Chilean Jack Mackerel
Workshop

Preliminary estimation of current state of Chilean Jack Mackerel (*Trachurus murphyi*) stock in the high seas in the South East Pacific

Dmitry Vasilyev[1], Alexander Glubokov[1], Doonam Kim[2]

1- Federal Research Institute of Fisheries&Oceanography, Moscow, Russia

2 – Distant Water Fisheries Resources Institute, Busan, Republic of Korea

Corresponding author: glubokov@vniro.ru

ABSTRACT

Clearly understanding that the amount of available information about the modern state of jack mackerel stocks in the high seas of the South East Pacific are close to lower limit for any stock assessment, we nevertheless made an attempt to make an assessment on the basis of very limited available data using a model from the TISVPA group, aiming at robustness of the analysis. According the results of the assessment (estimates of stock biomass (2+) in comparison to catches, recruitment at age 2, average fishing mortality (weighted by age groups abundance), and selection pattern by years) the stock biomass is relatively stable with the average about 7 million tones.

THE DATA AND THE MODEL

The state of the stock was estimated using the following data:

- catch-at-age (2003-2006), calculated from the total catch data, Vanuatu size structure of catches, the age-length key and average weight-at-age data from Russian surveys (2002-2003);
- Korean CPUE data (2003-2006);
- age structure of the stock for the beginning of 2003 from Russian surveys;
- the value of instantaneous natural mortality coefficient was assumed equal to 0.23 for all age groups.

The model used for analysis was a a separable cohort model of the TISVPA-group, implemented, among many other models, for stock assessment in frames of International Council for the Exploration of the Sea (ICES) and some other regions. The model is based on some principles of robust statistics what helps to extract weak signals from noisy data. The version of the model used for the assessment attributes residuals in cohort part of the model to errors in catch-at-age data, assuming that selection pattern is stable. This version is often more robust for noisy catch-at-age data. Additional robustness of cohort part of the model with respect to outliers in catch-at-age was attained 1) by minimization of the median of the distribution of squared residuals in logarithmic catch-at-age as a measure of closeness of the model fit to catch-at-age data, and 2) by condition of unbiased model description of logarithmic catch-at-age data. An additional signal about the size of the stock is taken from minimization of residuals between age proportions in the stock in 2003 taken from Russian surveys and from the cohort part of the model. Here the absolute median deviation (AMD) of logarithmic residuals in age proportions is used as a measure of closeness of fit. The AMD is the median of distribution of deviations of absolute logarithmic residual from their median value. The overall objective function of the model is composed as a weighted sum of catch-at-age-based and survey-based components (the weights reflect the relative efficient number of data points for each source of information).

As it was mentioned above, the main characteristic feature of the TISVPA model consists in intentional implementation of principles of robust statistics in procedures of estimation of the model parameters ((Vasilyev, 2005; 2006). The TISVPA includes several features which help it to operate with “real” (that is strongly noisy) data. Among them: robust loss functions, possibility to ensure unbiased solution, independence of estimated selection pattern upon user’s choice about its overall shape, implementation of different options concerning mutual validity of assumptions about quality of catch-at-age data and stability of selection pattern, possibility to exclude influence of year-to-year survey catchability variations caused by difference in survey conditions, etc

Since it is often believed that traditional separable models are too restrictive to describe some stocks, for example, with highly variable recruitment, the TISVPA model can represent fishing mortality coefficients (more precisely – exploitation rates) as a product of three parameters: f(year)*s(age)*g(cohort), that is it gives possibility to estimates within the model an additional set of generation-dependent parameters in separable representation of exploitation rates. This set of parameters is intended to adapt traditional separable representation of fishing mortality (as a product of age-dependent and year-dependent factors) to situations when several year classes may have peculiarities in their interaction with fishing fleets caused by different spatial distribution, higher attractiveness of more abundant schools to fishermen, or by some other reasons. Unfortunately, because of very limited information available, the traditional (not triple) separable representation was used for jack mackerel.

Brief description of the model is summarized in the table below.

Model	**TISVPA**
Version	**2006.1**
Model type	A separable model is applied to one or two periods, determined by the user. The separable model covers the whole assessment period. **It is possible to include the third, generation-dependent, factor into separable representation.**
Selection	The selection at oldest age is equal to that of previous age; selections as function of age s(a) are normalized by their sum to 1. For the plus group the same mortality as for the oldest true age. **If generation dependent factors are included**, then s(a,y)=s(a)g(cohort). s(a,y) can be normalized for each year by their sum to 1 – sub-model of "within-year effort redistribution by ages, or not – sub-model of "gain (loss) in selection". The matrix of g-factors is normalized to give global average = 1.
Estimated parameters	
Catchabilities	The catchabilities by ages and fleets can be estimated or assumed equal to 1. Catchabilities are derived analytically as exponents of the average logarithmic residuals between the catch-derived and the survey-derived estimates of abundance.
Plus group	The plus group is not modelled, but the abundance is derived from the catch assuming the same mortality as for the oldest true age.
SSB surveys	Considered as absolute or relative. If considered as relative, coefficient of proportionality is derived analytically as exponent of the average logarithmic residuals between the catch-derived and the survey estimates of SSB.
Surveys in year (terminal + 1)	Can be taken into account (in assumption that fishing pattern in the year (terminal+1) is equal to that of terminal year)
Objective function	The objective function is a weighted sum of terms (weights may be given by user). For the catch-at-age part of the model, the respective term is: • sum of squared residuals in logarithmic catches, or • median of distribution of squared residuals in logarithmic catches MDN(M, fn), or • absolute median deviation AMD(M, fn). For SSB surveys it is sum of squared residuals between logarithms of SSB from cohort part and from surveys. For age- structured surveys it is SS, or MDN, or AMD for logarithms of N(a,y) or for logarithms of proportions-at-age, or for logarithms of weighted (by abundance) proportions-at-age.
Variance estimates/ uncertainty	For estimation of uncertainty parametric conditional bootstrap with respect to catch-at-age, (assuming that errors in catch-at-age data are log-normally distributed, standard deviation is estimated in basic run), combined with adding noising to indexes (assuming that errors in indexes are log-normally distributed with specified values of standard deviation) is used.
Other issues	Three error models are available for the catch-at-age part of the model: • errors attributed to the catch-at-age data. This is a strictly separable model ("effort-controlled version") • errors attributed to the separable model of fishing mortality. This is effectively a VPA but uses the separable model to arrive at terminal fishing mortalities ("catch-controlled version") • errors attributed to both ("mixed version"). For each age and year, F is calculated from the separable model and from the VPA type approach (using Pope's approximation). The final estimate is an average between the two where the weighting is decided by the user or by the squared residual in that point. Four options are available for constraining the residuals on the catches: 1. Each row-sum and column-sum of the deviations between fishing mortalities derived from the separable model and derived from the VPA-type (effort controlled) model are forced to be zero.

	This is called "unbiased separabilization" 2. As option 1, but applied to logarithmic catch residuals. 3. As option 1, but the deviations are weighted by the selection-at-age. 4. No constraints on column-sums or row-sums of residuals. If "triple-separable" version is used, then option 2 also produces cohort-sum equal to zero. For options 1 and 2, as well as for option 3 if not the whole age range is chosen for application of g-factors, the listed above conditions are somewhat compromised, but they are still valid for generation-independent s(a).

Thus, according to the above mentioned, the TISVPA model is aimed at robustness, but what could be considered as a methodological difficulty, is the need to combine various loss functions, traditional or special robust ones. The formulations of data source – specific components of the overall model loss function could be quite different depending on properties of the data set under consideration. This causes some ideological problems of their mutual weighting while their mixing into overall loss function: it could be necessary to mix, for example, the sum of squared residuals for one data source with median of squared residuals or median absolute deviation for another one, etc., that is to combine values of different statistic nature.

To the contrary, implementation of likelihoods within the frames of Bayesian approach, as it is described, for example, in the paper concerning Chilean jack mackerel stock assessment (Canales & Serra, 2008), potentially allows to avoid this difficulty. Unfortunately, the classic likelihood functions are known to be extremely non-robust. For example, Y. Chen and D. Fournier (1999) reminded that "in formulating likelihood functions, data have been analyzed as if they are normally, identically, and independently distributed. It has come to be believed that the first two of the assumptions are frequently inappropriate in fisheries studies. In fact, data distributions are likely to be leptokurtic and (or) contaminated by occasional bad values giving rise to outliers in many fisheries studies." In their studies Y. Chen and D. Fournier showed that "the existence of outliers may greatly bias the derived posterior distributions. The likelihood of having outliers in fisheries studies implies that posterior distributions may be unreliable. This may lead to erroneous results on the dynamics of fish stocks and subsequently the adaptation of an inappropriate strategy in managing fisheries resources." Noel G. Cadigan and Ransom A. Myers (2001) examined two maximum likelihood estimators of the sequential population model (SPA) parameters. These estimators were based on assumption that the stock-size indices were described by the lognormal or gamma distributions. Using simulations, the authors found that both types of estimators could have significant biases; however, their results indicated that it could be preferable to use the gamma model, because it tended to have lower bias and variability, even when the true distribution of the stock-size indices was lognormal.

Generally speaking, the following approaches are commonly used in order to overcome non-robust properties of classic likelihoods:

- implementation of classic distributions with heavy tails (in order to better accommodate outliers);
- implementation of mixed, or so called mixture- distributions;
- implementation of exotic and extremely flexible distributions.

A lot of robust distributions are summarized, for example, by K.Passarin (2004), including: power-series distributions; extended power distribution; the Student-t distribution; elliptical distribution, which is considered to be a family of symmetric distributions including *inter alia* the normal and the Student−t distributions.

This group of approaches is aimed at a better accommodation of large errors in observation. But the question whether by doing this we can sufficiently reduce the influence of outliers on the solution still remains. F. Hampel (2002) underlines that …"The most common way out in practice still seems to be the replacement of the original parametric model, such as normality, by another, more complicated ad hoc model. *These models are, strictly speaking, as unrealistic as the original model*; if (as is frequently the case) they are chosen with good intuition, they do work for a full neighborhood of the original model, *but this can only be proven by robustness theory*."

Another group of approaches comprises construction of quasi-likelihoods based on reduced influence of "bad points" (M-estimates, etc.) (e.g. see Hampel et al., 1986). But when the model simultaneously includes a number of different quasi-likelihoods with artificially reduced influence of some points, the problem of mutual weighting of information from different sources rises again.

F. Hampel (2002) also pointed that "...some Bayesians may want to cling to their original model and to an unmodified likelihood function, yet be somewhat robust. For them I offer the following tentative suggestion. All they have to do is to replace the most extreme observations by pseudo-observations, which behave like data from the ideal model and do not contain dangerous outliers”. This seems to be a very useful generalized suggestion which in many cases could be easier to implement rather than to construct an exotic likelihood and then to prove that it really helps.

There are a lot of ways to detect outliers and then to do something with them. One of the most simple ones is α-winsorization (Huber, 1981) and procedures built on its basis. Unfortunately, the procedure of classic α-winsorization is itself not robust since it is based on such non-robust measure of scale, as standard deviation. However, it is not difficult to develop a more robust and efficient winsorization procedure (Vasilyev, 2004) on the basis, for example, of so called “X-84”-rule introduced by P.Huber (Hampel et al., 1986). For the simulated catch-at-age data containing 5% of outliers it was shown (Vasilyev, 2004) that “robustified” winsorization procedure gave almost twice better improvement of results of stock assessment compared to classic procedure.

It is also necessary to mention that dealing with procedures of iterative detection and improvement of outliers it is important to have a *sufficiently good initial estimate*, i.e. the one obtained using *all points*. Otherwise the procedure can run in the wrong direction. Therefore, the model itself should be already sufficiently robust, when it is applied to all data including not yet detected outliers. But, naturally, all of that - for future, when agreed international data for jack mackerel stock assessment in the high seas of the South East Pacific will be ready and used.

RESULTS OF THE ASSESSMENT

Profiles of components of the model objective function, as well as profile of the total objective function, are presented on Figure 1 as a function of the effort factor value in the terminal year (the effort factors are relative year-dependent factors in separable representation of fishing mortality coefficients). Despite of high level of noise in the data, the results of application of the model revealed rather coherent signals about the stock size from catch-at-age and surveys.

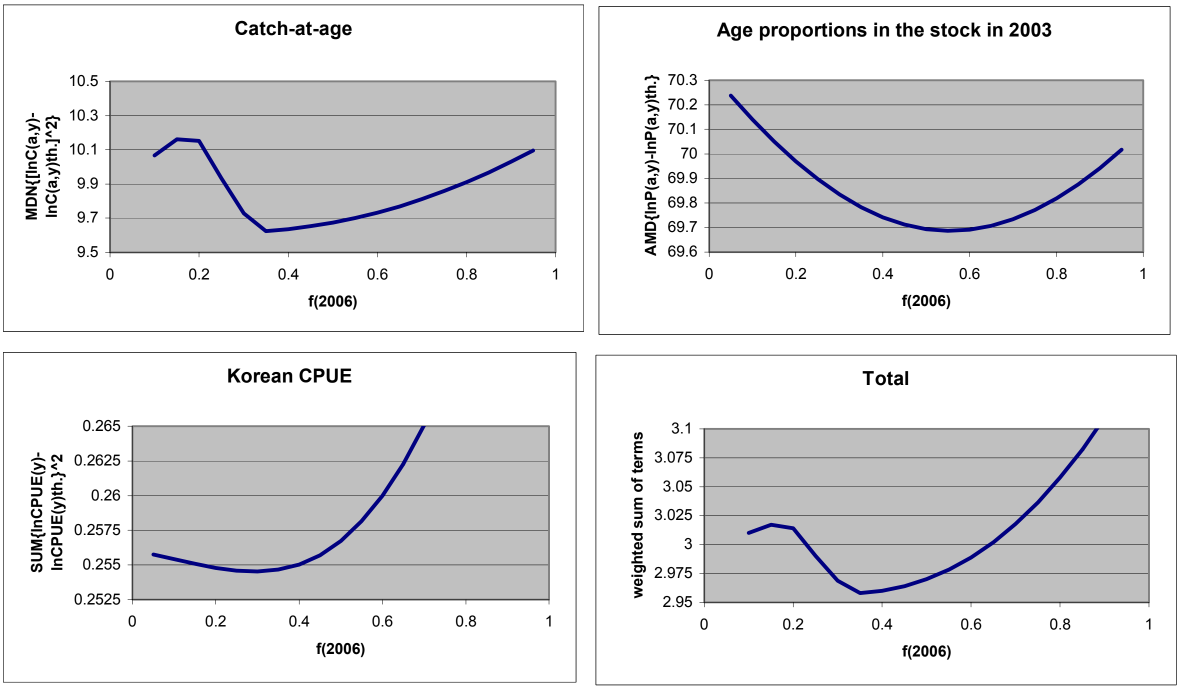

Figure 1. Profiles of components of the ISVPA objective function.

According the results of the assessment (estimates of stock biomass (2+) in comparison to catches, recruitment at age 2, average fishing mortality (weighted by age groups abundance), and selection pattern by years) the stock biomass is relatively stable with the average about 7 million tones (Figure 2).

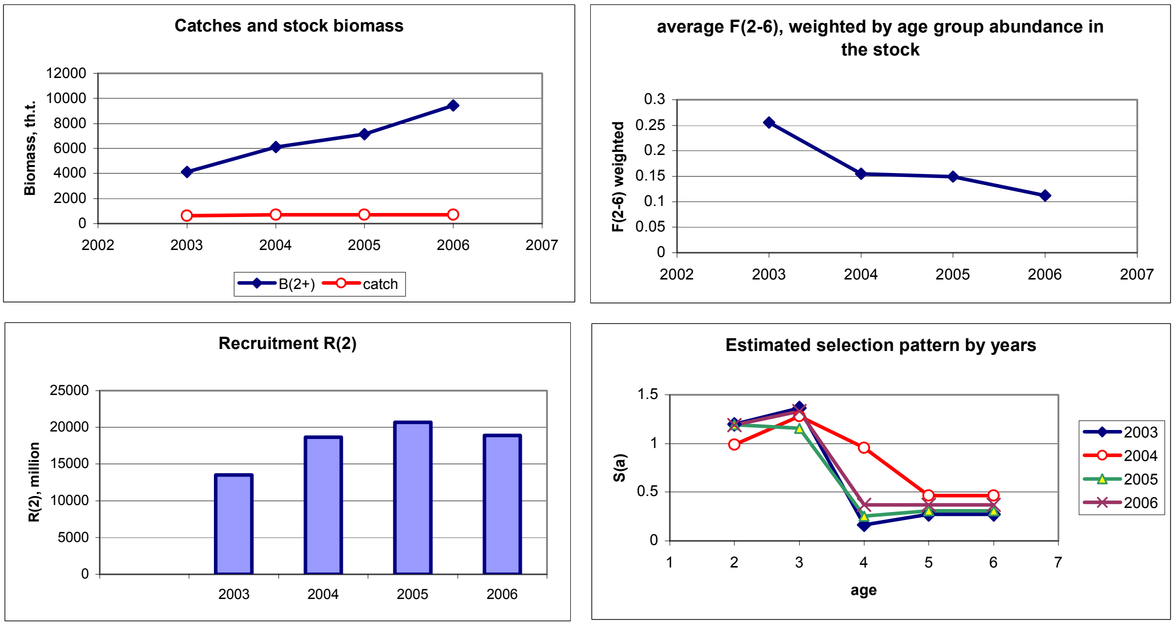

Figure 2. TISVPA-derived estimates of Chilean Jack Mackerel stock in the high seas of the South East Pacific.

Figure 3 represents the uncertainty in the stock biomass, selection-at-age and abundance in 2006 estimates obtained by means of parametric conditional bootstrap. Naturally, because of data-limited situation the uncertainty in the results is rather high.

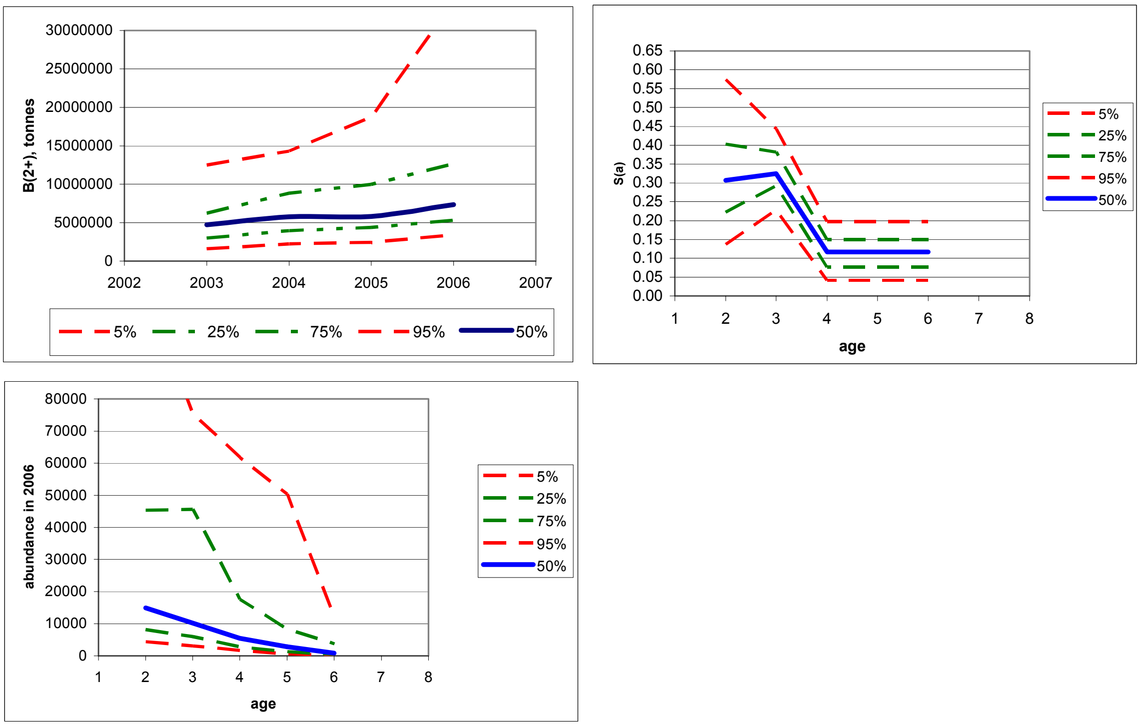

Figure 3.TISVPA: bootstrap estimates of uncertainty in the results.

CONCLUSION

The following conclusions can be drawn from our very preliminary assessment:

- the high seas stock biomass of the pacific jack mackerel in the South East Pacific in the area of available international catch data is relatively stable with the average about 7 million tones;
- uncertainty in the results is naturally high because of very restricted information available;
- it is strongly needed to compile agreed data for stock assessment on the basis of all available information from all countries about pacific jack mackerel stock and fishery in the high seas of the South East Pacific.

REFERENCES

Cadigan N.G., and R.A.Myers. 2001. A comparison of gamma and lognormal maximum likelihood estimators in a sequential population analysis. Can. J. Fish. Aquat. Sci. 58: p.560–567 (2001)

Canales, C. & R. Serra. 2008. Updated status of the Chilean jack mackerel stock. Chilean jack mackerel workshop. CHJMWS pap#25.

Chen, Y., and D. Fournier. 1999. Impacts of atypical data on Bayesian inference and robust Bayesian approach in fisheries. Can. J. Fish. Aquat. Sci. 56: p. 525–1533

Hampel, F. 2002. Some thoughts about classification. Research Report No. 102. January 2002. Seminar fur Statistik Eidgenossische Technische Hochschule (ETH) CH-8092 Zurich Switzerland. 19 p.

Hampel, F. R., Ronchetti, E.M., Rousseeuw, P. J., Stahel,W. A. 1986, Robust Statistics: The Approach Based on Influence Functions, Wiley, New York.

Passarin, K. 2004. Robust Bayesian estimation. UNIVERSITÀ DELL'INSUBRIA, FACOLTÀ DI ECONOMIA, 11/2004.

Vasilyev, D. 2004. Winsorization: does it help in cohort models? ICES CM2004/K:45.

Vasilyev D. 2005. Key aspects of robust fish stock assessment. M: VNIRO Publishing, 2005. 105 p.

Vasilyev D. 2006. Change in catchability caused by year class peculiarities: how stock assessment based on separable cohort models is able to take it into account? (Some illustrations for triple-separable case of the ISVPA model - TISVPA). ICES CM 2006/O:18. 35 p.