A SIMULATION FRAMEWORK FOR TESTING THE PROCEAN MODEL AND DEVELOPPING BAYESIAN PRIORS. PRELIMINARY RESULTS.

Olivier MAURY, Emmanuel CHASSOT
IRD (Institut de Recherche pour le Développement) UR THETIS
Centre de Recherche Halieutique Méditerranéenne et Tropicale
Avenue Jean Monnet - BP 171, 34203 Sète cedex
E-mail: *olivier.maury@mpl.ird.fr*

DRAFT

INTRODUCTION

The model PROCEAN (PROduction Catch-Effort Analysis) is a statistical catch/effort analysis framework based on a Pella and Tomlinson production model (Maury, 2001). It enables the use of a robust process error structure for the relationship linking fishing mortality to fishing effort, of random walks on catchabilities by fleet to model potential changes in fishing efficiency (Fournier *et al.*, 1998) and of random walks on the carrying capacity parameter. The model uses only catch and effort by fleet data. In its simplest form, without any process error, PROCEAN reduces to a standard non-equilibrium Pella and Tomlinson model. The use of such a production model could be usefull in IOTC where reliable size data are missing for stock assessment. Nevertheless, the highly complex characteristics of Indian Ocean tuna fisheries question the ability of PROCEAN to produce reliable results. The first IOTC working party on method (WPM Sète, 2001) identified different problems who may affect the PROCEAN model behaviour :

- the diversity of fishing patterns present in the fishery for fleets having different fishing effort history (*eg :* longliners fishing for old fish since the begining of the fishery and purse seiners appearing more recently and fishing for younger fish)
- the "one way trip" structure (continuous catch and effort increase over the whole period considered) of catch and effort data which does probably not provide enough contrast in the data to identify easily the production model parameters,
- the general poor quality of fishery data which are probably highly noisy,
- the possibility for important trends in fishing power for many fleets.

To adress those potential problems, the group recommended to test the PROCEAN methodology with realistic simulations. The group also recommended to develop methods for estimating priors for production models, such priors being probably a necessary condition for being able to use production models for Indian Ocean tropical tuna stock assessment.

The aim of this paper is to present the methodology adopted to test the PROCEAN model and to construct priors for its parameters. Because this study is still under progress, the results presented are partial and still provisional.

The questions we try to adress are the following :

1) Is there enough information in a typical «one-way trip » (continuous catch and effort increase over the whole period considered) data serie to estimate the parameters of a simple Pella and Tomlinson model ?
2) Does it make sense to use a simple surplus production model such as a Pella and Tomlinson model to represent a complex age-structured dynamics ?
3) How does the fishing pattern of the fishery change the characteristics (shape of the curve and MSY) of the production model ?

4) How does the PROCEAN model behave when different fleets with different fishing patterns and different effort series are present in the fishery (longline and purse seine for instance) ?
5) How can we determine bayesian priors for generalized production models ? Do such priors significantly improve their use ?
6) How does randomness in the data affect the fit of the PROCEAN model and the estimation of quantities needed for management (MSY and F_{MSY}) ?
7) How does the PROCEAN model behave when the different fleets exibit trends in catchability ?

Because this work is still under progress, only preliminary simulations are presented here.

THE PROCEAN MODEL

The PROCEAN model is based on the classic Pella and Tomlinson (1969) generalized production model which links the stock biomass B to the fishing mortality F by the mean of an ordinary differential equation continuous in time:

$$\frac{dB_t}{dt} = rB_t\left(1-\left(\frac{B_t}{K_t}\right)^{m-1}\right) - F_t B_t \qquad with \quad m > 1 \qquad (1)$$

With B_t, the biomass at time t; F_t, the instantaneous fishing mortality rate; K, the carrying capacity of the stock; r, the per capita intrinsic growth rate of the population and m, the shape parameter (the model becomes a simple Schaefer model when $m=2$).

To introduce catches and effort for multiple fleets into the model, the fishing mortality F_t is expressed as the sum of each fleet's instantaneous fishing mortality:

$$\begin{cases} \dfrac{dB_t}{dt} = rB_t\left(1-\left(\dfrac{B_t}{K_t}\right)^{m-1}\right) - \sum\limits_{i=1}^{n-1} q_{i,t}\, f_{i,t} B_t - C_{n,t} \\ \dfrac{dC_{i,t}}{dt} = q_{i,t}\, f_{i,t} B_t \quad 1 \le i < n \end{cases} \qquad (2)$$

With $n\text{-}1$, the total number of fleets for which fishing effort is available; $q_{i,\,t}$, the catchability coefficient for fleet i at time t, $f_{i,t}$ the mesured nominal fishing effort for fleet i at time t and $C_{i,t}$, the catches for fleet i at time t. $C_{n,t}$ represents the catches for all the fleets non documented in term of effort.
To take into account potential fluctuations of the carrying capacity due to environmental fluctuations or to modifications of the fishery configuration such as stock surface (process errors), the parameter K is assumed to be dependent of time. We assume that the parameters $log(K_t)$ has the structure of a random walk which allows to model slow variations over time (Fournier, 1996) :

$$K_{t+1} = K_t\, e^{J_t - \frac{s_{q_i}^2}{2}} \qquad J \sim N(0, s_J)$$

The local catchability by fleet is also supposed to vary slowly each year to take into account potential fluctuations of fishing power for each fleet (process errors). We assume a random walk structure to the catchability time series for each fleet (Fournier *et al.*, 1998):

$$q_{i,t+1} = q_{i,t}.e^{e_{i,t} - \frac{s_{e_i}^2}{2}} \text{ with } e \in N(0, s_{e_i})$$

To address high-frequency variability of the catchability coefficient, a lognormal process-error structure is assumed for the fishing mortality. Then, the fishing mortality of fleet k at time t is written Concerning the catchability coefficient, the fishing mortality error structure is assumed to be lognormal. Then, the fishing mortality of fleet i in year t is written $F_{i,t} = q_{i,t} f_{i,t} e^{h_{i,t} - \frac{s_i^2}{2}}$ where the h_i are robustified normally distributed random variables with mean 0.

THE SIMULATION MODEL

The model used for generating simulated data is a simple age structured model associated with a Beverton and Holt (1957) stock-recruitment relationship. Randomness is included at both recruitment and catchability levels. The equation for fish number at age a and time t is :

$$N_{a+1,t+1} = N_{a,t} e^{-(F_{a,t}+M_a)} \qquad a<A$$

A plus group accumulates the fish of age $A+$:

$$N_{A,t+1} = N_{A,t} e^{-(F_{A,t}+M_A)} + N_{A-1,t} e^{-(F_{A-1,t}+M_{A-1})}$$

Fig. 1 : Natural mortality at age M used in the simulations.

The total fishing mortality F is defined as the sum of of the fishing mortality for the n fleets i :

$$F_{a,t} = \sum_{i=1}^{n} F_{a,t}^{i}$$

According with the separability assumption, the catchability for each fleet is splitted into an age component, the selectivity, and a time component. Stochasticity is added to the effort/fishing mortality relationship with a multiplicative lognormal error :

$$F_{a,t}^{i} = q_t^i . s_a^i . f_t^i e^{e_{i,t} - \frac{s_q^2}{2}} \qquad e_{i,t} \approx N(0, s_q)$$

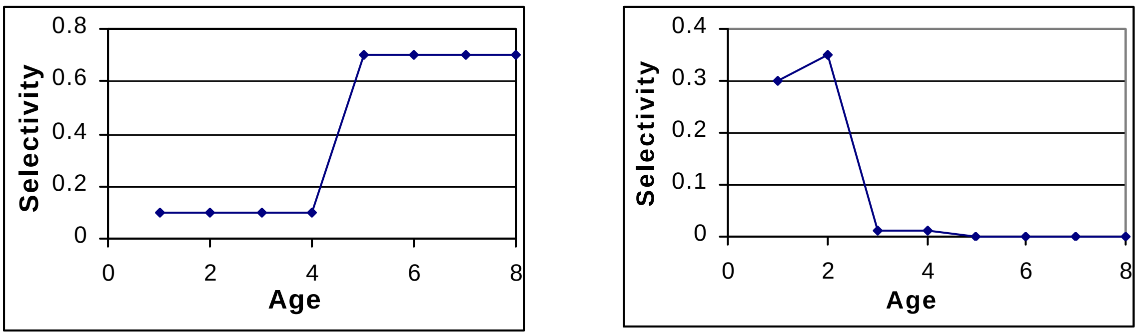

Fig. 2 : Selectivity at age s used in the simulations. Fleet 1 (left) represents a typical longline fishery and fleet 2 (right) represents a typical purse seine fishery.

The catches of age a fishes at time t are calculated with the usual catch equation for each fleet i :

$$C_{a,t}^{i}=\frac{F_{a,t}^{i}}{F_{a,t}+M_a}N_{a,t}^{i}\left(1-e^{-\left(F_{a,t}+M_a\right)}\right)$$

Catches in weight are simply derived by using a weight at age vector :

$$Y_t=\sum_{i=1}^{n}Y_t^i=\sum_{i=1}^{n}\sum_{a=0}^{A}C_{a,t}^{i}W_a$$

A Beverton and Holt stock-recruitment relationship incorporating a lognormaly distributed noise is used to calculate the fish number at age 0 :

$$N_{0,t}=\frac{aSSB_{t-1}}{b+SSB_{t-1}}e^{h_t-\frac{s_R^2}{2}} \quad h_t \approx N\left(0, s_R\right)$$

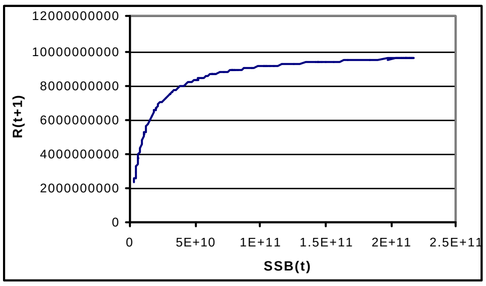

Fig. 3 : The Beverton and Holt stock-recruitment relationship used in the simulations (recruitment at time t+1 as a function of stock spawning biomass at time t).

With a and b, parameters and SSB, the stock spawning biomass calculated as follows :

$$SSB_t=\frac{1}{2}\sum_{a=0}^{A}N_{a,t}W_a\Phi_a$$

The PROCEAN model can be fitted to simulated data. To be able to compare its results to simulations, it it useful to calculate simulated values of parameters such as MSY and F_{MSY}. This is done by using the equilibrium relationship :

$$SSB_e=l_F R_e \Rightarrow R_e=\frac{l_F a-b}{l_F}$$

Where l_F is derived from the previous equations written for a pseudo-cohort and modified to express the terminal age group as follows :

$$N_A=N_{A-1}e^{-\left(F_{A-1}+M_{A-1}\right)}\sum_{j=0}^{+\infty}e^{-j\left(F_A+M_A\right)}$$

SIMULATIONS

In this chapter, the different question raised in the introduction are adressed through various simulations.

Is there enough information in a typical « one-way trip » (continuous catch and effort increase over the whole period considered) data serie to estimate the parameters of a simple Pella and Tomlinson model ?

It is well known that a « one way trip » data series do not provide enough information to estimate simultaneously all the parameters of a generalized production model (*e.g.* Hilborn and Walters, 1992). A simple data serie without any randomness has been constructed to check wether there is enough information in the data to estimate the parameters of the Pella and Tomlinson model in the case of short data series. The simulation has a maximum duration of a hundred years. The selectivity is the typical longline selectivity (fig. 1) and the effort serie increases continuously and reach regularly a plateau to let the catches come back to equilibrium (fig. 4). This serie is supposed to be more informative than a real one way trip with no plateau. PROCEAN is fitted to the simulated data set without any process error. Then, only parameters *r, K, q* and *m* are estimated.

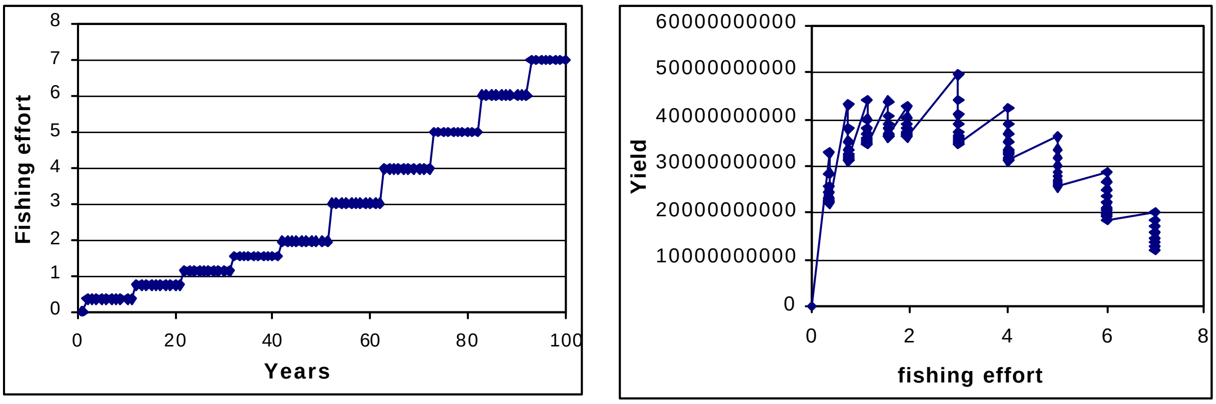

Fig.4 : Fishing effort time serie used in the simulations (left) and corresponding simulated yield as a function of fishing effort (right).

For the longer simulation (100 years), parameters *r, K, q* and *m* are estimated quite precisely (table 1). Because the estimated value for *m* is quite unstable when the data set is shortened, *m* is keeped fixed at 1.2 (1.18 being the estimated value for the longest data set used) and the other parameters are estimated for data set ranging from 10 years to 100 years (fig. 5).

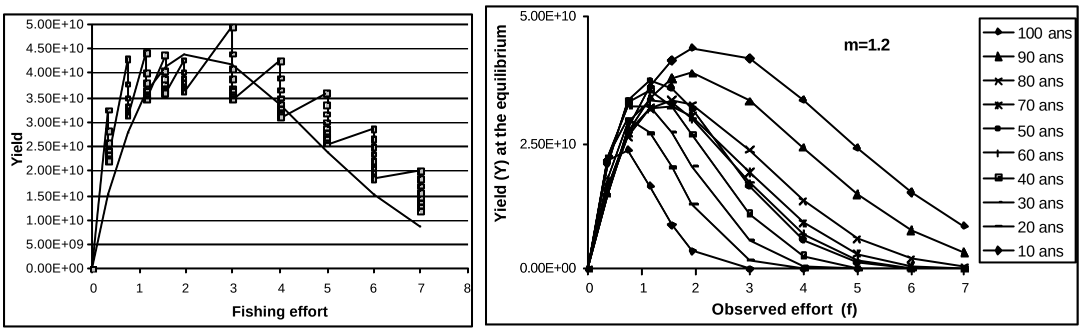

Fig.5 : Left, simulated yield and estimated equilibrium production curve for the longest data set used. Right, estimated production models with m fixed for data series from 10 to 100 years.

Looking at fig. 5, it is quite clear that the fit of the 100 years data serie under-estimate the production for low and high effort values and overestimate it for intermediate values (and overestimate MSY). Then, when the data serie is shortened, the weight of the high effort data progressively disappear and the fitted model modifies to get closer to intermediate and low effort yield data.

This results indicates that even with a quite informative data serie and with the shape parameter *m* fixed at its best value for the longest time serie, there is no stability in the estimated production curve when the data serie is shortened. This seems to indicate that the Pella and Tomlinson production model may not be flexible enough to represent precisely a complex age structured dynamics.

Does it make sense to use a simple surplus production model such as a Pella and Tomlinson model to represent complex age-structured dynamics ?

Surplus production models are based on very simple equations where the whole population is represented with a single state variable : the biomass of the population. This simplicity's counterpart is that surplus production models cannot produce dynamics as complex as age structured model's dynamics. In particular, surplus production models are not designed to treat explicitly the linkage between population dynamics, biomass productivity and the age structure. They are not able to distinguish the fished stock (the exploited fraction of the population) from the whole population which includes non fished age class.

Consider for instance two typical fisheries : a longline fishery fishing only for old large fish and a purse seine fishery fishing only for young small fish. For longliners, fishing will lead to a reduction of the stock relatively to the whole population. For purse seiners, fishing will have the opposite effect (fig. 6).

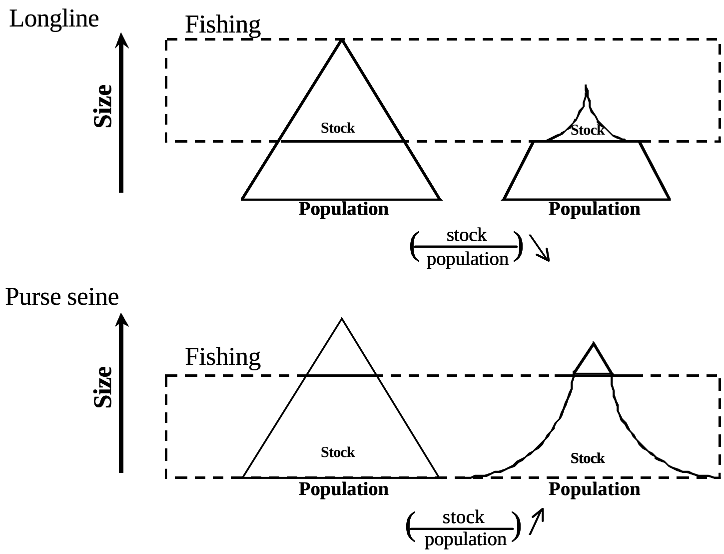

Fig.6 : Schematic population size structure for a typical longline fishery (top) and for a typical purse seine fishery (bottom). When the fishing pressure increases (right), the ratio stock/population decreases for the longliner fishery and it increases for the purse seiner fishery.

This age structured dynamics can be characterized in simulations by defining the catchability of the whole population as the ratio CPUE/population biomass which increases when effort increases for longliners and decreases for purse seiners (Fig. 7).

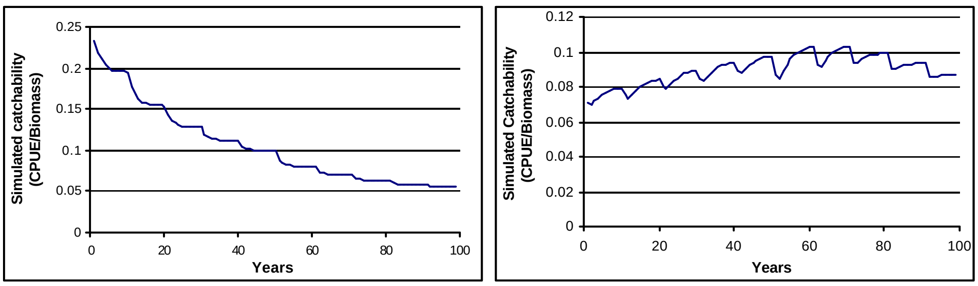

Fig.7 : Simulated population catchability calculated as the ratio stock/population. Left : longline fishery. Right : purse seine fishery.

This age structured processus is not deterministically taken into account in a surplus production model. But PROCEAN enables process error on catchability to be modelled with a random walk. This allows the model to be able to take implicitely into account the trend in catchability due to the effect of age-structure on fish population availability and then to estimate an unbiased population biomass. Fig 8 presents the simulated and the modeled (random walks) trends for catchability for both ll and ps simulated data. Concerning ll, the simulated and estimated curves match exactly. Concerning ps, the simulated and estimated curves do not match exactly because the random walk also compensates for differences between the production model (with m fixed at 1.2 which is the value for ll) and the real production curve.

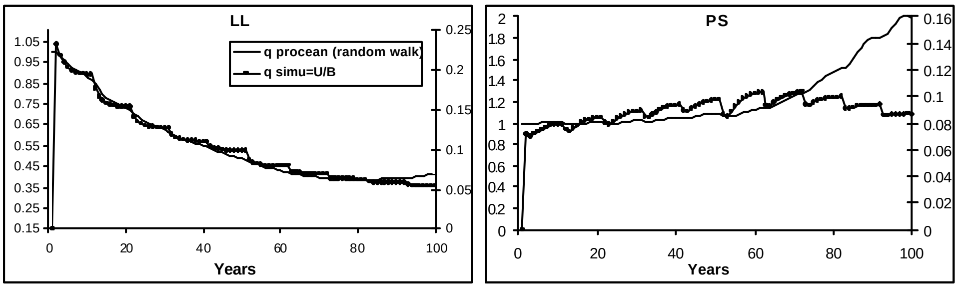

Fig.8 : Simulated catchability and PROCEAN random walk estimates for catchability.

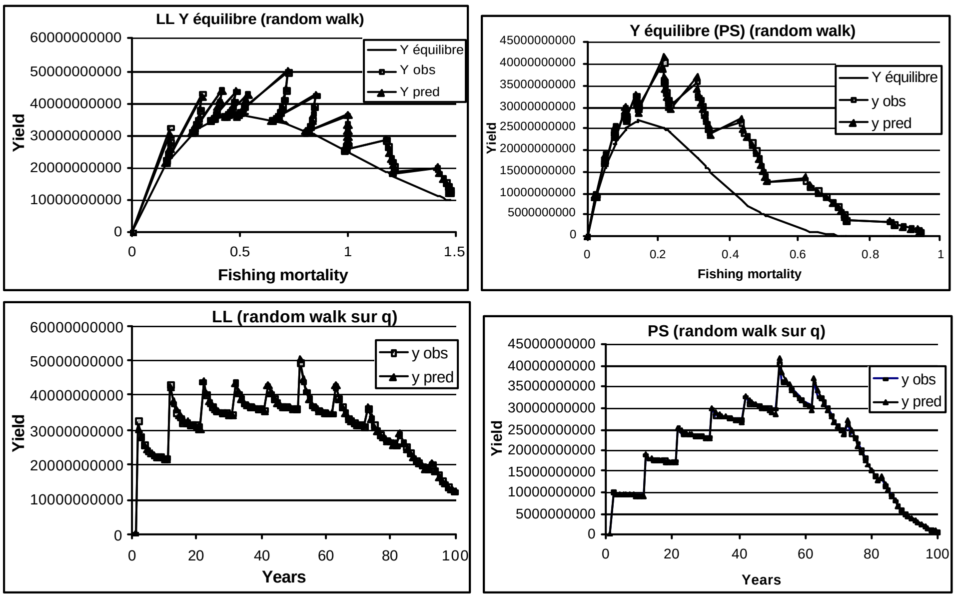

Fig.9 : PROCEAN models fitted to simulated ll (first line) and ps (second line) data taking into account process error in catchability with random walks.

How does the fishing pattern of the fishery change the characteristics (shape of the curve and MSY) of the production model ?

Simulation indicates that, according with the production model theory, the fishing pattern of the fishery do change the shape of the estimated production model.

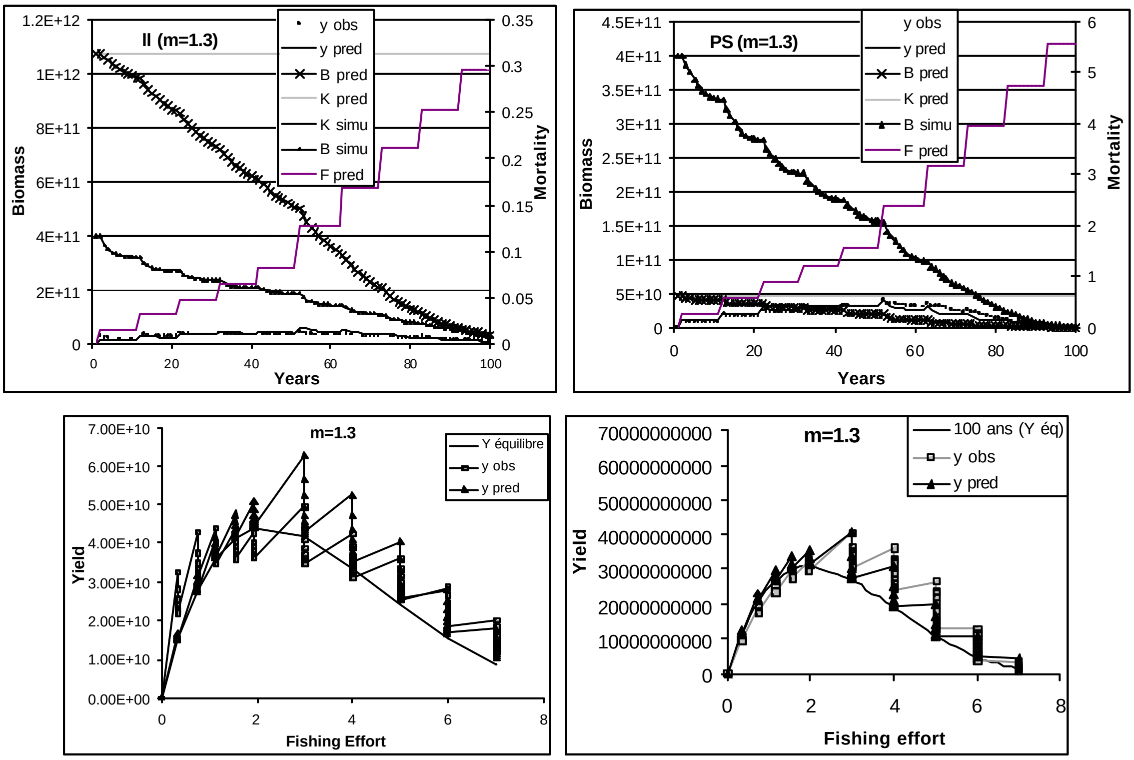

Fig.10 : Fit of the PROCEAN model without any process error. Left, longline fishery and right, purse seine fishery.

How does the PROCEAN model behaves when different fleets with different fishing patterns and different effort series are present in the fishery (longline and purse seine for instance) ?

To analyze the effects of changes of fishing patterns in a mixed fishery, a serie of simulations was realized by mixing two fisheries as follows :

$$F_t = aF_{1,t} + (1-a)F_{2,t}$$

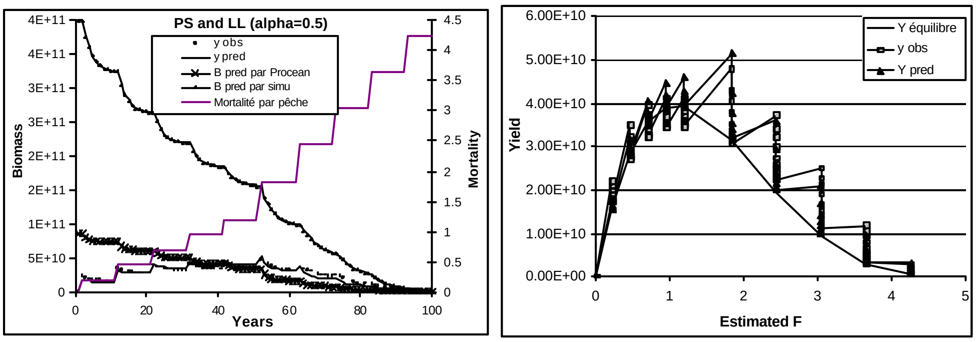

Fig.11 : Fit of the PROCEAN model without any process error for a mixed fishery (longline + purse seine).

How can we determine bayesian priors for generalized production models ? Do such priors significantly improve their use ?

Given the general « one way trip » structure of most tuna fishery data sets, bayesian priors could be profitably used for tropical tunas keeping in mind that priors on parameters distribution may have important consequences on the estimated values of the parameters (McAllister and Kirkwood, 1998). This should be carefully studied and can be estimated by comparing prior distribution with posteriors empirical distributions of the parameters (McAllister and Ianelli, 1997; Punt and Hilborn, 1997).

Given the high negative correlation of *r* and *m* parameters, it is highly desirable to use joint prior probability distributions for those parameters (McAllister and Kirkwood, 1998). Concerning the determination of priors itself, McAllister *et al.* (2000) propose a method for estimating priors on *r* and *m* parameters based on life history traits. Nevertheless, their method doesn't consider explicitely the dependance of the *r* and *m* parameters from the fishing pattern of the fishery. Maunder (*in prep.*) proposes a method for determining a fishery-dependent joint prior distribution for *r* and *m* parameters. His method rely on the calculation of B_{MSY}/K and MSY/B_{MSY} whose values are functions of *m* and *r*. The method requires assumptions about the value of the steepness of the stock-recruitment relationship, about the natural mortality at age, the fecondity at age, the weight at age and the selectivity of the fishery. Given a value for those parameters, Maunder (*in prep.*) uses an age-structured simulation similar to the one presented in this paper to calculate B_{MSY}/K and MSY/B_{MSY} and then to get a value for *r* and *m*. Randomly choosing a value for all the imput parameters in their respective distribution, estimating the corresponding *r* and *m* parameters and repeating the process a high number of times enable Maunder (*in prep.*) to reconstruct the joint prior distribution for *r* and *m*.

The method is consistent but it relies on the implicit assumption that a Pella and Tomlinson production model determined by B_{MSY}/K and MSY/B_{MSY} will characterize properly the production curve over the whole range of effort. We have seen that this may not always be the case (fig.5). Consequently, the production models determined using Maunder's method do not match with the simulated data over its whole span even if the MSY is exactly determined (fig. 12).

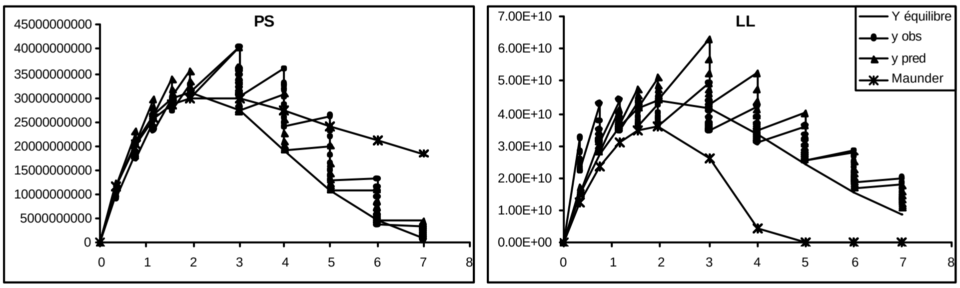

Fig.12 : Simulated data and associated production models fitted with Maunder's method and procean's fit for ps (left) and ll (right).

The method we propose for determining joint priors for *r* and *m* is based on Maunder's method. But instead of determining *r* and *m* parameters from B_{MSY}/K and MSY/B_{MSY}, it estimates it directly by fitting a production model to a highly informative simulated data set (fig12) based on the proper imput parameters (the steepness of the stock-recruitment relationship, the natural mortality at age, the fecondity at age, the weight at age and the selectivity of the fishery).

To account for possible bias due to age structured processus in the simulations, nominal fishing effort may be corrected by the calculated catchability to estimate an effective fishing effort (Fig.13) :

$$f_e = f_n \frac{CPUE}{population\ biomass}$$

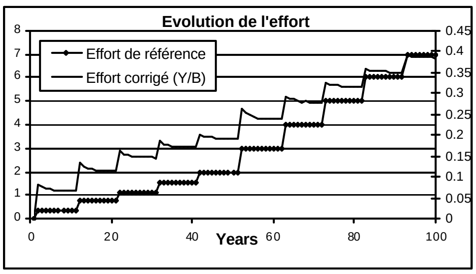

Fig.13 : Nominal fishing effort time serie used in the simulations for the longline fishery and effective fishing effort time serie corrected from the effect of age structure.

This correction seems to reduce the bias and improve the fit of the production model (fig.14).

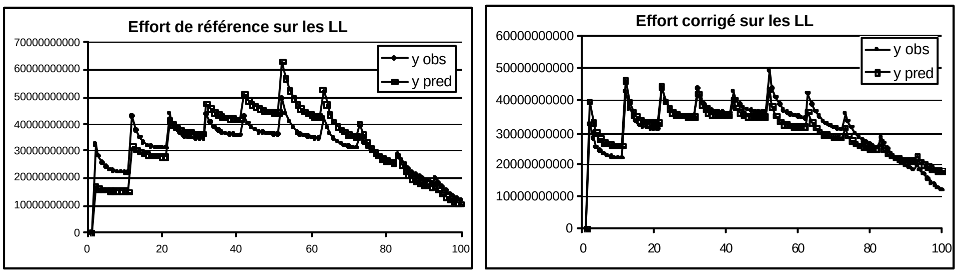

Fig.14 : On the left, simulated production and estimated one from the production model fitted using the nominal effort serie. On the right, simulated production and estimated one from the production model fitted using the corrected effort serie.

Preliminary trials seems to indicate that the absolute values of the estimated parameters are sensitive to the priors used but the value of the MSY and the trends obtained for the carrying capacity, the fishing mortality and the catchability by fleet seems to be robust. If confirmed with simulation trials, this could indicate that the method proposed is adapted to study the catchability evolution by fleet due to technical progress or to changes in fishing strategy and tactics such as changes in targeting practices.

How does randomness in the data affect the fit of the PROCEAN model and the estimation of quantities needed for management (MSY and F_{MSY}) ?

Real data are probably highly noisy, containing a lot of various process and measurement errors. The following simulations aim to measure the bias and estimation error on *r*, MSY and F_{MSY} when randomness is added in the simulated data, either on recruitment, either on catchabilities. A first set of simulations based on a simple production model without any process error is compared with a second set of simulations where the production model includes a robustified lognormal process errors on catchability.

When the PROCEAN model is fitted (m fixed) without any process error to simulated data generated with increasing recruitment variability, the estimated models seem to be increasingly biased (fig.15).

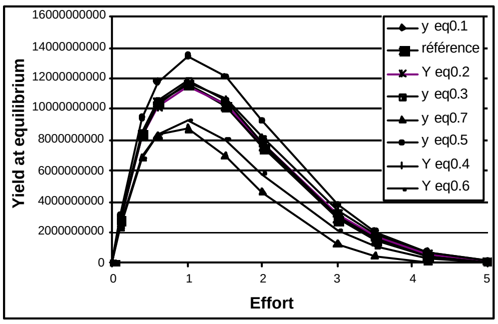

Fig.15 : Fit of PROCEAN model for various standard error of the lognormal recruitment variability.

To determine the effect of noise on recruitment and catchability on bias and variance on estimated quantities needed for management, the mean bias and variances over 100 simulations are presented fig. 16 and fig. 17.

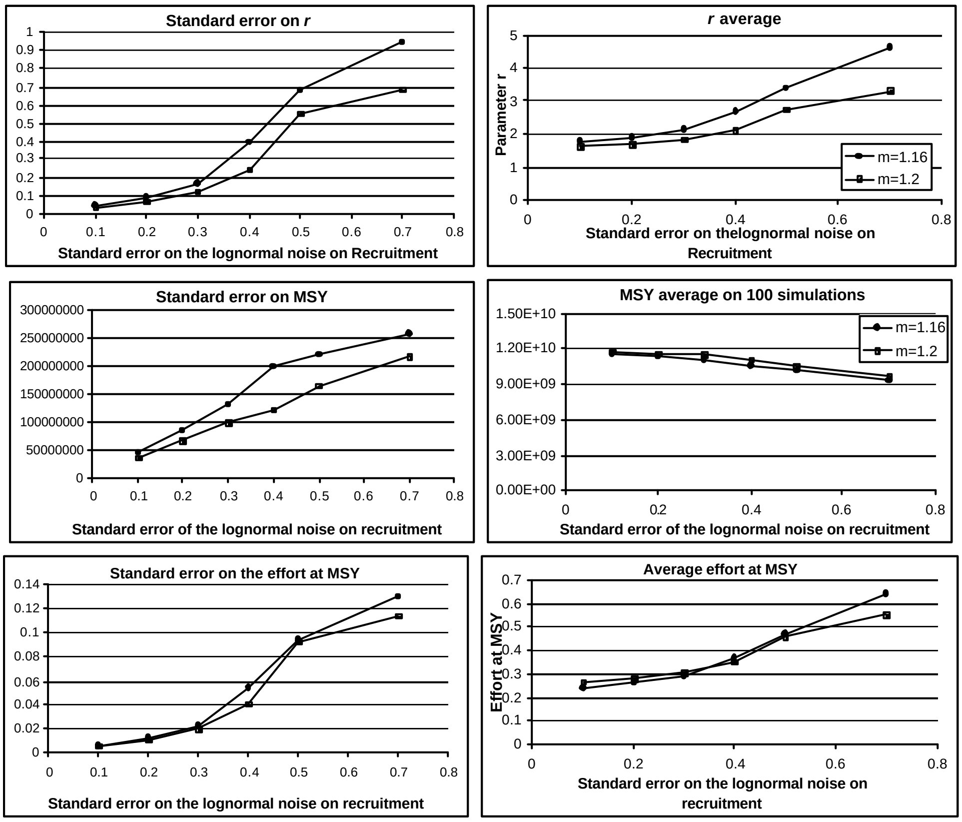

Fig.16 : Effect of noise on recruitment on bias and estimation variance of r *parameter, MSY and* F_{MSY}

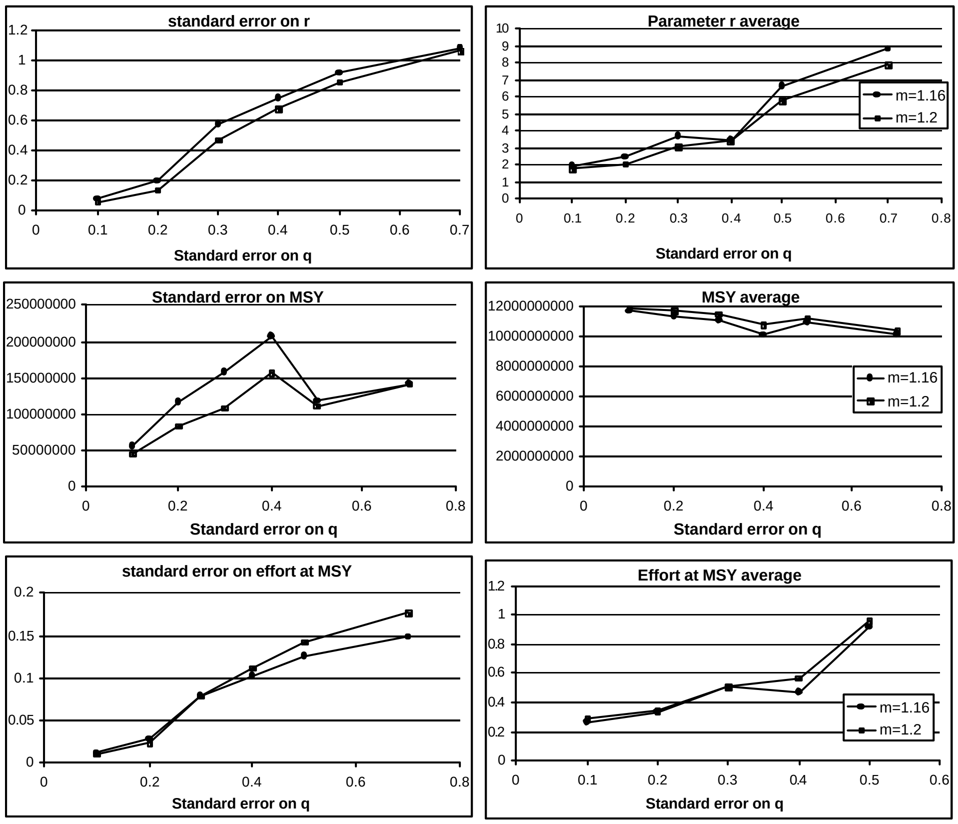

Fig.17 : Effect of noise on catchability on bias and estimation variance of r *parameter, MSY and* F_{MSY}.

How does the PROCEAN model behaves when the different fleets exibit trends in catchability ?

DISCUSSION-CONCLUSION

REFERENCES

Fournier D., 1996. An introduction to AD MODEL BUILDER for use in nonlinear modeling and statistics. Otter research Ltd.

Fournier D.A., J. Hampton and J.R. Sibert, 1998. MULTIFAN-CL : a length-based, age-structured model for fisheries stock assessment, with application to South Pacific albacore, *Thunnus alalunga. Can. J. Fish. Aquat Sci.* **55** : 2105-2116.

Maury O., 2001. PROCEAN, A PROduction Catch Effort Analysis framework to estimate catchability trends and fishery dynamics in a bayesian context. IOTC Working Party on Method, Sète, 23-27 april 2001. IOTC WPM.

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

McAllister M.K. and Kirkwood G.P., 1998. Bayesian stock assessment: a review and example application using the logistic model. *ICES Jour. of Mar Sci.* **55**: 1031-1060.

Pella J.J. and P.K. Tomlinson, 1969. A generalized stock production model. Bull. Inter. Am. Trop. Tuna Com.,**13**: 420-496.

Punt A.E. and Hilborn R., 1997. Fisheries stock assessment and decision analysis: the bayesian approach. *Rev. of Fish Biol. and Fisheries* **7**: 35-63.