A HIERARCHICAL FRAMEWORK FOR ANALYSING MULTIPLE INDICES; A WHITE MARLIN EXAMPLE

Laurence T. Kell[1], and Paul De Bruyn[1]

SUMMARY

In this paper we conduct a Bayesian hierarchical analysis of CPUE series. We do this in order to conduct an exploratory data analysis, i.e. to identify indices with similar trends.

RÉSUMÉ

Dans ce document, nous réalisons une analyse hiérarchique bayésienne des séries de CPUE. Ceci est effectué afin de réaliser une analyse exploratoire de données, c.-à-d. en vue d'identifier les indices présentant des tendances similaires.

RESUMEN

En este documento se realiza un análisis jerárquico bayesiano de las series de CPUE: Esto se realiza con miras a realizar un análisis exploratorio de datos, es decir, para identificar índices con tendencias similares.

KEYWORDS

Bayesian, Stock assessment, CPUE

1. Introduction

The Stock Assessment Methods Working Group (WGSAM) recommended that work be conducted on diagnostics and ways to combine multiple indices. Therefore, in this paper we conduct a Bayesian hierarchical analysis of CPUE series using the method of Hilborn and Conn (2009) to identify population trends from multiple, noisy indices.

The approach also allows differences between indices to be evaluated. These may be due to a variety of causes, e.g. random error (in either measurements or processes), changes in abundance or bias due to changes in population and fleet dynamics.

2. Material and Methods

The data used were the standardised indices of catch per unit effort (CPUE) used in the ASPIC assessment Prager (1994) for white marlin. An implicit assumption is that each index is measuring the same quantity (i.e., relative abundance) but is also subject to process error (attributable to variation in catchability, spatial distribution, etc.) an estimable level of within-survey variance (i.e., sampling or measurement error).

A lognormal error structure is often assumed for indices of abundance, adopting this convention and assuming that indices are subject to independent and multiplicative process and sampling errors, equation (1). Where $(\sigma p)2$ and $(\sigma s)2$ give the standard deviation associated with process and sampling errors, respectively. q_{it} is the catchability of index i in year t, N_t is vulnerable biomass or abundance in year t.

Assuming a lognormal error structure, the precision attributable to sampling error may be written as a function of the estimated coefficient of variation on the absolute scale using equation (2). Making the substitutions in

[1] CCAT Secretariat, C/Corazón de María, 8. 28002 Madrid, Spain.

equations (3) and (4) gives equations (5). If catchability is assumed to be stationary (e.g., no trends in catchability), the model may be further simplified to equation (6).

All analyses were conducted using R; i.e. the R2jags was used for fitting using Monte Carlo Markov Chain (MCMC) for simulation and coda for goodness of fit diagnostics.

2.1 Diagnostics

One of the main difficulties with MCMC methods is ensuring that the simulations have converged to a stationary distribution. The equilibrium distribution of the chain is the required posterior distribution but how do we know that the chain has reached equilibrium? A burn-in period where initial values are discarded helps. However, in complicated cases, i.e., where there is more than one local maximum or the posterior distribution is in the form of a ridge a chain, it can take a long time to move around the parameter space and a long burn-in period may be required. While a much larger sample may have to be taken to ensure that the chain has not just become temporarily stuck in one part of the parameter space.

For these reasons we use a variety of diagnostics to check convergence, see Best et al. (1995), e.g.

- Trace Plots are time series plot of a parameter as the Markov chain proceeds.
- Autocorrelation Plots measure the correlation between μt and $\mu t+1$ variable in a chain
- Correlation Plots can show if parameters are confounded
- Gelman-Rubin Diagnostic tests that the burn-in is adequate and requires that multiple starting points be used.
- Geweke Diagnostic, if burn-in is adequate, then the mean of the posterior distribution of μ from the first half of the chain should equal the mean from the second half of the chain.

2.2 Autocorrelation

Autocorrelation plots of the parameters of interested can help in deciding the level of thinning to apply to a chain in order to reduce correlation between successive values to an acceptable level (e.g., 0.2). The JAGS script can then be rerun using an appropriate thinning value.

2.3 Confounding

Model parameter might not be identifiable, either because a parameter is confounded with one or more other parameters or because the data are inadequate. Plotting the correlations between parameters is a way of seeing if there is confounding between parameters

2.4 Gelman-Rubin

If the chain has reached convergence, the Gelman-Rubin Gelman and Rubin [1992] test statistic R 1, then it can be concluded that the burn-in is adequate. While values above 1.05 indicate lack of convergence, since the distribution of R under the null hypothesis is essentially an F distribution. However the F-test for comparing two variances is not robust to violations of normality. Therefore using Gelman-Rubin diagnostic alone is not sufficient to ensure convergence

2.5 Geweke Diagnostic

For a parameter of interest, if the burn-in is adequate, the mean of the posterior distribution from the first half of the chain should equal the mean from the second half of the chain Geweke (1991). The Geweke statistic asymptotically has a standard normal distribution, so if the values from R are outside -2.5 or 2.5, this indicates non stationary of the chain and that burn-in is not sufficient.

3. Results

The indices of abundance are shown in **Figure 1,** points are the observed index values and the blue a lowess fit to the points by index. The red line is general additive model (GAM) fitted with a smooth term for year and a

categorical value by fleet. This allows a common year effective to be estimated and all indices to be plotted on a common scale.

Pairwise scatter plots of the indices of abundance are shown in **Figure 2**, blue lines are linear regressions fitted to the points, the shade area is the standard error of predicted means and the red line is the mean of the points on the y-axis. This shows that not all indices are correlated and that problems may be expected when fitting.

The presence of autocorrelation in the parameters was evaluated by plotting the autocorrelation function (ACF) by lag in **Figure 3** for each parameter. After a lag of 30 the ACF is 0.2 and so for thinning every 40th value in a chain was saved. The correlation between parameters is plotted in **Figure 4**; blue indicates a positive correlation and red negative. It can be seen that there is a positive correlation between the first 6 parameters, i.e, catchability and a negative correlation between catchability and the estimate of the expected values by year.

The Gewanke statistic is plotted by parameter in **Figure 5**, in only one case is a parameter greater than 2.5. The Gelman-Rubin statistic was 1 for all parameters, showing that approximate convergence had been achieved.

Figure 6 compares the combined time series (black line), where shaded areas represent the 95th and 50th probability regions, to the index fitted using the GAM. While **Figure 7** compares the individual indices (red lines) to the combined index

4. Discussion and Conclusions

A variety of diagnostics were used to check for convergence and confounding of parameters. These showed that there were problems with convergence, probably due to the fact that the indices were not all positively correlated.

Although a combined index was fitted using the hierarchical framework, the fact that the original indices were not all correlated implied that they may have be proxies of different stock components or ages and that some indices might be biased due to changes in targeting of the fleets or in management regulations, particularly since all indices are fishery dependent..

Schnute and Hilborn (1993) pointed out that fisheries stock assessments some- times prove, in retrospect, to be wrong, particularly acute when more than one data source is available and different data sets provide contradictory parameter estimates. Especially as traditional methods of stock assessment involve weighted averages of the contradictory data, and these generally produce parameter estimates intermediate to those obtained from the data sets individually. They also demonstrate that, when model or data errors are considered, the most likely parameter values are not intermediary to conflicting values; instead, they occur at one of the apparent extremes.

The Working Group on Stock Assessment Methods recommended that the method be evaluated before being used to generate a combined index in stock assessment. This will be best done through simulation, for example using cross-validation.

References

N. Best, M. K. Cowles, and K. Vines. 1955, Coda* convergence diagnosis and output analysis software for gibbs sampling output version 0.30. *MRC Biostatistics Unit, Cambridg*.

A. Gelman and D. B. Rubin, 1992, Inference from iterative simulation using multiple sequences. *Statistical science*, pages 457–472.

J. Geweke, 1991, Efficient simulation from the multivariate normal and student-t distributions subject to linear constraints and the evaluation of constraint probabilities. In *Computing science and statistics: Proceedings of the 23rd symposium on the interface*, pages 571–578. Citeseer.

R. Hilborn and P.B., 209, Conn. Hierarchical analysis of multiple noisy abundance indices. *Canadian Journal of Fisheries and Aquatic Sciences*, 67(1):108–120.

M. Prager, 1994, A suite of extensions to a nonequilibrium surplus-production model. *Fishery Bulletin*, 92(2): 374–389.

J. T. Schnute and R. Hilborn, 1993, Analysis of contradictory data sources in fish stock assessment. *Canadian Journal of Fisheries and Aquatic Sciences*, 50(9):1916–1923.
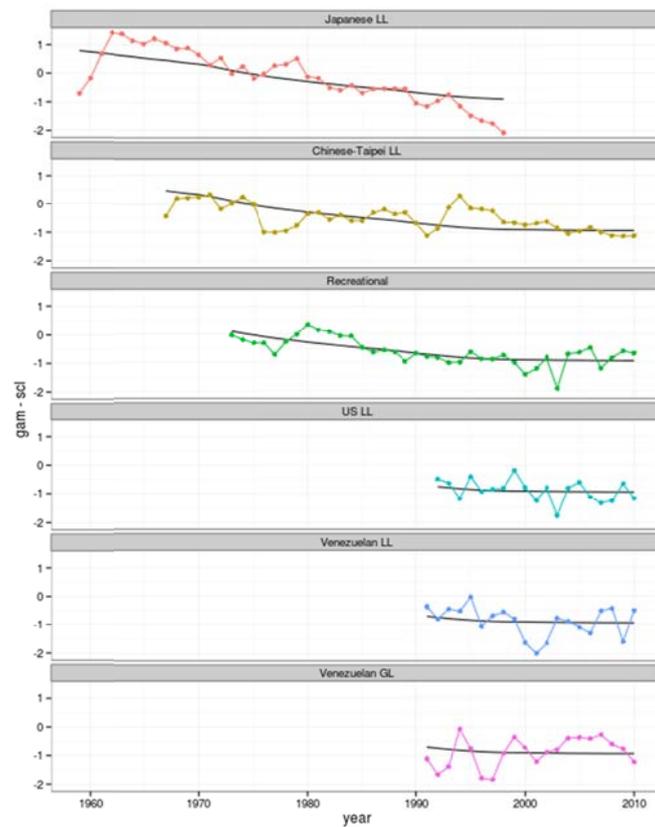
Figure 1. Plot of indices of abundance, points are the observed index values and the blue a lowes fit to the points by index. The red line is GAM fitted to lo(year) and fleet.
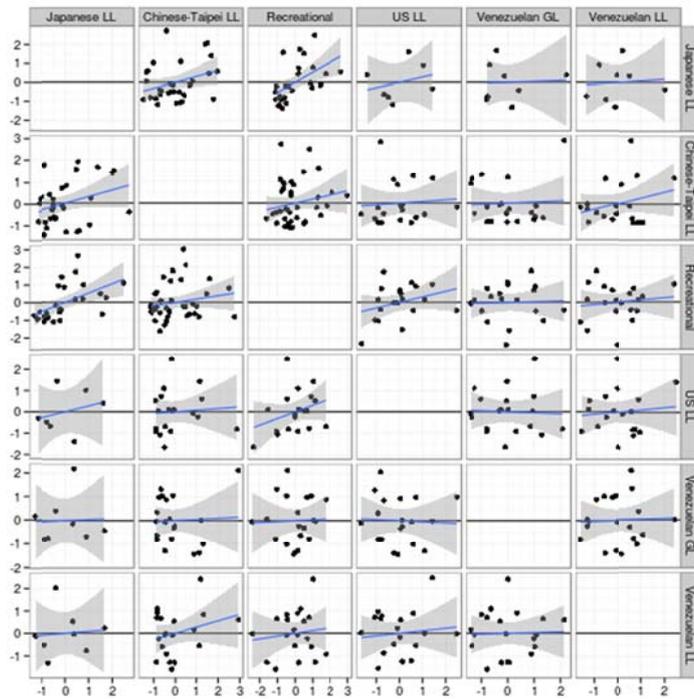
Figure 2. Pairwise scatter plots of the indices of abundance, blue lines are linear regressions fitted to the points, the shade area is the standard error of predicted means and the red line is the mean of the points on the y-axis.
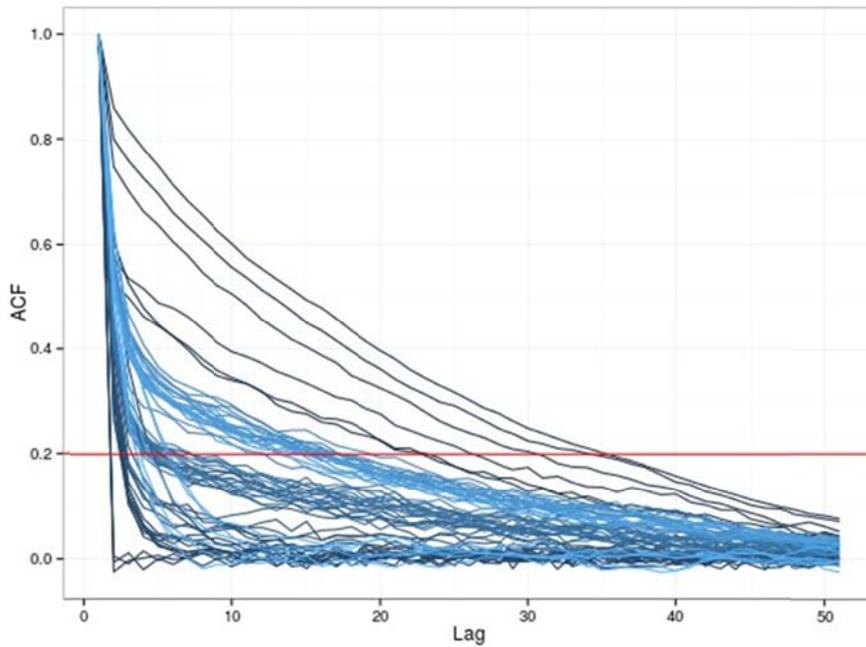
Figure 3. Autocorrelation in parameters by lag, each line represents a single parameter.
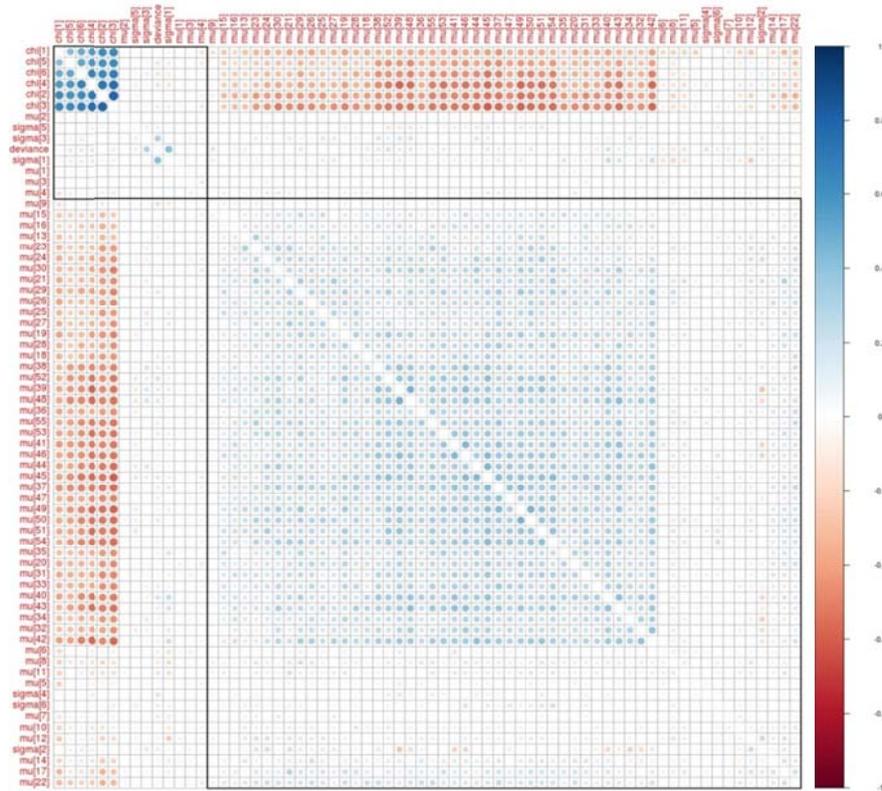
Figure 4. A plot of the correlation matrix for the parameters, blue indicate a positive correlation and red negative.
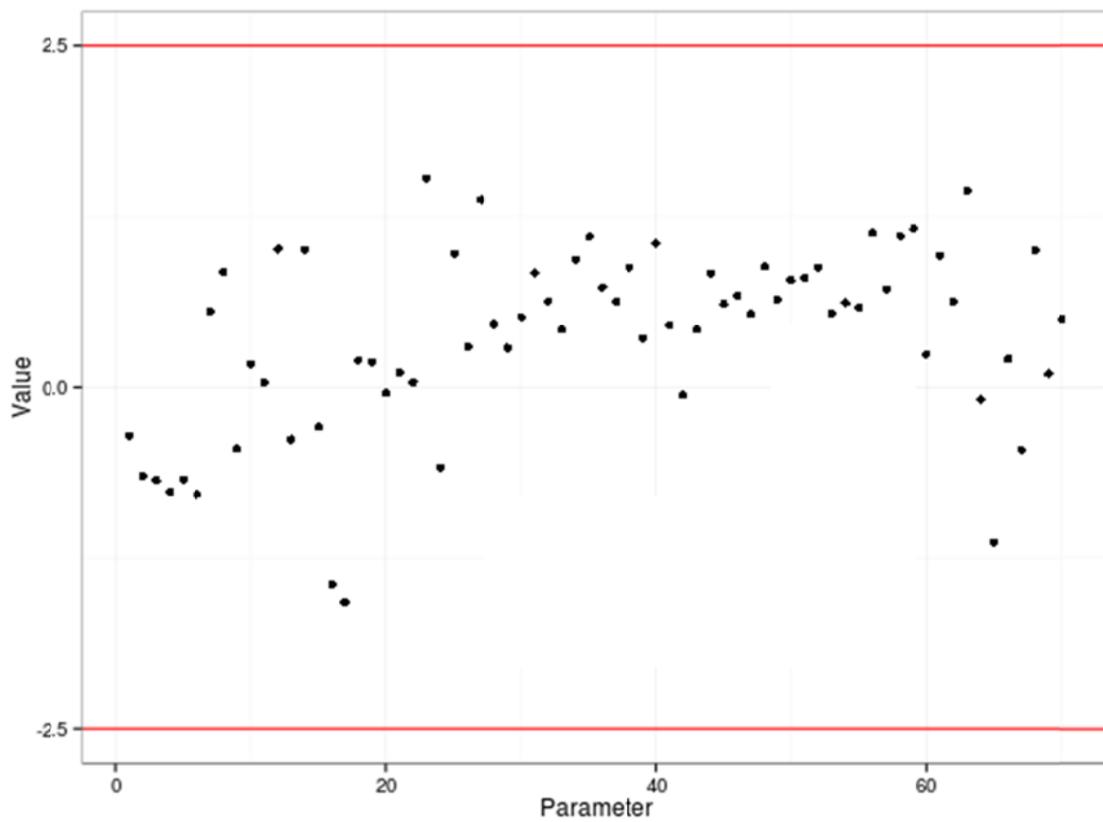
Figure 5. Plot of Gewanke statistic for all parameters.

1294
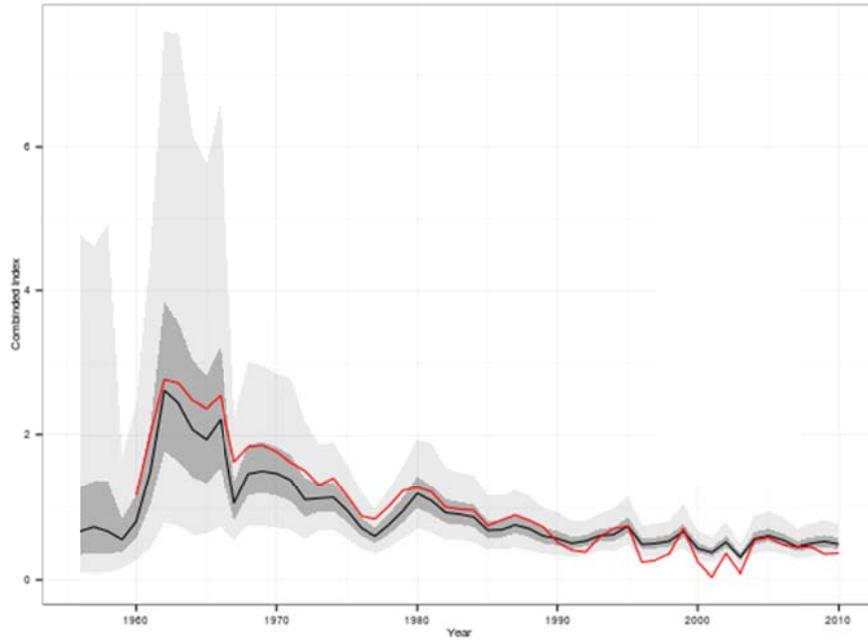
Figure 6. A comparison of combined index fitted using a GAM (red line) to the combined time series using the Bayesain analysis (black line), shaded areas respresent the 95th and 50th probability regions.
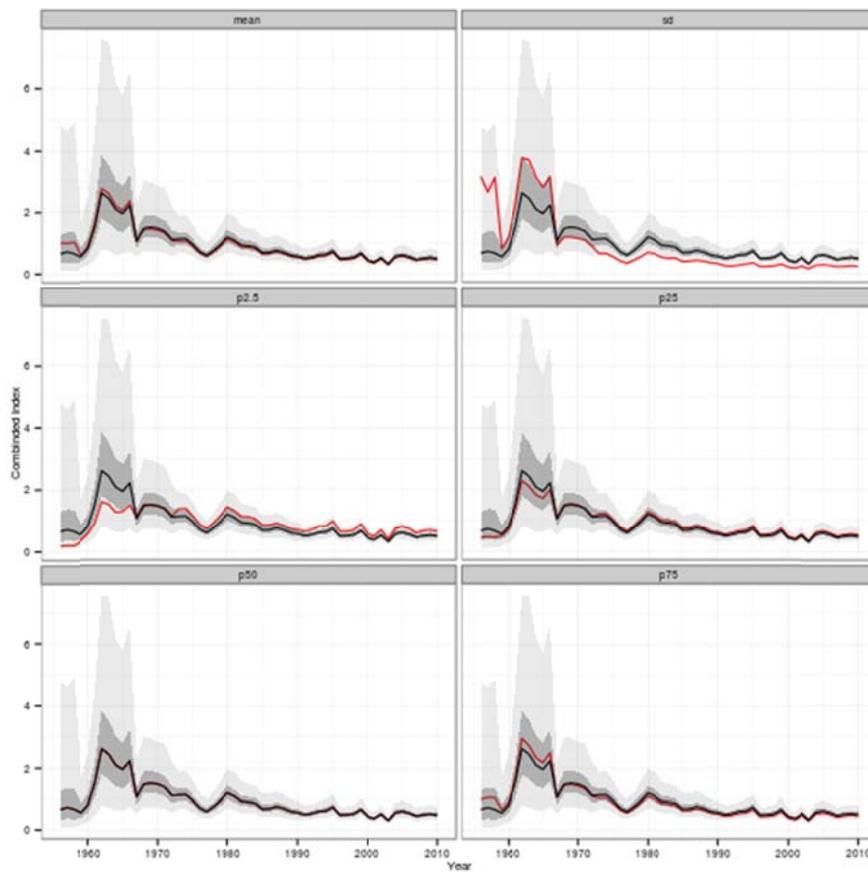
Figure 7. Time series of observed CPUE indices (red line) compared to the combined time series (black line), shaded areas represent the 95th and 50th probability regions.

1295