Approach to the implementation of hypothesis. A proposal for the Chilean jack mackerel integrated stock assessment

Cristian Canales (ccanales@ifop.cl) and Rodolfo Serra

Instituto de Fomento Pesquero (IFOP) - Chile

Abstract

Herein, we present a modeling approach in order to incorporate the different hypotheses to be evaluated. The largest model supposes the jack mackerel in the South Pacific Ocean is a single unit stock, and the nested models are the respective hypotheses proposed in the FAO Santiago Workshop. The model considers different fleets and different selectivity patterns that are operating in all distribution area of this species.

1. Introduction

In the Jack mackerel Workshop, held in the FAO regional office, Santiago 2008, four hypotheses were proposed regarding the jack mackerel stock structure, which are:

- H1: Jack mackerel caught off the coasts of Peru and Chile each constitute separate stocks which straddle the high seas.
- H2: Jack mackerel caught off the coasts of Peru and Chile constitute a single shared stock which straddles the high seas.
- H3: Jack mackerel caught off the Chilean area constitute a single straddling stock extending from the coast out to about 120°W.
- H4: Jack mackerel caught off the Chilean area constitute separate straddling and high seas stocks.

From a modeling perspective, hypotheses 1 and 3 could be considered as the same, ever since they refer to the main hypothesis of a single stock off Chile, beyond its EEZ. Hypothesis 2 could be considered an extension of the area where the main stock assessment is currently carried out (H1-H3). In the same way, hypothesis 4 could be considered (or not) another extension to the mentioned area, where an unknown flowing process has occurred, permitting the establishment of a probably isolated high-seas population. On the other hand, we could suppose that there is a larger stock unit in the South Pacific and the four hypotheses are nested models of the first one.

Population modeling requires making the main population processes related to the different proposed hypotheses explicit, in mathematical terms. In this paper, we show the characteristics of a modeling scheme that could be considered in an integrated jack mackerel stock assessment and the data requirements for such purposes.

2. Geographical representation

We recognize three units of the population, whose interconnection depend on the hypothesis: Peru, Chile and high-seas. In each of these units different fleets operate which have different selectivity or exploitation patterns. This means that the exploitation of the population has different impacts, depending on the fleets that are fishing. There are two acknowledged ways to recognize these exploitation patterns: a) the effects depend on each fleets (multi-fleet based model), or b) the effects depends on the migration process (spatial explicit model). These two alternatives require different data to test them.

The first alternative implies having information of the fleet only, while the second alternative implies having information on the migration process. This last requirement is very difficult to obtain, and needs a particular research program for collecting the information (marking). When that is the case, the assessments are based mainly on assumptions. Another option is the fleet-based assessment, where the most important consideration is that the population is exploited by different fleet-selectivity. So, each fleet has a specific impact on the population, particularly in the age-differentiated exploitation.

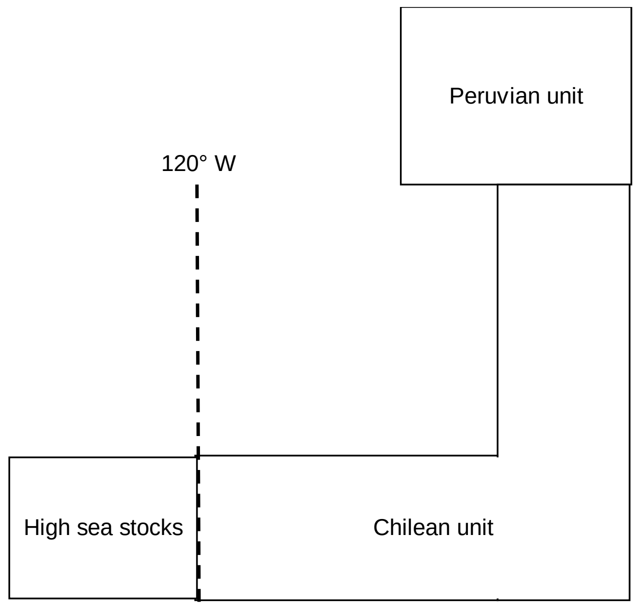

Figure 1. Representation of Chilean jack mackerel fishery units

3. Mathematical and statistical representation

If we consider a multi-fleet approach, a model could be represented by the next formulation:

3.1. Process model

a. Population dynamic

In order to avoid the over-parameterization of the model, we could consider Pope's approach, where the catch occurs instantaneously in the middle of each year, and births occur at the beginning of each year (i.e. January). So, the dynamics of the abundance-at-age (*a*) and per year (*y*) is represented by

$$N_{a,y} = \begin{cases} N_{a,y} & a = 2; y = 1 \\ N_{a-1,y-1}e^{-M} - \hat{C}^{tot}_{a-1,y-1}e^{-0,5M} & a = 2-11; y \geq 2 \\ (N_{a-1,y-1}e^{-M} - \hat{C}^{tot}_{a-1,y-1}e^{-0,5M}) + N_{a,y-1}e^{-M} - \hat{C}^{tot}_{a,y-1}e^{-0,5M} & a = 12; y \geq 2 \end{cases} \quad (1)$$

where C^{tot} is the total catch-at-age per year, and M is the natural mortality coefficient. The initial abundance-at-age could be considered around an equilibrium state, subject to stochastic noise:

$$N_{a,1} = R_o e^{-M(a-2)} e^{-\tau_a}$$

being R_o some equilibrium recruitment level and τ a stochastic variable defined as:

$$\tau \sim N(0, \sigma_N^2)$$

b. Recruitment

Recruitment come from some stock-recruitment relationship g (.) and is subject to process error:

$$R_1 = g(SB_{y-2})e^{-\varepsilon_y} \quad (2)$$

where $\varepsilon_y \sim N(0, \sigma_r{}^2)$.

c. Spawning biomass

The spawning biomass is estimated in mid-October as

$$SSB_y = \sum_a (N_{a,y}e^{-0.5M} - \hat{C}^{tot}{}_{a,t})e^{-3.5/12M} O_a w_{a,y} \quad (3)$$

where O is the maturity ogive at age (*a*) and w is the average weight-at-age and per year.

3.2. Observation models

a. Catches

The total catch-at-age by year corresponds to the sum of the catches by fleets:

$$\hat{C}_{a,y}^{tot} = \sum_f \lambda_f \hat{C}_{a,y}^f \tag{4}$$

Here, λ is a binary variable (0 or 1) and represents the catches of the *f*-fleet being considered (1) or not (0). In our case, we could consider at least 3 fleets: the Peruvian, the Chilean (northern and southern zones) and the international fleet operating in high sea (beyond 120°W). The catches are modeled in proportion to the stock at the middle of the year such that:

$$\hat{C}_{a,y}^f = \mu_{a,y}^f N_{a,y} e^{-0,5M} \tag{5}$$

$$\mu_{a,y}^f = \mu_y^f S^f{}_a \tag{6}$$

$$\mu_y^f = \frac{\lambda_f Y_y^f}{BV_y^f} \tag{7}$$

Where *Y* is the observed landing and the exploitable biomass by fleet corresponds to:

$$BV_y^f = \sum_a N_{a,y} S_a^f w_{a,y} e^{-0.5M} \tag{8}$$

Here, μ is the exploitation rate, Y is the landing, and S is age-specific selectivity modeled according to the fleet zone. Here we could propose a general selectivity model:

$$S_a^f = \begin{cases} e^{-\frac{1}{2\delta_1^2}(a-\beta)^2} & si \quad a < \beta \\ 1,0 & si \quad \beta < a < \gamma \\ e^{-\frac{1}{2\delta_2^2}(a-\gamma)^2} & si \quad a > \gamma \end{cases} \tag{9}$$

This model is composed by two normal distributions, and covers a wide range of selectivity forms, from trawl-type to gillnet-type. The model has four parameters: β and γ represent the reference ages where inflection points occur, and δ_1, δ_2 represent the two curve slopes (Canales et al., 2009).

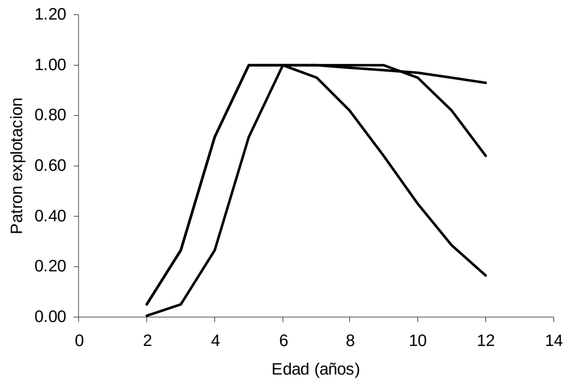

Figure 2. Different forms of jack mackerel selectivity model

For some fleets, we do not have available catch-at-age information; in this case we could use catch-at-length information considering some options:

- Using length-at-age keys from adjacent zones
- Using stochastic transformations in order to convert age to length

In the first option, the possibility of similar length compositions between zones would be a useful consideration, as it is the case of the Chilean fleet operating in the central-southern area, and the fleet operating offshore (off EEZ). In other cases, the length compositions are less similar as it is the case of the Peruvian fleet and the Chilean northern area; if that is the case, we must do some assumptions on the growth model in order to use it for this purpose.

In this sense, the predicted catch-at-length could be represented by;

$$\hat{C}_{l,y} = T\ \hat{C}_{a,y}$$

where T represents the probability of observing a fish of age "a" in the "l" length interval. This transformation matrix (in some cases called "transition" matrix) corresponds to:

$$T = \int_{l_1}^{l_2} g(l|a)dl$$

where the conditional probability function could be defined by:

$$g(l|a) = \frac{1}{\sqrt{2\pi}\sigma_a} e^{-\frac{(l-\bar{l}_a)}{2\sigma_a^2}}$$

Here, σ_a is the standard deviation in length for fish of "a" age class, and could be assumed to be proportional whit mean length in the form:

$$\sigma_a = \phi \bar{l}_a$$

being ϕ a fixed coefficient of variation to be resolved.

The mean length-at-age considers the VB growth parameter in the form:

$$l_a = l_{00}(1 - e^{-k(a-t0)})$$

Being l_{00}, k and t_0 the models parameters. These could be supposedly known or estimated on the model. For this last point, we need to establish some "prior" distribution, using, for example, the parameters available in literature.

b. Acoustic biomass

Several acoustic surveys have been carried out in Chile and Peru. If we suppose that they are carried out at middle of the year, the observation model could be represented by:

$$\hat{B}_y^f = q_y^f \sum_a (N_{a,y} e^{-0,5M} - \hat{C}_{a,y}) S_a^c w_{a,y} \qquad (10)$$

where q_y^f is the catchability coefficient for acoustic cruises and S_a^c is the age-specific availability that could be modeled considering (ec. 9). In Chile, the acoustic series is analyzed considering two hypotheses:

<u>Change in distribution</u>: this supposes that, as of 2002, the resource underwent a change in its distributional core beyond 200 nm. Thus, the estimated biomass in the acoustic surveys from 5 to 400 nm is comparable with the biomass estimated prior to 2002 between 5 and 200 nm.

$$q_y^c = \exp\left[\frac{1}{n}\sum_y \log\left(\frac{B_y}{\hat{B}_y}\right)\right] \tag{11}$$

<u>Contraction of biomass</u>: a population reduction starting in 2002 has been much faster (effect of hyper-reduction) along the distribution borders (first 200 nm from the coast) than at the nucleus of the population.

$$q_y^c = \begin{cases} \exp\left[\frac{1}{n}\sum_y \log\left(\frac{B_y}{\hat{B}_y}\right)\right] & \begin{array}{l} y < 2002 \\ B_y \in [5-400]mn \end{array} \\ \\ \eta \hat{B}_y^{\lambda} & \begin{array}{l} y \geq 2002 \\ B_y \in [5-200]mn \end{array} \end{cases} \tag{12}$$

c. Spawning biomass index

Chile has surveyed a couple of eggs in order to estimate the spawning biomass. The model for this observation is:

$$\hat{B}_y^{mph} = q^{mph} SSB_y \tag{13}$$

$$q^{mph} = \exp\left[\frac{1}{n_2}\sum_y \ln\left(\frac{B^{mph}{}_y}{\hat{B}^{mph}{}_y}\right)\right] \tag{14}$$

Here, B^{mph} is survey spawning biomass and n_2 the number of years with information from eggs/larvae surveys

d. Catch per Unit Effort (CPUE)

The CPUE from the central-southern zone of Chile is considered to be a good index of relative biomass between 1996 and 2003. Further changes in the fishery have affected this signal.

$$CP\hat{U}E_y = q \sum_a N_{a,y} e^{-0,5M} S^f_{a,y} \overline{w}_{a,y} \qquad (15)$$

$$q = \exp\left[\frac{1}{n}\sum_y \ln\left(\frac{CPUE_y}{\sum_a N_{a,y} e^{-0,5M} S^f_{a,y} w_{a,y}}\right)\right] \qquad (16)$$

3.3. Observation model errors (likelihood estimators)

There are several alternatives for modeling observation errors. Some of these are:

a. Age-at-catch compositions: a multinomial error distribution is assumed.

$$-\ln L_{p^f} = n \; p^f_{a,y} \ln(\hat{p}^f_{a,y}) \qquad (17)$$

where $p^f_{a,y} = \frac{C^f_{a,y}}{\sum_a C^f_{a,y}}$ and n is the size of the effective sample. $\hat{p}^f_{a,y} = \frac{\hat{C}^f_{a,y}}{\sum_a \hat{C}^f_{a,y}}$

b. **Acoustic-survey age compositions:** a multinomial error distribution is assumed.

$$-\ln L_{p^c} = n \; p^c_{a,y} \ln(\hat{p}^c_{a,y}) \tag{18}$$

where $p^c_{a,y} = \frac{N^c_{a,y}}{\sum_a N^c_{a,y}}$ and $\hat{p}^c_{a,y} = \frac{(N_{a,y}e^{-0,5M} - \hat{C}_{a,y})S^c_a}{\sum_a (N_{a,y}e^{-0,5M} - \hat{C}_{a,y})S^c_a}$. N^c is the abundance of the age observed.

The same form could be used if we are using catch-at-length information.

c. **Acoustic biomass:** a log-normal type of error distribution is assumed.

$$-\ln L_{B^c} = \frac{1}{2\sigma^2} \ln\left(\frac{B^c_y}{q^c \sum_a (N_{a,y}e^{-0,5M} - \hat{C}_{a,y})S^c_a \overline{w}_{a,y}} \right)^2 + cte_1 \tag{19}$$

d. **Spawning biomass index:** a log-normal type of error distribution is assumed.

$$-\ln L_{B^{mph}} = \frac{1}{2\sigma^2} \ln\left(\frac{B^{mph}_y}{q^{mph} \sum_a (N_{a,y}e^{-6/12M} - C_{a,t})e^{-3.5/12M} O_a w_{a,y}} \right)^2 + cte_2 \tag{20}$$

e. **CPUE:** a log-normal type of error distribution is assumed.

$$-\ln L_{CPUE} = \frac{1}{2\sigma^2} \ln\left(\frac{B^c_y}{q \sum_a N_{a,y}e^{-0,5M} S^f_{a,y} \overline{w}_{a,y}} \right)^2 + cte_3 \tag{21}$$

f. **Objective function:** corresponds to a Bayesian approach and considers the sum of the negative marginal contributions of the log-likelihood of the data and a *prior* (*p*) of the process error associated with the recruitment and the initial condition. The priors of parameters could be assumed as non-informative.

$$-\ln L_{tot} = (\ln L_{p^f} + \ln L_{p^c} + \ln L_{B^c} + \ln L_{B^{mph}} + \ln L_{CPUE}) + (\ln p_\tau + \ln p_\varepsilon) \qquad (22)$$

4. Some considerations

The general idea presented, was to propose a flexible modeling scheme, where the maximum information pieces could be considered under the respective hypothesis. This scheme should permit to incorporate the respective assumptions when there is no availability of knowledge or information. The state of art in stock assessment approach is based on statistical catch-at-age models when the information volume is important, as it is the case of jack mackerel in the South Pacific.

We showed a general model that could be improved with more information mainly related to biological process. This approach is based on a multi-fleet model, which is known to not having significant differences respect to a spatially-explicit model (Maunder et al, 2005)

In statistical literature, the "nested-models" term is used to refer to models from the same family applied to some data set. Here, the number of parameters could change. In this case we have the same model family, but different data sources. Thus, our idea is to use some statistical criteria to compare the behavior of different scale models. In this case, we could consider some statistical criteria like the Bayesian Information Criterion (BIC) (Swartz, 1978).

When estimating model parameters using maximum-likelihood estimation, it is possible to increase the likelihood by adding further parameters, which may result in overfitting. The BIC solves this problem by introducing a penalty term for the number of parameters in the model. This penalty for additional parameters is stronger than that of the AIC (Akaike Information Criterion).

Bibliography

Canales C R. Serra, L. Cubillos, L. Caballero, J. Cordoba, V. Bocic, A. Sepulveda, C. Gaticas, S. Nuñez, M. Arteaga y C. Gonzalez, 2009. Revision de metodologias de evaluacion del jurel a nivel subregional. Pre-informe final. Proyecto FIP 2007-23: 160 pp + anexos.

Canales C and R. Serra, 2008. Updated Status of the Chilean Jack Mackerel Stock. Technical Summary. Document SPRFMO-V-SWG. Fifth International Consultations on the Establishment of the South Pacific Regional Fisheries Management Organisation (SPRFMO). Guayaquil, March 2008.

Deriso, R.B., Quinn II, T.J., Neal, P.R., 1985. Catch-age analysis with auxiliary information. Can. J. Fish. Aquat. Sci. 42:815-824.

Elizarov, A.A., A.S. Grechina, B.N. Kotenev & A.N. Kuzetov. 1993. Peruvian jack mackerel, Trachurus symmetricus murphyi, in the open waters of the South Pacific. Journal of Ichthyology, 33(3): 86-104.

Fournier, D. & C.P. Archibald. 1982. A general theory for analyzing catch at age data. Can. J. Fish. Aquat. Sci., 39: 1195-1207.;

Hilborn R, Maunder M, Parma A, Ernst B, Payne J, Starr P., 2003. Coleraine: A Generalized Age-Structured Stock Assessment Model: User's Manual Version 2.0. Technical Report. School of Aquatic and Fishery Science, Fisheries Research Institute, Washington University [Rep. Fish. Res. Inst. Wash. Univ.], [np].;

Hilborn, R. 1990. Determination of fish movement patterns from tag recoveries using maximum likelihood estimators. Canadian Journal of Fisheries and Aquatic Sciences 47:635-643.

Hoyle S.D and M.Maunders., 2005. Status of yellowfin tuna in the eastern Pacific Ocean in 2004 and outlook for 2005. Inter-Amer. Trop. Tuna Comm., Stock Assessment Report, 4: 1-100.

Ianelli J, Lamberson RH., 2003. Introduction to special issue: Modeling in fisheries science, past, present and future. Natural Resource Modeling 16, 337-340.

Maunder MN, Watters GM., 2003 A-Scala: An age-structured statistical catch-at-length analysis for assessing tuna stocks in the Eastern pacific ocean. Bulletin. Inter-American Tropical Tuna Commission 22, 435-437.

Maunder MN, Starr PJ, Hilborn R., 2000. A Bayesian analysis to estimate loss in squid catch due to the implementation of a sea lion population management plan. Marine Mammal Science 16, 413-426.

Maunder MN, 2005 Workshop on stock assessment methods. Report. Inter-American Tropical Tuna Commission, La Jolla, California (USA) 7-11 November, 2005.

McAllister, M. & J. Ianelli. 1997. Bayesian stock assessment using catch-age data and the sampling-importance resampling algorithm. Can. J. Fisheries Aquat. Sci., 54: 284-300.

Quinn, T. J. and R. B. Deriso. 1999. Quantitative Fish Dynamics. Oxford University Press. 542 p.

Serra, J.R. 1991. Important life history aspects of the Chilean jack mackerel, Trachurus symmetricus murphyi. Investigación Pesqueras (Chile), 36: 67-83.

Serra, J.R y C. Canales, 2007. Evaluación y Captura Total Permisible de Jurel (Trachurus symmetricus murphyi) Sub Regional, 2008. Pre-Informe Final, Instituto de Fomento Pesquero, Valparaíso, Chile, 85 pp.

Schwarz, G. 1978. Estimating the dimension of a model. *Annals of Statistics* **6** (2): 461–464