

**ANALYSES RETROSPECTIVES
DES RESULTATS D'EVALUATIONS DE STOCKS**

- Ou les raisons d'un acharnement -

Benoît Mesnil

IFREMER
BP 7 - 17137 L'Hourmeau

RESUME

Les analyses rétrospectives sont un moyen pratique pour évaluer ou comparer les performances de diverses méthodes possibles d'estimation sur des jeux de données réels dont on ne peut savoir *a priori* s'ils respectent les hypothèses requises par les méthodes d'estimation considérées. La procédure est illustrée dans le cas des analyses de cohortes avec calibration, où elle sert à sélectionner la méthode ou les options annexes donnant les résultats les plus stables. Elle peut être utilisée dans tout contexte où des paramètres halieutiques sont estimés sur la base de séries pluri-annuelles de données.

ABSTRACT

Retrospective analyses are a pragmatic means of evaluating the properties of alternative assessment methods, or options thereof, on actual data sets whenever one cannot know whether the data comply with the underlying assumptions of the estimation methods. The procedure is exemplified with the case of VPA tuning, where it is commonly used to guide the final choice of method and options, but can also be applied in a variety of contexts where fishery or stock parameters are estimated using time series of data.

INTRODUCTION.

Les évaluations de stocks halieutiques en zones tempérées reposent encore largement sur les analyses de cohortes, tant pour l'analyse des évolutions historiques que pour l'ajustement des paramètres des modèles de simulation. Pour ce dernier objectif en particulier, il importe de disposer de la meilleure estimation possible des effectifs aux âges du stock présents en fin de la dernière année d'analyse, ce que ne permet pas par elle même la procédure classique d'analyse de cohortes n'utilisant que les matrices de captures aux âges.

On a donc assisté, au cours de la dernière décennie, au développement de diverses méthodes dites de calibration ("tuning") destinées à affiner les estimations en tirant parti d'informations supplémentaires disponibles, le plus souvent, sous forme de séries d'indices d'abondance; on peut citer par exemple les méthodes dites "ad hoc" du CIEM (Pope et Shepherd, 1985), CAGEAN (Deriso *et al*, 1985), ADAPT (Gavaris, 1988), XSA (Shepherd, ms, extension de Survivors de Doubleday, 1981), Stock Synthesis (Methot, 1990). Ces méthodes diffèrent surtout par la forme du modèle de capturabilité sous-jacent, la procédure d'estimation, le type d'erreurs prises en compte et les variables affectées (captures ou cpues, dernière année ou toutes), les contraintes. On notera que la plupart visent à combiner les informations fournies par plusieurs séries d'indices.

En principe, la meilleure façon d'évaluer les performances de ces méthodes consiste à les appliquer à des jeux de données générées artificiellement, avec des structures d'erreur ou de biais prédéterminées, puis à comparer les résultats aux "vraies" valeurs. Ceci fut fait, par exemple, lors d'un atelier CIEM en 1988 (CIEM, 1993) où il fut conclu que, parmi les 18 méthodes ou variantes testées, aucune ne pouvait prétendre être LA méthode universelle applicable en toutes circonstances. Elles ont toutes pu donner des résultats plus ou moins gravement erronés sur certains jeux, dès lors que les modèles et hypothèses utilisés pour la génération des données différaient significativement de ceux qui les sous-tendent.

Dans la réalité, le jeu de données relatif à chaque stock, voire à chaque année ou série d'indices additionnelle, constitue un cas particulier dont on ne sait jamais, au moment du traitement, s'il satisfait ou non les conditions requises par les méthodes d'estimation. La conséquence pratique est qu'il faut tester les méthodes d'évaluation et leurs diverses options sur chaque cas réel traité, et définir des critères pour sélectionner la méthode la plus appropriée au problème. Ceci peut être fait en procédant à des analyses rétrospectives, qui font désormais partie des protocoles réguliers des groupes de travail CIEM ou CAFSAC.

PROCEDURE.

Etant donné une méthode (*i.e.*, un modèle + une technique avec ses options) d'estimation de paramètres halieutiques et les données requises étant disponibles pour les années 1 à Y, la démarche générale consiste à appliquer la méthode à des séries temporelles de longueur croissante, en considérant successivement les années 1 à X, 1 à X+1, ..., 1 à Y, et à examiner comment les estimations successives des paramètres varient à mesure que l'on ajoute une année de données. Autrement dit, on compare les estimations basées sur la série de données complète à celles que l'on aurait obtenues avec la même méthode si l'évaluation avait été réalisée N années auparavant à partir des données supposées disponibles alors. Il faut noter toutefois que l'on ne simule pas exactement les circonstances d'une analyse faite il y a N années : dans la réalité, les données relatives à l'année terminale (X par exemple) sont souvent provisoires et mises à jour par la suite tandis que, dans les analyses rétrospectives, on utilise des sous-ensembles du jeu de données mais dans l'état où celui-ci est connu à ce jour. On supprime ainsi une des causes de déviation d'avec les estimations obtenues dans le passé pour s'attacher aux performances de chaque méthode dans l'utilisation d'un fonds de données dont une partie au moins est commune.

Le plus souvent, cette procédure est utilisée pour aider au choix d'une méthode de calibration des analyses de cohortes et à la sélection des paramètres optionnels, mais on peut aussi l'appliquer à toute technique d'estimation de paramètres à partir de séries temporelles de données, comme par exemple celles de prévision du recrutement (Anon., 1993b) ou d'ajustement de modèles globaux.

Normalement, on s'attend à ce que les estimations successives soient relativement stables mais, selon les circonstances (combinaisons données+méthodes+options), il se peut que différents types d'"effets rétrospectifs" se manifestent :

- des biais systématiques, lorsque les estimations relatives aux années "communes" sont toujours réévaluées dans le même sens à mesure que l'on allonge la série de données. On est généralement amené alors à remettre en cause la méthode utilisée, ou du moins le respect des hypothèses sous-jacentes. Il est clair, par exemple, qu'une flottille dont les capturabilités évoluent au fil du temps (ce qui est très difficile à évaluer en temps réel) pose des problèmes aux méthodes qui supposent une constance des capturabilités ; il faut alors éliminer cette flottille du jeu de données ou changer de méthode. Toutefois, certaines options telles que le rétrécissement vers la moyenne ("*shrinkage*") avec un poids approprié ou l'imposition de contraintes sur la forme des diagrammes d'exploitation peuvent modérer l'impact du biais.

- les déviations entre estimations successives sont de signe variable et l'on oscille entre sur-estimation et sous-estimation. Le plus souvent, cela traduit des incohérences dans les données (mauvaise quantification des captures ou de l'effort certaines années, par exemple). Plusieurs méthodes supposent que les indices relatifs à l'année terminale de l'analyse sont exacts; il est évident qu'elles sont particulièrement sensibles à des erreurs sur les indices d'une année lorsque celle-ci est la dernière traitée, alors que l'impact est moindre lorsque l'année en question est "banalisée".

EXEMPLE

Lors de sa réunion de 1992, le groupe de travail CIEM sur l'Evaluation des Stocks Démersaux du Plateau Sud (Anon., 1993a) a cherché à systématiser le recours aux analyses rétrospectives afin de choisir, pour chaque stock, la méthode de calibration des analyses de cohortes qui paraissait la plus pertinente. Ainsi, pour le stock nord de merlu, il disposait des données de captures aux âges pour les années 1978 à 1991 et de 5 séries d'indices d'abondance (4 basées sur les cpue de flottilles professionnelles, 1 sur des campagnes scientifiques). Chaque méthode de calibration a été appliquée en prenant successivement les années 1986 à 1991 comme années terminales de l'analyse. Chaque réplication fournit de nouvelles estimations des mortalités par pêche et des effectifs aux âges pour la période historique s'étendant de 1978 à l'année terminale considérée. Pour simplifier, on ne considèrera que les séquences des mortalités par pêche moyennes calculées entre les âges 1 à 4; on sait que, dans une analyse de cohortes, une sous-estimation des mortalités par pêche se traduit par une sur-estimation des effectifs (dont le recrutement des cohortes) et des biomasses, et vice-versa.

L'analyse rétrospective des estimations obtenues par la méthode de Laurec et Shepherd (1983) est illustrée sur la figure 1. Cette méthode suppose que les capturabilités aux âges pour chaque série d'indices sont constantes, et que les captures aux âges ainsi que les indices d'abondance relatifs à l'année terminale sont connus sans erreur. On remarque que la prise en compte de données additionnelles entraîne une remise en cause souvent notable des estimations obtenues précédemment. Ainsi, l'estimation de la mortalité par pêche moyenne pour 1989 passe d'environ 0,1 lorsque cette année est prise comme année terminale à environ 0,2 puis 0,25 lorsque l'on utilise le jeu de données complet. Par ailleurs, les estimations réalisées en partant des années 1986 à 1988 traduisent une forte instabilité qui, du fait de la faible convergence (mortalités par pêche proches de la mortalité naturelle supposée de 0,2) se répercute sur toute la série historique. Ici, l'analyse rétrospective a surtout confirmé l'existence de graves incohérences dans les données de captures aux âges disponibles pour les années 1985 et 1986; tant qu'elles ne sont pas corrigées, ces anomalies empêchent toute utilisation raisonnable de procédures basées sur les analyses de cohortes. On aurait pu, cependant, en réduire les effets en utilisant des pondérations appropriées.

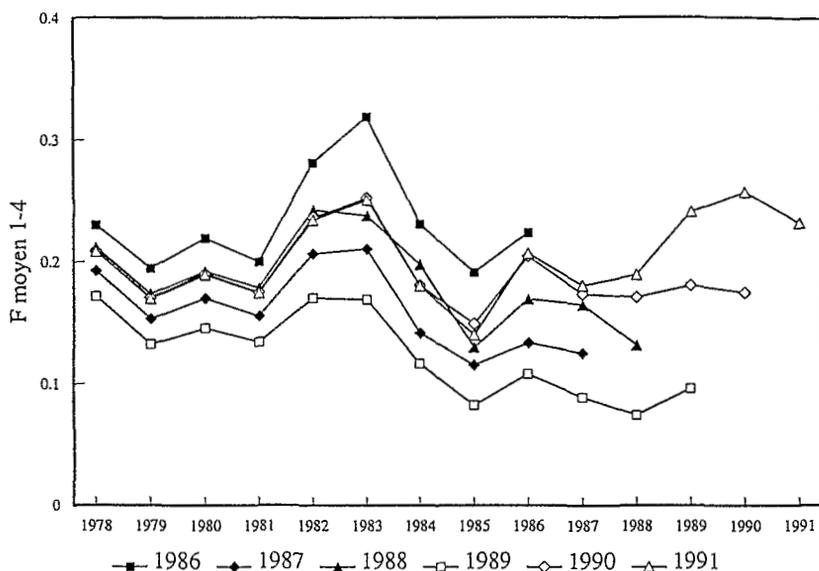


Figure 1 : Analyses rétrospectives des mortalités par pêche moyennes (âges 1 à 4) estimées par analyse de cohortes. Calibration par la méthode de Larec et Shepherd.

Le même exercice a été réalisé en utilisant la méthode XSA ("eXtended Survivors Analysis") qui diffère de la précédente en particulier par le fait qu'elle utilise une plus grande part de l'information contenue dans la matrice des captures et ne suppose pas que les indices relatifs à l'année terminale sont connus exactement. Deux variantes ont été utilisées, tenant compte ou non d'un rétrécissement vers la moyenne "historique" des mortalités par pêche estimées pour chaque âge (figures 2a et 2b). On remarque que les estimations sont plus durablement affectées par l'anomalie des données relatives à 1985 et 1986, surtout si l'on incorpore la moyenne historique. En revanche, les estimations obtenues en incluant progressivement les années 1989 à 1991 ne diffèrent pas significativement. Ceci a conduit à adopter cette méthode pour l'évaluation finale.

Cet exemple a été choisi délibérément pour illustrer la gravité potentielle des effets rétrospectifs lorsqu'ils surviennent. On doit toutefois à la vérité de signaler que l'évaluation de ce stock conduite en 1993 sur une base de données révisée, avec une meilleure définition des flottilles fournissant les indices d'abondance, a produit des analyses rétrospectives moins pathologiques, en particulier lorsque l'on assignait des valeurs appropriées au rétrécisseur. Ce n'est pas le moindre intérêt des analyses rétrospectives problématiques de 1992 que d'avoir suscité cette révision des données.

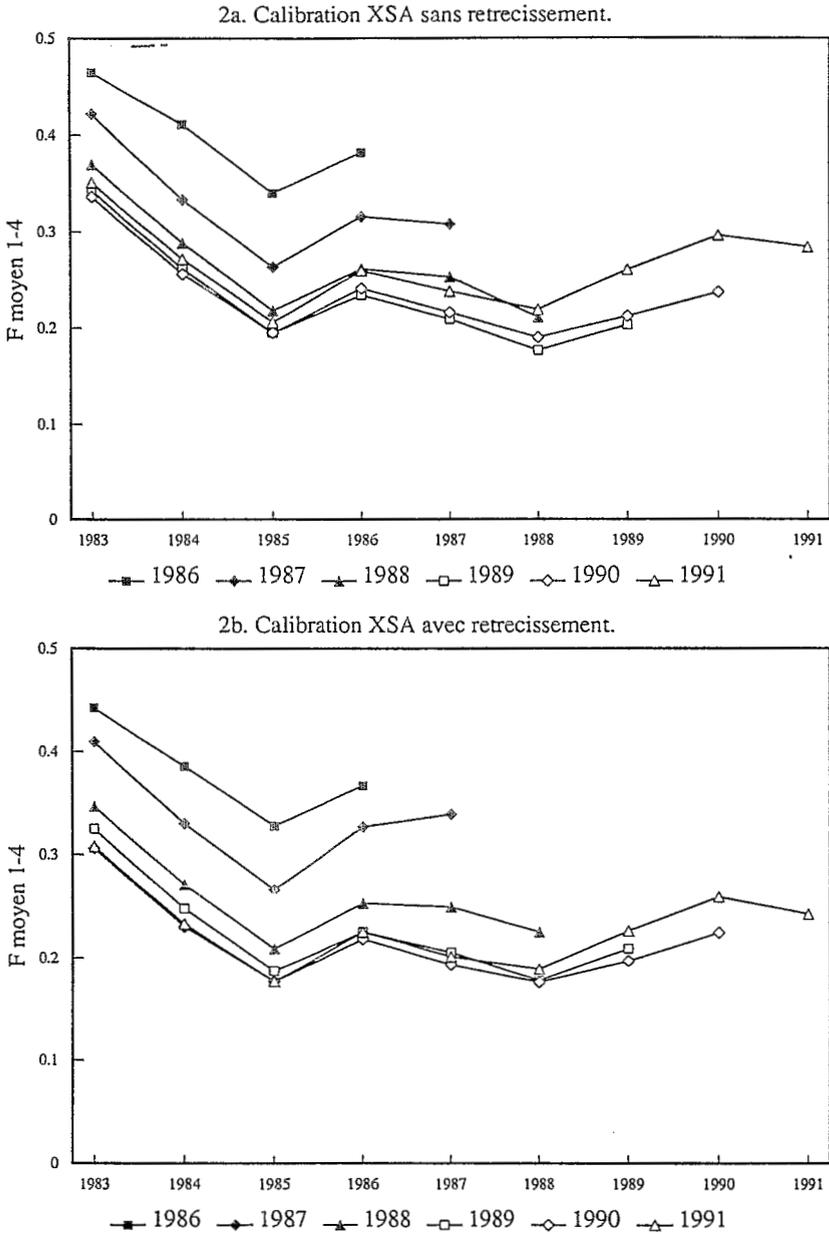


Figure 2. Analyses rétrospectives des mortalités par pêche moyennes (âges 1 à 4) estimées par analyse de cohortes. Calibration par la méthode XSA avec et sans rétrécissement vers la moyenne historique.

DISCUSSION

Du fait que l'adéquation entre les caractéristiques statistiques des données et les exigences des méthodes d'estimation ne peut être testée *a priori*, les analyses rétrospectives semblent être un des seuls moyens pratiques d'évaluer sur des cas concrets la validité des procédures couramment utilisées pour les évaluations de stocks halieutiques. De fait, plus elles sont appliquées, à des données de pêcheries plus diverses, plus elles mettent en évidence le fait que toutes les méthodes connues, aussi bien fondées soient-elles au plan théorique, peuvent buter sur des problèmes dans certaines circonstances. Aucune n'est suffisamment robuste pour admettre de graves violations des hypothèses qui la sous-tendent.

La démarche reste toutefois empirique et ne permet pas toujours d'identifier les causes d'effets rétrospectifs ou d'estimations aberrantes. Dans le cas des analyses de cohortes, celles-ci peuvent résulter de :

- la spécification du modèle: capturabilité constante ou variable au fil du temps, séparabilité des mortalités par pêche en effet-âge * effet-année, uniformité de la distribution du stock et des efforts de pêche, stabilité des mortalités naturelles,
- la structure des erreurs: origine (mesure ou processus), loi de distribution, type (systématique ou aléatoire),
- des erreurs sur certains paramètres clés (mortalité naturelle, migration).

Le plus souvent, hélas, on est amené à mettre en cause la qualité des données : erreurs d'âgeage ou de mesure de l'effort variables d'une année sur l'autre, incohérences diverses. On constate en effet que les effets rétrospectifs graves sont moins souvent un problème attaché à une méthode qu'à un jeu de données sur lequel ils tendent à persister quelle que soit la méthode d'estimation utilisée. Les analyses rétrospectives aident d'ailleurs, en complément des diagnostics (résidus, variances, etc.) fournis par les logiciels, à localiser les données anormales. Il suffit parfois de supprimer certaines séries d'indices ou de les restreindre à une partie des groupes d'âge pour améliorer notablement la stabilité des estimations. Il faut toutefois remarquer, comme on le fait pour certaines analyses de sensibilité ou de risque, que les analyses rétrospectives ne devraient pas s'arrêter à la phase d'estimation (l'analyse de cohortes) mais devraient s'intéresser aux résultats des simulations comme l'ont fait Rivard et Foy (1987) : en dernier ressort, c'est la stabilité des prévisions de capture qui importe. Soulignons enfin que l'absence de dérive rétrospective ne prouve aucunement que les estimations sont justes !

A l'expérience, il apparaît donc que l'on n'a pas tant besoin de développer des méthodes hyper-sophistiquées que d'améliorer la robustesse et les propriétés opérationnelles de celles qui existent, d'autant que ces dernières ont fini par converger sur des principes communs. C'est pourquoi tant d'efforts restent mobilisés au sein des commissions scientifiques pour introduire, puis valider expérimentalement, des options (pondérations, contraintes, rétrécisseurs, etc.) permettant de modérer l'impact de certaines violations des hypothèses par les données disponibles. Dans le même temps, il faut éviter que cela n'aboutisse à une trop grande flexibilité des procédures qui obligerait les pauvres utilisateurs, dans le contexte stressant d'un groupe de travail, à choisir parmi une débauche d'options dont ils auraient des difficultés à maîtriser les implications. On retrouve là un des problèmes fondamentaux de l'halieutique quantitative : les méthodes d'évaluation doivent à la fois posséder de sérieuses bases théoriques, être robustes aux imperfections des données, et être utilisables sans grand risque par les halieutes "de terrain".

REFERENCES

- Anonyme, 1993a. Report of the Working Group on the Assessment of Southern Shelf Demersal Stocks. Cons. int. Explor. Mer, CM 1993/Assess:3, 393 p.
- Anonyme, 1993b. Report of the Working Group on Methods of Fish Stock Assessment. Cons. int. Explor. Mer, CM 1993/Assess:3, 393 p.
- CIEM, 1993. Report of the Working Group on Methods of Fish Stock Assessment (1987, 1988, 1989). Cons. int. Explor. Mer, Coop. Res. Rep., 191: 249 p.
- Deriso, R.B., Quinn, T.J. II et Neal, P.R., 1985. Catch-age analysis with auxiliary information. Can. J. Fish. Aquat. Sci., 42(4): 815-824.
- Doubleday, W.G., 1981. A method of estimating the abundance of survivors of an exploited fish population using commercial catch at age and research vessel abundance indices. In: Doubleday, W.G. et Rivard, D. (Ed.), Bottom trawl surveys. Can. Spec. Publ. Fish. Aquat. Sci., 58: 273 p.
- Gavaris, S., 1988. An adaptive framework for the estimation of population size. CAFSAC Res. Doc. 88/29: 12 p.
- Laurec, A. et Shepherd, J.G., 1983. On the analysis of catch and effort data. J. Cons. int. Explor. Mer, 41: 81-84.
- Methot, R., 1990. Synthesis model: an adaptable framework for analysis of diverse stock assessment data. Int. N. Pacific Fish. Comm. Bull., 50: 259-277.
- Pope, J.G. et Shepherd, J.G., 1985. A comparison of the performance of various methods for tuning VPAs using effort data. J. Cons. int. Explor. Mer, 42: 176-184.
- Rivard, D. et Foy, M.G., 1987. An analysis of errors in catch projections for Canadian Atlantic fish stocks. Can. J. Fish. Aquat. Sci., 44: 967-981.