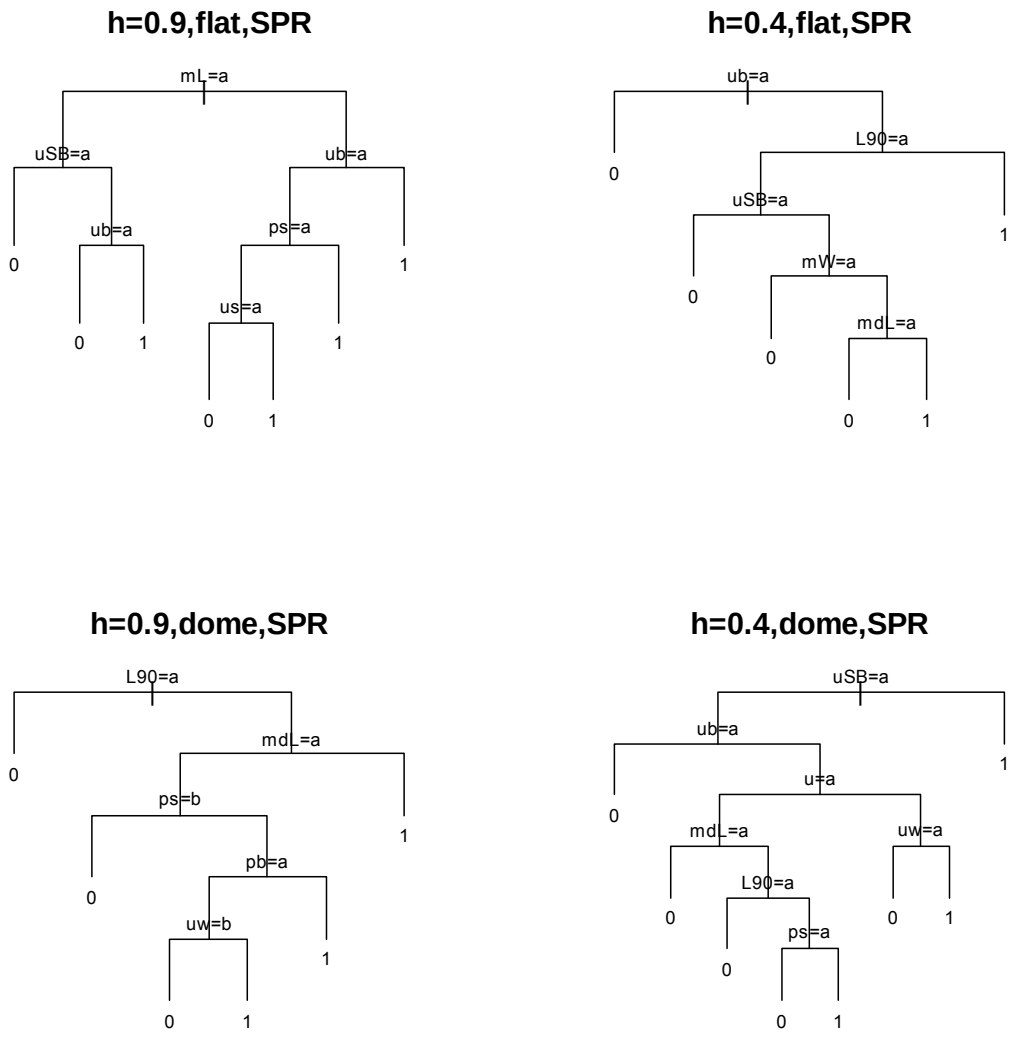


Figure 65. Structures of pruned trees for the low mortality scenario with high noise and using SPR trigger points. The four trees are for high and very low steepness (h=0.9 or h=0.4) combined with either flat or dome-shaped selectivity.

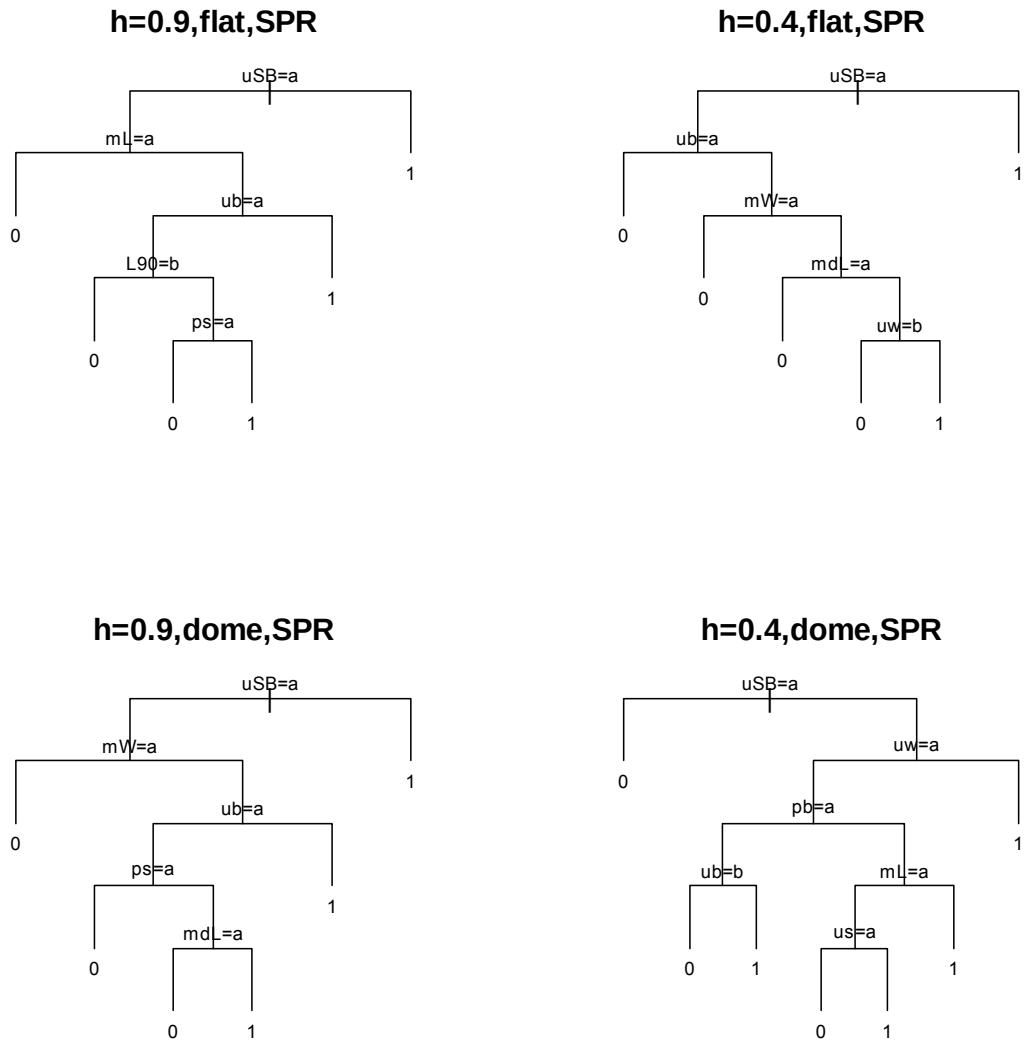


Figure 66. Structures of pruned trees for the high mortality scenario with high noise and using SPR trigger points. The four trees are for high and very low steepness (h=0.9 or h=0.4) combined with either flat or dome-shaped selectivity.

The prediction errors for best single indicator and for the trees again show that there is not a substantial difference between them when using SPR triggers (Table 14).

Table 14. Summary table of 'correct' tree-data matches prediction errors compared with best single indicator under SPR triggers

Tree Data	h09 flat	h04 flat	h09 dome	h04 dome
Low m = inf2				
best single	0.07	0.16	0.15	0.31
correct tree	0.06	0.14	0.15	0.30
High m = inf1				
best single	0.08	0.14	0.09	0.15
correct tree	0.07	0.13	0.08	0.15

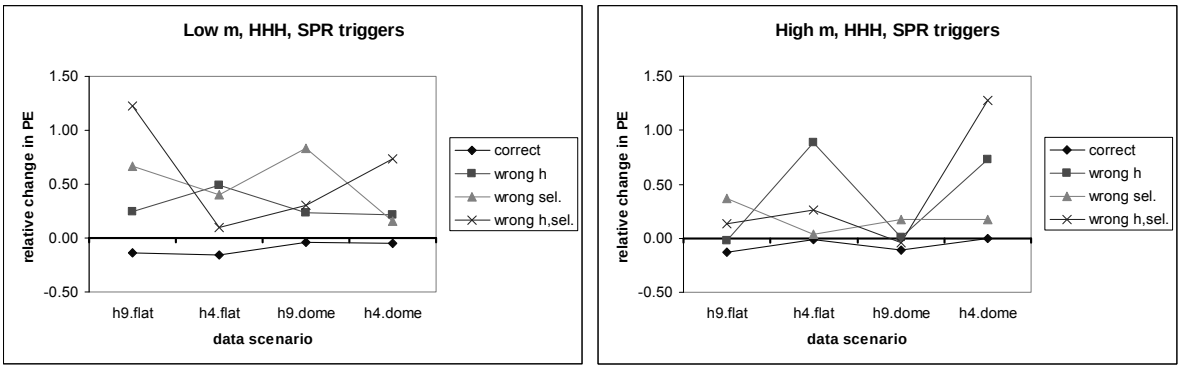


Figure 67. The relative change in prediction error between the best single indicator, the correct tree and several 'wrong' trees (used with 'wrong' triggers). The left panel is for low m, the right panel for high m. In both cases the HHH noise level was used. The x-axis identifies the combination of steepness (high, h9; very low, h4) and selectivity curve (flat; dome). See text for more detail.

Figure 67 shows the prediction errors for the correct tree and for several "wrong" trees (similar to that illustrated in Figure 62 for MSY triggers), and a comparison of these two figures shows that the affect of wrong assumptions is much less strong when SPR triggers are used. It also appears that when steepness is unknown, it is better to assume low steepness when generating data for a tree than to assume high steepness.

9.4.2.1 Surrogate indicators

Diagnostics associated with each fitted tree lists which alternative (surrogate) 'splits' would have been used if the original indicator were not available, or which alternatives give the same or very similar improvement in the relative error. This also illustrates how different indicators can potentially 'substitute' for one another in a classification tree. These diagnostics have been summarised in Table 15 (low m) and Table 16 (high m) which give the split level, the indicator at that split and the first surrogate (in some cases there is no surrogate).

Table 15. Summary of split level, indicator(s) at that split level and the primary surrogate indicator. These results are for trees based on data from the low m life history, high measurement error (HHH) and SPR triggers.

Tree	h9 flat	h4 flat	h9 dome	h4 dome
Split level	1 mL (mW)	1 ub (uSB)	1 L90 (W90)	1 uSB (uw)
	2 uSB (uw)	2 L90 (W90)	2 mdL (mdW)	2 ub
	2 ub (uSB)	3 uSB (uw)	3 ps (mL)	3 u (uw)
	3 ub (uw)	4 mW (mL)	4 pb (mW)	4 mdL (mdW)*
	3 ps (us)	5 mdL (mdW)	5 uw (u)	5 L90 (W90)

* 2nd level 4 split has uw with no surrogate

Table 16. Summary of split level, indicator(s) at that split level and the primary surrogate indicator. These results are for trees based on data from the high m life history, high measurement error (HHH) and SPR triggers.

Tree	h9 flat	h4 flat	h9 dome	h4 dome
Split level	1 uSB (uw)	1 uSB (uw)	1 uSB (uw)	1 uSB (uw)
	2 mL (mW)	2 ub	2 mW (mL)	2 uw (u)
	3 ub	3 mW (mL)	3 ub (uw)	3 pb (ub)
	4 L90 (W90)	4 mdL (mdW)	4 ps (us)	4 ub (L90)
	5 ps (mdL)	5 uw (u)	5 mdL (mdW)	4 mL (mW)

* the next level is 8 splits

These diagnostics on surrogate indicators really just reconfirms what was observed earlier:
a) the CPUE-based indicators tend to be substitutes for one another, particularly uSB, uw and ub; sometimes, but less often, u
b) the size-based indicators are substitutes in the following pairs (mL, mW), (mdL, mdW) and (L90,W90)
c) the means of (length or weight) often occur higher up the tree than the medians or 90th percentiles, but not always
d) pb and ps sometimes appear in the tree at split 3 or below and can be substituted by either CPUE or size based indicators; there is not a consistent pairing or surrogates for these two indicators

9.4.2.2. Effort creep

A very brief exploration of the effect of effort creep on the subsequent tree and its prediction error shows that it is better to assume effort creep in the data than not. If we apply a tree based on the assumption of 'no effort creep' to data where there is effort creep, the prediction error is 0.28 instead of 0.16 when the correct tree is used. On the other hand, when there is no effort creep in the data the prediction errors from the trees based on data with or without effort creep are quite similar (0.15 and 0.14 respectively).

Table 17. Prediction errors for trees based on data with and without effort creep and then applied to data with or without effort creep. The scenario is for low m, low steepness (h=0.4), flat selectivity and SPR triggers.

Tree data =	no effort creep in data	effort creep in data
no creep	0.14	**0.28**
creep	0.15	0.16

9.4.2.3 Tree with lagged indicators

Another very brief investigation was based on the observation in Section 8 that there is autocorrelation in the indicator time series. We created additional explanatory variables by lagging (by 1 time step) the top three indicators, mL, ub and uSB for the scenario with low m, high h, high measurement error and SPR triggers. The tree fitted to this expanded dataset does in fact include some of the lagged indicators near the top of the tree (Figure 68).

Although this is an interesting result and suggests a way in which the autocorrelation could potentially be taken into account, we note that the prediction errors for the two trees are very similar: 0.065 for the tree with no lagged indicators and 0.060 for the tree with lagged indicators.

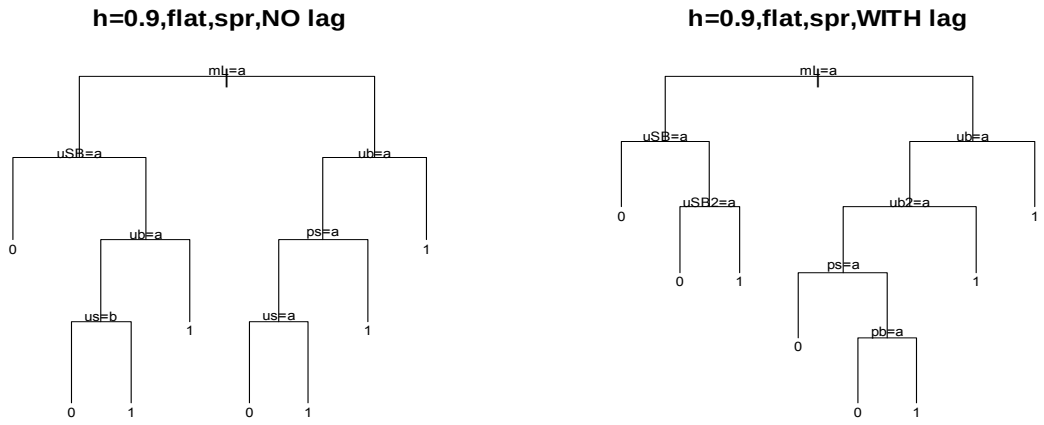


Figure 68. Comparison of pruned trees with no lagged indicators included in the fit (left panel) and with lagged versions of the indicators mL, ub and uSB (namely mL2, ub2 and uSB2; lagged by 1 time step; right panel) included in the fit. The scenario is for low m, high h, flat selectivity and SPR triggers.

9.4.2.4. Combined data for trees

We have emphasised the fact that steepness is unlikely to be known in reality. Although the SPR trigger points do not require an assumption about steepness, the underlying dynamics of the stock will depend on steepness and the performance of an indicator or classification tree will therefore depend on steepness. Here we explore two aspects: (i) prediction errors based on making the wrong assumption about steepness over a wider range of steepness values than done above and (ii) prediction errors of a tree based on simulated data from a much wider range of steepness values. This is again a very brief exploration and we have simply combined the datasets for four steepness ranges considered in sub-section 3 (covering values between 0.35 to 0.95), before fitting a big classification tree. Although there are gaps in this range (values in the intervals (0.75, 0.85) and (0.45, 0.55) are not present), the analysis is informative. We are obviously not proposing this implicit distribution of steepness as realistic, but merely as convenient for illustrating several points.

Results (Table 18) confirm that the worst problem arises when steepness is really low, but is assumed to be very high. In such a case, the prediction error could be as high as 0.25 compared to that from the 'correct' tree which is only 0.138 (see column under h=0.4). On the whole, it is better to base a classification tree (or choice of indicator, for that matter) on the assumption that steepness is low. Such a tree will not perform much more poorly than the 'correct' tree when steepness is really high; see for example the prediction errors under the h=0.9 column, or even the h=0.6 column.

Table 18. Prediction errors for trees and datasets derived from different assumptions about steepness, and for a tree based on the combined data for all steepness levels. Other quantities are: low m, flat selectivity and SPR triggers. As before, h=0.9 implies h~U(0.85, 0.95) and similarly for other h-values.

Data:	h=0.9	h=0.7	h=0.6	h=0.4
Tree based on				
h=0.9	**0.062 (0.08)**	0.088 (0.12)	0.147 (0.20)	0.245 (0.34)
h=0.7	0.075 (0.10)	**0.072 (0.11)**	0.101 (0.14)	0.163 (0.21)
h=0.6	0.085 (0.11)	0.081 (0.14)	**0.092 (0.15)**	0.150 (0.21)
h=0.4	0.090 (0.12)	0.094 (0.14)	0.106 (0.16)	**0.138 (0.21)**
all h levels	*0.075 (0.11)*	*0.079 (0.14)*	*0.095 (0.15)*	*0.142 (0.21)*

The tree based on the combined dataset (all h levels) also performs reasonably well in all cases.

As noted above, the SPR trigger points are not based on an assumption about steepness, so that the prediction errors given here, where we assume that the selectivity is correct, are realistic. The table shows that there are clear advantages in basing a tree on data simulated from a wide range of assumptions about steepness. The above is simply an illustration – one would ideally generate a continuous distribution rather than discrete (non-adjacent) ranges of h as here. The result would be somewhat sensitive to the assumption (e.g. whether it is uniform as here or whether it is centred around some value). The biggest risk is to assume a high steepness because if steepness is much lower, the prediction error would be rather high.

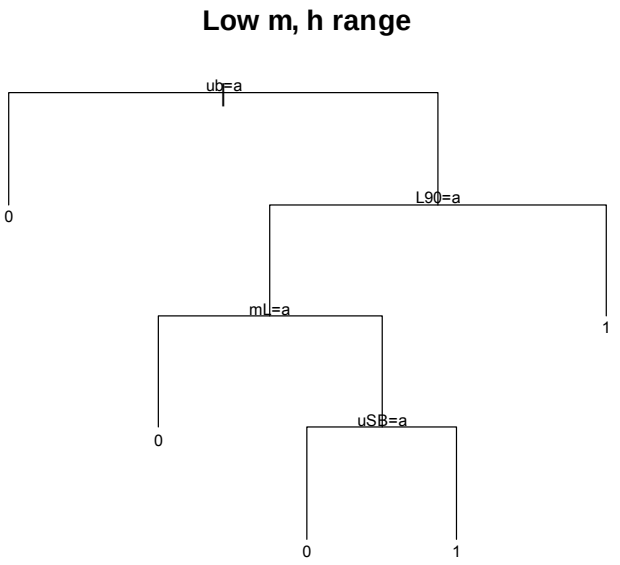


Figure 69. Tree based on data for 4 levels of h combined. The life-history is low m, selectivity flat, and SPR triggers were used.

9.5. Discussion and Conclusions

The magnitude of the prediction error for a given indicator depends on the choice of trigger for that indicator, the harvest rate(s) used in the simulation and the reference point used for spawning biomass. It is therefore important to take care when making comparisons between indicators to ensure that simulations imply a 'fair' comparison.

Using indicators and triggers relative to their unexploited values, lead to lower prediction errors than using absolute values. In the simulations it is assumed that the unexploited (f=0) values are perfectly known. In reality this may be more difficult, though it would still be worth considering how to construct a relative indicator and associated triggers.

The performance of an indicator is also affected by the level of noise in the data. The three main sources considered here are: catch-at-age error which affects CPUE- and size-based indicators; the CV of CPUE which only affects CPUE-based indicators; and sample size for the size frequency samples which affects the size-based indicators, as well as, the combined indicators such as cpueW, cpueB and cpueSB. Note that the catch-at-age error in simulations also reflect changes in selectivity from year to year, but it does not mimic trends or patterns (such as tracking of an abundant age class for example) in selectivity. The relative performance of size- and CPUE-based indicators will depend on the relative level of measurement error in these three quantities. It is, of course, not possible to directly compare error levels in CPUE with error levels in size frequencies.

When steepness is high, the best performing indicators (lowest prediction error) for the scenarios considered here are: ub, mL, mW and uSB. This is the case for the low and high error scenarios, but if CV of CPUE is increased above 40-50% it is very likely that mL and mW would start outperforming ub and uSB. If steepness is low, however, the CPUE-based indicators substantially outperform the size-based indicators. This is likely to be mostly due to the fact that the expected change in the size-based indicators is smaller when steepness is low (see Section 8).

If we assume dome-shaped selectivity in the real system, there is lower correlation between spawning biomass and indicators, and higher absolute prediction errors, than in the case of flat selectivity. Recall that dome-shaped selectivity can imply that a large part of the spawning population is not being sampled by the fishery. It seems that the predication errors are less sensitive to different levels of noise when selectivity is dome-shaped rather than flat. The combination of dome-shaped selectivity and high steepness, leads to the indicators pb, mL, mW and ub performing well. If steepness is low, the CPUE-based indicators (u, ub, uw, uSB) again outperform the size-based indicators.

This is convenient since it suggests that, in general, and in the absence of information on steepness and whether selectivity is likely to be flat or dome-shaped, the indicators ub, uSB (for CPUE) and mL, mW (for size) may well perform best.

Effort creep has the strongest effect on prediction errors of CPUE-based indicators when steepness is low and when selectivity is flat. Of the CPUE-based indicators, ub and uSB are less affected than u , uSB or uw. We have not explored the effects of density dependent changes in growth, but it is clear that this would affect size-based indicators in a similar way to that of effort creep on CPUE-based indicators.

The use of classification trees as a tool for finding combinations of indicators, which perform well as predictors of spawning biomass, proved to be informative and useful. It avoided the need to consider very large numbers of combinations of different numbers of, and different types of indicators. The results suggest that there is little to be gained in terms of lower prediction errors by having groups of more than about 4 or 5 decision nodes; in some cases, these numbers are as low as 2 or 3 decision nodes. This would equate to 5 or fewer indicators (since the same indicator can be used at more than one node). Results also show that, for the scenarios and levels of measurement error considered here, there is in fact only a very marginal benefit in using more than one indicator, compared to using the best performing single indicator. This result, should in our view be treated with some caution. It is likely that the benefit of multiple indicators only really comes into play when series are biased or prone to serious an unpredictable 'failure' (e.g. changes in the fishery operation which is not detected when an indicator is interpreted or standardised).

Prediction errors from trees based on one set of assumptions, and then applied to a simulated dataset obtained from a different set of assumptions provides guidance in the absence of knowledge about steepness and selectivity. Results suggest that it is prudent to assume a relatively low rather than a very high steepness, and to assume dome-shaped rather than flat selectivity. This approach will tend to favour the CPUE-based indicators as the top node, since the size-based indicators only form the top node when the simulated data are based on high steepness.

In Section 10 we use some of the trees fitted here as decision rules. In this context we also revisit the question of whether to base a tree on data generated from a simulation with widely varying F or with slightly varying F at an assumed sustainable level.

9.6. Annex of additional figures to Section 9

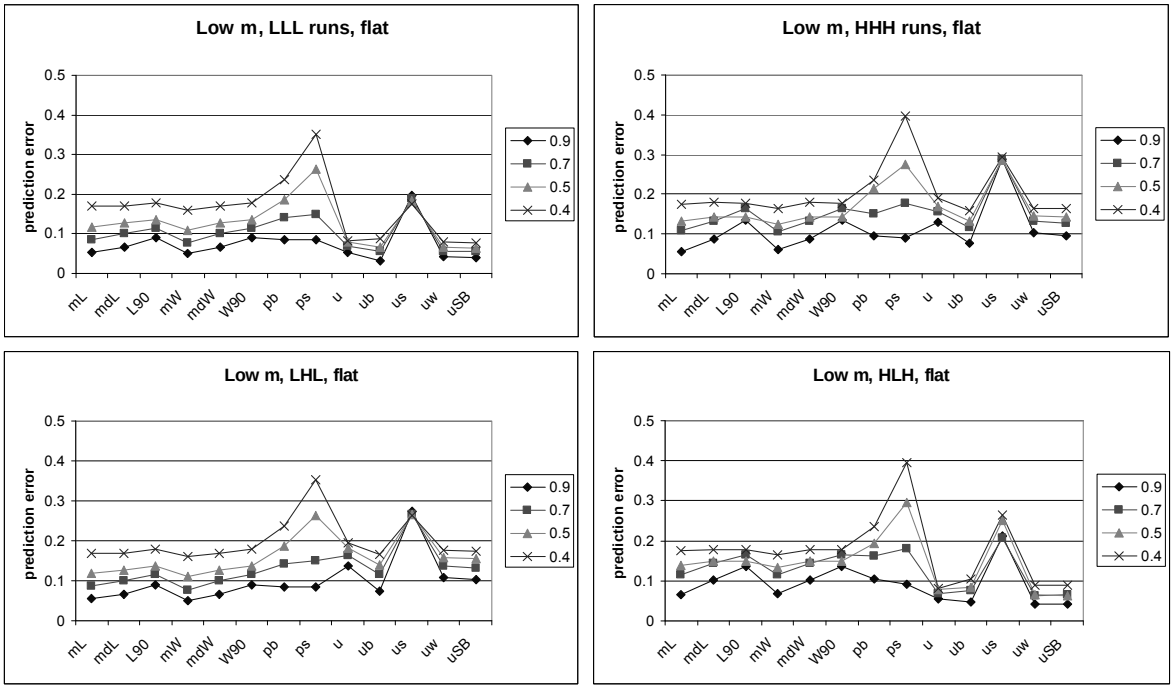


Figure A 14. Prediction errors for a range of steepness levels and for low noise (LLL; left panel) and high noise (HHH; right panel). All cases are for low m, FLAT selectivity and MSY triggers.

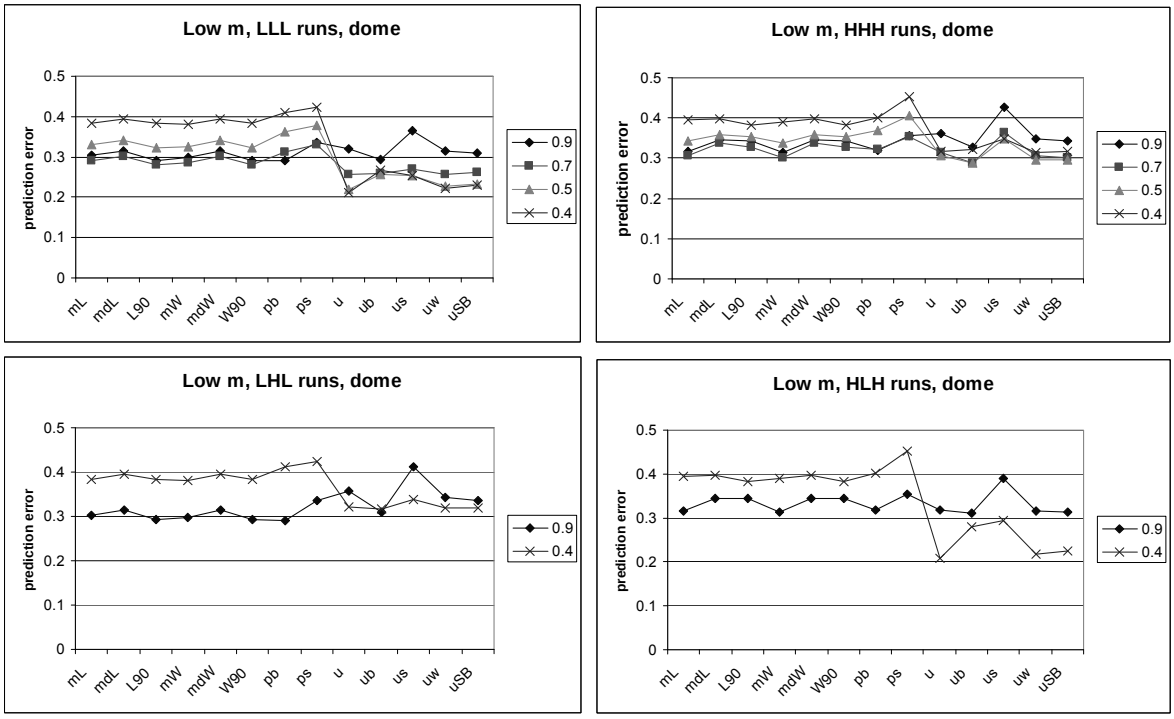


Figure A 15. Prediction errors for a range of steepness levels and for low noise (LLL; left panel) and high noise (HHH; right panel). All cases are for low m, DOME-shaped selectivity and MSY triggers.

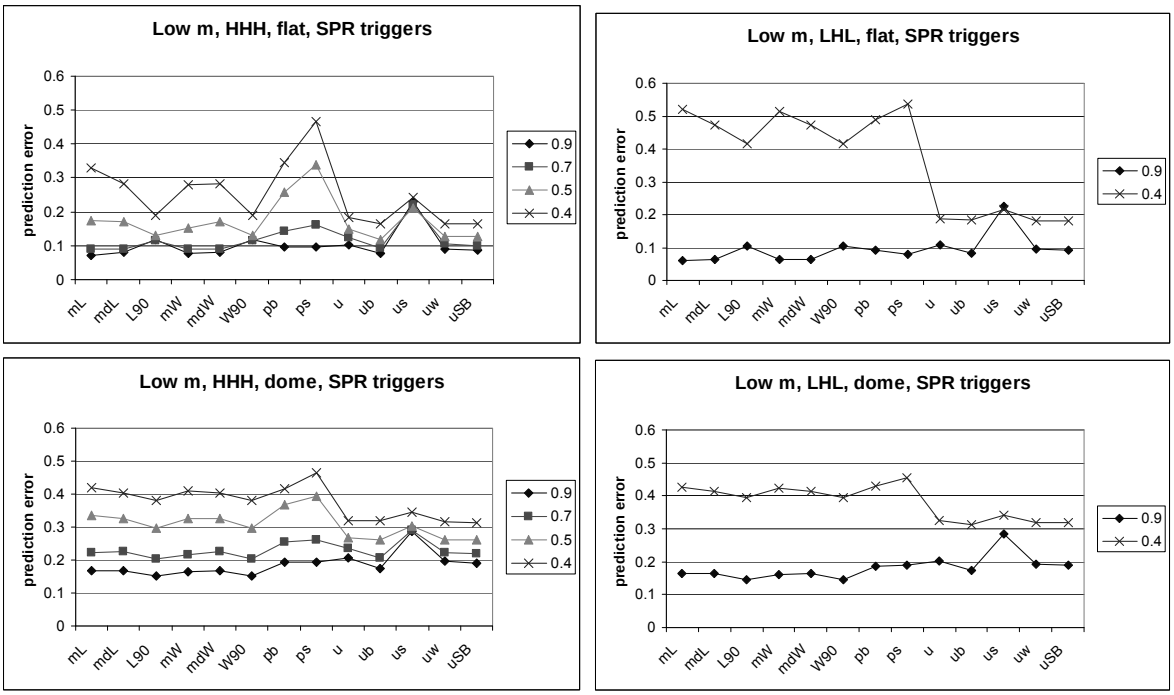


Figure A 16. Prediction errors for the long-lived life history at a range of steepness levels. The top row shows flat selectivity with high noise (HHH, left panel) and high CPUE CV (LHL, right panel). The bottom row shows dome-shaped selectivity with high noise (HHH, left panel) and high CPUE CV (LHL, right panel). SPR triggers based on 40% SPR and a relative biomass reference point 40% were used.

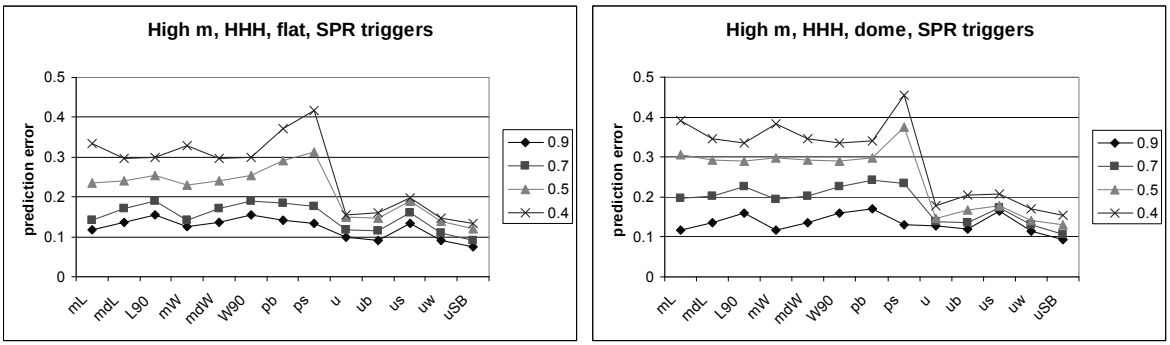


Figure A 17. Prediction errors for the short-lived life history at a range of steepness levels. The top row shows flat selectivity with high noise (HHH, left panel) and high CPUE CV (LHL, right panel). The bottom row shows dome-shaped selectivity with high noise (HHH, left panel) and high CPUE CV (LHL, right panel). SPR triggers based on 40% SPR and a relative biomass reference point 40% were used.

10. An evaluation of the performance of single, and groups of, indicators in feedback decision rules

Summary

The performance of feedback decision rules based on indicators – be they size-based or CPUE-based – is as much a function of the definition and parameters of the decision rule, as it is of the indicator. In many cases, the difference in performance between two types of decision rule based on the same indicator, is less than the difference in performance between the same type of decision rule based on two different indicators.

Decision rules based on just the slope of, or year-to-year change in, an indicator do not perform well, because it does not lead to rebuilding if the stock is already over-exploited, nor to increased exploitation if the stock is lightly exploited when the decision rule is first implemented. If such a rule also has a built in maximum harvest rate or catch, the rule performs better provided that the maximum has not been set too high.

A decision rule based on whether an indicator is above or below some trigger point, and with a built-in maximum harvest rate (or catch) seems to perform well. Similar rules based on groups of indicators either using classification trees or a traffic light approach also seem to perform well. In the case of traffic light type decision rules, the performance of the rule can be particularly sensitive to the number of 'red lights' that trigger action.

The familiar '20-40' type decision rules can easily be adapted to be used with indicators. Care needs to be taken when choosing the limit and target points that define such a rule, particularly when using size-based indicators. The reason is that if they are too close together, the decision rule can oscillate between a closure (or very low harvest rate) and the maximum harvest rate. Not only does this lead to very high interannual variability in the catch and effort, but this can also lead to large oscillations in the population size. The results in this Section suggest that the binary type rules actually perform better from that point of view.

Only harvest rate (or effort) based decision rules were compared in the feedback simulations in this Section. It is well-known, and illustrated in Section 7 using simpler models, that catch and harvest rate strategies can behave quite differently. Although some of the features of the decision rules, seen in these results, will persist under either strategy, other features may not. This could lead to different outcomes for comparisons between the different decision rules when based on total catch. In Section 9, we look at catch-based rules in the context of a spatial model.

It is worth recalling that one of the advantages of effort control is that the catch is likely to vary with the population size even if effort remains constant. This implies that there should, ideally, not be a need for an annual adjustment of effort. Although we have not explored the performance of decision rules invoked less frequently than annually, we consider this to be an important aspect to consider when designing decision rules for implementation in a real fishery, particularly if management is through effort control.

The comparisons between 20-40 and 10-40 versions of the decision rule illustrate an important point in our view. First we noted how the 20-40 rule often flip-flops between F_{max} and F_{min}, particularly since the limit and target points are quite close together (more so for ml than for CPUE). The 10-40 version of the rule reduces the frequency of this extreme flip-flopping, although the frequency of changes (of any magnitude) to the harvest rate is slightly higher.

Nonetheless, the lower choice of limit has only a small effect on the performance in terms of average biomass or minimum biomass. Besides, the value of F_{max} can be reduced slightly to offset this. The main point to emphasise here is that the objectives and reference points that might be desirable for biomass need NOT be directly translated to the trigger points for indicators used in a harvest rule. In this particular case, for example, it is in our view more important to avoid the extreme variability in catches by shifting the trigger points further apart by choosing a lower limit point (one can of course also choose a higher trigger point, and that may mean that Fmax can be slightly increased to achieve the same objectives in terms of average relative biomass.

One of the most important uncertainties that decision rules should ideally be robust to is steepness of the stock-recruit relationship. All the decision rules (binary, classification trees, traffic light and 10-40) respond with a reduction in harvest rate, but the average biomass is still lower when steepness is low. The CPUE-based decision rules, however, outperform both mean length-based rules and rules based on groups of indicators. Rules based on CPUE perform best in the sense that they show the least difference in performance measures such as average biomass and minimum biomass when steepness is very low compared to when it is very high. This is the case for a 10-40, as well as a binary rule based on CPUE.

Even groups of indicators, either in traffic light or classification trees, are out-performed by CPUE as a single indicator. Groups of indicators that performed well under high steepness (because they were obtained from analyses of simulated data based on high steepness) do not perform as well when steepness is low. They still achieve a drop in the harvest rate, but not very much. On the other hand, groups of indicators that were obtained from analyses of simulated data based on LOW steepness perform well when steepness is low, but are overly-conservative when steepness is actually high.

This underlines the importance of finding ways to obtain reliable information on catch, effort, and any additional quantitative or qualitative data which may help standardisation of CPUE. Fishery independent indicators of relative abundance or population density would also be excellent alternatives to strive for.

10.1. Introduction

In this Section we evaluate the performance of size-based and CPUE-based indicators in feedback decision rules. This adds another level of complexity because there are many different ways in which a single indicator can be used in a decision rule. For groups of indicators there are even more possibilities. This Section is therefore by no means an exhaustive investigation. The question we consider here is whether the form of the decision rule is important and, for a given form of decision rule, whether some indicators, or combinations of indicators, perform better than others.

Recall that a decision rule defines how inputs (which can be data, indicator values, or estimates of quantities, for example) are translated into outputs to management a fishery. The output from a decision rule can be a total allowable catch or an effort level, or it can specified as a change in the catch or effort. We only consider decision rules based on harvest rate in this Section.

Given that the whole project is aimed at data poor situations, or at least cases where there are unlikely to be reliable assessments, we limit the decision rules to relatively simple ones. The first set of decision rules simply respond to the trend in an indicator. If the indicator increases, the harvest rate (or catch) is increased, and vice versa. The second set consists of two related groups of decision rules. These rules are essentially an extension of the work presented in Section 9.

Decision rules are constructed from the binary versions of indicators. The action defined by the decision rule depends on whether the indicator is above or below its trigger point. The distinction between the two groups is only relevant where groups of indicators are used. One group is based on the classification trees developed in Section 9. The indicators are input to the classification tree which provides a 'prediction' of whether spawning biomass is above or below its reference point, and the decision tree action depends on this prediction.

The other group of decision rules are based on the so-called 'traffic light' concept (see e.g. Caddy, 2002). In this case the number of indicators that suggest spawning biomass is below its reference point (referred to as 'red lights') are counted and the decision rule action is based on the number (or proportion) of 'red lights'. In the original design of the traffic light approach, three levels were recognised:

a 'green light' for an indicator above its target reference point,
an 'amber light' for an indicator between its target and limit reference point, and
a 'red light' for an indicator below its limit reference point.

The three-level definition clearly needs two sets of reference (or trigger) points specified, and for simplicity and a more direct comparison with the binary-tree decision rules, we have implemented two-level traffic light decision rules.

The third set of decision rules are based on the familiar "20-40" rule concept (see Methods for full explanation and Appendix 1, Figure 1. In some cases this is also referred to as the '40/20' or '40-20' rule. They are identical). In this case there is essentially a predefined relationship between the value of the indicator (relative to a limit and target reference point) and the harvest rate.

The final set of decision rules are similar in concept to that developed in the Harvest Strategy Working Group (HSWG) for the Eastern and the Western tuna and billfish fisheries, and called a 'decision tree'. Note these decision rules are conceptually different from the 'classification trees'. Here we refer to 'decision trees' as hierarchical decision rules to distinguish between 'classification tree' rules. The methods section provides details of the differences.

Only the relatively long-lived, swordfish-like scenario is considered here. The other two quantities that define a scenario are as before, namely the selectivity pattern of the fishery and the measurement errors for indicators. We do not, however, look at all combinations of decision rules and scenarios. In many cases, comparisons of single realisations can illustrate the behaviour of a decision rule sufficiently, and it is often apparent that a change in the scenario is unlikely to change the behaviour of the decision rule.

10.2. Methods

The simulation model used for this set of investigations is essentially the same as that used, and described, in Section 9, with equations given in Appendix 1. It is a discrete time, age-specific, single species, single gender and single area model. There are, however, two important differences. Previously the harvest rate was pre-defined and the simulation was simply run forward with that pre-defined pattern of harvest rates. In this case, the simulation is started with a pre-defined harvest rate, but once the decision rule is activated, the values of the indicators are checked at each time-step and the decision rule determines what the catch or harvest rate in the next time-step should be.

The second difference is the way in which uncertainty has been treated in these simulations. In reality, there is one, unknown, set of actual, or 'true' underlying dynamics. The dynamics are defined by the population parameters such as growth rate, natural mortality, steepness (k,m,h),

and for the fishery it is essentially the 'true' selectivity pattern[21]. The trigger points and parameters in the decision rule would be based on a set of assumptions about these quantities (see Sections 8 and 9), because we do not know the true underlying dynamics. Ideally we want be robust to this uncertainty. We therefore mimic this by running many (50 - 200) realisations with the same underlying 'true' dynamics, but with different sets of trigger points. This is meant to reflect the potential for making wrong assumptions when constructing the decision rule. The question is then whether some indicators and/or decision rules perform better under this uncertainty than others.

10.2.1 Set of scenarios for 'true' dynamics

We only consider the long-lived life-history looked at in previous Sections (Table 7 in Section 9), under both high (h=0.9) and low steepness (h=0.4), to span the range of extremes. Note that in this case the true steepness is the single value 0.9 (or 0.4) for each realisation. The input parameters for the population dynamics aspect of the scenario are, as before:

m	natural mortality
k	von Bertalanffy growth rate
h	steepness in the Beverton-Holt stock-recruit relationship
t_m	age at maturity

All the other inputs are essentially scaling quantities (e.g. L_∞, length-weight relationship, R_0), and are as described in Section 9.

The 'truth' with respect to the fishery dynamics is just the definition of the selectivity pattern which is either flat or dome-shaped. The age-at-first capture (t_c) parameter defines where along the age-axis the selectivity pattern lies. Only one value of t_c is considered for the true dynamics. This leads to 4 scenarios by forming combinations of 1 life-history x 2 steepness values x 2 selectivity patterns.

In some scenarios we consider effort creep by increasing the catchability coefficient by some percentage each year. Details of how this was done are given in Section 2.3 below.

10.2.2 Decision rules

It is simpler to understand how the replicates for each of scenario of 'true dynamics' are constructed once the form of the decision rules has been explained. This is particularly true for decision rules that have built-in trigger points, or ones that have been constructed from simulated data such as the classification trees considered in Section 9. We only consider decision rules formulated in terms of harvest rate (i.e. effort in practice) here, but all the formulations given below can easily be changed to catch. It is worth noting that one of the advantages of effort control is that the catch varies with the population size. This implies that there should, ideally, not be a need for an annual adjustment of effort. As with any simulation study, there are always more and more questions that can be explored. Although we have not explored the performance of decision rules invoked less frequently than annually, we consider this to be an important aspect to consider when designing decision rules for implementation in a real fishery.

Before discussing the actual decision rules, we touch on the issue of 'tuning' which is relevant to all the decision rules considered in this study.

[21] In reality these quantities are also likely to vary from year to year or change over time, but here we assume constant values over the simulation.

10.2.2.1 'Tuning' decision rules

An important characteristic of all decision rules is the notion of 'tuning' the rule to achieve a desired outcome. Imagine two indicators; one which triggers action very frequently (O, for often) and another which does not trigger action as frequently (R, for rarely). A decision rule based on indicator O would lead to many instances of a reduction in the harvest rate; a decision rule based on indicator R would lead to relatively few instances. If the percentage reduction in harvest rate is the same in both cases, the two indicators would lead to different results for the overall or average catch and the relative spawning biomass. Indicator O would imply lower average catches than indicator R. It is, however, possible to specify different levels of reductions for the two indicators. If one applied a smaller % reduction to the rule based on indicator O and a larger % reduction to the rule based on indicator R, the resulting catch and spawning biomass levels would tend to become more similar. It is this potential to change parameters in a decision rule that is referred to as tuning.

Although 'tuning' is important when comparing the performance of decision rules, and particularly for actual use in management, it is time-consuming to do this for a large number of simulation trials. We therefore note the fact that the behaviour of decision rules can be changed through tuning, but in the simulation trials we mostly consider direct comparisons between rules that are constructed in similar ways and with comparable parameter values. Where relevant, we comment on the implications of not having tuned the decision rules. A simple illustration of tuning is given in section 2.2.5 below, under the heading 'Traffic light decision rules for groups of indicators'.

10.2.2.2. 'Up-Down' decision rules

A so-called 'up-down' decision rule simply responds to the change in an indicator. In very general terms, the rule is defined in terms of the slope of the indicator over (say) n years. Let b be the slope of the indicator time-series [i_{t-n}: i_t]. The decision rule then defines the harvest rate in year t+1, F_{t+1} as:

$$\begin{aligned} F_{t+1} &= d.F_t && \text{if } b < 0 \\ &= F_t && \text{if } b = 0 \\ &= (2\text{-}d).F_t && \text{if } b > 0 \end{aligned}$$

where d defines the proportional decrease (or increase), i.e. d<1, and F_t is the harvest rate in year t. We have mostly used an arbitrary value of d=0.9 which implies a 10% cut in effort when the slope is negative. This parameter is clearly one of the potential 'tuning' parameters of the rule. Also note that the proportion increase, when the slope is positive, can be defined to be different from the proportion decrease, in which case the rule will have an asymmetric response.

The number of years over which the slope is calculated can be 2 years or longer, and the best choice would usually be obtained through simulation. This is another potential 'tuning' parameter of the rule. If n=2, then the slope is deterministic, i.e. $(i_t - i_{t-1})$. If n>2 the slope is obtained by linear regression. The implementation we have used also considers whether the slope is significant or not, so that $b<0$ or $b>0$ only when the estimated slope is significantly different from zero.

Note that this rule has no upper limit on the harvest rate. On the other hand, any decrease will get smaller and smaller as F gets smaller, so although the harvest rate can get very close to 0, its

decline will slow down. This may not be a desirable feature, but as results will show, it is really the absence of an upper limit, or any built-in target that makes this a very poor candidate.

10.2.2.3. 'Up-Down with maximum' decision rule

The 'up-down' rule can easily be modified to have a built-in upper limit on the harvest rate. In this case the rule is:

$F_{t+1} = d.F_t$ if $b < 0$
$= F_t$ if $b = 0$
$= \min(F_{max}, (2-d).F_t)$ if $b > 0$

where F_{max} is the upper limit on the harvest rate (or catch, if the rule is specified in terms of catch), and all other parameters are as above for the 'up-down' rule. Note that F_{max} now adds a third potential tuning parameter to the decision rule.

'Binary' rules for single indicators

This set of decision rules are based on the binary value of a single indicator (i), depending on whether the indicator is at or above its trigger point (i=1) or below its trigger points (i=0). Conceptually, the decision rule is simply:

i=1 → no action required, keep harvest rate at its maximum (or target) level, or increase towards its maximum

i=0 → action required, reduce current harvest rate by some percentage

In order to implement this, there are obviously several additional things that need to be specified. The trigger point for the indicator needs to be defined. We only consider trigger points based on spawner per recruit (SPR) considerations, in particular those coinciding with F_{spr40} (i.e. the harvest rate which implies 40% of unexploited SPR; see Section 8 for more detail). The reason for only consider these trigger points is that the stock-recruit relationship and hence steepness (h) is highly likely to be unknown. MSY-based trigger points require an assumption about steepness, whereas SPR-based trigger points do not (Section 8).

In all simulations we have assumed that the indicator value at time step (year) t leads to a decision about harvest rate or catch in the next time step (year), t+1. The extent of reduction has again been set, arbitrarily, at 10% irrespective of the indicator being used. In terms of the notation used above, this rule is:

$F_{t+1} = d.F_t$ if i=0
$= \min(F_{max}, (2-d).F_t)$ if i=1

Under this rule, the harvest rate will continue to be reduced if the indicator remains 0, but will then increase back towards F_{max} if the indicator is 1. It is obvious that this rule could also have been constructed without the upper limit, but as will be shown through the results, this is not a good idea.

Although we have used the notation 'F_{max}' for the upper limit of F in a decision rule, this notation does not mean that F_{max} should be thought of as a limit reference point for F. F_{max} is conceptually closer to a target reference point for the harvest rate. However, as already noted, F_{max} can be one of the tuning parameters to achieve a desired outcome, and may end up being quite different from a value that one might choose *a priori* based on reference point considerations.

The binary rules have also been set up with a slightly different behaviour at the start of implementation of the decision rule, if the current harvest rate is well above F_{max}. In these cases, harvest rate is dropped to F_{max} (in the first year of the decision rule implementation) if the

indicator is 1, or dropped to $d.F_{max}$ if the indicator is 0). This scenario is meant to mimic the situation where there is concern that there is too much effort in the fishery, and a restructuring or buy out of effort leads to an initial and sudden reduction in effort. This sudden drop (rather than a gradual reduction) should not affect the general, and longer term, performance of the decision rules.

10.2.2.4. 'Classification tree' rules for groups of indicators

These decision rules are almost identical to the binary rules for single indicators, but now the 'indicator', *i*, is replaced by the prediction from a group of indicators through a classification tree. The classification trees are those that were obtained in Section 9 by fitting trees to (binary) simulated data for the full set of indicators. Here the trees are used as predictive tools. The simulated indicator data are again turned into binary quantities using the trigger points based on F_{SPR40}, and input to the tree. The prediction, *i*, is a single value – 0 or 1 – and should be interpreted as a prediction of whether spawning biomass is above or below its reference point. Recall that we used an SSB reference point of 40% of unexploited SSB ($B/B_0=0.4$) when constructing the trees (Section 9)[22].

Apart from the change in derivation of the indicator, *i*, the decision rule is identical:

$F_{t+1} = d.F_t$ if $i=0$

$= \min(F_{max}, (2-d).F_t)$ if $i=1$

The initial drop in harvest rate if it is above F_{max} has also been implemented with the classification tree rules.

10.2.2.5. 'Traffic light' decision rules for groups of indicators

So-called 'traffic light' rules evolved when the concepts of limit and target reference points were being developed for fisheries (see e.g. Caddy, 2002). The idea is that an indicator below its limit reference point is like a red light signifying 'stop' (fishing) or 'danger'; an indicator between the limit and target reference points is like an amber light signifying 'proceed with caution' or 'warning' of potential danger; an indicator above its target reference is like a green light signifying 'go' or 'continue'.

This concept is simple for a single indicator, but it was also considered a reasonable approach for combining different indicators by, for example, counting the number of red, amber and green lights for the group of indicators being used. The response is then a function of the number of lights of different colours, and this obviously needs to be defined. There are many possible ways in which the relationship between traffic light 'counts' and a response can be defined.

As indicated above, we have chosen to used the traffic light decision rules as comparisons to the binary classification tree rules. We therefore only use a two-light system with one set of trigger points. So, for indicator i_t, with trigger point $I_{trigger}$, the light, L_{it} is defined as:

$L_{it} = 1$ (green) if $i_t \geq I_{trigger}$

$= 0$ (red) if $i_t < I_{trigger}$

The sum of L_{it} over the set of, say N, indicators is then the number of green lights (n_g) and the number of red lights (n_r) is N minus the number of green lights. Given that we are deliberately

[22] As noted in Section 9, the harvest rate coinciding with SPR/SPR0=0.40 is not necessarily the same as that coinciding with B/B0=0.40, but in the scenarios considered here, they are similar enough that a spawning biomass reference point of 0.4B0 is appropriate.

constructing these rules to be similar to the classification tree rules, we define the response to be a reduction or increase in the harvest rate, with an upper limit on the harvest rate:

$$\begin{aligned} F_{t+1} &= d.F_t && \text{if } n_r \geq R \\ &= \min(F_{max}, (2-d).F_t) && \text{if } n_r < R \end{aligned}$$

where R is the number of red lights that triggers an action, i.e. distinguishes between a reduction or potential increase. In the above formulation, the minimum value of R should be 1 (not 0; and obviously $R \leq N$). In this rule, the choice of $I_{trigger}$ values, the reduction (d) and the number of red lights (R) are all tuning parameters. When comparing these rules with the classification trees, we primarily use R as the tuning parameter. We only use SPR-based trigger points for investigations with these rules.

The traffic light decision rule provides a convenient example for illustrating the concept of tuning, particularly through its 'R' parameter, the number of red lights that triggers action. Figure 70 shows relative spawning biomass and harvest rate for two 'tunings' of a traffic light rule using exactly the same set of indicators. It is unsurprising that a lower value of R leads to a more conservative decision rule which results in higher biomass and lower harvest rate. In this example, the version tuned to R=1, is in clearly unnecessarily conservative since biomass is in fact at or higher than its unexploited level, and the harvest rate has essentially dropped to zero.

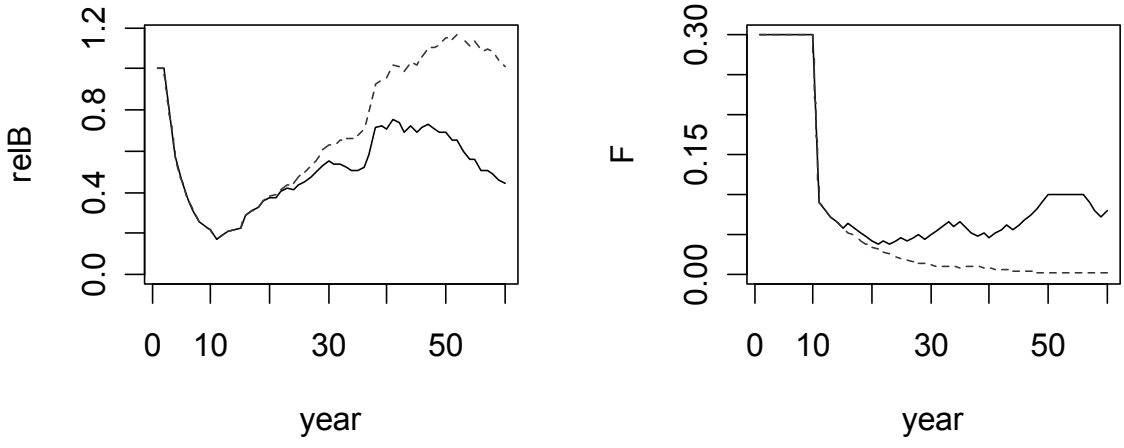


Figure 70. Illustration of a single realisation for two traffic light decision rules based on the same set of indicators (mL, uSB, ub, ps, us), but with either 3 'red lights' (out of 5) triggering a reduction in F (solid line), or 1 'red light' (out of 5) triggering a reduction in F (dashed line). The left panel is for spawning biomass relative to unexploited; the right panel for harvest rate, F.

10.2.2.6. '20-40' type decision rules (same as '20/40', '40-20' or '40/20')

This type of decision rule is based on two trigger points for each indicator. It originated as a decision rule that could be based on an estimate of relative spawning biomass. The basic idea was that if spawning biomass is below 20% of unexploited (considered a limit reference point), the harvest rate, or catch, should be zero (or very close to zero; essentially a fishery closure). If the estimate is above 40% of unexploited (considered a target reference point) then the harvest rate can be set at some target level which is considered sustainable. For estimates of biomass between the limit and target reference, the harvest rate should be set a some level below the

target harvest rate, depending on how far below its reference point the relative spawning biomass is. It is easy to see how one can translate this decision rule into one for use with a CPUE- or size-based indicator. The 20-40 rule is defined as follows for a single indicator *i* at time *t*, i_t:

$$\begin{aligned} F_{t+1} &= F_{max} && \text{if } i_t \geq I_{target} \\ &= F_{max}(i_t - I_{limit})/(I_{target} - I_{limit}) && \text{if } I_{limit} \leq i_t < I_{target} \\ &= 0 && \text{if } i_t < I_{limit} \end{aligned}$$

In this general notation, I_{target} and I_{limit} are the trigger points for the indicator. We again only consider trigger points based on spawner per recruit considerations. In all cases we use the expected value of the indicator coinciding with F_{spr40} as the target[23]. For the limit we consider two examples; one based on F_{spr20}, in which case we refer to the rule as a '20-40' decision rule, and the other bases on F_{spr10}, in which case we refer to the rule as a '10-40' decision rule. The reason for considering both versions is because 20% trigger point can be very close to the 40% trigger point for the size-based indicators (see Section 8). Theoretical considerations (see Section 12) suggest that a 20-40 type decision rule based on triggers that are quite close can lead to strong oscillations in the biomass and catch. The limit and target points are further apart in a '10-40' version of the decision rule.

If the harvest rate is exactly 0, there is no catch and therefore no data to construct the indicators. In the simulations we therefore constrain the harvest rate to a minimum of 0.005 to avoid this situation.

It is in fact possible to further expand this type of decision rule to use multiple indicators. The way we have done this is to calculate the quantity:

$$s_{i,t} = (i_t - I_{limit})/(I_{target} - I_{limit})$$

for each indicator. This essentially reflects how far above the limit the indicator is. In addition

$$\begin{aligned} s_{i,t} &= 1 && \text{if } i_t \geq I_{target} \\ &= 0 && \text{if } i_t < I_{limit} \end{aligned}$$

The $s_{i,t}$ values can then be combined over the chosen set of indicators by, for example, taking the average, the minimum, or a weighted sum, to obtain an overall statistic, say S_t. The decision rule is then:

$$\begin{aligned} F_{t+1} &= F_{max} && \text{if } S_t = 1 \\ &= F_{max}S_t && \text{if } 0 < S_t < 1 \\ &= 0 && \text{if } S_t = 0 \end{aligned}$$

We consider both the average and the minimum statistics, and explore how such decision rules perform compared to one based on a single indicator.

10.2.2.7. Ad-hoc hierarchical rules

It is worth briefly commenting on another type of rule which evolved from the work done within the Harvest strategy working group (HSWG) for the Eastern and the Western tuna and billfish fisheries. First we note that most of the work presented in this Section (i.e. feedback decision rules) had not yet been completed when the HSWG started its work to develop harvest strategies for the fishery (also see Section 5).

[23] Recall that this is the harvest rate which implies a spawner per recruit level of 40% of the unexploited spawner per recruit level.

The HSWG developed a hierarchical rule (called a 'decision tree' in the HSWG documents, but called a hierarchical rule here to avoid confusion with the classification tree rules), based on common-sense considerations, and noting that there may be several aspects of the stock and fishery that one would want to consider when setting a catch or harvest rate using a decision rule. The rationale and details of the Harvest strategy working group (HSWG) decision rule are given in the HSWG documents (Davies *et al.*, 2008). Here we present the concept in very general terms.

The main difference between this rule and the others presented so far is that there is a hierarchy or sequence in which indicators are considered and the response formulated. The idea is to first consider one indicator and determine a harvest rate from that indicator and an associated decision rule. Then several other indicators are considered and the harvest rate is further adjusted (reduced) based on these additional indicators. This approach can be presented in a general form as follows:

At time t, take the predefined set of indicators $I = (i_1, i_2, \ldots i_n)$ (we omit the subscript t for simplicity)
and their associated trigger points $(I_{1.trigger}, I_{2.trigger}, \ldots I_{n.trigger})$
Step 1. Use i_1 and $I_{1.trigger}$ with a decision rule, DR1, to determine a step 1 harvest rate F1

$F1 = DR1 (i_1, I_{1.trigger})$

here an example may be a 20-40 rule to give F1

Step 2. Use i_2 and $I_{2.trigger}$ with a decision rule, DR2, to determine a step 2 harvest rate based on F1

$F2 = DR2(i_2, I_{2.trigger}, F1)$

here an example may be a binary rule e.g. if $i_2 < I_{2.trigger}$ then F2 = d.F1, else F2=F1

Step 3. Use i_2 and $I_{2.trigger}$ with a decision rule, DR3, to determine a step 3 harvest rate based on F2

$F3 = DR3(i_3, I_{3.trigger}, F2)$

....

Step n. Use i_n and $I_{n.trigger}$ with a decision rule, DRn, to determine a step n harvest rate based on F(n-1)

$Fn = DRn(i_n, I_{n.trigger}, F(n-1))$

The management action is then given by the harvest rate from the last step, Fn (or catch, if formulated in terms of catch). The version presented here is a simple linear sequence. It is of course possible to construct a more complicated version where at some step one either goes down one 'branch' with a subset of indicators, or down another, different branch, with another subset of the indicators (see Appendix 1, and Davies *et al.*, 2007).

There are a few features of this decision rule that stand out even without any simulation results to look at. First, as already mentioned, this rule combines indicators in a sequential way; other rules based on groups of indicators, as described above, do not use such a sequential approach, but calculates some overall measure which defines the response. The classification trees can potentially be implemented in a sequential way, by tailoring the response to both the value of the prediction (0 or 1) AND to the node at which the prediction has occurred. A traffic light rule could also be constructed to behave in a similar fashion, if the strength of the response (harvest rate) varies with the number of 'red' lights (though this design would still not really have the sequential characteristic). The sequence of potential reductions in the hierarchical rule means that the more indicators showing 'poor' stock status, the lower the harvest rate will be. This approach does, however, suggest that the magnitude of reductions on F1 would have to be chosen carefully to avoid a frequency of (essentially) fishery 'closures'.

Second, this approach can potentially use a combination of different decision rules (e.g. a 20-40 rule together with a binary rule or the 'down' part of an 'up-down' rule. It is worth noting that although it may be tempting to suggest that there should be scope for both increases and decreases to the step1 harvest rate, this would change the nature of the logic behind the decision rule and should then also affect the construction and parameter values in the rule.

The third point that is worth noting, is that the set of indicators and the sequence in which they are applied can be 'ad hoc' (we include a 'common sense' approach under 'ad hoc'; this does not imply that it is necessarily a bad approach!). There is no associated theory about how to construct such a hierarchical tree, though common sense does in fact lead to strong suggestions. It is interesting to note that this built in common sense logic, which is arguably much easier to understand than, say, the theory behind fitting a classification tree, can make this type of decision rule attractive from a stakeholder point of view.

It is beyond the scope of this study to explore the design of this type of decision rule, but it is likely that there would be some sequences and sets of indicators that would perform better than others. Note that the HSWG decision rule has undergone preliminary simulation testing (Campbell *et al.*, 2007; Davies *et al.*, 2008) and will be further tested under a simulation model that has been parameterised to reflect swordfish dynamics more closely than the version used for preliminary tests (under an AFMA funded project). The rule is also used in some preliminary trials in a spatially explicit model in Section 13.

10.2.3 Generating realisations for each 'truth' and decision rule

First recall that simulations now assume a single 'true' dynamic for the population and for the fishery selectivity in each scenario. It is the details of the trigger points for indicators and parameters for decision rules that vary between realisations in a scenario. This is meant to mimic 'making errors' or 'wrong assumptions' when interpreting the indicators and defining the decision rules.

Each realisation contains 3 sources of error or noise which enter the construction of the inputs to the decision rule and/or the form of the decision rule itself. These are:

1. Assumption about life-history parameters
2. Assumption about selectivity: flat or dome-shaped, and value of t_c
3. Assumption about harvest rate
4. Measurement error and noise in indicators

10.2.4.1. Assumption about life-history parameters

First consider natural mortality, growth rate and age-at-maturity (m, k, t_m) parameters that are used to calculate trigger points and other inputs to the decision rules. For these three parameters we assume that there is sufficient knowledge that the assumed value is likely to be reasonably close to the true value. This is not an unreasonable assumption because even if, say m, is not well known, it is likely that one would have some idea about the average life-span or growth rate: is the species in question like skipjack? or is it more like southern bluefin tuna? Assumed mortality and growth rate values are therefore drawn as correlated quantities from log-normal distributions centred around the true value. The CVs for m and k are 0.02 and the covariance is 0.7. Figure 71 illustrates the values used in one example of 200 realisations.

As in previous analyses, values of age at maturity (t_m) were obtained from the sampled m and k values to give the same proportion alive at that age, as the true values imply. Although this may

seem restrictive, it still leads to a wide range of values (Figure 71) and avoids very unrealistic parameter combinations.

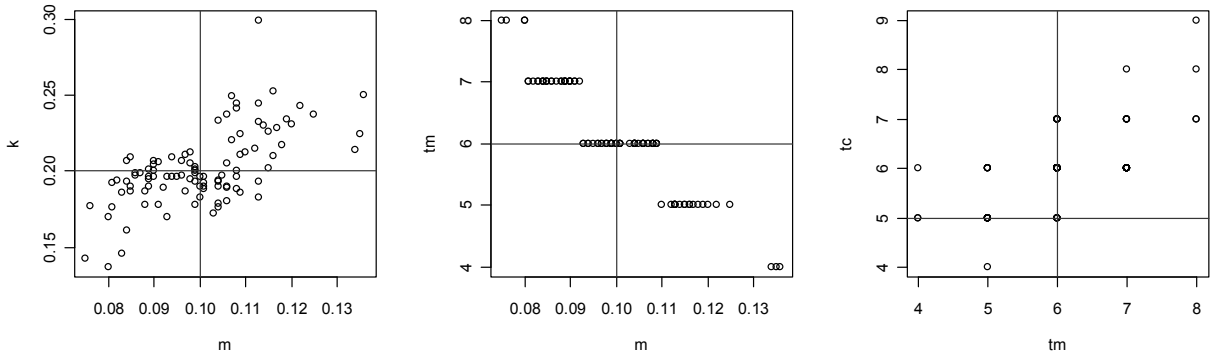


Figure 71. Example of values of m, k, tm and tc used in a set of 200 replicates. The (red) lines show the true values of the parameters.

10.2.4.2. Assumption about selectivity

The selectivity assumption affects the calculation of the trigger points, but would also affect the choice of tree structure if a classification tree is used as a decision rule. For example, if we assume dome-shaped selectivity here, then it would be consistent to use the classification tree structure from Section 9 that was based on data generated with dome-shaped selectivity.

The specification of selectivity also requires an age-at-first-capture (t_c) to position the selectivity curve along the age axis. This parameter is again obtained by taking the true age-at-first-capture (from the scenario specification) and turning it into a length-at-first capture. This length is then used with the sampled growth rate (and other von Bertalanffy parameters) for the realisation to obtain a new age-at-first capture. This is a reasonable assumption, since it is likely that one would have a reasonable idea of the size at first capture, but the associated age may be in error.

10.2.4.3. Measurement error and noise in indicators

We have retained the different noise levels (for the error on catch-at-age, the CV of CPUE and the proportion of the catch sampled for size) that were used in Section 9, but primarily look at the low (LLL) and high (HHH) noise levels. The three parameters that define the noise are again drawn from a uniform distribution with ranges defined by each level (L or H), as specified in Table 1, Section 9.

The simulated catch and size frequency data, used to calculate the indicators, are generated exactly as in Section 9 (see Methods sub-section of Section 9). For completeness, we repeat the list of indicators with their naming conventions here.

Table 19. List of indicators with associated abbreviated names (more detail in Section 8 and Appendix 1)

Description	Abbreviation / name
mean, median and 90th percentile of length	mL, mdL, L90 (or ml, mdl, l90)
mean, median and 90th percentile of weight	mW, mdW, W90 (or mw, mdw, w90)
proportion big, proportion small fish	pb, ps
cpue in numbers	u
cpue of big fish, cpue of small (in numbers)	ub, us
cpue in weight	uw
cpue of spawning fish (in weight)	uSB (or usb)

10.2.4.4. Effort creep

Results so far have not really shown any advantages of groups of indicators over single indicators. This is likely to be because (i) there is no implementation error in the scenarios considered so far, and (ii) no bias in any of the indicators. All indicators are potentially subject to some mechanism which could lead to bias. For example, a density dependent increase in growth rate, if undetected, would tend to lead to a positive bias in size-based indicators compared to their trigger points. So-called 'effort creep', which is essentially an increase in efficiency, leads to a positive bias in CPUE-based indicators unless it too is detected and removed from the time-series through standardisation.

In this section we simulate effort creep in the data, so that all CPUE-based indicators are positively biased, and more so over time. The catchability coefficient is increased by some percentage (we have used a modest figure of 2%) each year. Although the nominal effort is adjusted through the changes in the harvest rate obtained from the decision rule, the implemented (or 'real') effective effort is much higher because of the increase in catchability. This is reflected in the so-called implemented harvest rate (F_{imp}) which is the product of the (increasing) catchability coefficient and the effort. Note that scenarios which include effort creep, essentially imply some form of implementation error as well, given that the simulation assumes effort control rather than catch control.

10.2.4.5 Number of realisations

Since we are primarily interested the general and 'average' behaviour of decision rules, rather than in quantifying risks, we have only run 50 realisations in most cases. This does mean that quantities beyond the inter-quartile range should be interpreted with caution. In some cases (early in our investigations) we ran 200 replicates. Although these provide much better indications of the 'tails' of distributions for performance measures, we confirmed that subsets of 50 realisations are sufficient for means, medians, and quantiles of performance measures.

10.2.4 Performance measures

There are three types of quantities that are of interest when evaluating the performance of a decision rule: spawning biomass, catch and variability in catch. Within each category, it is important to look at short and long term behaviour and to look at average behaviour as well other measures, such as the minimum biomass. It is also informative to look at probabilities (or frequency counts) of, say, spawning biomass being below some level, though we have already noted that some of these quantities should ideally be based on more than 50 realisations. The simulation model generates a rather long list of such measures (
Table 20). In the results section, we mostly look at a subset of about 6 of these (given in **bold** the Table below). The performance measures are only calculated over those years that the

decision rule, or harvest strategy (hs), is implemented, i.e. from year 11 to year 60 in our simulations.

The average relative change in catch is based on the inter-annual relative change, i.e.

$$dC_{t+1} = (C_{t+1} - C_t) / C_t$$

where C_t is the catch in year t (C_{t+1}, the catch in year t+1). Simulations have been set up to mostly avoid zero catches, since none of the decision rules lead to total closures. However, if $C_t=0$, then dC_{t+1} is excluded from the calculation of the average relative change. This statistic is often calculated as the average absolute relative change, by taking the absolute value of the difference in catch. We, however prefer to retain the sign of the change here.

Table 20. List of performance measures generated by the simulation model for each scenario (of 'n' realisations). 'over hs period' means over the period that the decision rule was implemented – years 11 to 60.

	Short term statistics X = 10,20,30,40,50	Long term statistics atend=year 60, hs=years 11-60
BIOMASS		
B/Bmsy	every 10 years (B.Bmsy.atX)	in the last year (B.Bmsy.atend)
B/B0	every 10 years (B.B0.atX)	in the last year (B.B0.atend)
average B/B0	blocks of 10 years (B.B0.MeanX)	over hs period (**B.B0.Mean.hs**)
minimum B/B0		over hs period (**minB.B0**)
proportion of time B<0.2B0		over hs period (bbelow2.hs)
proportion of time B<0.3B0		over hs period (bbelow3.hs)
HARVEST RATE		
mean harvest rate		over hs period (**meanF.hs**)
proportion of time harvest rate was reduced		over hs period (**% times F reduced**)
proportion of time harvest rate was at its minimum ('closure' in 20-40 or 10-40 rules)		over hs period (minFp.hs)
CATCH		
C/Cmsy	every 10 years (C.Cmsy.atX)	in the last year (C.Cmsy.atend)
average C/Cmsy	blocks of 10 years (C.Cmsy.MeanX)	over hs period (**C.Cmsy.Mean.hs**)
average relative change in catch (inter-annual)		over hs period (**meanDelC**)

We summarise catch statistics relative to C_{msy}, only because the absolute values are arbitrary (and hence meaningless), and not because we attach any particular value to C_{msy} as a reference. The absolute values are meaningless, because the model has not been conditioned to historic data for a specific stock. For example, the level of recruitment is entirely arbitrary. In some cases (e.g. time-series plots of results) the catch is shown relative to the catch in the first year rather than C_{msy}.

10.3. Results: Decision rule performance for a long-lived life-history and high steepness

The first set of comparisons and explorations are done with one scenario, namely the long-lived life-history with high steepness and flat selectivity. High measurement errors are used on all

components, i.e. the catch-at-age sampling error, the CV of CPUE and the size-frequency sampling error (see Section 9). It is potentially confusing when decision rules and scenarios are changed at the same time! This scenario was chosen since it is the one where the indicators have the largest expected response and relatively low prediction errors (see Sections 8 and 9). Any decision rules that perform badly under this scenario are unlikely to perform much better under the other scenarios. In subsequent sections, a subset of decision rules are run under some of the other scenarios.

Note that, although we plot results for a constant F strategy, this should only be seen as illustrative and for comparison. The performance of this approach is only acceptable/good because the harvest rate has been set at an appropriate level. In reality the appropriate level may not be known, and a constant harvest strategy with F set too high would be a very bad approach.

10.3.1. 'Up-down' decision rules

First we look at a decision rule which increases or decreases the harvest rate by a pre-defined amount (always the same proportion), depending on whether the indicator is increasing or decreasing. For simplicity we illustrate behaviour by looking at single realisations at a time. These scenarios start with a harvest rate that is 3 times F_{spr40}, for the first 10 years. The decision rule is implemented in year 11.

Figure 72 shows that there is little difference in the overall biomass and catch for a decision rule based on mean length or on CPUE where both use two years to calculate the slope of the indicator. The details of the adjustments made to the harvest rate differs (middle panel in Figure 72), but on average the harvest rate remains at the level it was at before the decision rule was invoked, and there is no rebuilding of the spawning biomass. A scenario with constant harvest rate at F_{spr40} (=0.1) is shown for comparison. Although only one realisation is shown here, the patterns are very similar for different realisations (see below). The realisations are directly comparable because the recruitment deviations and other random numbers associated with the realisation are the same[24].

Changing the period over which the slope is calculated from 2 years to five years does not change this outcome (Figure 73). The harvest rate still remains too high and the spawning biomass does not rebuild. If the extent of the adjustment (d) was changed, to say a 20% reduction or increase, it would only lead to wider fluctuations in F, and consequently in Catch and SSB, but it would still not lead to rebuilding, because the fluctuations would still tend to be around the original level of $3.F_{spr40}$.

[24] Random numbers associated with sampling the size frequency would, however, be different because the catch and hence size frequency sample size would be different.

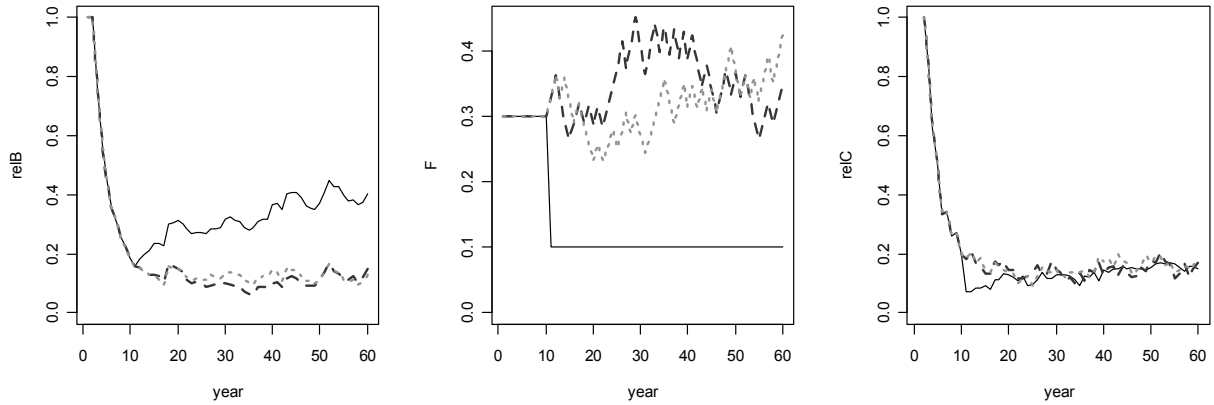


Figure 72. Results of a single realisation of the 'Up-down' decision rule based on 2 years' indicator values to calculate the slope, and an adjustment factor d=0.9; mean length as indicator (dashed line) and CPUE as indicator (grey dotted line). A constant harvest rate at Fspr40 from year 11 onward is shown for comparison (solid line). The left panel shows spawning biomass relative to unexploited (relB); middle panel shows harvest rate (F); right panel shows catch relative to that in the first year (relC).

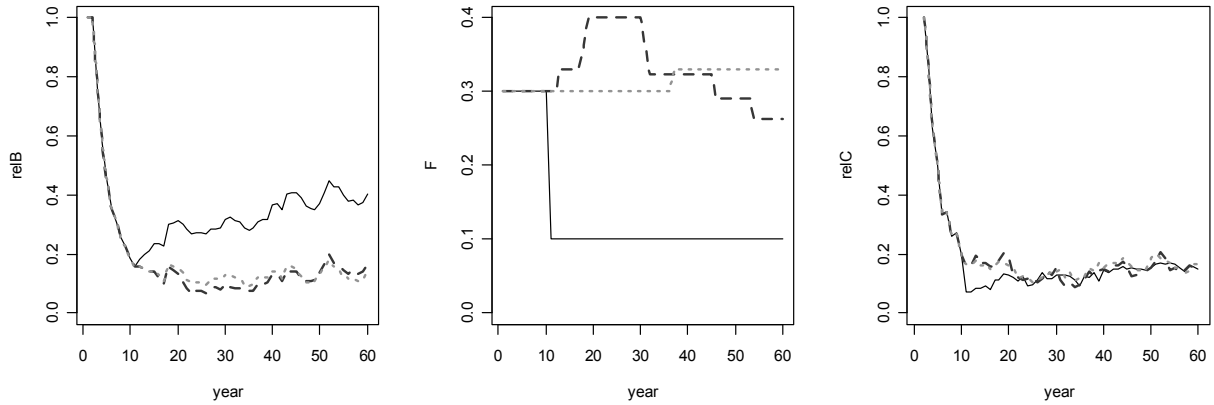


Figure 73. Results of a single realisation of the 'Up-down' decision rule based on 5 years' indicator values to calculate the slope, and an adjustment factor d=0.9; mean length as indicator (dashed line) and CPUE as indicator (grey dotted line). A constant harvest rate at Fspr40 from year 11 onward is shown for comparison (solid line). The left panel shows spawning biomass relative to unexploited (relB); middle panel shows harvest rate (F); right panel shows catch relative to that in the first year (relC).

It is interesting to note that this result is compatible with the theoretical results obtained with a simple stock-production model and an 'up-down' decision rule which only responds to the slope in an indicator (Section 11).

10.3.2. 'Up-down with maximum' decision rule.

The only difference between this rule and the previous one is that the harvest rate is now constrained to a maximum. This does, however, mean that another parameter (F_{max}) needs to be specified, and it is highly likely that there would be limited information on which to base the choice for F_{max}. In the first example, we set the maximum at the pre-decision rule level of $3.F_{spr40}$. The top set of panels in Figure 74, representing one realisation, now shows some recovery in biomass as a consequence of a drop in the harvest rate, for the rule with an upper limit on F (red, dashed line). The lower set of panels, representing a different realisation, also shows that the

harvest rate is now generally lower, but it does not show much of a recovery in the biomass. Summary results from a larger number of realisations are considered below.

If the maximum harvest rate is set at a much lower level (e.g. at F_{spr40}=0.1), there is a substantial increase in biomass as one would hope (Figure 75). Although the average catch is quite similar for the rule with low F_{max} and that without any limit (particularly towards the end of the time-period), the implied catch rates would be much higher for the rule with low F_{max} because the biomass is much higher.

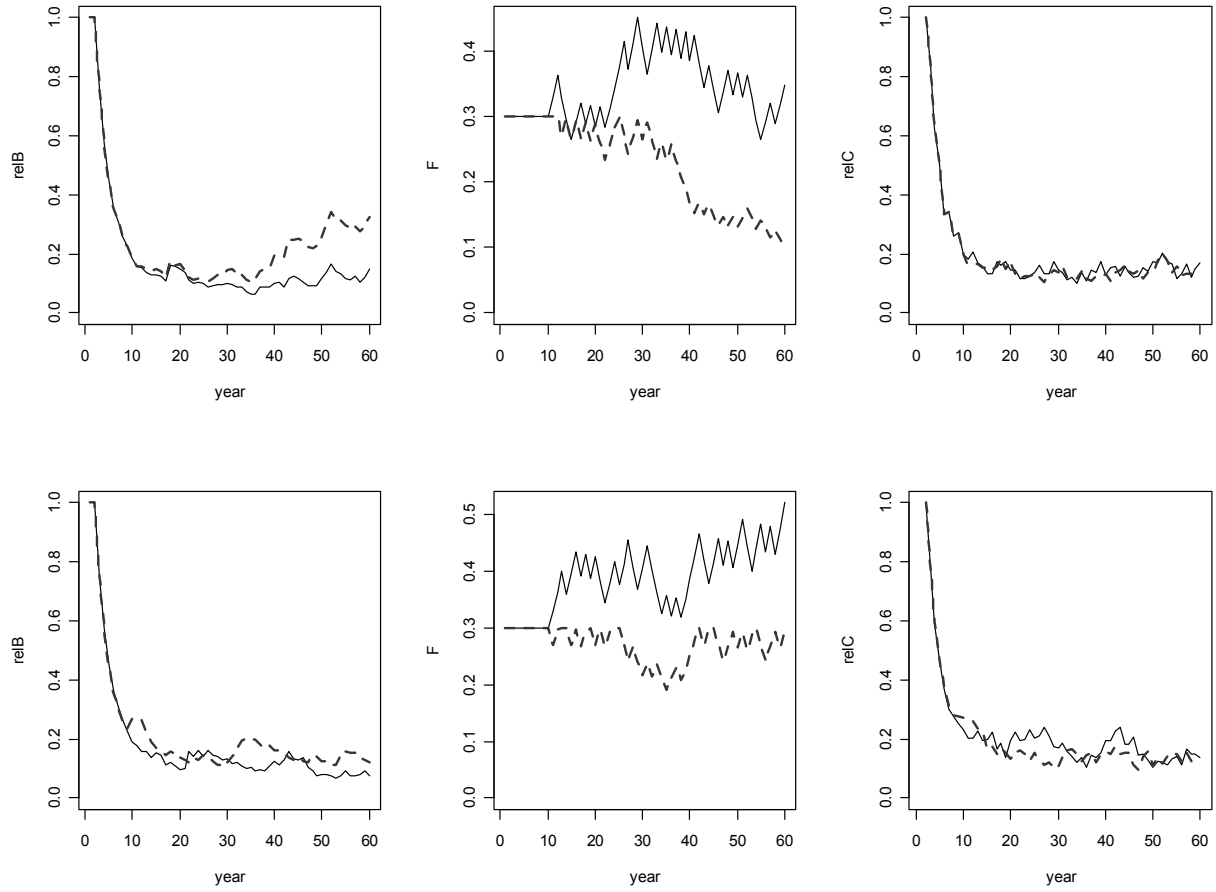


Figure 74. Results of two sets of realisation for the 'Up-down' decision rule without upper limit on F (solid line) and with an upper limit Fmax=3.Fspr40 on F (dashed line). In both cases the indicator is CPUE, the slope is based on 2 years and d=0.9. The top set of panels is for one pair of realisations, the lower set of panels for a second pair of realisations. The left panel shows spawning biomass relative to unexploited (relB); middle panel shows harvest rate (F); right panel shows catch relative to that in the first year (relC).

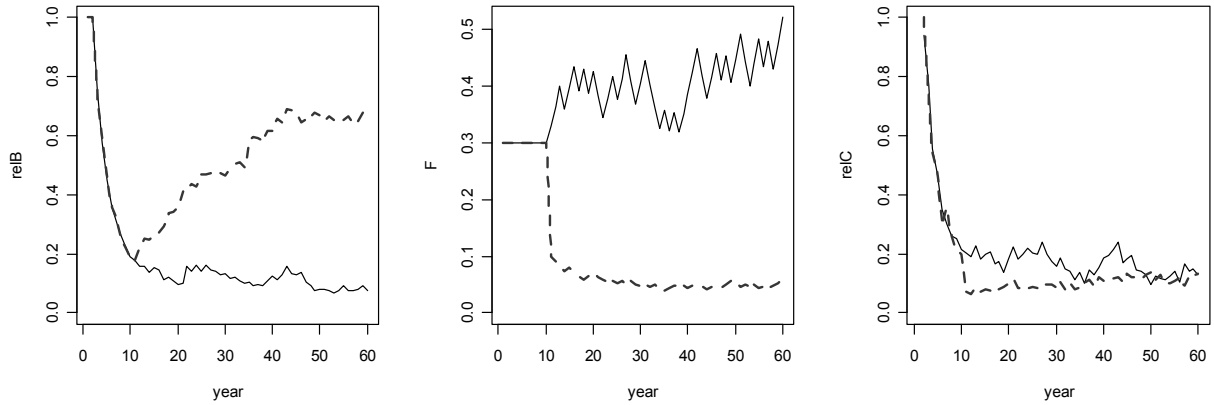


Figure 75. Results of one set of realisation for the 'Up-down' decision rule without an upper limit on F (solid line) and with Fmax=Fspr40 (dashed line). In both cases the indicator is CPUE, the slope is based on 2 years and d=0.9. The left panel shows spawning biomass relative to unexploited (relB); middle panel shows harvest rate (F); right panel shows catch relative to that in the first year (relC).

Results of 50 realisations confirm the overall patterns noted above. Figure 76 shows results in the form of boxplots for average values relating to biomass, harvest rate and catch over the period that the decision rule is implemented. (Recall that boxplots show the median and quantiles as horizontal lines, the whiskers are 1.5 times the interquartile range and any values beyond that are shown as open circles). Note that, although we plot results for constant F, this should only be seen as illustrative and for comparison. The performance of a constant F approach is only acceptable/good here because the harvest rate has been set at an appropriate level. In reality this level may not be known, and a constant harvest strategy with $F=3.F_{spr40}$, would be a very bad approach!

For these decision rules, there is much less difference between the two indicators (here mean length and CPUE) than between different decision rules. When there is no target built into the decision rule, the harvest rate and biomass tend to remain (on average) at pre-decision rule levels. If a stock is over-exploited at the start of such a decision rule being implemented, it will tend to remain over-exploited. On the other hand, if the stock is under-exploited at the start, it will tend to remain under-exploited under such an 'up-down' decision rule.

The situation is much better if there is a built-in target in the form of a maximum harvest rate (Figure 76). Even if the maximum harvest rate in the up-down rule is set too high, it is better than if there is no limit. Note for example that the lowest levels of average biomass for the 'ml' and 'u' scenarios in Figure 76 (i.e. mean length and CPUE as indicators; no limit on F) are lower than those for the mlx3 and ux3 scenarios (i.e. mean length and CPUE as indicators; $F_{max}=3.F_{spr40}$). This is also borne out in the minimum relative biomass (Figure 77), and implies that the risk of low spawning biomass is reduced when an upper limit on the harvest rate is included in the decision rule.

It is unsurprising that the lower value for F_{max} leads to the highest spawning biomass, lowest harvest rate, and lowest average catch. The differences between the two indicators are relatively subtle. For example, the average relative change in catch is slightly higher (and more often positive) for CPUE than for mean length as indicator.

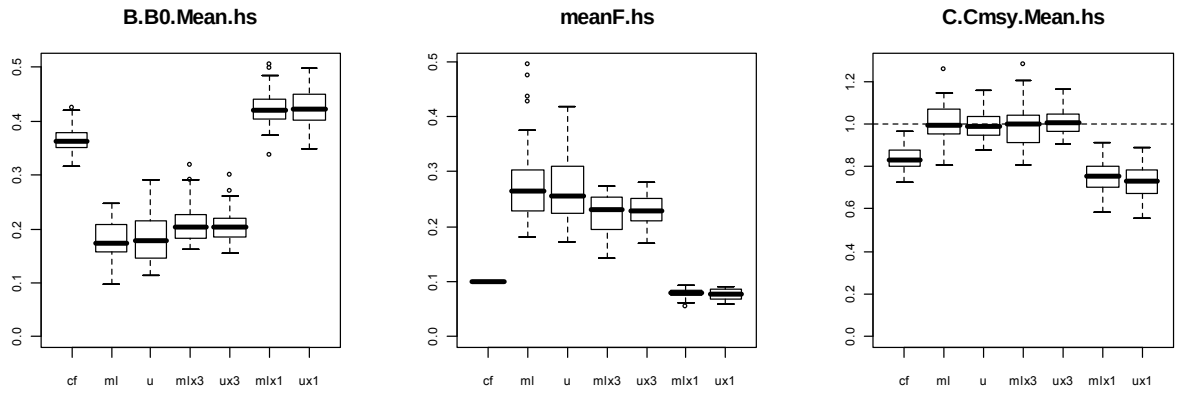


Figure 76. Boxplots of summary results of 50 realisations for 7 scenarios using the up-down decision rule without, or with a maximum on the harvest rate. The scenario is identified on the horizontal axis: cf=constant F=0.1; 'ml' refers to mean length as indicator, 'u' to CPUE as indicator. Further extensions to the names are: 'x3' for Fmax=3.Fspr40, and 'x1' for Fmax=Fspr40. The panels show the average spawning biomass relative to unexploited, the average F and the average catch relative to Cmsy; in all cases averaged over the period that the decision rule (or hs=harvest strategy) is applied (i.e. the first 10 years are omitted from calculations).

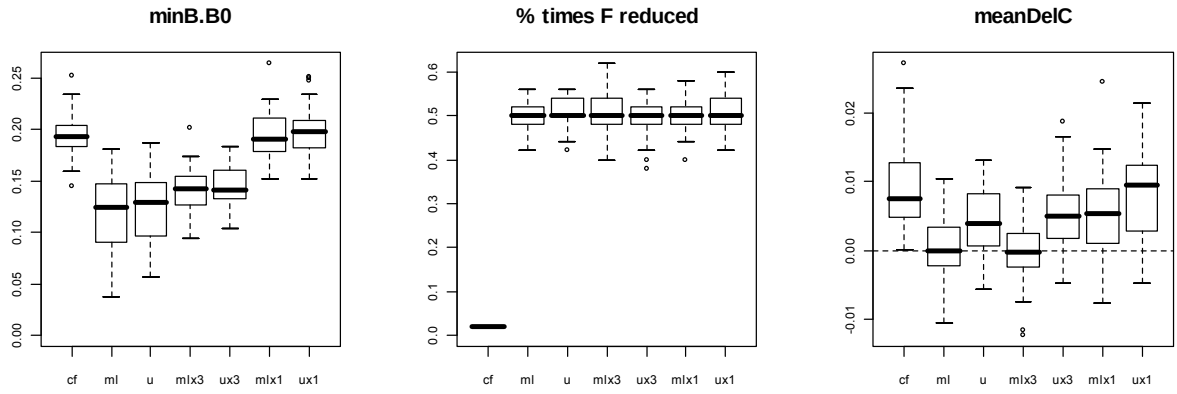


Figure 77. Boxplots of summary results for the same 7 scenarios shown in Figure 76. The panels show the minimum relative spawning biomass, the percentage time that F was reduced and the average relative change in catch; in all cases over the period that the decision rule is applied.

The features of the 'up-down' and 'up-down with maximum' decision rule that are illustrated through comparisons based on versions with mean length or CPUE, are very likely to persist if other indicators were to be used in these decision rules and with this scenario.

10.3.3. Results for binary decision rules and classification trees

Recall that the classification trees which we fitted to simulated data in Section 9 can be used as decision rules by taking the indicators (their binary values relative to their trigger points) as inputs and predicting a '0' or '1', i.e. whether spawning biomass is below or above its reference point. This means that a tree fitted to simulated data from whichever scenario, can be used with any other scenario. Here we compare two single-indicator binary rules, one based on mean length (ml) and another based on CPUE in weight (uw), and two classification tree rules. The first classification tree (h9.flat) was fitted to simulated data from the same scenario being used here.

The second classification tree (h4.flat) was fitted to simulated data which assumed low steepness (h=0.4), and therefore has a different structure. The two trees are shown in Figure 78.

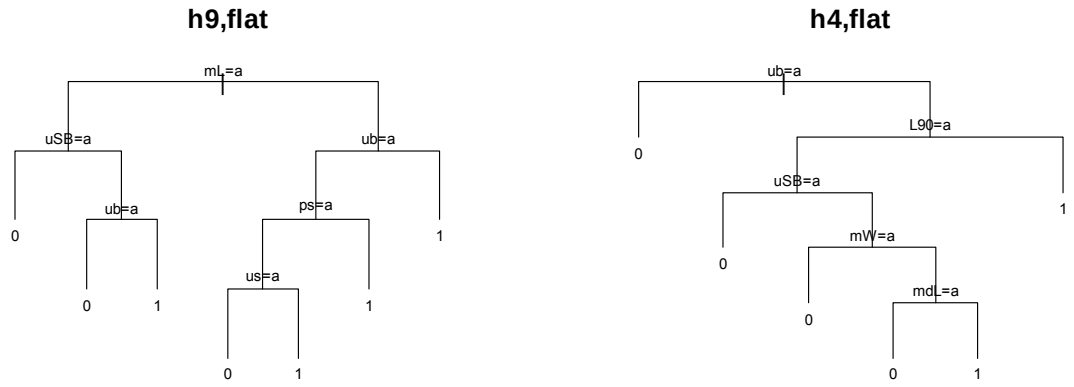


Figure 78. Classification trees fitted to simulated data based on high steepness (h9,flat) and on low steepness (h4,flat), for the long-lived life history with flat selectivity, high measurement error levels and SPR40 reference points. Repeated from Section 9; see text for more detail.

Time series plots of one realisation for the four decision rules show that they all behave reasonably similarly, particularly in terms of long-term averages, with the exception of the classification tree based on an assumption of low steepness (Figure 79). This decision rule appears to have a much lower harvest rate leading to much higher biomass. This behaviour is confirmed when we look at the summary results for 100 realisations

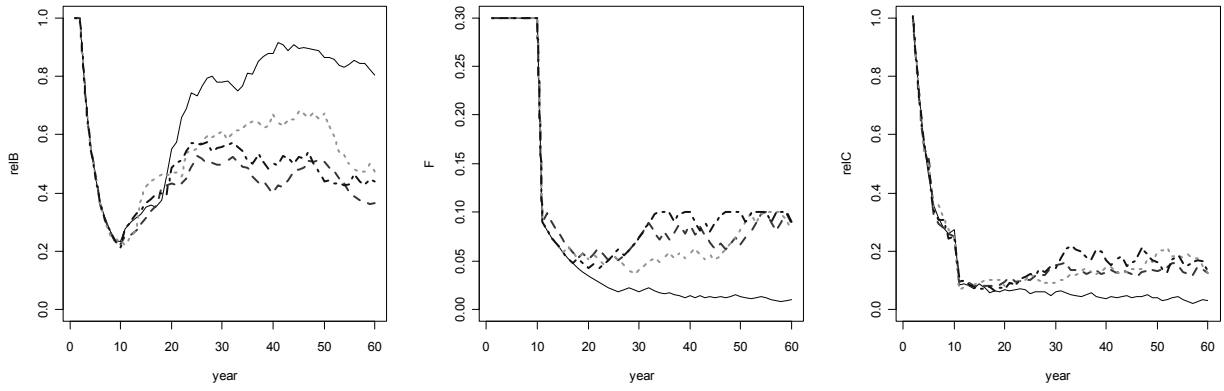


Figure 79. Time series plots of a single realisation for four decision rules: binary based on mean length (dashed line), binary based on CPUE in weight (dotted line), classification tree 'h9.flat' (dash-dot-dash line) and classification tree 'h4.flat' (solid line). See text for detail. Panels are for relative spawning biomass (relB), harvest rate (F) and relative catch (relC).

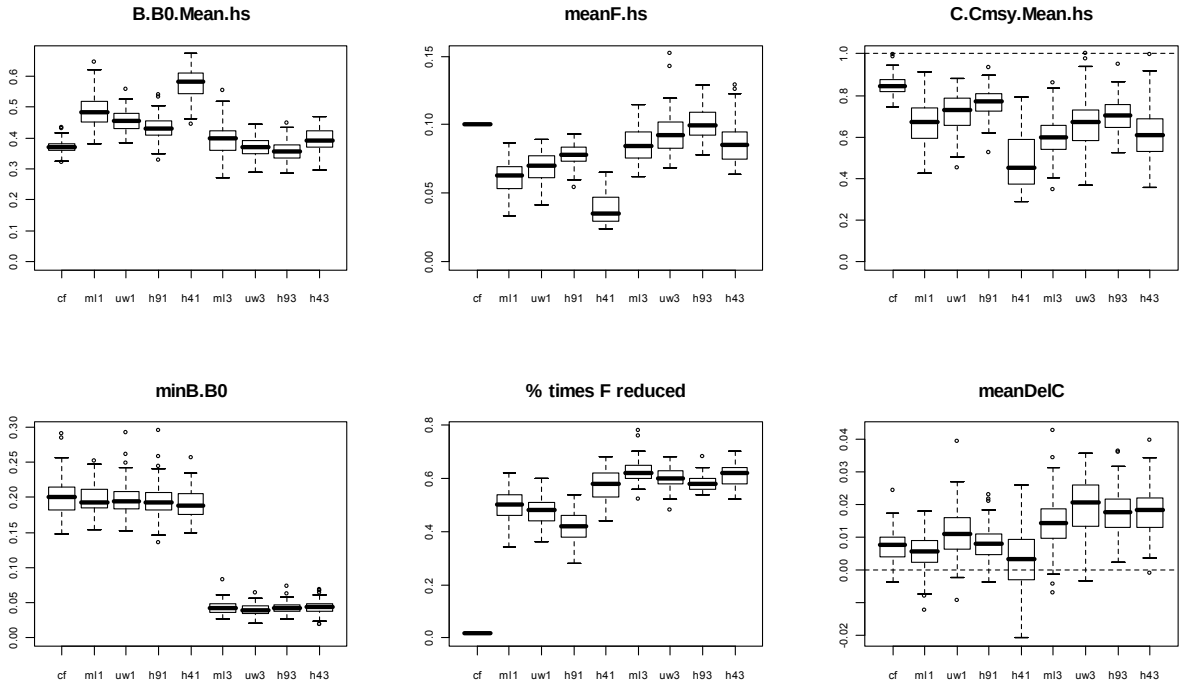


Figure 80. Summary results of 100 realisations for four decision rules: binary based on mean length (ml), binary based on cpue in weight (uw), classification tree 'h9.flat' (h9) and classification tree 'h4.flat (h4). The extension '1' or '3' indicates whether $F_{max}=1.F_{spr40}$, or $F_{max}=3.F_{spr40}$. Results for a constant harvest rate $F=F_{spr40}$ (cf) is plotted first for comparison. Panels are as in figures above, with averages taken over the period that the decision rule was operational.

First consider the first five 'boxes' in Figure 80, showing results for the case where $F_{max}=1.F_{spr40}$. As suggested in the time-series plots, the classification tree h4.flat leads to much higher biomass and lower harvest rate (and therefore catch) than the other decision rules. This does not mean it is better by definition; compared to the constant harvest rate at $F=F_{spr40}=0.1$, we note that it is arguably overly conservative. What is relevant here, is that without tuning of the rule, i.e. changing its parameters, it behaves in a different way compared to the others. It is very likely that this is due to the fact that the second node uses L90 as an indicator; L90 has a high prediction error relative to the other indicators for this particular life-history/selectivity scenario. The other three decision rules behave more similarly, though there is still a persistent pattern in the means; the mean length binary rule has a lower average F than the CPUE (in weight) binary rule, which in turn has a lower average F than the h9.flat decision tree.

When $F_{max}=3.F_{spr40}$ (the last four 'boxes' in Figure 80), the performance is still arguably quite acceptable. The average harvest rate, for the binary and tree-based rules, is generally well below F_{max} even if it is set at $3.F_{spr40}$ (Figure 80)[25]. The mean level of depletion for all four rules under the high F_{max} is around 0.4, whereas for the 'up-down' rules with high F_{max}, the mean level of depletion is around 0.2 (Figure 76). In spite of the fact that the rules have not been tuned, we consider the comparison reasonable because of the similarity in the construction of the rules (F_{max} and the reduction parameter 'd' are the same). The difference in performance between the two types of decision rule is most likely to lie in the fact that the 'up-down' rule only considers the

[25] It is probably worth recalling that although F_{max} is deliberately being set much too high when set to $3.F_{spr40}$, the trigger points are based on F_{spr40}. The potential errors in setting the trigger points are built into the simulations and are, in our view, likely to be less extreme than the potential error in the absolute level of the harvest rate or catch specified in the decision rule.

slope of the indicator, whereas the binary and tree rules consider the value of the indicator relative to a trigger point which is related to a 'reference' harvest rate.

The main problem is that the risk of low spawning biomass is much higher when F_{max} is high than when it is low, as is clear from the panel showing the minimum relative spawning biomass. This is also evident from statistics such as the proportion of time that relative biomass $B/B_0 <$ 0.3 (Figure 81), which is much higher when F_{max} is high than when it is low. Again, however, we note that the binary decision rule has a lower associated risk than its coinciding 'up-down' version with the same F_{max}.

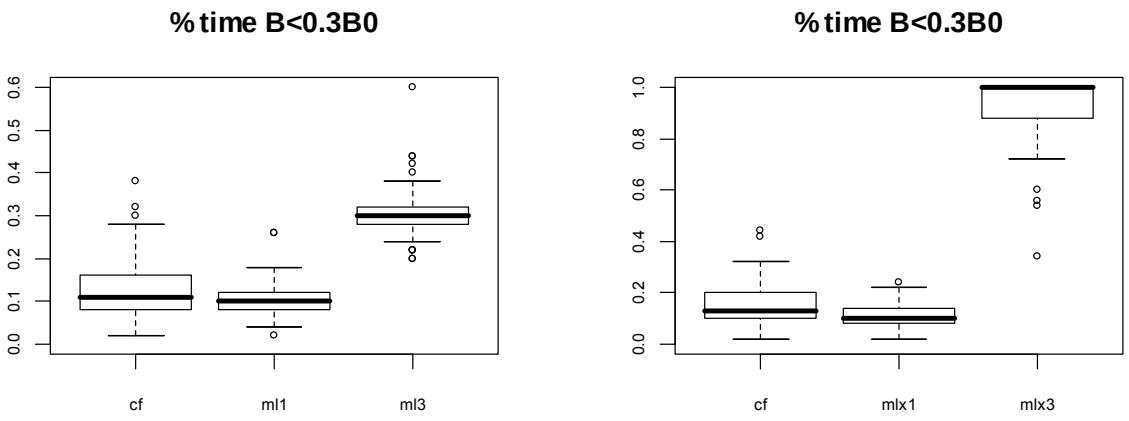


Figure 81. The percentage time that relative biomass is below 30% of unexploited under the constant harvest rate F=Fspr40 (cf) and for the binary rule based on mean length with Fmax=Fspr40 (ml1) or Fmax=3.Fspr40 (ml3) in the left panel. The right panel is again for constant F=Fspr40, and for the 'up-down' rule based on mean length with Fmax=Fspr40 (mlx1) or Fmax=3.Fspr40 (mlx3). Note the difference in scale on the y-axis for the two panels.

There are obviously still many questions to ask about the performance of the binary and tree-based decision rules, for example, under different steepness levels or other selectivity assumptions, but we first continue to look at the same scenario (low m, high h, flat selectivity), with a few more types of decision rule.

10.3.4 Traffic light decision rules

Traffic light decision rules use multiple indicators, but unlike the classification trees, there is not a particular sequence or logic in which the indicators are considered. The number of indicators that are below their trigger points ('red lights') are simply counted. If the critical number of red lights for invoking management action is 1 red light for example, then if any of the indicators is below its trigger, the management action will be triggered.

We look at two groups of indicators, chosen so that we can compare the performance of the classification tree with that of the traffic light rule. The first set consists of the five indicators (i.e. five 'traffic lights') in the 'h9.flat' decision tree and is referred to as 'set h9' = ml, usb, ub, ps, us. The second set of five is taken from the 'h4.flat' decision tree ('set h4' = ub, L90, usb, mw, mdl; see Section 2.4.3 for indicator abbreviations and names). Recall that, in addition to the parameters Fmax and d (reduction parameter), the traffic light rule also requires a parameter to specify how many 'red lights', should trigger a response. The behaviour of this type of decision rule can be sensitive to this parameter, as we illustrated in sub-section 2.2 above.

First we look at results for an arbitrary level of 3 'red lights' out of the total of 5. As before, the high F_{max} (=0.3) leads to lower average biomass and a higher average harvest rate which is nonetheless still substantially below F_{max} (the median is around 0.1). As was the case for the classification trees, there is again a difference in performance between the two sets of indicators; the h4 set is more conservative since it reduces F more often leading to a lower average harvest rate than the h9 set. It is interesting to go back to results in Section 9. Closer scrutiny reveals that for the scenario we are considering here, the h9 set has an average (over the 5 indicators) prediction error of 11.2%, whereas the h4 set has an average (over its 5 indicators) of 8.8%. The main difference is, however, that h4 set leads to a larger proportion of errors where the indicator is 0 but the biomass is 1 (in binary terms) than the h9 set. For example, the 8.8% is made up of a 6.1% type II error (indicator=0, biomass=1) and only a 2.7% Type I error (indicator=1, biomass=0). For the h9 set these figures are almost identical at about 5.6% for each type of error. It is clear how this would tend to lead to more conservative behaviour in a decision rule based on the h4 set of indicators.

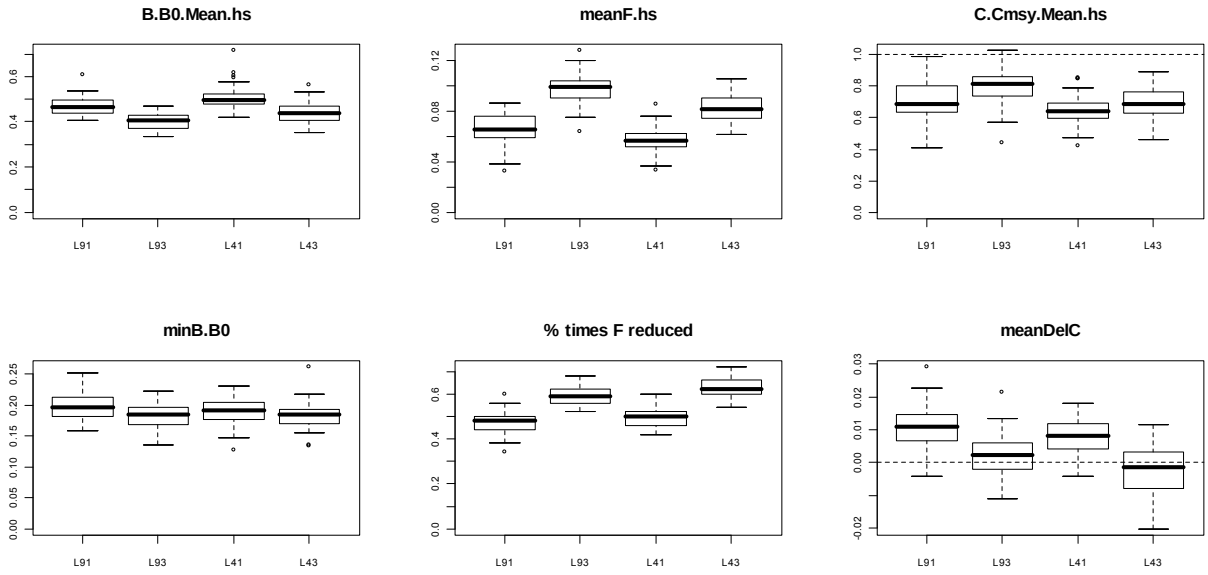


Figure 82. Summary results for traffic light rules triggered by 3 'red lights'. The first two boxes are for the traffic light rule with the h9 set of indicators and low F_{max} (L91) or high F_{max} (L93). The last two boxes are for the traffic light rule with the h4 set of indicators and low F_{max} (L41) or high F_{max} (L43).

Although we can compare the traffic light rule with the classification tree rule, this is a clear example where one should ideally first tune the traffic light rule to achieve the same objective with respect to one of the performance measures and then evaluate performance according to the other performance measures. We illustrate why this is important by showing results for three choices of the number of red lights that trigger action. Results of the classification tree tends to lie between results for 3 and 4 red lights triggering action, but a tuning with only 1 red light triggering action leads to much more conservative behaviour (Figure 83). This tuning parameter is obviously limited to integer values, but the reduction parameter 'd' can also be changed to achieve a more precise tuning if required.

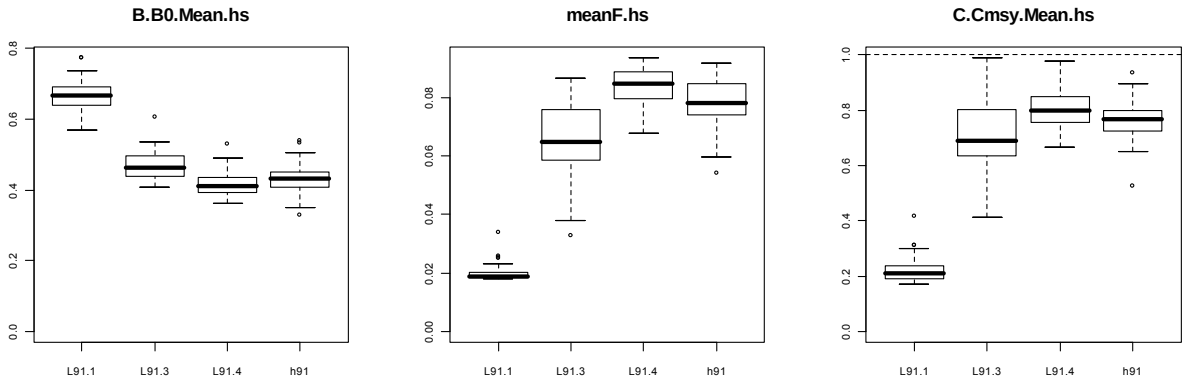


Figure 83. Results for the traffic light decision rule with h9 set of indicators, low Fmax, and different settings for the number of red lights that trigger action: 1 red light (L91.1); 3 red lights (L91.3) or 4 red lights (L91.4). The classification tree (h9.flat) is shown for comparison.

At this stage there is nothing to suggest that the traffic light rules (at least the binary versions of the rule) perform much better or worse than the classification tree rules, particularly if the set of indicators is well chosen.

10.3.5 The 20-40 and 10-40 decision rules

We now consider the performance of the 20-40 decision rules using mean length or CPUE, and with the F_{max} again either set at F_{spr40}=0.1 or at 3.F_{spr40}=0.3. It is interesting to note that the mean harvest rate is generally below F_{max}. For the case where F_{max}=3.F_{spr40}, the average is in fact only half that – around 0.15 (Figure 84, panel 'meanF.hs'), and although the average biomass is lower than for F_{max}=0.1, it is in fact somewhat higher than for the 'up-down' rule with F_{max}=0.3 (see scenario 'mlx3' and 'ux3' in Figure 76.

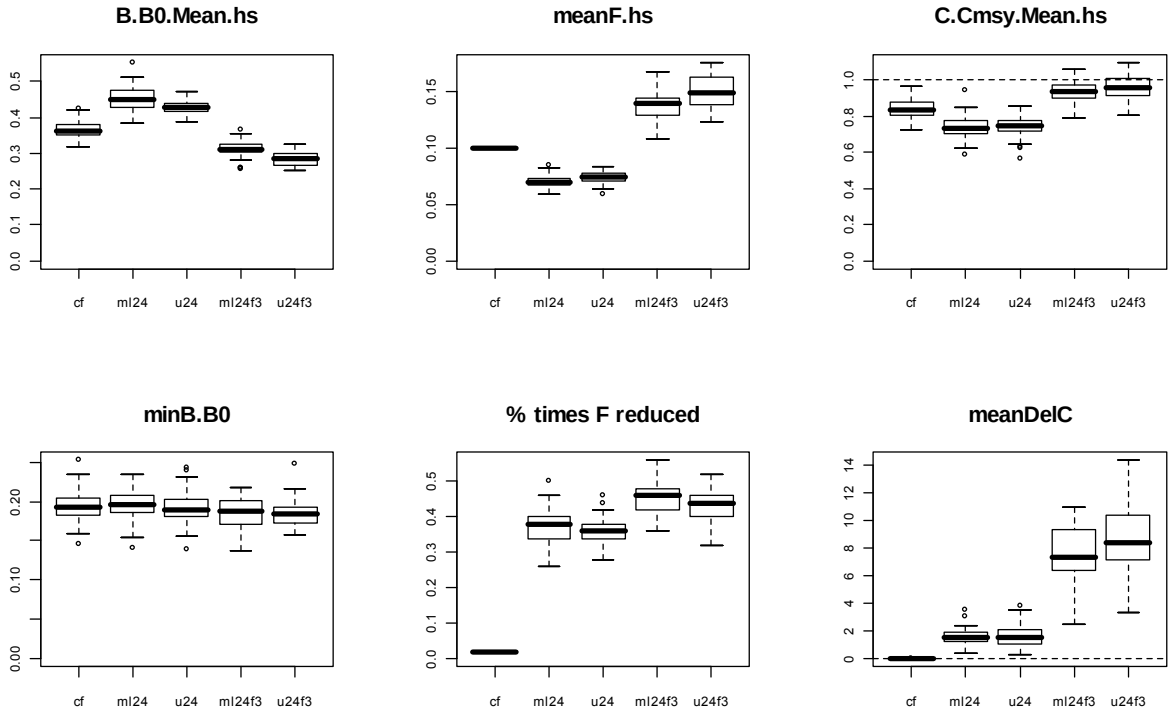


Figure 84. Results for the 20-40 decision rule with mean length as indicator (ml) or CPUE (u) as indicator. The constant harvest rate (Fspr40=0.1) is shown first, followed by the pair (ml24 and u24) with Fmax=Fspr40. The last pair (ml24f3, u24f3) is for Fmax=3.Fspr40. The performance statistics are as in the figures above.

A direct comparison between the 'up-down' and '20-40' rules with F_{max} set too high, shows that the '20-40' rule leads to a lower average F and hence a higher biomass than the 'up-down' rule. This is not surprising given the definition of the two rules.

Although the versions with low F_{max} and with high F_{max} still show greatest difference, there are now slightly more differences between results for mean length as indicator or CPUE as indicator. This is most likely driven by the number of times the harvest rate is reduced which is, in turn, a function of the measurement error in the indicator and also how close together the limit and target trigger points are.

What the summary statistics do not show, is the large magnitude of changes in the harvest rate under the 20-40 decision rule and the tendency for this to induce oscillations in the catch and, to a lesser extent, the biomass. Recall that the 'up-down' rule essentially has a fixed change built into it via the 'd' parameter. It is possible to build in such a constraint on the magnitude of change for the 20-40 rule, but there is a direct trade-off between the magnitude of change (and frequency of change) allowed and the average level of catch that can be taken to achieve the same conservation goals.

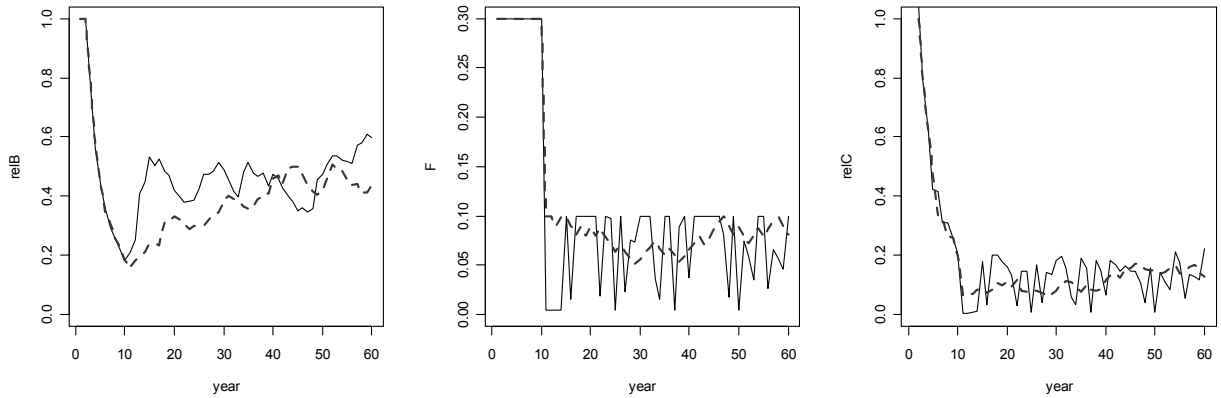


Figure 85. The 20-40 rule with mean length (solid line) and 'up-down with maximum' rule with mean length (dashed line); both rules have Fmax=Fspr40.

10.3.5.1. The 10-40 decision rule with a single indicator

The '10-40' version of the decision rule reduces the frequency of large fluctuations in F and hence reduces the average change in catch to some extent (Figure 86), though the overall frequency of changes (of any magnitude) in F is higher (Figure 87). The reason for this is most likely because the indicator is more often between the limit and target trigger points under the 10-40 rule than under the 20-40 rule, when it often falls below or very close to the limit (although this is not evident from the realisation plotted in Figure 86). Recall that if the indicator is below the limit point, the harvest rate remains unchanged at a very low level (0.005 in these simulations).

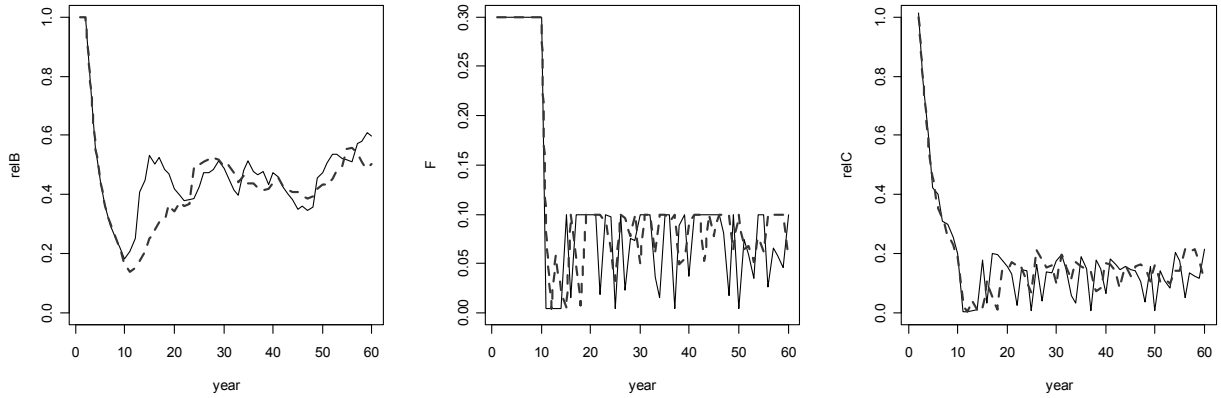


Figure 86. The 20-40 rule with mean length (solid line) and the 10-40 rule with mean length (dashed line); both rules have Fmax=Fspr40.

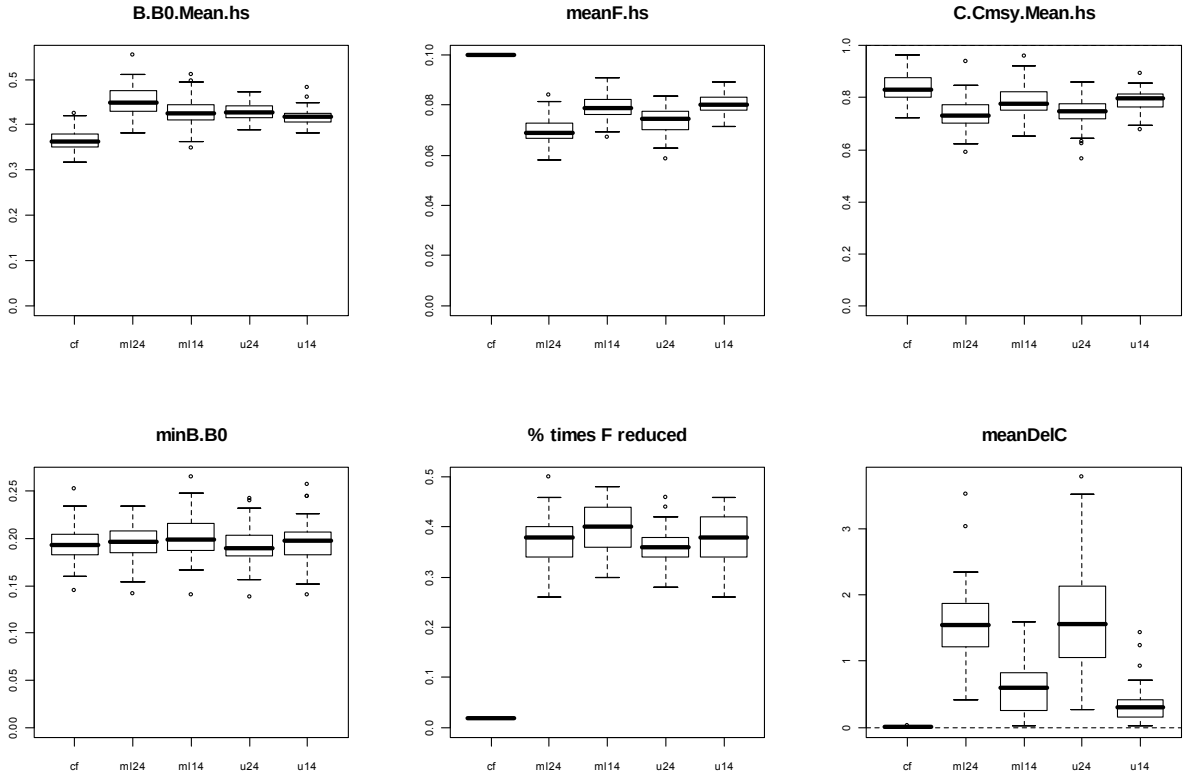


Figure 87. Results for the 20-40 ('24') and 10-40 ('14') decision rules with mean length as indicator (ml) or CPUE (u) as indicator. The constant harvest rate (Fspr40=0.1) is shown first, followed by the pair based on mean length (ml24 and ml14) and then the pair based on CPUE (u24 and u14). All four decision rules have Fmax=Fspr40. The performance statistics are as in the figures above.

There is relatively little difference in the median level of biomass or the minimum biomass between the 20-40 and the 10-40 versions of the rule. The average F and catch are slightly higher under the 10-40 rule. There is also a hint of better performance by the 10-40 rule, in terms of the lower quantile of the minimum relative biomass performance measure. Note though that the minima and maxima on the box-plots and the ranges they span should be interpreted with some caution because these results are based on only 50 realisations. Ideally, at least 200 realisations should be used to make inferences about statistics such as the minimum or about upper/lower percentiles of the distributions of results. At this stage, however, we are primarily considering average behaviour or characteristics of single realisations.

10.3.5.2. The '10-40' decision rule with multiple indicators

As noted in the methods section, it is possible to combine indicators in a 20-40 type rule. In the example shown below, we have combined the following five indicators in set A: mean length, CPUE of spawners, CPUE of big fish, the proportion small fish and the CPUE of small fish (mL, uSB, ub, ps, us). These indicators are the same ones that appear in the pruned tree for this scenario and that was used in a classification tree and a traffic light decision rule. For comparison, we also look at a random set of five indicators (by drawing 5 numbers from the sequence 1:13 without replacement!), referred to as set B: proportion big fish, CPUE, CPUE in weight, 90th percentile of weight and the 90th percentile of length (pb, u, uw, W90, L90).

As one might expect, the decision rule based on the minimum statistic, scenario A14min in (i.e. the lowest indicator), is highly conservative and when compared to the constant $F=F_{spr40}$

performance measures, it is arguably over-conservative. The version based on the average statistic (A14) shows similar performance to the rules based on just mean length or just CPUE. The main difference is that the harvest rate is changed more often, but the average magnitude of change in F (and therefore also the catch) is less. In addition, this decision rule does not invoke the 'fishery closure' (F=0.005 in the simulations) as often as the single indicators do. Time-series plots (not shown) confirmed that the version based on the minimum is not ideal because of the high frequency of 'fishery closures'.

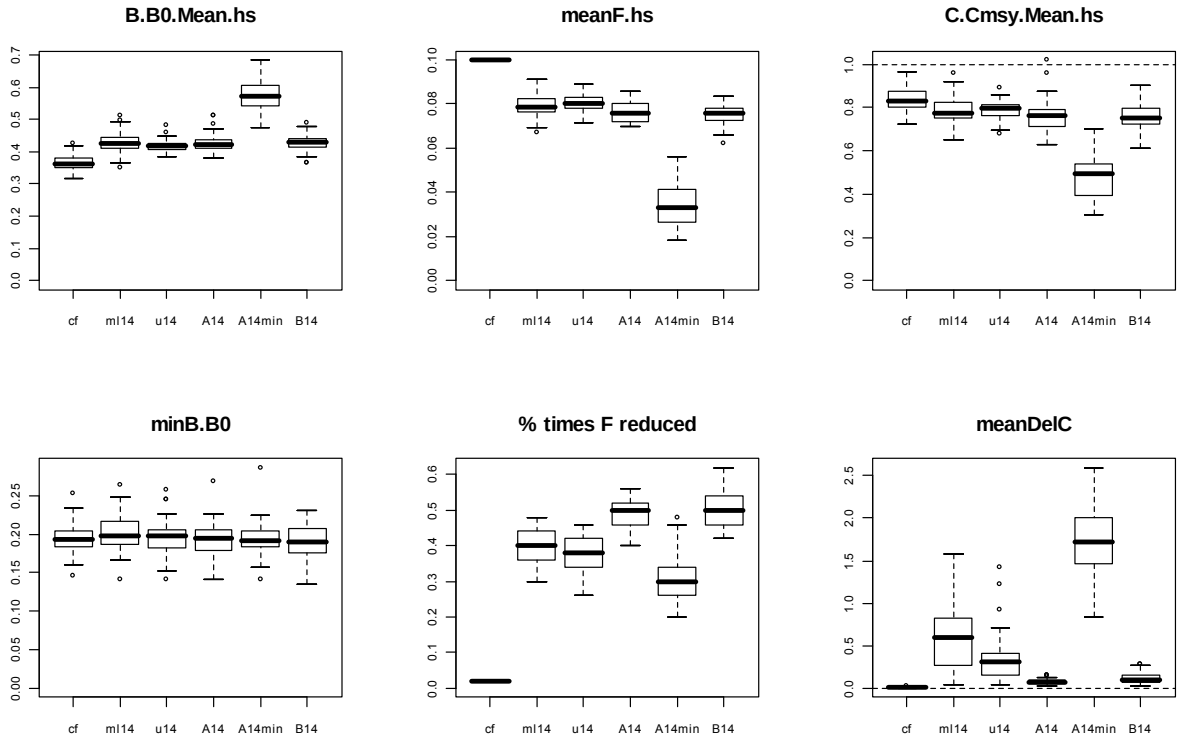


Figure 88. Results for 10-40 ('14') decision rules with mean length as indicator (ml), CPUE (u) as indicator or a set of indicators (A and B; see main text) combined using a mean statistic (A14 and B14) or a minimum statistic (A14min). The constant harvest rate (Fspr40=0.1) is shown for comparison. All five decision rules have Fmax=Fspr40. The performance statistics are as in the figures above.

It is amusing to compare results for the two different sets of indicators (A and B; scenarios A14 and B14 in Figure 88). The only noticeable different is in the frequency of reductions in F which is slightly higher for the random set (B). This is also reflected in the mean change in catch, but it is not obvious in the figure because of the much higher values for the A14min and ml14 scenarios.

We briefly return to a single-indicator version of the 10-40 decision rule based on the CPUE of small fish (indicator 'us'). In Section 9 we commented on the large prediction errors for this indicator relative to most of the others under many scenarios. We have already noted that the performance of mean length and CPUE as indicators are quite similar when used in comparable decision rules (though recall that this is for a high steepness scenario). Figure 89 (taken from Figure 24 in Section 9) shows the prediction errors of the full set of indicators under the same scenario that we've been considering here (low m, high steepness, high noise level=HHH, SPR trigger points), and show that the prediction errors for mean length and CPUE are quite similar at high steepness.

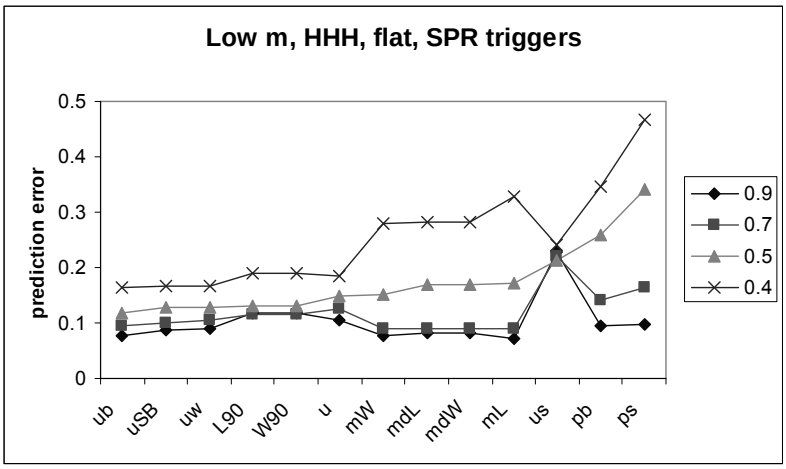


Figure 89. Prediction errors for a range of steepness levels with high noise (HHH), low m, FLAT selectivity and SPR triggers. Indicators are sorted by the prediction errors for h=0.5 (triangles).

If we use the CPUE of small fish, which has the highest prediction error (at high steepness), in a 10-40 decision rule, its performance is much poorer than that of mean length or CPUE (Figure 90) because it leads to a large number of reductions in F and hence a very low average harvest rate, low catch, and (not shown) very high average change in catch from year to year. Time-series plots confirm that this rule also invoke the 'fishery closure' very often. Although we only presented the overall prediction errors in Section 9, it is interesting that the 'us' indicator's 23% prediction error is made up of 15% type II error (the indicator predicts biomass below its reference point when the biomass is in fact above its reference point) and only 8% type I error (the indicator predicts biomass is above its reference point when the biomass is in fact below its reference point). This is consistent with the performance seen here, where the decision rule is essentially over-conservative.

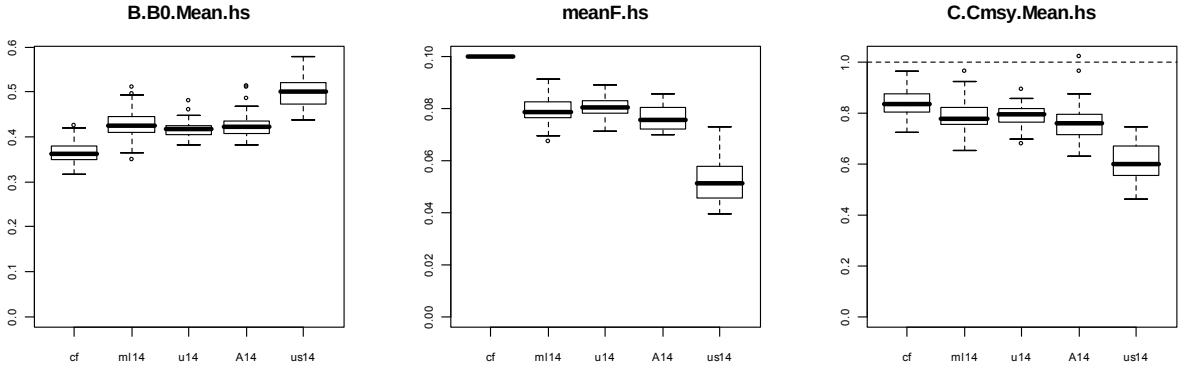


Figure 90. Results for 10-40 ('14') decision rules with mean length (ml), CPUE (u), set A and CPUE of small fish (us) as indicators. The constant harvest rate (Fspr40=0.1) is shown for comparison. All decision rules have Fmax=Fspr40. The performance statistics are as in the figures above.

10.4. Sensitivities: steepness, selectivity and effort creep

10.4.1 Low steepness

In the first part of this section (4.1.1) the true selectivity is assumed to be flat and all the decision rules are based on the assumption of flat (i.e. correct) selectivity. In the second part of the section (4.1.2) we look at a subset of decision rules for the case where true selectivity is assumed to be dome-shaped. Here too, we use decision rules based on the (correct) assumption of dome-shaped selectivity.

10.4.1.1. Flat selectivity

When considering the effect of different steepness values on the performance of decision rules, it is worth recalling that the SPR-based reference points themselves are not affected by steepness (i.e. make no assumption about steepness). For the very low steepness (h=0.4 in the 'true' dynamics), it is then not surprising that the biomass is more depleted under the same rule, than it is when steepness is high. Ideally, one is looking for a decision rule which shows least difference in the level of depletion and most difference in the mean harvest rate (with a lower harvest rate when steepness is low). Figure 91 shows results for the 'up-down' rule with F_{max}, and the 10-40 rule using either mean length or CPUE. These results suggests that the 10-40 rule based on CPUE performs better than the others since it shows the least difference in relative biomass for the two steepness levels (the two right-hand boxes in the first panel, compared to the other pairs). It also has the best performance in terms of the minimum ratio of B/B_0. The price that is paid to achieve this, is relatively high variability in catches under low steepness. This is in fact a very common trade-off.

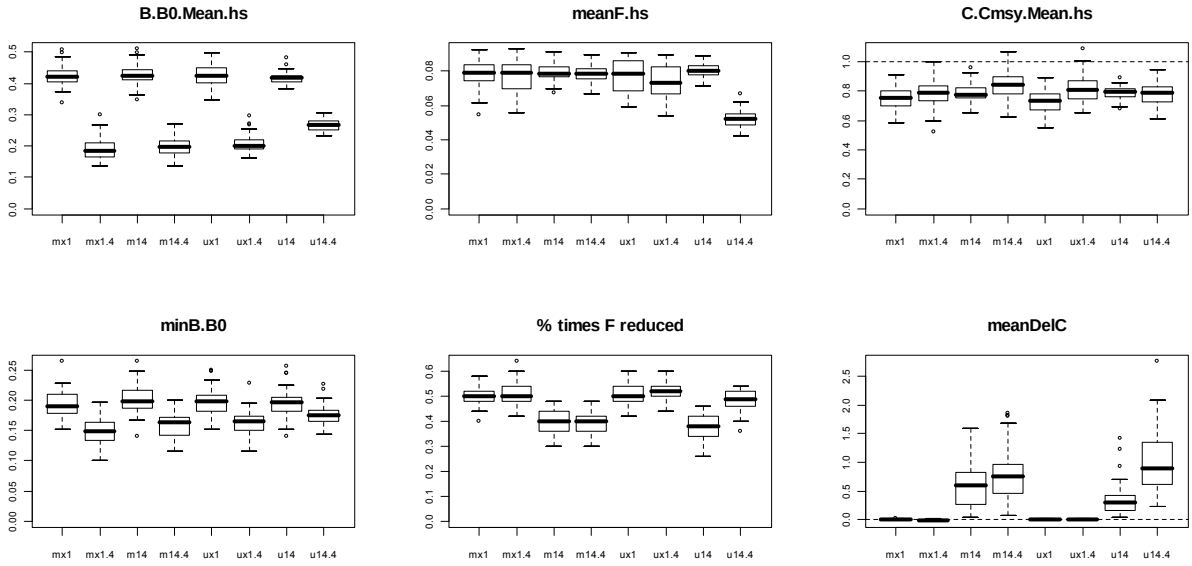


Figure 91. Comparisons of 'up-down with Fmax' and 10-40 rules under high steepness (h=0.9) and low steepness (h=0.4). Results are plotted in pairs for high, then low steepness, indicated by '.4' extension to the name. Rules are: 'up-down with Fmax=0.1 using mean length (mx1 and mx1.4); 10-40 rule with mean length (m14 and m14.4); 'up-down' with Fmax=0.1 using CPUE (ux1 and ux1.4); 10-40 rule with CPUE (u14 and u14.4).

For the two binary rules, the one based on cpue (in weight) performs well since it reduces the harvest rate and achieves a relatively high level of relative biomass. It also has the best

performance in terms of the minimum B/B_0 ratio. The catch is, however, much reduced as a result. The binary rule based on mean length does not perform as well and achieves only an average of 20% for B/B_0, compared to the cpue version which manages an average close to 35%. The classification tree which was based on simulated data with h=0.9 does relatively poorly when true steepness is low (compare 'h9' with 'h9.h4'). On the other hand, the classification tree which was based on simulated data with h=0.4 does relatively poorly when steepness is high (it has unnecessarily low harvest rate). It performs well when true steepness as low, as one would hope. On the whole, however, the cpue-based binary rule performs best when judged according to the smallest difference between depletion levels at the two steepness levels.

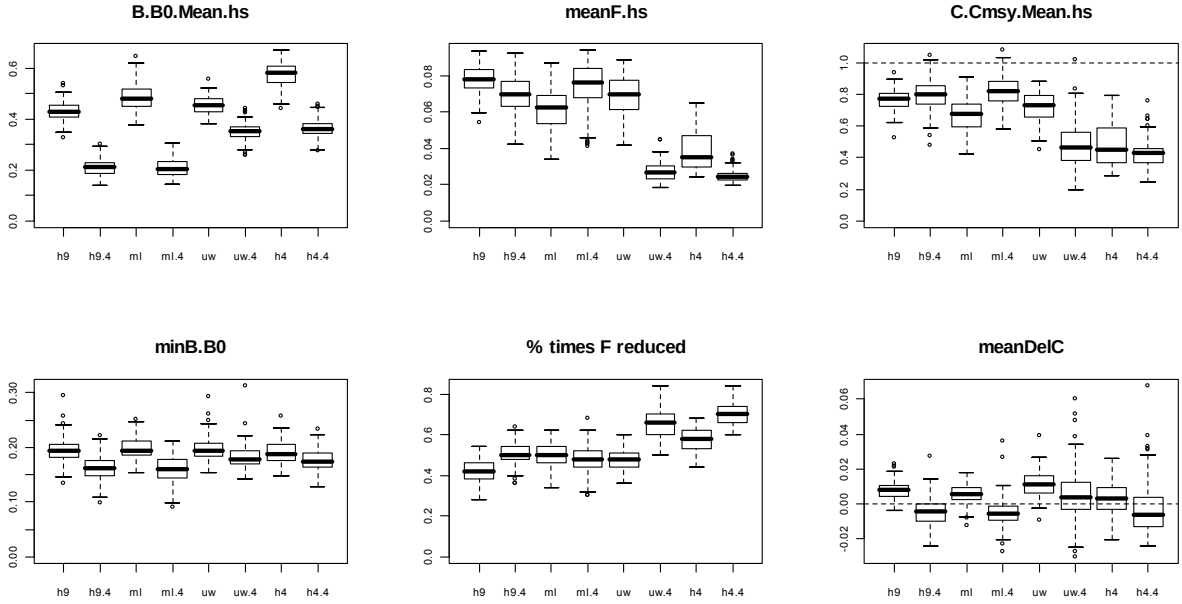


Figure 92. Comparisons of binary and classification tree rules under high steepness (h=0.9) and low steepness (h=0.4). Results are plotted in pairs for high, then low steepness, indicated by '.4' extension to the name. Rules are: classification tree h9 (h9 and h9.4); binary rule using mean length (ml and ml.4); binary rule using CPUE (in weight) (uw and uw.4); classification tree h4 (h4 and h4.4).

Results for the traffic light rule based on the h9 set of indicators (mL, uSB, ub, ps, us) leads to a similar drop in relative biomass when steepness is low, as the classification tree rule. Although the two rules have not been finely tuned, we illustrated above (Section 3.4) that under h=0.9, the traffic light rule with a setting of 4 red lights triggering action, leads to results that are quite similar to the classification tree decision rule in terms of relative biomass. A setting of 4 is, however, too high when the traffic light and classification tree are based on the h4 set of indicators; a setting of 2 gives more similar results and is used as an approximate tuning for these comparisons. Figure 93 shows results for the h9 set of indicators in the left section of each panel and the h4 set of indicators in the right section of each panel.

There is little difference in performance between the traffic light and classification tree rules when steepness is low rather than high. Both types of rules lead to lower harvest rates under the lower steepness, but not substantially so, and therefore they also lead to lower relative biomass. There may be some subtle differences. For example, the traffic light rule, particularly that based on the h9 set of indicators, seems to have a slightly tighter distribution of minimum B/B0 values and now quite as low a minimum (of the minimum) as the classification tree rule. Note though that these results are only based on 50 realisations which is not really sufficient for drawing conclusions about extremes.

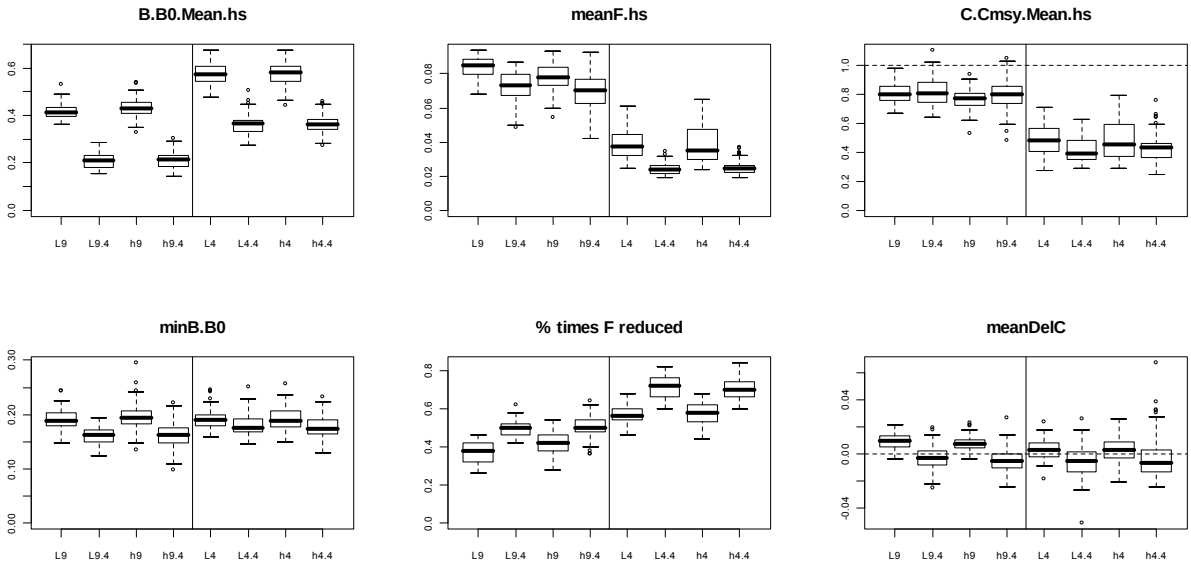


Figure 93. Comparisons of traffic light and classification tree rules under high steepness (h=0.9) and low steepness (h=0.4). Results are plotted in pairs for high, then low steepness, indicated by '.4' extension to the name. The left section of each panel is for rules based on the h9 set of indicators, i.e. traffic light rule (4 trigger lights) (L9 and L9.4) and classification tree rule (h9 and h9.4). The right section of each panel is for rules based on the h4 set of indicators, i.e. traffic light rule (2 trigger lights) (L4 and L4.4) and classification tree rule (h4 and h4.4).

10.4.1.2. Dome-shaped selectivity

Dome-shaped selectivity implies much lower (or zero) harvest rates on very large individuals. Since these individuals form an important part of the spawning biomass, it is evident that scenarios based on dome-shaped selectivity could behave differently at low steepness than the scenarios based on flat selectivity. Figure 94 does indeed show much less sensitivity to low steepness than the comparable set of results for flat selectivity (see Figure 92 above). For example, the panel for mean relative biomass shows that the pairs for h=0.9 and h=0.4 are much more similar in Figure 94 than in Figure 92. Judged on this criterion, the binary CPUE-based rule and that based on the low steepness decision tree (d4) perform best of the four shown here.

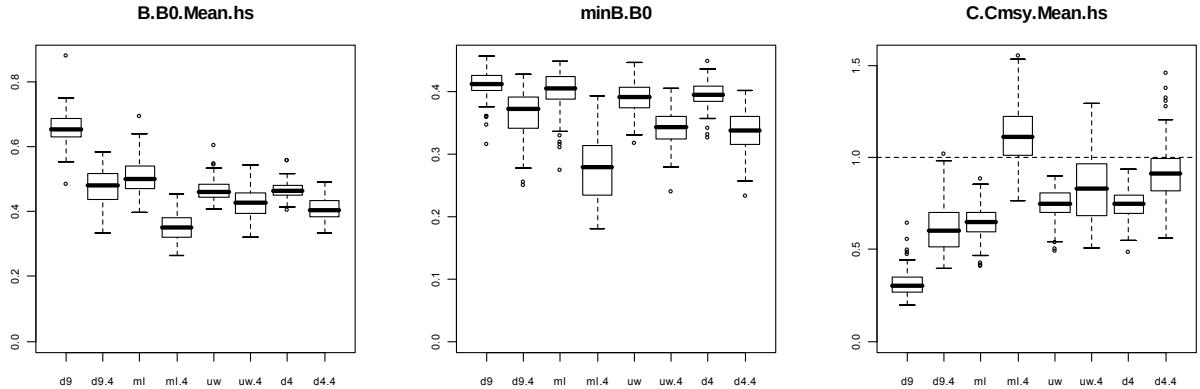


Figure 94. Comparisons of binary and classification tree rules under high steepness (h=0.9) and low steepness (h=0.4) when selectivity is dome-shaped. Results are plotted in pairs for high, then low steepness, indicated by '.4' extension to the name. Rules are: classification tree d9 (d9 and d9.4); binary rule using mean length (ml and ml.4); binary rule using CPUE (in weight) (uw and uw.4); classification tree d4 (d4 and d4.4).

It is very likely that the other types of rules (10-40 and traffic light) would also show less sensitivity to steepness when selectivity is dome-shaped than when it is flat.

10.4.2 Selectivity

In this section we first look at scenarios where true selectivity is flat (4.2.1), but the decision rules are based on the (wrong) assumption that selectivity is dome-shaped. We look at this for high and for low steepness in the true dynamics. In the second part (4.2.2) we look at scenarios where true selectivity is dome-shaped, but the decision rules are based on the (wrong) assumption that selectivity is dome-shaped.

10.4.2.1 True selectivity is flat

The assumption about selectivity affects any trigger or reference points that are built into a decision rule. This means that the 'up-down' rules would not directly be affected by selectivity (though one would need to explore this issue if tuning such a rule). The binary, classification tree, traffic light and 20-40 type decision rules are, however, all directly affected. Here we look at the results for simulations where the true selectivity is either flat (or dome-shaped, in the next sub-section) and where the construction of the decision rule is either based on the correct or incorrect assumption about selectivity. First we consider, as in all the explorations above, the case where true selectivity is flat. Figure 95 shows results for binary and classification trees. Results are plotted in pairs for each decision rule: first based on the correct (flat selectivity) assumption, and then based on the incorrect (dome-shaped selectivity). In these runs, however, the value of F_{max} was unchanged (F_{max}=0.1 in all scenarios); see below for runs where F_{max} was also changed. Generally, the second 'box' in each pair shows higher relative biomass and lower harvest rate. So, assuming dome-shaped selectivity when it is really flat, leads to conservative results, presenting a risk to the fishery (lower catches), but not to the stock. It is only the h4 classification tree that has a slightly lower relative biomass under the wrong assumption, but not substantially lower. Note also that the h9 classification tree is arguably the most sensitive (largest difference between results); the mean length and h4 classification tree are less sensitive.

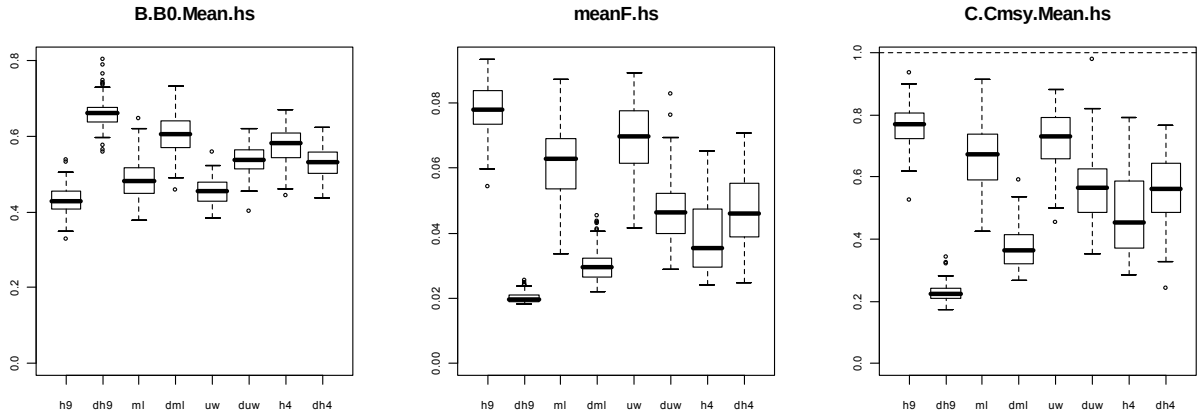


Figure 95. True selectivity is flat, steepness is high and Fmax=0.1 in all cases. Binary rules and classification trees either assume the correct (flat) selectivity or the incorrect (dome-shaped, prefix 'd.' to name) selectivity and results are plotted in these 'correct – incorrect' pairs for each rule. The rules are: classification tree h9, binary rule based on mean length (ml), binary rule based on cpue (in weight, uw) and classification tree h4. Measurement error was high on all components (HHH scenario).

Results above are for the high measurement error scenario. When the measurement error on all components is low, the overall patterns are very similar.

When true selectivity is flat and steepness is also low, results are as in Figure 96. Results are again more optimistic (or more conservative) when dome-shaped selectivity is assumed. The CPUE binary rule and the h4 classification tree rule shows least difference between the two assumptions.

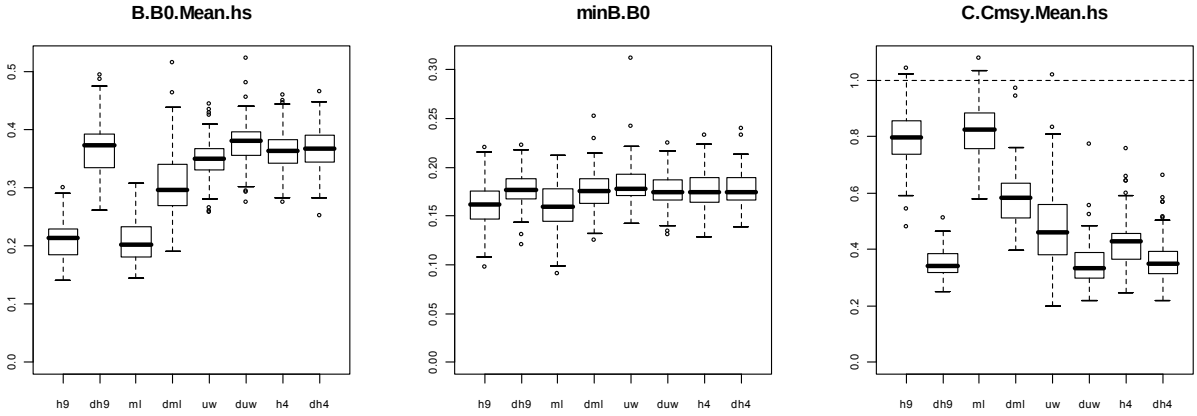


Figure 96. True selectivity is flat and true steepness is low; Fmax=0.1 in all cases. Binary rules and classification trees either assume the correct (flat) selectivity or the incorrect (dome-shaped, prefix 'd.' to name) selectivity and results are plotted in these 'correct – incorrect' pairs for each rule. The rules are: classification tree h9, binary rule based on mean length (ml), binary rule based on cpue (in weight, uw) and classification tree h4. Measurement error was high on all components (HHH scenario).

As pointed out above, the value of Fmax was not changed in the scenarios shown in Figure 95 and Figure 96. Fmax has, however, been based on Fspr40 which is, of course, different when dome-shaped selectivity is used. We deliberately separated out the effect of just changing the trigger points (as above) and changing the trigger points and the maximum harvest rate. Figure 97 shows comparisons, just for the binary CPUE rule, of all three scenarios, i.e. correct selectivity, incorrect selectivity (but correct Fmax), and incorrect selectivity and Fmax. Given that Fspr40 is higher under dome-shaped selectivity than under flat selectivity, it is not surprising that when both selectivity and Fmax are based on the dome-shaped selectivity assumption, relative biomass and minimum biomass are lower. When steepness is high, this error may not be a problem, but when steepness is low, it is more risky.

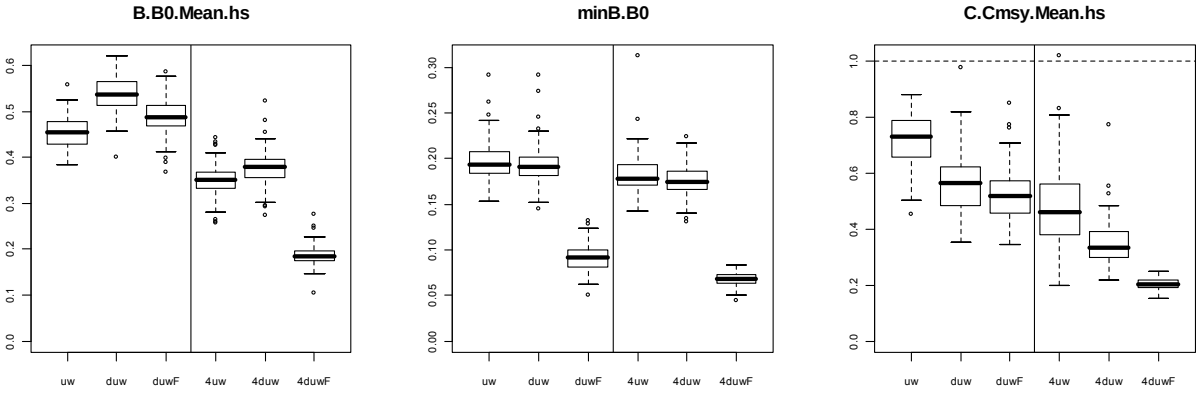


Figure 97. True selectivity is flat. The first 3 boxes are for true h=0.9, the second three are for true h=0.4 (prefix '4' to scenario name). In all cases it is the binary decision rule using CPUE. Assumed selectivity and Fmax for decision rules are: flat, Fmax=0.10 (uw, 4uw); dome-shaped, Fmax=0.10 (duw, 4duw); dome-shaped with Fmax=0.18 (duwF, 4duwF).

10.4.2.2. True selectivity is dome-shaped

Next we consider the opposite: true selectivity is dome-shaped. For these examples, the decision rules based on the correct assumption, are of course different from those used so far (i.e. those shown in Figure 78) , and we therefore repeat the figures for the classification tree structures here in Figure 98. In the rest of this section, the decision rule based on the 'h9,dome' tree in Figure 98 is referred to as the dh9 classification tree, and the 'h4,dome' tree is referred to as the dh4 classification tree.

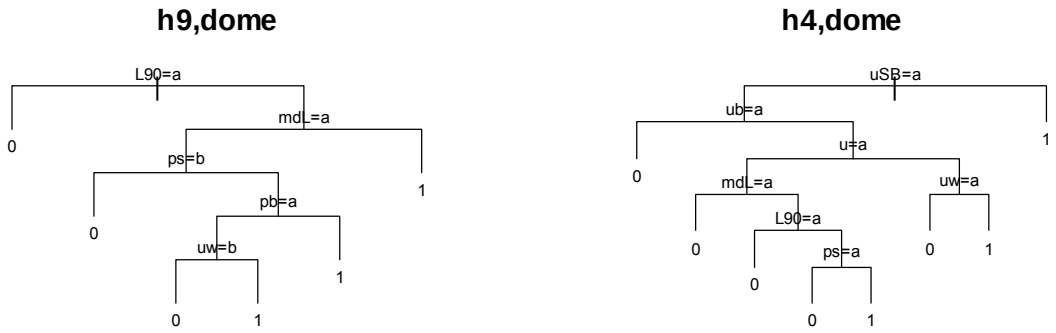


Figure 98. Classification trees fitted to simulated data based on high steepness (h9,dome) and on low steepness (h4,dome), for the long-lived life history with DOME-shaped selectivity, high measurement error levels and SPR40 reference points. Repeated from Section 9; see text for more detail.

The first set of comparisons is again for the case where F_{max} is not changed, so in all these scenarios, F_{max}=0.18. Now the wrong assumption (second 'box' in each pair, Figure 99) tends lead to less conservative results, though in some cases the differences are small. For example, results of relative biomass for the single-indicator rules, particularly for that based on CPUE, are not substantially different. Results of relative biomass are, however, rather different for the classification trees. Note in particular how the dh9 classification tree seems to be very conservative in its behaviour when true selectivity is dome shaped (the first 'box' in each panel of Figure 99). We consider this further below (under: Revisiting classification tree fitting), because it illustrates an interesting point[26].

The only pair that shows an opposite change is the pair dh4 and fh4. Here the version based on flat selectivity (and low h) is in fact more conservative than the one based on the correct selectivity.

[26] The refitted tree, dh9.C leads to results that are much more similar to those for the fh9 scenario in the above figures, making the pairwise comparisons more similar.

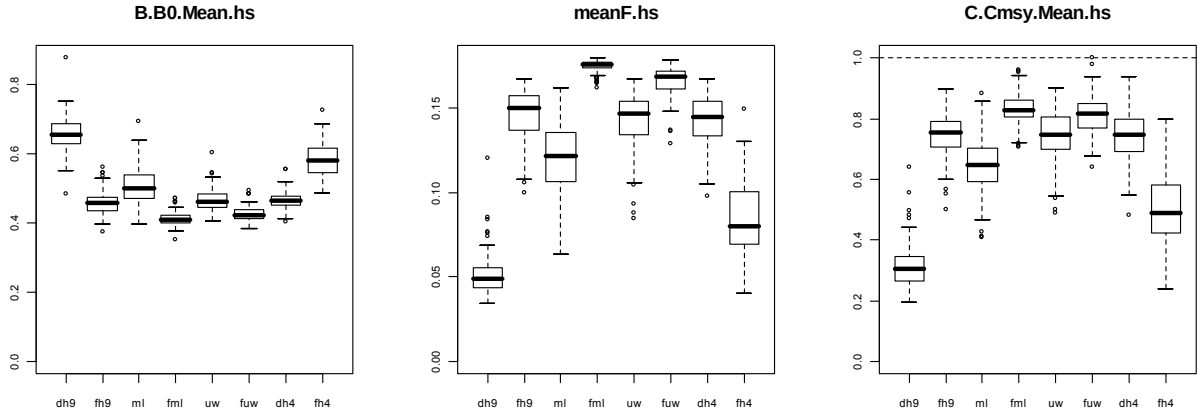


Figure 99. True selectivity is dome-shaped. Binary rules and classification trees either assume the correct (dome-shaped) selectivity or the incorrect (flat, prefix 'f.' to name) selectivity and results are plotted in these 'correct – incorrect' pairs for each rule. The rules are: classification tree dh9, binary rule based on mean length (ml), binary rule based on cpue (in weight, uw) and classification tree dh4. Measurement error was high on all components (HHH scenario).

When true selectivity is dome-shaped and true steepness is also low, results still show the same patterns; results are less conservative when flat selectivity is assumed.

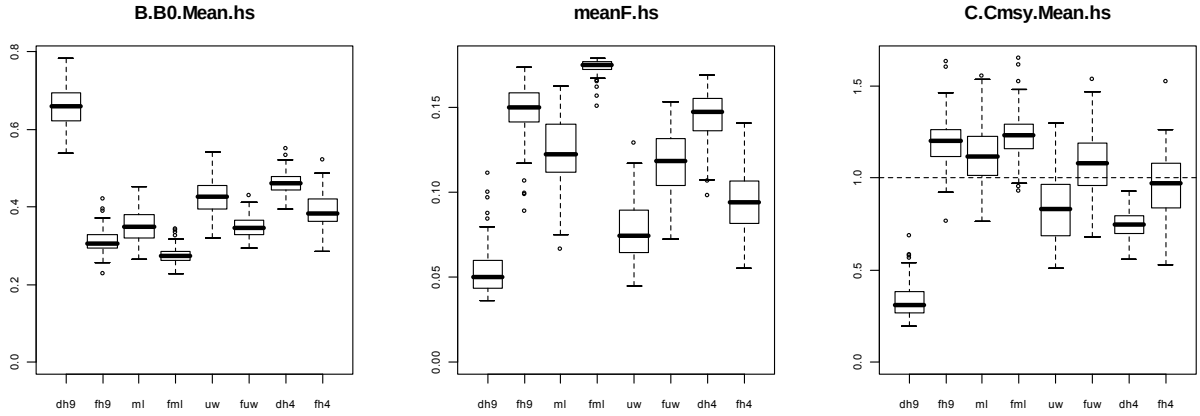


Figure 100. True selectivity is dome-shaped and steepness is very low (h=0.4). Binary rules and classification trees either assume the correct (dome-shaped) selectivity or the incorrect (flat, prefix 'f.' to name) selectivity and results are plotted in these 'correct – incorrect' pairs for each rule. The rules are: classification tree dh9, binary rule based on mean length (ml), binary rule based on cpue (in weight, uw) and classification tree dh4. Measurement error was high on all components (HHH scenario).

If the harvest rate, Fmax, is also changed in the decision rule, then results are more optimistic (more conservative), because Fmax=0.10 instead of 0.18 (Figure 101).

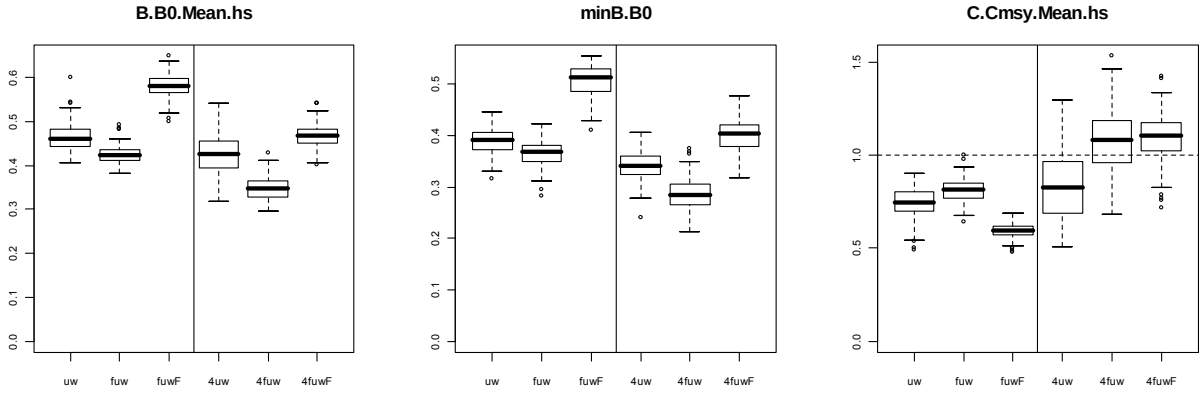


Figure 101. True selectivity is dome-shaped. The first 3 boxes are for true h=0.9, the second three are for true h=0.4 (prefix '4' to scenario name). In all cases it is the binary decision rule using CPUE. Assumed selectivity and Fmax for decision rules are: dome-shaped, Fmax=0.18 (uw, 4uw); flat, Fmax=0.18 (fuw, 4fuw); flat with Fmax=0.10 (fuwF, 4fuwF).

Although we have not explicitly also tested the traffic light and 10-40 decision rules, they are likely to show similar differences given that they too are based on the trigger points used (implicitly) in the binary and classification tree rules. One important caveat is that the above sensitivity trials were only based on changes in the trigger points for the indicators. The values of F_{max} were NOT changed. This is important, because F_{spr}-values are generally higher, possibly much higher, for dome-shaped than for flat selectivity. In this particular example, F_{spr40}=0.10 for flat selectivity and it is 0.18 for dome-shaped selectivity.

10.4.2.3. Revisiting classification tree fitting

We commented above on the fact that the dh9 classification tree rule appears to perform more conservatively than comparable single-indicator rules and the dh4 tree. This difference is very clearly shown in Figure 102. Note that the tree also performs conservatively under other scenarios for the 'true' dynamic, such as when true steepness is low. Although these rules are, of course, not tuned to achieve the same outcome for a chosen performance measure, it is the fact that it performs so differently from the other decision rules in this set, that seems strange.

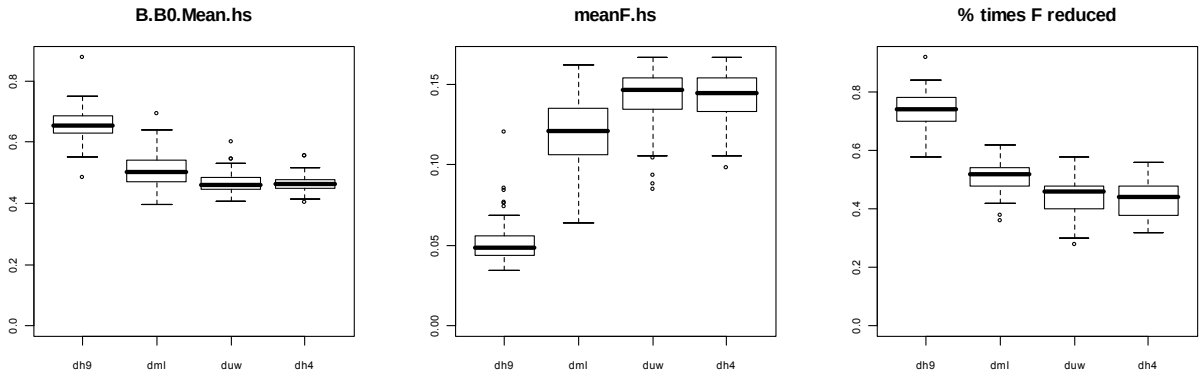


Figure 102. Performance of the dh9 classification tree rule compared to the binary rule with mean length (dml), binary rule with CPUE in weight (duw) and the dh4 classification tree. All the decision rules assume dome-shaped selectivity and true selectivity is dome-shaped.

We note that the dh9 tree has the 90th percentile of length, L90, as the first node of the tree. This indicator showed highly variable performance (i.e. in some cases very good and in other cases very poor) as a binary predictor of biomass (i.e. results in Section 9), and we consider that this may be part of the problem. We also recall that the performance of L90 was much more different than that of the other indicators in scenarios that were run with approximately constant F rather than with widely varying F.

A new tree was generated by rerunning one of the simulations to first generate a new dataset. Instead of allowing the harvest rate to vary widely by using a sine wave, the harvest rate was only allowed to vary slightly (with a CV of 0.015) around a constant level of $0.18=F_{spr40}$. Other quantities and assumptions are the same as for the original dh9 tree, i.e. high steepness, dome-shaped selectivity and high measurement error on all components.

A comparison of prediction errors show that although both L90 and W90 have very low prediction errors under the widely varying harvest rate scenario, they actually perform worse than the other indicators under the (approximately) constant harvest rate scenario. It is also interesting that the 'proportion big fish' indicator (pb) performs better under the constant harvest rate scenario.

A tree fitted to these data, based on approximately constant F, therefore looks quite different from the dh9 tree based on widely varying F. Both are shown in Figure 103 for direct comparison. The new tree (dh9, Fconstant) now has the proportion big indicator as its first node and the 90th percentile of length does not appear in the pruned tree.

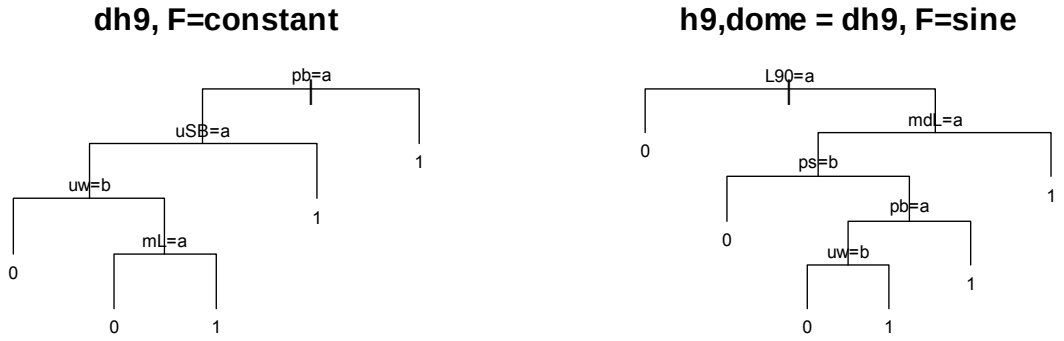


Figure 103. Classification trees fitted to simulated data based on F varying only slightly around a constant value (dh9,F=constant), and F varying widely according to a sine wave (h9,dome=dh9, F=sine). Both are based on the long-lived life history, dome-shaped selectivity, high measurement error levels and SPR40 reference points.

It is also interesting to compare prediction error for the two trees with the two associated simulated datasets (Table 21). It is unsurprising that the prediction errors are larger when the dataset has approximately constant F, because the contrast in the indicators would be much smaller. Note, though how much worse the prediction is for the original tree (based on sine F) with the data based on constant F (0.55), compared to its prediction error with the data to which it was fitted (0.15). The tree based on constant F, on the other hand, has only a prediction error of 0.31 with the data to which it was fitted and should therefore perform better as a decision rule.

Table 21. Prediction error for two trees and two datasets.

DATA >> TREE	data with F=sine	data with F=constant	
dh9, F=sine	0.15	0.55	tree with L90 at top
dh9.C, F=constant	0.22	0.31	tree with pb at top

This result suggests that it may well be better to better to choose groups of indicators for classification trees, and arguably for traffic light rules, from data generated with approximately constant harvest rate. There is, of course then the question about how sensitive results are to different choices of that constant harvest rate; a question which we (incorrectly) thought we had 'avoided' by using the widely ranging harvest rate time-series.

Having said that, we note that the same exercise was repeated for the low measurement error noise (LLL). In this case, the original tree had uSB (CPUE of spawners) as the first node, and L90 appeared as the third node in the tree. The new tree, fitted to data with approximately constant F still had uSB as the first node, but did not contain L90 at all. The differences in prediction error between the two trees and datasets were, however, much less pronounced than for the example given above. It may therefore only be in some specific cases where the choice of harvest rate pattern makes a substantial difference. In particular, we would suggest being cautious, and exploring different harvest rates values, if the upper percentiles (L90, W90) or the proportion big indicators appear at or near the top of a decision tree.

Results for the new classification tree, dh9.C, are quite similar to results of dh4 classification tree and the binary rules based on mean length or on CPUE. This is almost certainly because of the lower prediction error within the feedback simulation.

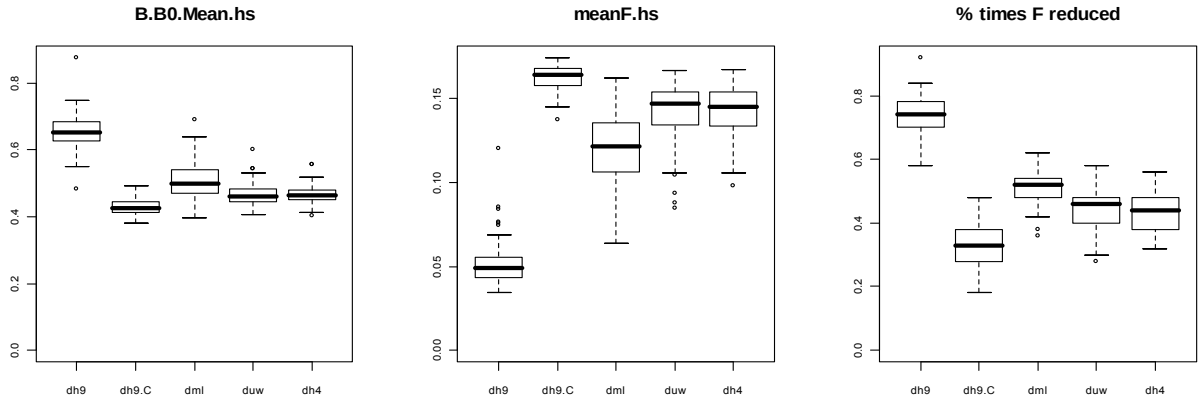


Figure 104. As Figure 102, but including the decision rule based on the new classification tree, dh9.C, fitted to data simulated under approximately constant F. The rules are: the dh9 classification tree rule (based on F=sine), the dh9.C classification tree rule (based on F=constant), the binary rule with mean length (dml), binary rule with CPUE in weight (duw) and the dh4 classification tree.

10.4.3. Effects of effort creep on decision rule performance

Results so far have not really shown any advantages of groups of indicators over single indicators. This is likely to be because (i) there is no implementation error in the scenarios considered so far, and (ii) no bias in any of the indicators. All indicators are potentially subject to some mechanism which could lead to bias. For example, a density dependent increase in growth rate, if undetected, would tend to lead to a positive bias in size-based indicators compared to their trigger points. So-called 'effort creep', which is essentially an increase in efficiency, leads to a positive bias in

CPUE-based indicators unless it is detected and removed from the time-series through standardisation.

In this section we simulate effort creep in the data, so that all CPUE-based indicators are positively biased. The catchability coefficient is increased by some percentage (we have used a modest figure of 2%) each year, so that catchability is almost 20% higher at the end of the series. Although the nominal effort is adjusted through the changes in the harvest rate obtained from the decision rule, the implemented (or 'real') effort is much higher because of the effort creep. This is reflected in the so-called implemented, or realised, harvest rate (F_{imp}). This way of modelling effort creep essentially implies some form of implementation error as well, given that the simulation assumes effort control rather than catch control.

This section is again split into two parts: the first considers effort creep when steepness is high (4.3.1) and the second, effort creep when steepness is low (4.3.2).

10.4.3.1. High steepness and Effort creep

It is useful to look at time-series plots comparing realisations with and without effort creep. The substantial increase in implemented (or realised) harvest rate under the effort creep scenario and a decision rule which used CPUE as an indicator is obvious in **Figure 105**. The relative biomass is therefore also lower when there is effort creep. The second row of panels in the same figure illustrate realisations for a decision rule based on mean length. As expected, the behaviour is much more similar[27], and the decision rule does not allow the harvest rate (even the implemented harvest rate) to increase as much because it keeps cutting back effort.

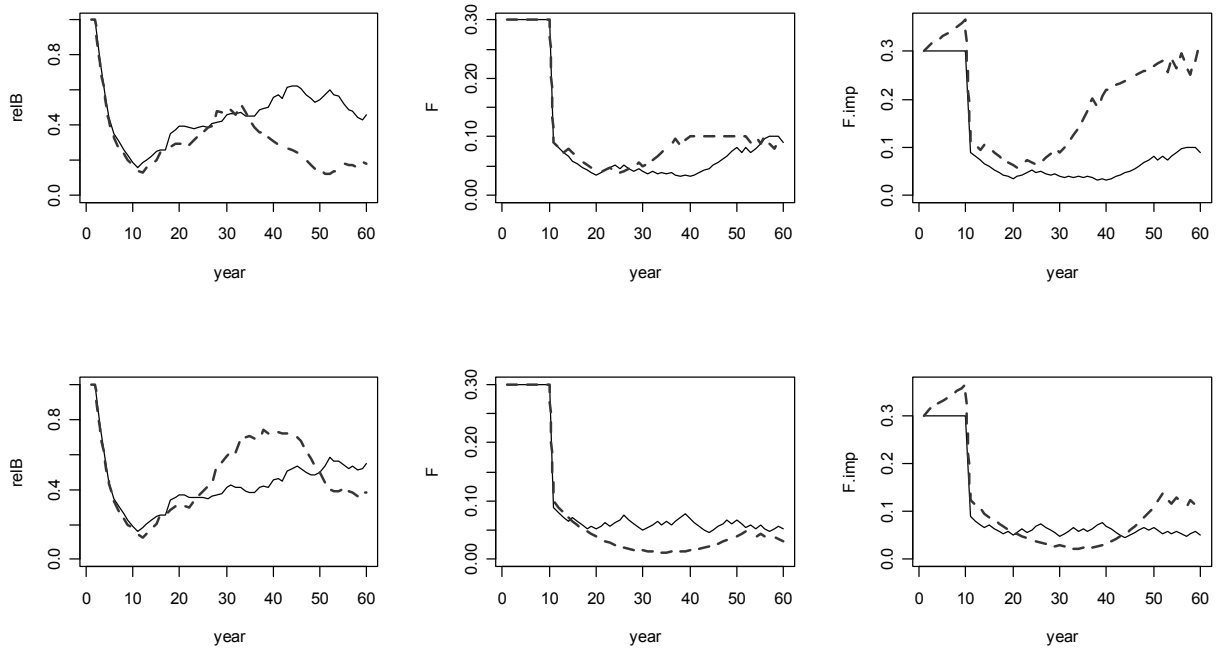


Figure 105. Time series plots for simulations with and without effort creep. The top row of panels are for the binary decision rule based on CPUE. Panels are relative SSB, the F that is generated by the decision rule, or assumed to be operating, and the F (F.imp) that is actually implemented; the solid line is for no effort creep; the dashed line is with effort creep. The bottom row is the same but for the binary rule based on mean length.

[27] The realisations are not identical because the catch numbers and therefore size-frequency samples are not identical.

Summary results for a range of decision rules under no effort creep and then with 2% effort creep per annum are shown in Figure 106. In all cases the relative biomass is lower when there is effort creep, than when there is no effort creep. The decision rule which only uses mean length, however, shows least difference between the 'with' and 'without' effort creep scenarios. Note also how this decision rule actually has a lower mean (decision rule) harvest rate under the effort creep scenario.

Although the h9 classification tree has a similar average harvest rate, the catches are still higher under effort creep, and this leads to lower biomass. It is worth recalling the design, or logic, of the h9 decision tree (Figure 78). It has mean length as its first node, but then branches to CPUE of spawners (if mean length is less than its trigger) or to CPUE of big fish (if mean length is greater than its trigger). This means that irrespective of the value of mean length, the CPUE-based indicators at the next level would tend to lead to wrong (positively biased) predictions.

When the same set of indicators are used as components of a traffic light decision rule, there is no such structure or hierarchy. The traffic light rule in Figure 106, for example, triggers a reduction in harvest rate when any 3 (or more) of the indicators in the set are below their trigger points. Given that the set of indicators consists of mean length, proportion small and three CPUE-based indicators, this does mean that the rule tends to be dominated by the CPUE-based indicators. The relative performance of a version based on 2 red lights triggering action, is likely to perform better.

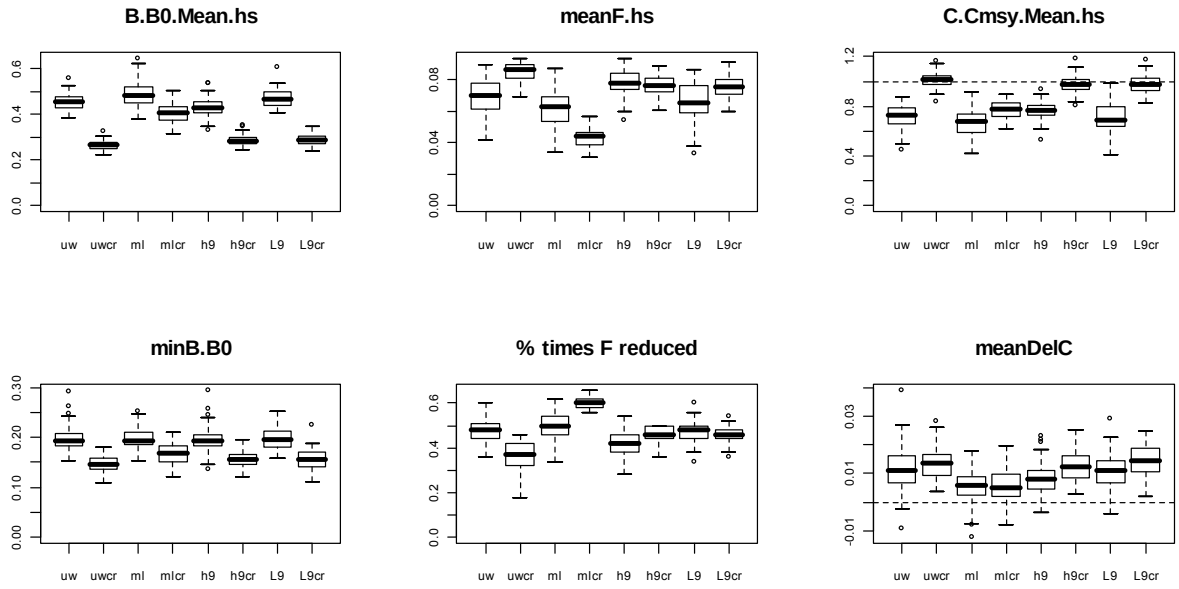


Figure 106. Summary results of six performance measures and for decision rules without and with effort creep, plotted in pairs; the version with effort creep has a 'cr' extension to the name. Decision rules are: binary rule based on CPUE in weight (uw and uwcr); binary rule based on mean length (ml and mlcr); classification tree h9 (h9 and h9cr) and traffic light rule based on indicator set h9 with 3 red lights triggering action. The harvest rates are those from the decision rule rather than the implemented Fs.

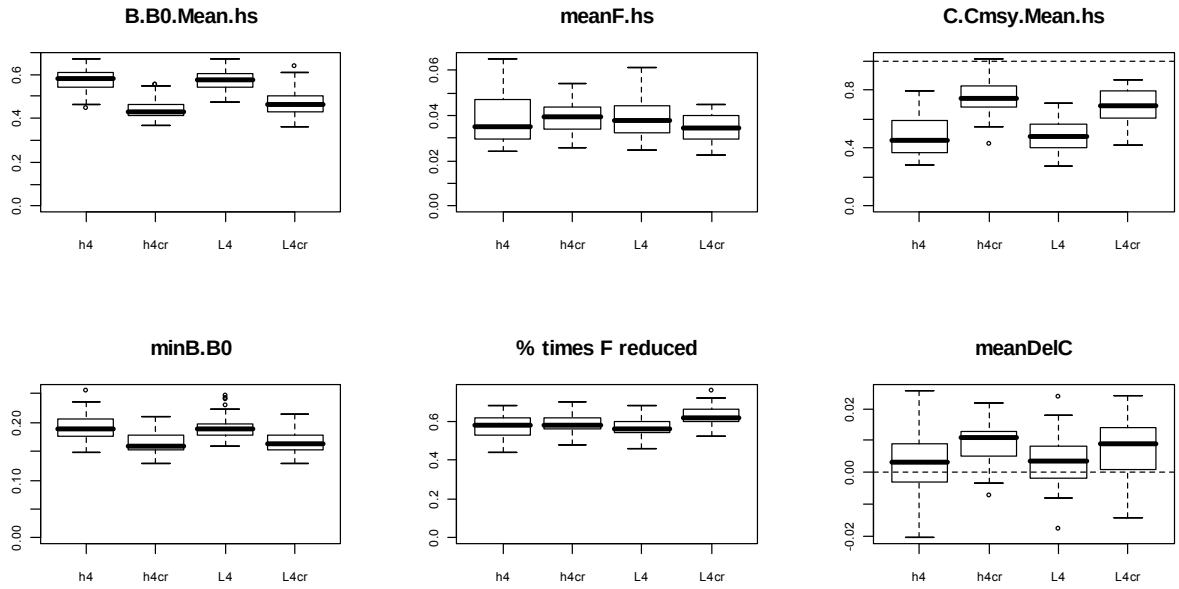


Figure 107. Summary results of six performance measures and for decision rules without and with effort creep, plotted in pairs; the version with effort creep has a 'cr' extension to the name. Decision rules are: classification tree h4 (h4 and h4cr) and traffic light rule based on indicator set h4 with 2 red lights triggering action. The harvest rates are those from the decision rule rather than the implemented Fs.

Results for the h4 classification tree and the traffic light rule based on the same set of indicators are plotted separately (Figure 107) because of slightly different implicit tuning levels (see text above). There is again, however, less difference between the performance of the classification tree and the traffic light rule when there is effort creep. Arguably, the traffic light rule performs slightly better because it leads to a slightly lower harvest rate and catch (than the classification tree) when there is effort creep. In this case, the indicator set is median length, 90th percentile of length, mean weight (i.e. 3 size-based indicators) and 2 CPUE-based indicators.

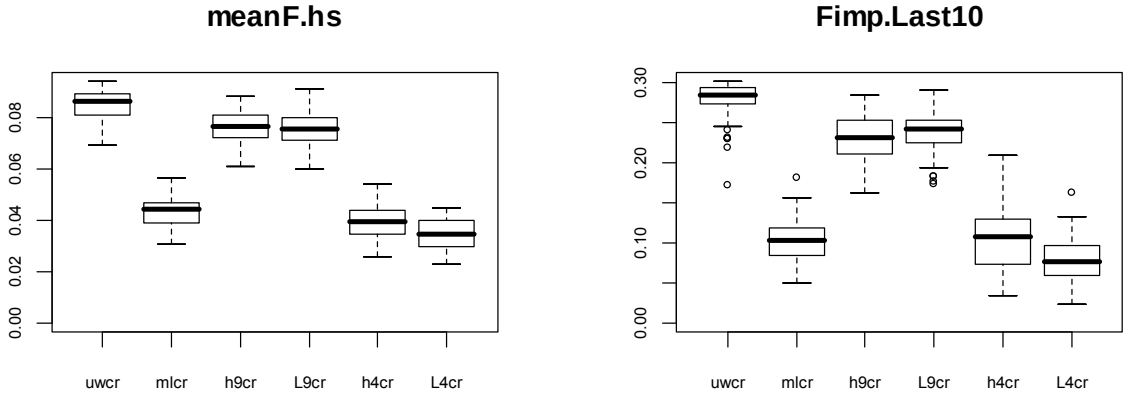


Figure 108. Summary results of the decision-rule harvest rates and implemented harvest rates under 6 decision rules and in the presence of effort creep (hence the 'cr' extension in the name). The rules are: binary rule based on CPUE (in weight, uwcr); binary rule based on mean length (mlcr); the classification tree h9 (h9cr); the traffic light rule based on the h9 indicator set and 3 lights to trigger action; the classification tree h4 (h4cr); and the traffic light rule based on the h4 indicator set and 2 lights to trigger action. The left panel is the mean harvest rate generated by the decision rule; the right panel is the mean implemented, or realised, harvest rate over the last 10 years of simulations.

It is also informative to just look at the harvest rates, both the decision-rule generated Fs and the actual implemented Fs. Recall that all these decision rules have an F_{max}=0.1. Figure 108 shows that it is only the binary decision rule based on mean length, and the classification tree and traffic light rule based on the h4 set of indicators that end up dropping the harvest rate sufficiently that the implemented harvest rate is also relatively low (e.g. around 0.1 rather than around 0.3). Results above illustrate that if there is effort creep (or, more generally, if one type of indicator becomes biased) then the choice of indicator(s) becomes much more relevant.

10.4.3.2. Low steepness and Effort creep

We have emphasised throughout this report that steepness is unlikely to be known. Here we look at the effect of effort creep when steepness is low (h=0.4). The same set of two binary rule, two classification tree rules and two traffic light rules are considered. Of main interest here is the change in performance of a decision rule between the 'with' and 'without' effort creep scenarios. Groups of decision rules are plotted together in Figure 109 and Figure 110, on the basis that the two groups are similarly tuned using the default parameter settings. The traffic light rules are the only ones which have been tuned to match the classification tree rule from which its set of indicators was taken. Under the 'no effort creep' scenario and low steepness, the difference in performance between this set of decision rules was already shown in Figure 92. Figure 109 shows no clear difference between, and therefore preference for, any of the three decision rules, namely the binary mean length rule, the classification tree 'h9' (based on mL, uSB, ub, ps and us) and the traffic light rule based on the same set of indicators. They all lead to lower average biomass, and lower harvest rates, but the drop in average biomass between the 'with' and 'without' effort creep scenarios is about the same. It is only in the extremes of the distributions (e.g. the minima) that there may be some differences, but this should be confirmed with a larger number of realisations than the 50 used here, before drawing any firm conclusions.

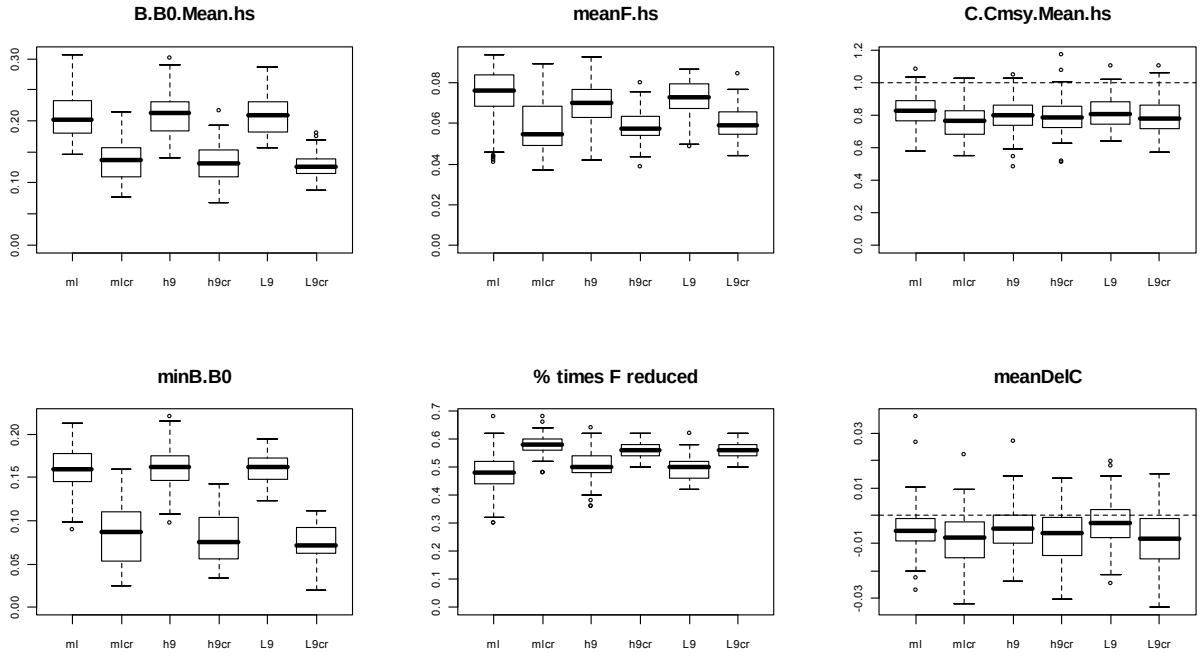


Figure 109. Low steepness in true dynamics. Summary results of six performance measures and for decision rules without and with effort creep, plotted in pairs; the version with effort creep has a 'cr' extension to the name. Decision rules are: binary rule based on mean length (ml and mlcr); classification tree h9 (h9 and h9cr) and traffic light rule based on indicator set h9 with 4 red lights triggering action. The harvest rates are those from the decision rule rather than the implemented Fs.

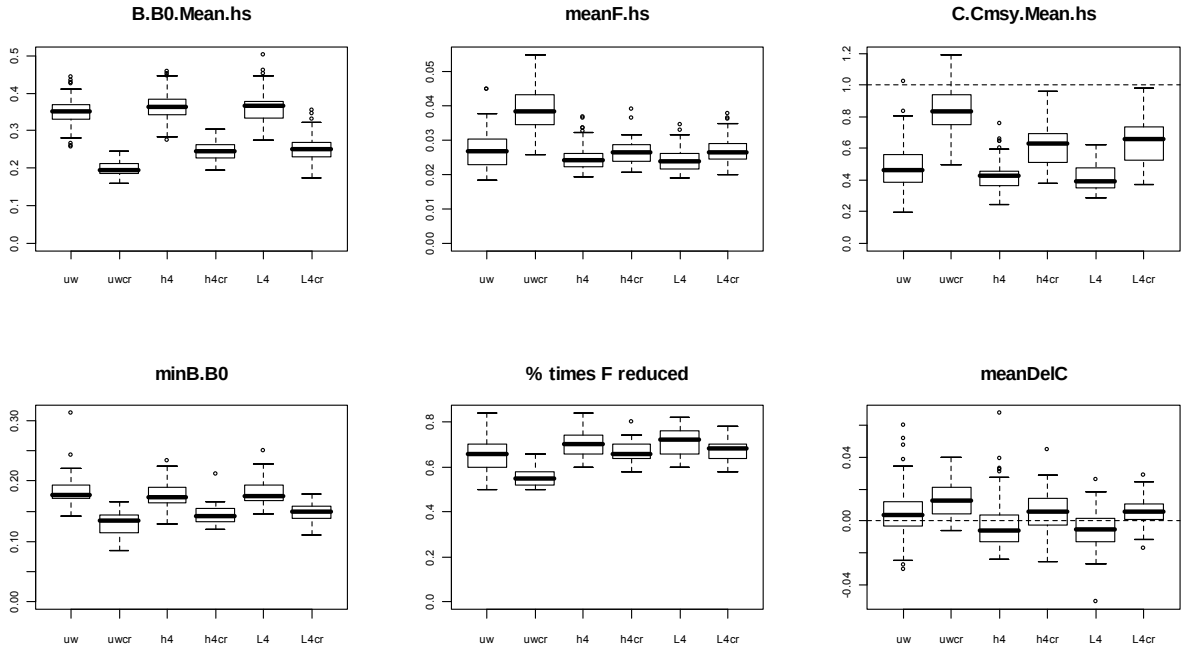


Figure 110. Summary results of six performance measures and for decision rules without and with effort creep, plotted in pairs; the version with effort creep has a 'cr' extension to the name. Decision rules are: binary rule based on CPUE in weight (uw and uwcr); classification tree h4 (h4 and h4cr) and traffic light rule based on indicator set h4 with 2 red lights triggering action. The harvest rates are those from the decision rule rather than the implemented Fs.

The second set of decision rules shown in Figure 110 also show only relatively small differences in the drop in average biomass. If we calculate the summary statistics for the ratios between each pair of realisations - one 'with' and one 'without' effort creep – we note, however, that the traffic light and classification tree rules perform slightly better than the binary CPUE rule. For example, the traffic light and classification rules both have a median ratio of 0.7 and a maximum of about 1. The binary CPUE rule has a lower median (0.6) and maximum (0.8), indicating more of a difference between the 'with' and 'without' effort creep scenarios.

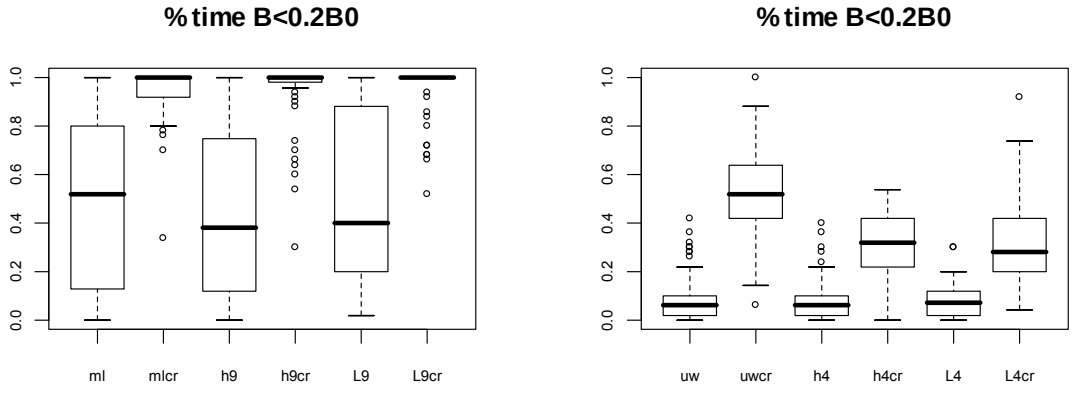


Figure 111. The percentage time that spawning biomass is below 20% of the unexploited level for the low steepness scenario, and for decision rules without, or with effort creep (plotted in pairs, with 'cr' indicating with effort creep). The left panel is for: binary mean length rule (ml, mlcr), classification tree h9 (h9, h9cr), and traffic light rule based on h9 set of indicators (L9,L9cr) and triggered by 4 red lights. The right panel is for: binary cpue rule (uw, uwcr), classification tree h4 (h4, h4cr) and traffic light rule based on the h4 set of indicators (L4, L4cr) triggered by 2 red lights.

This difference is arguably clearer when we look at the percentage time that spawning biomass is below 20% of the unexploited level (Figure 111). Note that results in the two panels are not directly comparable because rules are not tuned. Rules within each panel are similarly tuned with respect to relative biomass as noted above.

10.5. Discussion and Conclusions

The performance of feedback decision rules based on indicators – be they size-based or CPUE-based – is as much a function of the definition, or functional form, and parameters of the decision rule, as it is of the indicators. In many cases, the difference in performance between two types of decision rule based on the same indicator, is less than the difference in performance between the same type of decision rule based on two different indicators. It is, of course, possible to change the parameters in a decision rule in order to achieve a certain objective, and when comparing the performance of different decision rules this needs to be borne in mind in order to make fair comparisons.

Where assumptions are required to choose indicators or estimate trigger points, results suggest that it is safer to assume that selectivity is dome-shaped than flat. Care needs to be taken, however, not to then assume too high a harvest rate (or catch) given that dome-shaped selectivity would have a higher harvest rates (e.g. F_{spr40} or F_{msy}) than its related flat-topped version. It is also safer to assume a relatively conservative level of steepness in the stock-recruit relationship.

If simulations are used to find a set of indicators through the approach of fitting a classification tree, results also suggest that it is prudent to do this by simulating the data with an approximately constant harvest rate, though the sensitivity of outcomes to different values of harvest rate should be explored.

The structure of a decision rule also, however, affects its performance, and some structures have rather bad characteristics. Decision rules based on just the slope of, or year-to-year change in, an indicator do not perform well, because it does not lead to rebuilding if the stock is already over-exploited, nor to increased exploitation if the stock is lightly exploited when the decision rule is first implemented. If such a rule also has a built in maximum harvest rate or catch, the rule performs better provided that the maximum has not been set too high.

A decision rule based on whether an indicator is above or below some trigger point, and with a built-in maximum harvest rate (or catch) seems to perform well. Similar rules based on groups of indicators either using classification trees or a traffic light approach also seem to perform well. In the case of traffic light type decision rules, the performance of the rule can be particularly sensitive to the number of 'red lights' (one of the tuning parameters) that trigger action.

Given these observations, it is likely that decision rules based on considerations of both the level and the slope of an indicator, would perform even better, because they may avoid some of the unnecessary fluctuations in catch or effort.

The familiar '20-40' type decision rules can easily be adapted to be used with indicators. Care needs to be taken when choosing the limit and target points that define such a rule, particularly when using size-based indicators. The reason is that if they are too close together, the decision rule can oscillate between a closure (or very low harvest rate) and the maximum harvest rate. Not only does this lead to very high interannual variability in the catch and effort, but this can also lead to large oscillations in the population size. One approach to reducing the likelihood of oscillations is to ensure that the target and limit points in the rule are not too close together. A

'10-40' rule, for example, has a lower limit point and is therefore less likely to cause oscillations. Note, however, that the lower limit does not necessarily imply poorer (or less conservative) performance by a decision rule.

Results suggest that the binary type rules actually perform better than the '20-40' or '10-40' type of rule, from the point of view of catch variability. This is partly due to the fact that the binary rules have a limit on the percentage change in catch from one time-step to the next. In a real set of testing comparisons, one would first tune the decision rules to achieve similar performance with respect to one performance measure, and then consider the trade-offs between the other important performance measures.

For decision rules which have a built in maximum catch or harvest rate, there will always be the danger of setting this level too high. Some decision rules perform better than others when the maximum harvest rate is set too high. For example, binary rules and classification trees are better than the 'up-down' rules with a maximum, because they show less of a difference between average biomass under the correct and the 'too high' maximum F.

The comparisons between 20-40 and 10-40 versions of the decision rule illustrate an important point in our view. First we noted how the 20-40 rule often flip-flops between F_{max} and F_{min}, particularly since the limit and target points are quite close together (more so for ml than for CPUE). The 10-40 version of the rule reduces the frequency of this extreme flip-flopping, although the frequency of changes (of any magnitude) to the harvest rate is slightly higher. Nonetheless, the lower choice of limit has only a small effect on the performance in terms of average biomass or minimum biomass. Besides, the value of F_{max} can be reduced slightly to offset this. The main point to emphasise here is that the objectives and reference points that might be desirable for biomass need NOT be directly translated to the trigger points for indicators used in a harvest rule. In this particular case, for example, it is in our view more important to avoid the extreme variability in catches by shifting the trigger points further apart by choosing a lower limit point (one can of course also choose a higher trigger point, and that may mean that F_{max} can be slightly increased to achieve the same objectives in terms of average relative biomass.

When it comes to robustness to steepness, the performance of decision rules are in line with the prediction error results for indicators in Section 8. In all cases the relative and minimum biomass are lower. All the rules (binary, classification trees, traffic light and 10-40) respond with a reduction in harvest rate. The CPUE-based decision rules, however, outperform both mean length-based rules and rules based on groups of indicators. Rules based on CPUE perform best in the sense that they show the least difference in performance measures such as average biomass and minimum biomass when steepness is very low compared to when it is very high. This is the case for a 10-40, as well as a binary rule based on CPUE.

Even groups of indicators, either in traffic light or classification trees, are out-performed by CPUE as a single indicator. Groups of indicators that performed well under high steepness (because they were obtained from analyses of simulated data based on high steepness) do not perform as well when steepness is low. They still achieve a drop in the harvest rate, but not very much. On the other hand, groups of indicators that were obtained from analyses of simulated data based on LOW steepness perform well when steepness is low, but are overly-conservative when steepness is actually high.

This underlines the importance of finding ways to obtain reliable information on catch, effort, and any additional quantitative or qualitative data which may help standardisation of CPUE. Fishery independent indicators of relative abundance or population density would also be excellent alternatives to strive for.

11. Feedback dynamics of simple catch and harvest rate decision rules: a theoretical investigation

(M. Basson)

Summary

This Section relates to the objective of developing decision rules based on indicators, but here the focus is on the form, or design, of the decision rule rather than the specific indicator. Even a very good indicator can perform poorly if it is imbedded in a decision rule which has poor characteristics. What are poor characteristics? Decision rules which do not lead to rebuilding of an over-exploited stock, or ones which do not lead to any increases in catches of a very lightly fished stock, have poor characteristics. Also, decision rules which lead to large fluctuations in abundance (large decreases, followed by large increases and repeating of this pattern) have poor characteristics because the catches will end up fluctuating widely, and the stock could be at a serious risk of collapse when the abundance is at its lowest levels.

The approach used in this Section is rather technical because of its theoretical and mathematical nature. I have therefore attempted to give a flavour of the problem, the approach and the results in non-technical terms both here and in the Introduction (below). The approach used here is borrowed from the theoretical ecology literature, and offers a relatively simple and elegant way of extracting some of the important characteristics of a decision rule without having to build a detailed operating model and conduct a large number of simulations. This approach is therefore suggested as a useful first step to determine whether a particular type of decision rule is likely to perform well or not.

The results and main insights gleaned from this investigation are also rather technical. In non-technical terms, the results provide guidance about how to choose the form of a decision rule. In other words, how the rule translates the observed value of an indicator (or values of several indicators) into a total allowable catch (TAC) or total allowable effort (TAE). Results also provide insights about how to choose the parameters involved in this translation to ensure that the decision rule does not inadvertently drive the stock to very low levels, or introduce very large changes in the TAC or TAE from year to year. In particular, a DR which first allows the stock to be over-exploited then rebuilt to almost an unexploited level and then repeats this pattern, is to be avoided. The analyses presented in this section can identify, or diagnose, when a given decision rule is likely to exhibit such behaviour.

At a more technical level, the insights gained by this investigation can be described as follows:

- A decision rule (DR) which simply responds to increases or decreases in abundance (or an indicator of abundance), is very unlikely to lead to rebuilding a depleted stock, or to increased catches of a lightly fished stock.
- Such a DR will either keep the stock approximately at the same level as at the start of implementation of the DR, or it could lead to unexpected large changes in abundance.
- Apparently subtle differences in the formulation of the DR are important to the dynamics and stability conditions of the system.
- The multiplicative version (examples 3, 6; see Table 1) of a DR (where the catch in the following year is obtained by multiplying the catch in the current year by some factor based on the value of an indicator, say) has better characteristics than an additive version (where the catch in the following year is adjusted by adding or subtracting an amount based on the value of an indicator, say). The multiplicative version is less likely to lead to unexpected oscillations.

- There are also clear differences between the behaviour of a DR which is specified in terms of catch and one which is similar in all respects, but specified in terms of effort, or harvest rate (F). The details of the differences depend on how the rule is specified.

One common type of decision rule is the so-called 20-40 rule, which can also be formulated as a smooth sigmoid curve (see Figure 6 below for the shape and interpretation of this type of decision rule). When a sigmoid decision rule is constructed to mimic the familiar 20-40 or 10-40 type decision rule, the main insights are the following:

- This type of rule has a built-in target and equilibrium biomass which may or may not be stable. If it is not stable, the DR can lead to unexpected large oscillations or unpredictable behaviour.
- Depending on the parameters that specify the rule and the rate of increase in the stock, there may be multiple equilibria, which could mean that the stock decreases to a lower level than intended.
- This type of DR can also lead to complex behaviour, even for a stock with low or modest rate of increase, if the maximum catch (C_{max}) or harvest rate (F_{max}) is set too high, if the inflexion point ($\widetilde{B}$) is set too high, or the sigmoid curve is too steep (α is high). This can happen if the target and limit reference points on which the rule is based are too close together.
- This suggests that the maximum catch or harvest rate should be set conservatively, but care should be taken not to set $\widetilde{B}$ (or the Biomass reference points) too conservatively (too high) if point and limit cycles are to be avoided.
- For this type of rule, the catch-based version has better stability characteristics than the harvest rate version for similar specifications in other respects.

It is not surprising that the population model itself, for example the Schaefer or Fox model, makes a difference to the dynamics and stability characteristics, since the dynamics for the unharvested versions of the models differ. It is also not surprising that oscillatory behaviour, point or limit cycles may not matter if they are small relative the stochasticity in the system. It is also, however, easy to show that there may be cases where the cycles would not be obscured by the stochasticity. This means that the variability in the catches would not just be random, but follow cycles, possibly with a much wider range than natural variability alone would induce.

This work shows the potential value of quick, simple and very basic simulation (or analytical where feasible) exploration of proposed decision rules before implementation, or at the very least, during the first year of implementation. Not all feedback behaviours, of even such simple systems as those considered here, are intuitive or easy to foresee.

Finally, it is interesting to note that this work has pointed to the likely culprit of the observation which triggered this work: cycles in a simulated age-structured population model following the introduction of a decision rule. The decision rule was based on mean size as indicator and constructed like a 20-40 rule. The two reference points for mean size were based on the mean size coinciding with a harvest rate that would give 20% or 40% of unexploited spawner per recruit respectively. Since mean size changes very slowly as the harvest rate changes, the two reference points tend to be very close together. This in turn implies a high α value in the formulation used here, which tends to lead to complicated behaviour.

11.1. Introduction

Preamble
The work in this Section came about when we observed, during early trials of simple decision rules based on stock status indicators, the tendency of cyclical behaviour (that is, increases in abundance followed by decreases and then a repeat of this pattern over time) in the population abundance. Although the original set of objectives did not explicitly suggest exploration of different types of decision rule (the focus is more on the different indicators than the form of the decision rule), it has become evident that the form of the decision rule can play a very important part in the behaviour of the fishery and fish population system. Even a very good indicator of biomass is likely to have poor performance when embedded in a badly designed decision rule. This Section relates to objectives five and six (see Section 4. Objectives) of the project, with the focus being on the form of the decision rule rather than the specific indicator.

Background
One integral part of a fisheries harvest strategy (management strategy or procedure) is a decision rule which defines the catch level or a harvest rate, based on one or more inputs (Section 6 explains what a decision rule is and illustrates several different types of decision rule). Over the past two decades, the field of "management strategy evaluation" (MSE), where a harvest strategy is trialled, and its performance evaluated within a simulation framework, has grown. Evidence for this can be found in conference proceedings and special publications, such as Payne (1999); several papers therein provide detailed descriptions of the components of a harvest strategy and MSE.

Evaluating the performance of a management decision rule (DR) in a simulation framework is seen as a key part of the process to develop a management strategy for a fishery. Where a modelling framework and software already exists, testing may be relatively fast. In many cases however, this is not, or may not be, possible prior to deadlines for implementation. Also, for some species there may be no assessment, so that very simple DRs tend to be proposed. In such cases, the impetus for testing the rule before implementation may be somewhat diminished. Common sense may even suggest that simple rules are likely to behave well, but this is not always the case; simple rules can behave very poorly indeed, as is shown in this Section. In fact, in Section 10 we showed that the simple 'up-down' rule (increase catches when the indicator increases and decrease catches if the indicator decreases) has poor properties; it does not rebuild an over-exploited stock.

The problem
In the course of general simulation testing of decision rules based on stock indicators such as CPUE or mean size, and in the context of an age-structured population model, and including the work reported in Section 10, it became clear that:
(a) some DRs are ineffective at rebuilding a depleted stock or increasing catches in a lightly fished stock,
(b) some DRs can induce strong oscillations in the simulated population dynamics and
(c) the behaviour between catch-based and harvest rate (F)-based DRs can be quite different.

The issues mentioned under (a) and (b) above are obviously poor characteristics of a decision rule. Issue (c) is more of a warning that a rule which performs well when formulated in terms of catch is not guaranteed to perform well when formulated in terms of effort (or harvest rate).

The main aim of this section is to explore the theoretical characteristics of a system being harvested by a feed-back decision rule in order to gain an understanding of the likely dynamics of

different DRs and to identify potential pitfalls when designing simple decision rules. Such an investigation is particularly valuable if it can be done in a general way rather than for a specific set of parameter values in the DR, and with specific parameter values for the stock and fishery. I attempt to answer questions such as:

i) under which kinds of stock dynamics and decision rules might one expect to see oscillations or complex dynamics such as limit cycles or even chaotic behaviour?
ii) what are the differences in behaviour between catch- and F-based decision rules?
iii) can one identify good 'design features' for decision rules?

The approach to answers and solutions

Although one could choose to find this out by testing the behaviour of any given decision rule in a simulation model, there is in fact a mathematical technique (called local stability analysis) which allows one to 'diagnose' some of these likely problems in a simpler, more direct and more general way. This approach is similar to that used by May and others in the 1970s to study the theoretical characteristics of predator-prey systems and small ecosystems (e.g. May 1974; May and Oster, 1976; Hassell and Comins 1976). The local stability characteristic of equilibria in ecosystems provided insights into possible behaviours of such systems and the relationship between the system's behaviour and the values of key parameters, such as the rate of increase of the populations involved. May and Oster (1976) drew attention to the complicated behaviour (bifurcations, limit cycles and even chaos) that some such systems can display. Although early work focused on predator-prey systems, later work also considered harvested systems using the similar theoretical techniques (e.g. Getz and Haight, 1989; Basson and Fogarty, 1997).

There is a large and growing body of work on the characteristics of DRs based on simulation studies and done within the 'management strategy evaluation' context (e.g. De la Mare, 1986; Bergh and Butterworth, 1987; Punt, 1992; Polacheck *et al.* 1999, Geromont *et al.* 1999). The operating models used in these simulation studies are often complicated and are conditioned to the historic data from a particular fishery. The rules themselves are often also been complicated. For example, a decision rule might involve fitting a stock-production model to catch and CPUE data and then using estimated parameters to generate a total allowable catch (TAC) or harvest rate (Butterworth and Mori, 2003; Polacheck *et al.* 2003.). Although these simulation studies are essential for developing a management procedure that is robust to uncertainty, and provide valuable insights into the particular system being studied, it is much harder to draw general conclusions or gain generic insights from such studies. It is also not feasible to attempt analytical stability analyses on such complex systems.

The fact that catch and harvest rate strategies have different characteristics is well known, for example, from the work of Beddington (1978) who compared a constant catch policy to a constant harvest rate policies. General comparisons between the dynamics of similarly specified catch-based and F-based decision rules have not however been undertaken as far as I am aware, and it is therefore important, and relevant, to consider both types of decision rules.

This Section uses a single-species deterministic harvested system with relatively simple, but not unrealistic (particularly for data-poor fisheries) decision rules. Each system is then analysed to identify the equilibrium points and to characterise their local stability characteristics. The general methodology for doing this analysis is first set out. Several examples of decision rules, both in terms of catch and in terms of harvest rate, are then considered. Results from this exercise provide insight to the potential behaviour – particularly poor or undesirable behaviour – of different types of decision rules. The implications of stochasticity is briefly touched on at the end, as well as the likelihood of observing the complex dynamics in reality.

How does this work relate to Section 10?
By now the reader may well be asking why this Section is necessary given that decision rules have already been explored (quite extensively) in Section 10. In broad terms, the work in Section 10 is relatively specific in its exploration of life-history type (e.g. short-lived or long-lived), specific in terms of its choice of parameters in each decision rule (e.g. 10% change in catch or F per unit of change in the indicator). On the other hand, many different indicators are considers as candidates in any given type of decision rule.

In contrast, this Section uses a general formulation of population model and therefore includes a wide range of life-history types. It also uses a general formulation for the parameters in each decision rule and therefore, again, includes a wide range of rules – all of the same 'design', but with slightly different details. Finally, the assumption here is that a 'good' indicator of biomass (unbiased and of high precision) is used in the decision rule

The results obtained from analyses in these two Sections (10 and 11) are also very different. Section 10 provides results on the likely performance of decision rules (under different circumstances) in terms of the expected levels of catch and biomass after some time-period, and risks (e.g. probability of biomass being below some reference point). Although results in this Section are also focused on the likely biomass level(s) a system would, or could, end up at, it pertains more to the likely time-series dynamics of a system, for example, whether the system stays at one level or whether it fluctuates widely. In some sense, the interest is also about whether the system's behaviour can be controlled through the decision rule (i.e. the level of harvest), or not.

11.2. General methodology

All the analyses presented here are based on stock-production models of population biomass (B), and primarily on the logistic (or Schaefer) stock-production model in its discrete time form. A harvesting component is either specified as a catch (C) or as a harvest rate (F):

$$\Delta B_{t+1} = B_{t+1} - B_t = rB_t(1-B_t/K) - C_t \quad (1a)$$

$$\Delta B_{t+1} = B_{t+1} - B_t = rB_t(1-B_t/K) - F_tB_t \quad (1b)$$

where r is the rate of increase and K the carrying capacity, or unexploited biomass.

The catch or harvest rate is defined by a simple decision rule, generally defined as a function of the biomass or change in biomass. The details of the decision rules are presented below, but in general they can written as functions (g or h) of biomass:

$$\Delta C_{t+1} = C_{t+1} - C_t = g(\mathbf{B}) \quad (2a)$$

$$\Delta F_{t+1} = F_{t+1} - F_t = h(\mathbf{B}) \quad (2b)$$

Here **B** is meant to indicate either B_t or ΔB_{t+1} (i.e. B_{t+1} and B_t). In principle, longer subsets of the time-series of recent biomass can, of course, be used. The notion is that there is some indicator (e.g. CPUE) which is assumed to be proportional to **B**, or related to **B** in some known way. In reality, it is more likely that a decision rule would be based on an index of abundance (rather than a direct measure of biomass). The decision rule could therefore also have been formulated as a function of and index, I_t, instead of B_t, and the assumption that $I_t=qB_t$, where q is a constant of proportionality. For the purpose of this Section, the formulation in terms of I_t really only introduces another nuisance parameter, so the formulation in terms of biomass is preferred.

Although this appears to lead to a two-component dynamic system (biomass and catch, or biomass and harvest rate), all the examples shown here, can also be rewritten in terms of

biomass only, sometimes with an additional time lag in the difference equation. This is shown in sub-section 3 below. In the rest of this sub-section the system is formulated as a two-variable system to describe the general methodology.

The equilibrium points of the system (B*, C*; or B*, F*), where $B_{t+1}=B_t$ and $C_{t+1}=C_t$ (or $F_{t+1}=F_t$) are determined by setting $\Delta B=0$ and $\Delta C=0$ or $\Delta F=0$. Recall that the unexploited Schaefer production model has two equilibria, one at B*=0 and the other at B*=K. Local stability analysis is then used to infer the properties of the system in the neighbourhood of its equilibria. There are many excellent references on this type of analysis (e.g. May 1974, May and Oster 1976 and references in these documents), and details are not repeated here. In short, local stability analysis is based on assuming a small perturbation to an equilibrium point. A Taylor expansion of the basic equation around this equilibrium, with terms of second or higher order discarded, is then used as a linearised approximation of the population perturbation. The linearised approximation describes the population dynamics in the neighbourhood of the equilibrium point. The quantities that determine whether the perturbation dies away or grows are strongly related to the partial derivatives of the population equation(s), as shown below.

Consider a general system of population biomass and catch:

$$\Delta B_{t+1} = G_1(B_t,C_t) \qquad (3a)$$

$$\Delta C_{t+1} = G_2(B_t,C_t) \qquad (3b)$$

with equilibrium points (B*, C*) where $G_1(B^*,C^*)=0$ and $G_2(B^*,C^*)=0$. The first step in performing a local stability analysis around the equilibrium (B*, C*) is to construct the matrix, $A = [a_{ij}]$, of partial derivatives, evaluated at this equilibrium (indicated by the * in the notation):

$$a_{11} = (\partial G_1/\partial B)^* \qquad a_{12} = (\partial G_1/\partial C)^*$$

$$a_{21} = (\partial G_2/\partial C)^* \qquad a_{22} = (\partial G_2/\partial C)^* \qquad (4)$$

The two eigenvalues of A follow from the determinant, $\det|A-\lambda I| = 0$, i.e.

$$(a_{11}-\lambda)(a_{22}-\lambda) - a_{12}.a_{21} = 0 \qquad (5)$$

Let λ_1 and λ_2 be the two eigenvalues. The behaviour of the system is dominated by the (so-called) dominant eigenvalue, $\lambda = \lambda_{max} = \max(\lambda_1, \lambda_2)$. The stability criterion for this system of difference equations is:

$$|\lambda + 1| < 1 \quad \text{or} \quad |\mu| < 1 \text{ where } \mu = \lambda + 1 \qquad (6)$$

If one of the eigenvalues, say λ_1, is equal to zero, it means that $|\lambda_1 + 1| = 1$, which implies neutral stability when that eigenvalue is the dominant eigenvalue. The sign of the dominant eigenvalue determines whether the response to a perturbation is exponential (+) or oscillatory (-); this is discussed and illustrated further below.

In some cases it is not possible to find the equilibria analytically, and it is then necessary to investigate the stability characteristics numerically. In these cases, simple simulation models are used, and results from the time-series of population and catch illustrate the characteristics of the system under different parameter-values.

The examples considered below also include a case where, for some parameter combinations, there are multiple equilibria, and where classic bifurcations occur, along the lines found by May and Oster (1976) for the logistic model with very high rate of increase (r>2.69) and for simple predator-prey systems. In these examples of decision rules, limit cycles develop and, depending on the eigenvalues as before, such limit cycles may, or may not, be stable.

All analyses are done in the context of deterministic systems. Fisheries and fish stocks are manifestly not deterministic, but the insights obtained from considering a deterministic system

are still highly relevant and informative. Further comments on the implications of stochasticity are made in the Discussion sub-section.

11.3. Catch decision rules based on change in abundance

I consider three examples of a catch-based decision rule, with only subtle differences between them. The first example uses the absolute difference in biomass (or an indicator) from one time-step to the next to adjust the catch. This is essentially the same as the 'up-down' rule used in Section 10. The second example uses the relative difference in biomass (or an indicator) in an additive way, and the third example uses the relative difference in biomass (or an indicator) in a multiplicative way. The examples are first described, and results for all three are then compared. In the rest of this section, it is taken as understood that 'biomass' used in a decision rule essentially refers to an indicator of biomass.

It is possible to reduce the number of parameters in the system by rescaling the system relative to the carrying capacity, K, and defining a new variable $x_t = B_t/K$. Here, K has explicitly been retained to maintain the direct link to the stock production model being used. This makes the descriptions and interpretation of results somewhat less abstract. In the examples, K=100 has been used for illustrative purposes. Given that the unexploited stock would be at a level B=K, it would therefore be appropriate to think of other levels of B (e.g. equilibria B*) as %'s of the unexploited stock.

Example 1

Simple rules, particularly in the absence of reference points or knowledge about where the 'target' catch should be, are often proposed as simply responding to a change in an indicator, such as CPUE or a survey index of abundance (see e.g. Butterworth and Geromont, 1997). The notion is that if the indicator is decreasing, then catches should be decreased, whereas when it is increasing, catches could be increased. The definition of an increase or decrease is obviously flexible (e.g. one could use the slope over the most recent 3 years, or 5 years as a measure of increase or decrease). Here the difference in abundance from one year to the next is used as a measure of increase or decrease, to illustrate some of the fundamental dynamics of such a system. The decision rule can be formulated as:

$$C_{t+1} = C_t + \alpha(B_{t+1} - B_t)$$

which implies

$$\Delta C_{t+1} = C_{t+1} - C_t = \alpha(rB_t(1-B_t/K) - C_t) \qquad (7)$$

As noted above, the scaling between the indicator and biomass is ignored for simplicity and the decision rule is formulated directly in terms of biomass. Equation (7) clearly does not guarantee only positive catches, particularly if α is large. This is an obvious weakness which would need to be removed if such a DR were to be used in reality. It does not, however, hinder the further analyses for illustrative purposes, particularly since we are interested in the characteristics of the equilibrium (or multiple equilibria).

It is clear from equation (7) that at an equilibrium where $(B_{t+1} - B_t) = 0$, the catch will be constant from one year to the next. The unharvested population (equation 1a without the catch term) has equilibria at the trivial point, B=0 and at B=K. The set of equations (1a and 3) do not, however, have one (or even two) unique equilibrium solution(s), by many. For any level of biomass, B*, there is a constant catch,

$$C^*=rB^*(1-B^*/K) \qquad (8)$$

which would keep the system at the equilibrium (B*,C*).

This already hints at the first problem with such a DR: there is no implicit or explicit target biomass in the system, and the decision rule will not rebuild a heavily depleted stock or allow for increases in catches of a lightly harvested stock. If the stock is at a desirable level when the harvest strategy is introduced, it would tend to keep the stock at that level provided that equilibrium level is stable. This point is now further illustrated through stability analysis.

The matrix of first derivatives of the 2-variable system (equations 1a and 3, variables B and C), evaluated at the equilibrium (B*, C*) is given by:

$$A = \begin{bmatrix} r - 2rB^*/K & -1 \\ \alpha(r - 2rB^*/K) & -\alpha \end{bmatrix} \quad (9)$$

Note that the 2nd row of A is just a multiple of the first row which already indicates that one of the eigenvalues will be 0. The characteristic equation of A (i.e. det|A- λI|) is:

$\lambda(\lambda + (\alpha - r(1-2B^*/K))) = 0$

so that the eigenvalues are:

$\lambda_1 = 0$

$\lambda_2 = r(1-2B^*/K) - \alpha = (r - \alpha) - 2r/K.B^*$ (10)

Stability conditions require that $|\lambda + 1| < 1$ for a perturbation to die down. Note that λ_2 is a linear function of B* with a maximum value of $\lambda_2 + 1 = (r- \alpha) + 1$ at B*=0, and a minimum value of $\lambda_2 + 1 = (-r- \alpha) + 1$ at B*=K. This is illustrated in Figure 112 below.

As hinted above, this formulation of the decision rule, simply means the introduction of additional time-lags in terms of biomass. If we assume that catch and biomass in the first time-step are C_0 and B_0 respectively, then by repeatedly applying equations 1a and 3, we find that

$$\Delta B_{t+1} = \sum_{i=0}^{t} (-\alpha)^i rB_{t-i}(1 - B_{t-i}/K) + (-1)^{t+1}\alpha^t C_0 \quad (11)$$

which shows the dependence of the system on the starting values (B_0,C_0).

Example 2

If the catch is updated as a function of the relative change in biomass

$C_{t+1} = C_t + \alpha(\Delta B_t/B_t)$

$\Delta C_{t+1} = \alpha(r(1-B_t/K) - C_t/B_t)$ (12)

the basic dynamics are similar. There are multiple equilibria where $C^*=rB^*(1-B^*/K)$ and one of the eigenvalues is again zero ($\lambda_1 = 0$). The other eigenvalue, λ_2 is given by:

$\lambda_2 = r(1-2B^*/K) - \alpha/B^* = r - 2r/K.B^* - \alpha/B^*$ (13)

Now the function of λ_2+1 is not linear in B* because of the 1/B* term, and as B*->0, the eigenvalue will tend to –Infinity. At the equilibrium B*=K, the stability condition is

$\lambda_2+1 = 1 - (r + \alpha/K)$ (see Figure 112).

Example 3

In many ways, it makes more sense to formulate the DR in a multiplicative rather than additive way, i.e.

$C_{t+1} = C_t(1+ \alpha\Delta B/B)$

$\Delta C = \alpha.C.\ \Delta B/B = \alpha\ .C.(rB(1-B/K)-C)$ (14)

This formulation is intuitively appealing, because it implies that 'current catch' (C_t) is increased or reduced by x%, depending on whether the relative change in biomass is an increase or decrease, and depending on the value of α. If α=1, the catch is increased or reduced by the same % as the relative change in biomass.

For this example, the matrix of partial derivatives at the equilibrium (i.e. C*=rB*(1-B*/K)) is again a multiple of the first, so that one eigenvalue will always be 0. The other eigenvalue is given by:

$$\lambda_2 = r(1-2B^*/K) - \alpha r(1-B^*/K) = r(1-\alpha) + (r/K)(\alpha - 2)B^* \qquad (15)$$

Some interesting features are clear from equation (15). The eigenvalue is a linear function of B* with intercept r(1- α). There are two special cases: α =1 and α=2. If α =2 then λ_2 = r(1-2)=-r, irrespective of B*. This means that if 0<r<2, the system will be neutrally stable with the approach to the new equilibrium exponential if r<1, and oscillatory if 1<r<2.

If α =1 then λ_2 = - r/K.B*, so that for stability, we require B*/K < 2/r. For r<2, B*/K will always be <2/r since B*/K <1 in realistic space. Only when r>2, does unstable parameter space appear.

Results of local stability for examples 1,2 and 3

Figure 112 illustrates the stability criterion for examples 1 and 2 under low, moderate and high rate of increase (r) and for several different values of α, which can be thought of as the 'strength of response', since a small value of α implies a weak response to a given change in biomass, whereas a high value of α implies a strong response to a given change in biomass. Note though that the interpretation and hence absolute value of α is NOT the same in the different examples and can therefore not be compared directly (i.e. α=1 in example 1 is different from α=1 in example 2 or 3). The main point of the figures is to show how the stability condition changes as the rate of increase and strength of response increases, and at different levels of equilibrium biomass (B*).

The top row of panels in Figure 112 illustrates the stability condition for Example 1. The intercept of each line is given by (r- α) and the value at B*=K is (-r- α). Recall that stable equilibria are those where the stability criterion μ lies between -1 and 1 (the horizontal lines in Figure 112). Although not plotted here, it is important to recall that the other eigenvalue is always 0, so that one stability criterion is always met exactly. Note how the lines get steeper as r increases (the slope is -2r/K), and how there is therefore more chance of instable solutions at low biomass (particularly for low values of α) and at high biomass (particularly for high values of α).

The second row of panels in Figure 112 illustrates the stability condition for Example 2. The non-linearity is clear, as is the instability at low biomass because of the term -α/B* tending to – Infinity as B* tends to zero. There is again a higher chance of instability as r increases. For a given rate of increase, there is generally more chance of instability at low biomass than at high biomass in this example.

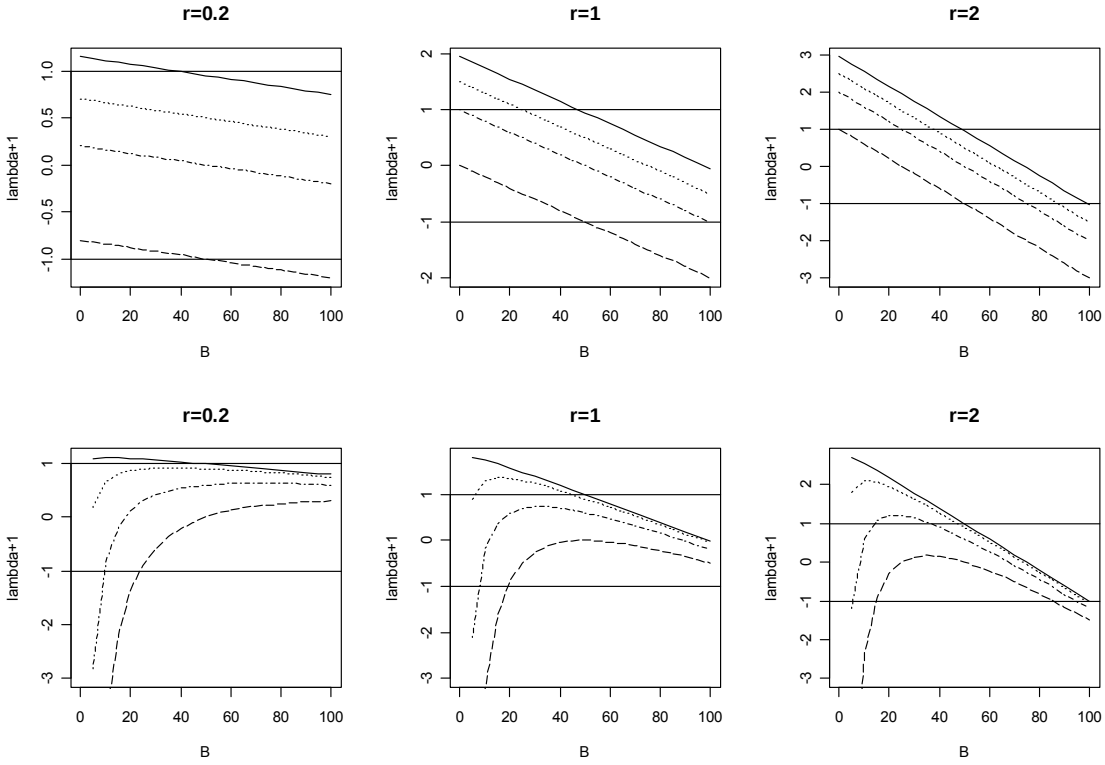


Figure 112. Stability criterion, μ = λ+1 for: top row=example 1 with lines (upper to lower) coinciding with α= (0.05,0.5, 1, 2) respectively and for 3 values or rate of increase, r; bottom row is the same but for example 2, and α= (0.5,5,20,50). In all cases K=100. Horizontal lines indicate the stability bounds at -1 and +1.

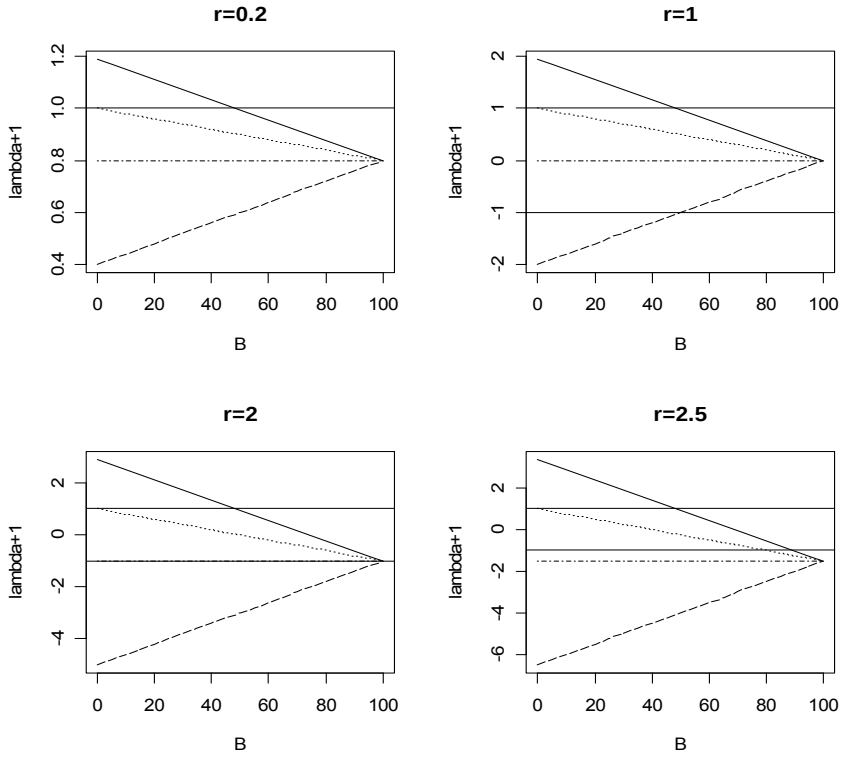


Figure 113. Stability criterion, μ = λ+1 for example 3 with lines (upper to lower) coinciding with α= (0.05, 1, 2,4) respectively and for 4 values or rate of increase, r. In all cases K=100. Horizontal lines indicate the stability bounds at -1 and +1.

Figure 113 (above) illustrates the stability condition for example 3, under different values of rate of increase, r, and for several values of α. The points made above are clear: when α =2, the stability condition is independent of where the equilibrium lies (horizontal line). When α =1, low biomass values will be neutrally stable, and higher levels will only be unstable if r>2.

Dynamics of neutrally stable, or unstable, equilibrium points

The time-series details of population biomass and catch will obviously be somewhat different under the three examples, but the type of behaviour following a small perturbation at an equilibrium point will be identical depending on the two eigenvalues of the system. Here only example 1 has been used to illustrate the types of behaviour. Similar 'portraits' can be obtained for the other two examples. Figure 114 shows examples of the biomass time-series following a small perturbation at some equilibrium biomass and catch (B* and C* = rB*(1-B*/K), are chosen to illustrate a particular behaviour). The perturbation was +0.01 in some cases and -0.01 in others, to show that the sign of the perturbation can sometimes matter. It is useful to refer back to panels in the top row of Figure 112 when looking at the panels in Figure 114.

The first 2 panels in the top row both started at a biomass level of B*=20, which according to Figure 112 is unstable with a positive value for λ+1. The first panel had a negative perturbation applied in the 2nd time-step and the population subsequently crashes to zero after just over 70 time-steps. The next panel had a positive perturbation applied and the system increased exponentially. The third panel in the top row was started at a (neutrally) stable equilibrium (B*=80), and the system goes to a new equilibrium – in this case very close to the original equilibrium. The second row shows oscillatory behaviour for: a neutrally stable equilibrium (oscillations because λ+1 is negative); an unstable equilibrium with oscillations of increasing magnitude; and a neutrally stable equilibrium but with one stability condition =1, the other =-1. In this last case, the perturbation simply persists as an oscillation around the original B*=50 level.

The last row of panels illustrate some examples for a high rate of increase (r=2.0). Note again how the sign of the perturbation (+0.01 or -0.01) leads either to a new equilibrium at high biomass (new B*=80) or a population crash. The final panel also shows a population crash, but following from oscillations with exponentially increasing amplitude.

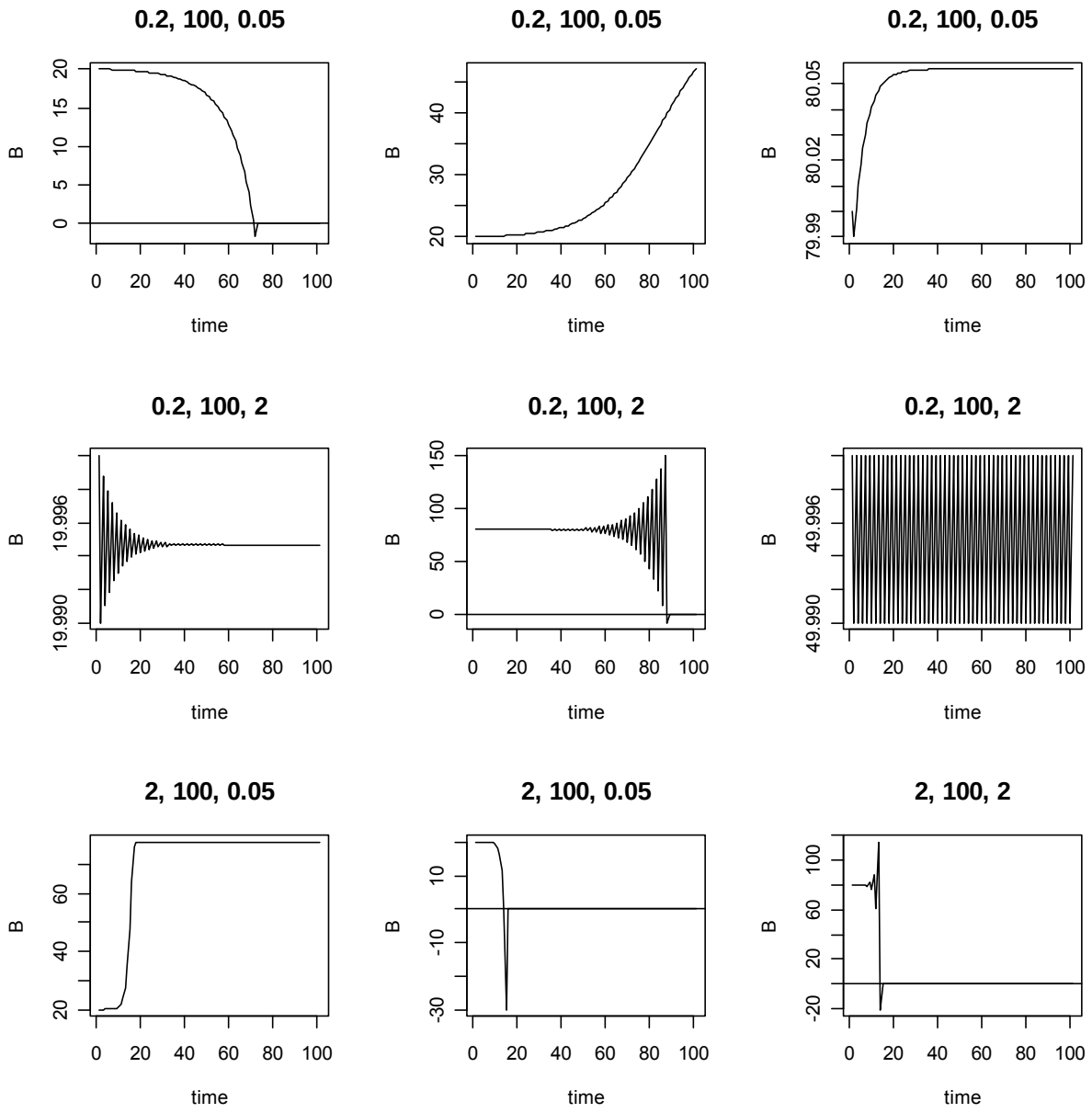


Figure 114. Time series of biomass following a small perturbation at an equilibrium value, for example 1 decision rule. The title line shows the parameter values for r, K, and α. Initial biomass (before the perturbation) and the sign of the perturbation is: top row, left to right, 20+, 20-, 80- ; middle row left to right, 20-, 80-, 50-; bottom row left to right 20+, 20-, 80-. The magnitude of the perturbation was 0.01 in all cases.

11.4. Harvest rate Decision rules based on change in abundance

Example 4

The only difference between example 1 and this example is that the decision rule is specified in terms of fishing mortality (F), or harvest rate, which can be thought of as the level of effort. In practice, management could, of course, be in terms of catch if an estimate of abundance is available to translate from F to C, though this is not relevant here. The decision rule is now:

$$\Delta F_{t+1} = F_{t+1} - F_t = \alpha(rB_t(1-B_t/K) - F_tB_t) \qquad (16)$$

Equation (16) again does not guarantee only positive harvest rates and this obvious weakness would need to be removed if such a DR were to be used in reality.

Analogous to the catch-based system in example 1, an equilibrium where $(B_{t+1} - B_t) = 0$, implies a constant harvest rate from one year to the next. The set of equations (1b and 16) also does not have a unique equilibrium solution. For any level of biomass, B*, there is a constant harvest rate, F*, which would keep the system at the equilibrium (B*,F*) where

$$F^* = r(1-B^*/K) \qquad (17)$$

There is again no implicit or explicit target biomass in the system, and the decision rule will not rebuild a heavily depleted stock or allow for increases in catches of a lightly harvested stock.

Analogous to the catch-based examples, the matrix of first derivatives with respect to (B,F) and evaluated at the equilibrium (B*,F*) is given by:

$$A = \begin{bmatrix} -rB^*/K & -B^* \\ -\alpha rB^*/K & -\alpha B^* \end{bmatrix} \qquad (18)$$

and the eigenvalues are:

$$\lambda_1 = 0$$
$$\lambda_2 = -rB^*/K - \alpha B^* = -(r/K + \alpha)B^* \qquad (19)$$

As before, the fact that the 2nd row of A is a multiple of the first row, indicates that one eigenvalue will be zero, and also that the system can be rewritten in terms of B alone with an additional time-lag (B_{t+1} depending on B_t and B_{t-1}). Since r and α are positive, for any feasible (i.e. positive) B*, $\lambda_2<0$ so that $\lambda_2+1<1$. This means that the response of this system to perturbations is generally oscillatory.

Example 5

If the relative change in biomass is used to update the harvest rate, i.e.

$$\Delta F_{t+1} = \alpha(r(1-B_t/K) - F_t) \qquad (20)$$

then the eigenvalues are:

$$\lambda_1 = 0$$
$$\lambda_2 = -rB^*/K - \alpha = -(r/K)B^* - \alpha \qquad (21)$$

Example 6

This example is similar to the third Catch example (example 3). The decision rule is specified in a multiplicative way (still in terms of fishing mortality, F). The decision rule is now based on:

$$F_{t+1} = F_t (1 + \alpha\Delta B_{t+1}/B_t)$$

so that

$$\Delta F_{t+1} = \alpha F_t(r(1-B_t/K) - F_t) \qquad (22)$$

There are still multiple equilibria where

$$F^* = r(1-B^*/K)$$

and the eigenvalues are:

$$\lambda_1 = 0$$
$$\lambda_2 = -rB^*/K - \alpha r(1-B^*/K) = -(r/K)(1-\alpha)B^* - \alpha r \qquad (23)$$

Note that since r and α are positive, for any feasible (i.e. positive) B*, $\lambda_2<0$ so that $\lambda_2+1<1$. This means that the response of this system to perturbations is also generally oscillatory. For B*=0 the non-zero eigenvalue is $\lambda_2=-\alpha r$, and for B*=K the non-zero eigenvalue is $\lambda_2=-r$.

Table 1 summarises the first 6 examples of DRs used with the Schaefer stock-production model. The non-zero eigenvalue is given in terms of the input parameters. The last two columns give the stability condition, $\mu = \lambda +1$ at the two extremes, B*=0 and B*=K. Recall that $|\mu|<1$ for an equilibrium to be locally stable. It is more convenient to consider the value of the stability condition directly, because the sign of μ indicates whether behaviour of the system following a perturbation would be exponential (+) or oscillatory (-).

Bear in mind that the interpretation, and hence scaling, of α is different in each example. For the harvest rate example 4 in particular, where F is adjusted by some fraction of the change in biomass, it is crucial that α is small enough given that sensible values of F are likely to be less than 1 and changes in biomass (in these examples) are of the order of 10's or more (in reality, likely to be even more). This already suggests that this rule is a lot less preferable than one based on relative (or % change) in biomass.

Figure 115 illustrates stability conditions for examples 4 (top row) ,5 (middle row) ,6 (bottom row) under low, moderate and high r and for a range of α values. Example 4 shows that equilibria at low biomass are generally neutrally stable. It is for high α and high r that instability appears at high biomass levels. Example 5 has similar stability patterns to those of example 1, but for low values of α, there is generally stability at low biomass. Instability tends to appear under high biomass (this may be because there is no limit on the harvest rate itself). Example 6 shows some similarity to example 3. Now instability tends to occur at low biomass levels (particularly for high r and high α).

The dynamics of equilibrium points under different stability conditions will again show the characteristic types of 'portraits' illustrated in Figure 114 in spite of the fact that the decision rules are here defined in terms of harvest rate rather than catch.

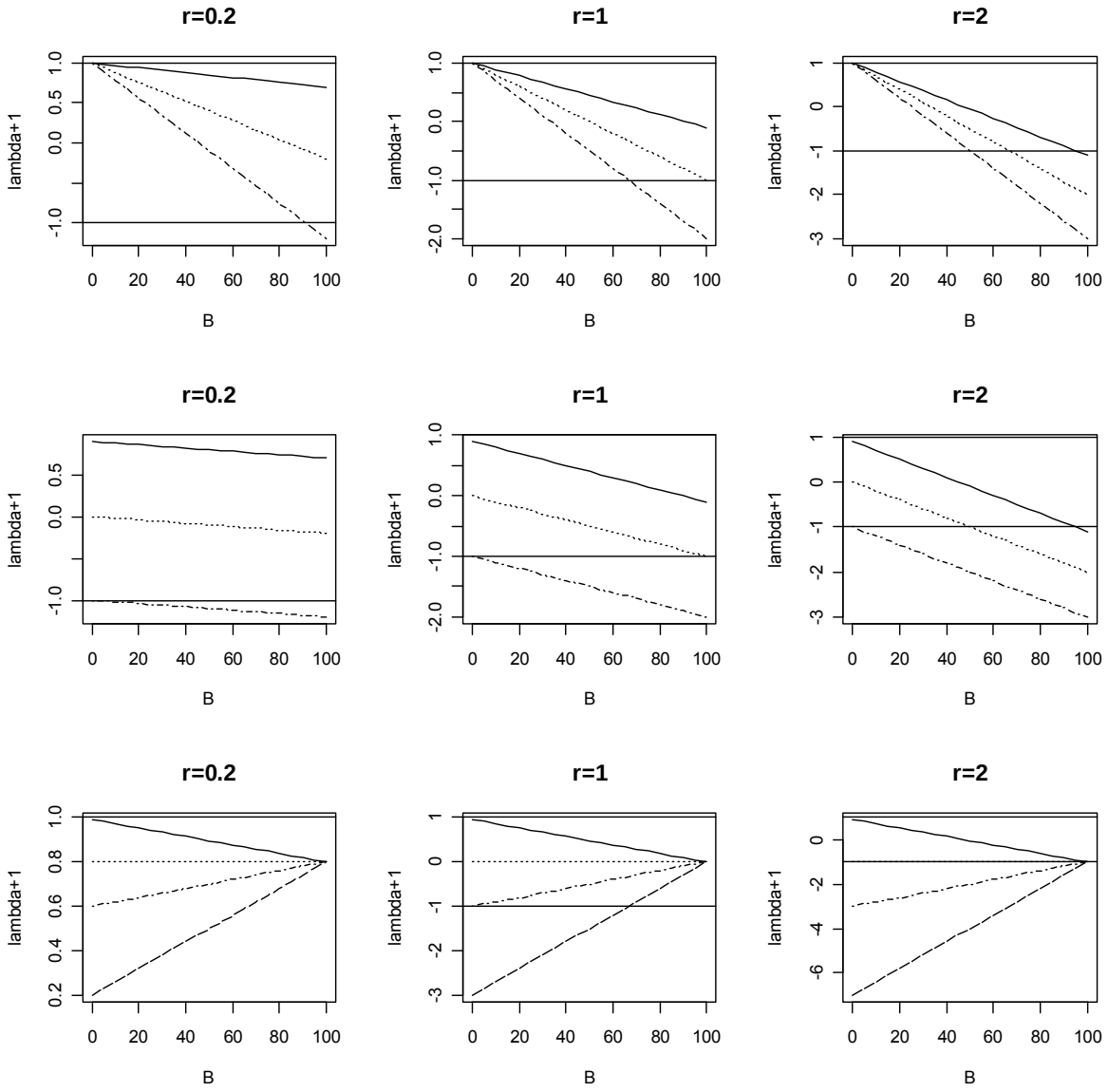


Figure 115. Stability criterion, μ = λ+1 for: top row=example 4 with lines (upper to lower) coinciding with α= (0.001,0.01, 0.02) respectively and for 3 values or rate of increase, r; middle row is the same but for example 5, and α= (0.1,1,2) and the bottom row is for example 6 and α= (0.05,1,2,4). In all cases K=100. Horizontal lines indicate the stability bounds at -1 and +1; (in the bottom right panel the line for α= 1 coincides with the horizontal line at -1.

Fox stock-production model

The shape of a stock-production model and particularly the ratio of $B_{msy}/B_0 = B_{msy}/K$ can be related to a steepness[28] level in the stock-recruit curve of a model based on numbers-at-age. It is easy to show, through a very simple age-structured population model, that a $B_{msy}/K=0.5$ is similar to a stock with very low steepness of around 0.3. At higher steepness (say 0.6-0.8) the ratio B_{msy}/B_0 is more likely to be around 0.3 to 0.4. This may suggest that the Fox stock-production model, with $B_{msy}/K=\exp(-1)\simeq 0.37$, would be generally more relevant or realistic than

[28] Steepness (h) of the Beverton-Holt stock recruit curve is defined as the ratio of recruitment (R) to unexploited recruitment (R0) when spawning biomass (SSB) is at 20% of unexploited spawning biomass (SSB0), i.e. h = R(SSB=0.2SSB0)/R0.

the Schaefer model. It is then worth asking what the dynamics of such a system would be with a decision rule specified in terms of catch or harvest rate.

I only consider the equivalent of examples 3 and 6 with the Fox model. These decision rules adjust catch or harvest rate multiplicatively by some percentage of the relative change in biomass. Table 1 shows results for these two cases:

Example 7:

$$\Delta B_{t+1} = B_{t+1} - B_t = rB_t(1 - \ln(B_t)/\ln(K)) - C_t \quad (24)$$
$$\Delta C_{t+1} = \alpha C_t \, \Delta B_{t+1}/B_t$$

Example 8:

$$\Delta B_{t+1} = B_{t+1} - B_t = rB_t(1 - \ln(B_t)/\ln(K)) - F_t B_t \quad (25)$$
$$\Delta F_{t+1} = \alpha F_t \, \Delta B_{t+1}/B_t$$

Although the stability condition is the same at B*=K for the Fox and Schaefer models, it is clear that the ln(B*) term introduces a nonlinearity in μ as a function of B*, and tends to – or + infinity as B* tends to zero.

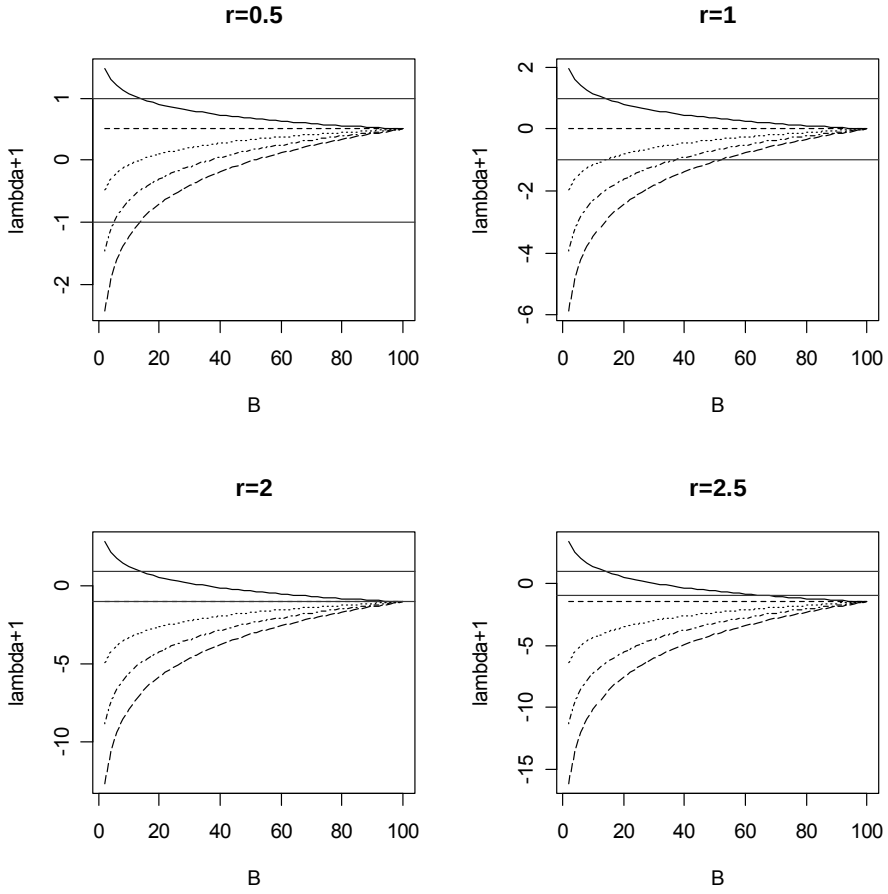


Figure 116. Stability criterion, μ = λ+1 for the Fox model with catch-based DR at different values of r and with lines (upper to lower) coinciding with α= (0.5, 1,1.5, 2, 2.5) respectively. In all cases K=100. Horizontal lines indicate the stability bounds at -1 and +1.

Figure 116 shows that with the Fox model the strength of response parameter, α probably needs to be chosen with more care. Note however, that because of the difference in parameterisation of the two models, the r-parameter does not have exactly the same interpretation in the two models.

The coinciding graph for the Fox model with F-based decision rule is very similar to Figure 116, but none of the curves increase to +Infinity (see Table 1, below). This implies that the dynamics outside the stable region are most likely to be oscillatory.

Table 22. Summary of equations for Decision rule examples used with the Schaefer and Fox stock production models, together with equations for the non-zero eigenvalue and the stability conditions at the extremes of possible biomass values, B*=0 and B*=K. For the Fox model the shorthand notation lnB*=ln(B*) and lnK=ln(K) are used.

Schaefer Stock-Production model				
Examples 1,2,3: multiple equilibria where C*=rB*(1-B*/K)				
Example	*Decision rule*	*non-zero eigenvalue*	*μ(B*=0)*	*μ(B*=K)*
1	$\Delta C_{t+1} = \alpha\, \Delta B_{t+1}$	$\lambda = r(1-2B^*/K) - \alpha$	$1 + r - \alpha$	$1 - r - \alpha$
2	$\Delta C_{t+1} = \alpha\, \Delta B_{t+1}/B_t$	$\lambda = r(1-2B^*/K) - \alpha/B^*$	$-\infty$	$1 - r - \alpha/K$
3	$\Delta C_{t+1} = \alpha C_t$ $\Delta B_{t+1}/B_t$	$\lambda = r(1-2B^*/K) - \alpha r(1-B^*/K)$	$1 + r - \alpha r$	$1 - r$
Examples 4,5,6: multiple equilibria where F*=r(1-B*/K)				
4	$\Delta F_{t+1} = \alpha\, \Delta B_{t+1}$	$\lambda = -rB^*/K - \alpha B^*$	1	$1 - r - \alpha K$
5	$\Delta F_{t+1} = \alpha\, \Delta B_{t+1}/B_t$	$\lambda = -rB^*/K - \alpha$	$1 - \alpha$	$1 - r - \alpha$
6	$\Delta F_{t+1} = \alpha F_t$ $\Delta B_{t+1}/B_t$	$\lambda = -rB^*/K - \alpha r(1-B^*/K)$	$1 - \alpha r$	$1 - r$
Fox Stock-Production model				
Example 7: multiple equilibria where C*=rB*(1 – lnB*/lnK)				
7	$\Delta C_{t+1} = \alpha C_t$ $\Delta B_{t+1}/B_t$	$\lambda = r(1-\alpha) - r/\ln K$ $- r(1-\alpha)\ln B^*/\ln K$	+ or - ∞	$1 - r/\ln K$
Example 8: multiple equilibria where F*=r(1 – lnB*/lnK)				
8	$\Delta F_{t+1} = \alpha F_t$ $\Delta B_{t+1}/B_t$	$\lambda = -r(1+\alpha) + (\alpha r)\ln B^*/\ln K$	$-\infty$	$1 - r$

11.5. Sigmoid Catch and Harvest rate rules

Example 9

This example is based on the concept of so-called 20-40 (or 10-40) decision rules which are generally formulated as piecewise linear functions (see Section 10). For example, the catch would be at some maximum value (e.g. MSY, or some fraction of MSY) when spawning biomass (SSB) is above, say, 40% of unexploited spawning biomass, SSB0; it would linearly decrease for SSB between 40% and 20% of SSB0; it would be zero for SSB below 20% of SSB0. One interpretation of (and basis for) the parameters in such a simple rule is that SSB40% acts as a target reference point and SSB20% as a limit reference point. The maximum catch, C_{max}, could be thought of as a limit catch reference point.

In the production model framework used here, I replace SSB by the total biomass, B. Instead of using the linear version of this rule, I have replaced it by a sigmoid function, with a parameterisation that relates well to those in the original rule. The catch, C_t in Eqn 1a is now specified as:

$$C_t = \frac{C_{\max}}{1 + \exp(\frac{\alpha}{K}(\tilde{B} - B_t))} \qquad (26)$$

The interpretation of the three parameters, C_{max}, α and $\tilde{B}$ are quite intuitive. C_{max} is the maximum catch; α governs the steepness of the sigmoid function (low α means a shallow curve, high α means a steep curve) and $\tilde{B}$ is the point where $C=C_{max}/2$. Figure 117 below illustrates this function for three values of α. The lowest value (α =1) does not in fact generate a sigmoid curve over the range of biomass values from 0 to K, but a linear curve. In later analyses, only values which generate an obviously sigmoid curve are considered. This decision rule's weakness is the fact that C_t is not exactly zero when B_t is zero, but tends to zero in the limit. In practice, this can be fixed, and it does not detract from illustrating the main dynamics of the system.

Although it may seem impractical to formulate catch at time-step t in terms of biomass (or an indicator) at time step t, in the context of the discrete-time model, B_t can be thought of as the biomass at the start of time step t. Similarly an indicator can be assumed to reflect the biomass at the start of the time period, so that it would be available to determine the catch (e.g. as a quota) before the catch is taken. This formulation is therefore not entirely impractical, though a formulation of C_t in terms of B_{t-1} may in some cases make more sense, depending on when the relevant data on the indicator becomes available. The additional time-lag does not affect the number or level of equilibria, but it will affect the dynamics as shown below.

The difference in biomass from one time-step to the next is now given by:

$$B_{t+1} - B_t = rB_t(1 - B_t / K) - \frac{C_{\max}}{1 + \exp(\frac{\alpha}{K}(\tilde{B} - B_t))} \qquad (27)$$

The maximum sustainable catch for the Schaefer model is MSY=rK/4 which occurs at $B_{msy}=K/2$. It is therefore useful to consider cases where, for example, $C_{max}=rK/4$ and $\tilde{B}=K/2$.

Ignoring the trivial (and undesirable) case where B=0, or approximately zero, there is now at least equilibrium (but possibly three or more) which needs to be found numerically. It is informative to plot ΔB_{t+1} against B_t to see where the equilibrium points are. Figure 117 also shows that as α increases, the curve for ΔB gets a 'kink'. This together with changes in $\tilde{B}$ can generate multiple equilibria.

How a combination of changes to α , $\tilde{B}$ and C_{max} affects the equilibria is less intuitive and best illustrated through graphical examples. Figure 118 below again shows how higher values of α generate more of a 'kink' in delB. When combined with a low value of $\tilde{B}$, it can lead to 3 or 4 equilibria. As $\tilde{B}$ is increased, the upper equilibrium shifts further to the right (i.e. at higher B). As C_{max} is increased, the equilibrium shifts to a lower value of B. The example shown here has r=0.2 (and K=100), with C_{max} set to MSY=5 and $\tilde{B}$ set to three different values: below B_{msy}, at $B_{msy}=50$, and above B_{msy}

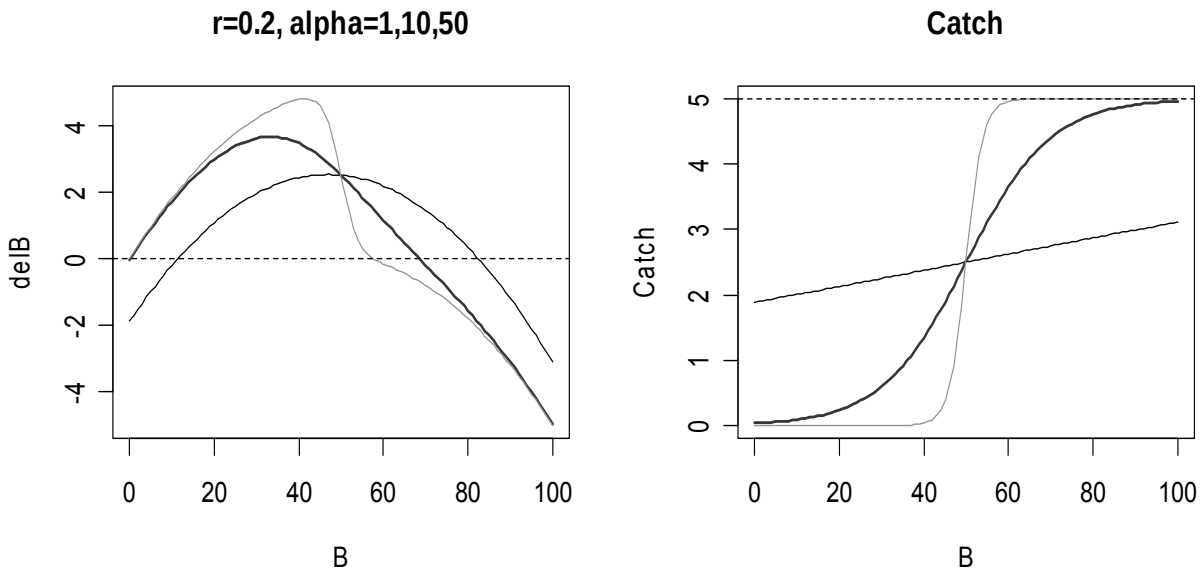


Figure 117. Change in Biomass as a function of the level of biomass and Catch as a function of biomass for a r=0.2, K=100, $\widetilde{B}$ =50 and three values of α: 1=thin black line; 10=bold black line; 50=grey line. C_{max}=5 in all cases (i.e. MSY=rK/4). Equilibrium points lie where ΔB=delB=0 on the left-hand panel.

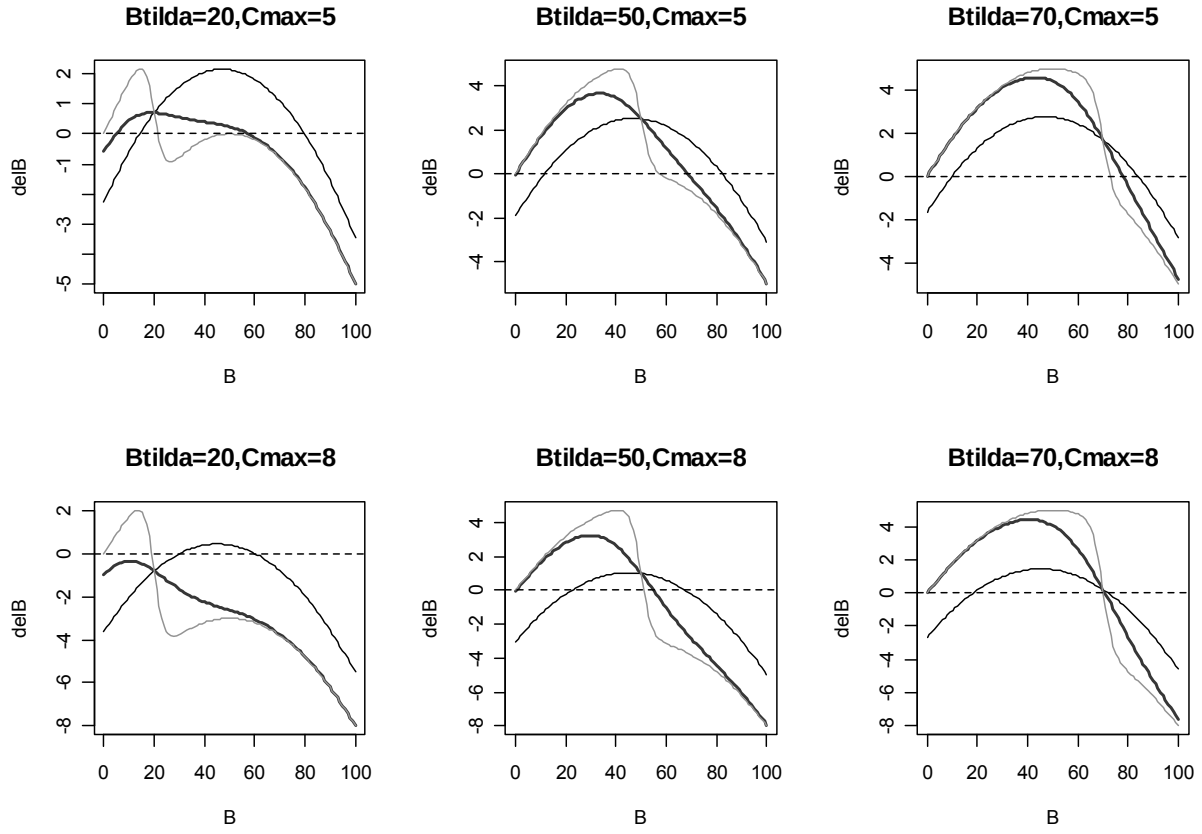


Figure 118. Change in biomass as a function of biomass level for r=0.2, K=100 and a range of values of Cmax and $\widetilde{B}$ as shown in the heading. In each panel, the lines are for α 1,10,50 (thin black, bold black, grey). Recall from Figure 6 that α=1 is not really sigmoid over the range of B-values.

In contrast to the other examples, delB is NOT exactly equal to 0 at B*=0, though the lowest equilibrium biomass is usually quite close to zero[29]. It is interesting to note that when $\widetilde{B}$ is set at or above B_{msy}, and C_{max} below or at MSY, the equilibrium Biomass is in fact ABOVE B_{msy}. This is explored and discussed further below.

The local stability characteristics for this decision rule at an equilibrium, B*, is based on the first derivative of Eq.27, evaluated at B*:

$$\lambda = r - 2rB^*/K - .\frac{(\alpha/K).C_{\max}\exp[(\alpha/K)(\widetilde{B}-B^*)]}{(1+\exp[(\alpha/K)(\widetilde{B}-B^*)])^2} \qquad (28)$$

As before, the condition $|\lambda + 1| < 1$ indicates a locally stable equilibrium.

In this case there are a countable number of equilibria, in contrast to examples 1-8 which could have an equilibrium point at any level of biomass. The equilibria associated with the top row of panels in Figure 118 above are identified by the symbols in Figure 119 (below), which also shows whether that equilibrium is stable or not. Note that the first panel in fact has four equilibrium points when α =50 (triangles), but there are two very close together either side of B=50 (B*=49.97 (unstable) and B*=50.03 (stable)).

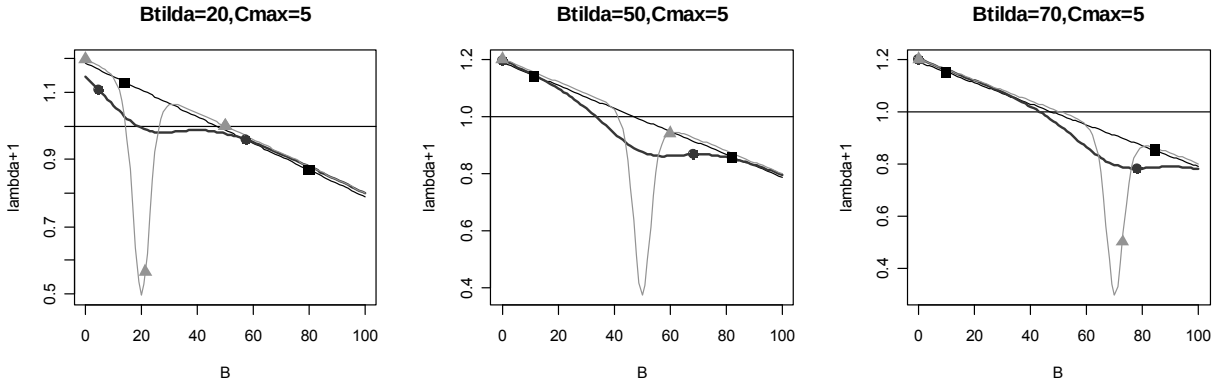


Figure 119. Stability criterion as a function of any biomass level with symbols indicating where the equilibria are (i.e. where delB=0). The horizontal line as at the stability boundary 1; equilibrium points above that are unstable. In each panel, the lines/symbols are for α 1,10,50 (thin black/squares, bold black/dots, grey/triangles).

It is again relevant to note that although each term B_t is written as a function of B_{t-1}, it is ultimately a function of the whole series of previous biomasses, including the starting value, B_0. It is easy to show by simulation that the starting condition (i.e. the value of B_0) matters for the ultimate outcome and at which stable equilibrium the system will end up.

Bifurcations, limit cycles and complex dynamics

May and Oster (1976) pointed out the 'disconcerting property' of some simple difference equation models, namely to go from a stable point through a sequence of bifurcations into stable cycles of increasing period. Table 1 in May and Oster (1976) includes the logistic model (without

[29] At B=0, delB = -Cmax/(1+exp(α* $\widetilde{B}$ /K))

harvest term) which displays these complex dynamics only at a very high rate of increase (r>2.69) Here the addition of a harvest term, and particularly of a sigmoid catch or harvest rate function, means that the complex dynamics can be observed even at a low rate of increase, depending on the choice of parameters in the decision rule. Some of these complicated dynamics are illustrated here.

There are essentially four parameters that affect the dynamics: r, α, C_{max} and $\widetilde{B}$. The illustrations below take a single value of r and explore the dynamics, particularly with respect to C_{max} and $\widetilde{B}$. A simulation model is used to take a set of input parameters and run the population (and catch) dynamics for, say 300, time steps. The population levels over the last 200 time-steps are then plotted – in this case, against the value of $\widetilde{B}$ on the x-axis. Where there is a single value, there is no bifurcation but a single equilibrium point. Where there are two values, there is a two-point limit cycle, 4 points imply a four-point limit cycle, etc. Where there is a broad band of points at a given $\widetilde{B}$ value it suggests a limit cycle of very high order. I should emphasise that the figures below are based on starting the population at its unexploited level, B_0=K. Some of the figures could look different if B_0 was set at a very different value[30].

Recall that this system is based on the decision rule with no time lag in the biomass (i.e. eqn 26, using B_t to determine C_t). The effect of a low versus a high value of α has already been illustrated in terms of a larger number of equilibrium points and the large change in the derivative in the vicinity of $\widetilde{B}$. It is unsurprising that for low r and low α, bifurcations do not really occur. Even for relatively high α and high maximum catch (e.g. C_{max}=2.C_{msy} or 4.C_{msy}), bifurcations do not occur (not illustrated).

As implied by the work of May and others on the logistic model without harvest, higher values of r is more likely to lead to instability or complex behaviour. For example, when r=1.0, then for the same value of α (=50), the dynamics are more complex (Figure 120 below) than for r=0.2. For r=1.0, C_{msy}=25, and even for C_{max}=26 (i.e. only 4% above C_{msy}), the complexity in dynamics at low $\widetilde{B}$ is striking.

Although one might ask why even consider such high levels of C_{max}, this is very relevant because in reality there may be limited knowledge and high uncertainty about where the true C_{max} (C_{msy}) really lies for a particular fish stock.

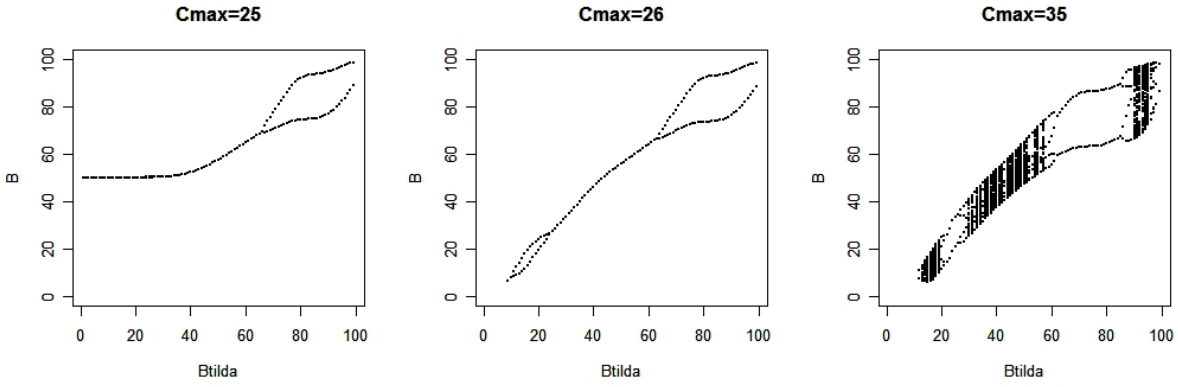


Figure 120. Bifurcation plots of biomass (B) as a function of $\widetilde{B}$ in the sigmoid decision rule based on B_t for different values of the maximum catch, Cmax. Other parameters are: r=1.0, K=100, α =50.

[30] This is particularly likely when there is a lower equilibrium point as in the case where r=0.2, K=100, $\widetilde{B}$ =20, Cmax=5.

The next set of figures take a particular combination of the four parameters (r, K, α, $\widetilde{B}$, C_{max}) and shows catch (C_t) versus biomass (B_t) for the last 200 time steps. These are not the standard phase-plots (of B_{t+1} versus B_t) that are often considered, but they have the advantage of showing the range (amplitude) of oscillations in both the biomass and catch. The examples below are taken from the plots above. Figure 121 (below) is for the low rate of increase (r=0.2) and Figure 122 is for the example where r=1.0. Note that in some cases the oscillations imply very large ranges of both biomass and catch (e.g. catches between 0 and 2.4 times C_{msy} and biomass between 15 and 30% of unexploited biomass, B_0; see middle panel of Figure 121). In other cases, the oscillations are converging and the ranges are very small (middle panel of Figure 122).

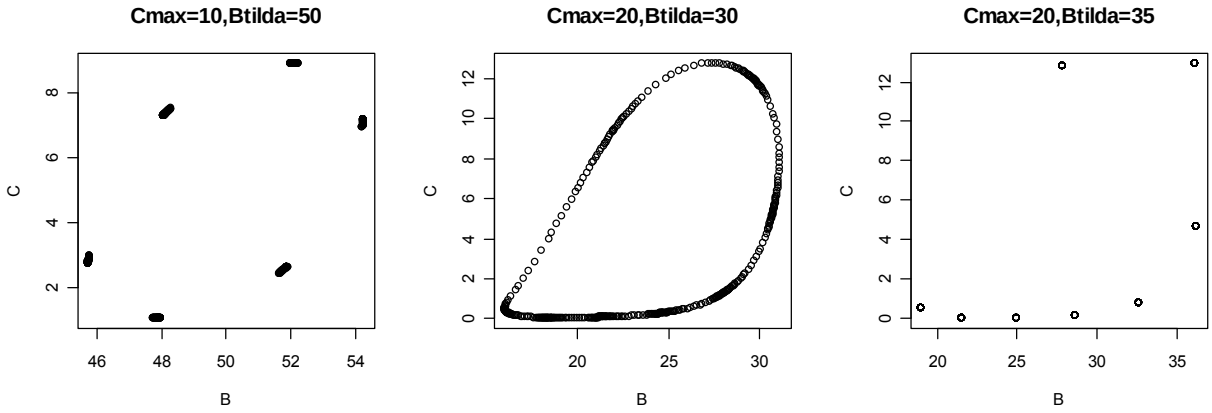


Figure 121. Catch versus biomass for the last 200 (out of 300) iterations. Values of Cmax and $\widetilde{B}$ are given above each panel. Other parameters are: r=0.2,K=100 and α =50 as in Figure 123.

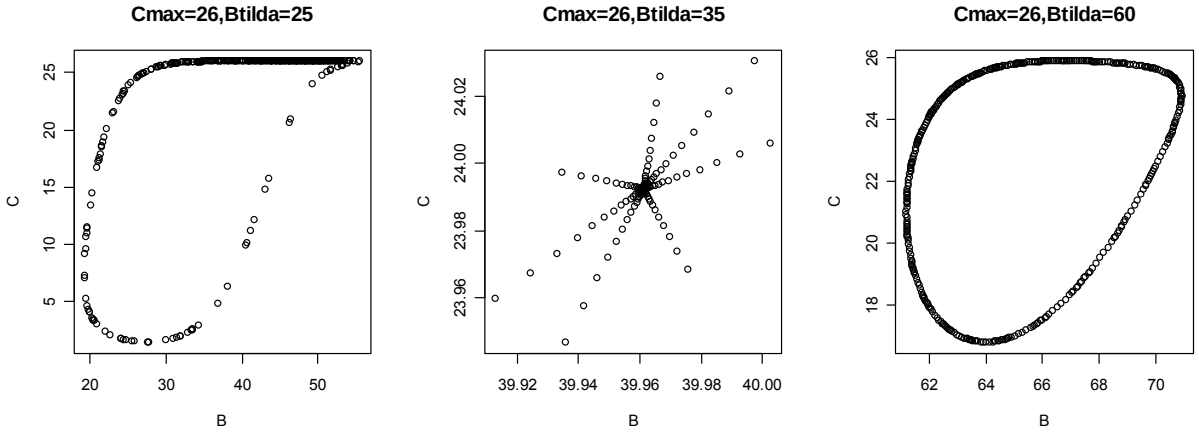


Figure 122. Catch versus biomass for the last 200 (out of 300) iterations. Values of Cmax and $\widetilde{B}$ are given above each panel. Other parameters are: r=1.0,K=100 and α =50 as in Figure 120.

The destabilising effect of additional time-lags found by May *et al.* (19..) is also evident here. If the term B_t in the decision rule given in equation 26 is replaced by B_{t-1}, then even at a low rate of increase, complex dynamics are observed when C_{max} is set well above C_{msy} (e.g. C_{max}=2.C_{msy} or 4.C_{msy}). Figure 123 shows the bifurcation plots for this example; the system with decision rule based on B_t does not show any of these complex behaviours at the same sets of parameter values.

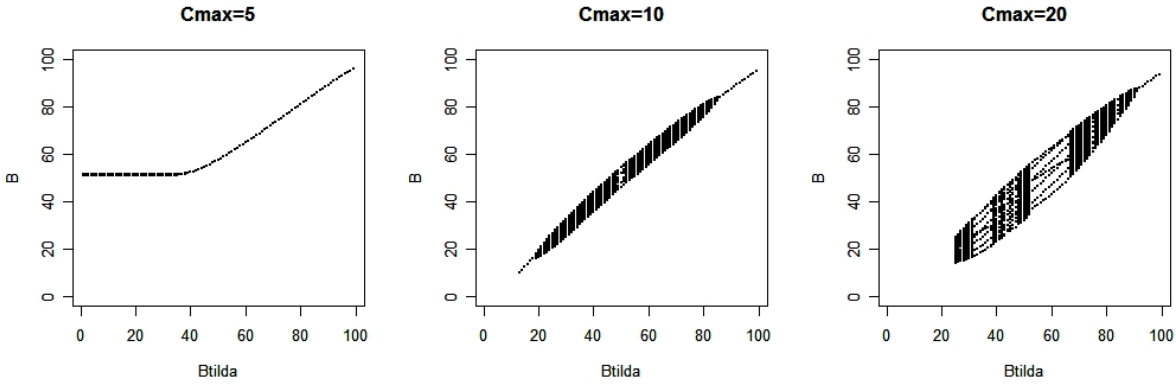


Figure 123. Bifurcation plots of biomass (B) as a function of $\tilde{B}$ in the sigmoid decision rule based on B_{t-1} instead of B_t for different values of the maximum catch, C_{max}. Other parameters are: r=0.2, K=100, α =50.

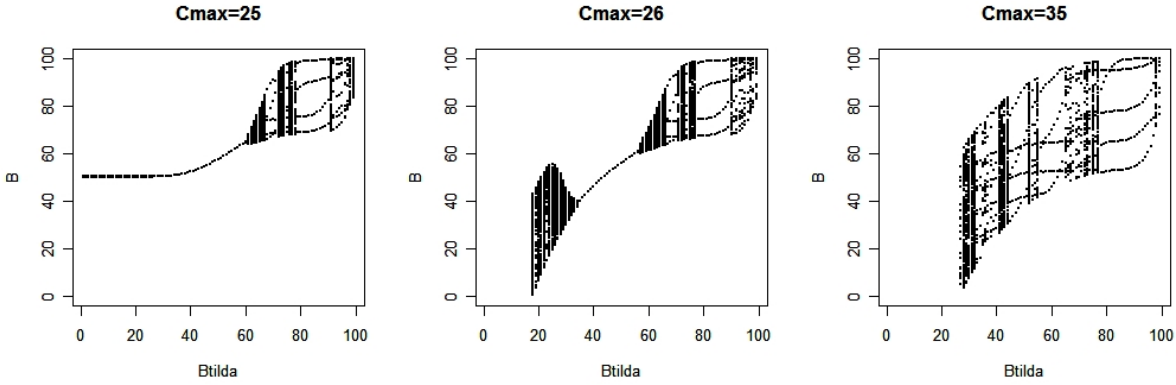


Figure 124. Bifurcation plots of biomass (B) as a function of $\tilde{B}$ in the sigmoid decision rule based on B_{t-1} instead of B_t for different values of the maximum catch, C_{max}. Other parameters are: r=1.0, K=100, α =50.

Example 10

The harvest rate (F) version of this model has the same population dynamics, but the decision rule is now specified in terms of the harvest rate rather than catch i.e.:

$$B_{t+1} - B_t = rB_t(1 - B_t / K) - \frac{F_{\max} B_t}{1 + \exp(\frac{\alpha}{K}(\tilde{B} - B_t))} \qquad (29)$$

The stability condition is analogous to that for the catch-based version, with

$$\lambda = r - 2rB^* / K - \frac{(\alpha / K).F_{\max} B^* \exp[(\alpha / K)(\tilde{B} - B^*)]}{(1 + \exp[(\alpha / K)(\tilde{B} - B^*)])^2} - \frac{F_{\max}}{(1 + \exp[(\alpha / K)(\tilde{B} - B^*)]} \qquad (30)$$

The sigmoid decision rule specified in terms of harvest rate, F, does not have the same tendency for multiple equilibria as the catch-based version. The lower equilibrium is now at B*=0. There is another important difference between the catch and harvest rate versions. The catch version has a maximum catch specified. The harvest rate version has a maximum harvest rate, F_{max} but not a maximum catch, so that even when F has reached the asymptote at F_{max}, the catch will increase if B_t increases further. This is likely to be one reason why in this example there is a tendency for complex dynamics over a wider range of parameter space for the harvest rate than the catch

version of the decision rule. This is illustrated in Figure 126 (below) which compares a bifurcation plot for the same values of r, K and α over a range of $\widetilde{B}$ values. The maximum catch, C_{max}, is set at C_{msy}=rK/4 for the catch-based DR; the maximum harvest rate, $F_{max} = F_{msy} = r/2$ for the harvest rate DR. In each set of comparisons, the harvest-rate version shows complex behaviour over a wider range of $\widetilde{B}$ values than the catch-based version.

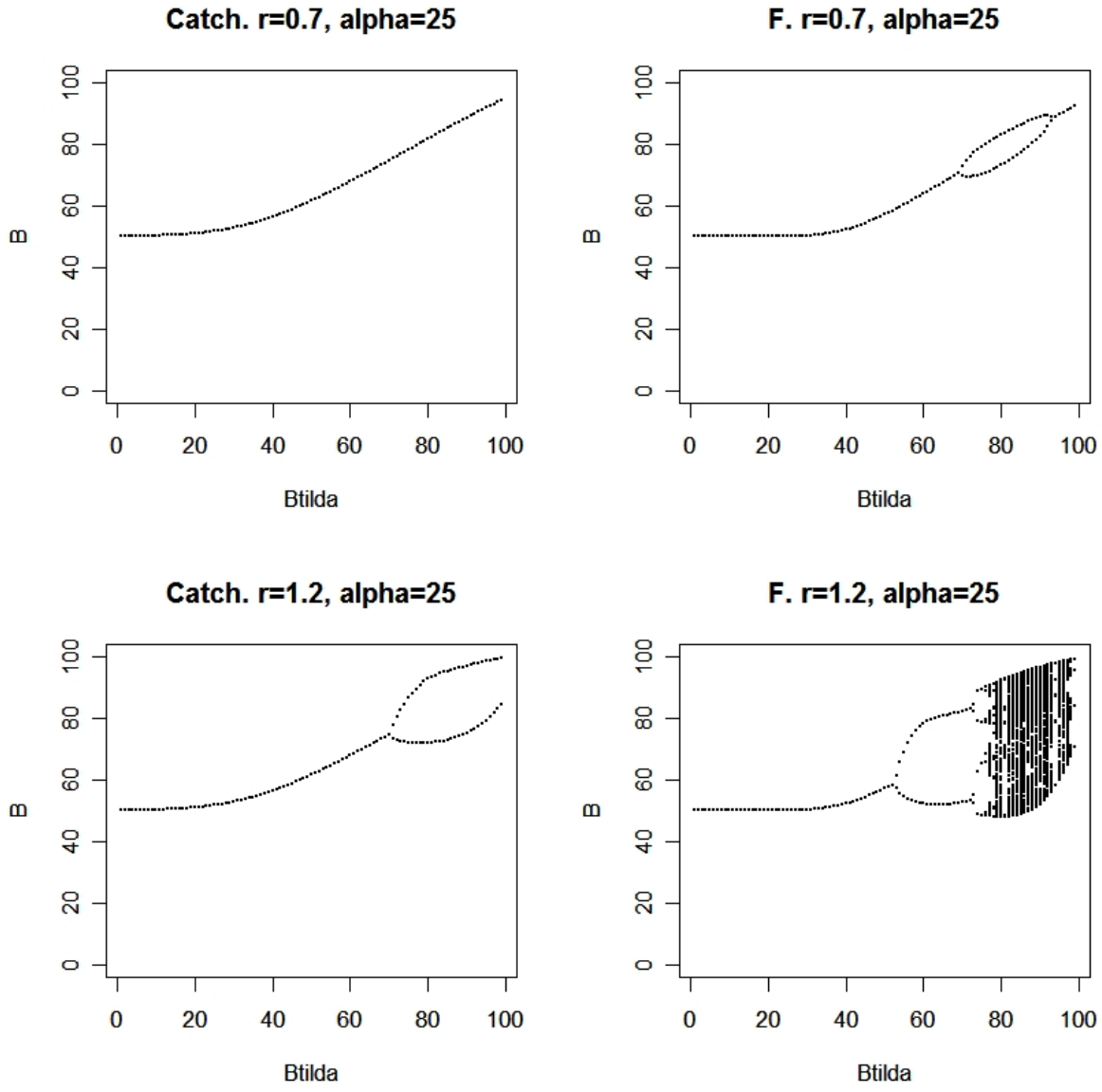


Figure 125. Bifurcation plots as a function of $\widetilde{B}$ with a sigmoidal catch-based decision rule (left-hand panels) or a harvest rate based decision rule (right-hand panels), in both cases based on B_t. The values of α and r are shown above each panel. Catch-based rules have C_{max} = rK/4; harvest rate-based rules have F_{max}=r/2.

When the harvest rate decision rule is based on B_{t-1} rather than B_t, there is again an increase in complex dynamics as was seen for the catch-based rules based on B_{t-1} rather than B_t. A comparison between the catch- and F-based rules with the time lag (i.e. based on B_{t-1}) again shows more complex dynamics (i.e. over a wider range of B values) for the harvest rate (F-based) version of the decision rule.

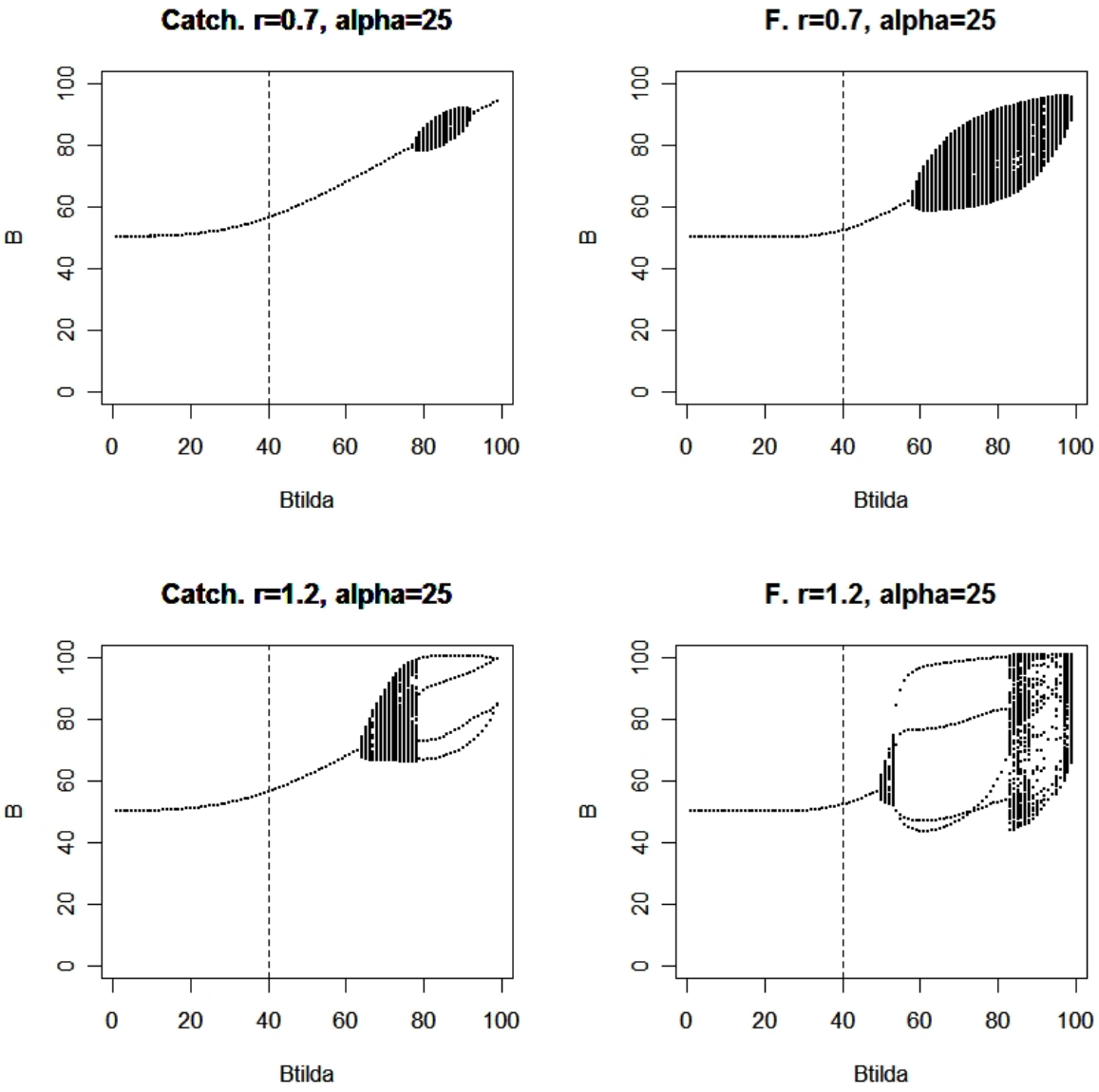


Figure 126. Bifurcation plots as a function of $\widetilde{B}$ with a sigmoidal catch-based decision rule (left-hand panels) or a harvest rate based decision rule (right-hand panels), in both cases based on B_{t-1} instead of B_t. The values of α and r are shown above each panel. Catch-based rules have C_{max} = rK/4; harvest rate-based rules have F_{max}=r/2.

Realism and stochasticity

As noted above, the systems considered here are all deterministic and there is little doubt that real fisheries systems are anything but deterministic. Looking at the cyclical behaviour illustrated above, two questions arise: First, is this likely to occur in realistic or "likely" parameter space or not? and second, does it matter if there is stochasticity in the system or would the oscillations be "swamped" by noise?

For the first question, regarding realism, it is informative to consider the proposed 10-40 or 20-40 decision rules mentioned above, and to find parameters for the sigmoid formulation that imply curves which map reasonably well onto the piecewise linear decision rules. Figure 127 below shows four curves based on two values of the α parameter (15,25) and two values of the

$\widetilde{B}$ parameter (30 and 40), where the catch or harvest rate is half that at the asymptote. Note that, since B_0=K=100 in these examples, $\widetilde{B}$ can be thought of as a percentage of unexploited biomass, B_0. We have already seen that the higher α, the more likely for bifurcations to occur, so only α =25 is further considered below to evaluate the likely occurrence of complex dynamics.

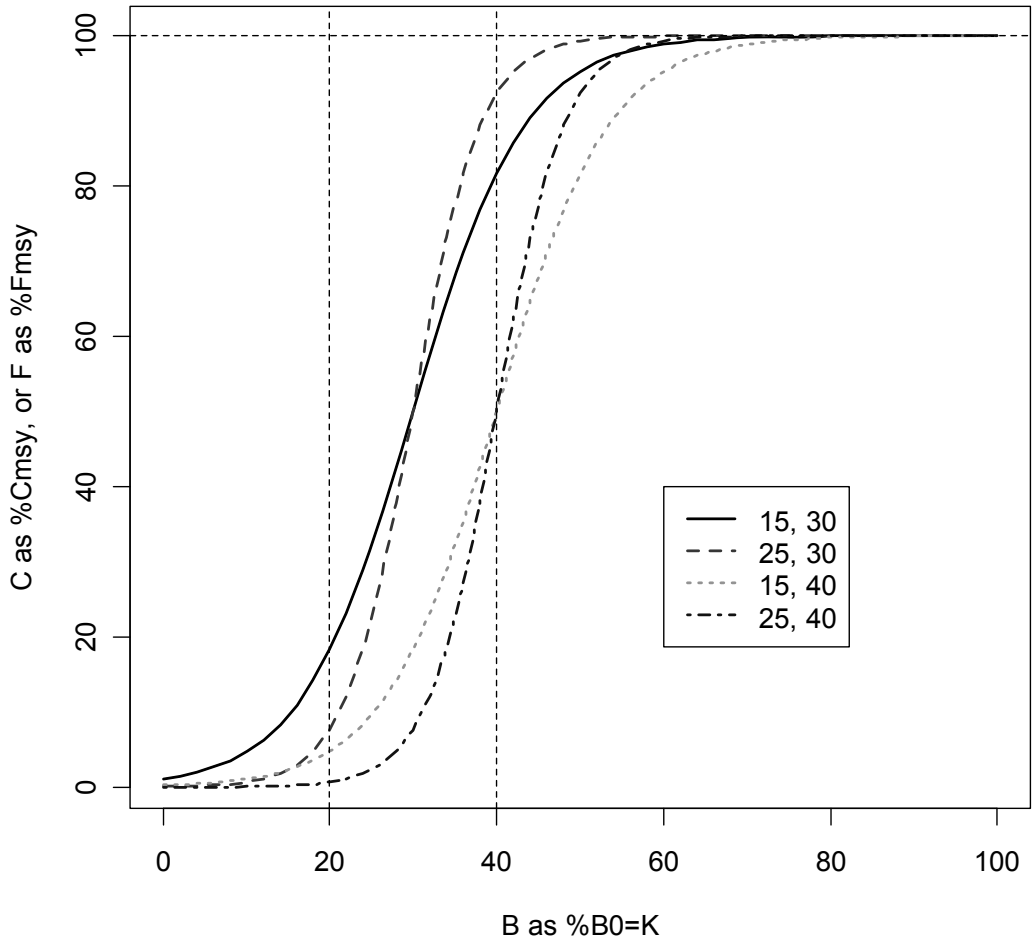


Figure 127. Sigmoid decision rule curves which could represent (and are similar to) 10-40 or 20-40 decision rules. The versions with $\widetilde{B}$ =30 are closer to the 10-40 or 20-40 rules, but the versions with $\widetilde{B}$ =40 are meant to reflect the notion of precaution which is sometimes implemented via a rule which reduces catches at higher levels of B/B_0.

Furthermore, C_{msy} or F_{msy} are used to set the maximum catch or harvest rate in the sigmoidal decision rules. These values can be thought of as upper 'extremes', given the notion that MSY-related quantities, such as C_{msy} or F_{msy}, should be treated as limit, not target, reference points. Note that I assume the decision rules are correctly specified, i.e. there is sufficient knowledge that C_{max} or F_{max} and $\widetilde{B}$ are reasonably close to the true correct values. It is of course likely that any of these parameters could inadvertently be set incorrectly, and as seen above, if F_{max} or C_{max} are set too high there is an increased likelihood of complex dynamics.

The versions of the decision rule with the additional time lag is used because this version is more likely to show complex dynamics and is also, arguably, likely to be more practical or feasible. (It is, for example, more likely that data from year t-1 would be used to determine a quota or level of effort for year t)

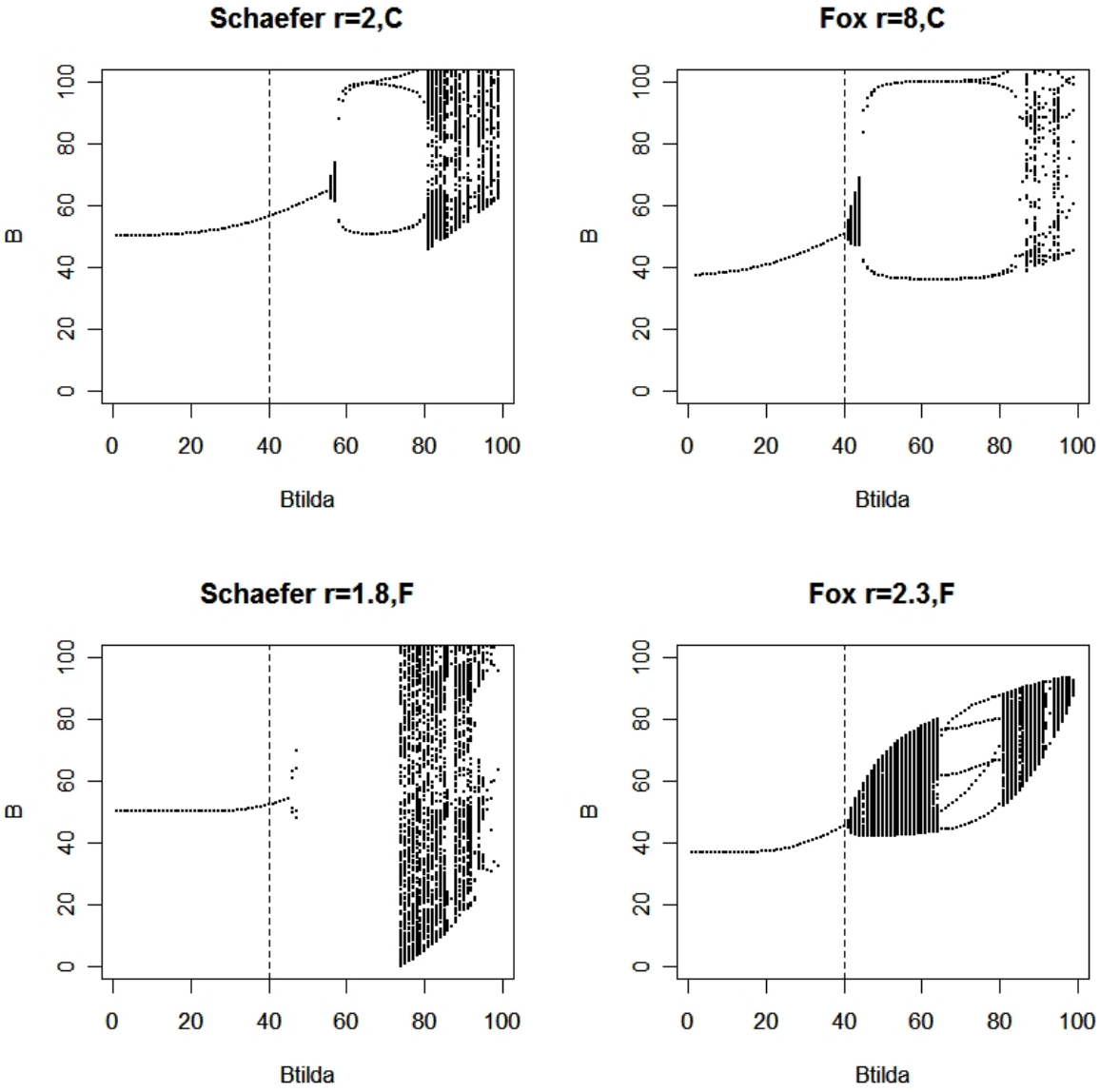


Figure 128. Bifurcation plots showing approximate maximum r-values (given in headers) which avoids limit cycles and complex behaviour for $\widetilde{B}$ <=40% of B_0 with α =25 in both the Schaefer and Fox stock-production models and for Catch-based rules (top row) and F-based rules (bottom row).

Figure 128 was obtained by taking the Schaefer and Fox models with either a catch or an F-based decision rule[31] and with the additional time-lag, i.e. based on B_{t-1}. Bifurcation plots where then generated for α =25 (and obviously over the full range of possible $\widetilde{B}$ - values) while r was increased to see at which value of r the complicated dynamics started just above $\widetilde{B}$ =40. This was simply done by eye for illustrative purposes. For the first panel (Schaefer, Catch) results at r=2 are shown since strange dynamics appear when r>2 (see May & Oster 1976), but even a

[31] Schaefer models had C_{max}=rK/4, and F_{max}=r/2; Fox models had C_{max}=(r/lnK).Ke^{-1} and F_{max}=r/lnK

value of r=2 is arguably on the high side for most harvested fish stocks. Results reconfirm that the Fox model starts showing complex dynamics at a higher value of r than the Schaefer model. This is related to the fact that the Schaefer model has a much higher maximum production for a given r-value than the Fox model although the slope at the origin (i.e. the rate of increase where the biomass is close to zero) is the same for the two models. The rate of increase in the Fox model needs to be much higher in order to achieve the same maximum production as a Schaefer model.

The harvest-rate version of the decision rule shows complex dynamics at a lower value of r than the catch-based version (for the same population model). Results suggest that, under most circumstances, for relatively modest rates of increase (below, say 1.5), the 20-40 rule, even with $\widetilde{B}$ set at 40% of B_0 rather than around 30% of B_0 should have a low probability of entering complex dynamics such as point or limit cycles. Care however needs to be taken when the underlying population model is thought to be more like the Fox than the Schaefer model (i.e. B_{msy} closer to 35% of B_0 than 50% of B_0), particularly if the DR is specified in terms of harvest rate rather than catch. Having note this, it is important to remember that the decision rule may not have been "correctly specified", because of the uncertainties about the main parameters in the system and/or lack of knowledge. It is under such circumstances that complex behaviours could still occur, particularly if C_{max}, F_{max}, and/or $\widetilde{B}$ are set too high.

A full exploration of the second question, relating to oscillations and stochasticity, is beyond the scope of this paper and is more usefully considered within the context of a particular fish stock or case study. It is, however, simple enough to show some examples.

The figures below were generated by including a log-normal error term on the population dynamics. This process error can be thought of as recruitment variability or variability in annual productivity. Fish stocks typically have recruitment CVs between 20 and 40%, or even higher. I have used 30% here for illustration, noting that this level is not insubstantial. In the first example, Figure 129, shows a comparison between a stable equilibrium with 30%CV process error, and a limit cycle with 30%CV process error. It is clear from the autocorrelation panel in Figure 129 that the periodicity of the limit cycle remains. In this example, the stochasticity does not 'swamp' the cycles.

Figure 130 shows an example where the process error does to some extent at least damp out or obscure the cycles as is most clearly seen in comparing the middle and right-hand autocorrelation panels.

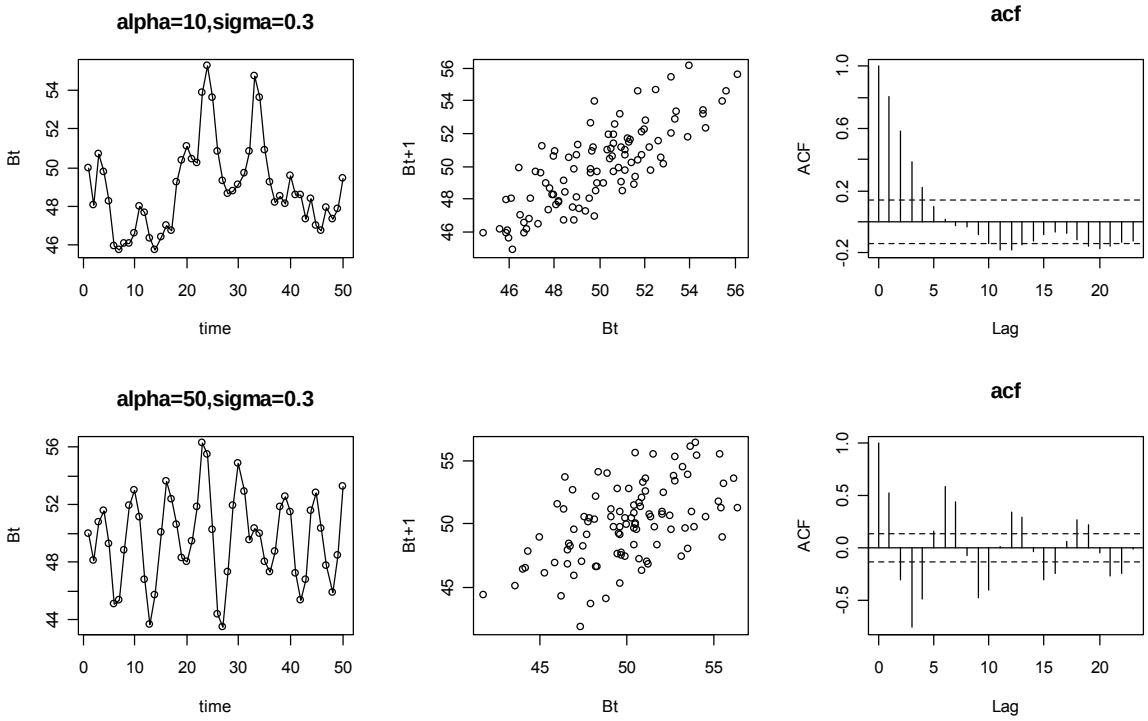


Figure 129. Schaefer model with sigmoid DR and 30% CV on process error. The top row is for a stable equilibrium and the bottom row for a limit cycle. The left-hand panels are time-series, the middle panels are phase plots of B_{t+1} versus B_t and the right hand panels are autocorrelation plots for a range of time lags.

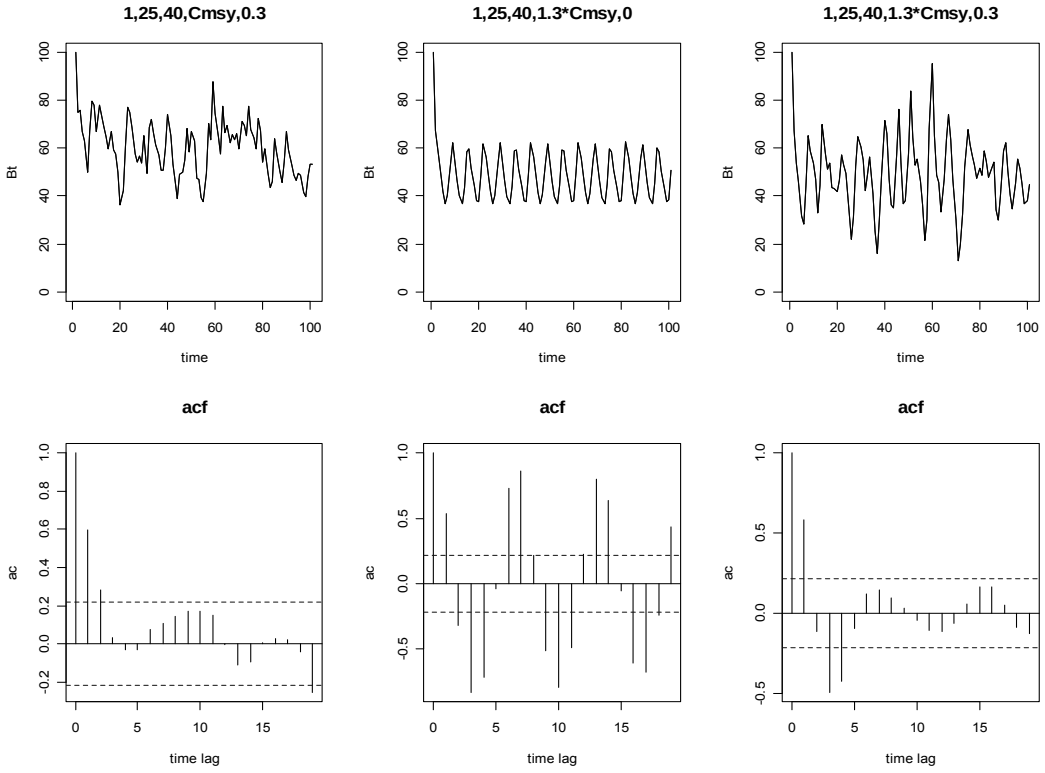


Figure 130. The pairs of panels (top and bottom) illustrate the time-series and autocorrelation plots for: left-hand pair= 30%CV process error on a stable equilibrium; middle pair = limit cycle with no process error; right-hand pair the same limit cycle with 30%CV process error. The titles above the top panels give the parameters: r, α, $\widetilde{B}$, C_{max}, σ. In all examples K=100.

11.6. Discussion and Conclusions

Although the above examples may seem abstract and theoretical, they provide important insights into the potential pitfalls of decision rule design. These insights are particularly relevant where there is insufficient time to test a decision rule before implementation, and in situations where (for whatever reason) very simple rules are being considered. This is not to say that simple rules perform poorly, but rather to highlight that simple rules cannot be assumed to perform well by the mere fact that they are 'simple' rather than complex. The insights are also highly relevant to cases where there is limited data, or very uncertain information about quantities such MSY, B_{msy}, or other related reference points. Also, the insights are not only relevant to situations where CPUE is used as an index of abundance, but also for situations where other indicators (e.g. mean size or mean weight in the catch) can be assumed to be approximately linearly related to biomass. As noted at the start, the proportionality coefficient which relates the indicator to the biomass (e.g. catchability, q, in the case of CPUE) is subsumed in the α parameter of the decision rule examples shown here.

The insights gained by this investigation include the following:

- A decision rule (DR) which simply responds to increases or decreases in abundance, is highly unlikely to lead to rebuilding a depleted stock, or to increased catches of a lightly fished stock
- such a DR will either keep the stock approximately at the same level as at the start of implementation of the DR,
- or could lead to unexpected large changes depending on whether that level coincides with a stable or unstable equilibrium.
- Apparently subtle differences in the formulation of the DR are important to the dynamics and stability conditions (e.g. comparison between examples 1,2,3 and between 4,5,6).
- The multiplicative version (examples 3, 6; see Table 1) has the best stability characteristics because for a given strength of response (α), any level of biomass could be a stable equilibrium if the inherent rate of increase, r, is not too high.
- There are also clear differences between a similarly specified catch- and F-based rules, but the details of the differences depend on how the rule is specified, and whether they can be compared directly.

When a sigmoid decision rule is constructed to mimic the familiar 20-40 or 10-40 type decision rule, the main insights are the following:

- This type of rule has a built-in target and equilibrium biomass which may or may not be stable.
- Depending on the parameters that specify the rule and the rate of increase in the stock, there may be multiple equilibria (this means that the stock could end up at a lower level than intended).
- This DR can also lead to complex behaviour, even for a stock with low or modest rate of increase, if the maximum catch (C_{max}) or harvest rate (F_{max}) is set too high, the inflexion point ($\widetilde{B}$) is set too high, or the sigmoid curve is too steep (α is high). The sigmoid curve can be too steep if the target and limit reference points are too close.
- This suggests that the maximum catch or harvest rate should be set conservatively, but care should be taken not to set $\widetilde{B}$ (or the Biomass reference points) too conservatively (i.e. too high) if point and limit cycles are to be avoided
- For this type of rule, the catch-based version has better stability characteristics than the harvest rate version for similar specifications in other respects.

It is not surprising that the population model itself, for example the Schaefer or Fox model, makes a difference to the dynamics and stability characteristics, since the dynamics for the unharvested versions of the models differ. It is also not surprising that oscillatory behaviour, point or limit cycles may not matter if they are small relative the stochasticity in the system. It is also, however, easy to show that there may be cases where the cycles would not be obscured by the stochasticity. This means that the variability in the catches would not just be random, but follow cycles, possibly with a much wider range than natural variability alone would induce.

This work shows the potential value of quick, simple and very basic simulation (or analytical where feasible) exploration of proposed decision rules before implementation, or at the very least, during the first year of implementation. Not all feedback behaviours, of even such simple systems as those considered here, are intuitive or easy to foresee.

Finally, it is interesting to note that this work has pointed to the likely culprit of the observation which triggered this work: cycles in a simulated age-structured population model following the introduction of a decision rule. The decision rule was based on mean size as indicator and constructed like a 20-40 rule. The two reference points for mean size were based on the mean size coinciding with a harvest rate that gives 20% or 40% of unexploited spawner per recruit respectively. Since mean size changes very slowly as the harvest rate changes, the two reference points tend to be very close together. This in turn implies a high α value in the formulation used here, which, as we've seen, tends to lead to complicated behaviour.

12. Implications of sexual dimorphism in swordfish growth for size-based indicators

Summary

This brief comparison of gender-specific and pooled size-based indicators did not suggest high priority for looking at gender-specific size frequency data when constructing indicators. The form of dimorphism, i.e. exactly how the male and female curves differ, seems important. Under some growth patterns (e.g. divergent), where the male and female curves are markedly different, it could be beneficial to consider indicators based on data for females only. This is because in swordfish, the female growth rates are generally lower (although their maximum size is usually higher) than the male growth rates. If this issue is to be explored further, it would be worth having reasonably good data on actual growth patterns.

Although results suggest that it is not essential to collect gender-specific size data, there are other important reasons why size data should be collected by gender if at all possible. This is because of the need for an indicator time-series to be standardised, and the fact that, in the presence of dimorphism in growth, the size-based indicators may be sensitive to changes in the sex ratio of the catch. If only pooled size frequency data are available, it may be impossible to standardise the size-based indicators. If information on the sex ratio in the catch is available, this may help to some extent.

It must also be emphasised that these analysis do not take into account issues such as spatial structure by size or gender. As hinted above, there are potential problems with the standardisation and use of size-based indicators if there is gender-based spatial and/or temporal disaggregation. For example, a time series of mean length would be invalid, and difficult to standardise if this was comprised of mean lengths obtained from different areas with very different gender ratios.

The brief sub-section on decision rules illustrates the way in which a size-based indicator could be used in a decision rule. The differences in performance of the stock under the 7 scenarios underline the importance of designing decision rules that are robust to uncertainty in the underlying dynamics. Although we have only illustrated this for uncertainty in growth, taking uncertainty in steepness and mortality into account will also be important.

12.1. Introduction

This Section briefly explores the likely effect of sexual dimorphism in swordfish growth on size-based indicators. This was not an explicit objective of the project, but we considered it a potentially important factor to flag, at the very least, and to explore briefly during the early stages of the project.

Given the evidence of potentially strong sexual dimorphism in growth of broadbill swordfish (e.g. Sun *et al.*, 2002), we have explored the behaviour of size-based indicators in this context, and we compare the behaviour of indices based on pooled data or data for males and females separately. Only a subset of results have been extracted for presentation here, and we have focused on those aspects most likely to be relevant to broadbill swordfish.

12.2. Methods

Descriptions of methods have only been summarised here. Full technical details of the simulation model may be found in Appendix 1 (and ignoring the spatial structure in the equations). A note with respect to F_{msy} is necessary, however. We have used F_{msy}, and proportions or multiples of F_{msy}, in many of our explorations. The main reason for this is that the response of an indicator to exploitation is directly and strongly affected by the harvest rate relative to the life-history parameters (Section 8). As such, F_{msy} is a convenient quantity which retains the scaling between harvest rate and life-history parameters and hence make results comparable. The fact that we have used F_{msy} in these explorations should not be interpreted as a suggestion that F_{msy} is an appropriate target harvest rate. The appropriateness of alternative harvest rates are best explored in the context of feedback decision rules using a range of performance measures, including the risk of low spawning biomass (SSB[32]).

Erratum: References to 'Mantiel' in figures should read 'Montiel'. References in the text are correct, and are to Montiel, 1996 (as cited in Ehrhardt 1999) and detailed in Section 18.

12.2.1 Indicators

The following indicators were calculated:

- the mean, median, 95th percentile of length and weight distributions in the catch
- the proportion of "small" fish in the catch (defined as the proportion of fish $\leq t_c+1$)
- the proportion of "large" fish in the catch (defined as the proportion of fish $\geq$170 cm TL[33])
- the proportion of mature fish in the catch (defined as the proportion of fish older than the age-at-maturity, t_m).

Results presented here do not include the 'proportion small fish' indicator because this indicator increases rather than decreases when biomass decreases and requires somewhat different treatment and interpretation.

12.2.2 Equilibrium considerations for sexual dimorphism in growth

For this investigation we include the complication of sexual dimorphism in growth and variability in length-at-age into the population model. It is highly likely that sexual dimorphism exists for Indian Ocean swordfish, although the extent and exact form of dimorphism in growth for Indian Ocean swordfish is not well known, as far as we are aware. The reasons for considering these models are the same as for the simpler deterministic models, but they should provide a somewhat more realistic reflection of the likely dynamics of swordfish.

The literature provides several estimates of growth for male and female swordfish from many oceans (see e.g. Radtke and Hurley, 1983; Tserpes and Tsimenides, 1995; Vanpouille *et al.*, 2001; Sun *et al.*, 2002). Young and Drake (2004) provide estimates for the west coast of Australia.

We examined these growth curves and noticed that male and female growth curves can be divergent (qv Berkeley and Houde, 1983), intersecting (qv Montiel, 1996, as cited in Ehrhardt, 1999) or parallel (qv Ehrhardt *et al.*, 1996). We chose the three cited sets of curves as being the most extreme examples of each type of pattern, and the Young and Drake (2004) growth curves (assuming age-at-maturity of 5 years) were included as being the only estimated curves for Indian

[32] We use SSB and B interchangeably to indicate spawning biomass unless otherwise stated

[33] This is an arbitrary choice and more comments are made regarding the choice of cut-off size in the results section.

Ocean swordfish (Figure 131; Table 23). We used these to evaluate the response of basic stock status indicators to alternative forms of sexually dimorphic growth. We also investigate the difference between male, female and pooled indicators, to determine whether sexual dimorphism implies that it better to consider gender-specific indicators. Note that the Ehrhardt et al. (1996) growth curves were not used to compare gender-specific indicators, since the male growth curve was such that no males ever reached fishable size under the selectivity used in the simulation.

Table 23. Parameters for von Bertalanffy growth curves from literature

Reference	Location	Female			Male		
		L_∞	k	t0	L_∞	k	t0
Young and Drake (2004)	Indian Ocean	296.51	0.0815	-3.0148	236.9	0.1096	-3.0118
Berkeley and Houde (1983)	Straits of Florida	340.04	0.09465	-2.5912	217.36	0.1948	-2.0444
Montiel (1996)	Chilean coast	282	0.2925	0.11	250	0.3216	-0.75
Ehrhardt et al. (1996)	Northwest Atlantic	364.69	0.0262	-0.556	189.58	0.0105	-0.41

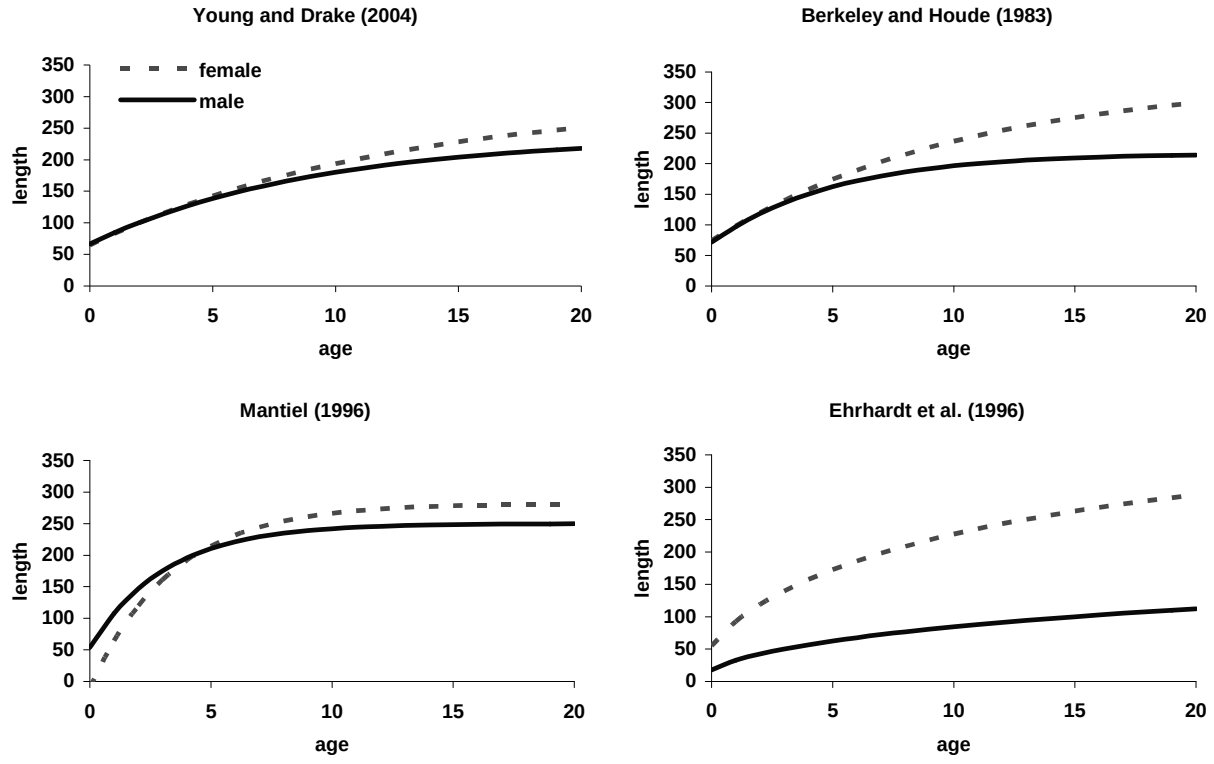


Figure 131: Male and female length-at-age for the 4 sets of swordfish growth curves from the literature

To cover the range of patterns seen in estimated sexually dimorphic growth curves, we used the 'Young and Drake' growth curves, which are in fact very similar for males and females, and we also constructed 3 'caricature' or stylised hypothetical dimorphic growth curves to explore. Given the divergent, intersecting and parallel patterns noticed from the curves in the literature for male and female growth curves, we have constructed the 'caricature' curves to represent those three patterns (Figure 132, Table 2), but we have exaggerated the extent of dimorphism. Of main concern, again, is (a) whether sexual dimorphism implies that it is always better to consider male and female indicators separately, or whether pooled data can be used (noting the difficulties associated with collecting size data by gender) and (b) whether the type of dimorphism (divergent, intersecting, parallel) matters.

In the results presented here, selectivity for "longline-like" fishing was knife-edged at 120cm TL. Length-at-(quarterly)-age for each gender was assumed to follow a normal distribution with mean calculated from the von-Bertalanffy growth curve, and standard deviation (σ_a) a linear function of the square root of the (annual) age a: $\sigma_a = 5.4 \cdot \sqrt{a/4} + 6.5$. This relationship was loosely based on information in Young and Drake 2004. Selectivity at age for each gender was then calculated as the integral of this normal distribution above the knife-edge length of 120cm TL. As such, the age at first capture depends on growth curve and gender.

Reproductive studies by Young and Drake (2002) reported size at 50% maturity for females to be approximately 200cm. For the Young and Drake growth curve (2004) this corresponds to an approximately 9 - 10 year old female, but for the stylised growth curves, this length corresponds to lower ages (3-6 years). As such, an age at maturity of 5 was used across all growth curves, for males and females, but for the Young and Drake growth curve, an additional model was run where the age at maturity was set to 9 years[34]. The model assumes knife-edged maturity at the gender-specific length corresponding to the age-at-maturity (5 years or 9 years). As with selectivity, the proportion mature-at-(quarterly)-age was then calculated for each gender as the integral of the normal distribution of length-at-age above the knife-edge length at maturity.

The sex ratio among recruits was assumed to be equal (50% per gender). Values for steepness in the stock-recruitment relationship were h=(0.3, 0.6, 0.9) and for natural mortality were M=(0.1, 0.25, 0.3). Results presented here are only for the mid-points: h=0.6 and M=0.25. These two parameters are likely to be the least well known of the life-history parameters and in the study as a whole, we therefore consider the ranges of values. F_{msy} was calculated for the combined catch of males and females.

The following indicators were derived from the "sampled" catch, in addition to those mentioned above:

- the mean length of "large" (≥170cm TL) fish in the catch
- the mean weight of "large" (≥170cm TL) fish in the catch

Indicators were calculated both separately for males and females, and for the catch sample pooled across genders.

To summarise, the five scenarios considered in this sub-section are:
1. Divergent growth, t_m=5

[34] There is clearly still large uncertainty about life-span and age at maturity for swordfish. For an assumed natural mortality rate of 0.25, which gives average life-span 1/m of 4 years, maturity at age 9 (when only about 10% of the year-class is still expected to be alive under no fishing) seemed rather high. This is why we included the scenario of the Young and Drake growth with age-at-maturity of 9.

2. Parallel growth, t_m=5
3. Intersecting growth, t_m=5
4. Young and Drake growth, t_m=5
5. Young and Drake growth, t_m=9

Table 2: Von Bertalanffy growth parameter values for the stylised sexually dimorphic growth curves, and the parameter values reported by Young and Drake (2004) for swordfish from the west coast of Australia

	Female			Male		
	L_∞	k	t0	L_∞	k	t0
Young and Drake (2004)	296.51	0.0815	-3.0148	236.9	0.1096	-3.0118
Divergent	350	0.1	-2	200	0.2	-2
Parallel	350	0.2	-1.5	200	0.2	-1
Intersecting	350	0.15	0	200	0.3	-1.5

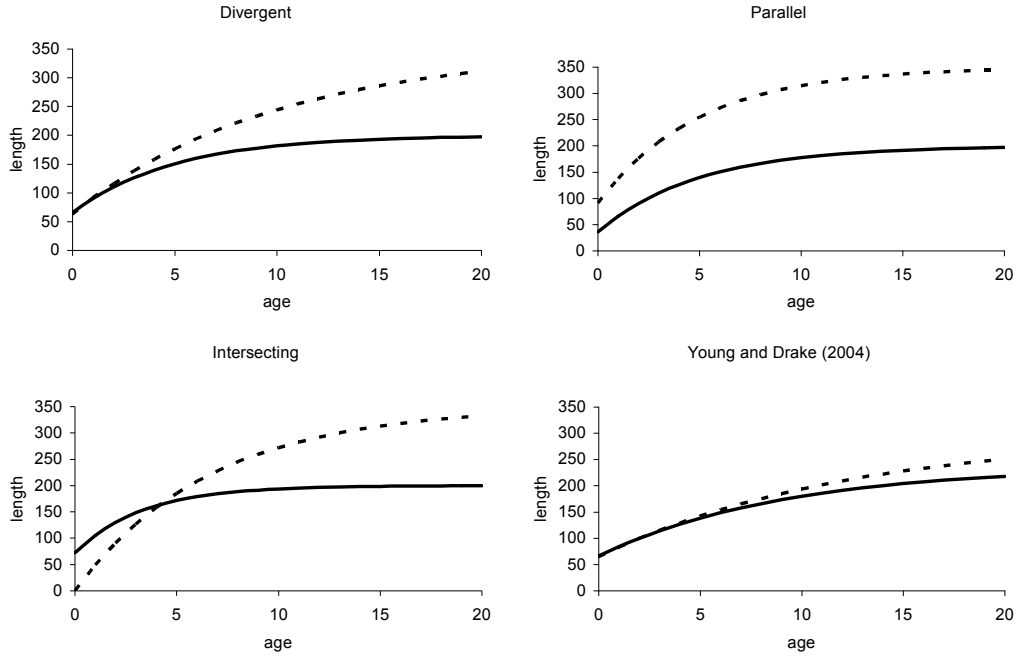


Figure 132. Stylised growth curves for sexually dimorphic male (solid line) and female (dashed line) growth, together with the growth curves reported by Young and Drake (2004) for swordfish from the west coast of Australia.

12.2.3 Behaviour of time-series of indicators

The models with sexual dimorphism in growth were next run with stochasticity introduced into recruitment (CVs of 5%, 20% and 40%) and with three sampling levels: good (100% of the catch sampled), poor (2% of the catch sampled) and very poor (100 individual fish sampled from each gender). For each model and sampling level, 100 replicates of 30 years each were run. The initial stock status was always unexploited, and levels of harvest were again $0.5F_{msy}$, F_{msy} and $2F_{msy}$, though results here focus on F_{msy}.

'No-change error' summaries

The 30-year time-series of each indicator was analysed by taking increasing periods of time (first 5 years, first 10, 15 and 20 years) and estimating the slope of the indicator over time by simple linear regression. The number of times the slope is NOT significantly different from 0 is then

counted and expressed as a % of the total number of replicates (100). A high count means that an indicator is unlikely to signal a change; a low count means that it usually picks up some signal. In the rest of this document, the number of non-significant slopes is called the 'no-change error'. Clearly, this approach does not reflect the magnitude of change – for example the mean length may only drop by 10% for a drop of, say, 60% in the SSB – but it is meant to show the difference in behaviour for different levels of (a) recruitment variability and (b) sampling (good, poor and very poor). It is again within the feedback rules that the issue of magnitude of change in an indicator relative to that in SSB will be addressed more fully.

This approach also makes a simplistic assumption of a linear change in indicators. Although the response is generally slightly curvilinear (changing faster at the start of exploitation and then slowing down), the deviation from linearity is relatively small compared to that for SSB, which is very obviously non-linear. This approach is simply being used as a tool for comparison and summary.

'GAM MSE' summaries

The time-series of indicators for 100 realisations were also analysed with respect to their relative ability to predict changes in biomass. As noted before, the biomass changes in a very non-linear way, particularly at the start of exploitation on a virgin stock. This analysis was therefore done by fitting a general additive model (GAM) to the indicator and simulated biomass, as follows:

$$dB_t \sim s(dI_t) + s(I_t)$$

where
B_t is the simulated spawning biomass at time t and
$dB_t = (B_{t+1} - B_t)/B_t$ is the relative change in simulated SSB from time t to t+1
I_t is the value of the indicator at time t and
$dI_t = (I_{t+1} - I_t)/I_t$ is the relative change in the indicator from time t to t+1
's' indicates a spline-based smooth in the GAM model

Both the change in the indicator and the level of the indicator is therefore used to predict the change in the SSB (this was found to be substantially better than just using the change in the indicator). Although this may not lead to the 'best' model fit in all cases, the approach is useful in a simulation context where it is impossible to fine-tune the fit for every realisation. This approach also leads to comparable results for all indicators and all realisations, and is therefore again a useful tool for comparison and summary. We use the mean squared error (MSE, i.e. the mean of the squared difference between the simulated SSB and model predicted SSB) as a measure of the performance of each indicator. Low values of the MSE indicate good performance.

12.2.4 Behaviour of single-indicator decision rules

Simple decision rules based on mean length were investigated to determine how well a single indicator might perform when used to manage a fishery. While various rules will ultimately be considered, that presented here is a variation on a simple "40/10" rule (usually defined in terms of biomass or CPUE), whereby fishing mortality is constant above a threshold indicator level, zero below a limit indicator level, and declines linearly between the two indicator values. The indicator, in this example mean length based on pooled data, is calculated at the end of each year and the coinciding harvest rate is determined from the decision rule. This harvest rate is then applied to the population to generate catch in the following year. Although the decision rule is formulated in terms of fishing mortality, a conversion to a TAC can easily be made in the model.

We assume that management is perfectly implemented (i.e. the harvest rate from the decision rule is applied to the population without error), and that estimates of quantities such as biomass are available (and unbiased) if management is through TAC. We recognise that this represents ideal and somewhat unrealistic conditions. The results presented are, however, intended to be comparative rather than absolute. We emphasise that the work on decision rules is still underway, and issues such as implementation error and the availability of estimates for key quantities are being considered.

In our case the decision rule was defined as shown in Figure 133. Fishing mortality was set to F_{msy} above a threshold value expressed as a fraction of the expected mean length of the unexploited stock. This fraction was chosen to be 0.88, corresponding to the ratio of the expected mean length at F_{msy} to the 'virgin' mean length for a population growing according to the Young and Drake von Bertalanffy growth parameters. The limit indicator level was set at the (arbitrary) relative mean length of 0.81, corresponding to a length of 130cm (i.e. 10cm above the knife-edged selectivity length assumed in the simulations).

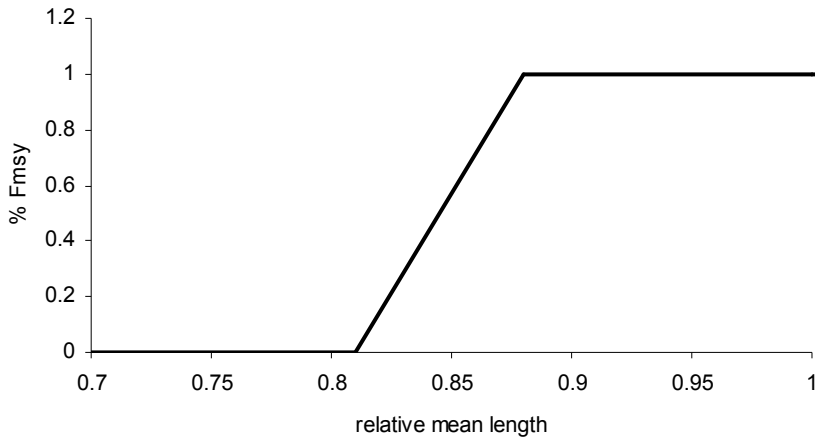


Figure 133. Schematic illustration of the simple decision rule based on mean length (relative to virgin mean length)

Additional scenarios were run where the threshold indicator level was increased to a relative mean length of 0.93. Under this decision rule, fishing mortality was lowered earlier and by a greater amount relative to the previous threshold value. The decision rule was invoked after 5 years of fishing at 1.6 F_{msy}. This level of fishing mortality was chosen so that the impact on the stock after 5 years would be substantial.

Scenarios were based on the two sexually dimorphic growth curves showing the most extremes in F_{msy}. Of the four growth curves considered, populations with the Young and Drake growth curve had the highest F_{msy} (0.235) while populations with the stylized "parallel" sexually dimorphic growth had the lowest F_{msy} (0.145). Note, however, that the relative thresholds, i.e. inflection points, for the decision rule remained at the values described above.

In order to gain insight into how a decision rule might behave when the population grows in a different way to that assumed in the construction of the decision rule, scenarios with the "correct" decision rule (i.e. F_{msy} coinciding with the underlying growth curve), as well as with a "wrong" decision rule (i.e. F_{msy} NOT coinciding with the underlying growth curve) were run (see below).

As a reference case, and for comparison, one scenario had no feedback decision rule and the population was harvested at $1.6F_{msy}$ for the entire 30 years:
Scenario 1: Young & Drake, $1.6F_{msy}$ (based on Young and Drake)

Six scenarios with decision rules were run. The scenarios are in pairs, with one using the lower threshold (0.88, as in Figure 2), the other using the higher (0.93) threshold, and with the following combinations of growth curve in the stock dynamics and F_{msy} in the decision rule:
Scenarios 2 and 3: Young & Drake growth with its own F_{msy}
Scenarios 4 and 5: Parallel growth curve with its own F_{msy}
Scenarios 6 and 7: Parallel growth curve with the Young and Drake-based F_{msy}

Results presented here are based on 100 replicates of 30 years with steepness of 0.6, natural mortality of 0.25, and the standard deviation for recruitment 0.4. Perfect (100%) catch sampling was assumed, and all outputs considered were those pooled across gender.

12.3. Results

12.3.1 Equilibrium considerations for sexual dimorphism in growth: growth curves from the literature

We first consider the effect of the form of the growth curves on the responsiveness of the pooled (i.e. combined male and female) indicators (assuming perfect sampling of the catch). Note that changes in indicators relative to their unfished values are plotted, and these should not be confused with the absolute values of the indicators. An indicator may be higher in value for a given set of growth curves than another, but if its relative change with fishing mortality is greater, its relative index value will be lower.

There are differences in the magnitude of the relative change in indicators between the growth curves (Figure 134). The magnitude of the relative changes increases with increasing fishing mortality. The difference in relative responsiveness between growth curves is strongest for the indicators proportion of "large" fish (which had over 20% difference in relative change between the Young and Drake (2004) and the Montiel (1996) growth curves at $F=2F_{msy}$), proportion of mature fish, and relative mean weight. It should be noted that the threshold defining "large" fish is fixed at 170cm, and this value will obviously have different implications for different growth curves.

Mean length and mean weight show the strongest responsiveness under the Montiel (1996) growth curves, and the least responsiveness under the Young and Drake (2004) growth curves (Figure 134). However, the proportion of mature and proportion of "large" fish show the most responsiveness for the Young and Drake (2004) growth curves and the least for the Montiel (1996) growth curves. It is noted that the Young and Drake (2004) growth curves showed the least response in terms of absolute spawner biomass, while the Montiel (1996) showed the most. Additionally, the growth rate (k) was highest for both males and females for the Montiel (1996) growth curves.

Relative changes in gender-specific indicators are plotted for each growth curve in Figure 135, Figure 136 and Figure 137. It is reiterated that these figures show the relative change in the indicator with respect to its value under no fishing. Values for absolute mean length, for example, would be higher for females than for males, but because the relative change in the indicator is larger, the relative index is smaller (hence lower values in Figure 135, Figure 136 and Figure 137).

The lower growth rate (k) but higher L_∞ for females than males leads to more response in indicators based on data for females than indicators based on data for males. The reverse would be true if both L_∞ and k for females were higher than for males. The indicator values based on pooled data (males and females) are intermediate. However, when catch is sampled perfectly, there was little difference in the responsiveness of gender-specific indicators for the Young and Drake (2004) (Figure 135) and Montiel (1996) (Figure 137) growth curves. This was particularly true for the proportion mature, proportion of 'large' fish, mean length and upper 95th length percentile indicators.

The Berkeley and Houde (1983) growth curves show more difference in responsiveness between gender-specific indicators, particularly for the upper 95th length percentile, mean weight and the upper 95th weight percentile (Figure 136). For these indicators, the relative difference in response between males and females is greater than that between the different growth curves (for the pooled indicators). The Berkeley and Houde (1983) growth curves have a divergent pattern (Figure 131), and hence show the most difference between males and females among the three sets of growth curves considered.

In general, however, the responsiveness of indicators, relative to their unexploited values, was much less between genders (male/female/pooled) within a set of growth curves than it was between the type of growth curve (for pooled indicators). This suggests that, in the context of indicator responsiveness, an understanding the nature/pattern of sexually dimorphic growth is more important than considering indicators on a gender-specific basis, although the latter may become important if the growth curves are divergent. However, this interpretation should be treated with caution, especially under high stochasticity or poor sampling of catch.

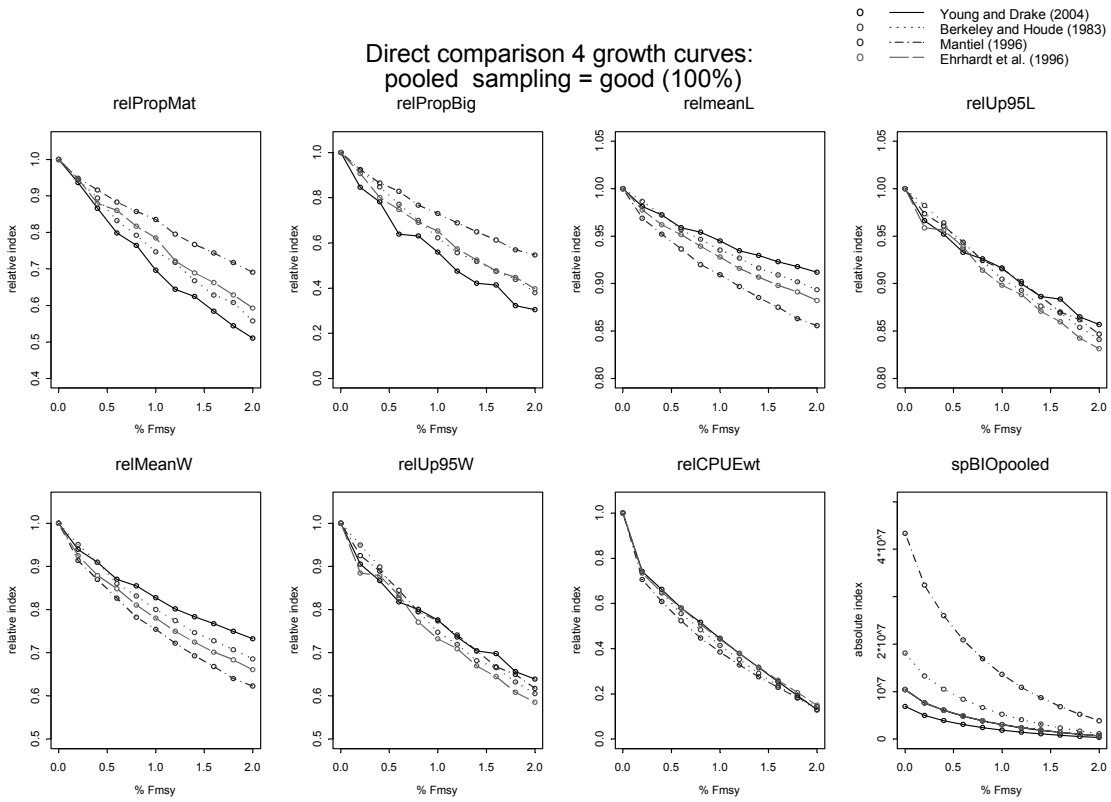


Figure 134. Expected values for indicators relative to their values at F=0 by growth curve for pooled genders. Absolute equilibrium spawner biomass is also plotted. Indicators are: proportion mature (PropMat), proportion big (PropBig), mean length (meanL), upper 95th percentile of length (Up95L), mean weight (MeanW), upper 95th percentile of weight (Up95W), CPUE in weight (CPUEwt).

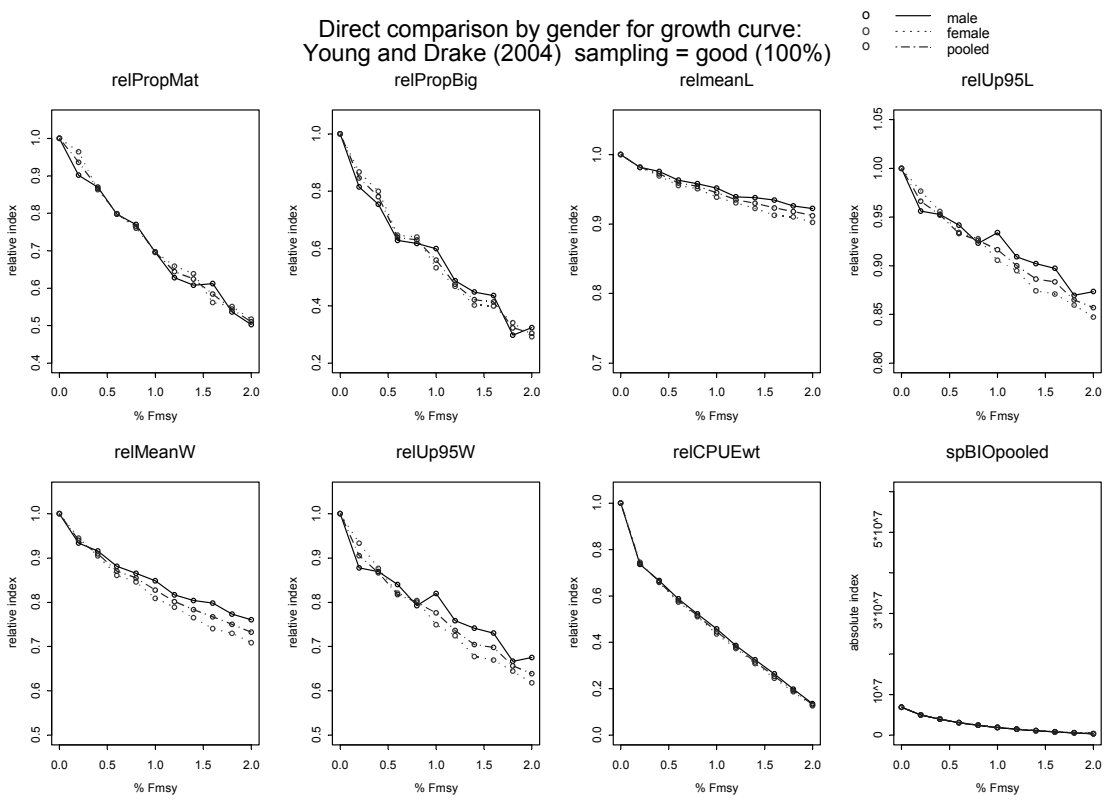


Figure 135. Expected values for indicators relative to their values at F=0 by gender (male, female, pooled), for the Young and Drake (2004) growth curves. Absolute equilibrium spawner biomass is also plotted. Indicators are as in Figure 134.

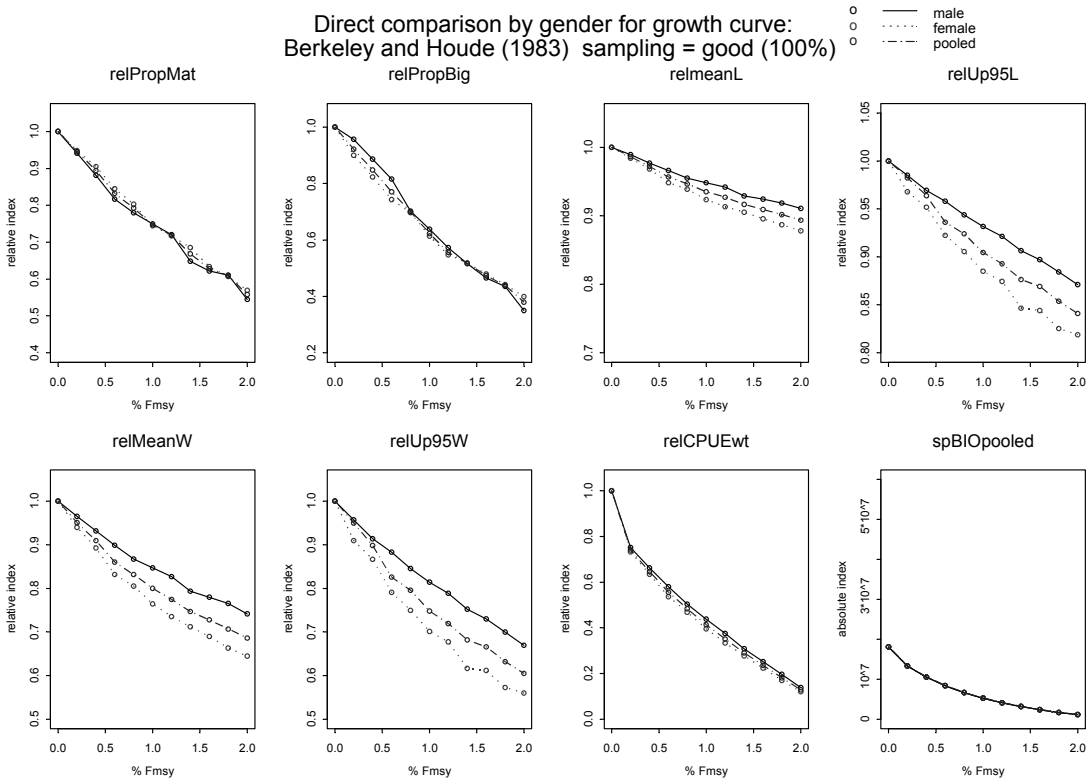


Figure 136. Expected values for indicators relative to their values at F=0 by gender (male, female, pooled), for the Berkeley and Houde (1983) growth curves. Absolute equilibrium spawner biomass is also plotted. Indicators are as in Figure 134.

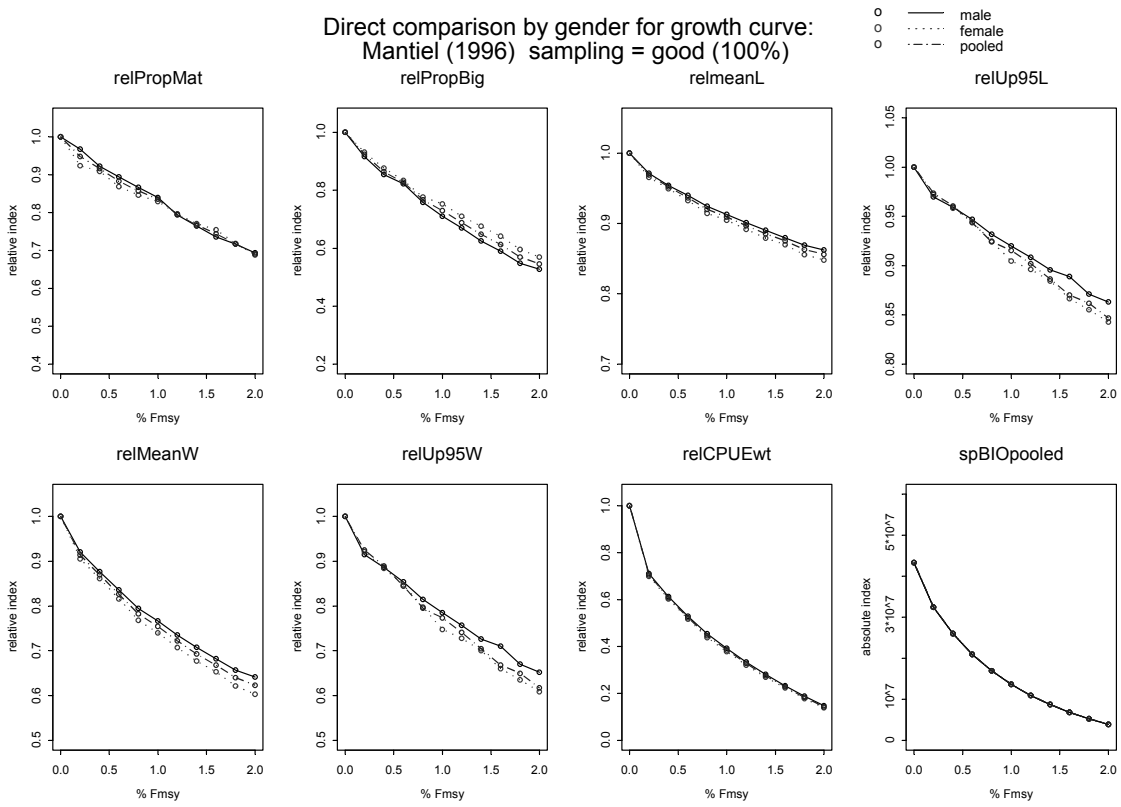


Figure 137 Expected values for indicators relative to their values at F=0 by gender (male, female, pooled), for the Mantiel (1996) growth curves. Absolute equilibrium spawner biomass is also plotted. Indicators are as in Figure 134.

12.3.2 Equilibrium considerations for sexual dimorphism in growth: caricature growth curves

Now we consider the growth curves illustrated in Figure 132. It is not surprising that the same life-history parameters that were important when no sexual dimorphism in growth was assumed (see Section 8), are important here. As with the growth curves from the literature, the suite of growth curves considered here generally have lower growth rate but higher L_∞ for females than males, and that leads to more response in indicators based on data for females than indicators based on data for males. The reverse would be true if both L_∞ and k for females were higher than for males. The indicator values based on pooled data (males and females) are intermediate. They are not, however, necessarily halfway between values for the male and female indicators.

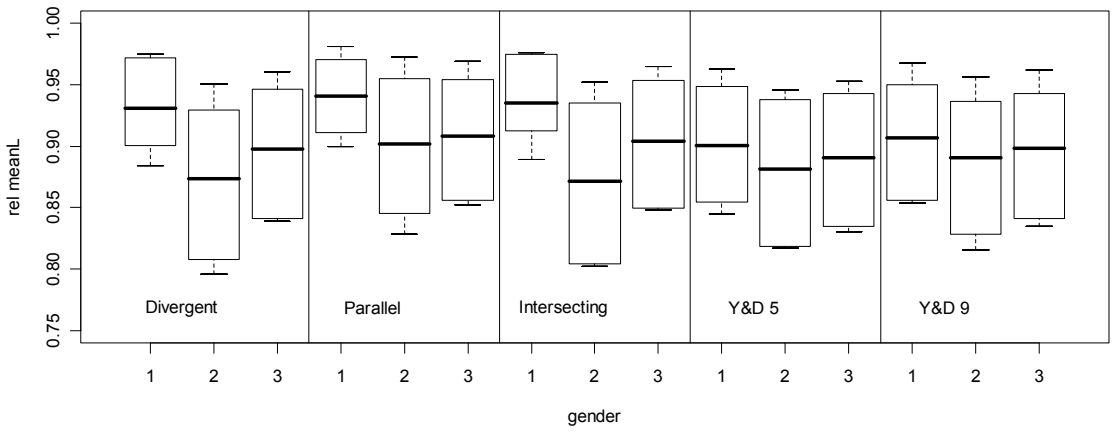


Figure 138. Expected values for relative mean length: meanL(Fmsy)/meanL(F=0) by growth curve and gender. Other factors are steepness, m, and implicitly age at first capture noting that size at first capture is identical in all cases. Gender is indicated by: 1=Male, 2= Female, 3=Pooled. Y&D refers to the Young and Drake growth curves with age-at-maturity of 5 years (Y&D 5), or 9 years (Y&D 9).

Absolute mean length of females would be larger than that for males, but the relative change in this indicator is bigger (hence lower values in Figure 138) for females than for males. The parallel growth curves have the same k for males and females (Table 2), but different L_∞ and t_0 parameters. Results in Figure 138 show that here these two quantities also play a role in the relative response of size-based indicators through their interaction with a single SIZE at first capture (and hence a different AGE at first capture for males and females).

It is also interesting that for the parallel growth curve, the pooled indicator behaves quite similarly to the female indicator. This is probably mainly due to the much lower age at first capture for females than for males which presumably affects the sex ratio in the fishable population. Consistent with the earlier results for the Berkeley and Houde (1994) growth curves (Figure 136), these results suggest that under some growth patterns (e.g. divergent and intersecting) it could be beneficial to consider indicators based on data for females only, though in other cases (e.g. parallel, or Young and Drake which has a low degree of dimorphism) there is apparently not much to be gained, in terms of responsiveness of the indicator, by considering gender-specific indicators. This analysis does not, however, take into account issues such as spatial structure by size or gender, and in practice there may be other important reasons why size data should be collected, and indicators constructed, by gender, if at all possible.

Other size-based indicators show similar patterns to those for relative mean length (Figure 138), though the 95th percentile can sometimes be more responsive for pooled rather than female data (not illustrated). This result should, however, be treated with caution, and the performance of the 95th percentile as indicator should rather be evaluated within the context of stochasticity, sampling and feedback decision rules.

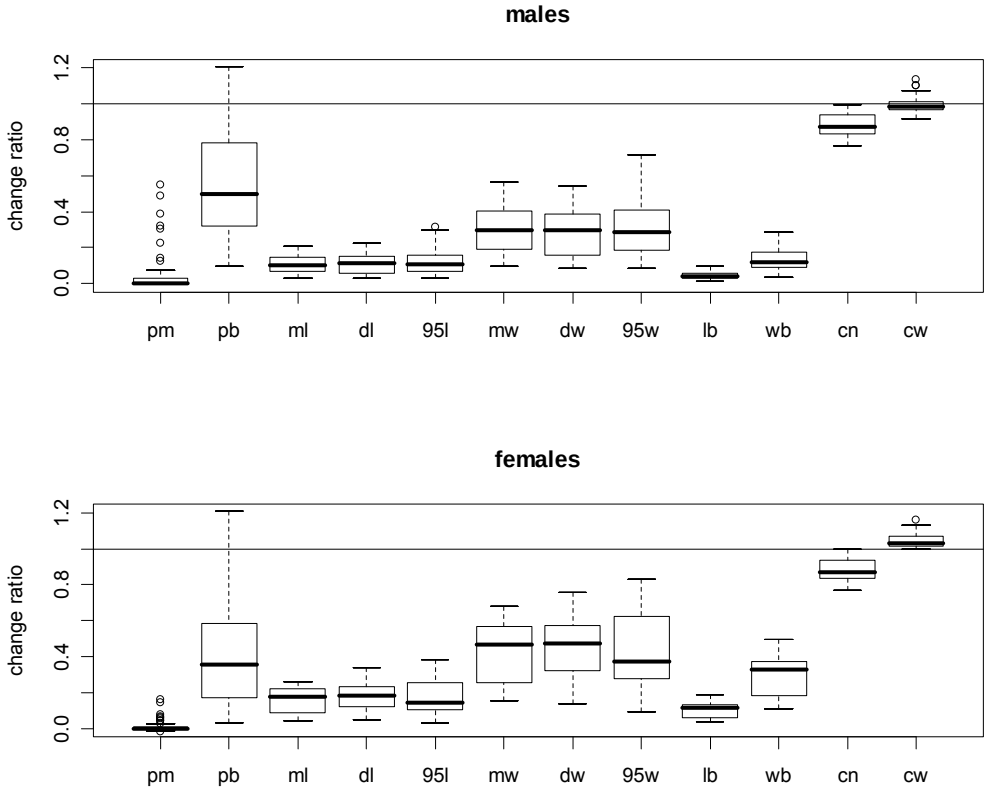


Figure 139 Change in indicator relative to change in biomass, [1 – indicator(Fmsy)/indicator(F=0)]/[1-Bmsy/B0], by gender, for 11 size-based indicators: proportion mature (pm), proportion big (pb), mean length (ml), median length (dl), 95th percentile of length (95l), mean weight (mw), median weight (dw) and 95th percentile of weight (95w), mean length of big fish (lb), mean weight of big fish (wb); and two CPUE indicators: cpue in numbers (cn) and cpue in weight (cw).

A comparison of the expected change in indicators relative to the change in biomass Figure 8is shown for the 5 scenarios combined in Figure 139. Although the responsiveness of a given indicator can differ between males and females, the relative responsiveness of indicators remains essentially the same (e.g. mean length is less responsive than mean weight for both males and females). Recall that quantities near 1 reflect a responsive indicator (values >1 imply that the indicator changed more than the biomass).

It is worth commenting on the 'proportion big' indicator. In all cases this was based on the proportion bigger than 170cm. As mentioned earlier, the choice of cut-off will have quite different implications for the different growth curves, and the results here, which suggest this indicator is only sometimes informative, should therefore be interpreted with caution. Figure 140 (below) shows the change in the 'proportion big' relative to change in SSB separately for each growth curve. This highlights that the wide range of performance is primarily driven by growth curve (and what the choice of size implies for that growth curve). It is likely that the cut-off used in constructing this indicator would have to be chosen with care.

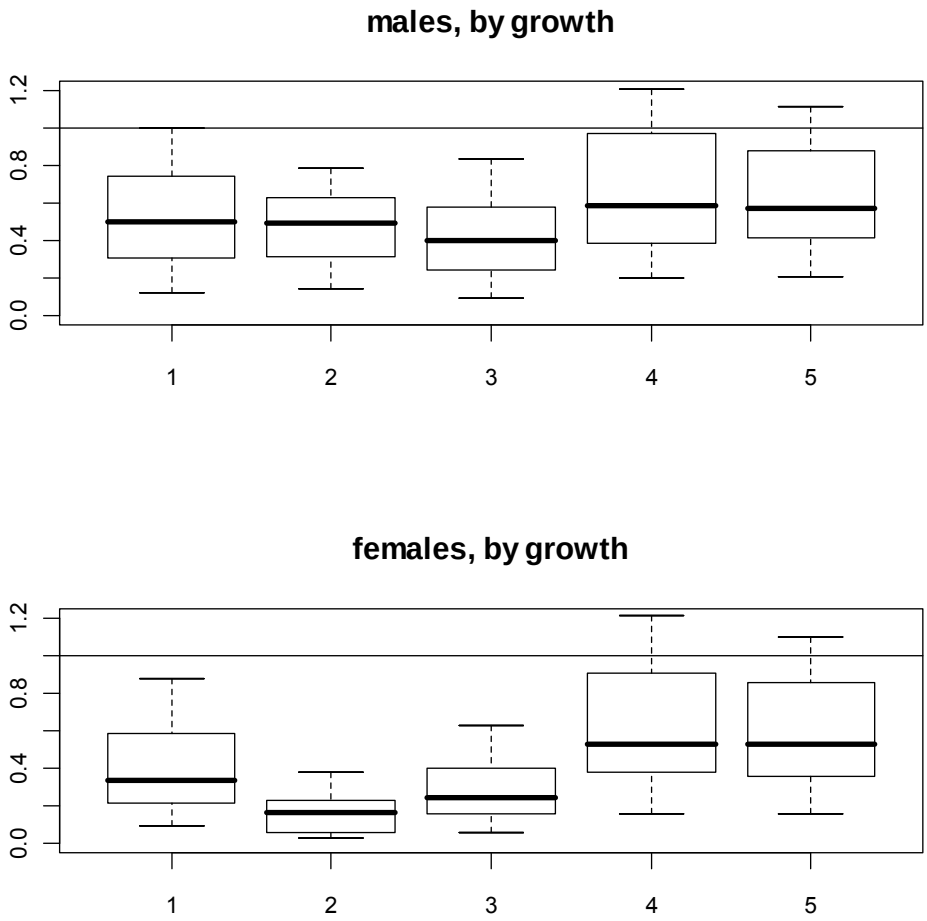


Figure 140 Change in the 'proportion big' indicator relative to change in biomass, [1 – indicator(Fmsy)/indicator(F=0)]/[1-Bmsy/B0], by gender and growth curve: 1=divergent, 2=parallel, 3=intersecting, 4=Young and Drake with age-at-maturity=5, and 5=Young and Drake with age-at-maturity=9.

12.3.3 Behaviour of time-series of Indicators

Here we explore the behaviour of the time-series of an indicator as a stock is exploited at a constant harvest rate, starting from an unexploited state. The equilibrium considerations in the previous sub-section were done without stochasticity in the dynamics, but here variability in recruitment is introduced to the scenarios with sexually dimorphic growth. We explored different harvest rates ($0.5F_{msy}$, F_{msy} and $2F_{msy}$) and different degrees of variability in recruitment (CVs of 5%, 20% and 40%) in the study, but only highlight a few results here. We note that the assumption of starting in an unexploited state may be rather unrealistic. Nonetheless, this provides insight into the extent of response that could be expected under such an 'extreme' scenario. More importantly, it provides a reasonable comparison between indicators. Further explorations of the performance of indicators in feedback decision rules are being conducted under alternative assumptions of starting state of the stock and of build-up of effort at the start of a fishery.

Some results are not surprising. For a given set of life-history parameters, the 'no-change error' (see Methods sub-section for definition):

- decreases as F increases,
- decreases when sampling level improves or increases,
- decreases as recruitment variability decreases

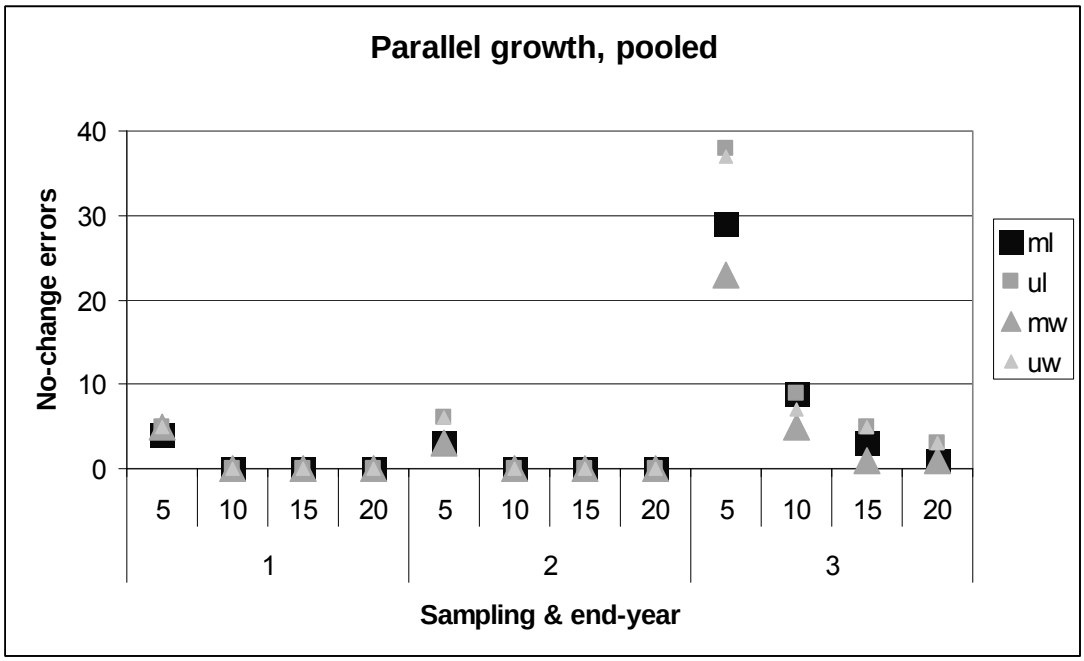


Figure 141. Percentage 'no-change errors' (from 100 runs) for ml=mean length, ul=95th percentile of length, mw=mean weight and uw=95th percentile of weight for 3 levels of sampling (1=good, 2=poor, 3= very poor) and 4 lengths of the time-series, all starting in year 1. The indicators are based on pooled data, the parallel growth curve was used with a recruitment CV of 40% and harvesting at Fmsy.

It is also not surprising that some indicators are more sensitive to the above factors than others. For example, the 95th percentile of length or weight is more sensitive to sampling level than mean length or weight, particularly over a short initial time-period. Figure 141 shows that when sampling is very poor (level 3), a significant change in the indicator is not detected in almost 40% of cases when the first 5 years' data are considered. This improves to about 10% when 10 years' data are used. One would expect indicators based on the 'tails' of distributions (e.g. the 95th percentile, or the proportion of individuals above some large size) to be more sensitive to sampling level than quantities such as the means.

Results with respect to gender-specific indicators are, however, less obvious. For the example based on pooled data and 'parallel' growth shown in Figure 141, the 95th percentile performs even more poorly when based on size data for males only. When based on data for females only, it appears to performs better than the mean length. It is not yet clear why this is the case – it could be spurious and due to the relatively small number of replicates (100 time series), it could be spuriously poor fits of the linear model in the case of the data for males only, or it could be a real effect. We note that this only occurs for the parallel growth curve, and we are exploring this further.

To consider the relative sensitivity of different indicators to sampling level, we have calculated the total 'no-change' errors over all the time-periods (5,10,15, and 20 year periods) and expressed them as a proportion. When indicators are based on pooled data, the proportion big, median

length, median weight and the mean length and weight of big fish are particularly sensitive to sampling level, and more so for the parallel growth curves than the divergent or intersecting (Figure 142).

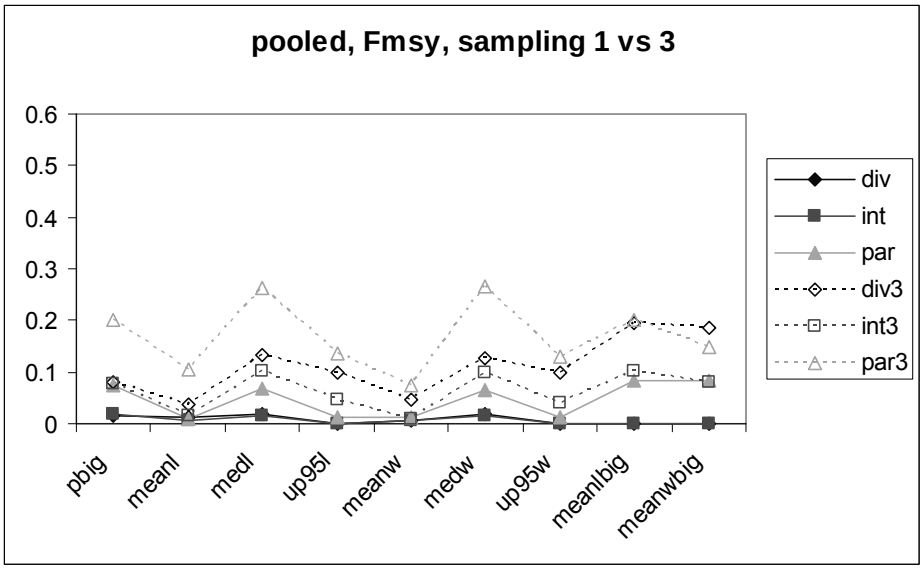


Figure 142. Overall proportion of 'no change' errors for pooled data, over periods of 5,10,15 and 20 years, for harvesting at F_{msy} and for the 3 growth curves with good sampling (1, indicated by 'div','int','par') and very poor sampling (3, indicated by 'div3','int3','par3'). See text for more detail on sampling level. Indicators are: proportion big (pbig), mean length (meanl), median length (medl), upper 95th percentile of length (up95l), mean weight (meanw), median weight (medw), upper 95th percentile of weight (up95w), mean length of big fish (meanlbig) and mean weight of big fish (meanwbig).

Figure 142 also shows that the median length and weight appears to perform more poorly overall than the mean and the 95th percentiles (NOTE here results have been combined over all time-periods, whereas Figure 141 shows performance separately for different time-periods, highlighting that the 95th percentiles perform better over longer time-periods).

The 'no-change errors' for gender-specific indicators show a somewhat different picture. Figure 143 shows that the 'males only' indicators (top panel) are much more sensitive to poor sampling than the 'females only' indicators (bottom panel).

This is likely to be a combination of the higher growth rate, lower L_∞ for males than females, and the expected sensitivity of indicators based on the 'tails' of the distribution to poor sampling.

It is not surprising that the means (length and weight) are the least sensitive of the suite of indicators shown.

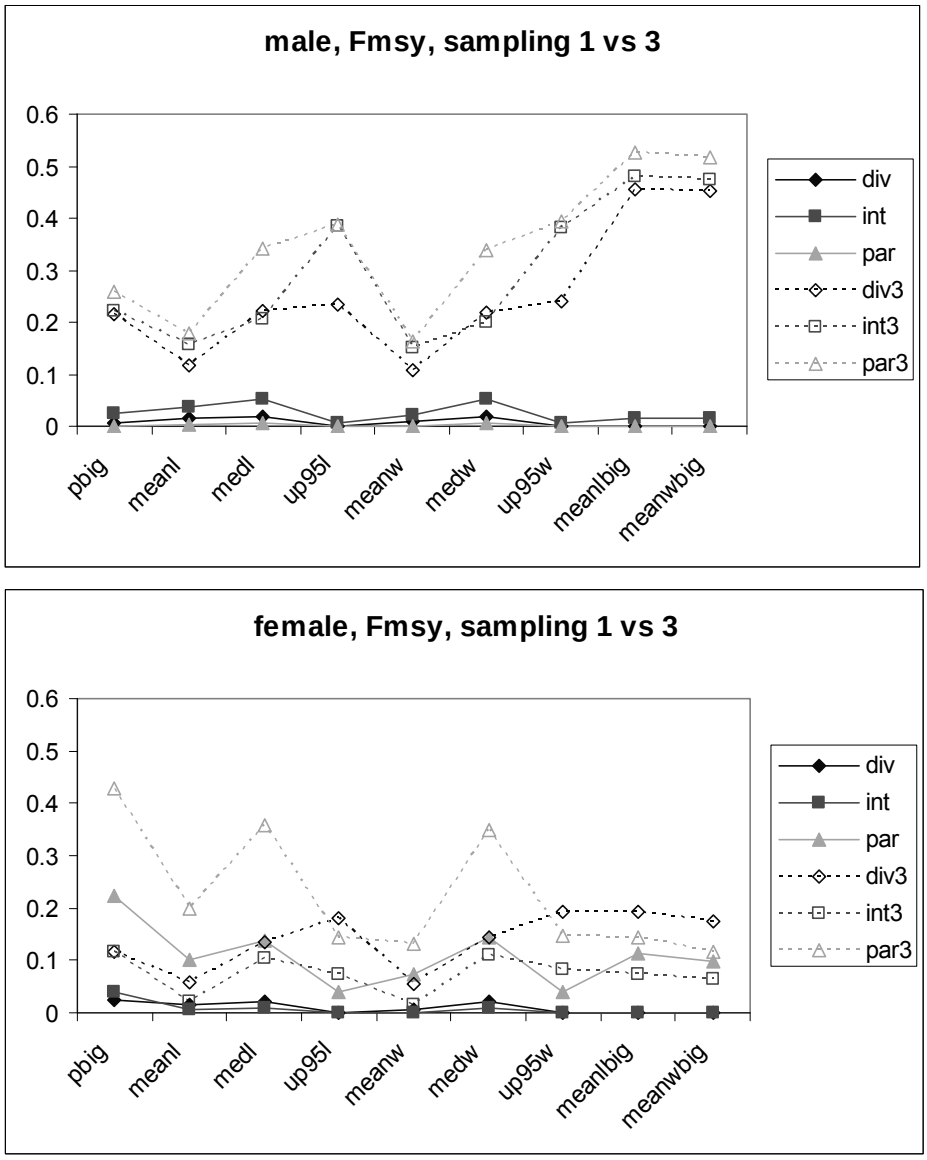


Figure 143. As Figure 142, but separately for males and females.

It is also interesting to note that the 'female' indicators under the parallel growth curve perform more poorly than under the other growth curves even when sampling is good. As mentioned above, the main reason for this is not obvious, and it is being investigated further. We also considered how well the time-series of indicators can predict changes in spawning biomass, using the mean squared errors from GAM models (see methods sub-section). For each indicator, 100 replicates of 30-year time-series were used, but the analysis could also be done on shorter subsets of the time-series. The mean squared error (MSE) over all replicates is used to evaluate performance; low values of the MSE suggest a potentially good indicator.

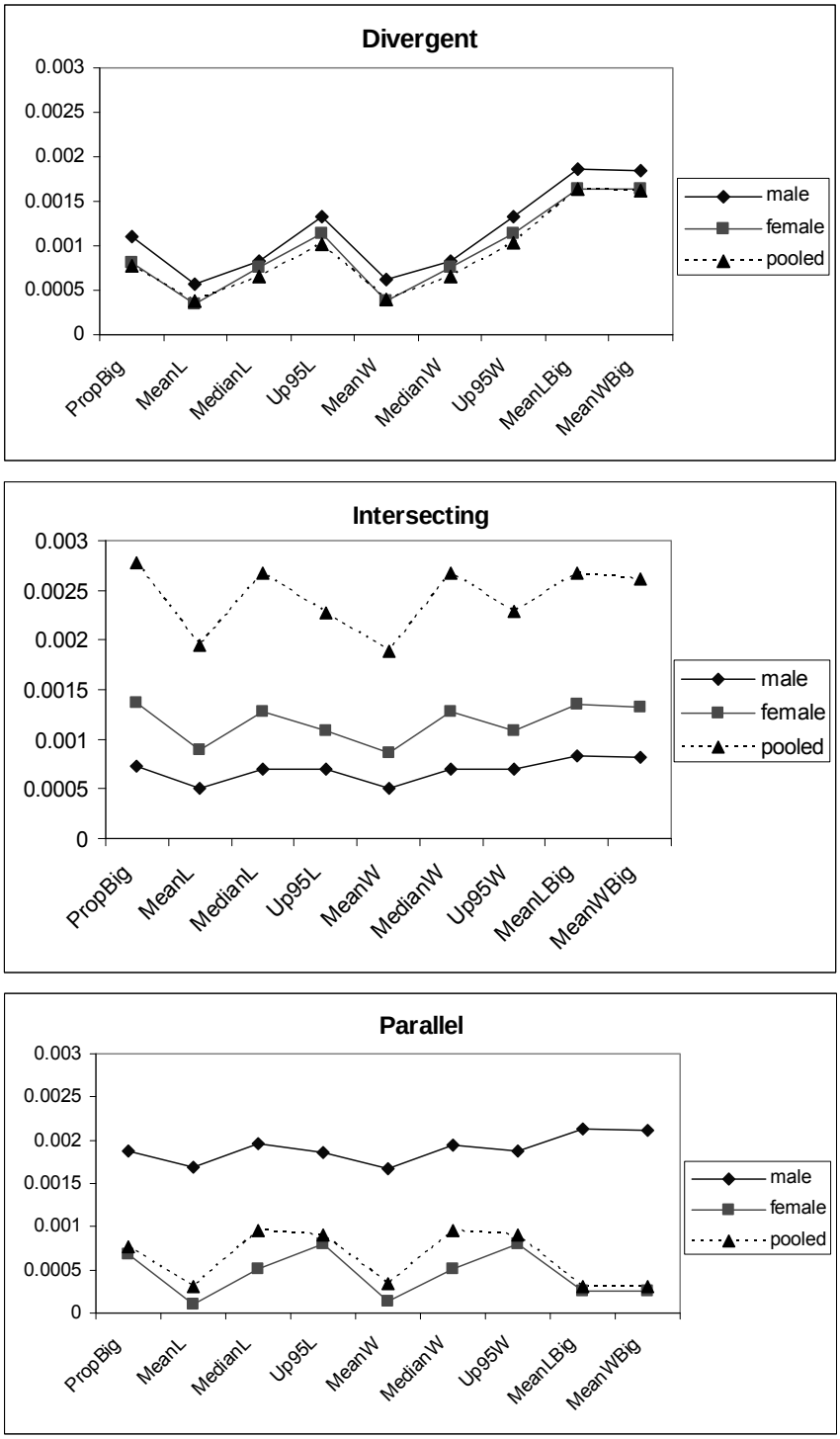


Figure 144. Mean squared error from GAM models for indicators by gender and pooled data. Other assumptions are: recruitment variability CV of 40%, F=Fmsy and poor sampling (level 2). Indicators are as in Figure 142.

Results (Figure 144) show the differences between the growth curves and choice of gender-based or pooled-data indices. There is little difference for the divergent growth pattern, but for the intersecting curves, the indicators based on pooled data perform more poorly than either of the gender-based indices. For the parallel growth pattern, however, the pooled indices perform much like the female-based indicators, and both perform better than indicators based on the data for males only. Whether the differences in MSE shown above are sufficient to have a strong effect on the performance of indicators in a feedback decision-rule context still has to be explored. The results confirm the good performance of mean length and mean weight relative

to the other indicators. In some cases the mean length and/or weight of large fish also perform well, but as noted before, the performance of this indicator is sensitive to the choice of cut-off relative to growth parameters.

12.3.4 Behaviour of single-indicator decision rules

Summary results of the decision-rule scenarios, in terms of the change in the indicator and performance of catch and spawning biomass, are shown below. All quantities are relative to their initial values, except for the number of years SSB is below 0.4.SSB0, and catch which is expressed relative to the catch at MSY, C_{msy}, (see [35]) for the growth curve used in the population dynamics. Although it is customary to consider the proportion of time that spawning biomass is below 20% of unexploited SSB as an indication of risk, these proportions were almost always zero for scenarios other than scenario 1. Instead, we have shown the number of years (out of 30) that spawning biomass is below 40% of unexploited spawning biomass. Given that B_{msy}/B_0 is around 0.35, this statistic should not be over-interpreted since one would expect the stock to be below 40%B0 for some years without necessarily being over-exploited.

Figure 145 shows the strong effect a decision rule based on an incorrect assumption about growth can have on the performance of the stock. For scenario 6 where growth is 'parallel' but the decision rule is based on F_{msy} associated with the Young and Drake growth curves, the relative spawning biomass is almost as low as for the unmanaged and over-exploited scenario 1 (where $F=1.6F_{msy}$). When the threshold or inflection point of the decision rule is increased, there is a slight improvement (scenario 7) in relative biomass, but it is not as good as for the case where the Parallel growth curve is used with the decision rule based on its associated F_{msy} (Scenarios 3 and 4).

An alternative way of looking at results is in terms of the trade-off between relative spawning biomass and catch relative to MSY (Figure 146). This figure more clearly shows that, in these examples, relative catch changes more slowly than relative SSB (e.g. there is less difference between relative catch for the 7 scenarios than for the relative SSB). It is therefore not necessarily the case that a large amount of catch needs to be forgone for an increase in spawning biomass which could imply a much lower risk of low SSB and possible recruitment failure.

[35] Since the actual level of recruitment is entirely arbitrary, the absolute levels of biomass and catch are not meaningful, and it is therefore necessary to rescale quantities here. Under circumstances where stock-recruit information was available from an assessment, for example, one would consider unscaled catch.

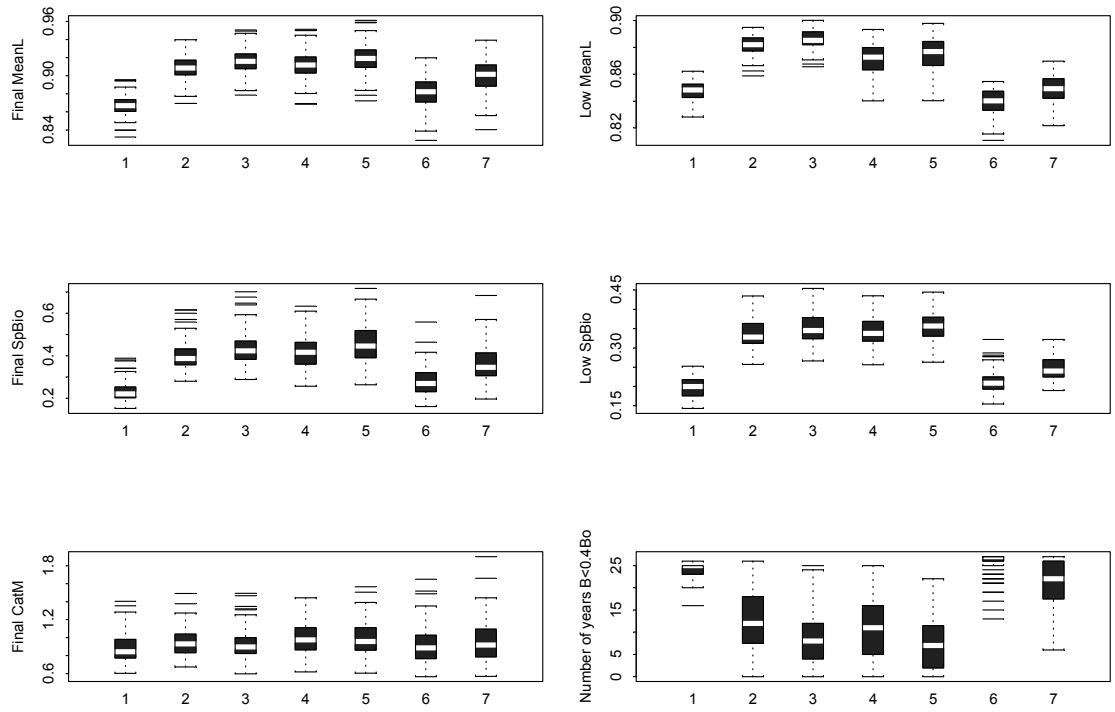


Figure 145. Summary results for 100 replicates of the 7 scenarios (numbered on the x-axis; see text in sub-section 2.4 above for definitions). All quantities are relative to their initial values (unexploited stock), except for final catch which is relative to the MSY of the underlying growth curve, and last panel which is the number of years out of 30 that SSB is below 40% of unexploited SSB.

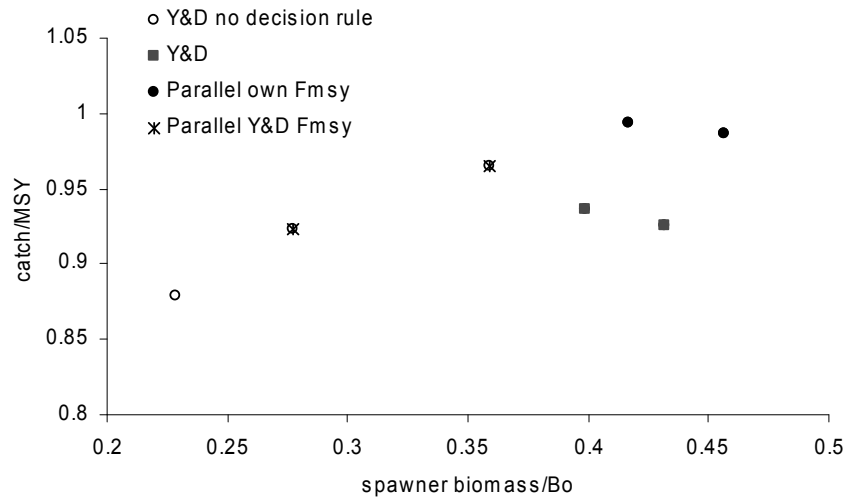


Figure 146. Trade-off plot in terms of spawning biomass (B/B0) and catch (Catch/MSY), averaged over the 100 replicates, for the 7 scenarios (see text). The right-hand points in each pair are for the decision rule with higher threshold (0.93), those on the left, for the decision rule with lower threshold (0.88).

12.4. Discussion and conclusions

This brief exploration of sexually dimorphic growth did not suggest a high priority for looking at gender-specific size frequency data, particularly since, generally, considerations of a broad range of life-history parameters (including growth rates) (Section 8) can be used in this context.

Simulations using sexually dimorphic growth curves taken from the literature suggest that, in terms of responsiveness of indicators, it is potentially more important to understand the form of the sexually dimorphic growth curves than to consider gender-specific indicators. In general, different types of growth curve resulted in greater differences in responsiveness of indicators than that between gender-specific indicators within a pair of sexually dimorphic growth curves. However, under some growth patterns (e.g. divergent), where the male and female curves are markedly different, it could be beneficial to consider indicators based on data for females only. In other cases (e.g. Young and Drake which has a low degree of dimorphism) there is apparently not much to be gained, in terms of responsiveness of the indicator, by considering gender-specific indicators.

However, in practice there are other important reasons why size data should be collected by gender if at all possible. This is because of the need for an indicator time-series to be standardised, and the fact that, in the presence of dimorphism in growth, the size-based indicators may be sensitive to changes in the sex ratio of the catch. If only pooled size frequency data are available, it may be impossible to standardise the size-based indicators. If information on the sex ratio in the catch is available, this may help to some extent. In some cases, it may be possible to use the fact that females grow larger than males to construct an indicator which is likely to mostly reflect the female size distribution (for example the mean length of fish larger than L_∞ for males). It is obvious that such an indicator will only work if there is a big enough difference between the maximum sizes for the two genders, and if there are sufficient individuals in the catch.

The results from the exaggerated growth curves are consistent with those from the literature-based growth curves. It is not surprising that sexual dimorphism in growth implies a stronger response in an indicator based on the data from the gender with lower growth rate. In our swordfish examples, this is always females (unless growth rates are very similar). Pooled data can lead to an indicator behaving very much like the female-based indicator or something in between the male- and female-based indicators. It is the detail of the form of the male and female growth curves (as well as other factors, including size at first capture relative to L_∞, and t_0) that determines whether the gender-specific or pooled indicators are preferable.

When "noise" is added to the indicators via recruitment variability and sampling level, it is more difficult to detect a change in an indicator as the noise-to-signal ratio increases. What is interesting, however, is that the gender-specific indicators respond differently to an increase in "noise", particularly to sampling level. The type of sexually dimorphic growth (divergent, intersecting or parallel) also affects the response.

It must be emphasised that these analysis do not take into account issues such as spatial structure by size or gender. As hinted above, there are potential problems with the standardisation and use of size-based indicators if there is gender-based spatial and/or temporal disaggregation. For example, a time series of mean length would be invalid, and difficult to standardise if this was comprised of mean lengths obtained from different areas with very different gender ratios.

The brief sub-section on decision rules illustrates the way in which a size-based indicator could be used in a decision rule. The differences in performance of the stock under the 7 scenarios underline the importance of designing decision rules that are robust to uncertainty in the underlying dynamics. Although we have only illustrated this for uncertainty in growth, taking uncertainty in steepness and mortality into account will also be important.

13. Swordfish: The performance of decision rules in the presence of fish and fleet movement

Summary

The two main issues considered in this Section are local depletion and stock structure. These issues relate to objectives 5 and 6 (see Section 4) of the project, although they are not explicitly stated as such.

The simulation model and decision rules used in this study are not tuned to actual data, and this means that the results should only be considered in relative terms. In particular, results from the decision tree rule should be interpreted with great caution, and should only be seen as illustrative.

Given the tendency of some decision rules to exhibit oscillations in simulations, and given that this is supported by theory (Section 11), we consider it prudent to always run some realisations for long time-periods, of say 100 years or so, to check for long term oscillations or strange behaviour. This is relevant for spatial and non-spatial simulation models. Testing should also explore the effects of the frequency of implementation of the decision rule, as well as the number of years of catch rates used in calculating key quantities used in the decision rule.

The spatial model allows the fleet to move between areas, and in some cases the fleet does not fish in all areas. This creates a problem of 'missing' catch rate data from areas not fished, which then affects the overall simulated catch rates. In reality, this can also affect the standardisation of CPUE which may be used in a decision rule. The potential for 'missing data' needs to be considered if spatial management is envisaged, because it can affect the performance of a decision rule.

The inclusion of a spatial indicator (which in these simulations reflect how far offshore the fleet is fishing) in a decision rule, can play a useful role. The simulations show that, if the catch is reduced when the spatial indicator is triggered, the overall biomass is higher (the catch is obviously lower), and the CPUE is generally higher. It is interesting to consider the notion that a spatial indicator could be used to 'trigger' fleet movement rather than a cut in catches, as modelled here. Fleet movement can be thought of as a re-distribution of effort and there would not necessarily be a need to close any areas, which would mean that catch rate data can still be collected from that area.

There is a strong link between the definition of 'local depletion', the details of fleet dynamics, and details of fish movement. If the details of these mechanisms and associated parameters are mismatched or incompatible in the specification of a spatial model, the results from simulations are likely to be of limited use. This suggests that the most productive way to explore the behaviour of management strategies in a spatial context is to use a reasonably case-specific model, ideally tuned to actual data. This does not exclude the possibility of exploring uncertainty in any of the components, or the model, but it will help focus on the most realistic or relevant subset of scenarios.

It is not surprising that the performance of decision rules are particularly sensitive to the assumption about stock structure, if that implies additional uncontrolled harvesting. The issue of stock structure and additional uncontrolled harvesting elsewhere, is likely to be particularly important when a decision rule is being tuned to achieve a specific objective, for example, with respect to average biomass levels. In this regard we make the, possibly obvious, comment that

the issue of additional uncontrolled harvesting can easily be incorporated into a non-spatial model. The tuning of a decision rule taking this into account does therefore not have to wait until a full spatial model has been set up and parameterised (conditioned) for a specific case.

The limited explorations considered here, do not point to a strong difference between (or preference for) the 40/20 decision rule using CPUE and that using mean weight. We would, however, caution against extrapolating from these results to other life-histories or levels of variability.

We only tested one decision rule based on multiple indicators, namely the decision tree rule. As already noted, this rule appears to be particularly sensitive to the implications of a spatial model on the variability of CPUE. However, the spatial model and (particularly) the fleet dynamics, used here are relatively simplistic. The model is also not tuned to any data (so not necessarily realistic enough), which is why we consider that results should be interpreted with care. This may be one of those cases (i.e. spatial model with fleet dynamics and feed-back decision rule) where a generic investigation is less useful than a case-specific investigation. We therefore suggest that further investigations, with a case-specific model incorporating more realistic fleet and movement dynamics, would be required to better assess the behaviour of the decision tree rule.

13.1. Introduction

In the case of pelagic fish stocks, there are often large uncertainties about spatial movement, migration patterns and stock structure. How these uncertainties might affect the performance of decision rules based on indicators, is therefore a very relevant question. In the case of swordfish, there are two particular issues which we explore here: local depletion and additional uncontrolled ('international') harvesting.

There is evidence that at least a proportion of the swordfish stock, off the East Coast of Australia, has high residency, generally about seamounts (Campbell and Hobday 2003), although the species is capable of ocean basin-scale migration. Fishing fleets tend to track fish "hot spots", and prefer areas of high local abundance to obtain high (or at least economic) catch rates. The Eastern Tuna and Billfish Fishery has experienced localized depletion of swordfish, leading to progressive offshore fishing as seamounts have become fished down. Localised depletion can occur even when the overall stock is sustainably harvested, but it may still pose a risk to the stock (for example, in terms of maintaining migration paths to feeding or breeding grounds), a risk to the local ecosystem, and it clearly is a risk to the viability of the fishery.

This Section uses a spatially explicit simulation framework to evaluate the performance of decision rules based on indicators in the context of fish residency and fleet movement. The spatial framework also allows for the effectiveness of a spatial indicator (the average distance offshore of the fishing fleet) to be evaluated when it is used to inform management action. Here the management action, or decision rules, is formulated in terms of total allowable annual catches (TACs). In some cases we consider a constant harvest rate scenario for comparison. Although there is no explicit attempt at simulating spatial management, the fleet dynamics are implemented in such a way that it can be thought of as some form of spatial management.

The first set of evaluations is based on simulation scenarios which assume that the decision rule has control over all the effort in the fishery. Put another way, it can represent the case where other fleets fishing on the stock has a negligible, very low harvest rate. Or, it can represent the case where the Australian fleet which is assumed to be managed by the decision rule (on the East or the West coast) is harvesting a separate, 'local' stock which is not harvested by anyone else. At

this stage, the first interpretation is unrealistic, both in the Indian and the Western Central Pacific (WCP) Oceans. The second interpretation is clearly unrealistic, particularly for the WCP[36]. The third interpretation (separate 'local' stock) is considered unlikely, given what is known about swordfish spawning and movement patterns, and genetics (Palko *et al.* 1981, Grall *et al.* 1983, Young and Drake 2002, Sun *et al.* 2000 Reeb *et al.* 2000, Ward *et al.* 2001).

A second set of evaluations are therefore constructed to represent the situation where there is a non-domestic (i.e. international) fishery which is not controlled by the decision rule. It is obvious that additional uncontrolled fishing would compromise the performance of a decision rule. The question is whether, and under which circumstances, the indicators obtained from the domestic fleet can detect signals of low stock size and hence lead to a response in the decision rule. The issue of local depletion also remains an important focus of this set of scenarios.

A subset of the scenarios is then used to evaluate the sensitivity of the decision rules to some unknown population parameters and incorrect assumptions when constructing decision rules. As we have emphasised throughout this study, it is unlikely that steepness of the stock-recruit curve would be known. It is also unlikely that reliable estimates of the maximum sustainable catch or harvest rate would be known.

This Section is by no means an exhaustive exploration of possible or even plausible scenarios. Many permutations of assumptions about stock and fishery dynamics and about indicators and decision rules are possible. Here, we have focused on swordfish-like population dynamics, and attempted to cover a range of extremes with the combinations of parameters chosen. In some of the sensitivity trials, the implications of one assumption instead of another will be predictable in terms of the direction of change in, say average biomass. The magnitude of change under different decision rules and indicators, is however, of interest. These simulations provide initial insights into the performance of different decision rules, with or without a spatial indicator, in the context of fish and fleet movement, and external fishing mortality.

After the describing the methods (sub-section 2), results are presented. The first part of sub-section 3 shows results for scenarios without external fishing, and the second part of sub-section 3 shows results for scenarios where there is additional external fishing. Results for sensitivity trials and additional scenarios are presented in sub-section 4.

13.2. Methods

13.2.1 Model specifications

The basic single area simulation model used in earlier Sections (e.g. Sections 9 and 10) was expanded to incorporate area-specific dynamics (see Appendix 1 for full details). The conceptual model was based on the evidence of local depletion in the ETBF and since this manifested itself (in broad terms) as a progressive offshore movement of the fleet, the spatial simulation model was constructed with areas analogous to longitudinal "bands" progressing from inshore to offshore. An arbitrary number of four areas was used, but this choice should not be crucial to the outcomes. A choice of two areas was considered insufficient to fully explore the spatial behaviour of the system, and any more than four was considered excessive. For simulations assuming external fishing by a foreign fleet, there was an additional, i.e. fifth, "outside" area.

[36] Currently, the Australian catch in the Indian Ocean is in fact negligible compared to the catch by other fleets.

Although the work is being conducted within the context of the Indian Ocean and/or the Western Central Pacific, a relatively general spatial model which can represent alternative hypotheses is used, as opposed to a specific 'Indian Ocean' operating model for billfish (or the tropical tunas). The main reason for this is the lack of direct data on movement and spatial structure, particularly in the Indian Ocean. This work should, however, lay the groundwork for further evaluation of indicators once the development of more realistic and data-based operating models for swordfish (or tropical tunas) in the Indian Ocean become possible.

The population dynamics are essentially as before, with natural mortality, recruitment and growth treated as overall (rather than area-specific) processes, but now catches are taken within a given area at the start of each quarter, after which (but prior to aging) the population is re-distributed by area according to the assumed fish movement dynamics. This means that we keep track of the numbers at age in the population in each area. Although overall recruitment is still based on the overall spawning biomass, recruits are also assigned to areas when they enter the population. The fishing fleet (and hence effort) is also redistributed according to a set of fleet movement dynamics. This is also done by quarterly time-steps.

The area-specific population and fishery equations are given in Appendix 1, but a prose description of the fish and fleet movement dynamics is provided here. At the start of the simulation, the unexploited population is distributed between areas using an assumed inshore-offshore gradient pattern, with the highest density in the inshore area. This can be thought of as a kind of habitat preference. Again, the choice of proportions in each area is arbitrary; for the purposes of this investigation it is sufficient that there is some gradient.

We have mentioned the notion that some pelagic species appears to have relatively low viscosity in terms of movement, and this appears to be the case for swordfish (Campbell and Hobday, 2003). In this terminology, high viscosity would imply that, if the original spatial distribution is drastically changed, for example, by a large catch being taken out of one area, then it could take some years before the population returns to the original spatial distribution. Here it is helpful to think of 'spatial distribution' as the proportion of the population in each area. A low viscosity would imply a rapid return to the original distribution. In order to model this notion of a slow return to the original distribution, we introduce the concept of residency. Assume that some proportion of the stock is resident in a given area, but the remainder moves freely among the areas. When the fish are re-distributed in the simulation, it is only the non-resident proportion in any area that can move to another area.

The simulated fish movement can therefore be summarised as follows:

at the start of the simulation, the population is distributed among areas according to the pre-defined ('habitat preference') proportions. Note that this applies to all age classes, so there is no age- or size-specific distribution

at the beginning of each quarterly time-step (after removals due to mortality, but prior to aging), the non-resident proportion in each area is redistributed according to the pre-defined ('habitat preference') proportions. If it is quarter 1 or 2, the new recruits are distributed at the end of the quarter, according to the pre-defined proportions.

The actual level of residency of swordfish is unknown and we consider two scenarios: very high residency, where the resident proportion is 90%, and moderate residency, where the resident proportion is 50%. We do not consider a much lower level of residency, because the spatial dynamics become almost irrelevant in that case.

For scenarios with only four areas, proportions used at the start and for the redistribution of fish are: 0.6, 0.2, 0.15, 0.05. These (arbitrary) values were chosen to give a reasonably steep gradient

of abundance over the four areas. For scenarios with an additional external area, we assume that 60% of the population is in the four 'domestic' areas and 40% is in the fifth, 'international' area. Using the same gradient in the distribution as above, and renormalising proportions, this leads to the following proportions for the five areas: 0.43, 0.14, 0.11, 0.04; 0.29 (inshore-offshore; outside (or external)).

Fleet movement

The fleet dynamics used in this study are admittedly very simple, and not based on actual data or economic modelling. However, the concepts are based on experience in, and qualitative evidence from, the ETBF. The specific scenarios should therefore be seen as illustrative. The main driver for the fleet dynamics is the notion that fishing closer inshore is preferable to fishing further offshore, and that good (high) catch rates are desirable. The simulation model assumes that if catch rates, relative to the highest catch rate experienced, drop below some level, then a proportion of the fleet will move to the next area further offshore. If catch rates in the inshore area recover, the fleet will move back to that inshore area.

The fleet dynamics were modelled as follows:

at the start of the simulation, all "domestic" fishing mortality is located in the most inshore area.

if the catch rate drops to some pre-defined proportion of the initial catch rate, a certain proportion of the fleet moves to the adjacent area further offshore.

if the catch rate drops to an even lower pre-defined proportion of the initial catch rate, fishing ceases in that area. This offshore movement is repeated for the other areas

if the catch rate in inshore areas recover, the fleet moves back to the most inshore area(s) in preference.

We consider two scenarios for the fleet dynamics, defined by three parameters, namely: the critical catch rate in numbers (CPUEnum) at which offshore movement is triggered, the CPUEnum below which fishing in that area ceases, and the proportion of the fleet that moves offshore. The definitions are referred to as "move on" criteria.

"Move on" criteria:

i) Gradual: 25% of the effort moves to the adjacent offshore area if CPUEnum in the area drops to below 70% of its initial value; fishing ceases in the area if the CPUEnum drops below 50% of its initial value;

ii) Stronger: 90% of effort moved to adjacent offshore area if CPUEnum in the area drops to below 50% of its initial value; fishing ceases in the area if the CPUEnum drops below 30% of its initial value.

The first set of criteria invokes a gradual move offshore which commences earlier than the second set of criteria, which mimics stronger local depletion. This is because, in the second set, catch rates are allowed to drop further before offshore movement occurs, but when this takes place, the majority of the fleet relocate.

The movement back into previously abandoned areas work in a similar way, but in reverse. This implicitly means that we assume the fleet has knowledge of catch rates in areas that they are not currently fishing. The 'domestic' fleet always only fishes in, and moves between, the domestic areas, 1 to 4, even when there is a fifth 'international' area. Similarly, the 'international' fleet, only fishes in the fifth area.

The definition of the fleet movement is in terms of relative CPUE, and model diagnostics and performance measures, are output in terms of the relative area-specific CPUEs. Unfortunately,

we belatedly realised that, in terms of absolute CPUE, the particular choices of "Move on" criteria combined with the particular proportions of abundance in each area, imply somewhat peculiar fleet behaviour if seen in terms of a single species fishery. In some cases, part of the fleet may move offshore to catch rates which are in fact lower than those in the area from which they have moved[37]. This should however, primarily occur in the offshore areas, and given the dominance of the inshore area in terms of contributions to biomass and catch, we consider that this issue should not seriously affect the general results. Also, note that only catch rates in area 1 are considered when exploring the issue of local depletion (i.e. the part of the fleet that remains in area 1) for decision rules with or without a spatial indicator.

It is also worth noting that these move-on criteria together with a high harvest rate at the start, leads to very rapid offshore movement – even by the third or fourth quarter of the first year.

The above set of fish and fleet movement dynamics do not cover the theoretical extremes. For example, it does not include a scenario with no fish movement between areas (i.e. 100% residency). Neither does it include 0% residency. Similarly, for the fishing fleet, it does not include random distribution of fishing effort among areas, or fishing effort distributed in direct proportion to abundance. The reason for this is two-fold. Some scenarios are unrealistic, for example, random distribution of effort. Also, in some cases scenarios are simply not interesting, because the spatial detail do not affect the main quantities of interest.

13.2.2 A simple example illustrating implications of fish movement

It is useful to illustrate the implications of fish movement with a simple example, which will also help understand how the distribution of the fishing effort across areas interact with the spatial distribution of the fish. This is included here rather than in the results sub-section, since it is based on a much simpler model than that used in the simulations and it is only for illustrative purposes.

Consider a simple population model with constant recruitment and constant natural mortality (m) over time. In the unexploited population, recruitment and natural mortality are balanced so that the population is in equilibrium. Assume there are three adjacent areas, and the assume the 'habitat preference' proportions are (0.6,0.3,0.1). Also assume that there is some level of residency, as described above. For example, if residency is 0.9, then (1-0.9)=0.1 of the population is redistributed at the end of each time-step, into the three areas using the proportions (0.6,0.3,0.1). Recruits are also distributed according to (0.6,0.3,0.1).

We examine four simple cases, to see how (and whether) these proportions are different when the population is harvested in different ways. First, however, we consider the case where there is no harvest.

Case 0: No harvesting.
In this case, the population is distributed according to the proportions (0.6, 0.3, 0.1) in the 3 areas at each time step, irrespective of the level of residency. This is because recruitment is distributed according to the same set of proportions and natural mortality is the same in each area.

[37] This may not be totally unrealistic if they move elsewhere to target a different species, with the original species becoming more of a bycatch in the new area

Case 1. Harvesting at the same rate (i.e. same F) in each area.
This scenario also leads to the same proportions, (0.6, 0.3, 0.1) in each area at each time step, irrespective of the level of residency. This is because total mortality (m+F) is the same in each area, and recruitment is also distributed according to the same set of proportions.

Case 2. Harvest rate, F, is proportional to the initial distribution in each area
Now we assume that the harvest rate, or effort, is split between areas according to the proportions 0.6, 0.3, 0.1. If total F=0.2, it means that F by area is the vector (0.2*0.6, 0.2*0.3, 0.2*0.1). Now the proportions change over time until a new equilibrium is reached. If residency is low, there is little difference between the original and the new equilibrium proportions, but when residency is high, there is a more marked difference (Figure 147). There is also a bigger change in the proportions when F is high (=0.6) than when it is low (F=0.2).

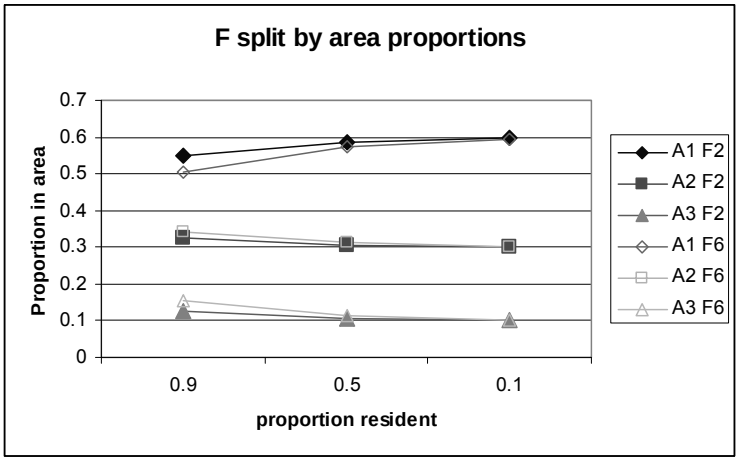


Figure 147. The proportion of the stock by area (A1, A2, A3) for 2 levels of harvest rate: F=0.2 i.e. relatively low (closed symbols), and F=0.6 which is very high (open symbols). The horizontal axis indicates the proportion of the population assumed to be resident. (Legend: A1 F6 implies Area 1, F=0.6 etc.)

Case 3. Harvesting only in area 1.
If all the effort is in area 1, then there is again little change in the proportions when residency is low. When residency is high, however, there is a big change in the proportions by area; the proportion in area 1 drops to just above 0.3 when F=0.6. The change is bigger than for Case 2. This is because the catches that are taken out of area 1 are bigger than in Case 2, and there is not sufficient fish movement to replace the catch. As expected, the effect is strongest for the high harvest rate.

There is the obvious caveat to the comparison between Case 2 and 3, that by putting all the effort (or F) into area 1, the total mortality on the stock is different from the case where that same effort (or F) is distributed over the 3 areas. In order to make the two cases more directly comparable, the harvest rate to apply to area 1 (in Case 3) should really be F=0.157 instead of F=0.2. This leads to the following split for 0.9 residency: 0.478, 0.392, 0.131 (compared to what is shown in Figure 148, which is 0.455, 0.409, 0.136; clearly these values are not very different).

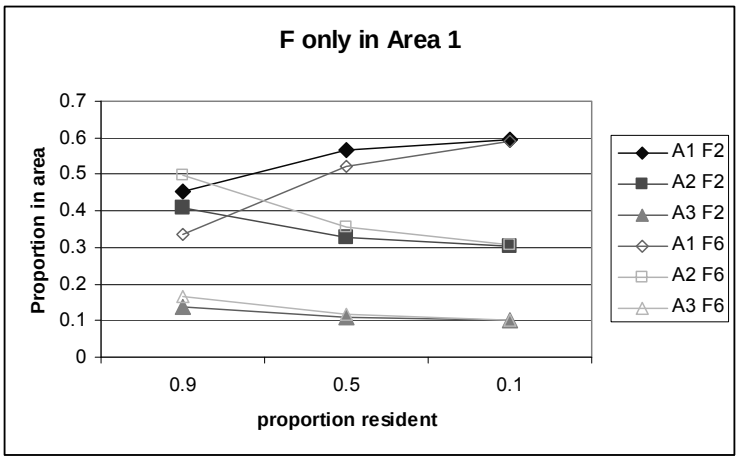


Figure 148. The proportion of the stock by area (A1, A2, A3) for 2 levels of harvest rate: F=0.2 i.e. relatively low (closed symbols), and F=0.6 which is very high (open symbols). The horizontal axis indicates the proportion of the population assumed to be resident. (Legend: A1 F6 implies Area 1, F=0.6 etc.).

13.2.3 Running the simulations

Simulations were run for 30 years with an annual decision rule being invoked after the first five years. The stock was taken to be at virgin levels in the initial year. To test the ability of the decision rule to recover the stock from heavy overfishing, a fishing mortality of $2.F_{msy}$ was applied for the first five years. Given that the whole fleet starts fishing in the inshore area, this high harvest rate is applied entirely within the most inshore area (area 1), with the move-on criteria (see below) applied to all subsequent time steps, and the same harvest rate then distributed over those areas that are being fished.

The decision rules control the fishery through catch quotas, which are translated to fishing mortalities by numerically finding the harvest rate which coincides with that catch (see Appendix 1; note that the quarterly time-step makes this slightly more complicated).

Reference-scenario population and fishery dynamics were as described earlier:

- growth according to a von-Bertalanffy length-at-age vector (no sexual dimorphism)
- Natural mortality, $m = 0.2$
- Almost "knife-edged" selectivity-at-age – initially selected at age 3 years; fully selected from age 6 years onwards
- Steepness $h = 0.65$
- Age-at-maturity 9 years
- standard deviation of the logarithm of the fluctuations in recruitment $\sigma_R = 0.2$

Additionally, small lognormal random errors were applied to catch and CPUE as follows:

- standard deviation of the logarithm of the fluctuations in catch $\sigma_C = 0.05$
- standard deviation of the logarithm of the fluctuations in catch rate $\sigma_{CPUE} = 0.1$

For each scenario tested, 10 realisations were run to encapsulate some of the stochasticity associated with the random errors on recruitment, catch and CPUE. Ideally, more realizations would have been preferable, with each over a longer time series. However, this was compromised by the run-time of the simulation given the large range of scenarios examined. Given the small number of realisations, we only show the medians and inter-quartile range on boxplots of results for performance measures.

When 'external' fishing mortality was applied, in the outside area of the spatial model, this was set at a level of 3.Fmsy for the entire time period. The rationale for this high level is to evaluate the performance of domestic indicators and decision rules when confronted with heavy external fishing pressure. Selectivity was identical to that of the "domestic" fleet. In these scenarios, the indicators used in the domestic decision rule are those based only on the domestic catch.

Results are presented for the 70% sampling regime only (i.e. assume 70% of the catch is sampled for size and used in calculating indicators such as mean length in the catch). This is a realistic level of sampling given the current port/processor-based monitoring within the ETBF.

Scenarios tested

The suite of scenarios was designed to investigate three main issues:

- the spatially explicit issues of: i) fish residency and movement, ii) fleet movement as characterized by "move-on" criteria, and iii) the effectiveness of a spatial indicator when combined with a decision rule to avoid localised depletion.
- the effect of external fishing mortality on the effectiveness of indicators and decision rules
- sensitivity of indicators and decision rules to i) effort creep, ii) too high a maximum catch level in a decision rule, iii) steepness assumed to be too high.

Alternative assumptions combined with alternative parameter values for a large number of factors quickly results in an unmanageable and impractically large potential set of scenarios. Scenarios and values for parameters were therefore chosen to span plausible ranges, or to focus on those most likely to have large effects on results. For example, scenarios tested but considered to have minimal impact (e.g. a lower level of external fishing mortality) or indicators likely to show similar performance (e.g. gleaned from results on prediction errors in Section 9) were omitted from the suite of scenarios.

The following inputs were crossed to form the initial suite of 8 scenarios (summarized in

Table 24):

Residency:

i) 50% in each area
ii) 90% in each area

"Move on" criteria (as defined above):

i) Gradual move-on
ii) Stronger move-on

The Decision rules tested were as follows:

i) "40/20" type rule with mean weight as the indicator
ii) "40/20" type rule with CPUE in numbers (across all "domestic" fishing) as the indicator;
iii) the "decision tree" developed for the ETBF (Appendix 1; Davies *et al.,* 2008).

Spatial indicator:

i) Spatial indicator combined with the decision rule: a reduction of 30% of the catch quota resulting from the decision rule, if the average distance offshore in the previous year exceeded the maximum that occurred over years 2 to 4 of the fishery.
ii) no spatial indicator applied

The "40/20" type rule is described in more detail in Appendix 1. In these simulations, the target is the indicator value corresponding to the fishing mortality, F_{spr40}, that would drive the stock to 40% of the unexploited spawner-biomass-per-recruit level. The limit is the indicator corresponding to the fishing mortality that would drive the stock to 10% of the unexploited spawner-biomass-per-recruit, F_{spr10}. This is therefore really a '40/10' decision rule, but we have kept the nomenclature '40/20' in here. The reason for using the 10% level as a limit rather than the 20% level lies in results in Sections 9 and 11 where we noted that for size-based indicators, the 20% and 40% reference points can lie very close together. This can then lead to oscillations in the population dynamics. The 10% and 40% levels are further apart and oscillations are therefore less likely.

The 40/20 rule also requires a maximum catch level to apply when the indicator is at or above its target. We have set $C_{max} = C_{msy}$. For the indicator below its limit, the catch is set to 5% of C_{max} to avoid total closures and absence of data in the simulations. The choice of C_{max} will in practice be difficult, and the choice of C_{msy} in a real situation would be a bad idea, first, because MSY-related quantities are difficult to estimate (particularly since steepness is usually highly uncertain), and second, because MSY-related quantities should be seen as limits rather than targets.

Mean weight was chosen as an indicator since swordfish catch is reported in weight. Length-based indicators were not tested in this context. The version based on CPUE uses the overall CPUE (in numbers) from all the "domestic" areas fished in that time-step. This raises one issue which we return to in the results sub-section, namely standardisation of the CPUE series input to the decision rule. Here we simply note that if an area is not being fished, there are no data from that area. On the other hand, if the fleet is moving between areas, and not always fishing in all areas, then the combined CPUE is, in some sense, not really standardised (... but without any data from the unfished areas, how does one standardise?).

The "decision tree" rule was developed in the HSWG for the ETBF (Campbell *et al.* 2007; Davies *et al.* 2008). This rule is based on recent CPUE trends, where CPUE is size-specific for "prime-sized" individuals. The first decision point in the tree considers the level and trend in this CPUE, and either increases or decreases the current catch accordingly (see Appendix 1). The extent of the adjustment is also related to the specified timeframe in which the target should be reached. The resulting catch quota is then modified according to additional indicators - the CPUE of large fish, the proportion of large fish, and the CPUE of "recruiting" fish in the catch, relative to their values at 40% SPR_0 (SPR_0=unexploited spawner-biomass-per-recruit).

The HSWG definition of the decision tree rule is such that the target CPUE is specified as 'number of fish per 1000 hooks' (i.e. in absolute terms) although it is based on 50% of the CPUE achieved during the first year of the fishery. Although the same approach could have been used in the simulations, the fact that we start with an unexploited stock, means that the target CPUE ends up being very high. Instead we used the expected CPUE (of prime fish) at F_{msy}. (As it turns out, this is also in fact rather high, and leads to quite conservative behaviour).

It is important to note that results from the above two forms of decision rule are not directly comparable because we have not 'tuned' them (i.e. changed their defining parameters) to achieve the same outcome relative to one performance measure. The 40/20 rules based on mean weight or based on CPUE are reasonably comparable, because their default settings (i.e. definition of trigger points and C_{max}) are based on the same logic.

Table 24. The 8 base case scenarios, run for each of the 3 decision rules, both with no external fishing mortality, and with external fishing mortality equating to 3.F_{msy}.

Residency (%)	Spatial indicator	Move-on criteria	40/20 MeanW	40/20 CPUEnum	Decision tree
50	N	Gradual	Y	Y	Y
50	N	Stronger	Y	Y	Y
50	Y	Gradual	Y	Y	Y
50	Y	Stronger	Y	Y	Y
90	N	Gradual	Y	Y	Y
90	N	Stronger	Y	Y	Y
90	Y	Gradual	Y	Y	Y
90	Y	Stronger	Y	Y	Y

These 8 "base-case" scenarios (Table 24) were run using each of three decision rules, both with and without the "external" area/fishing mortality. Additionally, the 8 scenarios were run with no underlying decision rule, but a constant harvest rate (F_{msy}) instead.

To obtain an insight into the overall longer-term behaviour of the population under each decision rule, single realizations were run over 100 or 200 years for the scenario assuming 50% residency and the "gradual" move-on criteria (both with and without spatial indicators). Illustrating long-term behaviour was particularly important given the cyclical nature of the behaviour of the population under the decision tree framework (Davies *et al.*, 2008). Interpretations made from a 30-year projection can be misleading without an awareness of long-term dynamics.

Sensitivity trials

Exploration of the sensitivity of indicators and decision rules is limited to three aspects which are likely to have the strongest effect on results: i) effort creep, ii) too high a maximum catch level in a "40/20 style" decision rule, iii) steepness assumed to be too high.

In addition, and given the overall similarities in results over some of the assumptions, the sensitivity trials are only done for the scenario assuming 90% residency and 'gradual' move-on criteria. Versions of the decision rule with and without a spatial indicator are considered for all the sensitivity runs. The rationale behind these choices is again primarily related to limiting the vast number of scenario runs that could be constructed.

Many of the sensitivity tests will have obvious implications in terms of the direction of change in biomass and catch. That is, the manner in which the alternative dynamics will impact on the fishery is predictable. The question is the likely magnitude of the change, and whether the rules respond similarly or differently under the alternative set of assumptions.

We did in fact also run some trials to examine robustness to uncertainty in the decision rule and its associated inputs (e.g. assuming incorrect population biology when calculating spawner-per-recruit (SPR) reference points) and to increased variability and measurement errors in the system. Results are not presented here because the effects were minimal compared to those for the three issues identified above. The list of sensitivity scenarios is summarised in Table 25

Table 25. List of scenarios considered in sensitivity trials; a blank cell means no runs were done for that scenario. The default value for steepness is h=0.65 unless otherwise stated, e.g. in scenarios 6-9.

Scenario	No F outside			F outside		
ALL ARE 90% res, Gradual move-on	40/20 MeanW	40/20 CPUE	Decision tree	40/20 MeanW	40/20 CPUE	Decision tree
1.1 NO spatial indicator (SI)	Y	Y	Y			
1.2 WITH spatial indicator	Y	Y	Y	Y	Y	Y
2. 2% p.a. effort creep, WITH SI	Y	Y	Y	Y	Y	Y
3. 2% p.a. effort creep, NO SI	Y	Y	Y			
4. $C_{max} = 1.5C_{msy}$, WITH SI	Y	Y		Y	Y	
5. $C_{max} = 1.5C_{msy}$, NO SI	Y	Y				
6. 'True' h = 0.3 but MSY-quantities in rules based on h=0.65. WITH SI 40/20 rules: $C_{max} = 1.5C_{msy}(h=0.65)$ Decision Tree: $CPUE_{msy}(h=0.65)$	Y	Y	Y			
7. 'True' h = 0.3 but MSY-quantities in rules based on h=0.65. NO SI 40/20 rules: $C_{max} = 1.5C_{msy}(h=0.65)$ D.Tree: $CPUE_{msy}(h=0.65)$	Y	Y	Y			
8. 'True' h=0.3 and MSY-quantities in rules based on h=0.3; NO SI	Y	Y	Y			
9. 'True' h=0.3 and MSY-quantities in rules based on h=0.3; WITH SI	Y	Y	Y			

13.3. Results: No additional uncontrolled harvest

Recall that this scenario, where there is no additional uncontrolled harvest in an 'outside' or 'international waters' area, also has an alternative interpretation. The scenario can also imply that the fishery being managed by the decision rule is a separate stock from that being harvested in the 'outside' area.

When looking at results for the decision tree it is particularly important to note several caveats and to consider results as illustrative only. Results are only being used in a comparative sense, for example comparing results for the decision tree with and without a spatial indicator, and not in an absolute sense. The simulation model has not been conditioned to actual catch and effort data, so we cannot draw conclusions about the actual level of depletion or catch it would achieve. Also, the decision rule has been designed using information from actual catch-effort and size frequency data. This means that the decision tree has not been 'tuned' to this model, so it may, for example, perform overly conservatively, or overly aggressively. This issue is in fact illustrated below. In addition, the spatial detail of this particular model (fish and fleet movement) would have an effect on the behaviour of the decision rule, and these details are also not based on actual data, but on scenarios. The spatial nature of the model also affects how relative abundance is 'measured' through CPUE and this appears to affect the decision tree very strongly. This is discussed further below.

Before considering the issue of local depletion, we briefly look at the long term behaviour of the decision rules with and without a spatial indicator.

13.3.1 Long term behaviour of decision rules.

Although all the explorations considered here were done on a more sensible time-scale of 30 years, we ran a single long term realisation for each decision rule (and for a constant harvest rate), because we were aware of the potential for the decision tree to exhibit oscillatory behaviour. This occurred in some of the trials conducted in a simpler model without spatial structure, fish movement or fleet movement. In that case, however, oscillations did not persist, but were damped out over time (Campbell *et al.* 2007). Apart from the fact that wide-ranging oscillations in the population dynamics imply unacceptably high variability in catch, and a high risk of low biomass, it also makes comparisons between decision rules rather difficult. For example, if performance is compared using the average biomass over some 10 year period, say, then if the scenario for one rule happens to be in a trough of its oscillations and another happens to be at a peak, the comparison is entirely unfair and inappropriate. This implies that performance measures need to be chosen with some care. Given the tendency of some decision rules to exhibit oscillations in simulations, and given that this is supported by theory (Section 11), we consider it prudent to always run some realisations for long time-periods, of say 100 years or so, to check for long term oscillations or strange behaviour.

In one set of realisations (Figure 149), all rules show only moderate long term fluctuations, though the decision tree leads to substantially higher biomass than the other rules. This is explored further below. Catches for the 40/20 decision rules based on mean weight are almost constant somewhat below the maximum catch, C_{max}, in this realisation (following the initial reduction and slow increase while indicators recover to above their target levels). This outcome is not entirely surprising if we note that the three harvest rates involved in the design of the rule have the following values: F_{msy} = 0.11, F_{spr40} = 0.09 and F_{spr10}=0.3. It is because F_{msy} (on which C_{max} is based) is close to and slightly above F_{spr40} (on which the target point for the indicator is based) that the catch is unlikely to be as high as C_{max} much of the time.

The constant harvest rate policy, which is primarily shown for illustrative purposes, leads to more variable catches, because the catches vary with the population size. Catches are also likely to be more variable than the population size, because of the spatial nature of the model, and the movement of the fleet between areas of high and low density. The decision tree has the lowest catch and shows quite large fluctuations towards the end of the time period.

We were concerned about the very high biomass (almost at the unexploited level), with almost no catch for a long period. Several more long realisations were therefore run with the decision tree, also using a lower value for the target CPUE which forms the first decision point of the tree.

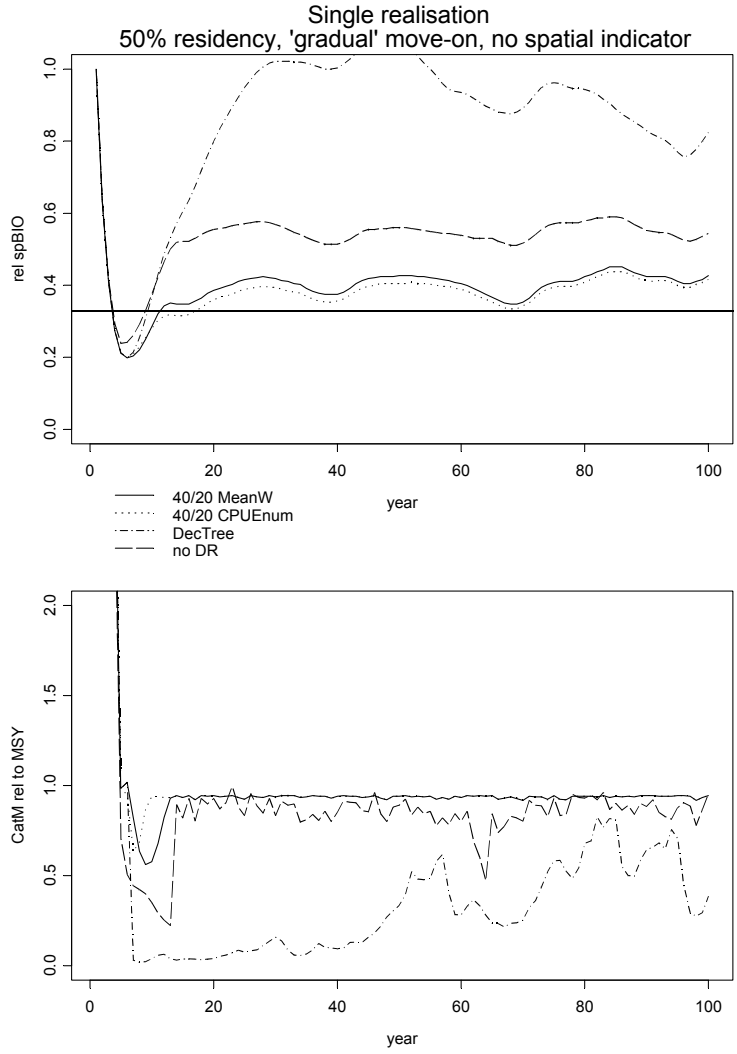


Figure 149. Single, long time-series, realisations for 3 decision rules and a constant harvest rate (without a spatial indicator). All scenarios start with a harvest rate of 2.F_{msy} over the first 5 years. The decision rule, or constant harvest rate (F_{msy}) starts in year 6. The top panel is for the total (all areas) relative spawning biomass; the lower panel is the total catch (in weight; all areas) relative to C_{msy}.

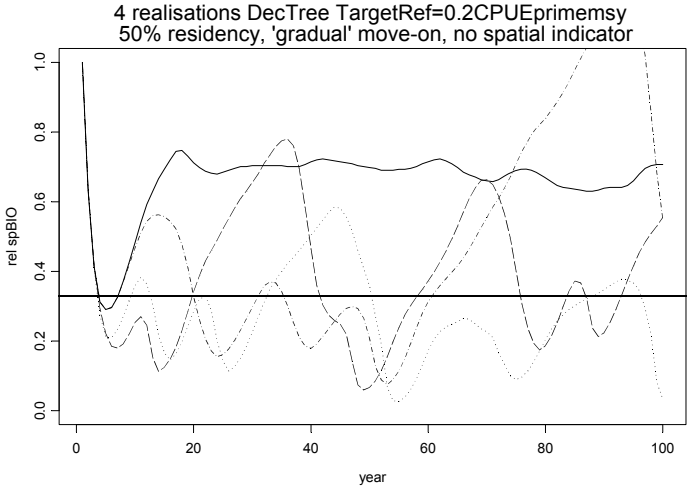


Figure 150. Four realisations of the decision tree rule with a lower target reference for CPUE. The scenario is for 50% residency, 'gradual' move-on criteria, and no spatial indicator was used. Spawning biomass relative to unexploited is plotted over time.

Two points are clear from Figure 150, which illustrates results for the lower target reference CPUE in the decision tree rule. There is a very big difference between realisations, and in some cases they exhibit large and relatively long term oscillations. The lower target level also leads to more reasonable performance; note that B_{msy} is around $0.3B_0$ for the life-history parameters in this scenario (recall that steepness is assumed to be 0.65). We emphasise that the performance of the decision tree with this particular spatial model should not be over-interpreted given that the model has not been conditioned to data, and the decision tree rule has not been tuned in any way.

Two of the most obvious ways of tuning the decision tree rule is to change the CPUE target at the first decision point, as shown in Figure 150, and/or to change the percentage additional reduction in catch that can occur at any of the subsequent decision points. Figure 151 (below) shows that a change in the target CPUE has the biggest effect; halving the subsequent reductions in catch from 10% to 5% has only a relatively small effect.

Figure 151 also shows the rather worryingly large variability and oscillations in the catch. Unfortunately, we do not have direct comparisons of trajectories for this model and a version with no spatial structure, so it is not possible to say to what extent the behaviour is being driven by the detail of the spatial aspects of the model, or to what extent it is the parameter values and nature of the decision rule itself. It is most likely that the behaviour is being caused by a mixture of factors, two of which are likely to be: (i) the way CPUE is 'measured' and (ii) the frequency of catch updating.

First consider CPUE. In the spatial model, the fleet can leave one area and move to another, and there are instances when the fleet is not fishing in all areas. This means there are no data from those areas. Given the very steep gradient in abundance, an absence of fishing in area 1 leads to a rather substantial drop in CPUE – essentially a discontinuous drop. Similarly, when the fleet moves back into area 1, there is a substantial increase in the CPUE. The decision tree does not have a built in limit on the amount of change in the catch from one year to the next. (The real CPUE data may well imply that such a limit is not needed). The simulated data from the spatial model therefore leads to the massive adjustments in catch that are seen in some of the trajectories in Figure 151. One could argue that the CPUE from the simulation, particularly when the fleet moves in and out of areas, has not been 'standardised' before being used in the decision tree. On the other hand, if the fleet is not fishing in a given area, and there are no data from that area, standardisation becomes rather difficult (at times impossible), in simulations and in reality! This is a rather important observation and somewhat of a conundrum which would need to be carefully considered if spatial management is envisaged, or if there are strong spatial patterns in fleet behaviour.

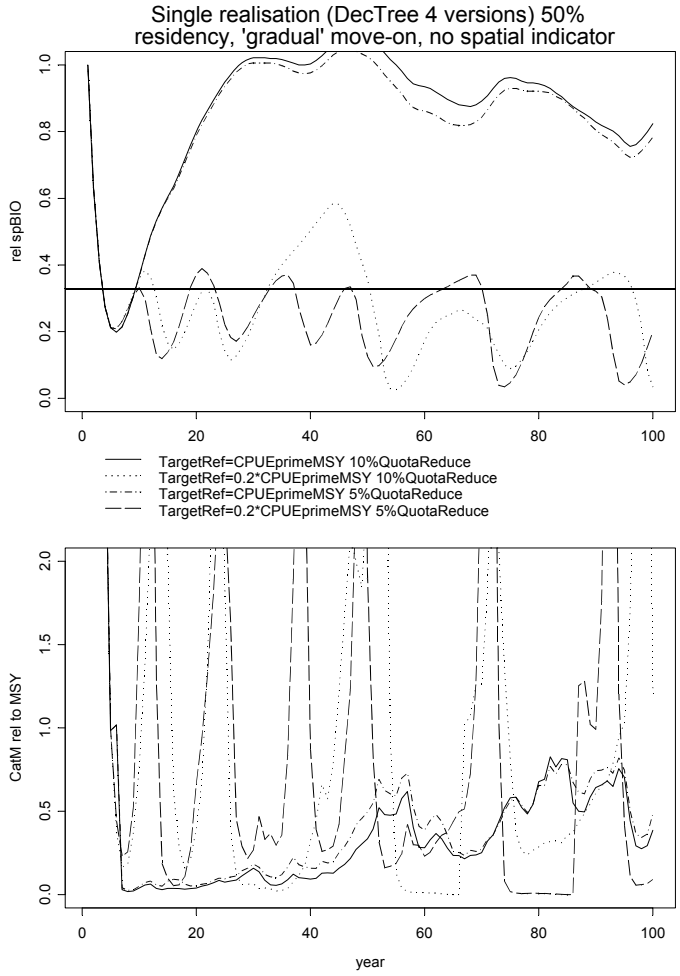


Figure 151. Four different tunings, or parameter settings, for the decision tree, run without a spatial indicator and assumptions of 50% residency and 'gradual move-on criteria. The top panel shows spawning biomass relative to unexploited; the lower panel shows total catch in weight relative to MSY.

It is worth adding that we ran some of the simulations 'pretending' that the CPUE was being sampled in all areas, including those that were NOT being fished. These simulations also included some oscillatory behaviour and large variability in catches under the decision tree rule. The 40/20 CPUE rule appears to be far less sensitive to this issue, most likely because it has a built-in maximum catch, and a relationship between the absolute catch and the indicator.

The second issue is the frequency of catch adjustments. The simulations all involved adjusting the catch annually. The decision tree uses 5 years' CPUE to estimate the slope in CPUE and hence the 'slope-to-target' quantity which determines whether the catch is increased or decreased. Based on results in Section 8, it is highly likely that the CPUE is autocorrelated, and that there is a lag between the management action and a response detected in the CPUE. Such time-lags can easily generate oscillations in a system. It may not in fact be necessary (or desirable) to adjust the catch, or effort, annually.

Preliminary investigations of the decision tree with a non-spatial model and with annual catch adjustments, or adjustments every 5 years, were conducted by the Harvest strategy working group (HSWG). The long term realisations with annual adjustments showed some tendency for damped oscillations (Appendix D in Campbell *et al.*, 2007), but nothing as extreme as seen with the spatial model. For practical reasons, only a subset of results are illustrated in Campbell *et al.* (2007), and

runs with catches adjusted every 5 years are not shown. Results for these runs mainly showed a slower increase in biomass at the start of the time-period than the annual version of the rule (as one would expect). In the longer term, performance seemed similar to that of the annual version. (We had access to these results since one of the authors of this report, Natalie Dowling, was also conducting the preliminary investigations of the decision tree for the HSWG).

The decision tree rule is undergoing further simulation testing on a simulation model conditioned to historic data (under an existing FRDC project 2007/017: "Integrated evaluation of management strategies for tropical multi-species long-line fisheries"), and long term oscillations or very large catch variability may, or may not be a feature under that set of simulation parameters. Results so far do, however, strongly suggest that the testing should include some long (say 100 year) runs to evaluate the long term dynamics of the decision tree rule, and that the frequency of implementation of the decision rule, as well as the number of years of catch rates used in calculating the slope, should be further explored.

In the rest of this Section we have retained the original parameter settings for the decision tree, and as will be noted below, the average relative biomass, over 10 realisations, is not as extreme as suggested by the single realisation shown in Figure 149.

When a spatial indicator is also used in each of the decision rules, the biomass is generally somewhat higher, but more importantly, the catch can be much more variable, particularly if the triggering of the spatial indicator leads to quite a big adjustment in the catch 'quota'. The percentage change (30% cut in catch) used in these simulations is arbitrary and admittedly rather large, but it serves to illustrate the effect; a small change of, say, 5%, is unlikely to show much of an effect in comparisons. The increase in catch variability when a spatial indicator is included, is obviously a combination of the magnitude of change in the catch and the frequency with which the spatial indicator is triggered. This highlights the importance of quantifying the trade-offs between the level of biomass, the level of catch and the variability in catch, as well as relevant measures of 'local depletion' when evaluating the performance of decision rules and their parameters.

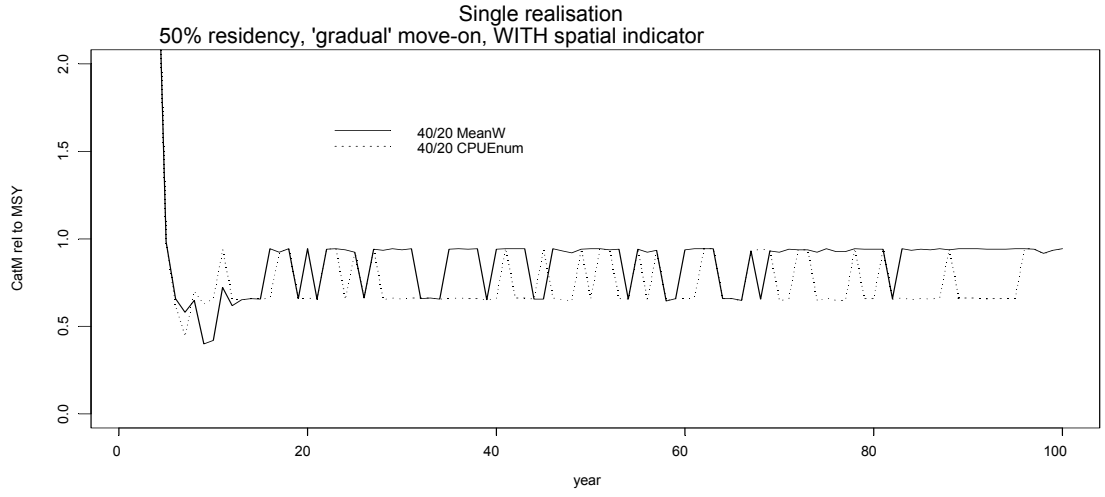


Figure 152. Single, long time-series, realisations of catch in weight (relative to C_{msy}) for the two 40/20 decision rules used with a spatial indicator which reduces the catch by 30% when it is triggered. All scenarios start with a harvest rate of $2.F_{msy}$ over the first 5 years. The decision rule starts in year 6.

The catch time series (one realisation) of the two 40/20 decision rules with a spatial indicator is shown in Figure 152. In this realisation the catch mostly fluctuates between a value close to C_{max} and a value 30% below that when the spatial indicator is triggered. It is interesting to note again that, in some sense, the fleet movement can be thought of as some form of spatial management (a redistribution of effort) which may obviate the need for a cut in catch as well.

More substantial differences between realisations, particularly for the decision tree under some scenarios will be evident in further results. The potential for long-term oscillations and the fact that it was not practical to test for this in each scenario (only 30-year time series were run), means that some performance measures, such as final biomass, or even average biomass at the end of the short term (30-year) simulations, should be interpreted with caution. The issue of local depletion also requires a focus on slightly different performance measures. Since the fleet dynamics are modelled to reflect a strong preference to fish in the inshore area (area 1), we consider the catch rates in area 1 as a relevant performance measure, particularly in comparisons between versions of a decision rule with or without the spatial indicator.

13.3.2 Overall short term behaviour

The first point to note is that biomass declines faster at the start of the time-series when the residency is low (50%) than when it is high (90%) (Figure 153). This feature is most marked when looking at relative biomass in the offshore areas (not shown). The result is due to a combination of the level of residency and the fact that the harvest starts in one area only. The extent of the difference in depletion between the two residency levels is also a function of the magnitude of the harvest rate; a higher harvest rate leads to a bigger difference. (These results, in terms of biomass relative to unexploited, should not be confused with the illustrated effect of residency on the proportions of the population in different areas, sub-section 2.2). For a given residency, the difference in depletion between the two move-on criteria is very small.

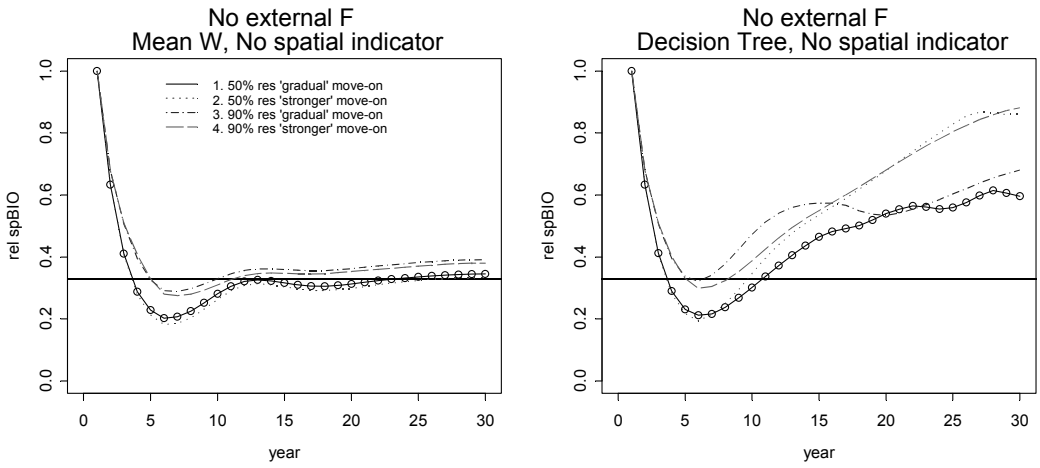


Figure 153. Average relative biomass (over 10 realisations) for the 40/20 rule based on mean weight (left panel) and the decision tree (right panel), and for two residency levels and move-on criteria.

Based on the average of 10 realisations, all the decision rules lead to a recovery of the spawning biomass to B_{msy} or above over the 30 year period. With the 40/20 decision rules, the 90% residency scenarios generally have slightly higher relative biomass at the end of the 30 years than

the 50% residency scenarios. The decision tree dynamics are more complex though, as illustrated above, and we have noted the need for caution when interpreting the average biomass trajectories based on only 10 realisations.

When a spatial indicator is also incorporated into the decision rule, the catches are, as expected, lower and the biomass higher. The spatial indicator has the strongest effect on the decision tree, and this is considered in more detail below. The decision rules have not been tuned, and their default settings seem to imply relatively conservative behaviour. It is therefore tempting to ask whether a spatial indicator is even necessary with respect to maintaining overall biomass at a desired level. Under these default settings, the answer is almost certainly 'no'. The question is, however, whether there are other reasons, associated with area-specific issues and catch rates, which may benefit from such an indicator, particularly if a less conservative tuning of the rule were to be used.

13.3.3 Observations : effects of fish and fleet movement

Although we are partly interested in the overall behaviour of the decision rules in the context of the spatial model, it became clear early on, that there are details due to interactions between the fleet dynamics, fish movement (or rather level of residency assumed) and the decision rules, that are worth noting.

At 50% residency, the average time series of total catch shows a drop in the early years of the decision rule when mean weight is used as the indicator in a 40/20 rule. This decline is not seen when CPUE is used in the 40/20 rule (Figure 154, left panel). It occurs because mean weight drops below the target reference point more often than CPUE. At 50% residency, the relative proportion of small fish in the catch in the inshore area is higher, as the fleet remains longer inshore under low residency and hence removes more of the larger fish, than is the case for 90% residency. Since the catch from the inshore area dominates the total catch, this leads to a low mean weight in the catch at the time the decision rule is implemented (Figure 155). It is, however, interesting to note that when residency is 90%, the CPUE-based 40/20 rule leads to a decrease in catch at the start of decision rule implementation, whereas the mean weight-based rule does not (Figure 154, right panel). Given that these figures are based on only 10 realisations, it is necessary to note that when individual realisations were considered, they all showed this pattern (Figure 156).

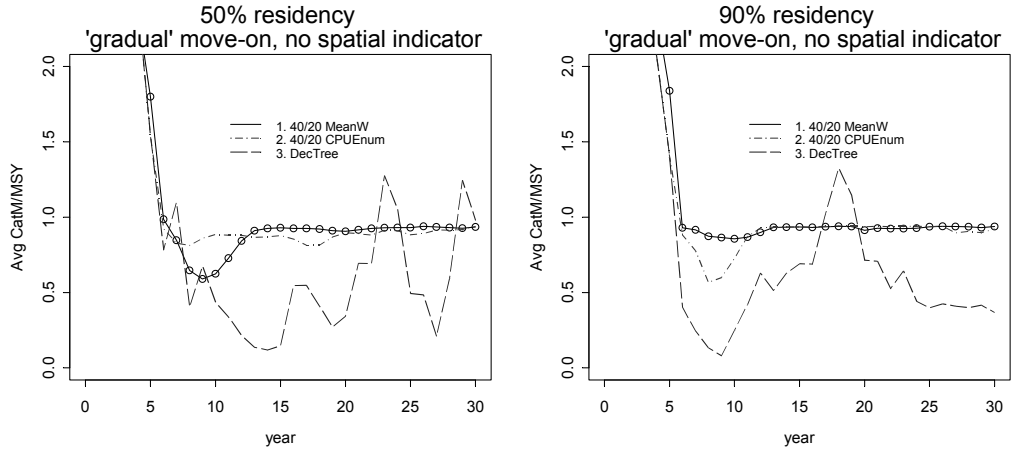


Figure 154. Average catch times series for two 40/20 decision rules, one using mean weight, the other using CPUE, and the decision tree. The left panel is for 50% residency, the right panel for 90% residency. The move-on criteria are 'gradual' and there is no spatial indicator.

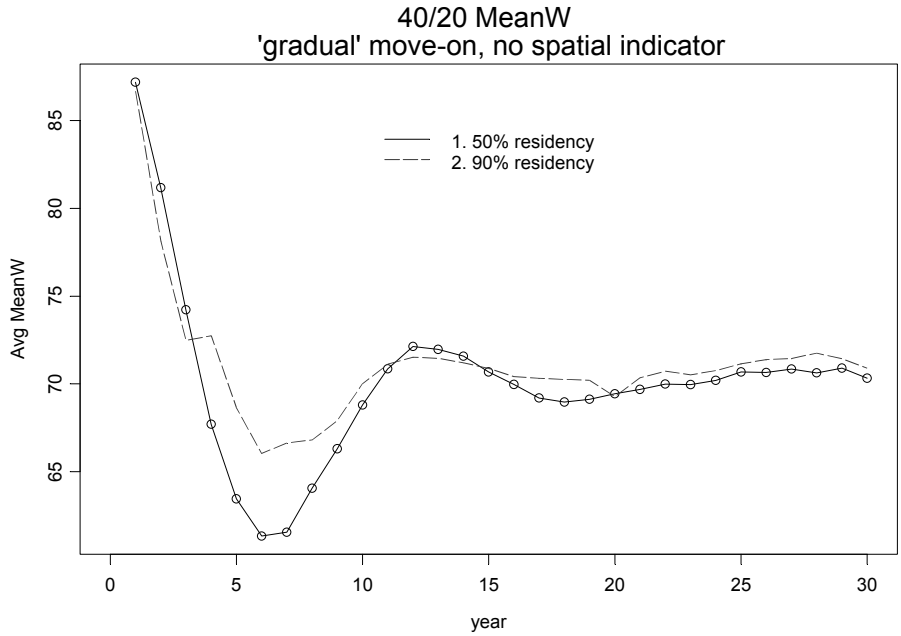


Figure 155. Average mean weight used in the 40/20 decision rule under the 'gradual' move-on criteria with no spatial indicator, and for 50% and 90% residency.

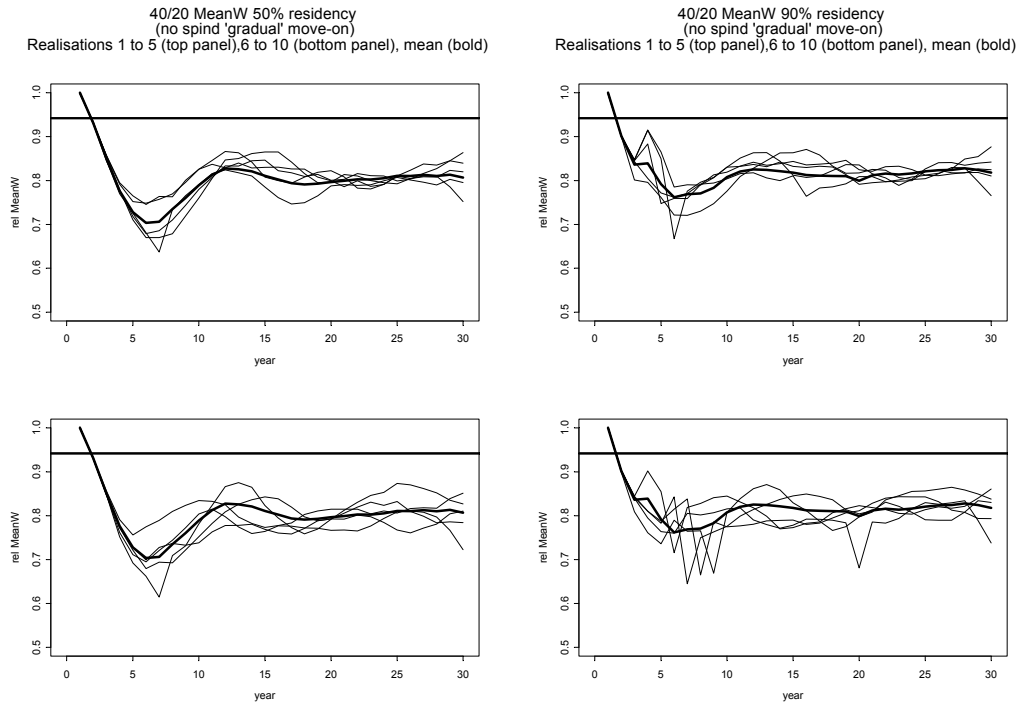


Figure 156. Time series of Mean weight from the 10 realisations in two scenarios: 50% residency (left panels) and 90% residency (right panels). In each case, the first 5 realisations are plotted in the top panel and the second 5 in the lower panel. The mean (over all 10) is plotted in bold, and the horizontal line is the target reference level.

Many other detailed observations regarding differences between areas under different decision rules, move-on criteria and residency can be made. These observations are, however, very often case-specific or driven by choice of parameters – for example the CPUE levels in the move-on criteria relative to the CPUE trigger points in the decision rule. We therefore consider that these detailed issues, which may well be very important and relevant in reality, are best explored in a conditioned simulation model.

13.3.4 Spatial depletion

In order to explore the issue of local depletion, we focus on the catch rates in the inshore (preferred) area, and compare the minimum and average catch rates for a decision rule with and without a spatial indicator. Recall that the (arbitrary) response to the spatial indicator being triggered is a rather large reduction (30%) in the total allowable catch (TAC). A comparison of the variability in the catch between the versions of the decision rule 'with' and 'without' a spatial indicator, will therefore reflect the number of times the spatial indicator was triggered.

In the rest of this sub-section, results for two versions of each rule – one without a spatial indicator, the other with a spatial indicator – are shown in pairs in boxplots. The first pair is usually for the constant harvest rate policy (i.e. no decision rule as such). Subsequent pairs are for the two 40/20 decision rules (using mean weight, or using CPUE) and for the decision tree. The legend on the boxplot confirms which is which. The important point to note is that the main comparisons should be between the pairs of 'boxes' in each panel, since the rules have not been tuned. Comparisons between the mean weight and CPUE-based 40/20 rules are not, however, unreasonable since the only difference between them is the indicator used.

Also note that the two CPUE panels in the plots are only for catch rates in the inshore area, but the catch variability and total catch panels are for all areas combined. Only the medians and inter-quartile ranges are shown given the small number of realisations (10) for each scenario.

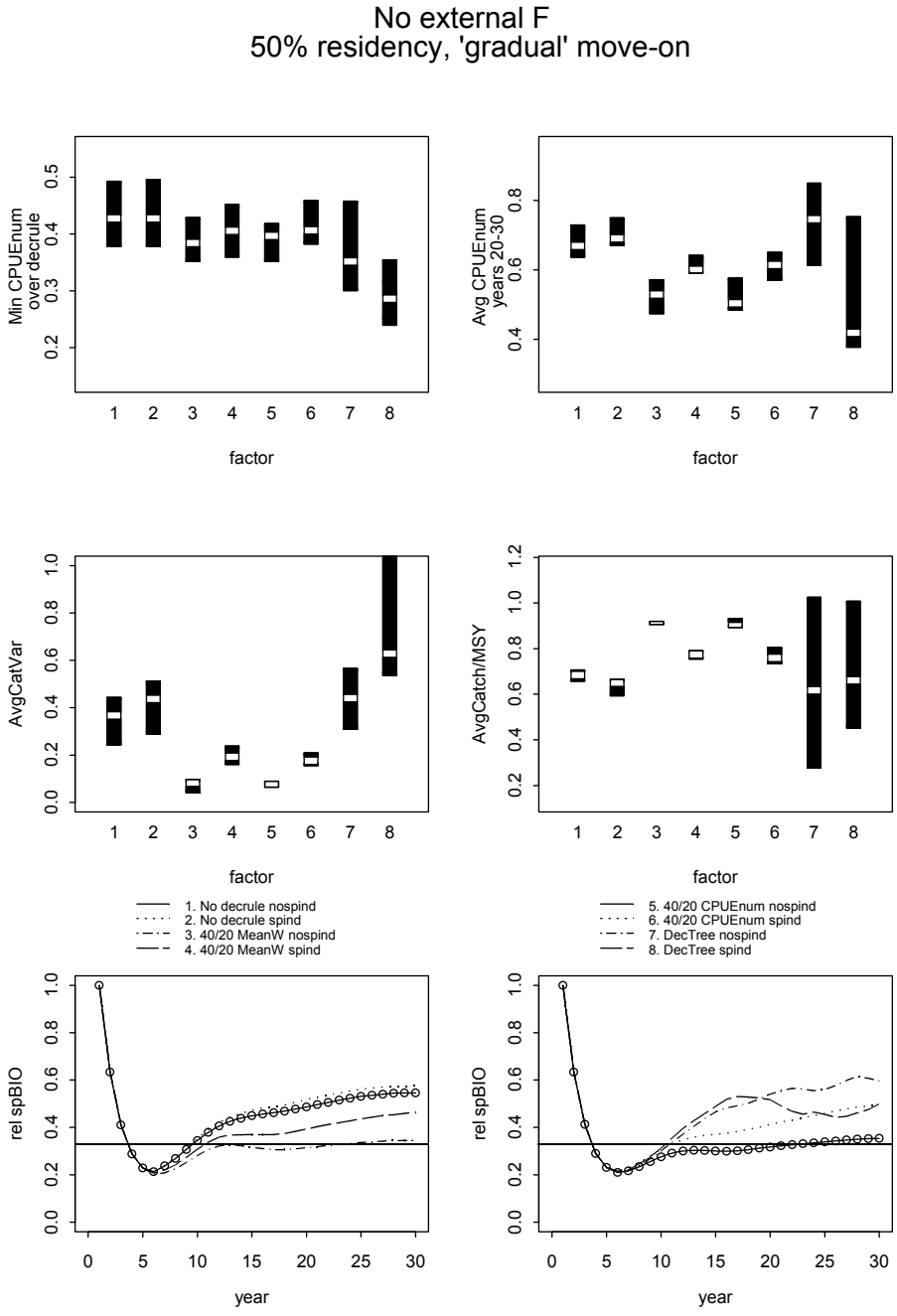


Figure 157. Scenario: 50% residency, gradual move-on criteria. Boxplots of catch rates (minimum and average over years 20-30) in the inshore area, and catches (variability and total, averaged over the decision rule period) in all areas, over 10 realisations. The 'factor' number identifies the decision rule used, as indicated in the legend. Time series of the average relative biomass (all areas) over the 10 realisations is also shown.

Figure 157 shows comparisons between scenarios without a spatial indicator ('nospind' in the legend of Figure 157) and with a spatial indicator ('spind' in the legend of Figure 157). The high variability between realisations for the decision tree, compared to the other rules, is clear from

the boxplots. Comparisons between pairs of results show that, for the constant harvest rate ('No decrule') and the two 40/20 decision rules, the inclusion of a spatial indicator leads to more variability in the catch, lower average catch, but higher average CPUE in the inshore area. The minimum (lowest) CPUE in the inshore area is also somewhat higher, but not markedly so. Note though that the minimum CPUE is strongly related to the move-on criteria and the relative CPUEs implied by the parameters (e.g. trigger points or targets) in the decision rules. Results for the two 40/20 rules are in fact also quite similar to one another.

Results for the decision tree are somewhat counter-intuitive. The minimum and average CPUE in the inshore area are lower when the spatial indicator is also used. The catch variability is increased, but the average catch is about the same. It is not obvious why this is occurring, but we do note that the response of the decision tree to the slope in CPUE is quite complex (see Appendix 1; for example, the CPUE can be above the target level, but if the slope over the previous 5 years is negative, the catch can still be reduced rather than increased).

In these results, as well as any of those shown further below, where the mean biomass pattern for the decision tree looks quite different from that of the other rules, and where there is large variability in the other performance measures, the average biomass trajectories need to be interpreted with caution. Figure 158 shows the very large differences between individual realisations, particularly towards the end of the 30 year period.

It may seem curious that some trajectories have very sudden and sharp decreases in biomass. Detailed consideration of those realisations showed that this tends to happen when the fleet moves back into area 1 where the catch rates are highest. The overall CPUE 'jumps' up, almost in a discontinuous manner, which affects the level and slope of the CPUE relative to the target. This then leads to a very large increase in the catch (recalling that the decision tree does not have an upper limit on the catch), and a subsequent sudden drop in biomass.

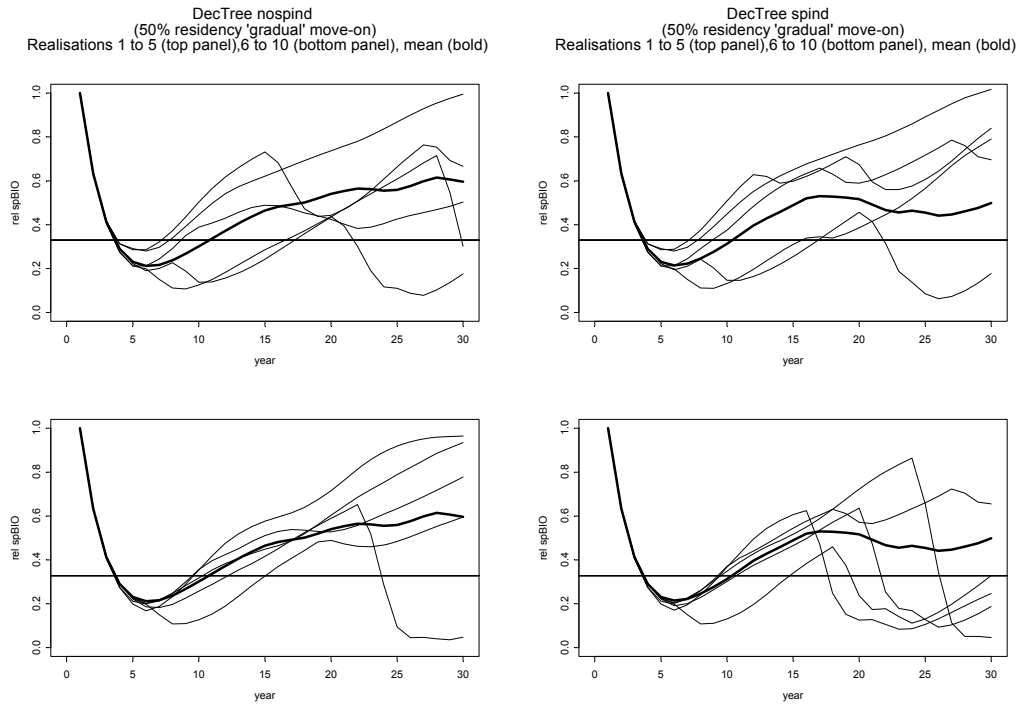


Figure 158. Time series of relative biomass from individual realisations for the decision tree without (left panels) and with a spatial indicator (right panels). The scenario is 50% residency and gradual move-on criteria. In each case, the first 5 realisations are plotted in the top panel and the second 5 in the lower panel. The mean (over all 10) is plotted in bold.

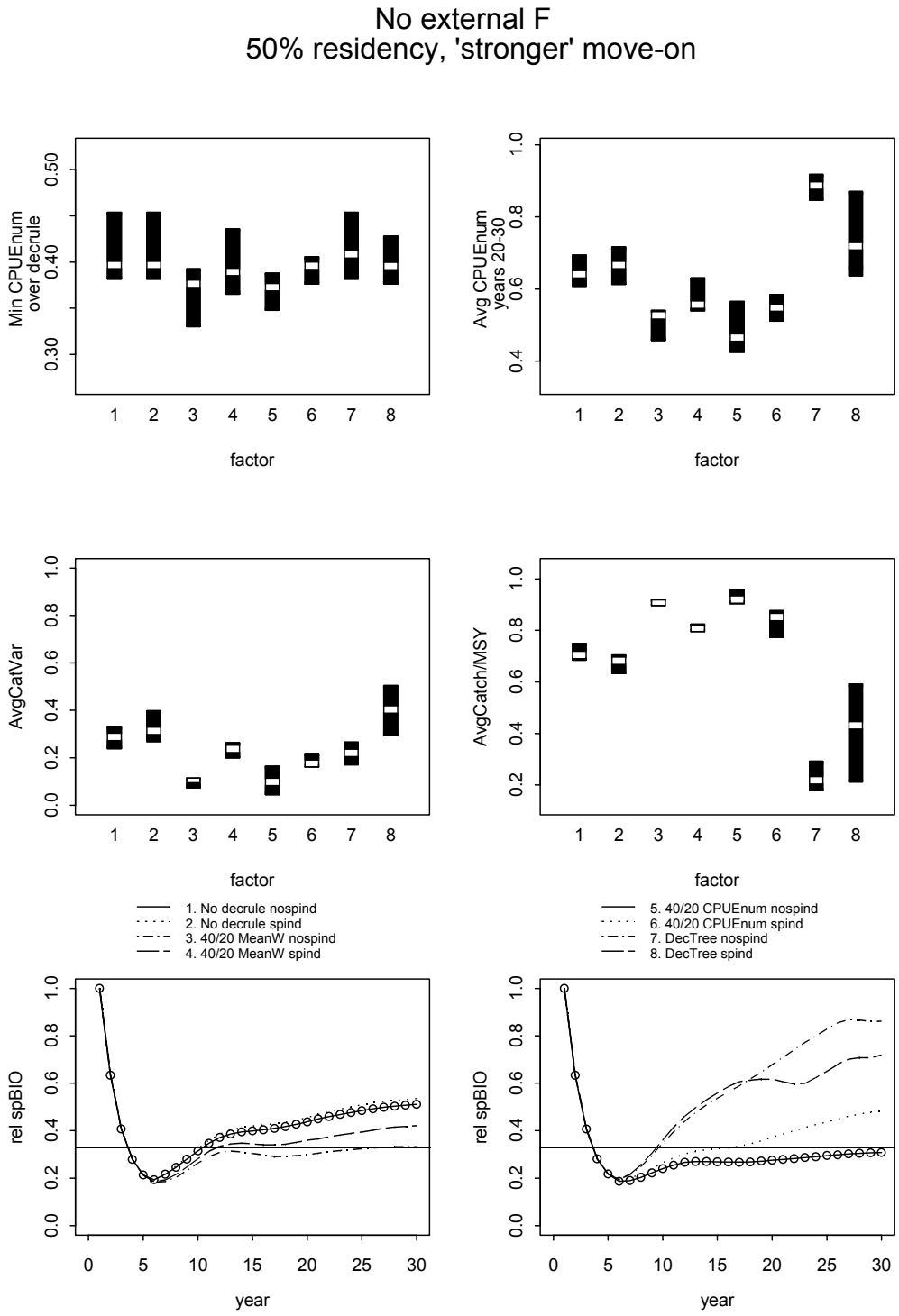


Figure 159. Scenario: 50% residency, stronger move-on criteria. Boxplots of catch rates (minimum and average over years 20-30) in the inshore area, and catches (variability and total, averaged over the decision rule period) in all areas, over 10 realisations. The 'factor' number identifies the decision rule used, as indicated in the legend. Time series of the average relative biomass (all areas) over the 10 realisations is also shown.

When the move-on criteria are such that a larger proportion of the fleet moves offshore at a smaller drop in the catch rate, results are similar in terms of the direction of change in the four

performance measures (Figure 159). The changes in performance measures for the decision tree are again in the opposite direction compared to the other rules. Note also how the variability between realisations of the decision tree rule has dropped substantially under this ('stronger') move-on criteria, compared to the gradual move-on criteria (see above).

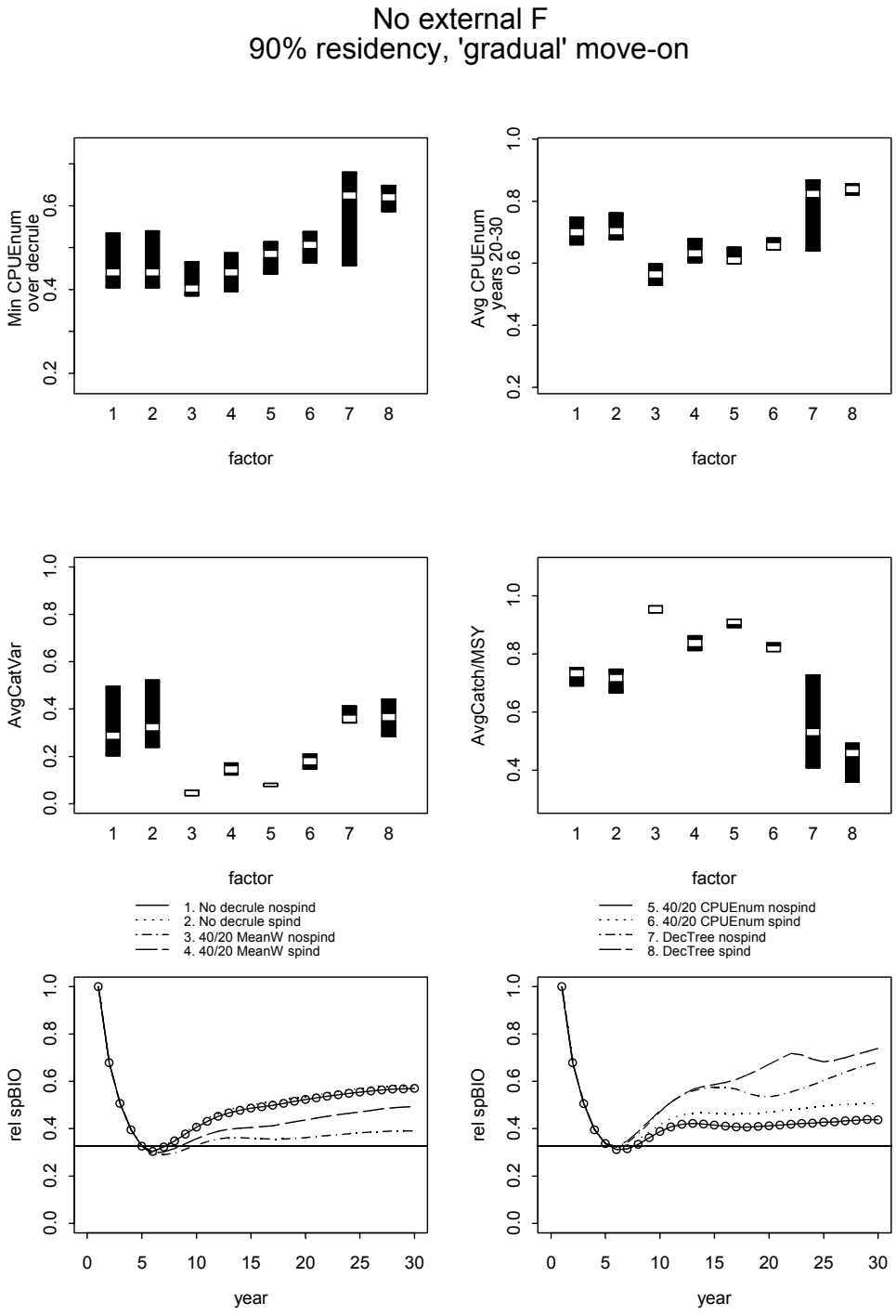


Figure 160. Scenario: 90% residency, gradual move-on criteria. Boxplots of catch rates (minimum and average over years 20-30) in the inshore area, and catches (variability and total, averaged over the decision rule period) in all areas, over 10 realisations. The 'factor' number identifies the decision rule used, as indicated in the legend. Time series of the average relative biomass (all areas) over the 10 realisations is also shown.

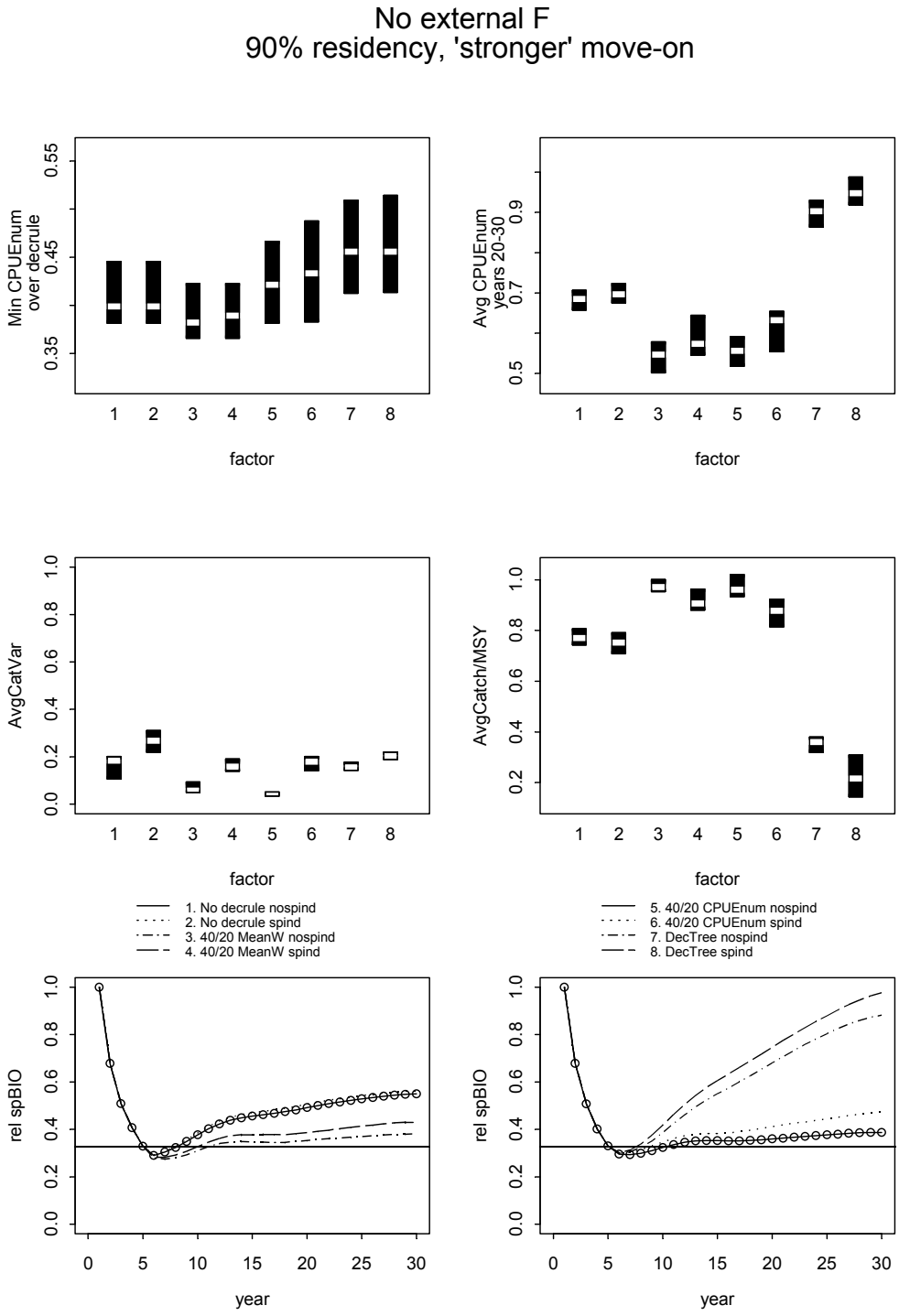


Figure 161. Scenario: 90% residency, stronger move-on criteria. Boxplots of catch rates (minimum and average over years 20-30) in the inshore area, and catches (variability and total, averaged over the decision rule period) in all areas, over 10 realisations. The 'factor' number identifies the decision rule used, as indicated in the legend. Time series of the average relative biomass (all areas) over the 10 realisations is also shown.

When higher residency (90%) is assumed, the decision tree behaves in a similar way to the other decision rules, with higher catch rates but lower average catch when the rule is used with a spatial indicator (compared to without a spatial indicator; Figure 160 and Figure 161). It also seems that

the 'stronger' move-on criterion leads to lower variability between realisations for the decision tree rule, given that the ranges of the 'boxes' are much smaller. This is only an observation, however, and would have to be confirmed by running a larger number of realisations.

The overall behaviour of the decision rules using their default parameter settings, is such that there is little need for a spatial indicator from the point of view of overall biomass. The inclusion of a spatial indicator can, however, increase the average CPUE in the preferred area (though note that this is partly due to the fact that when triggered, the spatial indicator leads to a cut in the catch and hence to higher biomass in general). It is interesting to consider the notion that a spatial indicator could be used to trigger fleet movement (or effort re-distribution as some form of spatial management; not necessarily area closure), rather than a cut in the catch. In the simulation, this could be modelled by activating additional fleet movement when the spatial indicator is triggered rather than through the catch rate mechanism we have used here.

13.4. Results: Additional uncontrolled harvest

The scenarios in this sub-section are meant to mimic the situation where the same stock that is being harvested by the 'domestic' fleets (managed by the decision rule), is also being harvested by an 'international' fleet which is not controlled. The additional harvest occurs in an area adjacent to the most offshore domestic area, and is referred to as the 'outside' area.

13.4.1 Overall short term behaviour

In this set of scenarios, the main effect of an additional, uncontrolled, harvest rate on the stock is lower overall biomass relative to B_{msy} (particularly since we have simulated a very high level of additional fishing to exaggerate and illustrate the effect; $F_{outside}=3.F_{msy}$), lower relative biomass in the domestic areas and hence also lower catches and catch rates.

The relative performance of the decision rules is similar to the case where there is no outside harvest. In this regard, therefore, the main issue is not 'which rule' to use, but rather how to tune a chosen rule under this particular uncertainty of stock structure. It is very likely that there would be area-specific detail which would differ between rules (as we saw in sub-section 3 above for scenarios without additional uncontrolled harvest), but again we consider that this is best explored in a conditioned model with more realistic fleet dynamics.

There is also no strong overall effect of the different residency levels. For any given decision rule, the differences between relative catches without 'outside' fishing and with 'outside' fishing are similar irrespective of the assumed residency. Again, there are likely to be differences in the spatial details. Note, however, that fish movement in these simulations are such that fish move from any of the 5 areas TO any of the 5 areas. Results are likely to be quite different if there is a constraint on distance that fish can moved between time-steps, so that they can only move to, say, adjacent areas.

13.4.2. Spatial indicator

The inclusion of a spatial indicator also has a similar effect on results as it does in the scenarios where there is no additional harvest. The decision tree now only shows its counter-intuitive behaviour when residency is 90% and the move-on criteria are 'gradual'.

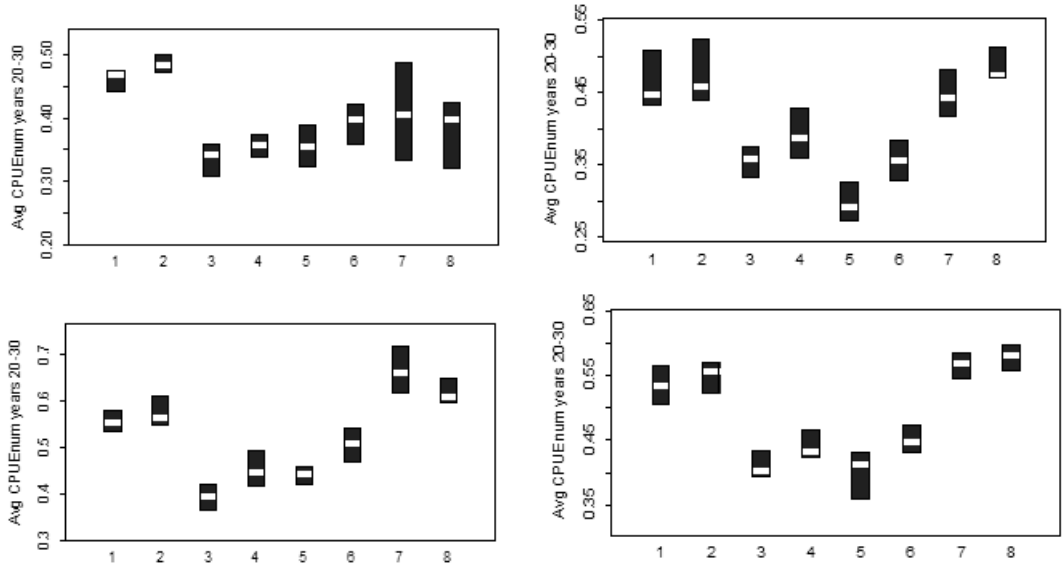


Figure 162. Average CPUE in area 1 over the last 11 years of the decision rule, for scenarios with uncontrolled harvest outside. Panels are: 50% residency, gradual move-on (top left); 50% residency, stronger move-on (top right); 90% residency, gradual move-on (bottom left); and 90% residency, stronger move-on. Within each panel, the decision rules are as in Figure 161, i.e. in pairs for without and with the spatial indicator, for constant harvest rate (1,2), 40/20 mean weight (3,4), 40/20 CPUE (5,6) and the decision tree (7,8).

Additional figures for scenarios with additional uncontrolled harvest can be found in sub-section 7.

13.5. Results: Sensitivity trials

There are many uncertainties and assumptions which could be explored. For practical reasons, we focus on three which are likely to have the strongest effect on results: effort creep, unknown steepness (in the stock-recruit relationship), and a poor choice of the maximum catch in the 40/20 decision rules. The decision tree does not have an explicit maximum catch, but its target CPUE, used at the first decision point, is based on MSY considerations and therefore affected by assumptions about steepness. We therefore include the decision tree in a set of trials where steepness is assumed to be lower. Given the relatively small effect of residency and move-on criteria on the performance measures of interest for the main set of scenarios, only the scenario with 90% residency and gradual move-on criteria is used in these trials.

Implicitly there is another issue, namely, sensitivity to assumptions about stock structure. This has already been looked at through the 8 basic set of scenarios which were run with and without additional effort in an 'outside' area, meant to mimic the notions of a single or a shared stock. This is not re-addressed here, but further comments on results are made in the discussion sub-section.

Some of the other quantities that should ideally also be included in sensitivity analyses for a real situation, are the assumptions about selectivity (flat or dome-shaped, for example), and the parameters used to calculate the trigger points used in decision rules together with assumptions about the actual population dynamics. A limited number of runs exploring these issues with the spatial model, as well as some of the outcomes from the non-spatial analyses (Section 10), suggested that these issues are likely to have less impact on performance than the three mentioned above.

13.5.1. Effort Creep

Trials were only done for the 40/20 CPUE rule and the decision tree, since the 40/20 mean weight decision rule should not be affected by effort creep. Also recall that management is through a catch quota, not effort, in these simulations, so increased catchability does not lead to an increase in the catch.

The minimum and average inshore CPUE under the 40/20 CPUE rule behave as one might expect, with a higher minimum and average when there is effort creep (Figure 163). For the 40/20 CPUE rule, the catches are slightly higher. This is to be expected, but note that catches are constrained to C_{max} by the design of the rule. The same figure shows that the decision tree again responds in a much more complicated way, particularly with respect to the minimum CPUE. Without the spatial indicator, the average catch is almost the same irrespective of whether there is effort creep or not.

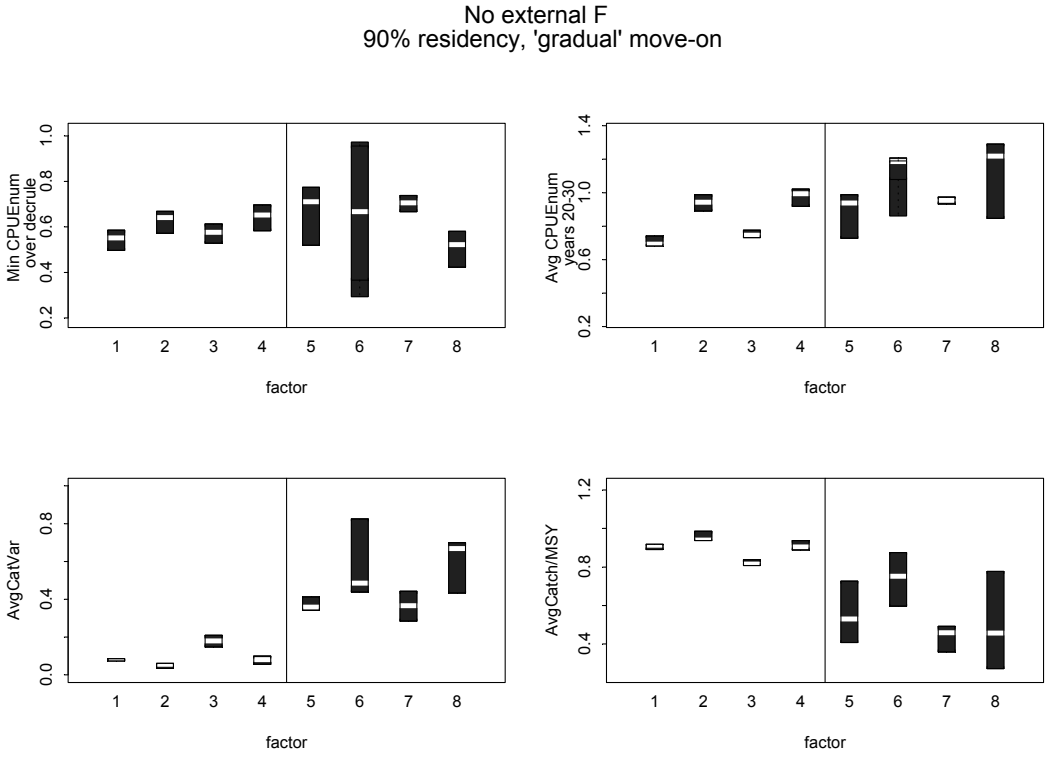


Figure 163. Scenario: 90% residency, gradual move-on criteria. Boxplots of catch rates (minimum and average over years 20-30) in the inshore area, and catches (variability and total, averaged over the decision rule period) in all areas, over 10 realisations. Results are plotted in pairs of scenarios without and with effort creep for the following decision rules: 40/20 CPUE rule without spatial indicator (1,2) and with spatial indicator (3,4); to the right of the vertical line are the decision tree rule without spatial indicator (5,6) and with spatial indicator (7,8).

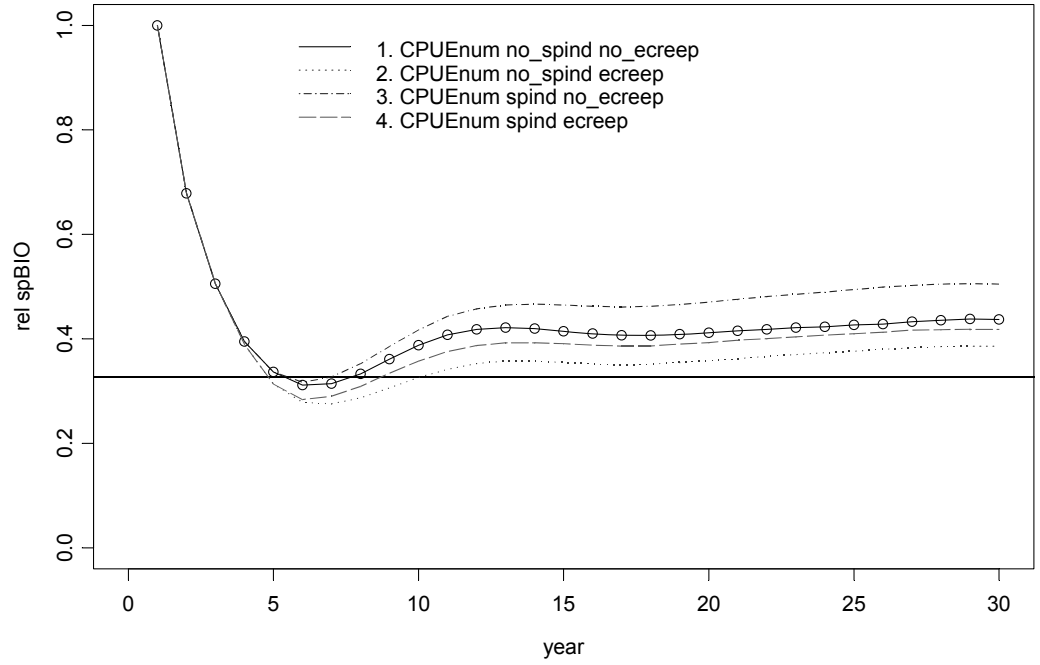


Figure 164. Time series of average (10 realisations) relative biomass for the 40/20 CPUE decision rule without and with effort creep, and without and with a spatial indicator. Residency is 90% and the move-on criteria are gradual.

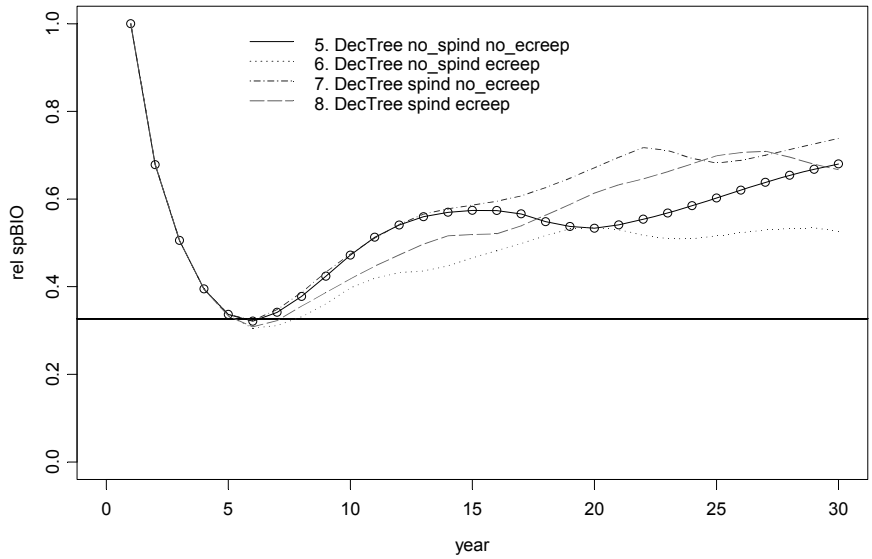


Figure 165. Time series of average (10 realisations) relative biomass for the decision tree rule without and with effort creep, and without and with a spatial indicator. Residency is 90% and the move-on criteria are gradual. Average trajectories should be interpreted with caution for this decision rule.

The summary biomass plots for the 40/20 CPUE rule (Figure 164) also show expected outcomes: when there is effort creep, the biomass is generally lower than when there is no effort creep because the catches are kept somewhat higher. It is harder to interpret the average biomass trajectories for the decision tree, given the high variability in some scenarios, and the possibilities

of oscillations. Results for the decision tree with spatial indicator are more intuitive, since the average trajectories are similar but at different levels, with a higher biomass for the scenario without effort creep.

Plots of single realisations for the 40/20 CPUE rule (Figure 166) show that there is relatively little variability between realisations (recall that random and measurement errors have deliberately been set quite low in these scenarios). The averages of 10 realisations are therefore reasonable representations of the dynamics.

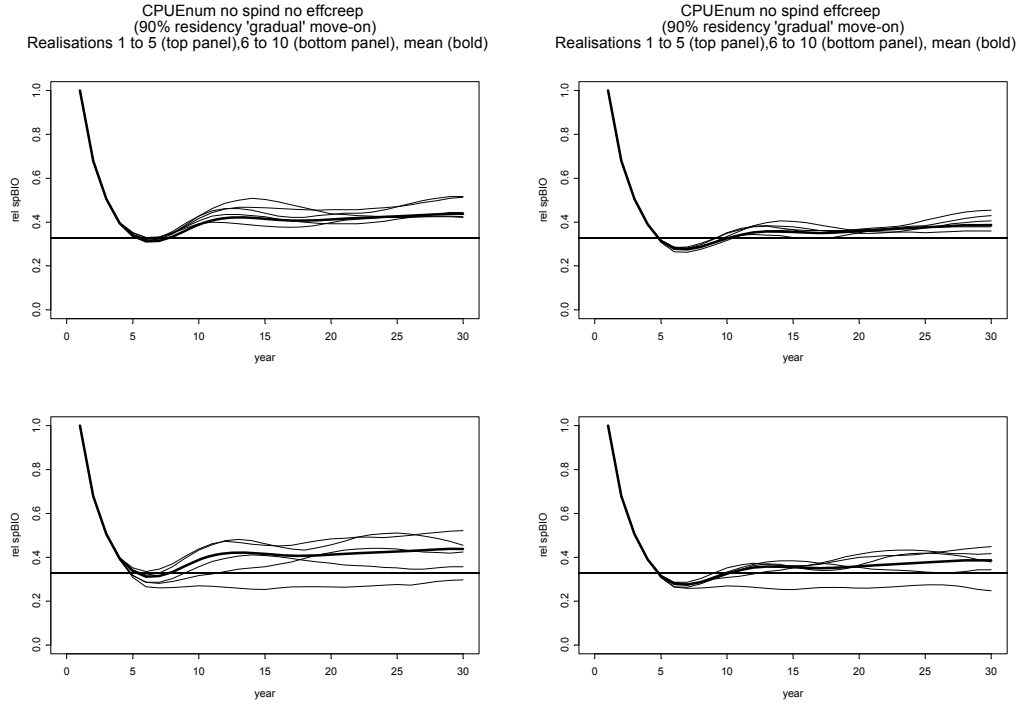


Figure 166. Time series of relative biomass from individual realisations for the 40/20 CPUE rule without effort creep (left panels) and with effort creep (right panels). The scenario is 90% residency and gradual move-on criteria, and no spatial indicator was used. In each case, the first 5 realisations are plotted in the top panel and the second 5 in the lower panel. The mean (over all 10) is plotted in bold.

Plots of single realisations for the decision tree rule (Figure 167), however, confirm concerns. There is very large differences between realisations, and the averages of 10 realisations is not really representative of any of the dynamics seen in the individual realisations, with the possible exception of the first ten years or so. In a case like this, it would be better to run many more realisations, or to do the comparisons between each pair of realisations.

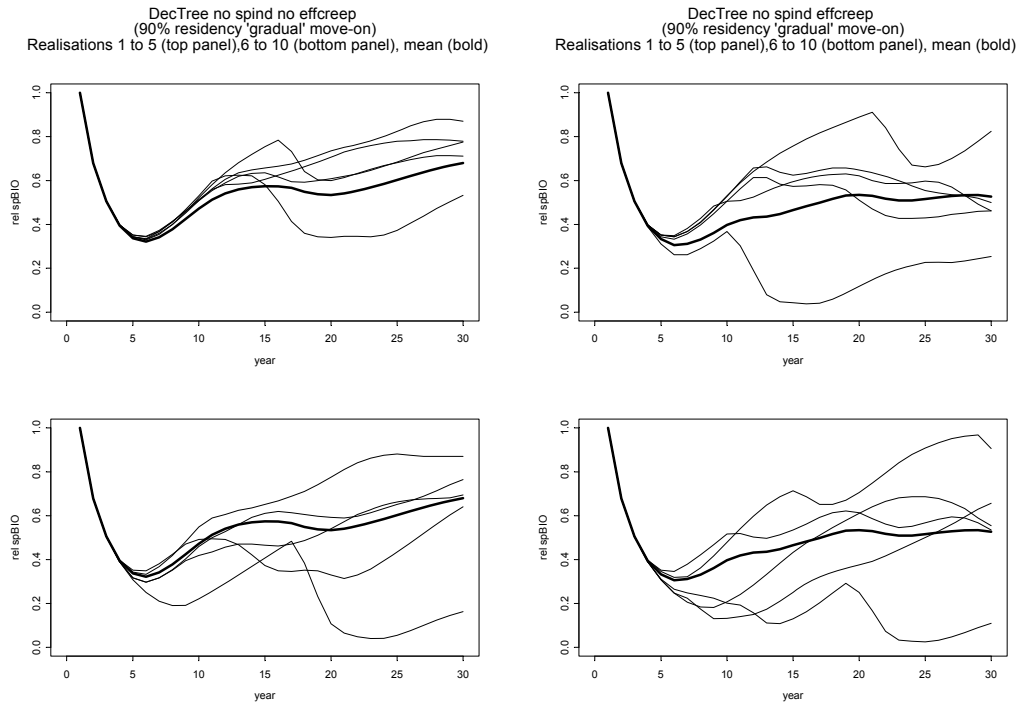


Figure 167. Time series of relative biomass from individual realisations for the decision tree rule without effort creep (left panels) and with effort creep (right panels). The scenario is 90% residency and gradual move-on criteria, and no spatial indicator was used. In each case, the first 5 realisations are plotted in the top panel and the second 5 in the lower panel. The mean (over all 10) is plotted in bold.

13.5.2 Maximum catch in 40/20 decision rules.

In the base case scenarios, the maximum catch for the two 40/20 decision rules was C_{msy}. This choice is relatively arbitrary, and should not be seen as a suggestion that this should be done in reality. We are aware of the risks of using MSY-related quantities as target reference points, and the 'best practice' approach of using these quantities as limit reference points instead. More importantly though, in reality C_{msy} is unlikely to be known or well estimated. The issue of how to set a maximum catch or effort level in a decision rule is not an easy one. There is a real danger that the maximum could be set incorrectly – too high, or even unnecessarily low. Here we only look at the case where it is set too high, since this implies a higher risk to the stock.

Some rules, such as the decision tree, avoids this difficult issue by not specifying a maximum, and the decision tree is therefore not included in these trials.

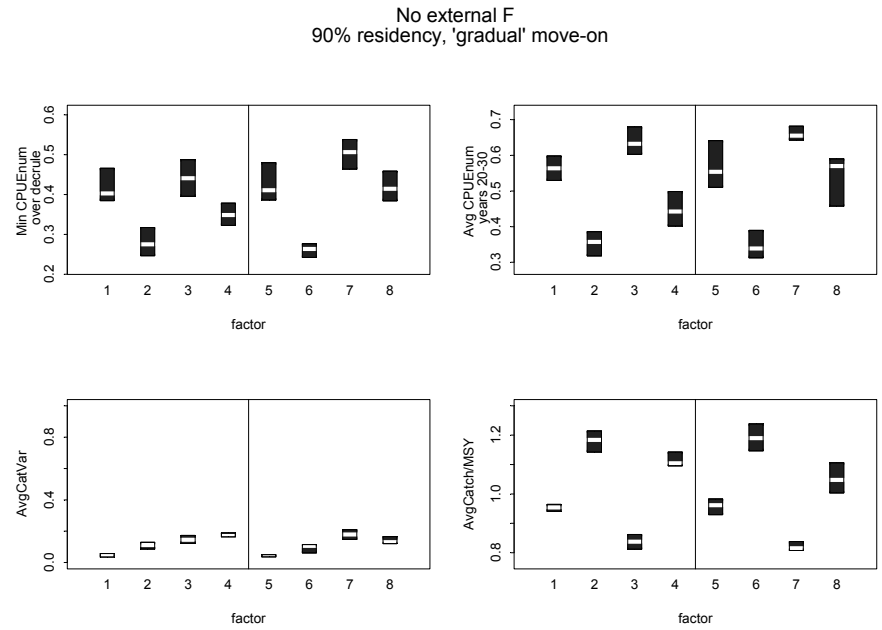


Figure 168. Scenario: 90% residency, gradual move-on criteria. Boxplots of catch rates (minimum and average over years 20-30) in the inshore area, and catches (variability and total, averaged over the decision rule period) in all areas, over 10 realisations. Results are plotted in pairs of scenarios with Cmax set correctly and with Cmax set 50% too high for the following decision rules: 40/20 mean weight rule without spatial indicator (1,2) and with spatial indicator (3,4); to the right of the vertical line is the 40/20 CPUE rule without spatial indicator (5,6) and with spatial indicator (7,8).

It is interesting to note how, when C_{max} is set too high, the inclusion of a spatial indicator results in the CPUE not dropping quite as low (minimum CPUE is higher; Figure 168). The biomass plots (Figure 169 and Figure 170) also suggest that the relative biomass at the end of the period is not quite as much lower (for C_{max} set too high, compared to C_{max} set correctly) when the spatial indicator is included. This effect is, however, not very strong; the ratio is about 0.5 without spatial indicator, and about 0.6 to 0.7 with spatial indicator. The benefit probably comes primarily from the fact that higher catches (for high C_{max}) lead to lower biomass, implying lower, and faster declining CPUE, which invokes the spatial indicator earlier and more often. The average catch panel (Figure 168) shows that, when C_{max} is too high, the catches are markedly lower when the spatial indicator is included in the rule, than when it is not; compare scenarios (2 and 4) and (6 and 8).

There is not much noticeable difference between the overall performance of the mean weight and the CPUE versions of the rule under the relative low noise and measurement error assumptions used here.

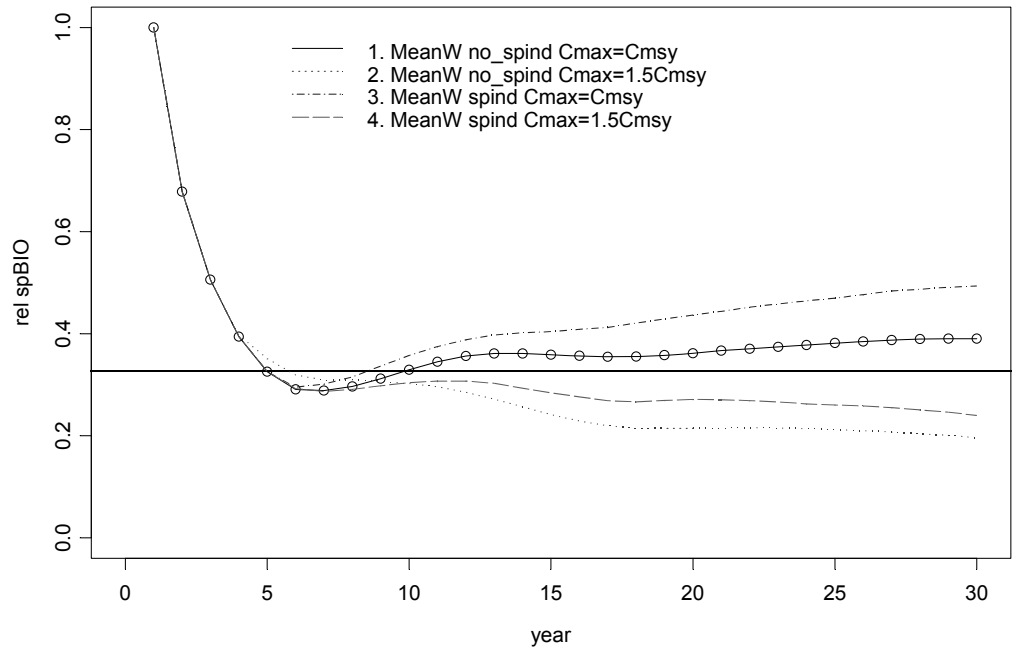


Figure 169. Time series of average (10 realisations) relative biomass for the 40/20 mean weight decision rule with Cmax set correctly or incorrectly, and without and with a spatial indicator. Residency is 90% and the move-on criteria are gradual.

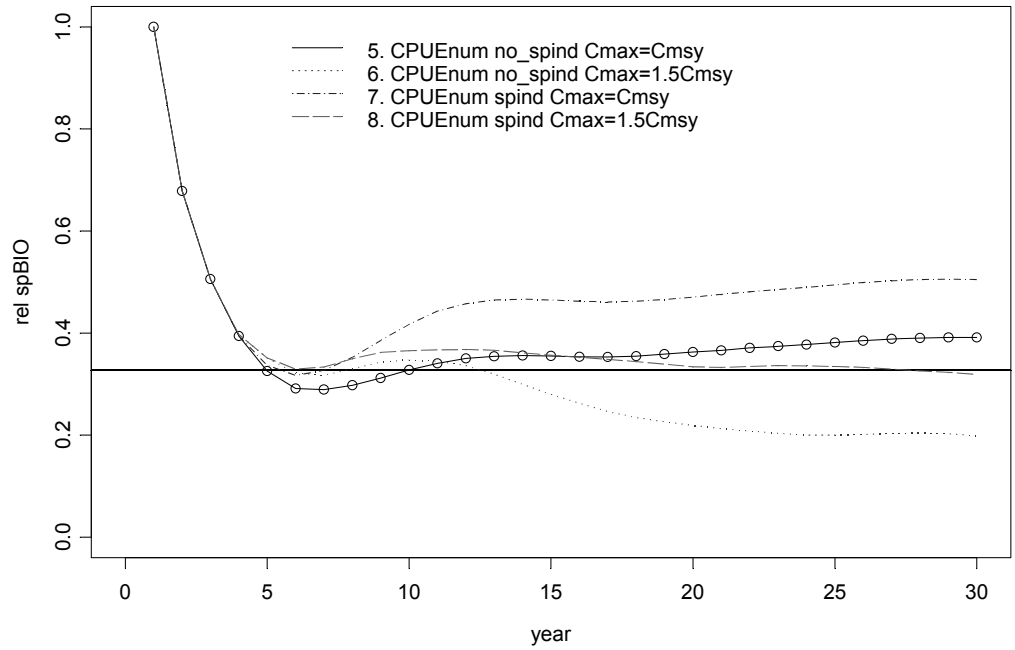


Figure 170. Time series of average (10 realisations) relative biomass for the 40/20 CPUE decision rule with Cmax set correctly or incorrectly, and without and with a spatial indicator. Residency is 90% and the move-on criteria are gradual.

13.5.3 Low steepness

In this set of scenarios, the true population dynamics are simulated using a very low steepness of h=0.3. In the 40/20 decision rules, C_{msy} is based on the assumption that steepness is 0.65, and then C_{max} is set to $1.5C_{msy}$. (Recall that the calculation of spawner per recruit trigger points for mean weight and CPUE are independent of steepness.) It is obvious, *a priori*, that this could easily lead to very poor performance, and a badly over-exploited stock, or even stock collapse. These scenarios are compared with ones where C_{max} is set too high ($C_{max}=1.5C_{msy}$), but C_{msy} is based on the correct, low, steepness.

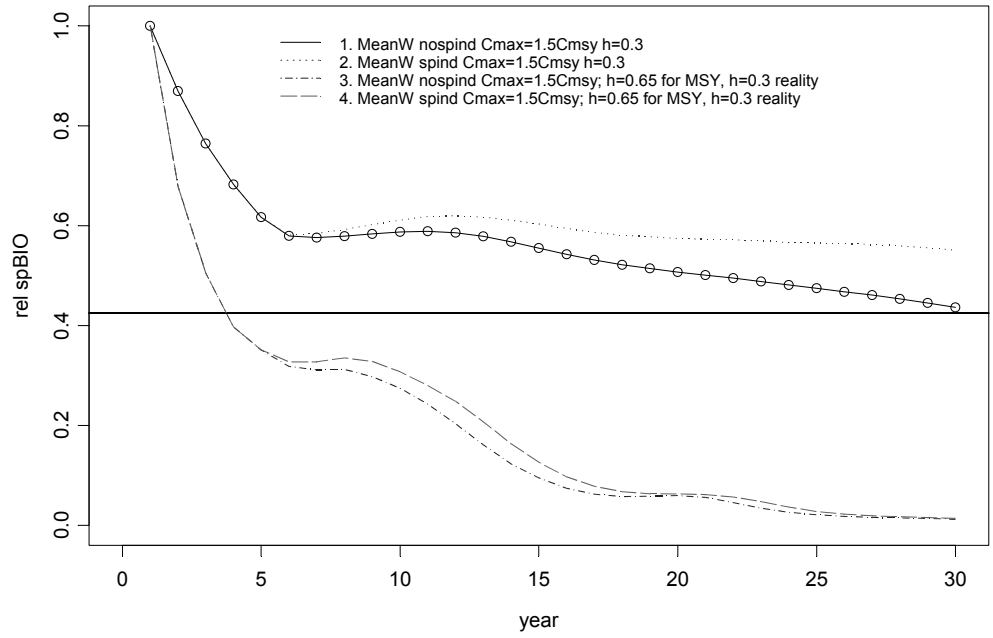


Figure 171. Time series of average (10 realisations) relative biomass for the 40/20 mean weight decision rule with Cmax set too high, but its calculation based on the correct steepness (h=0.3), and without and with a spatial indicator (1,2); and Cmax set too high, and based on the incorrect steepness (h=0.65), without and with a spatial indicator (3,4). In all cases true steepness is 0.3. Residency is 90% and the move-on criteria are gradual.

Figure 171 shows that the 40/20 mean weight decision rule cannot avoid a stock collapse, irrespective of whether a spatial indicator is used or not. Here it is useful to recall the result from Sections 8 and 9, where we show that F_{SPR40}, and hence its associated trigger points for indicators, are conservative for steepness values down to about 0.6, but below that, F_{SPR40} is higher than F_{msy} and therefore too high. Although the decline would be slower if $C_{max} = C_{msy}$, based on h=0.65, (instead of **1.5**C_{msy} based on h=0.65), it is very likely that the simulations would still show stock collapse because of the choice of trigger points for the indicators. This can be addressed by choosing more conservative spawner per recruit ratios (for example, constructing a '60/30' rule, say) when defining triggers, but the size-based indicators can then lead to poor performance because if these trigger points are too close to the unexploited values (see Sections 8 and 9).

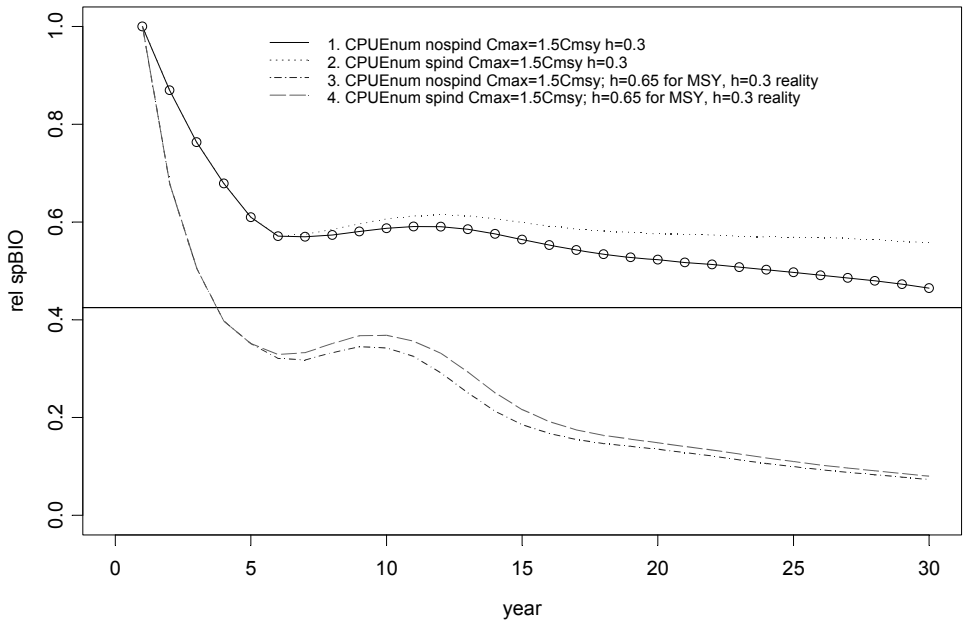


Figure 172. Time series of average (10 realisations) relative biomass for the 40/20 CPUE decision rule with Cmax set too high, but its calculation based on the correct steepness (h=0.3), and without and with a spatial indicator (1,2); and Cmax set too high, and based on the incorrect steepness (h=0.65), without and with a spatial indicator (3,4). In all cases true steepness is 0.3. Residency is 90% and the move-on criteria are gradual.

The decision tree rule is constructed quite differently from the 40/20 rules. The scenarios we have run to explore the sensitivity of the decision tree to steepness are therefore not directly comparable those for the 40/20 rule. The only quantity in the decision tree that depends on MSY-type quantities is the target CPUE (referred to as $CPUEprime_{MSY}$) which is used in the 'slope-to-target' calculations (see Appendix 1). This target CPUE is higher when steepness is low (obviously, the ratio B_{msy}/B_0 is also higher when steepness is low). Other trigger points in the decision rule are based on spawner-per-recruit considerations and are unchanged in these trials.

Figure 173 shows the time-series of relative biomass, averaged over 10 realisations. There are again some signs of oscillatory behaviour, in some scenarios and hence some caution is required in interpreting these results. Figure 174 shows individual realisations for the scenarios where steepness is low and where the target CPUE in the decision tree is based on that correct, low, steepness. It is in fact interesting to note that the spatial indicator (right panels) now has quite a strong effect on the dynamics, and its inclusion leads to higher biomass in all realisations. The sharp drop seen in many realisations towards the end of the time-period suggests that longer simulations runs should be considered before drawing firmer conclusions.

Figure 175 shows realisations for the case where 'true' steepness is low, but the decision tree's target CPUE is based on high steepness (h=0.65). There is so much variability, with some trajectories being well above B_{msy} and others well below it, that the average based on only 10 realisations cannot be considered reliable. It is however fair to say that, in some cases, the decision tree seems to be capable of halting, or even avoiding, a decline over the 30 year period.

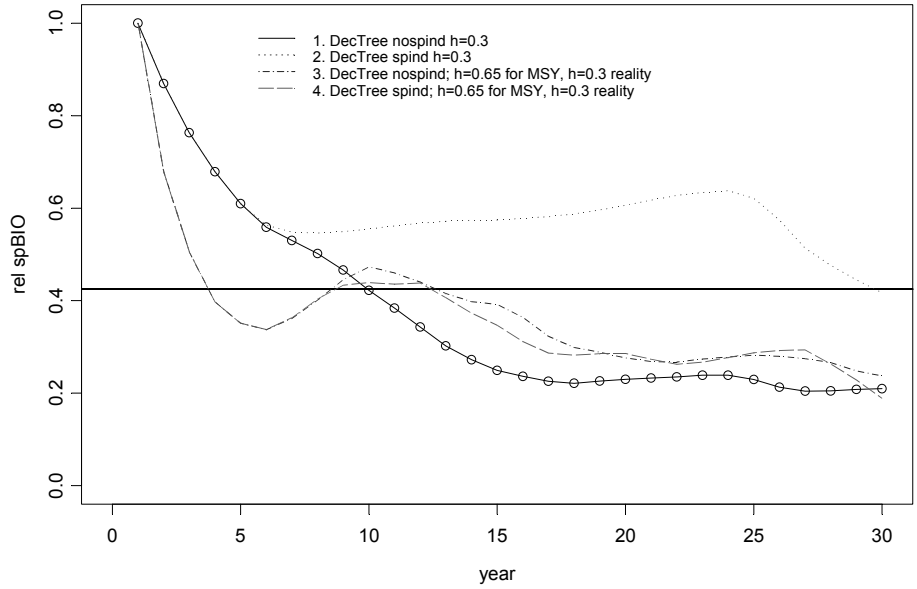


Figure 173. Time series of average (10 realisations) relative biomass for the decision tree rule with target CPUE based on the correct steepness (h=0.3), and without and with a spatial indicator (1,2); and target CPUE based on the incorrect steepness (h=0.65), and hence set too low, without and with a spatial indicator (3,4). In all cases 'true' steepness is 0.3. Residency is 90% and the move-on criteria are gradual.

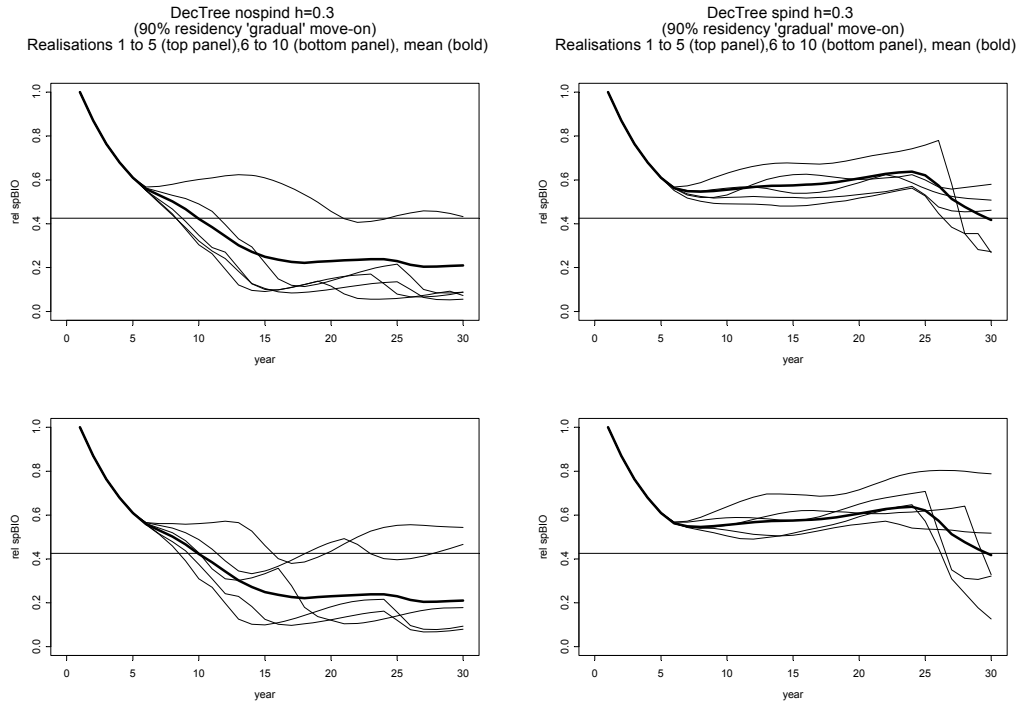


Figure 174. Time series of relative biomass from individual realisations for a scenario with low steepness (h=0.3), 90% residency and gradual move-on criteria. The decision tree rule has a CPUE target based on the correct (low) steepness and is used without a spatial indicator (left panels) and with a spatial indicator (right panels). In each case, the first 5 realisations are plotted in the top panel and the second 5 in the lower panel. The mean (over all 10) is plotted in bold.

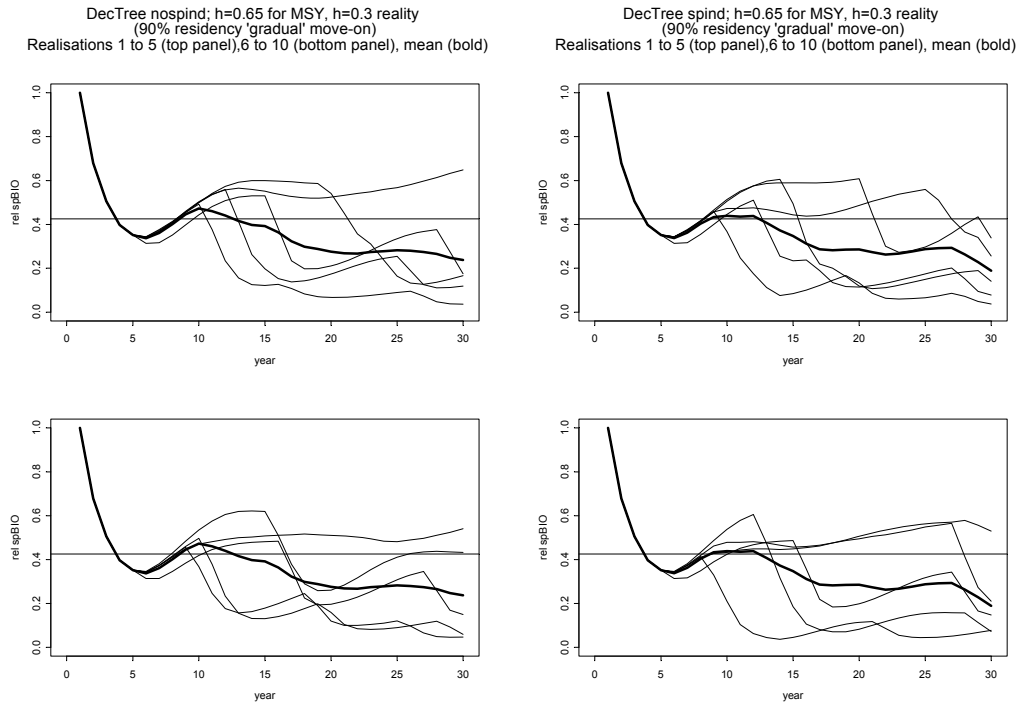


Figure 175. Time series of relative biomass from individual realisations for a scenario with low steepness (h=0.3), 90% residency and gradual move-on criteria. The decision tree rule has a CPUE target based on the incorrect (high) steepness, h=0.65, and is used without a spatial indicator (left panels) and with a spatial indicator (right panels). In each case, the first 5 realisations are plotted in the top panel and the second 5 in the lower panel. The mean (over all 10) is plotted in bold.

13.5.4 Sensitivity trials for scenarios with additional uncontrolled harvest.

Only a very limited number of sensitivity trials regarding effort creep and with the maximum catch set too high in the 40/20 decision rules were conducted (see Table 25), partly due to time limitations and partly because results for the base case scenarios strongly suggested that the patterns of change and the broad effects would be similar.

The relative effect of effort creep is essentially the same for these scenarios as for the ones where there is no additional uncontrolled harvest. The decision tree realisations are again very variable, making it difficult to draw firm conclusions about overall behaviour on the basis of only 10 realisations. The relative effect of setting the maximum catch in the 40/20 decision rules too high is also very similar for these scenarios as for the ones where there is no additional uncontrolled harvest.

13.6. Discussion and conclusions

The two main issues considered here are local depletion and stock structure. The investigations are only preliminary and by no means exhaustive. In the process of our evaluations with the spatially explicit model which includes fish movement and fleet movement, many useful observations were made regarding the decision tree rule which was developed by the ETBF and WTBF Harvest strategy working group (HSWG). We start by highlighting these with the hope that our observations will be useful for the further simulation testing being conducted under an FRDC-funded project.

The second part of this sub-section contains general observations regarding the complexities of (even simple) spatial models with fish and fleet movement, and the interactions between the parameterisation of these dynamics and that of the decision rules.

It is worth briefly reflecting on the approach we chose for this 'spatial' investigation. The notion was to use a relatively generic model which nonetheless related to the broad characteristics of (and qualitative observations from) the swordfish fisheries off the East and West coasts of Australia. The notion was also to run scenarios that could generate strong spatial signals. If spatial indicators appear to be of little benefit under such circumstances, we considered that their use in scenarios with weak spatial signals would be questionable. Finally, we were primarily interested in the performance of the model and decision rules on an annual and long term basis, rather than, say, a 'month-to-month' timeframe. In retrospect, the model may have been a little too coarse in its time-step (quarterly) and a little to 'sudden' its movement dynamics between the few areas. This may have caused the relatively sudden, sometimes almost step-wise changes in some quantities rather than relatively smooth changes. This may also be part of the reason for some of the wide fluctuations in abundance and catches seen in some scenarios. It would be interesting to model the fleet and fish movement dynamics in a more realistic and case-specific (ie based on actual data) way than we have done here and to compare results from such an exercise with those we have observed here. The comments below should be seen in the light of these observations.

13.76.1 Comments on the characteristics of the decision tree rule

Over the course of this project, we have become more and more aware of the potential for some decision rules to exhibit oscillatory behaviour involving strong fluctuations in biomass and very high catch variability. When the decision tree is implemented in the spatially explicit simulation model, oscillations in biomass are a feature of many scenarios. Even when there are no obvious oscillations, there is very large variability between realisations. We consider that the oscillatory behaviour in the spatial model is most likely being caused by a mixture of factors, two of which are likely to be: (i) the way CPUE is 'measured' and (ii) the frequency of catch updating.

The fleet dynamics in the spatial model implies that the simulated CPUE data are not from all (or the same set of) areas at each time step. The steep gradient in abundance used in the model means that CPUE is sensitive to which areas are included[38], and the CPUE itself becomes rather variable. The structure of the decision tree is such that it is sensitive to this (e.g. it does not have a built-in limit on the amount of change in the catch from one year to the next, or a maximum catch level as the 40/20 rules do). The simulated data from the spatial model therefore leads to massive adjustments in catch obtained from the decision tree rule.

Although one could argue that we have not 'standardised' the CPUE from the simulation when the fleet moves in and out of areas, one could also argue that, if the fleet is not fishing in a given area, and there are no data from that area, standardisation becomes rather difficult (or impossible), in simulations and in reality! In spite of the fact that the fleet dynamics in the simulation are not based on real data, this is an important observation, particularly if some form of spatial management is envisaged, or if there are strong spatial patterns in fleet behaviour.

Even if the CPUE is simulated as if it is sampled from all areas (including those that were NOT being fished) the decision tree still exhibits some oscillatory behaviour and large variability in

[38] Here the parameterisation of the move-on criteria is also relevant and we are aware that the parameterisation we used is somewhat extreme; see comments in Methods section.

catches. It is likely that the 40/20 rule based on CPUE is much less sensitive to this because it has a built-in maximum catch.

The second issue is the frequency of catch adjustments. The decision tree uses 5 years' CPUE to estimate the slope which determines whether the catch is increased or decreased, and by how much. It is very likely that this quantity is autocorrelated, and that there is a lag between the management action and a response detected in the CPUE. The simulations all involved adjusting the catch annually, which does not give the population a chance to 'signal' increase (for example) before another cut is applied to the catch. This can easily lead to sequences of overshooting and undershooting the target. When combined with the CPUE issue, this effect is likely to be exacerbated.

From these observations we conclude the following. Given the tendency of some decision rules to exhibit oscillations in simulations, and given that this is supported by theory (Section 11), we consider it prudent to always run some realisations for long time-periods, of say 100 years or so, to check for long term oscillations or strange behaviour. Testing should also explore the effects of the frequency of implementation of the decision rule, as well as the number of years of catch rates used in calculating key quantities, such as the slope in CPUE.

In addition to these observations, it is crucial to emphasise the need for caution when looking at results for the decision tree in this Section. First, note that results are only meaningful in a comparative sense (e.g. comparing results for the decision tree with and without a spatial indicator), and not in an absolute sense. The simulation model has not been conditioned to actual catch and effort data, so we cannot draw conclusions about the actual level of depletion or catch it would achieve. Also, the HSWG decision tree has been designed using information from actual catch-effort and size frequency data. This means that the decision tree and the simulation model have not been 'tuned' to one another. The fact that it tends to perform very conservatively should not be seen as an inherent characteristic of the decision rule, but rather due to the lack of tuning to a particular objective in this simulation study.

13.6.2 General Insights from Simulations

Spatial models with fish and fleet movement behave in a complex, sometimes counter-intuitive manner. There is a vast amount of detailed observations that can be made regarding differences in biomass, catch and CPUE patterns between areas under different decision rules, move-on criteria and residency assumptions. These observations are, however, very often case-specific or driven by choice of parameters, for example the CPUE levels in the move-on criteria relative to the CPUE trigger points in the decision rule. In the case of the decision tree rule, observations may even be 'realisation' specific. For these reasons, we have not focused much on the spatial details in results. Such details, may well be very important and relevant in reality, but are best explored in a conditioned simulation model.

The notion of including a spatial indicator (which is meant to reflect when the fleet starts moving offshore due to low catch rates inshore) in a decision rule is to help avoid very low CPUE. Depending on when it is triggered, it could end up only being activated after the CPUE has dropped and the fleet has moved on. In these simulations the fleet dynamics were such that the fleet starts moving on before the CPUE drops to very low levels, and the point at which the spatial indicator is triggered is not necessarily properly matched to the move-on criteria. The simulations therefore do not show a very strong advantage in using a spatial indicator together with a decision rule. In fact, the un-tuned decision rules are so conservative in most cases that there is essentially no need for a spatial indicator.

The inclusion of a spatial indicator can, however, play a useful role in some cases. The simulations show that, if the triggering of the spatial indicator implies a catch cut, the overall biomass is higher (the catch is obviously lower), and the CPUE is generally higher. It is, however, interesting to consider the notion that a spatial indicator could be used to trigger further fleet movement rather than a cut in catches. This can be thought of as a re-distribution of effort and there would not necessarily be a need to close any areas, which would mean that catch rate data can still be collected from that area.

There is a strong link between: (i) the actual definition of 'local depletion' (i.e. at what level of absolute CPUE is it un-economical for the fleet to fish in that area?), (ii) the way in which fleet dynamics are modelled, and the associated parameters, and (iii) the assumptions about fish movement and associated parameters. If the details of these mechanisms are mismatched or incompatible in the specification of a spatial model, the results from simulations are likely to be of limited use. This suggests that the most productive way to explore the behaviour of management strategies in a spatial context is to use a reasonably case-specific model, ideally tuned to actual data, as far as possible. Note that this does not exclude the possibility of exploring uncertainty in any of the components, or the model, but it will help focus on the most realistic or relevant subset of scenarios.

It is not surprising that the performance of decision rules are particularly sensitive to the assumption about stock structure, if that implies the possibility of additional uncontrolled harvesting. Although the decision rules can, to some extent, counter the negative effects of over-exploitation elsewhere, this is usually achieved by cutting back the domestic catch. This observation should be tied to the caveat that, although we considered different levels of residency, we have not considered other more complex movement patterns or 'connectivities' between what one might call the 'domestic' part of the stock and the 'international' part of the stock. The issue of stock structure and additional uncontrolled harvesting elsewhere, is likely to be important when a decision rule is being tuned to achieve a specific objective, for example, with respect to average biomass levels.

In this regard we make the, possibly obvious, comment that the issue of additional uncontrolled harvesting can easily be incorporated into a non-spatial model. The tuning of a decision rule taking this into account does therefore not have to wait until a full spatial model has been set up and parameterised (conditioned) for a specific case.

The limited explorations considered here, do not point to a strong difference between (or preference for) the 40/20 decision rule using CPUE and that using mean weight. For example, in broad terms, they respond to alternative assumptions in a very similar way. The life-history parameters and low levels of noise and measurement error used in these scenarios do, however, fall into the area of parameter space where results from Sections 8 and 9 suggest that this is likely to be the case. We would definitely caution against extrapolating from these results to other life-histories or levels of variability.

We only tested one decision rule based on multiple indicators, namely the decision tree rule. As noted in Section 11, the decision tree does not simply combine results from indicators, but follows a sequence of further reductions to an initial catch level depending on the values of indicators relative to target points. As already noted, this rule appears to be particularly sensitive to the implications of a spatial model on the variability of CPUE. Further investigations would be required to determine which factors play the largest role in this behaviour.

13.7. Annex of additional Figures to Section 13.

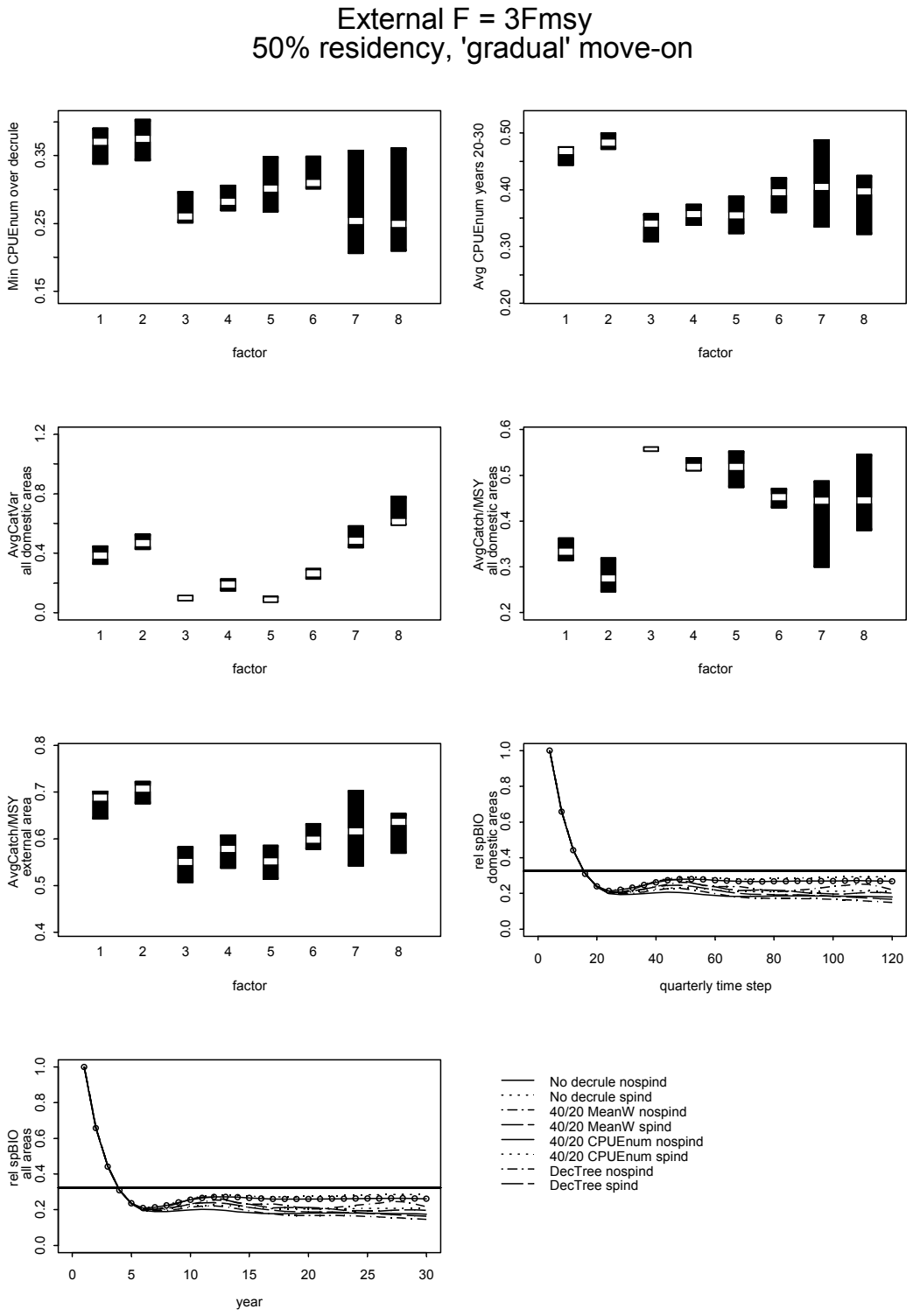


Figure A 18. Scenario: 50% residency, gradual move-on criteria, additional uncontrolled harvest. Boxplots of catch rates (minimum and average over years 20-30) in the inshore area, and catches (variability and total, averaged over the decision rule period) in all areas, and time series of the average relative biomass in all 'domestic' areas, and all areas (including the 'outside' area) over 10 realisations. The 'factor' number identifies the decision rule used, as indicated in the legend (factor numbers are 1:8 from top to bottom).

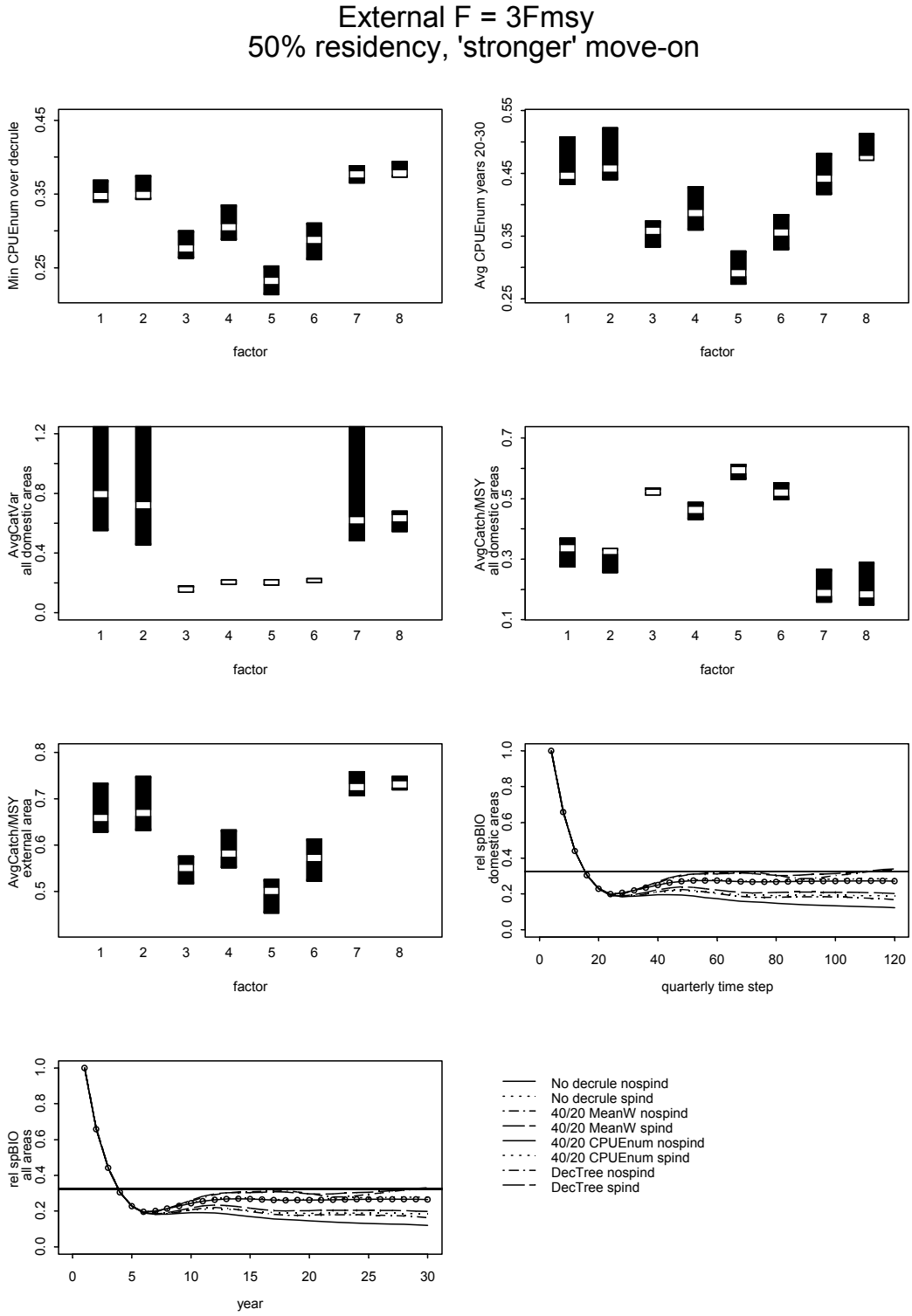


Figure A 19. Scenario: 50% residency, stronger move-on criteria, additional uncontrolled harvest. Boxplots of catch rates (minimum and average over years 20-30) in the inshore area, and catches (variability and total, averaged over the decision rule period) in all areas, and time series of the average relative biomass in all 'domestic' areas, and all areas (including the 'outside' area) over 10 realisations. The 'factor' number identifies the decision rule used, as indicated in the legend (factor numbers are 1:8 from top to bottom).

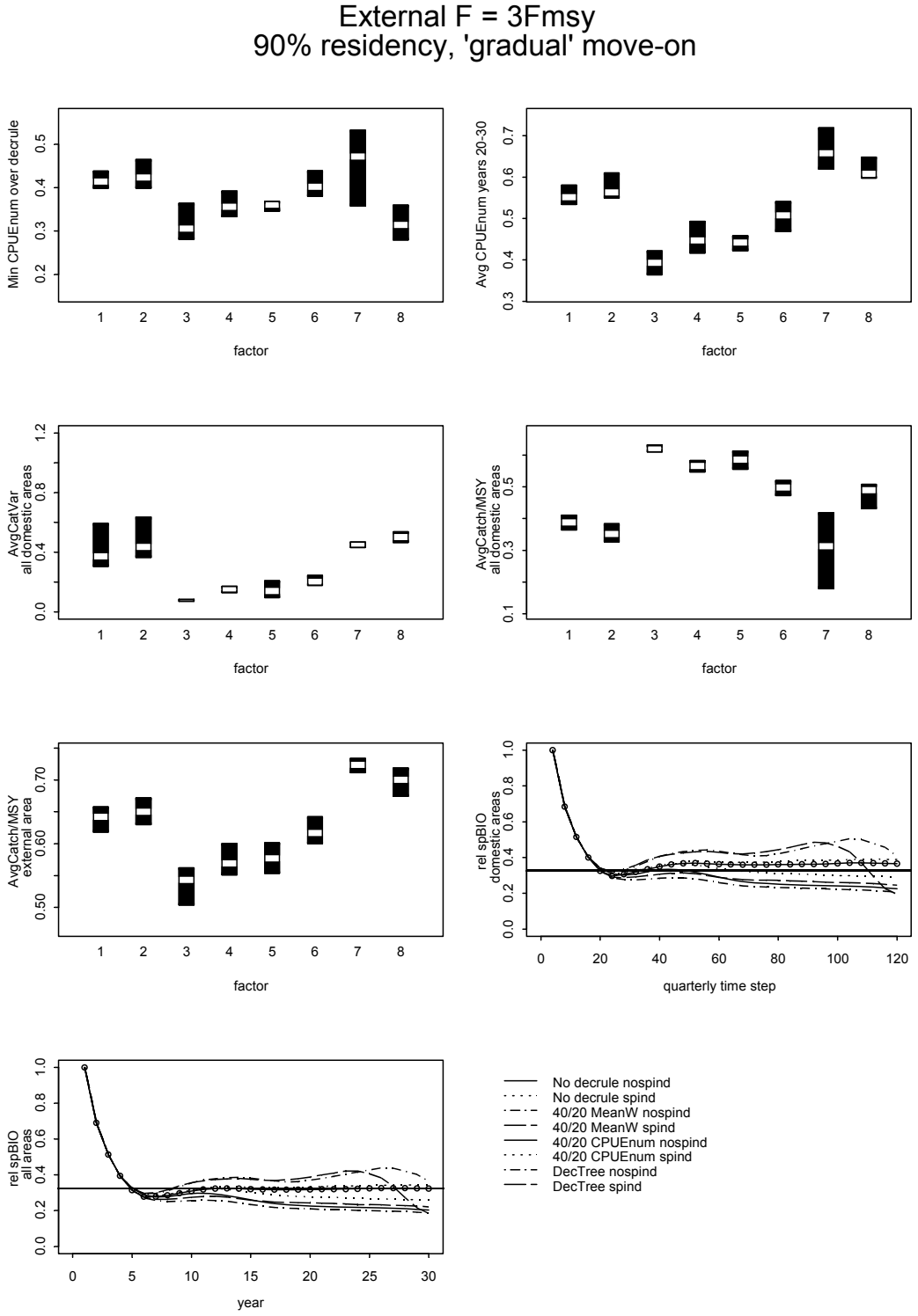


Figure A 20. Scenario: 90% residency, gradual move-on criteria, additional uncontrolled harvest. Boxplots of catch rates (minimum and average over years 20-30) in the inshore area, and catches (variability and total, averaged over the decision rule period) in all areas, and time series of the average relative biomass in all 'domestic' areas, and all areas (including the 'outside' area) over 10 realisations. The 'factor' number identifies the decision rule used, as indicated in the legend (factor numbers are 1:8 from top to bottom).

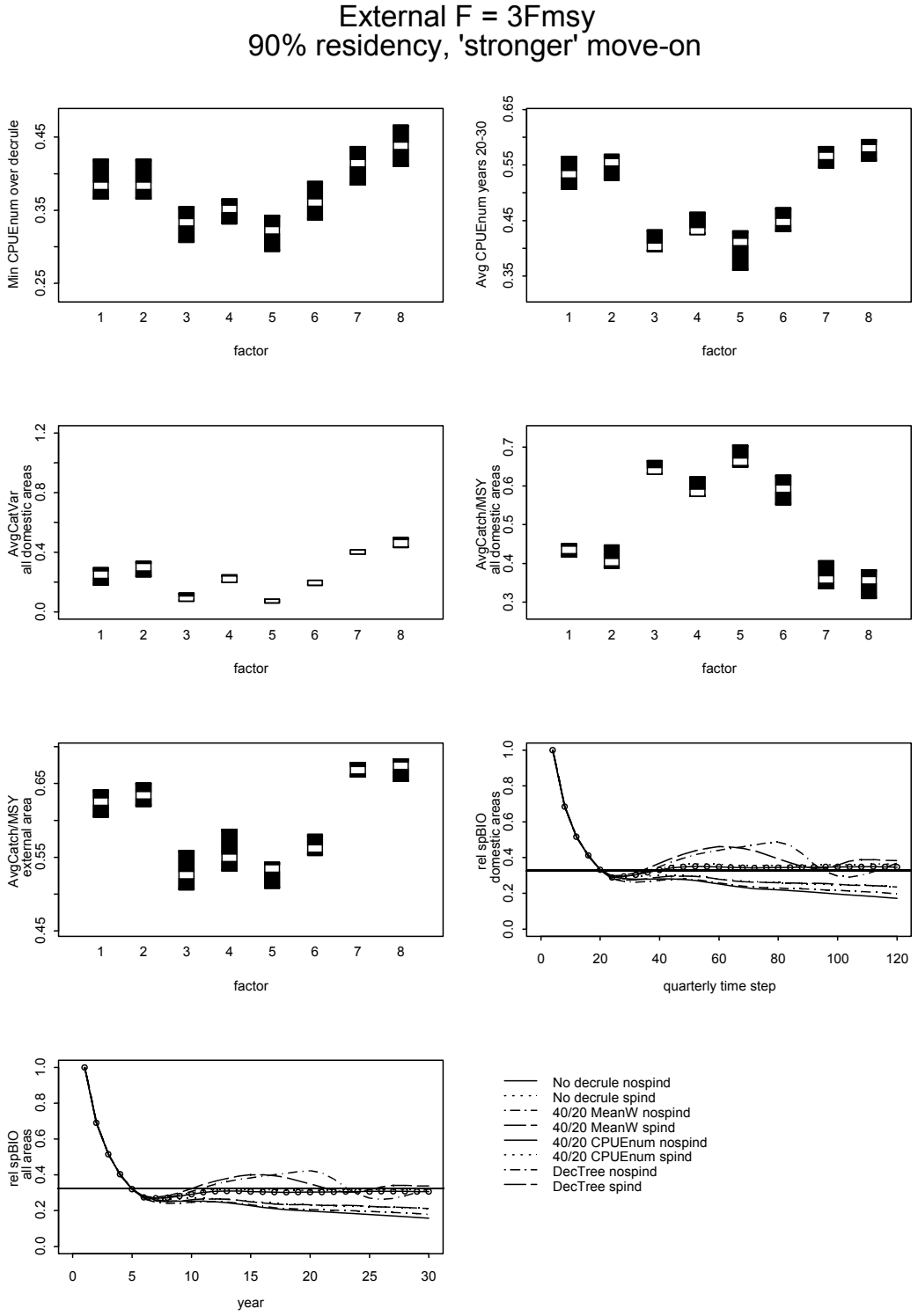


Figure A 21. Scenario: 90% residency, stronger move-on criteria, additional uncontrolled harvest. Boxplots of catch rates (minimum and average over years 20-30) in the inshore area, and catches (variability and total, averaged over the decision rule period) in all areas, and time series of the average relative biomass in all 'domestic' areas, and all areas (including the 'outside' area) over 10 realisations. The 'factor' number identifies the decision rule used, as indicated in the legend (factor numbers are 1:8 from top to bottom).

Sensitivity trials – additional figures

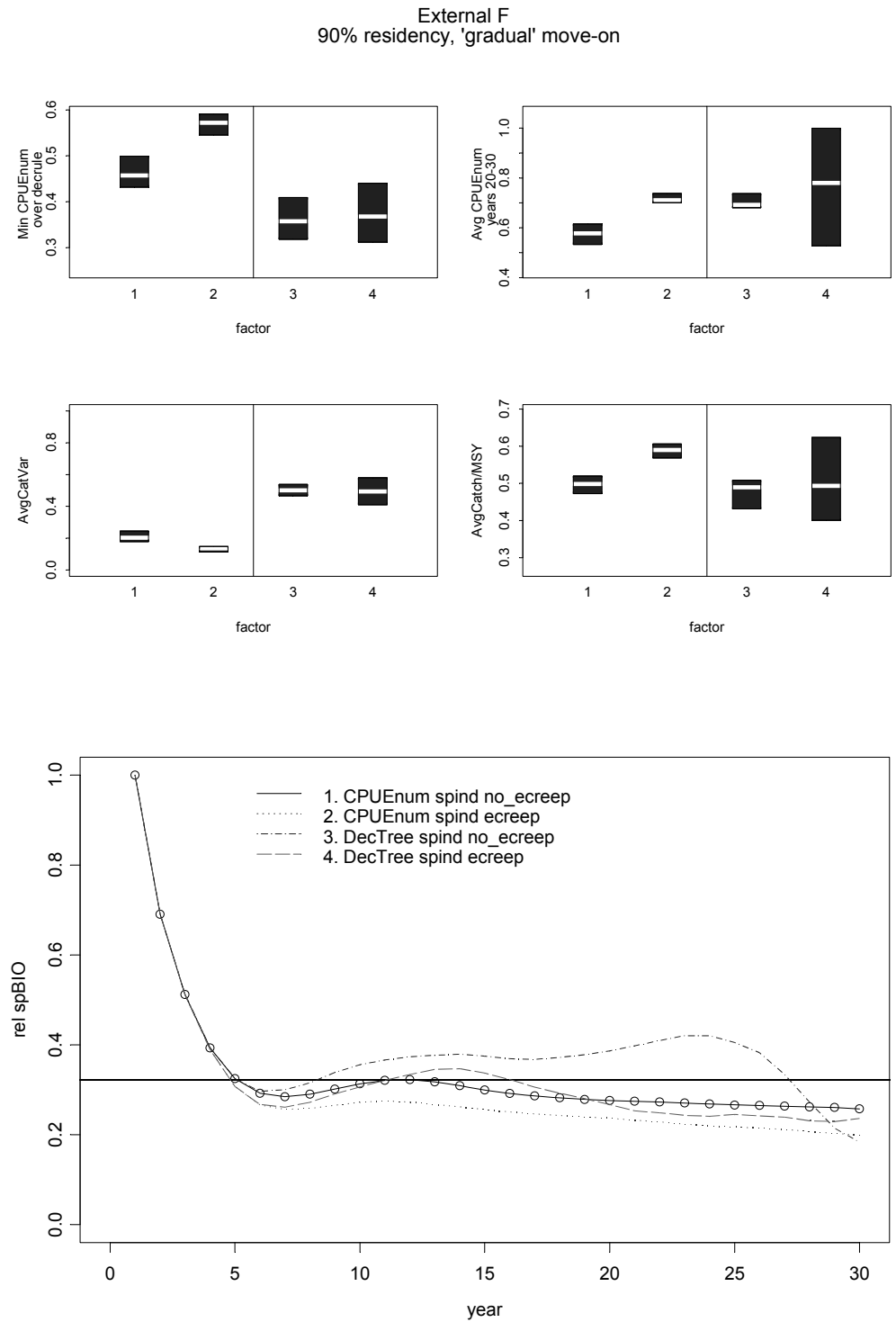


Figure A 22. Effort Creep. Scenario: 90% residency, gradual move-on criteria, with spatial indicator and additional uncontrolled harvest. Boxplots of catch rates (minimum and average over years 20-30) in the inshore area, and catches (variability and total, averaged over the decision rule period) in all domestic areas, over 10 realisations. Results are plotted in pairs of scenarios without and with effort creep for the following decision rules: 40/20 CPUE rule with spatial indicator (1,2); the decision tree rule with spatial indicator (3,4). The bottom panel shows relative spawning biomass in all areas over 10 realisations.

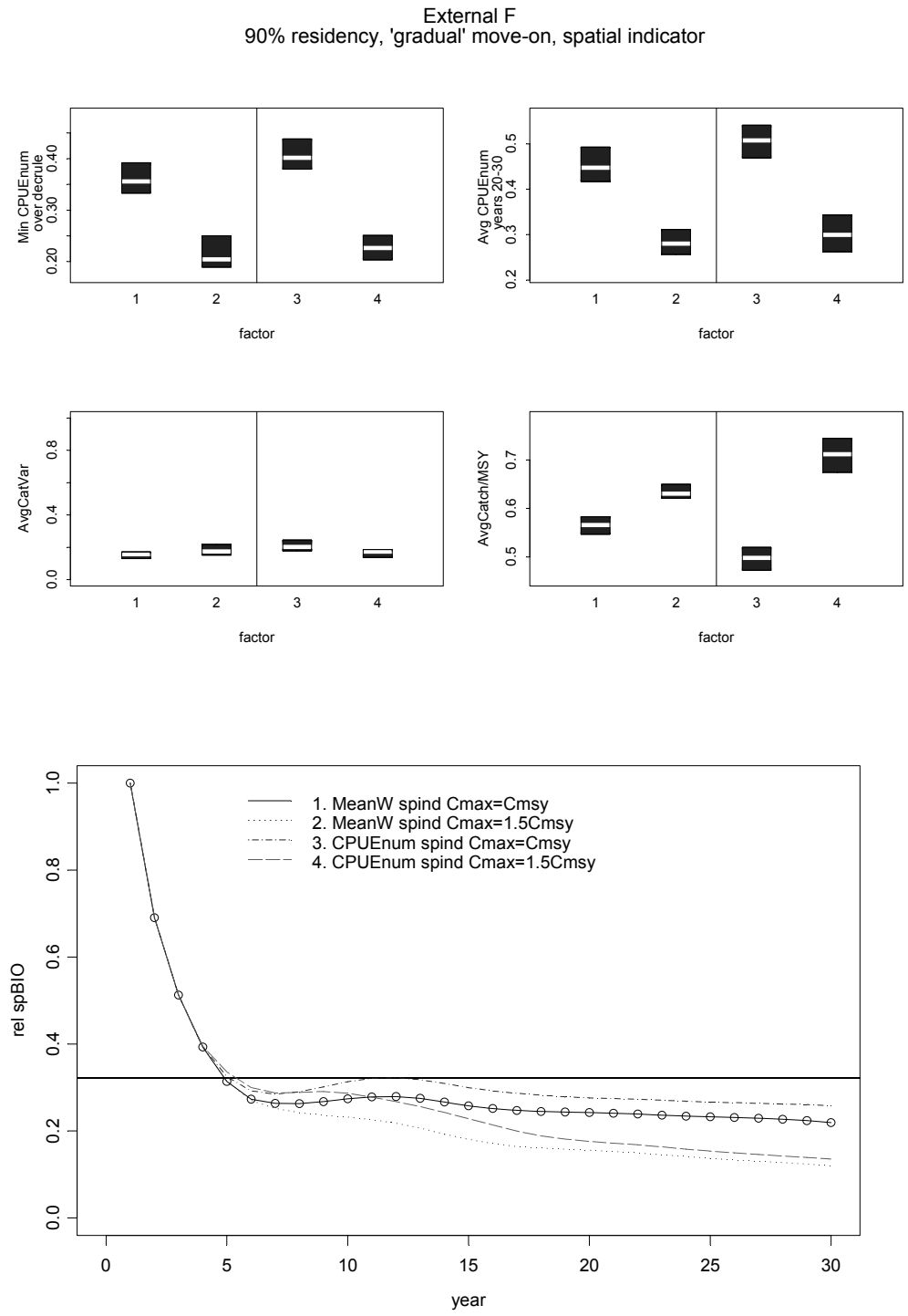


Figure A 23. C_{max} too high. Scenario: 90% residency, gradual move-on criteria, with spatial indicator and additional uncontrolled harvest. Boxplots of catch rates (minimum and average over years 20-30) in the inshore area, and catches (variability and total, averaged over the decision rule period) in domestic areas, over 10 realisations. Results are plotted in pairs of scenarios with [C_{max}=C_{msy}]and [C_{max}=1.5C_{msy}] for the following decision rules: 40/20 mean weight rule with spatial indicator (1,2); 40/20 CPUE rule with spatial indicator (3,4). The bottom panel shows relative spawning biomass in all areas over 10 realisations.

14. Benefits and adoption

Two working papers on approaches for standardisation of size-based indicators, and analyses of yellowfin tuna, bigeye tuna and swordfish size frequency were presented at IOTC Working party meetings. Although the results did not suggest that standardisation was essential for the datasets we considered, it lay the groundwork for subsequent analyses that can be undertaken as more data are collected, and for other species.

The Policy documents that accompanied, and followed, the Ministerial directive (Australian Government, 2007) set the reference points for biomass and harvest rates and this issue was therefore taken up by the Harvest strategy working group (HSWG). The choice of biomass and harvest rate reference points, as well as associated trigger points for the indicators, was agreed in the context of the HSWG. Both authors of this report were members of the group and provided input from results in this project. For example, the uncertainty about steepness in the stock-recruit relationship and the advantages of reference points based on spawner per recruit considerations rather than maximum sustainable yield considerations were taken on board.

We would, however, like to explicitly acknowledge that the work done in the HSWG was a joint effort by all members, and many of the points mentioned here were in fact shared views and experience. In addition, work undertaken in the HSWG, inspired some of the subsequent, and parallel work conducted under this project[39].

The Harvest strategy working group has developed a decision rule, primarily for the ETBF, and has conducted preliminary testing of the decision rule. The design of the rule was influenced by the outcomes from this project that were available at the time. For example, CPUE of a certain size range of fish forms the top node of the decision tree. One reason for this was concern about the fact that size-based indicators (mean weight, for example) are not responsive if steepness is low, and/or growth rate is high.

The first decision point in the tree, i.e. whether catch should be decreased or can be increased, is not just based on the slope of CPUE. It is also based on the level of CPUE and a built-in target CPUE. This stems from observations from simulations and theoretical considerations in this project, which showed that a decision rule which simply responds to a change in an indicator is unlikely to perform well.

The decision tree uses multiple indicators; a suite of CPUE and size-based indicators. Although the simulations did not clearly show substantially better performance by groups of indicators, our concerns that this result may be due to the simulations not including sufficiently poor, or erratic, behaviour by indicators meant that multiple indicators were retained.

The preliminary testing could not have been undertaken as quickly as it did, had it not been for the general simulation model developed under this project. The decision tree is complicated enough that it is not easy to see how it will behave under different circumstances. Early simulation testing was very useful for checking that it did behave as intended, and where necessary, the rule was modified. The simulation testing also greatly assisted the choice of parameters, at least for the first version of the decision rule.

[39] For example, the development of a decision tree in the HSWG, inspired the thought of fitting classification trees to simulated data, as a tool for estimating prediction errors of groups of indicators and as a potential decision rule.

Further testing of the decision tree using the same simulation model (and code developed in this project), but with additional code to allow conditioning of the model to actual historic data is now being undertaken (under an AFMA funded project).

General results and observations from the exploratory runs with the spatially explicit version of the simulation model and the HSWG decision tree, are being provided to the project team doing further testing. These results highlight some of the types of investigations which should be conducted to ensure robustness of the decision tree.

The direct benefits to AFMA and the Western and the Eastern tuna and billfish fisheries (WTBF and ETBF) are the development and even initial testing of Harvest Strategies which were subsequently agreed and adopted. We again acknowledge that this was achieved jointly through this project and the work of the Harvest Strategy Working Group.

15. Further Development

We intend to publish several peer-reviewed papers based on the work in this report in order to reach a wider fisheries science audience. We also intend to disseminate results more broadly to international regional fisheries management organisations (RFMO's) of tunas and billfish, particularly the IOTC given its interest in this work from the start.

The groundwork laid in this project will also form the basis for further explorations of harvest rate indicators (obtained from tagging data, for example) as a basis for decision rules and management procedures.

16. Planned Outcomes

One of the planned outcomes of this project was that results would provide guidance on the kinds and combinations of indicators, the frameworks for combination, together with reference points and decision rules that are most likely to be robust for management of the various large pelagic tunas and billfish in the WTBF and the ETBF. This was achieved through the inputs from this project to the development of harvest strategies for the ETBF and WTBF, conducted by the Harvest Strategy working group (HSWG) of which both authors of this report were members.

A second planned outcome was the improvement of procedures for setting and adjusting TACs/TAEs in order to meet management objectives within the fisheries' management plans, in other words, improved decision rules. This was again achieved through input to the HSWG and the availability of a simulation model which could be used for direct testing and evaluation of candidate decision rules. These quantitative evaluations helped in refining the details and parameter values of decision rule. In addition, because they could illustrate the likely performance of the decision rules, we consider that they played a significant role in assisting the acceptance and adoption of the harvest strategy.

The major findings of the project should assist the IOTC in developing approaches for monitor stock status and managing catches (or effort) in the Indian Ocean context, particularly where more sophisticated assessments are infeasible, or highly uncertain.

17. Conclusion

All the objectives of this project have been completed. A set of potential stock status indicators were identified. Approaches for the standardisation of time-series of size-based indicators were illustrated using size frequency data from bigeye tuna, yellowfin tuna and broadbill swordfish catches taken in the Indian Ocean.

A spatial simulation model was developed by modifying an existing population model. Initial explorations were conducted using a non-spatial version of the model to explore the characteristics of the various indicators. The responsiveness of the indicators were compared and tested under different assumptions about the life-history of the fish stock involved, to provide guidance for a wider range of species than just bigeye and yellowfin tuna, and swordfish.

We developed a novel framework for using multiple indicators by fitting classification trees to simulated time-series of indicators and using these trees as decision rules. In addition, we used the well-known 'traffic light' framework for combining indicators. Indicators were tested, both singly, and in these frameworks, against reference points to evaluate their effectiveness in predicting whether spawning biomass was above or below a given reference point.

Decision rules were then also tested in a feedback simulation set-up where either catch or effort was adjusted according to indicator values. This highlighted the importance of the design, or form, of the decision rule. A theoretical investigation was used to further explore the characteristics of decision rules.

Actual reference points and the details of the decision rule that has been developed for the ETBF and WTBF took place within the Harvest strategy working group (HSWG) rather than solely within this project as had initially been planned. However, as noted above, the results from this project played an important and valuable role in the work of the HSWG.

In summary, the results from this project suggest the following main conclusions (also see the Non-Technical summary and Section Summaries):

1. The need for standardisation should be considered before an indicator is used to monitor and manage a fish stock.
2. CPUE is potentially the most responsive indicator (since it changes in proportion to the stock size) and is also more robust to uncertainty in the stock-recruit relationship than size-based indicators. This emphasises the importance of obtaining reliable catch-effort data, and additional data on fishing operations for standardisation purposes. Obtaining fishery independent indices of relative abundance are even more important to strive for.
3. Size-based indicators are much less responsive (change by far less than a given change in the population), and are therefore much more sensitive to measurement error or random noise, e.g. due to recruitment variability.
4. Results strongly suggest that size-based indicators, in terms of length or weight, are only likely to be informative for populations where individual growth is slow, and of all the size-based indicators considered, the means (of length or weight) generally perform best.
6. When groups of indicators are used, there is little to be gained by using more than about 4 to 5 indicators together; as few as 2 or 3 could be sufficient.
7. Decision rules based on groups of indicators did not generally outperform those based on the best single indicator in the group, but we consider it risky to over-interpret this result. The benefit of multiple indicators may only really come into play when indicator time series are biased or prone to serious unpredictable 'failures'.
8. Even good indicators can perform poorly when used in a decision rule that is badly designed; even very simple decision rules can have bad characteristics. Theoretical or simulation

approaches should be used to ensure that a decision rule does not induce large-scale fluctuations or oscillations in abundance and catches.

9. Experience with a relatively general spatial model of fish and fleet dynamics suggests that it may be more productive to use a case-specific, data driven model, particularly for the fleet dynamics, to explore the implications of and solutions to local depletion.

18. References

Australian Government. 2007. Commonwealth Fisheries Harvest Strategy Policy Guidelines. Australian Government Department of Agriculture, Fisheries and Forestry, Canberra, Australia, 55 pp.

Basson, M and M.J. Fogarty. 1997. Harvesting in discrete-time predator-prey systems. Mathematical Biosciences 141:41-74.

Beddington, J.R. 1978. On the risks associated with different harvesting strategies. Rep. Int.Whal. Commn 28; SC/29/Doc 6;165-167

Bergh, M.P. and Butterworth, D.S. 1987. Towards rational harvesting of South African anchovy considering survey imprecision and recruitment variability. South African Journal of Marine Science, 5: 937-951.

Berkeley, S.A. and Houde, E.D. (1983). Age determination of broadbill swordfish, *Xiphias gladius*, from the Straits of Florida, using anal fin spline sections. U.S. Nat. Mar. Fish. Serv., NOAA Technical Report. NMFS 8: 37-143.

Butterworth, D.S. and Geromont, H.F. 1997. Evaluation of a range of possible simple interim management procedures of the Namibian hake fishery. Report to the Ministry of Fisheries and Marine Resources, Namibia. 28pp.

Butterworth, D.S. and M.Mori. 2003. Some initial investigations of possible Management Procedures for SBT based upon age-aggregated production models. CCSBT 2nd Management Procedure Workshop, document CCSBT-MP/0304/12.

Caddy, J.F. 2002. Limit reference points, traffic lights, and holistic approaches to fisheries management with minimal stock assessment input. Fisheries Research 56(2) 133-137.

Campbell, R., M.Basson and N.Dowling. 2003. Review and analysis of information for determination of Total Allowable Catches and decision rules for the Southern and Western Tuna and Billfish Fishery. Compilation of related project papers. Final report to AFMA.

Campbell, R.A. and N. Dowling. 2003. Development of an operating model and evaluation of harvest strategies for the eastern tuna and billfish fishery. Final report for project 1999/107, Fisheries Research Development Corporation, Canberra, Australia.

Campbell, R and Hobday, A. 2003. Swordfish-Environment-Seamount-Fishery Interactions off eastern Australia. Report to the Australian Fisheries Management Authority, Canberra, Australia. 97pp.

Campbell, R., Davies, C., Prince, J., Dowling, N., Kolody, D., Basson, M., McLoughlin, K., Ward, P., Freeman, I., and A. Bodsworth. 2007. Development and preliminary testing of harvest strategies for the Eastern and Western Tuna and Billfish Fisheries. Draft report to the Australian Fisheries Management Authority.

Chambers, J.M. and Hastie, T.J. 1993. Statistical Models in S (Edited by Chambers and Hastie). Chapman and Hall Computer Science Series, Chapman and Hall, New York.

CCSBT, 2003. Report of the Fourth Meeting of the Stock Assessment Group. 25-29 August 2003, Christchurch, New Zealand. 36pp.

CCSBT, 2004. Report of the Fifth Meeting of the Stock Assessment Group. 6-11 September 2004, Seogwipo City, Jeju, Republic of Korea. 62pp.

Davies, Nick., Robert Campbell, Dale Kolody. 2006. CASAL Stock assessment for South-West Pacific broadbill swordfish 1952-2004. Working paper to the Western Central Pacific Fisheries Commission, WCPFC-SC2-2006/ME WP-4.

Davies, C., Campbell, R., Prince, J., Dowling, N., Kolody, D., Basson, M., McLoughlin, K., Ward, P., Freeman, I., Bodsworth, A., 2008. Development and preliminary testing of the harvest strategy framework for the Eastern and Western Tuna and Billfish Fisheries. Report to the Australian Fisheries Management Authority, Canberra.

De la Mare, W.K. 1986. Simulation studies on management procedures. Thirty Sixth Report of the International Whaling Commission, Document SC/37/O14.

Ehrhardt, N.M., Robbins, R.J. and Arocha, F. (1996). Age validation and growth of swordfish, Xiphias gladius, in the northwestern Atlantic. ACCAT (International Commission for the Conservation of Tunas) Col. Vol. Sci. Pap. 45(2): 383-367.

Ehrhardt, N.M. (1999) Review of age and growth of swordfish, Xiphias gladius, using methods other than otoliths session. Background paper presented at the Second International Pacific Swordfish Symposium, Kahuku, Hawaii, 3–6 March 1997.

Fonteneau, A. and Gaertner, D. (2002). Analysis of trend of total yearly catches of yellowfin tuna in the Indian Ocean and status of stock. Working Paper WPTT-02-17; IOTC Proceedings (5) 2002 p 310-31

Gaertner, D., Fonteneau, A. and Laloe, F. (2001). Approximate estimate of the maximum sustainable yield from catch data without detailed effort information: application to tuna fisheries. Aquatic Living Resources 14: 1-9

Getz, W.M. and R.G. Haight. Population harvesting: demographic models of fish, forest and animal resources. Monographs in Population Biology 27, Princeton University Press, 391pp.

Grall, C. De Sylva D.P. and Houde, E.D. (1983) Distribution , relative abundance, and seasonality of swordfish larvae. Trans. AM. Fish. Soc. 112: 235-246.

Hassell, M.P. and H.N. Comins. 1976. Discrete time models for two-species competition. Theoretical Population Biology 9: 202-221.

ICCAT. 2006a. SCRS/2006/010 Report of the 2006 ICCAT Swordfish stock structure workshop. Heraklion, Crete, Greece, 13-15 March, 2006.

ICCAT. 2006b. SCRS/2006/015 Report of the 2006 Atlantic swordfish stock assessment**.** Madrid, September 4 to 8, 2006. 66pp.

ICCAT. 2007. SCRS/2007/016 2007 Mediterranean swordfish stock assessment session. Madrid, Spain - September 3 to 7, 2007. 55pp.

IOTC Commission Resolution 07/01. 2007. See www.iotc.org: Management and Conservation measures adopted at the 11th Session of the Commission (texts of the Resolutions adopted by the Commission in May 2007).

IOTC. 2004. Report of the 4th Session of the IOTC Working Party on Billfish. Mauritius, 27 September – 1 October 2004 (IOTC-2004-WPB-R). 36pp.

IOTC Scientific Committee. 2002. Report of the Fifth Session of the Scientific Committee. IOTC-S7-02-03.

Lu, Ching-Ping., Chaolun Allen Chen, Cho-Fat Hui, Tzong-Der Tzeng, and Shean-Ya Yeh (2006) Population genetic structure of the swordfish, *Xiphias gladius* (Linnaeus, 1758), in the Indian Ocean and West Pacific inferred from the complete DNA sequence of the mitochondrial control region. Zoological Studies **45**(2): 269-279.

Macías, D., A. Hattour, J.M. de la Serna, M.J. Gómez-Vives, and D. Godoy. 2005. Reproductive characteristics of swordfish (*Xiphias gladius*) caught in the southwestern Mediterranean during 2003. SCRS/2004/068 Col. Vol. Sci. Pap. ICCAT, 58(2): 454-469

May, R.M. 1974. Stability and Complexity in Model Ecosystems. Monographs in Population Biology 6, Princeton University Press, 265pp.

May, R.M. and G.F. Oster. 1976. Bifurcations and Dynamic Complexity in Simple Ecological Models. The American Naturalist. Vol 110, No. 974: 573-599.

Montiel, A.V. (1996). Determinación de la edad y del crecimiento del pez espada (*Xiphias gladius*) explotado frente a las costas de Chile. Fisheries Engineering Thesis. Catholic Univ. Valparaiso, Chile. 56p.

Palko, B.J., Beardsley, G.L., Richards, W.J. (1981) Synopsis of the biology of the swordfish, *Xiphias gladius* Linnaeus. NOAA Technical Report, NMFS Cir. 441, pp1-21.

Payne, A.I.L. (Guest Editor). 1999. Confronting Uncertainty in the Evaluation and Implementation of Fisheries-Management Systems. Proceedings of an ICES Symposium held in Cape Town, South Africa, 16-19 November 1998.

Polacheck, T., D. Ricard, P. Eveson, M. Basson and D. Kolody. 2003. Results from initial testing of some candidate management procedures for Southern Bluefin Tuna. CCSBT 2nd Management Procedure Workshop, document CCSBT-MP/0304/06

Punt, A.E. 1992. Selecting management methodologies for marine resources, with an illustration for southern African hake. South African Journal of Marine Science, 12:943-958.

Punt, A.E., R.A. Campbell, A.D.M. Smith. 1999. Evaluation of Performance Indicators in the Eastern Tuna and Billfish Fishery - A preliminary study. Final Report to AFMA.

Radtke, R.L. and Hurley, P.C.F. (1983). Age estimation and growth of broadbill swordfish, *Xiphias gladius*, from the northwest Atlantic based on external features of otoliths. U.S. Dep. Commer., NOAA Tech. Rep. NMFS 8:145-150

Reeb, C.A., Arcangeli, L. and Block, B.A. (2000) Structure and migration corridors in Pacific populations of the swordfish, Xiphias gladius, as inferred through analysis of mitochondrial

DNA. Working paper BBRG-13 presented at the 13th meeting of the Standing Committee on Tuna and Billfish, held 5-12 July 2000, Noumea, New Caledonia.

Rochet, M. and Trenkel, V.M. (2003). Which community indicators can measure the impact of fishing? A review and proposals. Canadian Journal of Fisheries and Aquatic Sciences 60: 86-99

Sun, C-L., Wang, S-P. and Yeh., S-Z. (2000) Note on Taiwan water's swordfish spawning as indicated by gonad indices. Working paper BBRG-9 presented at the 13th meeting of the Standing Committee on Tuna and Billfish, held 5-12 July 2000, Noumea, New Caledonia.

Sun, C-L, Wang, S-P. and Yeh, S-Z. (2002) Age and growth of the swordfish (*Xiphias gladius* L.) in the waters around Taiwan determined from anal-fin rays. Fishery Bulletin 100(4): 822-835.

Tserpes, G. and Tsimenides, N. (1995). Determination of age and growth of swordfish, *gladius* L., 1758, in the eastern Mediterranean using anal-fin spines. Fishery Bulletin 93: 594-602.

Vanpouille, K., Poisson, R., Taquet, M., Ogor, A. and Troadec, H. (2001). Etude de la croissance de l'espadon. *In* 'L'espadon: de la recherché a l'exploitation durable'. (eds. Poisson, F. and Taquet, M.) pp 170-211. Programme Palangre Reunionnais, Rapport finale.

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

Ward, R.D., Reeb, C.A. and Block, B.A. (2001) Population structure of Australian swordfish, *Xiphias gladius*. Final Report to the Australian Fisheries Management Authority, Canberra. 39pp.

Young, J.W. and Drake, A. (2002). Reproductive dynamics of broadbill swordfish (*Xiphias gladius*) in the domestic longline fishery off eastern Australia. Final report to the Fisheries Research and Development Corporation Project No. 1999/108.

Young, J.W. and Drake, A. (2004). Age and growth of broadbill swordfish (*Xiphias gladius*) from Australian waters. Final report to the Fisheries Research and Development Corporation Project No. 2001/014.

19. Intellectual Property

No commercial intellectual property arose from this work.

20. Staff

Marinelle Basson
Natalie A. Dowling

21. Appendix 1: Technical description of the operating model

This appendix provides the technical details of the operating models, the stock status indicators and the form of harvest strategies used across the various simulations.

Different aspects of this project used slightly different simulation models, which range from a simple single gender, single area, age-structured model to a spatially explicit, gender-specific, age-structured model. Here we describe and give equations for the general full model, noting that the simpler versions use the same equations, but without the relevant subscripts (e.g. for area or gender) and associated parameters. Additional specific details of each model are given in each of the chapters. Details about parameter values and scenarios are also primarily discussed under each of the relevant chapters rather than here.

The full operating model is an age-structured, life-history driven model, designed to characterise a harvested fish species; the specified inputs define whether it is, for example, swordfish-like or yellowfin-like. It is based on a single stock and explicitly considers the age and sex structure of the population. The model has quarterly time steps, with recruitment occurring in quarters 1 and 2. The model assumes that natural and fishing mortality occur at the start of each quarter. The population and fleet may move among multiple areas. This is modelled to occur at the end of each quarter. The spatial model has primarily been constructed to explore issues of local depletion in swordfish, and the most basic spatial representation essentially only includes the notional area where Australian vessels fish. There is also, however, the option to define an 'outside' area adjacent to the furthest offshore area where Australian vessels fish, to represent additional "external" fishing mortality from foreign fleets.

1. General population dynamics

The population dynamics of the full spatial model are essentially an expansion of the very familiar one-area population equation for non-recruits; the numbers at age a+1 in year t+1 are the survivors from age a-1 in year t-1, where the total (natural and fishing) mortality on that age class is $Z_{t,a}$:

$$N_{t+1,a+1} = N_{t,a}.\mathrm{e}^{(-Z_{t,a})} \qquad (\mathbf{1})$$

For the spatial version of the model, we keep track of the numbers at age (and gender) in each area, so the movement dynamics of the population between areas needs to be contained in the population equations. We assume that, at each time step, a proportion r of the numbers in a given area will stay in that area, and that the remainder (1-r) will move to another area according to defined area-specific proportions, n_A. The population numbers at time t+1, age a+1, gender g, in area A is given by equation (**2**):

$$N_{t+1,a+1,g,A} = \begin{cases} r \cdot N_{t,0,g,A} + (1-r) n_A N_{t,0,g,A} & a = 0 \\ r \cdot N_{t,a,g,A} e^{-Z_{t,a,g,A}} + (1-r) n_A N_{t,a,g,A} e^{-Z_{t,,a,g,A}} & 1 \le a < x \\ r \cdot (N_{t,x-1,g,A} e^{-Z_{t,x-1,g,A}} + N_{t,x,g,A} e^{-Z_{t,x,g,A}}) + (1-r) n_A \cdot (N_{t,x-1,g,A} e^{-Z_{t,x-1,g,A}} + N_{t,x,g,A} e^{-Z_{t,x,g,A}}) & a = x \end{cases}$$

where
r is the resident proportion
n_A is the proportion of the moving stock assigned to area A
x is the maximum age (above which all fish are combined as a "plus group") chosen so that selectivity, fecundity and natural mortality for ages x and greater can reasonably be assumed to be the same
Z is the total (natural and fishing) mortality at age a and time t of gender g in area A (see under Fishery section below for more detail).

Movement, as described by the proportions, n_A, of the moving stock assigned to each area may be: i) uniform (i.e. the same proportion to each area), or ii) pre-specified proportions (typically an offshore gradient of high to low density). In the second case, these proportions can be thought of as quantities that represent some kind of habitat preference, for example. If the residency parameter r=1, there is no movement, and in this case the proportions are set to zero. This version of the model is only implemented when assuming a single area.

Recruitment

Recruitment occurs in the first two quarters of the year and is calculated for each quarter using a Beverton-Holt stock-recruit relationship with lognormal error. Total recruitment in a year is therefore $N_{1,0,g} + N_{2,0,g}$ (note that, this sum does not strictly equate to a Beverton-Holt relationship, but it is close if the spawning biomass does not change much between the two quarters). The total number of recruits of each gender, g, in any time step t is given by equation (**3**):

$$N_{t,0,g} = K_g \alpha B_t (\beta + B_t)^{-1} e^{\varepsilon_t - \sigma_R^2/2} \qquad \varepsilon_t \sim N(0;\sigma_R^2) \qquad \text{for quarter=1 or 2}$$
$$= 0 \qquad \text{for quarter=3 or 4}$$

where
α, β are the parameters of the stock-recruitment relationship,
B_t is the spawning biomass at time (quarter) t
K_g is the fraction of recruits assigned to gender group g (for an equal proportion of males and females, $K_1 = K_2 = 0.5$)
σ_R is the standard deviation of the logarithm of the fluctuations in recruitment (note that the error term can be omitted to yield deterministic recruitment).

It is convenient to rewrite the stock-recruit relationship in terms of steepness[40], h, unexploited spawning biomass, B_0 and recruitment of the unexploited stock, R_0, so that

$$\alpha = R_0 \frac{(\beta + B_0)}{B_0} \qquad (\mathbf{4a})$$

$$\beta = \frac{B_0(1-h)}{5h-1} \qquad (\mathbf{4b})$$

[40] Steepness is defined as the ratio between recruitment (R) when spawning biomass is at 20% of unexploited spawning biomass (B_0), and recruitment at the unexploited spawning biomass i.e. $h = R(0.2B_0) / R(B_0)$.

We consider this further in the section on running the simulations.

Recruitment is assigned to each area using the same area-specific proportions, n_A so that the number of recruits of gender g in area A at time t is given by:

$$N_{t,0,A,g} = n_A.N_{t,0,g}$$

The spawning biomass, B_t , is calculated as the sum of the weight of all mature individuals (males and females, g=1 and g=2, in this case):

$$B_t = \sum_{a=1}^{x}\sum_{g=1}^{2} p_{a,g} w_{a,g} N_{t,a,g} \quad (\mathbf{5})$$

where

$$N_{t,a,g} = \sum_{A=1}^{y} N_{t,a,g,A} \quad (\mathbf{6})$$

and

y is the number of areas
$p_{a,g}$ is the proportion of the population at age a, of gender g, that is mature, and
$w_{a,g}$ is the body weight of an individual of age a, gender g
maturity-at-age is assumed to be knife-edged at age $t_{m,g}$.

Growth

The mean body weight of individuals of each gender and in each age class is obtained from a von Bertalanffy growth curve and a length-weight relationship. The size of an individual in gender group g , age a is:

$$L_{a,g} = L_\infty^g (1 - e^{-k^g (a - t_0^g)}) \quad (\mathbf{7})$$

Note that the von-Bertalanffy parameters, L_∞ ,k, t_0, can be gender-specific (indicated by the superscript 'g'), allowing us to explore the effects of sexual dimorphism in growth. This equation gives the mean length at age. We discuss the sampling of the catch and its associated size-frequency distribution, which consists of a mixed normal distribution, below under 'General fishery dynamics'.

The mean weight-at-age was derived from a power relationship to mean length-at-age:

$$w_{a,g} = a(L_{a,g})^b \quad (\mathbf{8})$$

Assumptions about values for parameters are discussed further below and under the descriptions of specific scenarios.

2. General fishery dynamics

As hinted above, the spatial model was constructed to explore the potential issue of local depletion and the performance of spatial indicators in decision rules. The relatively simple model

used here has notional areas (i.e. not actual geographic areas) meant to mimic 3 longitudinal bands starting with an inshore band, a near-shore band and an offshore band. There is also the option to include an additional offshore band in which unregulated international fishing can be simulated.

The choice of inshore-offshore bands are based on experience in the ETBF (Eastern Tuna and billfish fishery) where the fleet started fishing for swordfish mostly in inshore waters and then moved further offshore as swordfish catch rates dropped. Although there may well be additional north-south dynamics of swordfish in the real systems on either, or both, the East and West Coast of Australia, the simplified model used here is sufficient for a general exploration of the questions posed above. The choice of an additional offshore fishing area with unregulated effort is based on the fact that in both the Western Central Pacific and the Indian Ocean there is additional effort on swordfish by non-Australian fleets. Although the stock structure of swordfish is still poorly known, the possibility that one stock is harvested by Australian vessels and other fleets in international waters cannot be ruled out.

The fishing, and hence total, mortality is area-specific. The fishing mortality at time t, age, a, of gender g, in area A, $Z_{t,a,g,A}$, is the sum of the fishing mortality-at-age by area, $F_{t,a,g,A}$, and the instantaneous rate of age-specific natural mortality, M_a (note, however, that for most simulations, natural mortality was constant over all ages and the same for both genders) :

$$Z_{t,a,g,A} = F_{t,a,g,A} + M_a \qquad (\mathbf{9})$$

Natural mortality is assumed constant over time. The fishing mortality at time t, (quarterly)-age a for gender g in area A, $F_{t,a,g,A}$ is calculated:

$$F_{t,a,g,A} = S_{a,g} \cdot P_{t,A} \cdot f_t \qquad (\mathbf{10})$$

where $S_{a,g}$ is the selectivity-at-age a for gender g, $P_{t,A}$ is the proportion of the fishing mortality assigned to area A and f_t is the total fishing mortality in time step t.

Selectivity is essentially specified in terms of length and then converted to an equivalent selectivity-at-age. For "long-line" like fishing, selectivity is knife-edged, with fish being 100% selected above the length at first capture, l_c and 0% below it. Selectivity for "purse-seine" like fishing is 1 between specified selectivity sizes, and 0 otherwise. The length-based selectivity curves were converted into age-based selectivity curves using the von Bertalanffy growth curve. Age-specific selectivity curves are therefore different for males and females when sexual dimorphism (different growth curves) are assumed.

There are many possible approaches to defining the proportions of overall fishing mortality in each area (i.e. defining the values of the proportions, $P_{t,A}$). The proportions should be thought of as the outcome of the spatial distribution of the fishing fleet. Three options were considered:
i) random distribution of vessels – a proportion is randomly generated
ii) effort (and therefore F) proportional to the relative density of fish in each area, or
iii) vessels tracking relative fish density, but with a preference for inshore fishing and near-shore fishing.

The latter occurs as follows. All fishing commences in the most inshore area (i.e. the proportions from inshore to offshore areas are [1,0,0,0] respectively). If the catch rate in any area drops below a specified fraction of its initial value, then a specified fraction of the fishing mortality (F) is relocated into the next adjacent offshore area. If the catch rate in this new area then drops below

the fraction of its initial value, the specified fraction of the new F in this area is again relocated, and so on. Once the catch rate drops below a specified limit threshold, fishing ceases in that area and all remaining fishing mortality is relocated into the next adjacent offshore area. Again, this is applied in a cascading manner through the offshore areas. In the most extreme outcome, all fishing mortality would be located in the most offshore area. If, in the meantime, an area recovers such that its potential catch rate is higher than the initial threshold, then any effort further offshore will be relocated back into this area, so that fishing will always occur preferentially inshore. The description of actual scenarios and choices of values for the various quantities is given in Section 13.

The catch numbers in each quarterly time step t for quarterly age a and area A is calculated using the traditional catch equation with a lognormal error term:

$$C_{t,a,g,A} = \frac{F_{t,a,g,A}}{Z_{t,a,g,A}}\left(1 - e^{-Z_{t,a,g,A}}\right)N_{t,a,g,A} \cdot e^{\mu_t - \sigma_c^2/2} \qquad \mu_t \sim N(0;\sigma_c^2) \qquad (\mathbf{11})$$

where

σ_c is the standard deviation of the logarithm of the error in catch numbers (noting that this error term can be omitted). Only a small error ($\sigma_c = 0.05$) was typically imposed.
$N_{t,a,g,A}$ is the number of fish of (quarterly) age a of gender g in area A at the start of quarterly time step t

In the spatial model, the population numbers are updated using total mortality as per equation 1 above rather than by directly deducting catch. The calculated catch (from equation 11) is used in two ways: 1) for sampling purposes, in order to obtain size-based indicators and other catch-derived diagnostics, and 2) where management is by catch, and that 'catch taken' matches with the F that updates the population (see section below on "Translating catch quota into fishing mortality") . In this context, the error term is a (small) implementation error. In the non-spatial version of the model, however, the numbers are updated using the catch sampled with error. This is highlighted in the methods sub-sections of Sections 9 and 10.

Obtaining size-frequency samples from the catch

The data from which the stock status indicators are derived (for example, CPUE, mean length etc), are generated by simulating sampling of the catch. The proportion of the catch to be sampled is specified. If 100% of the catch is sampled, the total number of fish in the catch-at-age (equation 11) forms the sample. If <100% of the catch is sampled, a new set of 'proportions at age' in the sampled catch is obtained from the true 'proportions at age' in the catch. A multinomial random variable is used to generate ages to which each fish in the catch sample is assigned.

A size distribution is sampled from the catch-at-age vector by assuming that length at (quarterly) age for each gender follows a normal distribution with mean as calculated from the von-Bertalanffy growth curve, and standard deviation (σ_a) a linear function of the square root of the (annual) age $\tilde{a}$:

$$\sigma_a = 5.4 \cdot \sqrt{\tilde{a}/4} + 6.5 \qquad (\mathbf{12})$$

This relationship is loosely based on summaries of swordfish mean length and variance of length by age class in Young and Drake (2004). A weight distribution is obtained by converting each

length to a weight using the length-weight relationship given above (equation 8). In some of the non-spatial model explorations there were subtle differences to this sampling approach to introduce additional error into the process. This is discussed in the methods sub-sections of Sections 9 and 10.

3. Stock Status Indicators

A range of catch-per-unit-effort (CPUE)-based and size-based indicators were calculated from the simulated catch, and the length or weight frequencies sampled from the catch. For the simple (non-spatial) versions of the model, this was done for the overall catch (i.e. there were no gender- or area-specific indices). Even for the spatial version of the model, the size-based indicators were also calculated for the catch from all areas combined.

The set of size-based indicators that were used in most of the explorations are:

mL, mdL, L90 or L95 - the mean, median, 90th, or 95th, percentile of length

mW, mdW, W90 or W95 - the mean, median, 90th, or 95th, percentile of weight

ps - the proportion of "small" fish in the catch (defined as the proportion of fish $\leq t_c+1$)

pb - the proportion of "large" fish in the catch (defined as the proportion of fish $\geq t_c+3$)

pm - the proportion of mature fish in that catch (defined as the proportion of fish older than the age-at-maturity, t_m).

The exact specifications of the cut-off sizes (t_c, t_m) are given in each relevant Section. It is self-evident how the above indicators can be obtained from the length or weight frequency distributions.

The set of CPUE-based indicators that were used in most of the explorations are:

u - CPUE in numbers

us - CPUE of small fish in numbers (u x ps)

ub - CPUE of big fish numbers (u x pb)

uw - CPUE in weight

uSB - CPUE of spawning fish (weight)

The CPUE of small and big fish are calculated by multiplying the CPUE in numbers by the proportion of fish in the catch in the desired size range. The CPUE of spawning fish is the product of the CPUE in numbers, the proportion mature, and the mean weight of spawners (from the weight frequency distribution).

Note that the most appropriate way of calculating each index depends somewhat on the way in which the 'data' are simulated and the level of detail in the data. The descriptions below are therefore, in some cases, modified for specific simulation trials; such modifications are described in the individual Sections.

CPUE in numbers was obtained for each area by multiplying a constant catchability (q) by the total fishable numbers (N^f_A) in that area. The fishable numbers are simply the numbers at age multiplied by the selectivity curve (S). This calculation of CPUE is the same as first calculating a nominal effort as:

$$E_A = C_A/(q.N^f_A) \qquad (\mathbf{13})$$

and then dividing the catch in numbers by that effort. Total CPUE across all areas was obtained by applying the same catchability to the total numbers summed across all areas:

$$u = q \cdot \sum_{A=1}^{y} N_A^f \quad \text{where } y \text{ is the number of areas.}$$

If fishing did not occur in all areas, CPUE was calculated using the fishable numbers summed only across the areas in which fishing occurred. A multiplicative lognormal random error, with standard deviation σ_{CPUE} was generally applied to the CPUE. When effort creep (i.e. an increase in catchability) was modelled, the same approach was used, but q was assumed to increase over time by some specified percentage.

In the non-spatial version of the model, where decision rules used effort (or essentially harvest rate, F) rather than catch controls, the catch rates were based on the sampled catch, with error where applicable, divided by the nominal effort obtained from F/q.

CPUE by weight was obtained by multiplying CPUE in numbers by the average individual fish weight across the relevant size category (average weight calculated as the total weight of the sampled catch in the relevant size range, divided by the number of fish in the sample for that size range).

Three additional 'spatial' indicators, intended as indicators of how far off-shore the fleet is fishing and/or how spread out or how aggregated the fleet is operating, are constructed for the spatial model:

- Number of vessels per area; calculated by scaling as set number of vessels by the proportion of the fishing mortality in each area
- Mean distance offshore; calculated by assigning each "vessel" a random distance within the area.
- Proportion of total fishing mortality, F, in each area, $P_{t,A}$

These indicators are examples which should be informative in the context of the simulated fish and fleet dynamics. In reality, care would need to be taken when constructing such indicators if there are strong features (e.g. sea mounts or fronts) or strong fish aggregation or habitat preference dynamics.

4. Inputs to the simulation model

The main input parameters to the area-specific simulation are listed in this section.

1. Population model

- steepness (h) defining the Beverton-Holt stock-recruitment curve
- growth rate, k, in the von Bertalanffy growth curve, and the less influential parameters L_∞ and t_0
- natural mortality at age (M_a) vector
- age at maturity (t_m) . Maturity is knife-edged, with 100% maturity assumed above the age of maturity, t_m, and 0% below it.
- parameters a and b in the weight-at-length power relationship (typically a = 0.00002, b = 2.9)
- the proportion of female and male recruits, K_g

- the resident proportion of the stock, *r*, assumed to be the same for each area
- the proportions of the (non-resident proportion) of the stock assigned to each area, n_A
- an initial number *InitN* of total animals with which to start the simulation
- standard deviation of the logarithm of the fluctuations in recruitment, σ_R

These quantities are assumed to be constant over time. If sexual dimorphism is assumed, the growth parameters and age at maturity are gender-specific.

2. Fishery dynamics

- magnitude and duration of fishing mortality *F* prior to implementation of decision rule
- total fishing mortality *F*, or catch (either from the decision rule, or a constant fixed *F*)
- catchability (*q*) for calculation of CPUE
- length-at-first-capture, l_c
- selectivity-at-age, S_a
- initial proportion of fishing mortality by area (which is then modified via move-on criteria): $P_{0,A}$
- move-on criteria in area-specific model:
  - % of the initial CPUE below which fishing mortality relocates offshore
  - % of fishing mortality relocated offshore once the CPUE threshold is breached
  - % of the initial CPUE below which fishing ceases in the area and all fishing mortality is relocated offshore
- standard deviation of the logarithm of the fluctuations in catch and CPUE

Note that scaling quantities such as recruitment *R*, L_∞, and *a* in the weight-length relationship, are of little relevance because we are primarily interested in the change in an indicator at some harvest rate relative to its value at $F=0$. Setting these quantities at arbitrary values should therefore not affect results of relative change in indicators.

Inputs to define decision rules are considered in sub-section 6 below.

5. Starting a simulation run

The population was assumed to be unexploited at the start of the simulations. The initial spawner biomass is obtained via a spawner-biomass-per-recruit (*SPR*) calculation. As the time-steps are quarterly, and recruitment occurs in the first two quarters of the year, this in turn is an average of the spawner-biomass-per-recruit for the quarter 1 recruitment and quarter 2 recruitment (assuming these are spawned from the spawner biomass in the previous quarter 4, and quarter 1, respectively):

$$SPR = \left(\sum_{qtr=1}^{2} \sum_{a=1}^{x} \sum_{g=1}^{2} V_{a,qtr} K_g p_{a,g} w_{a,g} \right) / 2 \qquad (\mathbf{14})$$

where

if $a < x$ (recall x is the maximum "plus-group" quarterly age)

$$V_{a,qtr} = \begin{cases} 0 & a = 0 \\ 1 & a = 1 \\ e^{-\sum_{i=1}^{a} M} & a > 1 \end{cases} \quad (\mathbf{15})$$

if $a = x$

$$V_x = \begin{cases} e^{-\sum_{i=1}^{x} M} \cdot (e^{-3\sum_{i=1}^{x} M} + 1)/(1 - e^{-4\sum_{i=1}^{x} M}) & qtr = 1 \\ e^{-\sum_{i=1}^{x} M} \cdot (e^{-\sum_{i=1}^{x} M} + 1)/(1 - e^{-4\sum_{i=1}^{x} M}) & qtr = 2 \end{cases} \quad (\mathbf{16})$$

where

qtr is the quarter of the year in which the recruitment is occurring. For the recruitment occurring in quarters 1 and 2, V_x is the plus-group equation for the spawner biomass from the preceding quarter (i.e. quarter 4 in the previous year, and quarter 1 in the current year, respectively).

Also recall that $p_{a,g}$ is the proportion of the population of age *a*, gender *g* that is mature, and $w_{a,g}$ is the body weight of an individual of age *a,* gender *g*. The initial spawner biomass, B_0, was then obtained by multiplying the unexploited spawner-biomass-per-recruit by the pre-specified initial numbers *InitNum:*

$$B_0 = SPR \cdot InitNum \quad (\mathbf{17})$$

The initial numbers-at-age by gender and area was calculated as

$$N_{0,a,g,A} = K_g \cdot InitNum \cdot V_a \cdot n_A \quad (\mathbf{18})$$

In this context, when calculating $N_{0,x,g,A}$, V_x is the plus-group equation associated with quarter 4 (i.e. the last quarter immediately prior to the commencement of the simulated dynamics, assuming these commence in quarter 1 of the first simulated year), i.e.:

$$V_x = e^{-\sum_{i=1}^{x} M} \cdot (e^{-3\sum_{i=1}^{x} M} + 1)/(1 - e^{-4\sum_{i=1}^{x} M}) \quad (\mathbf{19})$$

Initial biomass by age, gender and area was simply calculated by multiplying the above numbers by the corresponding weight-at-age.

In commencing the area-specific simulations, the virgin stock was fished for a number of years (generally 5) at a fixed harvest rate, which was generally some multiple of MSY designed to fish the stock down below MSY. A decision rule was then applied annually, to generate a (generally 30 year) time series. For each spatial simulation, a small number of realisations (generally 10) were run, wherein random variability was applied to recruitment, catch and CPUE.

6. Decision rules

Various forms of decision rules were used across the study to test the performance of indicators/suites of indicators in a management feedback context. All decision rules require specification of the time step in which to commence the decision rule and how often it should be invoked.

'40/10'-style rule
Note that this type of rule is sometimes referred to as '40-10', or as '10/40' or '10-40'.

This type of decision rule was originally defined in terms of the estimated level of spawning biomass (SSB). The logic was that the catch should be set at a minimum (usually zero) if SSB is below 10% of unexploited SSB. For SSB above 40% of unexploited SSB the catch would be set some maximum level. For SSB between 10% and 40% of unexploited, the catch would increase linearly from the minimum to the maximum. The name '40/10' reflects the choice of 10% and 40% as the change points in the relationship between catch and SSB. A similar rule can be defined in terms of other percentages; for example, 20/40 rules are quite common.

Here we define this rule in terms of an indicator instead of an estimate of SSB, but we use the same logic (A1.1). Change points (or reference points) are based on spawner-per-recruit (SPR) considerations. For example, the expected value of the indicator is calculated for a harvest rate at which SPR would be 10% of unexploited SPR; this provides the limit reference point. The expected value of the indicator is also calculated for a harvest rate at which SPR would be 40% of unexploited SPR; this provides the target reference point. If the indicator value is below the limit reference point, the catch or fishing mortality is set at zero, or, as is illustrated in Figure A1.1, a fixed very low level. When the indicator value is greater than that target reference point, the catch or fishing mortality is set at some maximum which could, for example, be MSY or F_{msy}. A less risky option is to set it at a lower level (e.g. $0.8MSY$ or $0.8F_{msy}$). When the indicator lies between the two reference points, the catch or fishing mortality follows a straight line relationship between the minimum and maximum. In this study, the reason for not setting the minimum catch or harvest rate to zero is to avoid the total absence of simulated 'data'.

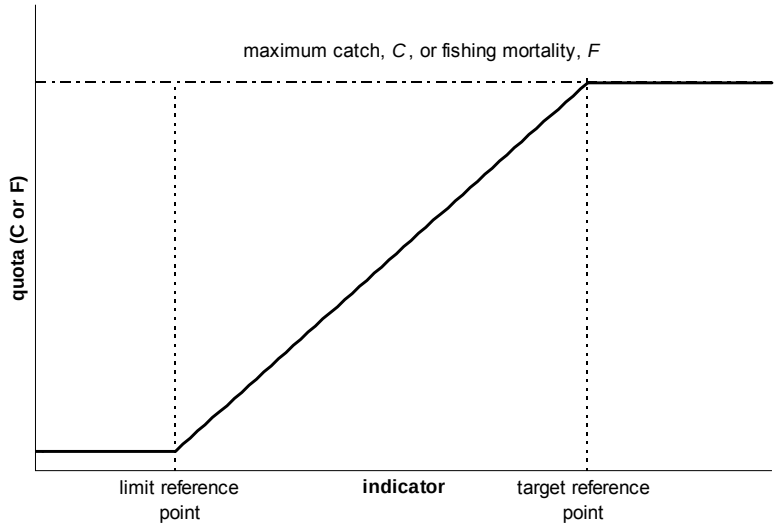


Figure A1.1 Representation of the "40/10" style decision rule used for adjusting quota or fishing mortality.

The quantities that need to be specified for this type of rule are:

- maximum catch or fishing mortality
- minimum catch or fishing mortality
- reference points for indicator
- period over which indicator is averaged, unless a single year is used

The Decision tree

An empirical "decision tree" was developed as part of the Harvest Strategy for the ETBF (Campbell *et al.*, 2007; Davies *et al.*, 2008), consistent with the Harvest Strategy Policy resulting from the Ministerial Directive that all Commonwealth fisheries are required to have formal harvest strategies in place by January 1, 2008 (Anon, 2007). For a full description see Campbell *et al.* (2007); we provide a shortened, modified version here for completeness.

The decision tree consists of four levels and uses several indicators. It is therefore complicated and probably best illustrated by the flow-diagram shown in Figure A1.2. The logic is, however, relatively straightforward.

i) Level 1: Primary Determination of Recommended Biological Catch (RBC)

At the first level, the question is: where is the CPUE of prime-sized (commercially most valuable) fish relative to its target level? The decision rule increases or decreases catch depending on the answer to this question, and with respect to the pre-defined target and a time-frame within which the target should be reached. The *RBC* is set by the angle between the current $CPUE_{prime}$ and the target $CPUE_{prime}$ an agreed number of years ahead (the so-called 'Slope-to-target). The catch is increased when the current $CPUE_{prime}$ is above the target and decreased when the current $CPUE_{prime}$ is below the target. The slope of the $CPUE_{prime}$ is taken over the previous 5 years and the nominated number of years to reach $CPUE_{targ}$ is also taken to be 5 years. The *RBC* is adjusted using the following rule, referred to as the primary decision rule:

$$RBC_{t+1} = RBC_t * [1 + \beta * \text{Slope-to-Target}_{CPUEprime}] \qquad (\mathbf{20})$$

where β is a specified parameter which reflects the responsiveness of the decision rule (Note this β should not be confused with β in the stock-recruit relationship; this notation has been retained to be consistent with the original documents from which this decision rule was taken).

This value will be modified or affirmed by application of the subsequent levels of the Decision Tree. The subsequent levels attempt to identify and resolve situations which would remain undetected or ambiguous if $CPUE_{prime}$ alone were used (e.g. stock declines masked by effort creep, recruitment failure, and/or strong recruitment pulses).

ii) Second Level of the Decision Tree based on $\text{CPUE}_{\text{prime}}$

In the second level of the Decision-Tree the question asked is whether $CPUE_{prime}$:

- Rising,
- Stable, or
- Falling.

Stability is assessed on whether or not the annual rate-of-change in $CPUE_{prime}$ is within a given limit (in this case, 5% of the average $CPUE_{prime}$ over 5 years). The trend in $CPUE_{prime}$ then determines which of the three main limbs of the Decision Tree will be used in each year's

assessment. The three main limbs are shown by the three arrows and the three boxes across the bottom half of Figure A1.2. There is no adjustment of the RBC at this level of the tree.

iii) Third Level of the Decision Tree based on $CPUE_{Old}$ and Proportion Old

In the third level of the Decision Tree $CPUE_{Old}$ and *Proportion-Old* are considered. Both $CPUE_{Old}$ and Proportion of these fish are used, since they can be highly informative when used together. Four questions are considered. All take the general form: "Is $CPUE_{Old}$ is above/below Target Level and is *Proportion Old* above/below Target Level?" The four answers (A-D) cover all four possible true combinations of this generalized question:

A. CPUE of Old fish is above and Proportion is above the reference point of SPR_{40}.
B. CPUE of Old fish is above but Proportion is below the reference point of SPR_{40}.
C. CPUE of Old fish is below but Proportion is above the reference point of SPR_{40}.
D. CPUE of Old fish is below and Proportion is below the reference point of SPR_{40}.

In the tabular form of the Decision Tree shown in A1.2 this third level of the Decision Tree is represented by the four questions contained in the box that stretches horizontally across the middle of the figure. Depending on the answer to this question (A-D) the assessment proceeds down the limb (Rising, Stable, Falling) of the Decision tree determined at the second level of the Decision Tree, to the matching category (A, B, C or D) at the fourth level of the tree. There is again no adjustment to the RBC at this level, but together with level 2, the route to follow at level 4 is defined.

iv) Fourth Level of Decision Tree based on CPUE-Recruits

The fourth level of the decision tree uses $CPUE_{Recruits}$. The idea with this smallest class of fish is that they are sub-optimal for the market, and biologically speaking best kept out of the fishery until they are larger. However, trends in their catch rate can provide information about recruitment trends to the fishery. Recruitment pulses, increases or declines can be very informative. In some cases changes in the *Proportion-Old* may be produced by an increase in recruitment rather than a change in the actual number of old fish. In this way $CPUE_{Recruits}$ provides a final test for distinguishing between some otherwise potentially ambiguous possibilities. It is at this fourth level that there are potentially further adjustments to the RBC, but if there is an adjustment, it is always in the form of a reduction (see Figure A1.2, level 4)

Application of each of the levels in the Decision-Tree is contingent on target values against which the corresponding indicator can be assessed, and there are additional parameters for adjusting the RBC. Table A1.1 lists these input parameters together with the values used in the application of the decision tree in Section 13 of this report.

Table A1.1 Listing of the parameters used in the decision tree

Decision Level	Parameter	Suggested Value
Level 1	Number of years over which the slope of $CPUE_{prime}$ is calculated	5 years
	Target value for $CPUE_{prime}$	$CPUE_{prime}$ at *MSY*
	Feedback gain/responsiveness factor, β	1.2
Level 2	Bound on the percentage annual change in $CPUE_{prime}$ to define stability in this indicator (Note: change is relative to the mean value of $CPUE_{prime}$ over the previous y years – see next entry)	5% per year
	Number of years mean $CPUE_{prime}$ is calculated over	5 years
Level 3	Target value for $CPUE_{Old}$	SPR_{40}
	Target value for *Proportion-Old*	SPR_{40}
Level 4	Value of *CPUE-Recruits* to define high recruitment	70% $CPUE_{Rec\ (F=0)}$
	Decrease in *CPUE-Recruits* to define declining recruitment	10% per year[#]
	Reduction factor on *RBC*	10%
	Number of years mean $CPUE_{Recruits}$ is calculated over	5 years

Change defined in a similar manner as for the Level 2 parameter

$CPUE_{Rec(F=0)}$ is the expected (or theoretical) CPUE of recruits when F=0

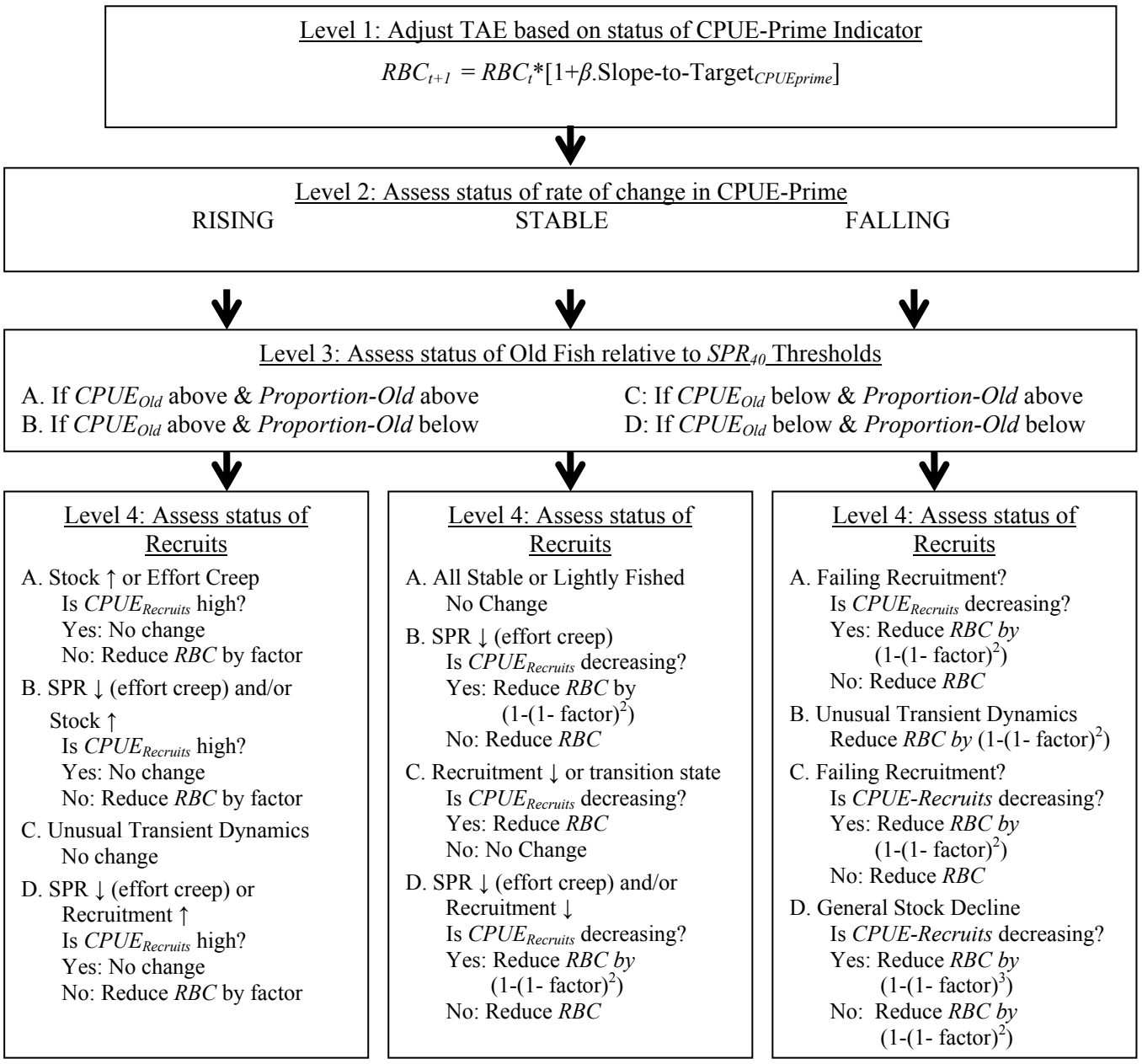


Figure A1.2 Flowchart diagram of the four levels of the decision tree

7. Translating catch quota fishing mortality

Decision rules manage the fishery either by specifying an annual catch quota or a fishing mortality. If the decision rule yields fishing mortalities, then these feed directly into the population dynamics equations. However, if managing by catch, the specified catch quota is translated to a fishing mortality.

The annual fishing mortality corresponding to this catch quota for the population at that time is determined via the bisection method. This trials various levels of F in the catch equation given the current stock abundance (summing over areas) until the F corresponding to the specified quota is found. If the population is updated quarterly, this fishing mortality is then divided by 4. However, applying this same F every quarter will not necessarily yield a total annual catch that

equates to the catch quota due to the dynamics affecting the total fishable numbers throughout the year. To address this (albeit inelegantly), the quartered *F* is applied for the first 3 quarters of the year, and the associated catches are summed over these first 3 quarters. The difference between this subtotal and the annual catch quota forms the remaining quota for the year, and a corresponding fishing mortality is again calculated (via bisection) corresponding to this remaining quota.

8. Spatial triggers and external fishing

Spatial triggers are unusual because, unlike indicators such as catch rates and mean length (for example), they do not have an associated biological expected level when fishing at a given *F*. In this study, they are tested as a simple proxy for localised depletion. In testing the use of a spatial indicator, the more formal decision rules (such as the 40/20 style rule, or the decision tree, as defined above) are applied first, but then a spatial trigger is used to further modify the catch or the fishing mortality.

In the area-specific models, a spatial trigger may be invoked whereby the overall catch quota or fishing mortality is reduced by a set percentage if the average distance offshore exceeds the furthest average distance offshore occurring over a specified number of early years of the fishery. The "move-on" criteria for the fishing fleet are described in detail in the Section 13.

In some spatial simulations, external fishing is modelled by having some fixed fishing mortality in an additional area (typically, a multiple of the F_{msy} calculated for the stock as a whole, assuming that selectivities are identical between the domestic and foreign fleets). When external fishing is occurring, the optimal catch in the decision rule is modified by the proportion of the total population occurring within the domestically fished areas.

9. Performance indicators

For the spatial model, four types of quantities were considered when evaluating the performance of a decision rule: spawning biomass, CPUE, catch and variability in catch. There were also the fleet-related spatial indicators (number of vessels in each area, mean distance offshore). Within each category, it is important to look at area-specific and temporal behaviour and to look at overall summaries such as means, as well other measures such as lower CPUE percentiles. It should be noted though that lower or upper percentiles are unreliable for a low number of realisations, such as were run for the spatial simulations. It is also informative to look at probabilities (or frequency counts) of, say, spawning biomass being below some level. The simulation model generates a rather long list of such measures, which are summarised in Table A1.2. In the spatial model, we mostly look at a subset of about 6 of these (see Section 13). The performance measures are only calculated over those years that the decision rule, or harvest strategy, is implemented, i.e. from year 6 to year 30 in the spatial simulations.

We summarise catch statistics relative to C_{msy} or, in some cases such as time-series plots of results, relative to the catch in the first year. The reason for this is because the absolute values of catch are arbitrary since they are based on an arbitrary absolute population size. The choice of C_{msy} is for convenience and not because we attach any particular value to C_{msy}.

It is important that summary statistics are interpreted with caution for rules that yield cyclical behaviour, such as the decision tree framework.

Table A1.2. Performance measures for the Spatial model simulations. Time-series are generated for all quantities and in some cases summary statistics are derived (2nd column)

	Summary statistics
BIOMASS	
B/B_{msy}	Value in year 30 Average proportion of time $B/B_{msy}<1$ (ie $B<B_{msy}$)
B/B0	average over final 10 years minimum
B relative to 0.4B0	proportion of time B<0.4B0
CATCH RATE	
CPUE	average 10th percentile minimum
CATCH	
C/C_{msy}	average minimum
C	variability
SPATIAL INDICATORS	
Number of areas fished	
Number of vessels in each area	
Mean distance offshore	
Proportion of total F in each area	

22. Appendix 2. Reference points in fisheries management: a background paper

Preamble

This appendix contains information and comments on reference points in fisheries management, prepared as a background paper to address part of objective 5 of the project: "Develop meaningful reference points, in consultation with FAGs". The Ministerial Directive with its associated Harvest Policy documents and processes, however replaced this particular part of the objective. Strong guidelines for choice of reference points were given and there was little point in running a parallel process to that of the HSWG. The general background and explanations of what reference points are, and what their purpose is, are still relevant. This information is therefore included as this appendix to the full report.

Summary

Reference points are tools for implementing management objectives. A reference point is an estimated or quantified value that corresponds to a specific state of the resource or the fishery, and can be used as a guide for fisheries management. Most fisheries management frameworks now recognise the need for two types of reference points: Target reference points (TRP) that reflect a desirable state of the fishery or resource (where we want to be / what we want to aim for), and Limit reference points (LRP) that reflect states of the fishery or resource to avoid. Reference points are most often defined in terms of fishing mortality, yield or spawning biomass.

There are many candidate reference points based on a range of fisheries concepts, including maximum sustainable yield (MSY), maximum yield per recruit and 'safe' levels of spawner per recruit. There is a large body of literature on the performance and merits of the various candidate reference points. The increasing use of simulation studies and Management Strategy Evaluation approaches, makes it feasible, and sensible, to construct 'custom-made' reference points which best match the management objectives of the fishery. Where this is not possible, there is also a large body of information to guide the choice of reference points based on experience in a wide range of fisheries.

The concept of Escapement (number or proportion of spawners that 'escape' the fishery) is less well known as a basis for reference points, but it has potential benefits for management of short-lived species, and has been applied to the management of squid and salmon fisheries.

The concept of MSY, on the other hand, is very familiar and has frequently been used as a basis for management. On the basis of theoretical work and practical experience, the scientific consensus is now that MSY-based quantities (e.g. F_{msy}, B_{msy}) are not appropriate as target reference points because of the high associated risk of overfishing and stock depletion. Instead, these quantities are considered appropriate as limit reference points.

Reference points (both TRPs and LRPs) should be compatible with the management objectives for the fishery and compatible with one another. A management framework that uses reference points requires active monitoring and (usually) annual readjustment of the management measures, e.g. the catch quota or effort level. One could argue that a fishery managed via effort should require less frequent adjustments than a fishery managed via catch quotas. This is because the catch associated with a fixed level of effort should vary with the stock size. In such cases, it would, of course, be particularly important to monitor the efficiency of the fleet for 'effort creep'.

There should be a pre-specified, agreed action, or process for arriving at action, when an LRP is exceeded and the stock or fishery is deemed to be in the area which should be avoided. The action may be defined in terms of a 'rebuilding plan' with associated time-frame, and there may be additional management measures that are invoked when the resource is in the 'to be avoided' area.

Reference points sit very naturally in any management procedure (MP) framework, and alternative reference points can be evaluated, their performance compared, and details fine-tuned to best meet management objectives. Instead of taking a reference point and having to 'live' with its associated risk, one can start by defining the risk you are prepared to take, and find the point which gives you that level of risk. It should be noted that the decision about level of risk is really a policy choice. Science can, however, provide guidance about the likely consequences of different levels of risk to the harvested population.

NOTE: In this Appendix, references to the literature are given at the end of the Appendix rather than in Section 18.

1. Introduction

This document is intended as a non-technical background paper on the need for, choice and implications of reference points in fisheries management. I have relied heavily on existing scientific literature, in particular the FAO Technical Paper by Caddy and Mahon (1995; referred to as "Caddy & Mahon" in the rest of the document), and on my experience elsewhere on a wide range of fisheries in different management contexts: squid fisheries around the Falkland Islands (managed by the UK), krill and icefish fisheries in the Antarctic (CCAMLR[41]), groundfish trawl fisheries in the USA (NMFS[42]), trawl fisheries for herring, cod-like species and flatfish in the North Sea (ICES[43]), and more recently, tropical tunas and billfish in the Indian Ocean (IOTC[44]) and southern bluefin tuna (CCSBT[45]).

2. What are reference points and why do we need them?

Reference points are tools for implementing management objectives. Caddy & Mahon suggest that "reference points begin as conceptual criteria which capture in broad terms the management objective for the fishery". The notion that one begins with a management objective is important, because that should play a role in the choice of a reference point. To make this a little more concrete, consider the example where the management objective is to maximise yield. A possible (and frequently used) conceptual reference point which would capture this objective is MSY (maximum sustainable yield). Another candidate would be MEY (maximum sustainable economic yield). So for a single objective, there could be more than one potential conceptual criterion which could form the basis for a reference point.

In order to implement this into management, a conceptual reference point is not enough. It must be possible to convert the conceptual reference point into a 'Technical Reference point', which can be calculated or quantified on the basis of biological or economic characteristics of the

[41] Convention for the Conservation of Antarctic Marine Living Resources
[42] National Marine Fisheries Service of the USA
[43] International Council for the Exploration of the Seas; advisory body for fisheries management in Europe.
[44] Indian Ocean Tuna Commission
[45] Commission for the Conservation of Southern Bluefin Tuna

fishery.[46] From a technical point of view, MSY has been interpreted in various ways: (a) the peak of the surplus production curve, (b) the point of maximum surplus reproduction on a stock recruitment curve or (c) a more literal interpretation as the maximum constant yield that can be taken year after year. Therefore, to finish the task of defining a reference point, we also need to say whether we are talking about definition (a), (b) or (c).

To summarise: A reference point is an estimated or quantified value corresponding to a specific state of the resource and/or the fishery, is related to a management objective, and can be used as a guide for fisheries management. The definition should become clearer once we've looked at the next section and at some examples.

In the rest of this paper, the distinction between 'conceptual' and 'technical' reference points is dropped. Instead, we look at a more important distinction between Target and Limit reference points.

3. Types of reference points: targets and limits.

Target reference points (TRP)

The first types of reference points developed for fisheries management were, what is now known as, target reference points. They are essentially aimed at answering the question: 'what are we trying to achieve in the fishery and for the resource?' At what level would we like the fishery effort or catch to be on average? At what level would we like the resource to be on average? These questions essentially aim to define a target (or targets) to aim for, and the reference points that we define to implement these management objectives are therefore called target reference points.

Caddy & Mahon's definition is: "A Target Reference Point indicates a state of a fishery and/or resource which is considered to be desirable and at which management action, whether during development or stock rebuilding, should aim." The 'state' can be defined and measured in terms of the level of effort in the fishery, the average level of catch, and the average level of spawning biomass (SSB), for example.

A hypothetical and simplified example might be:

Management objective:	maximise economic yield
Target reference point:	F_{mey} (with method for estimating it clearly defined)
Management action:	set effort to be compatible with F_{mey} or determine quota on the basis of F_{mey} and current estimate of fishable biomass.

(This hypothetical example is highly simplified to illustrate broad concepts; there are many details which would need to be considered to make something like this operational).

Limit reference points (LRP)

Together with the development of and work on reference points, there has been increasing recognition of the importance of uncertainty, and the need to take this into account when designing management frameworks. There are many sources of uncertainty: uncertainty in estimates of current stock status or fishing mortality, uncertainty about the underlying dynamics of the resource (e.g. productivity or stock-recruit relationship), uncertainty about next year's recruitment (random variability). The presence of uncertainty in natural systems means that even if we are aiming at some target, we might miss. If this happens often, or if we miss the target by a

[46] Reference points based on biological characteristics of the resource are sometimes referred to as Biological Reference Points to distinguish them from other types, e.g. economic reference points.

large amount, it can have serious consequences, such as stock collapse, which of course also implies fishery collapse.

Most management frameworks now recognise the need for management objectives which relate to 'areas' we want to avoid, such as very low spawning biomass which could lead to stock collapse, or very high fishing mortality which could lead to overfishing. In the USA, fisheries management plans require that a reference point be provided for overfishing for each stock. In the ICES area, this conceptual point is referred to as the Minimum Biologically Acceptable Level (*of SSB*), MBAL, for the fishery. 'Overfishing' and MBALs must be technically defined, i.e. quantified in order to be implemented. These types of reference points are called limit reference points[47], and they are usually used together with a (or more than one) target reference point.

Caddy & Mahon's definition is: "A Limit Reference Point indicates a state of a fishery and/or a resource which is considered to be undesirable and which management action should avoid." In most implementations the implication of entering the 'red/danger' area, or area to avoid, by exceeding an LRP is drastic management action. This could be closure of the fishery, or substantial reductions in quota or effort. It is therefore sensible to ensure that there is a LOW probability of exceeding the LRP.

A hypothetical and simplified example might be:

Management objective:	ensure a low probability of stock collapse
Limit reference point:	$SSB_{0.2}$ i.e. SSB at 20% of what the SSB for the unfished stock would be
Management:	If SSB is estimated to be below $SSB_{0.2}$, take drastic action (e.g. reduce catch or effort by 50%)

The management action (hypothetical in the above example) should be clearly pre-defined and agreed. If the LRP was defined in terms of fishing mortality, then one would take drastic action if the estimated fishing mortality was above its limit reference point.

4. Multiple objectives - multiple reference points

Fisheries management is intimately familiar with multiples objectives. Three familiar ones are: optimise yield, minimise the risk of stock collapse, and ensure high stability in the fishery. Fisheries management is also familiar with the fact that it is almost always impossible to achieve all objectives simultaneously. In other words, there are trade-offs between the objectives. As the yield or level of effort is increased, the risk of low SSB is increased. If there is total stability by having a constant catch year after year, the average yield will almost certainly have to be lower than if catch was allowed to vary from year in order to achieve the same level of risk to the stock.

Multiple objectives imply more than one reference point and we have already noted the need for two types of reference points, targets and limits. A commonly used framework is to define a target and a limit reference point for spawning biomass, and a target and limit reference point for the harvest rate. The fact that there are trade-offs between different objectives means that a set of reference points need to be chosen carefully to ensure that they do not contradict one another. It is worth noting that the approach of Management Strategy Evaluation (MSE; where a simulation model of the fishery and resource is used to test the performance of management

47 Limit reference points are also sometimes called 'Threshold' reference points, but to avoid confusion with the acronym Target RP, the word 'limit' tends to be preferred.

procedures, including decision rules, reference points etc.) is potentially very useful for defining and choosing compatible reference points[48].

A simple example may highlight the potential problem. Assume there are two objectives: maximise yield and minimise the chances of stock collapse or very low SSB. Assume that a target reference point for the first is objective is chosen as F_{msy}, and that a LRP for the second objective is chosen as SSB at 20% of unexploited SSB (call it $SSB_{0.2}$). Recall that the LRP implies that we want to avoid, or have a very low probability of, SSB levels below $SSB_{0.2.}$ If we made these choices without careful consideration of what harvesting at F_{msy} implies for the stock, we may find that SSB_{msy}[49] is so close to $SSB_{0.2}$, that there is in fact a very high probability of falling below the LRP. This is admittedly a slightly contrived example, because one would NOT (should not) make choices of reference points without careful consideration of the connections between them.

In addition to multiple objectives there may, of course, be more than one stakeholder group with different objectives. For example, recreational fisheries prefer high strike rates (high CPUE) and large sized fish, which would usually imply lower harvest rates than those associated with, say, maximising yield. Another potential complication arises in mixed fisheries, particularly when there is a large number of species with very different life history characteristics (e.g. fast growing short-lived and slow growing long-lived; Where species are relatively similar, the problem is likely to be less serious). The science to underpin choices of reference points for such situations is still relatively young. This paper focuses on the simpler single stock, single stakeholder (commercial fishing) scenario.

5. What 'currency' should we use to define reference points?

Some of the above examples of reference points were in terms of spawning stock biomass, and others in terms of fishing mortality (F) or level of catch. In the management framework adopted by ICES, reference points (TRPs and LRPs) are defined in terms of both SSB and F. In the many parts of the world (e.g. Australia, New Zealand, the USA), a distinction is made between an overfished stock, measured in terms of spawning biomass, and a stock that is BEING overfished, measured in terms of fishing mortality[50]. The relationship between fishing mortality (F), stock biomass (SSB) and yield can provide the basis for discussion of most reference points. Reference Points on these variables can be set using various criteria e.g. the F which, if applied over a number of years, produces an average yield equivalent to MSY; the F which maximizes the average yield per recruit; the biomass which will produce a desired level of recruitment.

Whatever 'currency' is used, it needs to be something that is meaningful and can be estimated. In a fishery managed by effort, it makes little sense to define a reference point in terms of yield, but using F or Effort directly makes sense (though in practice, it may still be difficult to estimate the harvest rate implied by a given level of effort). With regard to estimation, studies have shown that biomass ratios (current biomass as a proportion of unexploited biomass, or as a proportion of the biomass in some year in the past) can often be much better estimated, and more robust to

[48] In theory it is not difficult to design a compatible set of reference points. However, the non-linearities in harvested systems and the potentially skewed nature of random noise and errors on estimated quantities could lead to incompatibilities between such reference points, or to objectives not being met. This can only be tested through simulation studies.

[49] SSBmsy is the average SSB expected when harvesting at Fmsy.

[50] This recognises that, even if current SSB is at a 'safe' level, but the current F is too high, the stock will tend to decline to 'dangerously low' levels unless some management action is taken. The time-frame of the decline and urgency for action will depend, amongst other things, on the life-span of the resource. Short lived species will show a faster decline and require action sooner than a very long lived species. Note that a stock can be overfished and also in the process of being overfished, i.e. when SSB is too low and F is too high.

alternative model assumptions, than absolute biomass. Biomass ratios are often unbiased, whereas absolute biomass tends to be over-estimated. It may therefore be sensible to construct a reference point in terms of a biomass ratio.

6. Candidates for reference points

It is beyond the scope of this background paper to discuss in detail the merits or disadvantages of the large range of reference points that have been developed and used in fisheries management. Instead, three broad groups of reference points are briefly considered. The first group are based on the concept of 'escapement', an approach used with squid fisheries in the Falkland Islands. The potential relevance of this for tuna-like stocks is discussed below! The second group are based on the concept of MSY, and the third group on the concept of spawner per recruit (SPR).

6.1. Reference points based on the concept of Escapement

Much of the classic stock assessment and reference point work has been done with so-called groundfish stocks in mind (e.g. cod, haddock, flatfish etc., usually taken in trawl fisheries). Most of these species have life-spans of several years, and a spawning stock consisting of several age classes. Species that only live a year or two are often not well represented. The implications of a single year's failure in recruitment can obviously have serious, if not disastrous, effects on the fishery and future viability of a short-lived stock. The 'buffering' effect of several age classes in the spawning component of a longer-lived (and not overfished!) stock makes it more robust to a single year's recruitment failure.

Many squid species are annual or bi-annual, for example the shortfin squid (*Illex argentinus)* and the longfin squid *(Loligo gahi)* harvested by squid jiggers off Argentina and the Falkland Islands. The approaches taken to develop reference points for such stocks are discussed in detail in Basson and Beddington (1993). The fishery for the shortfin squid harvests immature animals, and spawning occurs after the end of the fishing season (about March to May). After spawning, the squid die. It is clear that next year's recruits, and hence fishery, depend on the number of spawners that survive until after this year's fishery. There are obviously also other factors, such as oceanography and natural variability, which affect recruitment strength. This means that even for the same number of spawners, subsequent recruitment can be highly variable.

It would be inappropriate to manage such a fishery by catch quotas, unless stock size can be predicted with some accuracy in advance. It is more appropriate to use effort control, and this Falkland Islands squid fishery is therefore managed by an annual licensing system and a fixed season. The approach we took to decide on an appropriate level of effort was to focus on escapement. Escapement is the number, or proportion, of spawners that survive to spawn (i.e. escape the fishery). In the early years of the fishery, the level of effort was set to achieve a 40% proportional escapement[51]. This means that the spawning biomass at the end of the fishing season should be 40% of the spawning biomass that WOULD have been there under no fishing (essentially, $SSB_{\text{with fishing}}/SSB_{\text{no fishing}} = 0.4$). This was treated as a target reference point.

We recognised, however, that if it was a particularly poor year (due to whatever circumstances), then even 40% of a small number of spawners could be too low to ensure viable recruitment for the following year. As our time series of spawners and subsequent recruitment grew, we were

[51] The figure of 40% was based on conventions in other major squid fisheries, since there was no information or data on this particular stock at the start of the fishery. Studies on longer lived fish species have shown that figures around 20-30% was likely to be relatively 'safe' for many fish species. These levels would be too risky for annual species, so the figure had to be above 20-30%.

able to also define an absolute level of escapement, in terms of a tonnage SSB, to use more like a limit reference point.

This particular fishery does not have an explicit management objective to maximise yield. Instead, the focus is on maintaining SSB at a productive level and minimising the risk of stock collapse. It is informative to evaluate the likely yields, levels of effort, and associated risks for different levels of escapement in an MSE framework. In the squid fishery, we used simulation studies to ensure that the reference points (proportional escapement TRP and the absolute escapement LRP) were set in such a way that there would be a low probability of requiring early closure of the fishery, because early closures are disruptive to the industry and costly for the management agency.

Caddy & Mahon note that salmon management has classically been based on attempting to achieve a minimum escapement to spawning, and many such fisheries in western North America are managed with fixed escapement objectives.

The potential relevance to tuna- and billfish-like stocks is two-fold. First, the approach may be suitable for relatively short-lived species like skipjack. Second, the approach may be useful in cases where the harvest occurs primarily on non-spawners. In this regard it is useful to note that if the catch does not take spawners, and there is no other fishery or fishery-independent means of getting information about spawners, then it would be very difficult to operationalise a management framework with reference points based on spawning biomass. The current suite of species in the WTBF and ETBF catches (swordfish, yellowfin, bigeye) are not really good candidates for this approach, but it is worth documenting and bearing in mind as a possibility in the future.

6.2. Reference points based on the concept of Maximum Sustainable Yield

The 1982 Convention on the Law of the Sea specifies only one technical Reference Point, notably the Maximum Sustainable Yield (MSY), but adds: '...as qualified by relevant environmental and economic factors'. Caddy & Mahon comment that this is, at first sight, an obvious target for management of a single species fishery and that it was widely used for this purpose by fishery commissions in the 1960s and 1970s. Subsequent developments in the theory, and perhaps more so, practical experience in fishery management, have cast doubts on the usefulness of MSY as a safe TRP (e.g. Larkin 1977, and references below).

The distinction between, and advantages of, management by constant harvest rate versus management by constant catch also apply to reference points. Some criticisms that can be levelled at MSY as a TRP may not be applicable to F_{msy} as a TRP. Properly implemented management with F_{msy} as the TRP would, for example, allow catches to vary as the stock varies in contrast to a constant catch at MSY, but would unfortunately still not avoid the risk of overfishing.

Mace (2001) re-considers the role of MSY in management and reminds us of the basis for criticism of MSY as a management target. Criticisms of MSY can be divided into three categories (Punt and Smith, 2001): (i) estimation problems, (ii) the appropriateness of MSY as a management goal; and (iii) the ability to effectively implement harvest strategies based on MSY. Mace notes that estimation problems arise due to poor assumptions in some models, lack of contrast in data, and lack of reliability of the data. As a management goal Mace suggests that F_{msy} could be close to $F_{extinction}$, the fishing mortality that will lead to stock extinction (e.g. Cook et. al. 1997), in which case, management based on F_{msy} is unlikely to guard against recruitment failure. Second, she notes that MSY-based management ignores economic and social considerations (Larkin 1977) and it is impossible to attain MSY, or harvest at F_{msy}, for all species in an

assemblage simultaneously. The main criticism in terms of implementation relate to the tendency in stock assessments to under-estimate F and over-estimate SSB. This, together with effort creep and imperfect implementation tends to lead to management targets being overshot. Although this criticism can also be levelled at other F-based targets, the implications of several years of overshooting F_{msy} is likely to be worse than several years of overshooting a lower, more conservative, fishing mortality target.

A few more comments on estimation problems are worth noting. The classic way of estimating MSY quantities is via production models, but where information on age or size is available, MSY quantities can be estimated via an age structured model. Such models explicitly estimate spawning biomass and recruitment, and the shape of the stock-recruit relationship has a strong effect on the MSY quantities, including the biomass where MSY is achieved (B_{msy}/B_0). The problem is that unless a fishery has historically been driven to very low levels (in fact, the low levels that we want to avoid!), there is usually very little information in the stock-recruit estimates to indicate the level of productivity or density dependence. High uncertainty in this relationship then translates to high uncertainty in MSY-related quantities, including F_{msy}.

Die and Caddy (1997) reviewed a range of methods for calculating MSY, and discussed their applicability to species with high values of natural mortality. They conclude that using MSY or F_{msy} as targets causes significant problems, and alternatives based on more conservative harvest rates (like $F_{0.1}$ or F_{MBP}=fishing mortality rate leading to the Maximum Biological Production of the stock) come closer to both ecological and economic optima. They suggest that F_{msy} can be useful as an upper limit to exploitation, i.e. as an LRP.

Caddy & Mahon comment on an alternative to F_{msy}: "Doubleday (1976) postulated that fishing at the effort level that corresponded to 2/3 of the effort needed to produce MSY would allow a very large fraction (about 80%) of the MSY to be harvested with a significantly reduced risk of stock collapse. This target, although safer than F_{msy}, has been criticised, in our view unfairly, as arbitrary, empirical and insensitive to changes in recruitment. It may be useful to note that this approach to setting reference levels for F may be generalized for other commonly used population models and can be regarded as a precautionary approach to using the results of production modelling."

A concluding point from Mace (2001) is worth repeating here (my addition in italics): "I believe the use of biological reference points as indicators of overfishing (*essentially LRPs*) has been taken far more seriously and has achieved more success in ensuring rational exploitation of marine resources than the use of MSY or its proxies as fishing targets. Part of the reason is that the consequences of overfishing (exceeding biological limits) can be readily portrayed and appreciated in terms of risk to the resource, while the consequences of somewhat exceeding annual fishing targets are generally considered inconsequential and the cumulative effects of annual overshoots are underappreciated."

6.3. Reference points based on the concept of Spawner-per-Recruit

As noted above, one of the main problems associated with MSY-based quantities is the estimation. If a stock-production model is used, there is usually a built-in assumption about the relative biomass at MSY. For example, the familiar Schaefer stock-production model has B_{msy} at half the unexploited level (i.e. $B_{msy}/B_0 = 0.5$). If an age-structured model is used, the MSY-based quantities would depend heavily on the stock-recruit relationship and this is almost always highly uncertain.

Spawner-per-recruit (SPR) calculations do not require any assumption about the stock-recruit relationship. The concept of SPR is analogous to the more familiar yield-per-recruit (YPR). In each case, one starts with one recruit and projects forward for a given harvest rate (and natural mortality rate) to obtain the total expected spawning potential (or biomass), or the expected yield, over the life-time of the recruit[52]. SPR can be calculated for a wide range of harvest rates, and it is then convenient to consider these SPR values (at harvest rate F) relative to the SPR when there is no harvest, SPR_F/SPR_0. A reference point can then be based on some ratio of SPR_F/SPR_0. Given that these quantities are likely to be used in cases where a stock-recruit relationship is highly uncertain, it is common to choose SPR reference levels relatively conservatively. Note that if one chose, for example, 40% as a target for relative spawner per recruit, then this implies an associated harvest rate, say F_{40spr}, which can be used as a target reference point for the harvest rate – at least conceptually.

Looked at from the yield point of view, the YPR can also be calculated for a wide range of different harvest rates, and in most cases, there will be some harvest rate (F_{mypr}) where the yield per recruit is a maximum. One can then obviously also calculate what the SPR would be at F_{mypr}. Clearly, if this ratio is low, then F_{mypr} would not be an appropriate candidate for a target reference point. SPR and YPR calculations need to be treated with particular caution given that they essentially ignore the possibility of recruitment failure at low spawning biomass.

6.4. Other reference points

There are potentially many other ways of defining reference points, either for the performance of the fishery or the status of the stock. There may be case-specific reasons why one approach is more applicable or appropriate for one fishery than another. In the case of Southern Bluefin Tuna, for example, a spawning biomass reference was based on a historic estimate of SSB, namely, SSB in 1980. Some of the main reasons for this choice was that recruitment seemed to drop substantially after that (when SSB was lower), and in addition there seemed to be a contraction of the spatial extent of the stock distribution. Most importantly, the estimates of MSY-related quantities were highly uncertain because of the difficulties associated with estimation of the stock-recruit relationship. More detail is given in Polacheck (2003) and Basson and Polacheck (2004). This is simply mentioned here as an example of an alternative approach to the more familiar ones outlined above.

7. Implementation

Choosing reference points

Much has been written about choosing reference points, and sections above have already noted some considerations. Most important in my view is the need to first consider the management objectives and then choose reference points accordingly. This approach is more likely to lead to consideration of not just biologically based reference points, but also ones which take economic factors into consideration. A word of caution, however, economic reference points on their own are unlikely to be a 'safe' option. In the spirit of the precautionary approach, it would be necessary to include a limit reference point related to the resource to minimise the chances of low spawning biomass or stock collapse.

[52] Conceptually it is sometimes easier to think of the calculation as being done by starting with 100 (say) recruits, calculating the total spawning potential or yield over their lifetime and then, at the end, dividing by 100 to get 'per recruit' quantities.

The choice should, however, take into account how feasible it is to estimate and implement the reference points for the particular fishery. Uncertainty in estimation of (a) the reference point itself and (b) the quantity which will be compared to the reference point should also be considered. Recall the comment about ratios of SSB which may be more robust than absolute levels of SSB.

Calculating/estimating reference point values

In well studied fisheries with good historical data it is likely that there will be at least one model-based TRP, and possibly several LRPs; some model-based, others empirical. For data poor fisheries, it may be necessary to adopt TRPs and LRPs, which are simple criteria based on experience derived from other similar fisheries or from generalisations about many fisheries.

In the first instance, model-based reference points may have to be implemented without simulation testing, but there is enormous value in using the MSE approach, or even much simpler simulation trials to test the likely performance of reference points within a management framework. This approach would allow us to evaluate the effects of choosing some fraction of F_{msy} as a target reference point, and therefore quantify WHAT fraction of F_{msy} would meet management objectives in a less arbitrary way than picking, say, $2/3F_{msy}$.

The difficulties associated with estimating MSY-based reference points has already been noted. The potential for using SPR-based reference points, which do not make assumptions about the stock-recruit relationship (or productivity) and are therefore simpler to calculate, was also noted with some caveats.

Implementing reference points

Ideally, reference points should be relatively stable quantities. Model-based reference points can, however, turn out to be highly unstable – i.e. vary greatly from one assessment to the next - if the associated assessment is (for example) still based on a short time-series of data, or is sensitive to model assumptions. If the reference point itself varies greatly then the system could switch far too often between the states of being in the 'safe' zone to being in the 'danger' zone. This will clearly have a knock-on effect on the fishery. The likely stability of assessments should therefore be taken into account before model-based reference points are implemented. In this regard it is also worth noting that one can design a management system where the assessment is not redone annually, but on a longer time-scale, for example, every second, or every third year.

In its most common implementation, the target reference point in a quota managed fishery defines how to determine the annual quota if the state of the fishery and resource is such that no LRPs are triggered (i.e. it is not in the area to be avoided). If specified in terms yield, that yield becomes the quota, and the quota will therefore be constant over time unless an LRP is triggered. If specified in terms of a fishing mortality, then that is usually used with an estimate of current biomass to calculate the quota. In this case the quota will vary from year to year tracking fluctuations in the biomass.

It is important to remember that an effort managed fishery has the advantage that the realised catch is likely to track fluctuations in the biomass. A constant effort policy does not suffer from some of the problems that a constant catch policy suffers from. One could therefore argue that the implementation of a target reference point in an effort managed fishery need not lead to annual adjustments or changes to effort. An alternative may be only to adjust effort if the effective fishing mortality is a certain percentage outside the target fishing mortality. Recall that fishing mortality is the combination of catchability or efficiency and effort. If a target reference point is specified in terms of effort, it is therefore necessary to include a statement about the efficiency associated with it, and to carefully monitor efficiency.

As already noted, there should be a pre-specified action, or process for arriving at action, when an LRP is triggered. This may be defined in terms of a 'rebuilding plan' with associated time-frame, and there may be additional management measures that are invoked when the resource is in the 'red' area.

There is another important factor which is more evident when talking about LRPs than about TRPs: uncertainty. Let's say we have an LRP which says that SSB should not be below 20,000t. If it is, the catch or effort needs to be reduced by 50% for that year. Let's also say that we've done an assessment and the point-estimate of current SSB is 19,990t. Should this case lead to the same response, a 50% cut, as one where the point estimate of SSB is 15,000t? Intuitively one would consider that the 15,000t case is worse than the 19,990t case. One way of dealing with this is to re-define the LRP to be something like this instead: the probability that SSB is below 20,000t should be less than 50%[53] (i.e. if it's <50% then the LRP is NOT triggered, but if it's 50% or more, then it IS triggered). Now we also need to look at the uncertainty around those point estimates (and this is good practice in any case!). If the confidence interval around the point estimate is not symmetric, we may even find that the LRP is NOT triggered for the 19,990t case.

There are many different ways of dealing with uncertainty in this context, and the above example is merely intended to raise awareness and emphasise the need to explicitly think about how uncertainty will interact with reference points and their implementation.

8. Management procedures, Decision Rules and Reference points

Finally, a few words about management procedures and reference points. The CCSBT is in the process of developing a management procedure (MP) for Southern Bluefin Tuna, and some of the lessons learnt from that process are relevant here. A management procedure is a bit like a management plan, but it has some components which a management plan might not have. It is essentially an agreed set of rules that define management action (e.g. setting quotas or effort levels) in the future, based on changes in the stock status with a goal to achieve specified management objectives. The rules need to include specification of the data that will be used to estimate changes in stock status and the mathematical procedure, also called the decision rule, used to calculated what the quota or effort will be given the agreed these data. The rules should also specify what happens when there are exceptional circumstances, for example, when a LRP is triggered.

One of the main components of a management procedure is the so-called decision rule which determines how the quota or effort should be set. A very simple example of a decision rule for an effort managed fishery could simply be to set effort at a level E_{normal}, unless the spawning biomass is below SSB_{min}, in which case set effort at E_{min}. In this context, one could think of E_{normal} as acting like a target reference point. The way of finding the appropriate level of E_{normal}, would however tend to be via a simulation study which mimics the resource and fishery dynamics, as well as year-to year implementation of the management procedure. The key quantities of interest, for example, average yield, average SSB, lowest SSB, number of times that E_{min} is invoked are output for different values of inputs such as E_{normal}, SSB_{min} etc., and also for different, but plausible, assumptions about the underlying dynamics of the stock and fishery. This then allows us to find the reference points which best meet our management objectives, and

[53] It is more common to define LRPs so that there is a very low probability, e.g. 5% or 10% that SSB is below some level, e.g. 20,000t. In such cases, the LRP could be triggered even if the point estimate is well above the 20,000t, but the associated uncertainty is great.

to quantify clearly the trade-offs between them (loss in yield for reduced risk of low recruitment, say).

A very important characteristic of this approach is that the aim is NOT to find an optimal management strategy, but rather to find a strategy, including the decision rule and reference points, which is robust to the uncertainties about the resource, fishery, data, our ability to estimate quantities, and likely future dynamics.

There are many advantages of the MP approach over the more traditional annual assessment approach, but what is relevant here is the advantages regarding choice and estimation of reference points. The first point to make is that reference points sit very naturally in any management procedure framework. In some cases it is not at first glance obvious exactly what component is playing the role of a reference point, but any well designed MP will have implicit or explicit target and limit reference points. Second, there are major advantages when working in the MSE (Management strategy evaluation) framework because alternative reference points can be evaluated, their performance compared, and the MP can be fine-tuned to best meet management objectives and ensure robustness to uncertainty.

References (relevant to Appendix 2)

Basson, M. and Beddington, J.R. 1993. Risks and Uncertainties in the Management of a Single-Cohort Squid Fishery: the Falkland Islands Illex fishery as an example. p253-259. In S.J. Smith, J.J. Hunt and D.Rivard [Ed.] Risk evaluation and biological reference points for fisheries management. Can. Spec. Publ. Fish. Aquat. Sci. 120.

Basson, M. and T. Polacheck. 2004. Further consideration of issues related to setting rebuilding objectives for southern bluefin tuna in the context of management procedures. Working paper to the CCSBT Scientific Committee meeting; CCSBT-ESC/0409/22.

Caddy, J.F. and Mahon, R. 1995. Reference points for fishery management. FAO Fisheries Technical Paper No. 347, 82pp.

Cooke, R.M., Sinclair, A. and Stefansson, G. 1997. Potential collapse of North Sea Cod stocks. Nature 385, p521-522.

Die, D.J., Caddy, J.F. 1997. Sustainable yield indicators from biomass: are they appropriate reference points for use in tropical fisheries? Fisheries Research 32, p69-79.

Doubleday, W.G. (1976). Environmental fluctuations and fisheries management. ICNAF Sel. Pap. 1: 141–50.

Larkin, P.A. 1977 An epitaph for the concept of maximum sustainable yield. Transactions of the American Fisheries Society V 106, p1-11.

Mace, P.M. 2001. A new role for MSY in single-species and ecosystem approaches to fisheries stock assessment and management. Fish and Fisheries, Vol.2, p2-32.

Polacheck, T. 2003. Issues related to setting rebuilding objectives for southern bluefin tuna. Working paper to the CCSBT Scientific Committee meeting; CCSBT-ESC/0309/30.

Punt, A.E. and Smith, A.D.M. 2001. The gospel of maximum sustainable yield in fisheries management: birth, crucifixion and reincarnation. In: Conservation of Exploited Species (Eds. J.D. Reynolds, G.M. Mace, K.R. Redford and J.R. Robinson. Cambridge University Press, UK.

Contact Us

Phone: 1300 363 400
+61 3 9545 2176

Email: enquiries@csiro.au

Web: www.csiro.au

Your CSIRO

Australia is founding its future on science and innovation. Its national science agency, CSIRO, is a powerhouse of ideas, technologies and skills for building prosperity, growth, health and sustainability. It serves governments, industries, business and communities across the nation.