 Collect. Vol. Sci. Pap. ICCAT, 74(6): 3037-3051 (2018)

AN EXPLORATORY DATA ANALYSIS OF THE EAST ATLANTIC BLUEFIN STOCK ASSESSMENT DATASET

L.T. Kell[1], A. Ben Mhamed[2], T. Rouyer[3] and A. Kimoto[4]

SUMMARY

An exploratory data analysis is conducted of the Eastern Atlantic and Mediterranean bluefin dataset prepared for the Virtual Population Analysis. These data include the catch-at-age of the whole stock, catch per unit effort and their partial catches. The analysis explored correlations and conflicts between the CPUE series, the selection patterns of the main fleets and fishing mortality of the terminal ages the main parameter estimated by VPA. The consequences for developing scenarios for use in the assessment are discussed.

RÉSUMÉ

Une analyse exploratoire des données est menée en ce qui concerne le jeu de données du thon rouge de l'Atlantique Est et de la Méditerranée préparé pour l'analyse de population virtuelle. Ces données incluent la prise par âge de l'ensemble du stock, la capture par unité d'effort et les prises partielles. L'analyse a exploré les corrélations et les conflits entre la série de CPUE, les schémas de sélection des principales flottilles et la mortalité par pêche des âges terminaux, le paramètre principal estimé par la VPA. Les conséquences de l'élaboration de scénarios à des fins d'utilisation dans l'évaluation sont discutées.

RESUMEN

Se llevó a cabo un análisis exploratorio del conjunto de datos de atún rojo del Atlántico este y Mediterráneo preparado para el análisis de población virtual. Estos datos incluyen la captura por edad de todo el stock, la captura por unidad de esfuerzo y sus capturas parciales. El análisis exploraba las correlaciones y los conflictos entre las series de CPUE, los patrones de selección de las principales flotas y la mortalidad por pesca de las edades terminales, el principal parámetro estimado por el VPA. Se discuten las consecuencias de desarrollar escenarios para utilizarlos en la evaluación.

KEYWORDS

Catch Curve Analysis, Catch Per Unit Effort,
F Ratio, Selectivity, Stock Assessment,

[1] ICCAT Secretariat, C/Corazón de María, 8. 28002 Madrid, Spain; Laurie.Kell@iccat.int; Phone: +34 914 165 600; Fax: +34 914 152 612.
[2] INRH, Boulevard Sidi Abderrahman Ain Diab 2; a.benmhamed@mail.com
[3] Tristan Rouyer, Ifremer Sète, UMR MARBEC, Avenue Jean Monnet 34200 Sète
[4] NRIFSF;5-7-1 Orido Shimizu Shizuoka Japan

Introduction

Recent bluefin assessments have been based on continuity, which involve keeping the same scenarios (i.e. assessment choices and settings) and updating the last assessment by adding new years of data before conducting Virtual Population Analysis. In the 2017 assessment, however, historical as well as recent years of data have been updated, biological parameters changed, management measure have resulted in changes in selection pattern and the relationships between catch, effort and stock abundance, and five different stock assessment models will be applied. Therefore it is helpful to conduct an exploratory data analysis of the main datasets, e.g. catch per unit effort (CPUE) and catch-at-age data, to help in agree in advance how hypotheses will be tested and how models will be validated.

Material and Methods

Material

The data are those used of the VPA-2box base case, i.e. as found in the input data files.

These contain two main types of data namely i) the standardised CPUE indices for the fleets used for calibration, their partial catches and mass-at-age for indices based on biomass, and ii) the catch-at-age (CAA) for the entire stock.

Methods

The CPUE indices are assumed to represent an index of relative abundance. However, they represent different fleets, which may operate in different areas and seasons from each other, target different age classes, and cover different historical periods. If the CPUE series are not correlated or negatively correlated with each other this may mean that there are problems in fitting, e.g. the maximum likelihood is not well defined or that the residuals are not independent and identically distribute (I.I.D). Therefore we look at the correlations between series, the age composition, differences in selection patterns between series, and changes in selection patterns over time.

A catch curve analysis is conducted for the catch-at-age from the stock to explore changes in selection pattern over time (Jensen, A.L., Millar, 1992). When fitting stock assessment models the selection pattern of the terminal ages, e.g. in recent years and the ratio between F in the last true age and F of the plus group is of particular importance although often to estimate in practice. While changes in selection patterns are important for determining maximum sustainable reference points.

Results

The CPUE time series are plotted in **Figures 1 and 2** to compare trends by stock, in **Figure 2** a lowess smoother is fitted to year using a general additive model (GAM).

To look at deviations from the overall trends the residuals from the fits are compared in **Figure 3**. This allows conflicts between indices (e.g. highlighted by patterns in the residuals), autocorrelation within indices which may be due to year-class effects or the importance of factors not included in the standardisation of the CPUE to be identified.

Next the correlation between the indices was evaluated for the Northern Indices in **Figure 4**, the lower triangle show the pairwise scatter plots between the indices with a regression line, the upper triangle the correlation coefficients and the diagonal the range of observations. A single influential point may cause a strong spurious correlation there it is important to look at the plots as well as the correlation coefficients. Also a strong correlation could be found by chance if two series only overlap for a few years.

If indices represent the same stock components then it is reasonable to expect them to be correlated, if indices are not correlated or negatively correlated, i.e. they show conflicting trends, this may result in poor fits to the data and bias in the estimates. Therefore the correlations can be used to select groups that represent a common hypotheses about the evolution of the stock (ICCAT 2016, 2017). **Figure 5** shows the results from a hierarchical cluster analysis using a set of dissimilarities.

Next the cross-correlations are plotted in **Figure 6**, i.e. the correlations between series when they are lagged (i.e. by -10 to 10 years). The diagonals show the autocorrelations as an index is lagged against itself.

Figure 7 shows the length frequencies by age of the partial catches and **Figure 8** the mass frequencies for the CPUE series. While **Figure 9** shows the selection patterns by CPUE series derived from a catch curve analysis, and **Figure 10** that of the entire catch.

The estimated selectivity of age 9 and ages 10 and older are plotted in **Figure 11** and their ratio in **Figure 12**.

The CPUE time series are plotted in **Figures 1 and 2** to compare trends by stock, in **Figure 2** a lowess smoother is fitted to year using a general additive model (GAM).

To look at deviations from the overall trends the residuals from the fits are compared in **figure 3**. This allows conflicts between indices (e.g. highlighted by patterns in the residuals), autocorrelation within indices which may be due to year-class effects or the importance of factors not included in the standardisation of the CPUE to be identified.

Next the correlation between the indices was evaluated for the Northern Indices in **Figure 4**, the lower triangle show the pairwise scatter plots between the indices with a regression line, the upper triangle the correlation coefficients and the diagonal the range of observations. A single influential point may cause a strong spurious correlation there it is important to look at the plots as well as the correlation coefficients. Also a strong correlation could be found by chance if two series only overlap for a few years.

If indices represent the same stock components then it is reasonable to expect them to be correlated, if indices are not correlated or negatively correlated, i.e. they show conflicting trends, this may result in poor fits to the data and bias in the estimates. Therefore it the correlations can be used to select groups that represent a common hypotheses about the evolution of the stock (ICCAT 2016, 2017). **Figure 5** shows the results from a hierarchical cluster analysis using a set of dissimilarities.

Next the cross-correlations are plotted in **Figure 6**, i.e. the correlations between series when they are lagged (i.e. by -10 to 10 years). The diagonals show the autocorrelations as an index is lagged against itself.

Figure 7 shows the length frequencies by age of the partial catches and **Figure 8** the mass frequencies for the CPUE series. While **Figure 9** shows the selection patterns by CPUE series derived from a catch curve analysis, and **Figure 10** that of the entire catch.

The estimated selectivity of age 9 and ages 10 and older are plotted in **Figure 11** and their ratio in **Figure 12**.

All analysis was conducted using R and FLR and the *diags* package which provides a set of common methods for reading these data into R, plotting and summarising them (http://www.flr-project.org/).

Discussion

The exploratory data analysis showed that the most recent CPUE series are only for a short time period. These indices will be important in providing estimates of current stock status and reference points, however since the samples are small there will be sensitive to abnormal points and the predictive ability of all models needs to be fully evaluated (SCRS/2017/2017).

The aerial survey appeared to be correlated with the Japanese long line indices and the early Spanish bait boat indices, although again only a few points are available for the aerial survey index.

There appear to have been large changes in the stock selection pattern over time, i.e. fisheries originally targeting smaller and larger individuals, which has now changed with highest selectivity been found on ages 4 to 9.

The F ratio (i.e. ratio of fishing mortality at age 9 to 10) obtained from the catch curve analysis is very different from the fixed values used in the last assessment.

Conclusions

- There appear to be conflicts in the CPUE series, i.e. indices are not correlated with each other, which may result in residuals not being IID. This may make it difficult to fit the VPA to data and create bias in the estimates.
- Alternative scenarios based on different sets of CPUE indices could be evaluated
- The most recent CPUE series are only for a limited number of years and the small sample sizes may result in biased and unstable estimates of current and projected stock status. It will be important to validate any models using methods based on prediction residuals
- The estimates of the F ratio is very different from the fixed values used in the last assessment.

References

Jensen, A.L. 1985. Comparison of catch curve methods for estimation of mortality. Trans. Amer. Fish. Soc. 114: 743-747.

Millar, R.B. 1992. Estimating the size selectivity of fishing gear by conditioning on the total catch. J. Amer. Stat. Assoc. 87:962-968.

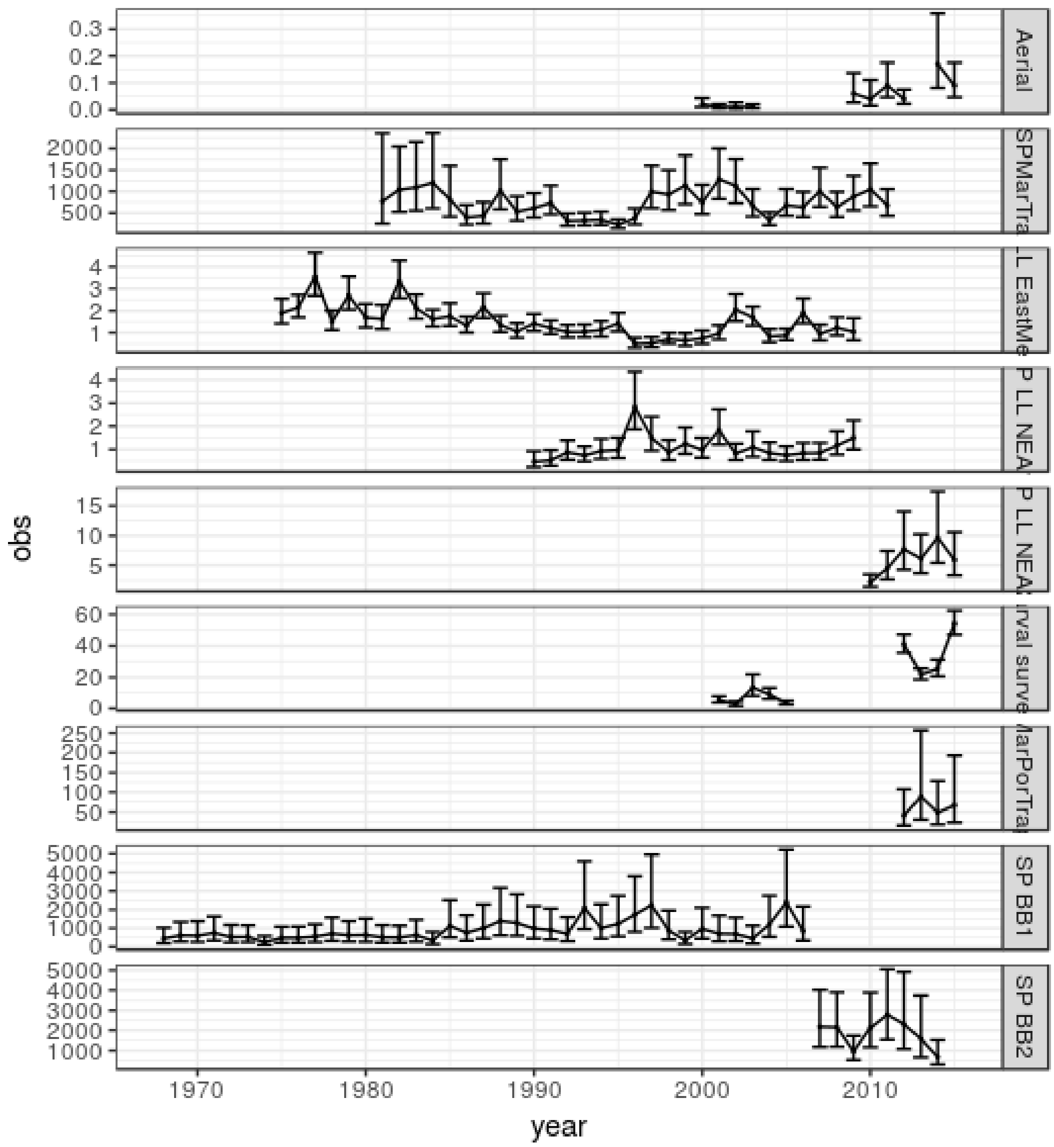

Figure 1. Time series of CPUE indices with 95% confidence intervals.

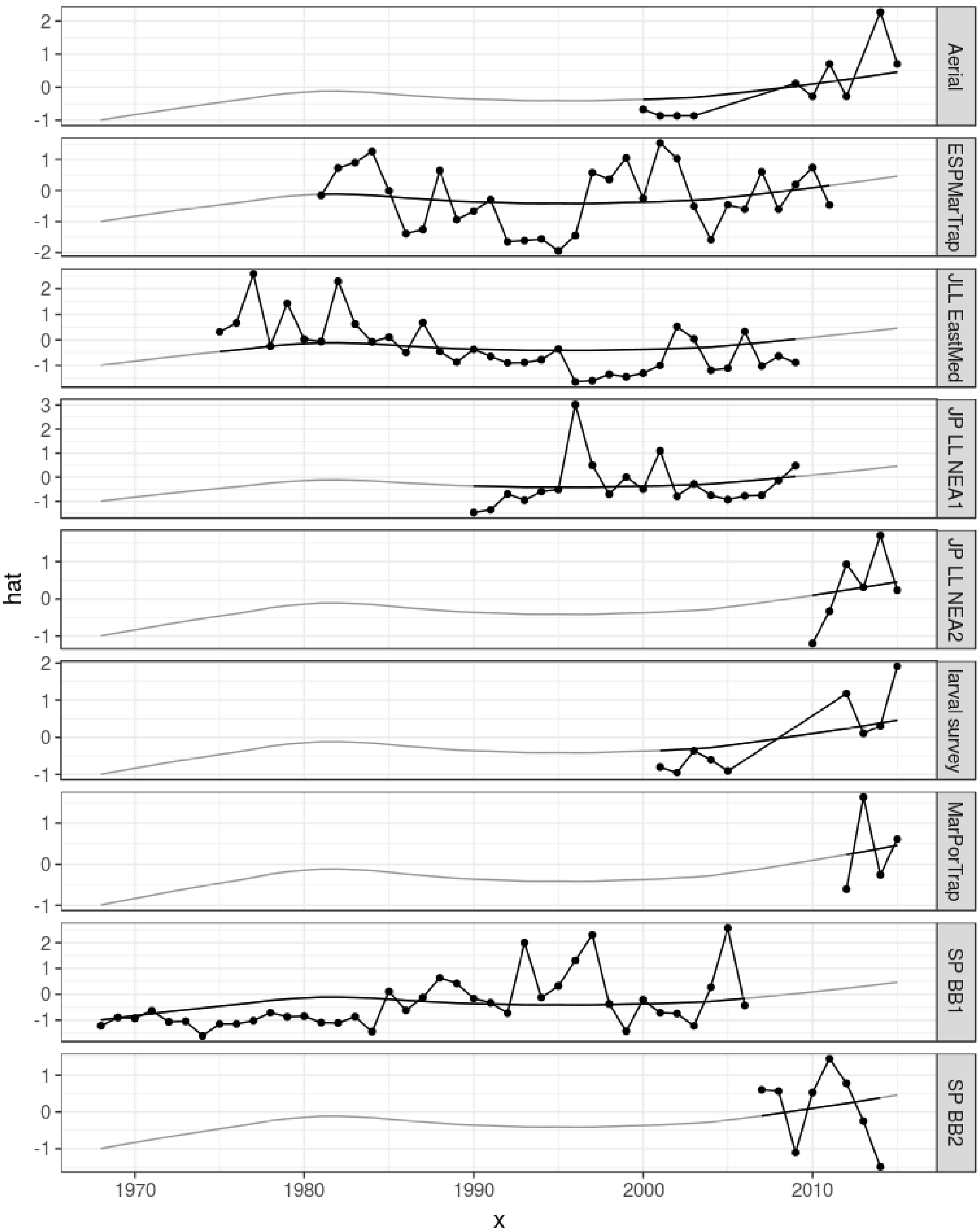

Figure 2. Time series of CPUE indices, continuous black line is a lowess smother showing the average trend by area (i.e. fitted to year for each area with series as a factor)

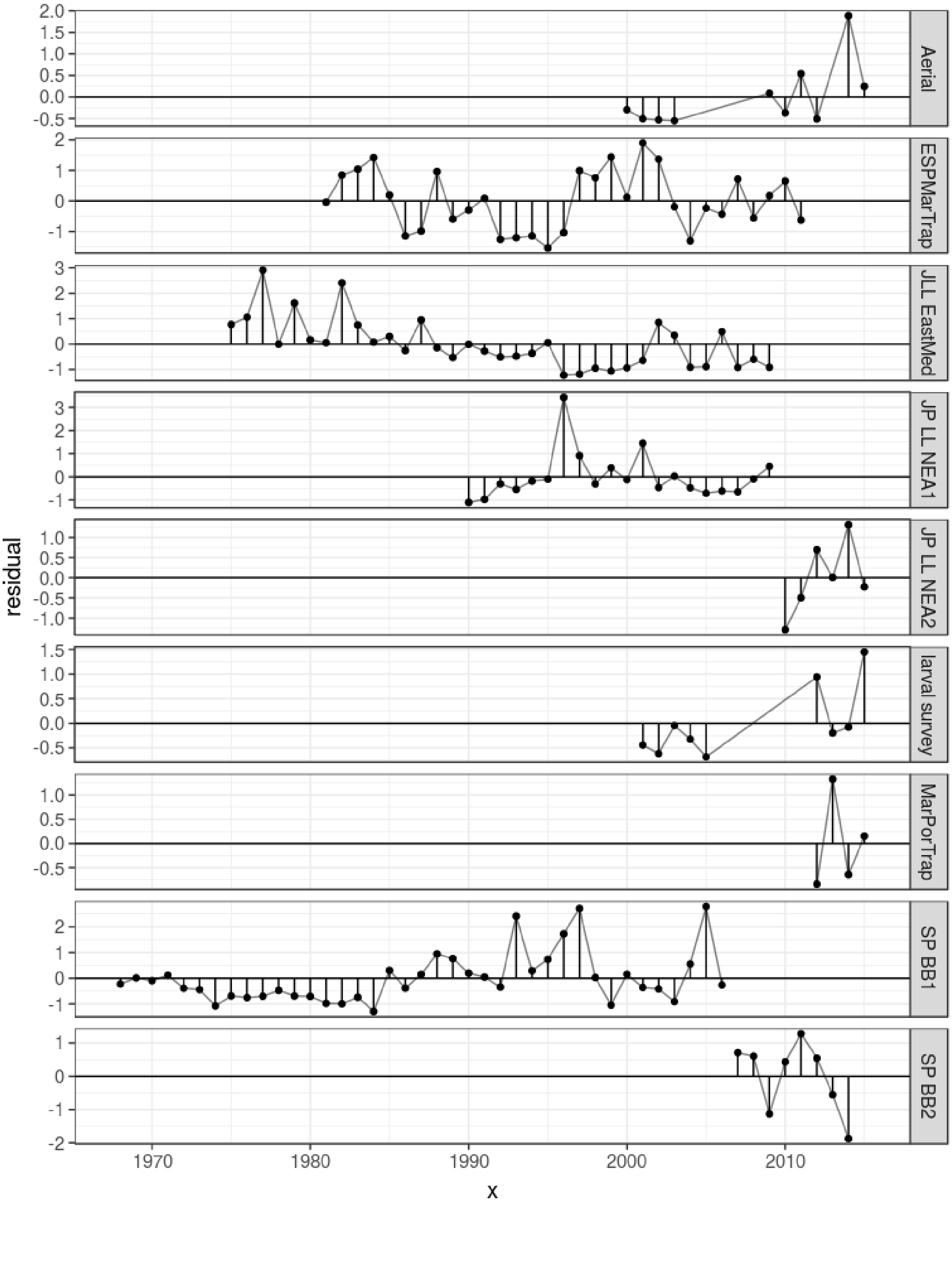

Figure 3. Time series of residuals from the lowess fit.

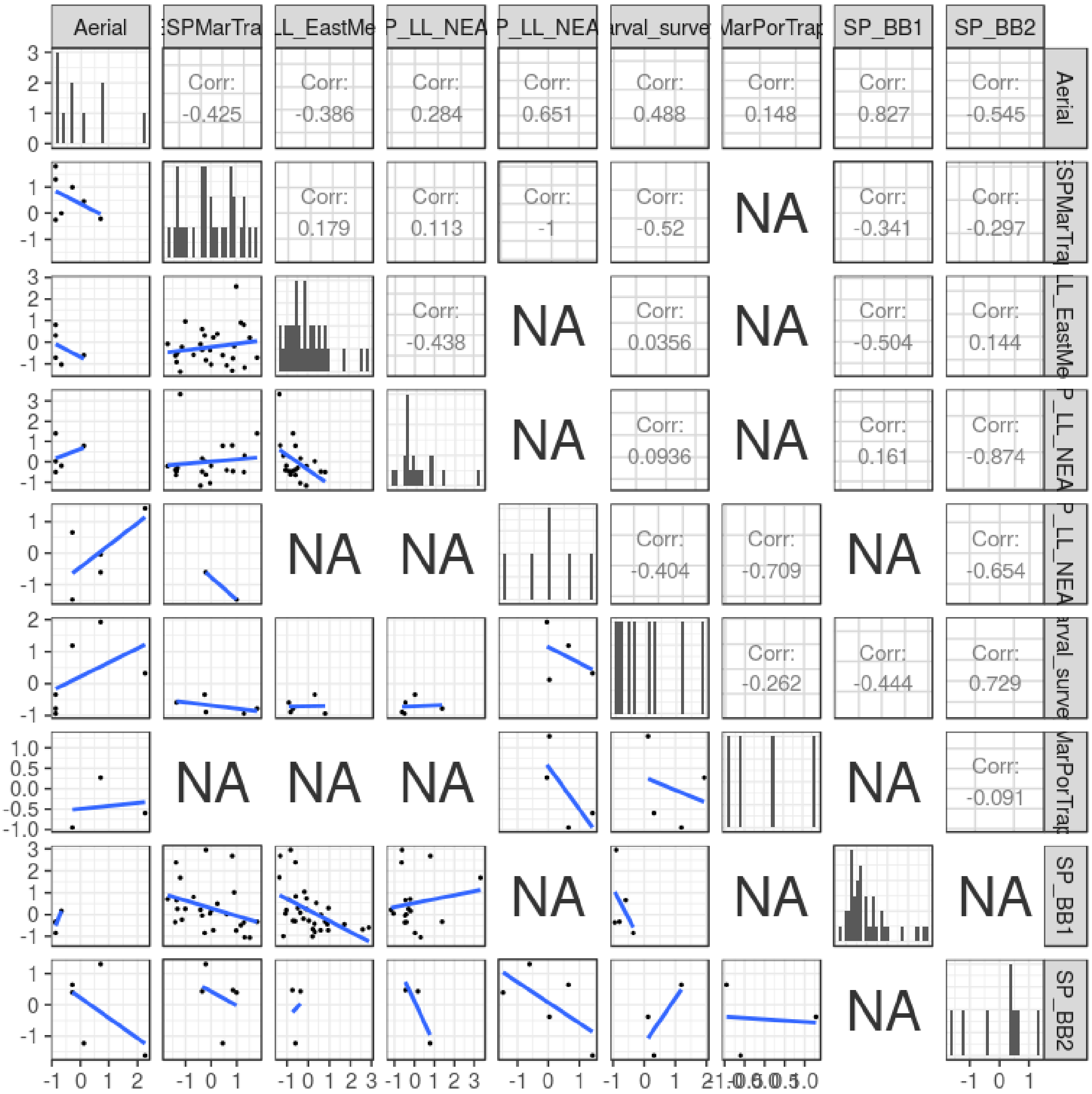

Figure 4. Pairwise scatter plots to look at correlations between Indices, North.

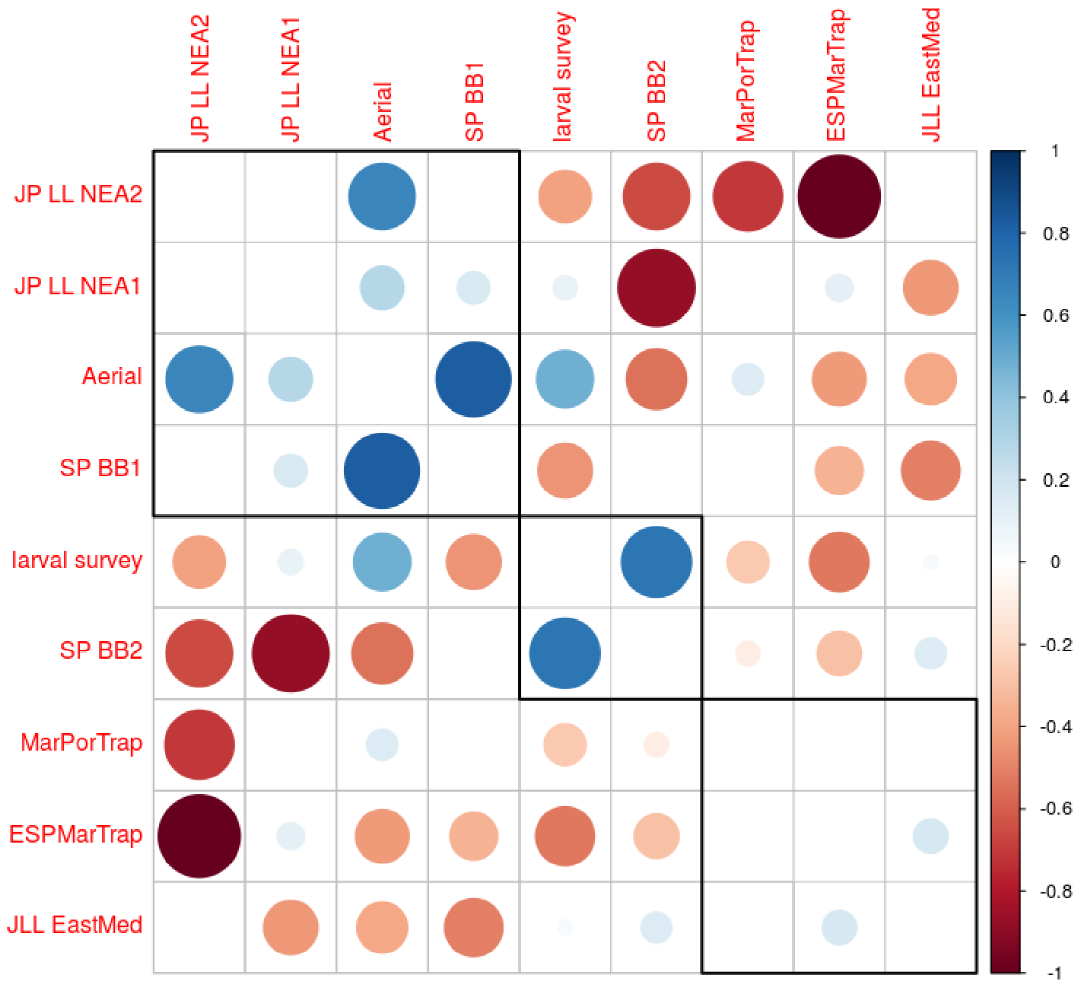

Figure 5. Plot of the correlation matrix for the Southern CPUE indices, blue indicate a positive correlation and red negative. The order of the indices and the rectangular boxes are chosen based on a hierarchical cluster analysis using a set of dissimilarities for the indices being clustered.

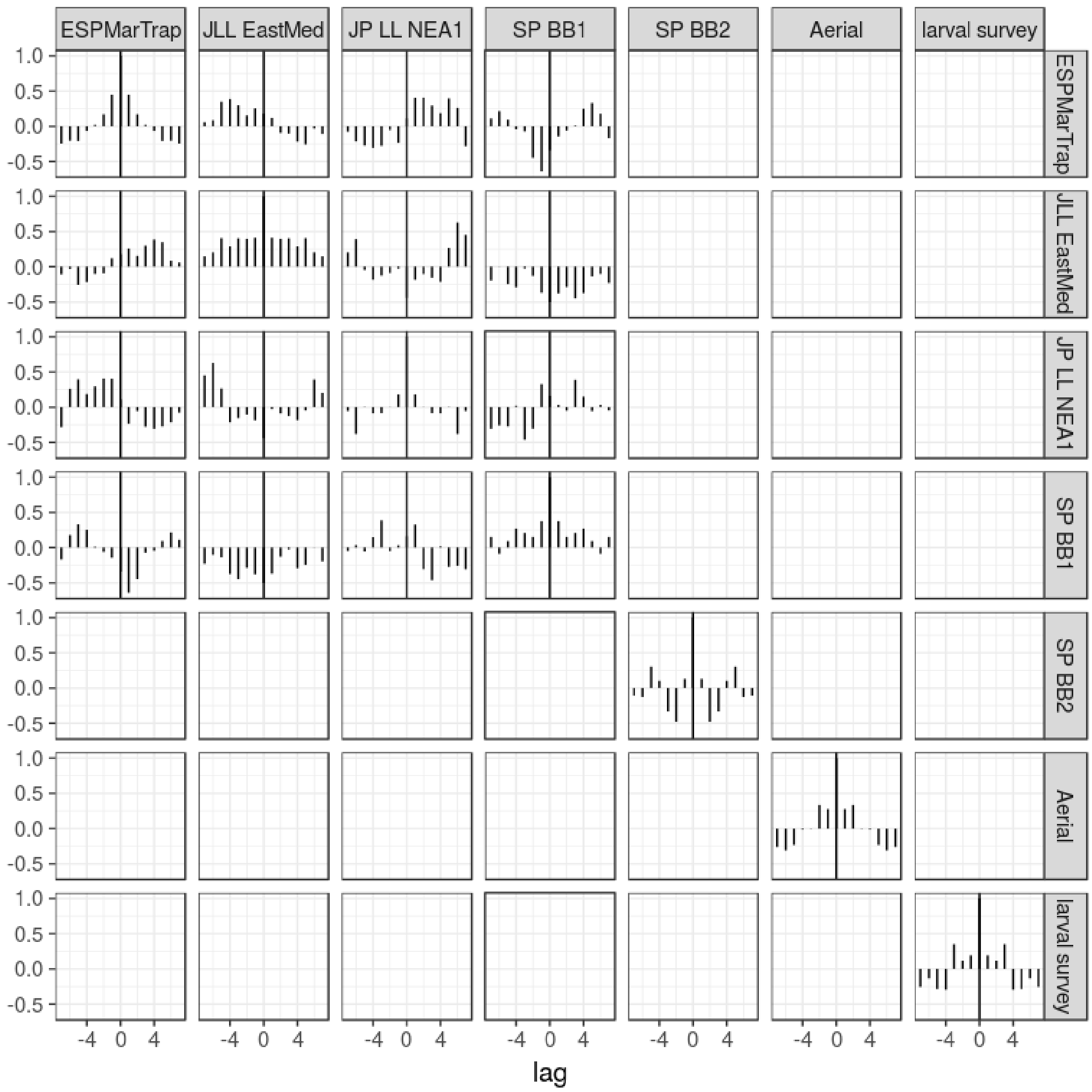

Figure 6. Cross correlations between Northern indices, to identify potential lags due to year-class effects.

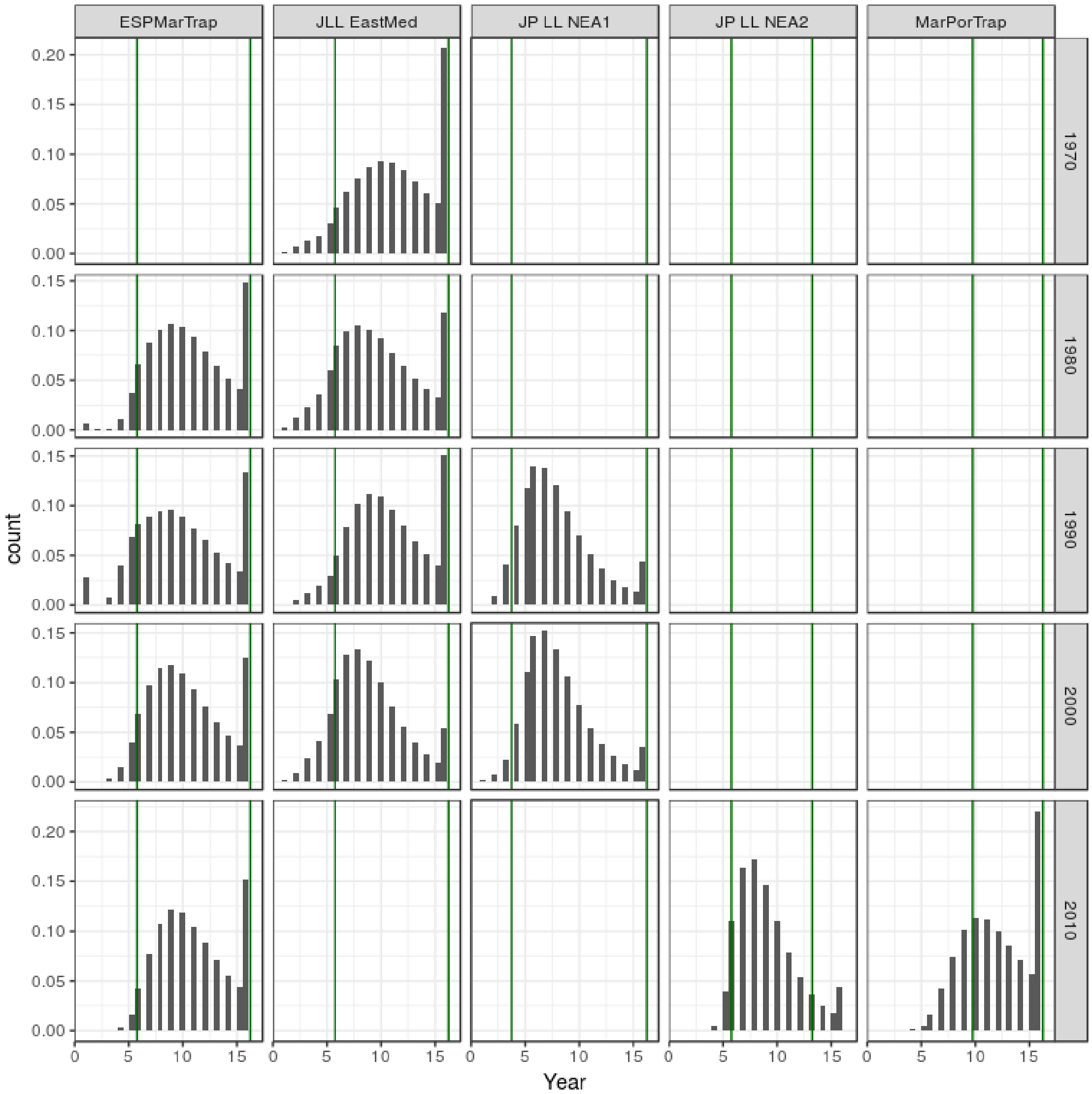

Figure 7. Catch composition, vertical lines show the ages used in calibration.

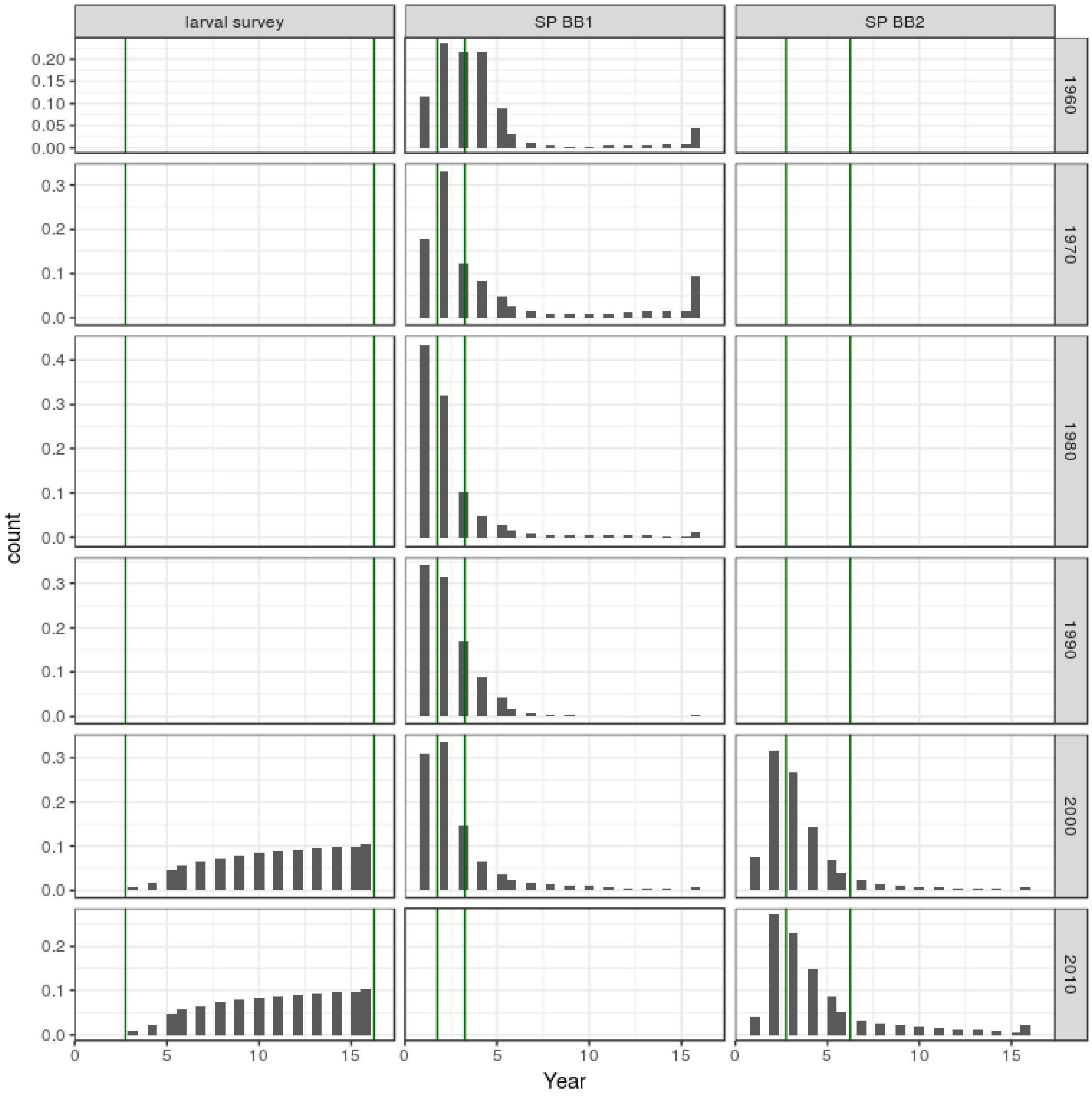

Figure 8. Catch biomass composition, vertical lines show the ages used in calibration.

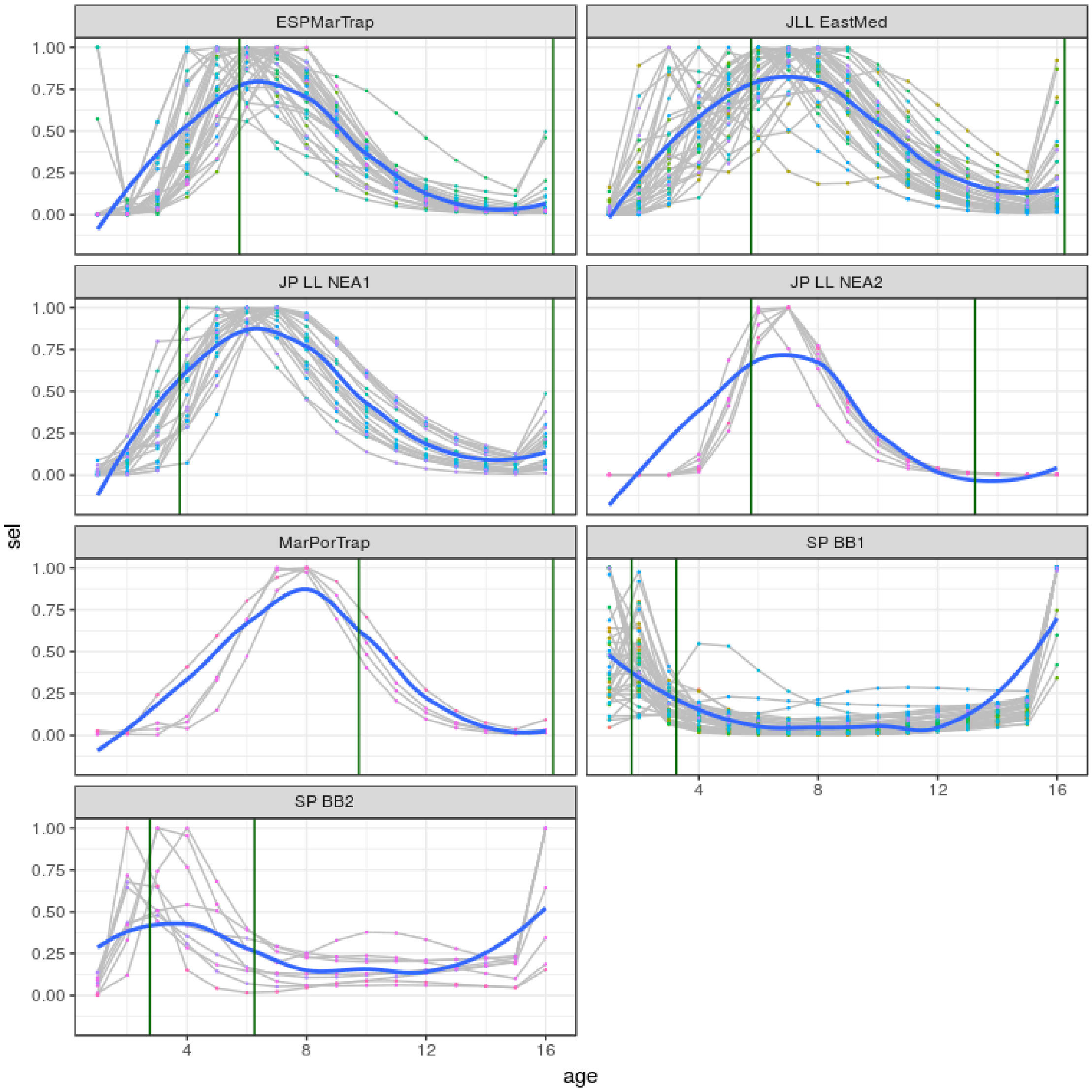

Figure 9. Catch curve analyses, vertical lines show the ages used in calibration.

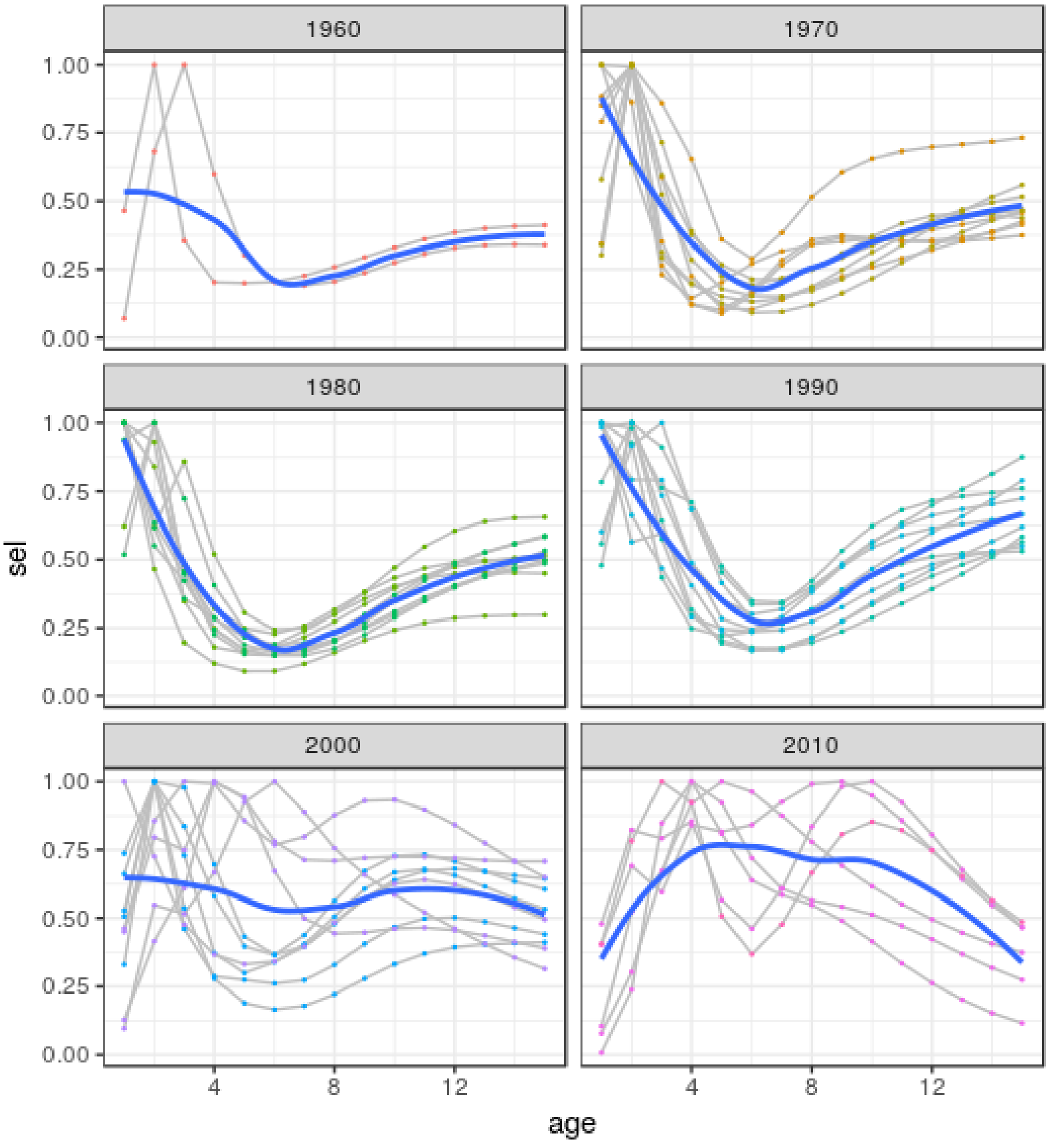

Figure 10. Selection pattern of the entire catch-at-age.

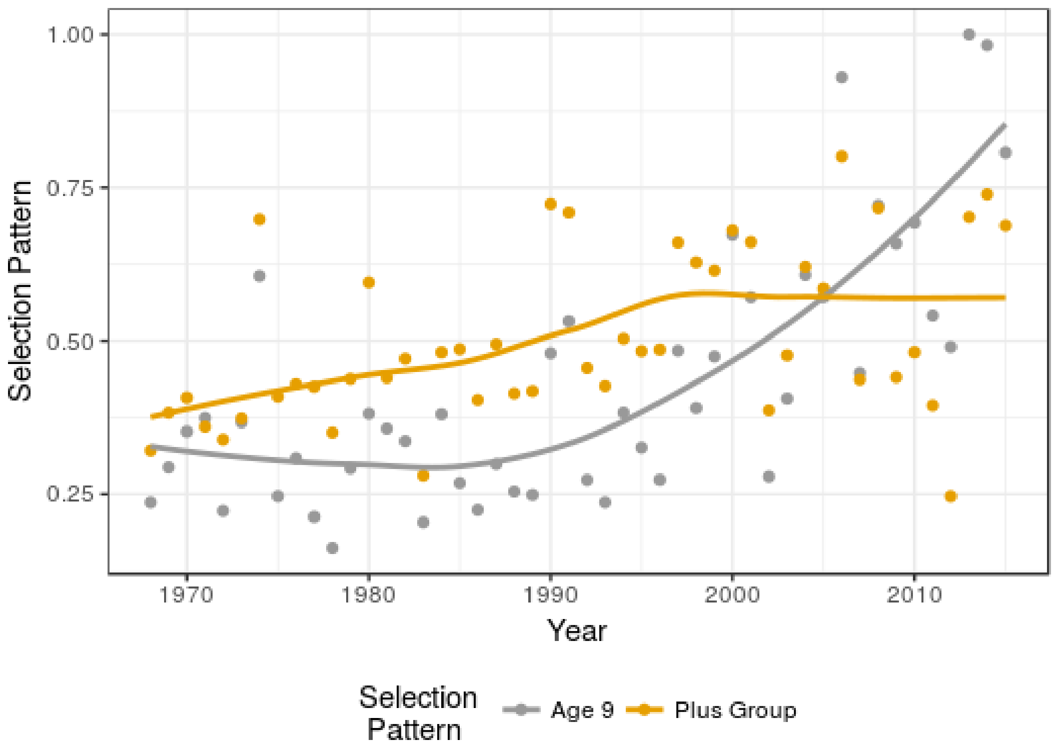

Figure 11. Plus group and age 9 selection patterns.

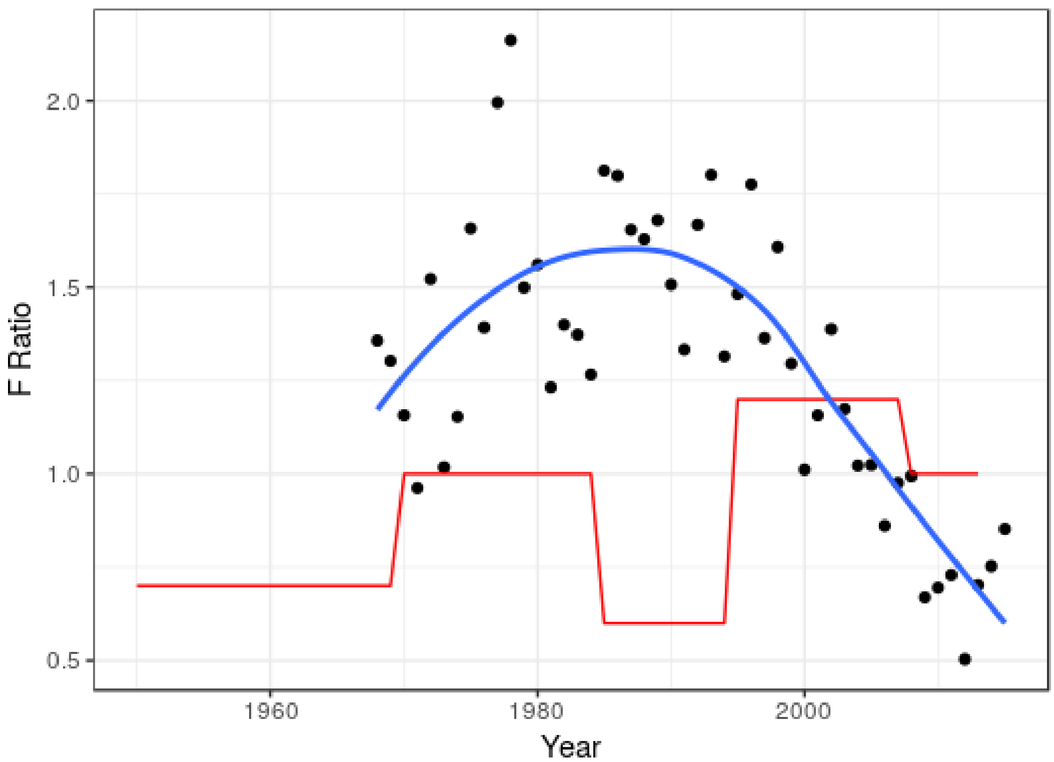

Figure 12. Plus group selection patterns.