Albacore MSE

Parsimony in Operating Model Designs

Laurence Kell
21/07/2014

Outline

For each of the Indian Ocean and North Atlantic Albacore Operating Model 1440 scenarios were conditioned on fisheries dependent data using Stock Syntheses. The design includes a variety of factors for difficult to estimate parampers such as the steepnes of the stock recruitment relationship and natural mortality as well as choices related to data weighting (**Table 1**). When conducting MSE full factorial designs have commonly been used, i.e. all interactions are run. In scientific experiments, however, often fractional factorial designs, consisting of a carefully chosen subset (fraction) of the experimental runs from the full factorial design, are prfered. The subset is chosen so as to exploit the sparsity-of-effects principle to expose information about the most important features of the problem studied, while using a fraction of the effort of a full factorial design in terms of experimental runs and resources.

In the case of Albacore many of the OMs had similar characteristics and not all interactions were important. Rather than running all the proposed OM scenarios when conducting an MSE a more efficient design would be to run the main effects and only the important interactions. Or in other words if OMs share the same characteristics then we dont need to duplicate them. Therefore first we summarise the 2880 OMs with respect to current status (F/F_{MSY} , SSB/B_{MSY}), population growth rate (r), carrying capacity (K), shape of the production function (B_{MSY}/K), MSY , time to recovery (i.e. number of years to recover to $B_{MSY}, this can be negative if $B>B_{MSY}$) and the variability in the time series.

Based on an analysis of the characteristics of the OM, it is possible to postulate for a given MP which OMs will provide similar performance. If the MPs are then evaluated for the full factorial design the predictions can be evaluated. If similar OMs provide result in similar performance for an MP we dont need to run all 3000. Having 2 stocks allows us to evaluate the generality of our analysis

Other questions

- MSE is about feedback control, therefore the variability in the time series is more important. I.e. we need to look in the frequency and not the time domain. We can do this by computing the transfer functions.

- Another question (i.e. a 2nd paper) is does the performance depend on the MP, i.e. if you use an MP based on VPA, a biomass dynamic model, or an empirical HCR do the important characteristics of the OM change.

Operating Model Structure

- The most important factor is selection pattern followed by M. I.e. 1st two terms are the main effects for Sel and M. Then interactions become important, i.e. steepnesss is only important for double normal selection (i.e. dome shaped) with low M. For flat topped selection patterns then Sample size is important.

Methods

Although estimates of F/F_{MSY} and S/B_{MSY} showed a wide range of values points appeared to be grouped and within a group points appeared to be correlated. Therefore to explore the structure of the estimates and reduce the dimensionality model based clustering and pitch classification was used to automatically chooses the covariance structure. This allows both the number of groups and the covariance across groups to vary. Selection of the groups is based on maximization of the Bayesian Information Criterion (BIC). Although prior data can be included for more accurate estimation, this was not done in this case which is primarily an exploratory analysis.

Once the groups are chosen it is necessary to explain the cause of the difference between the clusters, to do this CHi-squared Automatical Interaction Detection (CHAID) was used. CHAID is a decision tree technique for nominal dependent variables (Kass, 1980) that detects interactions between categorical variables. CHAID uses an algorithm for recursive partitioning based on maximising the significance of a chi-squared statistic for cross-tabulations between the dependent variable and the predictors at each partition. The data are partitioned into mutually exclusive, exhaustive subsets that best describe the dependent variable. In this case the dependent variable was the group derived above and the predictors were the scenarios (i.e. factors and levels of the treatments).

Results

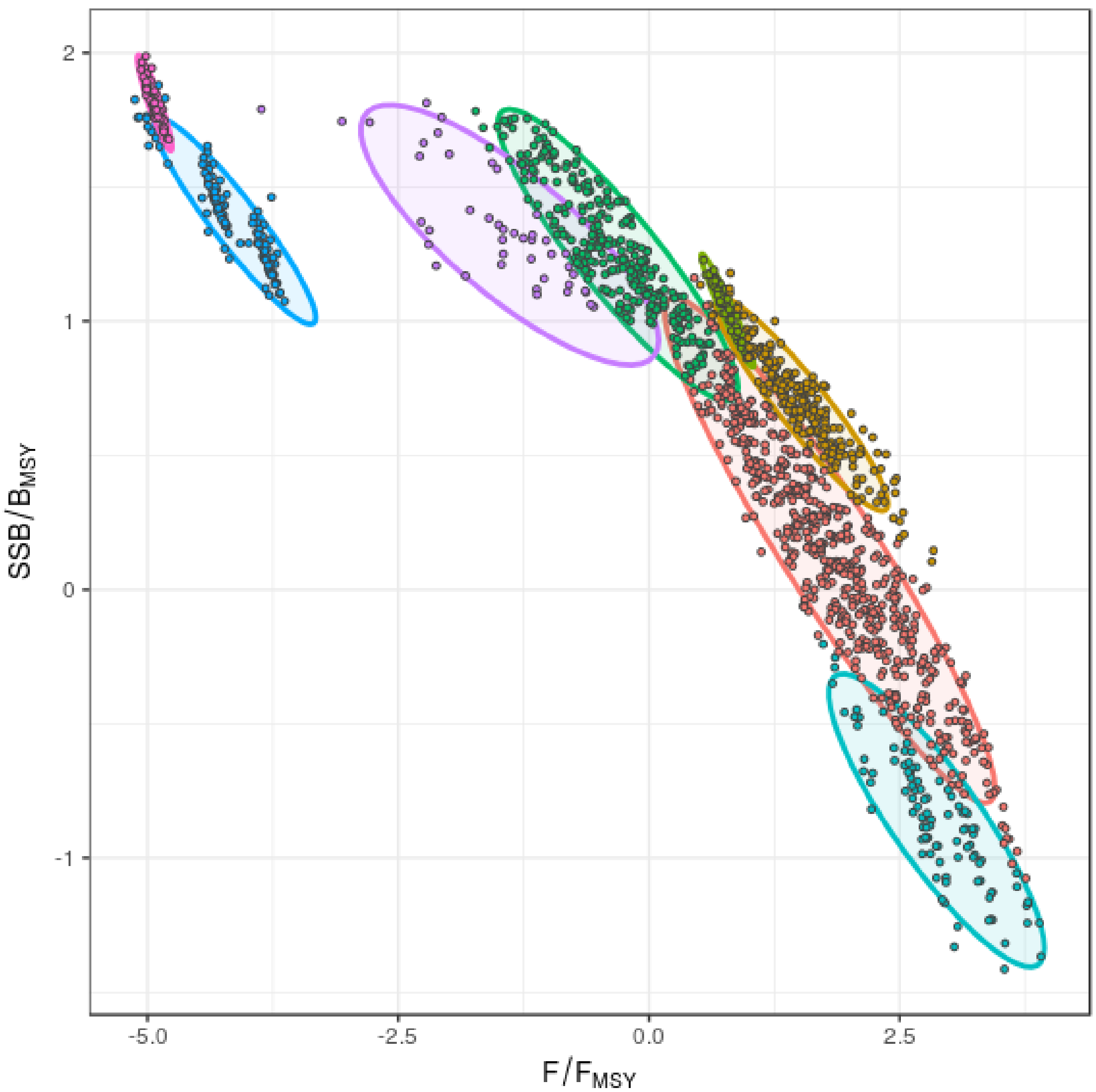


Figure 1. Scatter plot of $SSB:B_{MSY}$ $F:F_{MSY}$ for 1440 Indiian Ocean OM Scenarios, ellipses are clusters with common covariance structure.

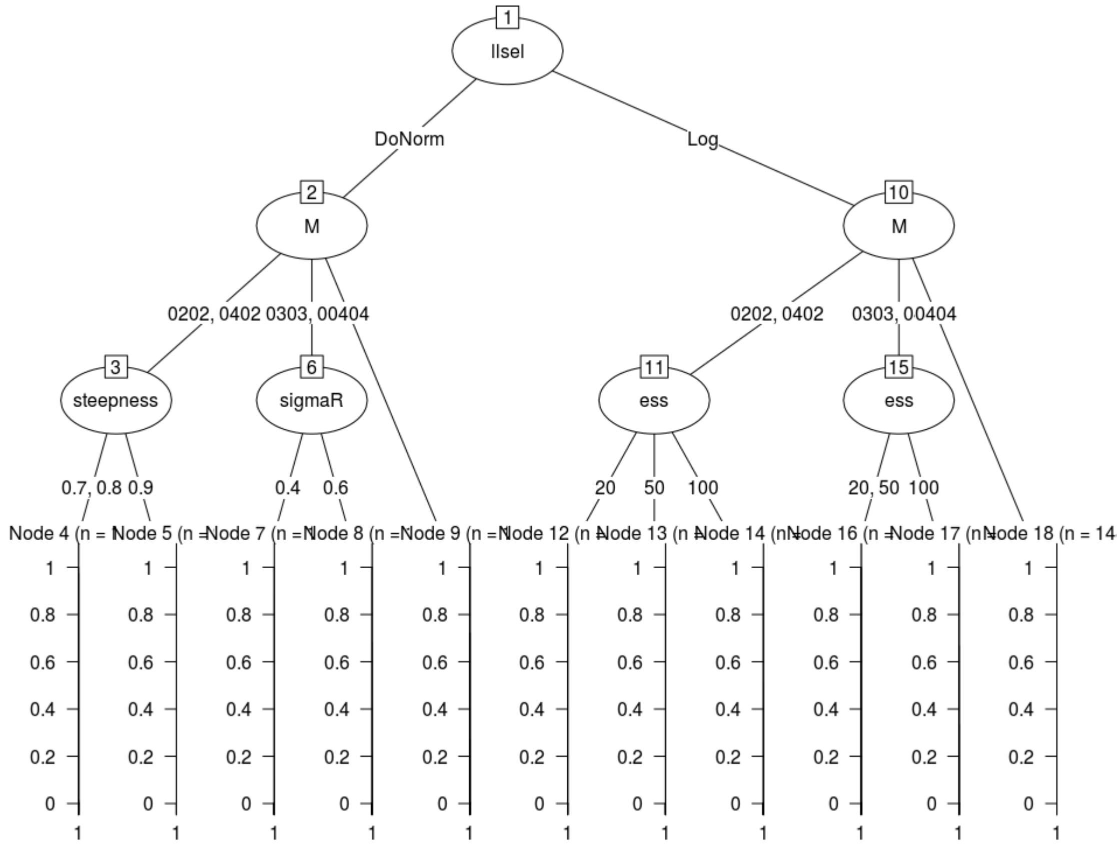
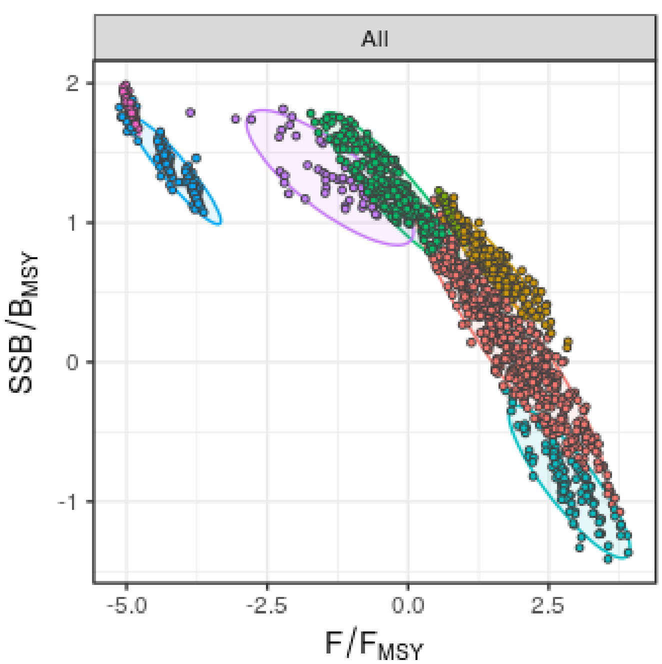
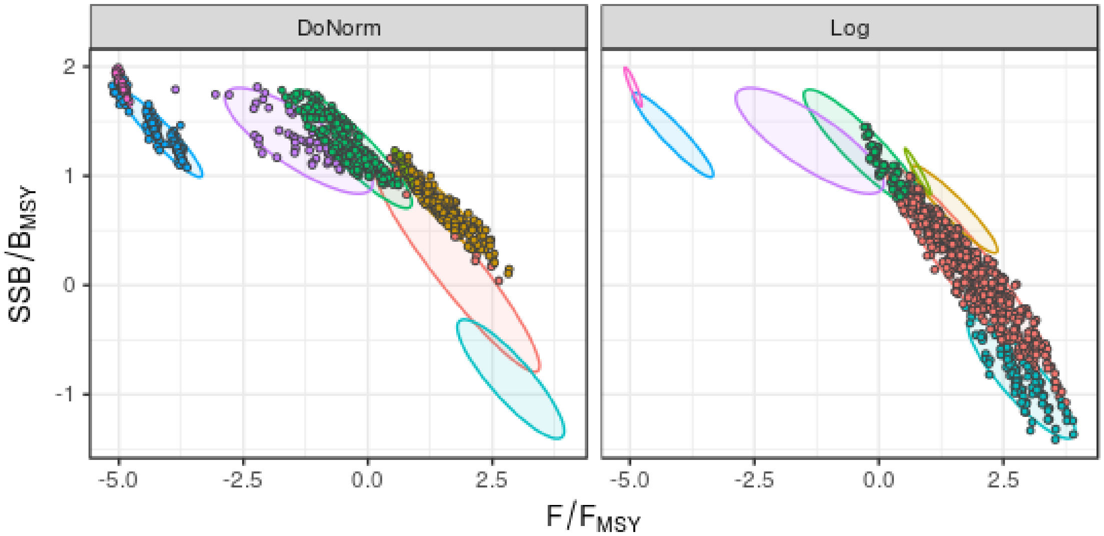
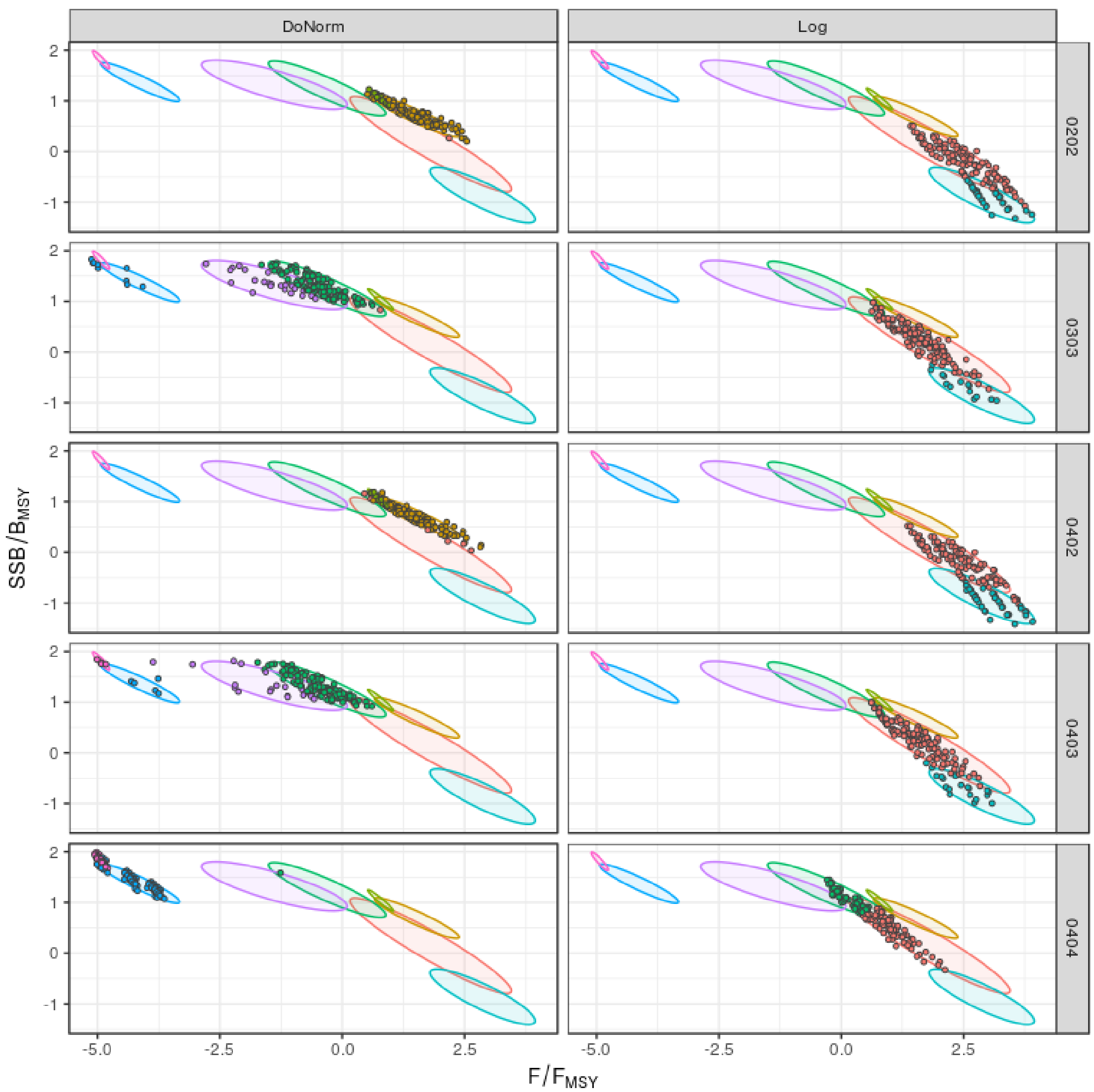


Figure 2. Decision tree that explains the cause of the differences between the clusters, using CHi-squared Automatical Interaction Detection





DoNorm
Log
SSB/B$_{MSY}$
F/F$_{MSY}$



DoNorm
Log
0202
0303
0402
0403
0404
SSB/B$_{MSY}$
F/F$_{MSY}$

Figure 3. Bivariate plots of F/F_{MSY} against S/B_{MSY}, showing the clusters split by two levels.

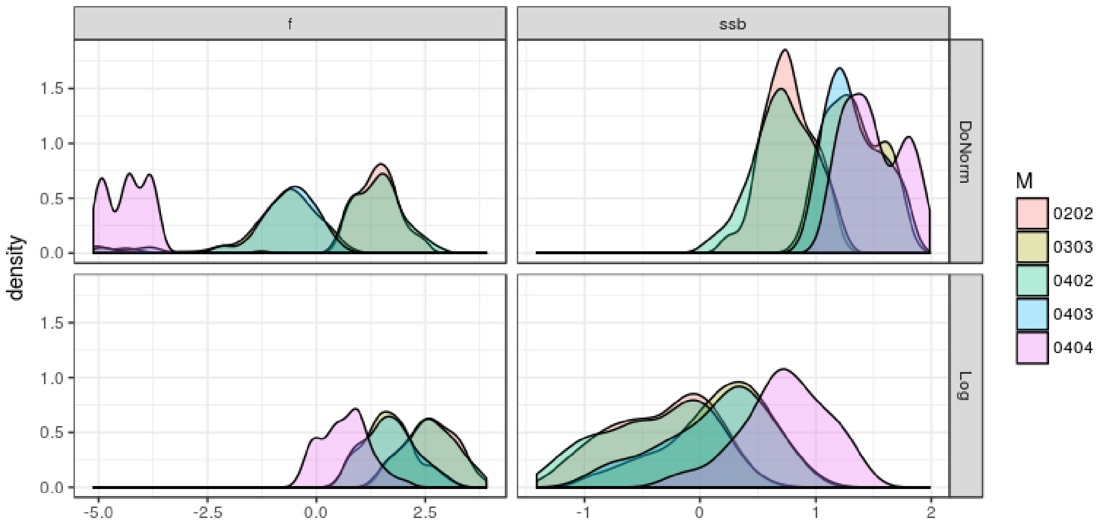


Figure 4. Univariate plots of F/F_{MSY} against S/B_{MSY}, showing the clusters split by two levels.

References

G. V. Kass (1980). An Exploratory Technique for Investigating Large Quantities of Categorical Data. Applied Statistics, 29(2), 119–127.