CLIMPROD: A FULLY INTERACTIVE EXPERT-SYSTEM SOFTWARE FOR CHOOSING AND ADJUSTING A GLOBAL PRODUCTION MODEL WHICH ACCOUNTS FOR CHANGES IN ENVIRONMENTAL FACTORS

P. FREON(*1), C. MULLON(*2) and G. PICHON(*2)

(1*) ORSTOM, Pôle Caraïbe, BP 81, 97256 Fort-de-France Cédex, Martinique, French W.I., France. (2*) ORSTOM, 70-74 Route d'Aulnay, 93149 Bondy, France.

ABSTRACT

Various equations allowing for the introduction of an environmental variable into different surplus production models were proposed by the first author (1). CLIMPROD is an experimental expert-system, using artificial intelligence, which provides a statistical and graphical description of the data set and helps the user to select the model corresponding to his case according to objective criteria. The software fits the model to the data set using a non-linear regression routine, and assesses the fit with parametric and non-parametric tests, and provides a graphical representation of the results. An example is provided on a pelagic stock, showing the strong effect of environment on long-term variability.

(1)Fréon P., 1983. In: Int. Symp. Long Term Changes Mar. Fish Pop., Vigo. (T. Wyatt and M.G. Larrañeta Eds.), pp. 481-528.

KEYWORDS

Modelling; production model; ptock assessment; environmental effects; population dynamics; computer program; upwelling; expert-system.

INTRODUCTION

Conventional surplus production models are not suitable for certain stocks because fishing effort (E) variations explain only a small part of the total variability of annual production. Often the residual variability originates from the influence of environmental phenomena, which affects either the abundance or the catchability of a stock from one year to the next. Previous attempts to incorporate environmental data in these models were purely statistical, and empirically used multivariate regression analysis, which was unsatisfactory since it did not represent a real modeling approach.

During the Vigo Symposium on Long Term Changes in Marine Fish Populations, various equations allowing the introduction of an environmental variable into different surplus production models were presented (Fréon, 1988). One (sometimes two) additional environmental variable (V) has been inserted into the conventional models in order to improve their accuracy. These variables appear in simple formulae, either at the level of stock abundance, or at the level of the catchability coefficient, or at both levels.

The limitations of this kind of model have been considered. Among various difficulties of application, the risk of obtaining spurious correlations was underlined, as well as the importance of an objective choice of the suitable model. CLIMPROD is an experience in artificial intelligence for choosing the model adapted to each situation, and for assessing the fit. It is designed as an expert-system. Its conception was aided by FAO grants.

METHOD

General presentation

The software is written for PC/XT/AT compatible micro computer using MS-DOS version 3.0 (at least). It is fully interactive and has two main objectives: firstly a normal data management function, whose statistical and graphical utilities use TURBO C language; secondly a guided selection of the appropriate model, showing information path. This part of the model uses an inference motor, written in TURBO PROLOG. It applies about one hundred rules which are interactive with information provided by:

-questions to the user on the stock, independently from the data set (example: life-span of the species?),

-statistics on the data set (example: ratio of effort range on minimum effort value),

-graphical deduction from the data set (examples: does this time series look unstable? Do you see a decreasing relationship on this plot?)

As can be seen in the previous exemple, some questions are asked even when the answer could be obtained from a simple statistical data analysis. In such cases, graphical and/or statistical help is provided but, in order to make the program more clear and pedagogical, the user is required to give an answer. Answering "I don't know" is allowed.

The program is structured and does not necessarily uses the whole set of questions. An exemple of order in the application of the rules is presented in Figure 1.

From the main menu, the user is allowed to open or to select a data file; to update it with a full screen editor; to search for the most suitable model, or to choose one directly; to validate the model, and finally to see the path of the expert decisions.

It must be noted that in order to choose among 30 multivariate models (see appendix 1), the program first performs a regression using the catch per unit of effort (c.p.u.e.) as dependent variable and the effort (or the environment in some cases) as independant variable. From the graphic display of the residuals of this regression against the environmental variable, the user may determine which kind of relationship will link environment and c.p.u.e. in the final multivariate model. This procedure provides an easy interpretation and visualisation of the process for the model choice, and allows for interactive dialogue with the user which can introduce additional information. Nevertheless, recent statistical techniques of optimal transformation for multiple regression (Breiman and Friedman, 1985; Cury and Roy, 1989), could be more powerful and optimal for choosing the model from a strictly statistical point of view, but using nothing other than the multivariate data set (which is often too small for this technique).

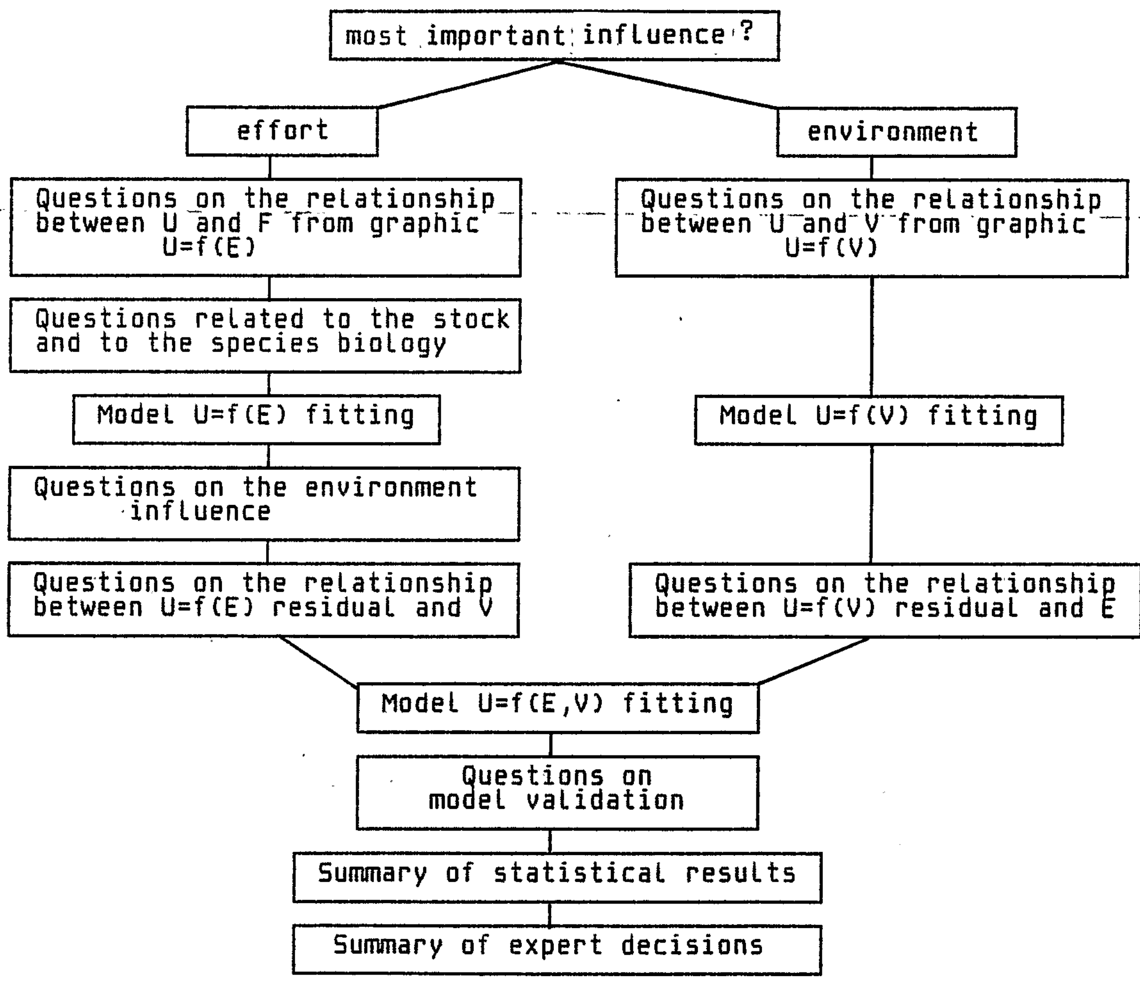

Fig. 1. Partial and simplified flow diagram of CLIMPROD.

Data entry and update

The basic set of data used by CLIMPROD include time series of catch (Y), fishing effort (E), c.p.u.e. (U = Y/E), and one or two environmental variable(s) (V). A full screen editor allows for data entry, correction and updating.

Univariate statistics and graphics

The following statistics are computed for each variable: sample size, average, variance, standard deviation, coefficient of variation, coefficient of skewness and kurtosis, minimum and maximum values, range, median. The distribution of the individual values are plotted along a line in order to show eventual outliers. Although no fishery data could be used if normality were strictly required for modelling, these results may give an idea of the data structure. CLIMPROD stops the analysis, and/or displays advice or warnings, according to the distribution of the values in the different variables.

Bivariate graphics

First each variable is plotted against time (years) in order to detect any strong instability in the series which would hinder the interpretation of the results in some cases (influence of E or V over several years). Then, later in the program, the following relations are plotted: U against E and U against V. According to the apparent type of relation between these variables, it will be decided to adjust first a bivariate model from the function U=f(E) or the function U=f(V).

Questions guiding choice of model

Questions on basic assumptions of surplus production models (see appendix 2) are systematically asked; also the following questions:

-do you think that the effort influence on c.p.u.e is more important than the environmental one (if unknown, yes is assumed)? The answer, guided by statistical and graphical helps, orients first the program either to U=f(E) or to U=f(V) models;

-does the environment influence the abundance, the catchability or both? At that moment the program does not provide any help to answer this question. It is supposed that the user knows the mecanism of action of the environment on the stock, or has already performed time series analyses using a monthly or weekly time interval (Fréon, 1986; 1988) to determine if the environment present a unlagged or short-lagged (influence on catchability) or a lagged relationship with c.p.u.e.

Between these two questions, the program will ask one or several questions in order to determine which relationship is more suitable between U and F (Schaefer's linear model, Fox and Garrod's exponential model or Pella and Tomlinson's generalized model), and between U and V (linear, exponential, general or quadratic). Formulae are presented in appendix 1 and the whole set of questions appears in appendix 2.

Model fitting

In case of non-equilibrium conditions (transitional cases), the equilibrium approximation approach is used (Fox, 1975): a weighted average of E and/or V is computed. In case of delayed influence of the environment on abundance, a lag is inserted between the weighted average of V and U (see Fréon, 1988, for further details). The Marquardt's algorithm is used for least-square estimation of nonlinear parameters. According to the model, the parameter initial values are 1, 0 or computed from the original data set before running the algorithm. As a first result, the percentage of variation explained by the model (R^2) is given. The following steps depend on the quality of the fit, that is:

-after the step of bivariate model estimation, if $R^2<40\%$, the program stops or invite the user to give new answers to the previously unanswered questions. If $R^2>90\%$ a validation of the bivariate model can be tried. If $40<R^2<90\%$, the program will try to find a multivariate U=f(E,V) model providing a better correlation than the bivariate one;

-after a multivariate model estimation , validation intent is possible if $R^2>70\%$.

Fit assessment

The fit assessment is mainly based on a jackknife estimation of the parameters and of R^2 (Ducan, 1978; Efron and Gong, 1983), but also on the residual analysis and on the data set characteristics. Graphics of the predicted values of the model and of its residuals are displayed.

Summary of expert decision

At the end of every step of the main menu, the user may display the path followed by the program at each level of decision, with the corresponding rule number.

RESULTS

An example of application is presented which corresponds to the Senegalese sardine fishery (Fréon, 1983; 1988). A tentative new abundance index has been used i.e. the mean annual weight per set when a single successful set is performed per trip (Fréon, this meeting). The environmental variable influencing the stock abundance is the mean wind speed during the upwelling season.

According to stock and species knowledge, CLIMPROD chooses first to fit the exponential model for the function U=f(F) and find R^2 = 86%. The relationship between the residuals of this model and V is linear (Fig. 2) and therefore the linear-exponential model is fitted and provides an R^2 value equal to 95% (Fig. 3). The jack-knife validation indicates that all parameters are significant at a 5% level, and that no single year contributes to more than 35% of any coefficient estimation, which is relatively satisfactory. The residuals of the model are not autocorrelated.

DISCUSSION

Although the results obtained on the Senegalese stock seem satisfactory, further studies are necessary before accepting the catch per set as a representative index of abundance. Therefore this particular example -given only as an illustration of the program capabilities- will no more be dis-

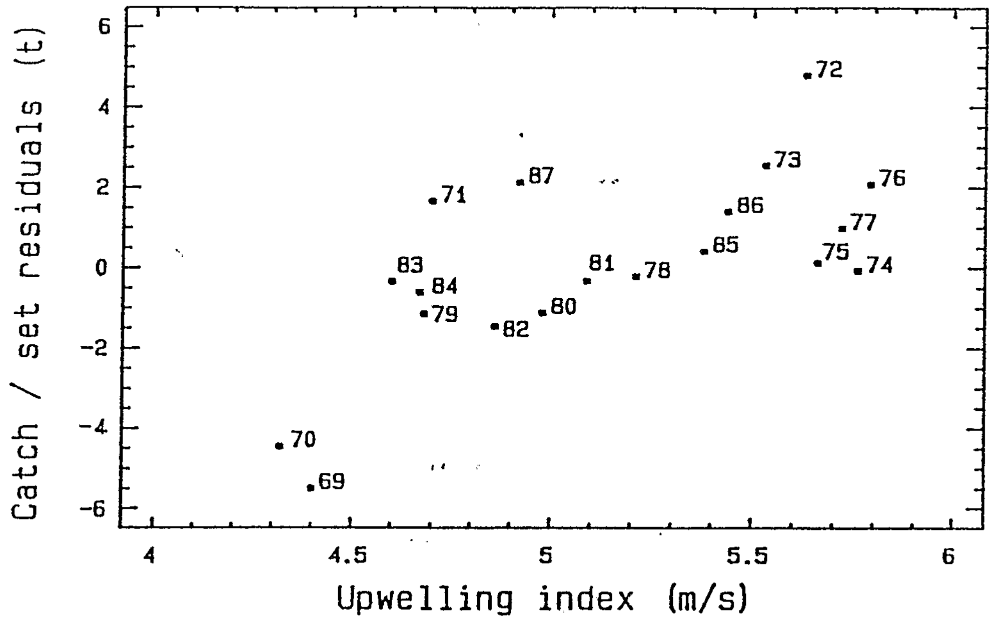

Fig. 2. Coded scatterplot between the upwelling index (two years weighted average of wind speed during the upwelling season) and the catch per set residuals (from a conventional surplus production model; see text). Numbers represent years.

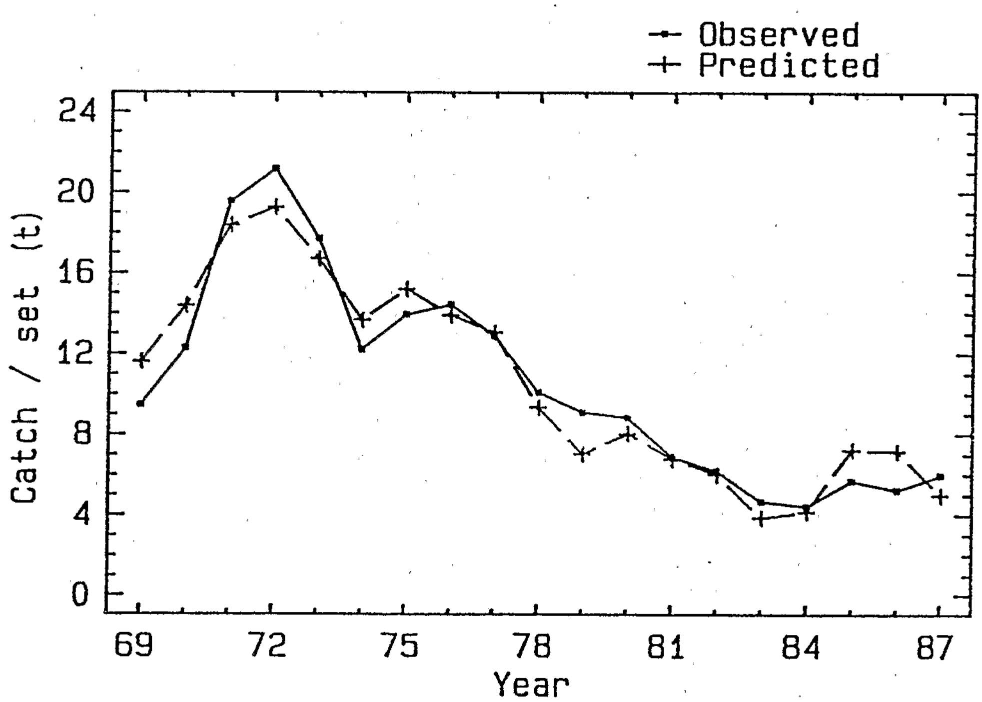

Fig. 3. Predicted, and observed catch per set when introducing an environmental variable in the surplus production model.

Of most interest is the discussion on the improvements, limitations and risks brought by the expert-system approach. Simple surplus production models have been criticized because they suffer from lack of biological realism. Nevertheless, in many instances more sofisticated age-structured models, as proposed by Deriso (1980), do not perform better owing to difficulties in additionnal parameters estimation (Ludwig and Walters, 1985). CLIMPROD use only one additionnal variable and zero to three (but more often one) additionnal parameter to the conventional surplus production models. Moreover, the artificial intelligence allows for using additional quantitative or qualitative data which are not included in the model as variables, but help the user to choose the best model equation according to the stock characteristics, and not only to the criterium of the best fit. This last criterium has been demonstrated to not necessarily provide the more realistic policy prescription (Uhler, 1980). The present approach can provide better assessment and managment of the stock by taking into account the user's knowledge of the stock biology or structure, and the experts' experience on other stocks.

Some negative aspects of CLIMPROD must also be underlined. This tool will be made available to fishery biologists or fishery managers, and can be used to fit any model without special knowledge of population dynamics. Even though, by way of precaution, some questions concerning the basic assumptions are included, the user remains free and responsible for errors. Moreover, the choice of the environmental variable is often the key factor to avoid spurious correlations. In general, even in the case of surplus production models, a minimum knowledge of the stock and of the species biology is required.

CONCLUSION

This experience on artificial intelligence, through the necessary dialog between computer and biology sciences, leads to the formulation of modelling rules, which are often empirical and crude. Such a simplification of the biologist's way of thinking is not devoided of interest. It allows for the exchange of ideas between experts. Moreover, the software itself could be an interesting pedagogic tool, either when used with real data sets or with simulated ones. In this latter case, the limits and interest of the surplus production models -with or without introducing an environmental variable- can be tested (in preparation). It would then appear that the models provide a greater interest where the environmental influence affects catchability (for any species), or in the case of short-lived species, where it affects abundance.

Examples of application were presented in upwelling areas (Fréon, 1988) showing that the CLIMPROD models can provide a fairly good interpretation of fishery history, particularly when a stock collapses unexpectedly without any appreciable increase in the nominal fishing effort. These models can also provide a useful tool for the efficient management of a fishery, particularly in cases where climatic phenomena can be forecast, or when their influence is restricted to the preceding year(s) recruitment. Nevertheless the use of CLIMPROD is not devoid of risks, and careful use must be made by persons having a general knowledge of population dynamics and particular knowledge of the stock and of those species concerned by these models.

REFERENCES

Cury P. and Roy C. (1989). Optimal environmental window and pelagic fish recruitment success in upwelling areas. Can. J. Fish. Aquat. Sci., 46, 670-680.

Ducan, G.T. (1978.). An empirical study of jackknife-constructed confidence regions in nonlinear regression. Technometrics, 20, 123-129.

Efron, B. and G. Gong (1983). A leisurely look at the Bootstrap, the Jack-knife and Cross-validation. The American Statistician, 37, 36-48.

Fox, W.W. (1975). Fitting the generalized stock production model by least squares and equilibrium approximation. Fish. Bull. (U.S.), 73, 23-36.

Fréon, P. (1983). Introduction of environmental variables into global pro duction models. In: Int. Symp. Long Term Changes Mar. Fish Pop., Vigo, (Spain) (T. Wyatt and M.G. Larrañeta, Eds), pp. 481-528.

Ludwig, D. and C.J. Walters (1985). Are age-structured models appropriate for catch-effort data? Can. J. Fish. Aquat. Sci., 42, 1066-1072.

Uhler, R.S. (1980). Least squares regression estimates of the Schaefer pro-duction model : some Monte Carlo simulation results. Can. J. Fish. Aquat. Sci., 37, 1284-1292.

APPENDIX 1: CLIMPROD available models

PUE = f(E) models

PUE=a+b.E	(linear)
PUE=a.exp(b.E)	(exponential)
PUE=(a+b.E)^(1/c-1)	(generalized)

PUE = f(V) models

PUE=a+b.V	(linear)
PUE=a.V^b.	(exponential)
PUE=a+b.V^c	(exponential)
PUE=a.V+b.V^2	(quadratic)

PUE = f(E,V) models; influence of V on abundance

PUE=a.V+b.E	(linear-linear)
PUE=a+b.V+c.E	(linear-linear)
PUE=a.V^b+c.E	(linear-exponential)
PUE=a.V+b.V^2+c.E	(linear-quadratic)
PUE=(a+b.V).exp(c.E)	(exponential-linear)
PUE=a.V.exp(b.E)")	(exponential-linear)
PUE=a.exp(b.E)+c.V+d	(exponential-linear)
PUE=a.V^b.exp(c.E)	(exponential-exponential)
PUE=a.V^b.exp(c.V^d.E) without constraints	(exponential-exponential)
PUE=a.V+b.V^2.exp(c.E)	(exponential-quadratic)
PUE=((a.V^b)+c.E)^(1/(d-1))	(generalized-exponential)

PUE=(a+b.V^2)^(d-1)+c.E)^(1/d-1)) (generalized-quadratic)

D - PUE = f(E,V) models; influence of V on catchability

PUE=a.V+b.V^2.E (linear-linear)
PUE=a+b.V+(c+d.V)^2.E (linear-linear)
PUE=a.V^b+c.V^(2.b).E (linear-exponential)
PUE=a.V.exp(b.V.E) (exponential-linear)
PUE=(a+b.V).exp(c.V.E) (exponential-linear)
PUE=a.V^b.exp(c.E.V^b) (exponential-exponential)
PUE=a.V.(b+c.V)+(b^2.d.V^2+2.b.c.d.V^3+c^2.d.V^4).E (linear-quadratic)
PUE=a.V.(b+c.V)+exp((b.d.V - c.d.V^2).E) (exponential-quadratic)

E - PUE = f(E,V) models; influence of V on both abundance and catchability

PUE=a.V^(b+c)+d.V^(2.b).E (linear-exponential)
PUE=a.V^(1+b)+c.V^(2+b)+d.V^(2.b).E (linear-quadratic)
PUE=a.V^b.exp(c.V^d.E) with sign constraint (exponential-exponential)
PUE=(a.V^(1+b)+c.V^(2+b)).exp(d.V^b.E) (exponential-quadratic)

APPENDIX 2: Questions asked by CLIMPROD

The following questions are asked by CLIMPROD, except the 11 last ones which are not yet incorporated in this experimental version of the software.

-Do you see any abnormal statistics in this table?
-Do you see outlier points on the graphic (E, U and V distribution)?
-Do you see instability on time series graphic (E, U, V)?
-Does the data apply to a single stock?
-Are there sub-stocks?
-Is the considered sub-stock well isolated (with few exchanges) from others?
-May external fisheries considerably affect the stock production?
-Did the exploitation pattern remain the same during the years of observation?
-Is the fishing effort unit standardized, and is the c.p.u.e. proportional to abundance?
-Are you sure that the increasing c.p.u.e. during the first years of exploitation is not due to fishermen learning, to changes in the fleet composition or to technological improvements?
-Were there some important management decisions during the observation period (quota, effort regulation, mesh-size reglementation ...)?
-Have the time lags in processes associated with effort and population changes, and deviations from the stable age structure at any population level, produced negligible effects on production rate?
-Is the influence of effort on c.p.u.e. more important than the environmental influence?
-Does the relationship between effort and c.p.u.e. seem to be decreasing on the graphic?
-Does the relationship between effort and c.p.u.e. seem obviously linear or exponential on the graphic?
-Does the relationship between environment and c.p.u.e. seem monotonic on the graphic?

-Does the relationship between environment and c.p.u.e. seem linear on the graphic? (The two previous questions are the same for the relationship between environment and the residuals of the model U = f(E))
-Do you think that the stock-recruitment relationship has a strong influence?
-Do you think there is a strong parental predation on larvae and/or juveniles?
-Is the fishing effort oriented toward a single target species?
-Is the fishing effort dynamics able to provide sharp increases?
-Did stock already collapse or exhibit drastic catch decrease?
-What is the life-span of the species (0,1,2,3,4,5,6,7,8,9,10,>10 years)?
-Which year classes are significantly exploited?
-Is the ratio life-span/number of exploited year-classes lower than two?
-Are there natural protected areas for the stock, or inaccessible biomass?
-Are there one or several non-negligible spawning before recruitement?
-Is the fecundity of the species very low (sharks, mammals)?
-Do you have any additional reason suggesting a pessimistic modelisation of production?
-Does the environment influence abundance, catchability or both?
-Does the stock exhibit collapse during some years and massive overproduction during others?
-Does environment influence take place before or after the recruitement?
-At which age does the environment influence start?
-How long does the environment influence last?
-Do you see any trend in the model residuals?
-Is ZERO included in the confidence interval of the constant of your model?
-Is ONE included in the confidence interval of any multiplicative or exponential parameter?
-In the jackknife tables, are there some years whose contribution accounts for more than 40% in the model?
-Do you think that during the observed period the stock has been always under-exploited, has at some time reached the optimal exploitation, or has been sometimes over-exploited?
-Does the effort time series exhibit a rather constant increase?
-Does the effort time series exhibit an increasing phase followed by a decreasing one?
-On the U=f(E) graphic, are the decreasing effort points above, under, or superimposed, to the increasing effort points?
-Will the exploitation pattern remain the same during the period of prediction?

FREON P. 3

INTERNATIONAL SYMPOSIUM ON THE LONG-TERM VARIABILITIY OF PELAGIC FISH POPULATIONS AND THEIR ENVIRONMENT

14-18 NOVEMBER, 1989

SENDAI, JAPAN

SECOND CIRCULAR

September, 1988

Organizing Committee of the Symposium

c/o Faculty of Agriculture, Tohoku University, Sendai 980, Japan

Phone: (022)-272-4321 ext. 295   Telefax: (022)-272-1870