Chilean Jack Mackerel Workshop

pap#22

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

Clearly understanding that the amount of available information about the modern state of jack mackerel stock in the high seas is close to lower limit needed for stock assessment, we however consider it is necessary to begin the process.

Input data

- *catch-at-age (2003-2006), calculated from the total catch data, Vanuatu and EU size structure of catches (2003-2006), 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;*
- *M = 0.23 for all age groups.*

BX - 783 «Jan Maria» №1066.652-401A

«Atlantida»

The model: TISVPA

- separable (ordinary or "triple")
- based on principles of robust statistics what helps to extract weak signals from noisy data
- robust objective functions – instead of likelihoods
- possibility to ensure unbiased solution
- implemented for stock assessment in frames of the International Council for the Exploration of the Sea (ICES).

description of the model

TISVPA

W(a,y); mat(a,y)

C(a,y)

M(a)

C(a,y) and surveys filtration (Kriging, robust winsorization, etc.)

Choice of the option for M

1. M= const - ? 2. M(a) - ? 3. M(a) is known

Choice of properties of the solution

1. unbiased separable representation of F(a,y)
2. unbiased weighted separable representation of F(a,y)
3. unbiased model description of logarithmic C(a,y)

Choice of the separable model (double or triple) and its age range

Choice of error model

1. Errors – in catch-at-age
2. Errors – separable representation of F(a,y)
3. Errors – in both

What to minimize for log. C(a,y)? (SSE, MDN(SE) or AMD(E))

To scan or to look for precise solution ?

{lnC-lnC*}

bootstrap

Auxiliary information

1. Integrated SSB (or FSB) indexes
2. Age-structured abundance indices (or CPUE) for mature, immature, or total stock (SSE, MDN(SE) or AMD(E)->min. for log. N(a,y)) or log. P(a,y)

Results:{N(a,y)}; {B(a,y)};{SSB(a,y)}{s(a)};f(y);{F(a,y)}; M; q(a)

The TISVPA idea: F(a,y)=f(y)s(a)G(cohort)

-age-range of estimation and application of G-factors can be optimized (to make it "physically" relevant and from point of view of minimization)

- two sub-versions with respect to G-factors:
  - model of "within-year effort redistribution by ages" (s(a,y)=s(a)G - normalization is hold for each year)
  - model of "gain (loss) in selection" (only s(a) is normalized, but not s(a,y))

Robustness of likelihood functions ?
(some experience)

Y. *Chen and D. Fournier. Impacts of atypical data on Bayesian inference and robust Bayesian approach in fisheries. Can. J. Fish. Aquat. Sci. 56: 1525–1533 (1999):*

"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"….
"*This study shows 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*."

Robustness of likelihood functions ?

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

“We examine two maximum likelihood estimators of SPA parameters. These estimators are based on assuming that the stock-size indices are from lognormal or gamma distributions. Using simulations, we find that both types of estimators can have significant biases; however, our results indicate that it is preferable to use the gamma model, because it tends to have lower bias and variability, even when the true distribution of the stock-size indices is lognormal.”

Robustness of likelihood functions ?

(classic likelihoods are known to be extremely non-robust)

common ways:

- classic distributions with heavy tails (to accommodate outliers)
- mixed ("mixture") distributions
-exotic (and extremely flexible) distributions

(what we are really doing by this?)

-*quasi-likelihoods* based on reduced influence of "bad points" (M-estimates) (Huber, Hampel, etc)

(*but here the question of weighting of inputs from different sources of information rising again*)

A lot of robust distributions are summarized, for example, in:

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

Robustness of likelihood functions ?

1

"Bayesians seem to have problems with robustness, especially with robustness against deviations from the parametric model and against changes of the prior distribution. 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."

Frank Hampel. Some thoughts about classification. Research Report No. 102. January 2002. Seminar f¨ur Statistik Eidgen¨ossische Technische Hochschule (ETH) CH-8092 Z¨urich Switzerland

Robustness of likelihood functions ?

2 *(about exotic distributions)*

"...Such models cannot claim either to be the exact "true model"; they are more complicated, mathematically less nice and harder to interpret; they either lose efficiency by switching between simple models, or they try to estimate ill-determined parameters and thus are in danger of doing overfitting (which may be a partial explanation for their surprisingly mediocre performance); and they contradict one of the deepest principles of experienced data analysis: use (and first search for) the simplest model reasonably possible, even if it is "significantly wrong"(!), because it is more useful, more reliable and better generalizable than a more complicated one.."

Frank Hampel. Some thoughts about classification. Research Report No. 102. January 2002. Seminar f¨ur Statistik Eidgen¨ossische Technische Hochschule (ETH) CH-8092 Z¨urich Switzerland

Robustness of likelihood functions ?

3

"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".

Frank Hampel. Some thoughts about classification. Research Report No. 102. January 2002. Seminar f¨ur Statistik Eidgen¨ossische Technische Hochschule (ETH) CH-8092 Z¨urich Switzerland

Summary:

Model parameters can be estimated via :

Special objective functions | **Likelihoods**

+ Easy to make it robust and distribution – free

+ *apparent* easiness of combining signals from different data sources
- But extremely low robustness

Rational approach: data censoring based on robust winsorization (detection and correction of "bad points"), e.g. "X-84 rule" by Huber

Frequent (but often deadlock) approach:
-Over-flexible and exotic distributions
-Mixed (mixture) distributions

Robust initial estimates of the parameters

The TISVPA setting used in the assessment:

• separable model – “ordinary” (not triple), applied to all years and ages

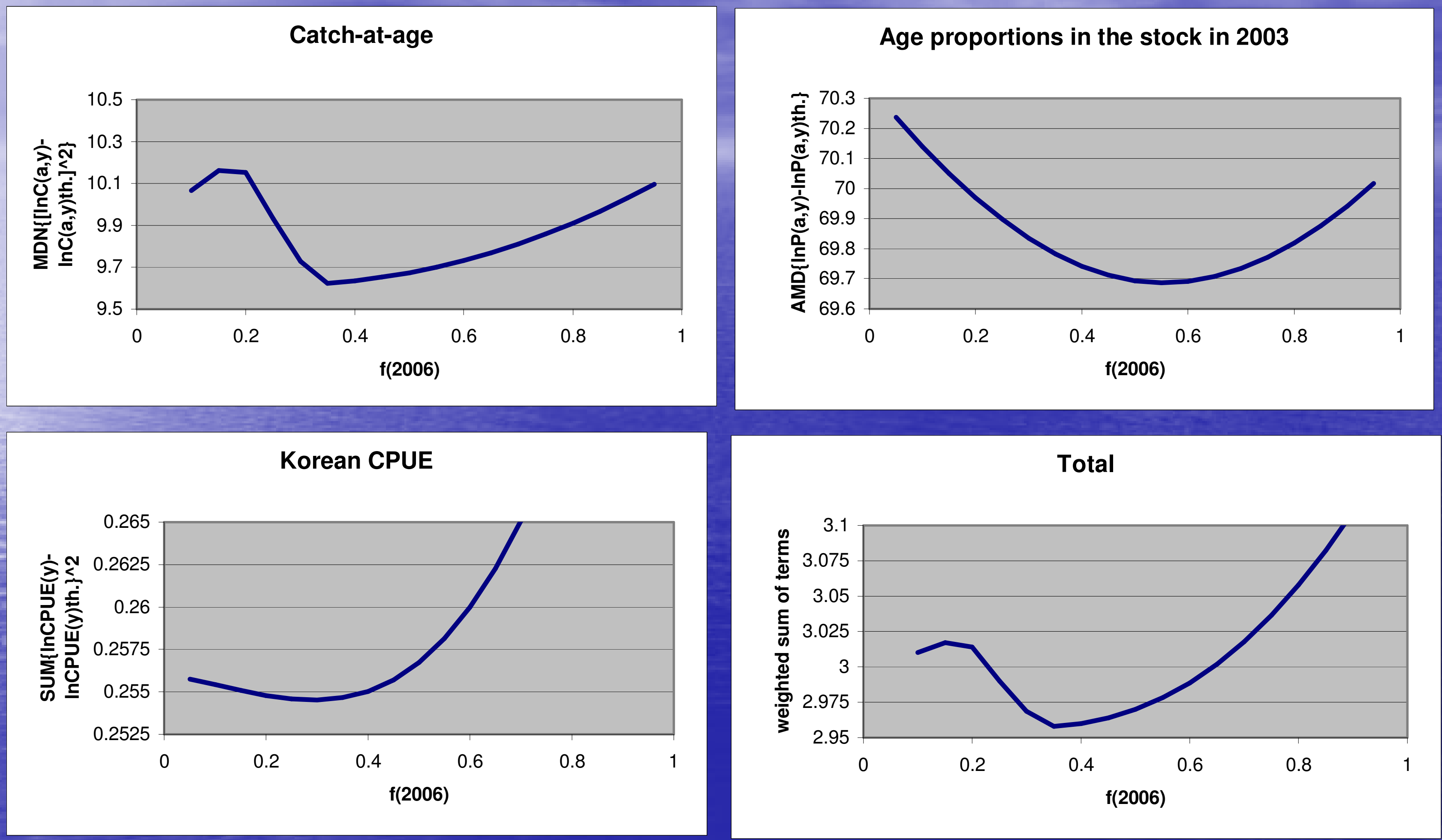
• residuals in cohort part of the model are attributed to errors in catch-at-age data. This version is often more robust for noisy catch-at-age data

• 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

• the condition of *unbiased* model description of logarithmic catch-at-age data

• the absolute median deviation (AMD) of logarithmic residuals in age proportions is used as a measure of closeness of fit to data on stock age structure from Russian surveys

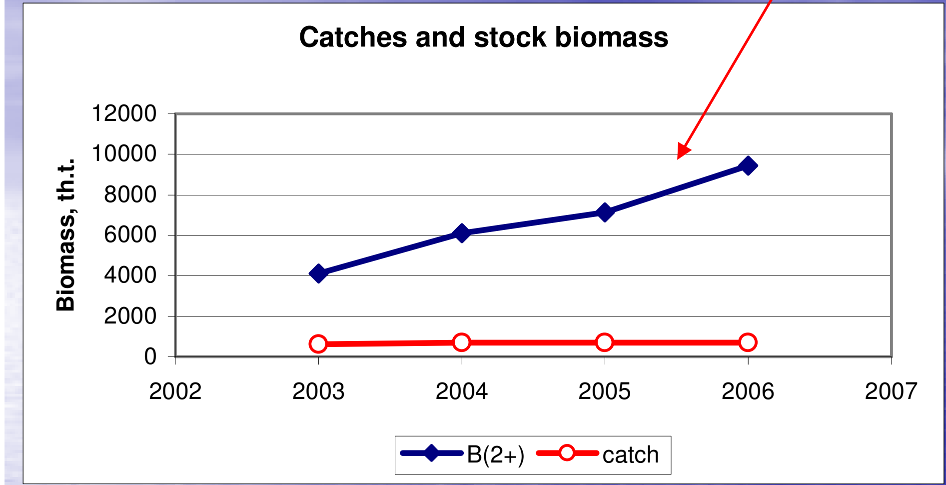
Results

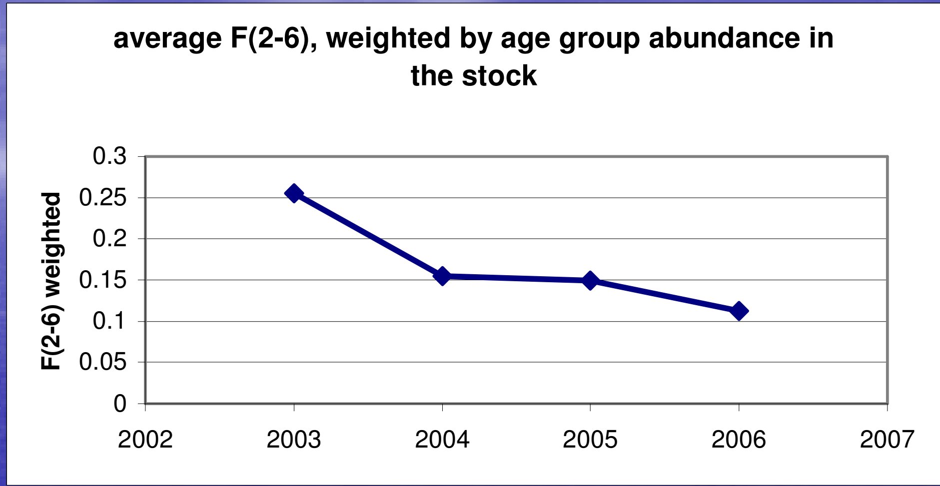


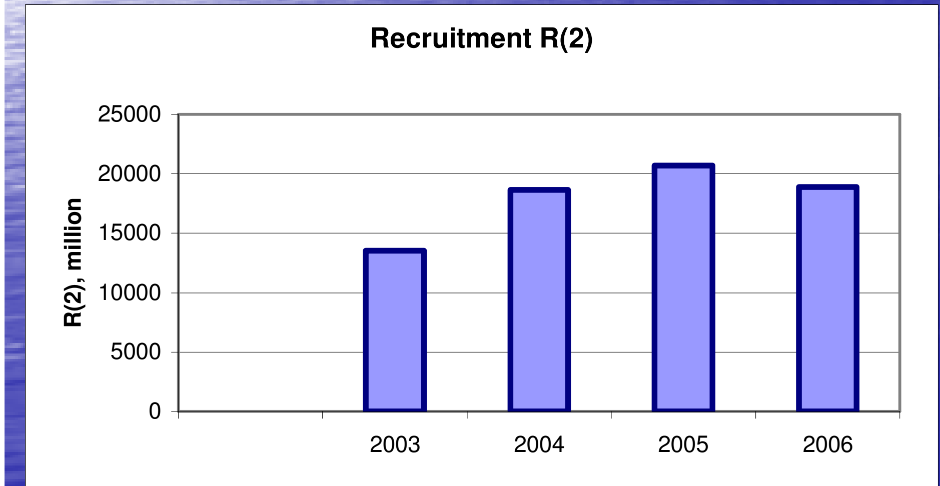
Profiles of the TISVPA objective functions

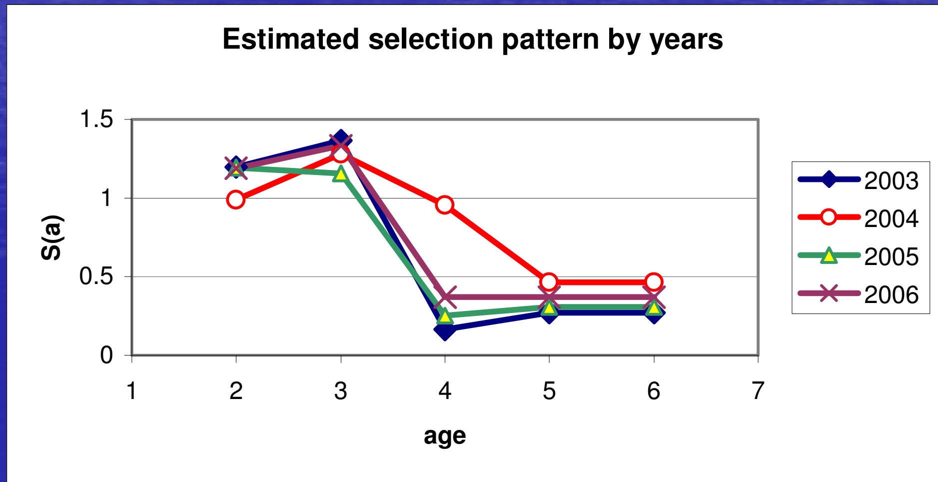
Results

compare to rising trends in biomass and cpue from #11,#10



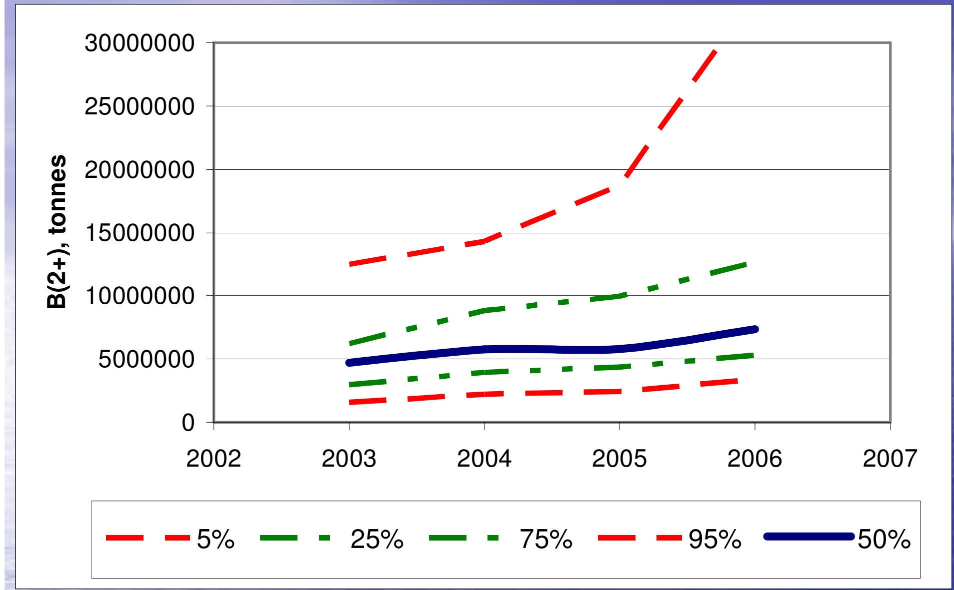


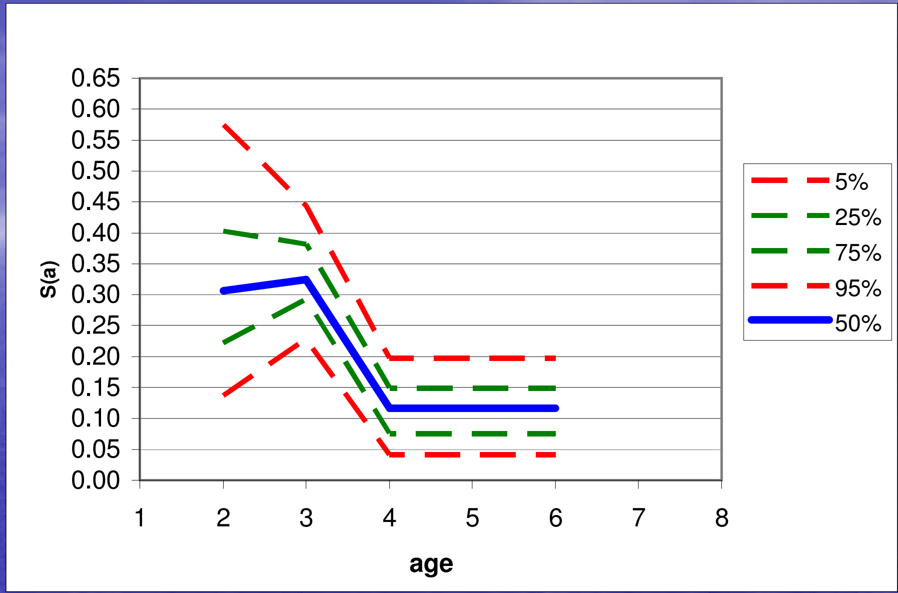


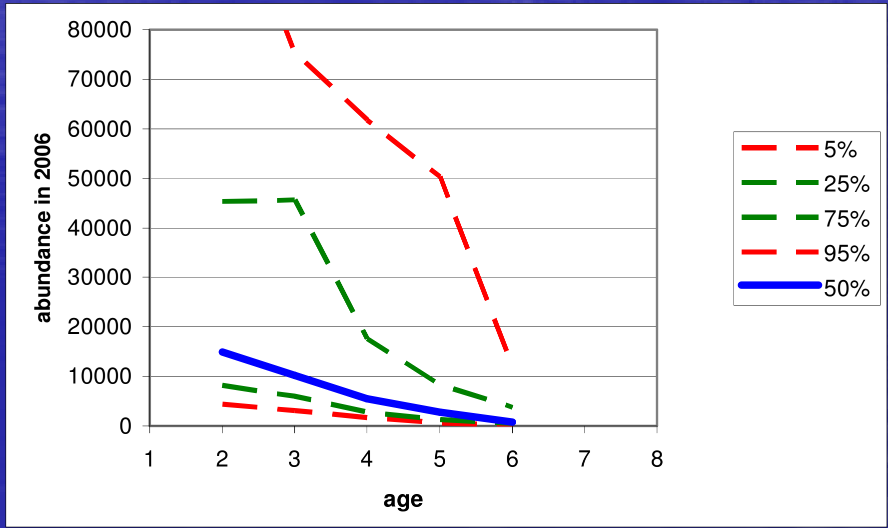


TISVPA – derived estimates of the stock biomass, F, R(2) and selection pattern

Results







TISVPA – estimates of uncertainty (conditional parametric bootstrap)

Conclusions

- *the stock biomass is relatively stable with the average about 7 million tones*
- *uncertainty in the results is naturally high because of very restricted information available*
- *the need in agreed data for stock assessment on the basis of all available information from all countries about Chilean Jack Mackerel stock and fishery in the high seas of the South East Pacific*

The authors would like to extend their gratitude to the Pelagic Freezer Trawler Association which is funding the Dutch research in the Pacific, and the company Parlevliet and Van der Plas, the owner of the vessel BX783 Jan Maria, as well as Fishing company Unimed Glory SA.