

REGARDS STATISTIQUES ET INFORMATIQUES SUR

L'ANALYSE DES PUISSANCES DE PECHE ET DES ABONDANCES APPARENTES

A. LAUREC ⁽¹⁾ et J.B. PERODOU ⁽²⁾

Résumé : l'application du modèle linéaire à l'estimation des puissances de pêche et abondances mène à des systèmes d'une dimension qui peut être élevée. Il est possible de réduire la taille des matrices manipulées (MORRISON, 1976). Les auteurs présentent ainsi l'adaptation de la technique de substitution aux systèmes associés aux méthodes de Robson (1966), Laurec et Fonteneau (1979). Le problème de l'estimation des variances-covariances des estimateurs est également discuté, l'accent étant mis sur la distinction des techniques de résolution à appliquer, selon que les variances-covariances sont ou non recherchées. Deux annexes sont consacrées respectivement à l'évolution de l'importance pratique de la correction de Laurent (1963), et au traitement d'un cas où sont connues captures cumulées par marée et temps de pêche dans les différentes strates, mais non les captures par strate.

SUMMARY : multiplicative models, applied to the estimation of fishing powers lead after logarithmic transformations to linear systems, the dimension of which can be large. It is however possible to decrease the dimensions of the matrices corresponding to the linear systems created (Morrison, 1976) through a substitution technique. This technique is applied to the classical methods due to Robson (1966) Laurec et Fonteneau, 1979) Gavaris, (1980). The estimation of the variance-covariance of the output is discussed, the duality between techniques calculating such quantities, and simpler ones providing just a best fit solution. The practical influence of the correction factor (Laurent, 1963) discussed in a first appendix. A second appendix is devoted to a situation where catches by vessel and stratum are not available, but only catches by trip and vessel, as well as a the fishing time in the different strata.

(1) Laboratoire ERHAL - Centre IFREMER de NANTES
BP 1049 44037 NANTES CEDEX

(2) Laboratoire R.H - Station IFREMER de LORIENT
8, rue F. Toullec 56100 LORIENT

I N T R O D U C T I O N

L'application du modèle linéaire à l'estimation des puissances de pêche et abondances apparentes est relativement ancienne. La séparation de l'impact des puissances de pêche et abondances apparentes par strate est en effet suggérée par GÜLLAND (1956), ROBSON (1966). Par transformation logarithmique, on se ramène au modèle linéaire à deux facteurs. Il est alors possible de résoudre directement le système linéaire ainsi créé, par exemple, en inversant la matrice ad hoc.

Cette procédure a été à l'origine du programme FORTRAN développé par BERUDE (in ABRAMSON, 1971 et SIMS, 1986). Ce logiciel nécessite l'inversion d'une matrice symétrique dont la dimension est $N_b + N_s - 1$, si N_b est le nombre de bateaux, N_s le nombre total de strates spatio-temporelles. L'analyse de cas concrets (Charuau et al 1985, 1986) où N_b avoisinait 400, N_s 100, a créé des problèmes pratiques, liés aux moyens informatiques requis. Modérées en termes d'espace mémoire requis, ces difficultés sont beaucoup plus graves en matière de temps de calcul. On sait en effet que le temps nécessaire à l'inversion d'une matrice croît comme le cube de sa dimension. La définition d'une technique de calcul permettant de réduire les dimensions mises en cause est ainsi d'un intérêt majeur. La technique présentée dans les pages qui suivent fera que le système linéaire à résoudre aura pour dimension le plus petit des nombres N_b et N_s . Dans l'exemple évoqué, la réduction du temps de calcul sera égale à $5^3 = 125$. Les calculs deviennent ainsi possibles, même sur micro ordinateurs dans un très grand nombre de situations. La procédure utilisée, classique pour les statisticiens spécialisés dans le modèle linéaire (cf. par exemple MORRISON, 1976) n'a pas fait à notre connaissance l'objet d'un exposé adapté aux problèmes de puissances de pêche. C'est cette lacune que le présent document s'efforce de combler.

La procédure est adaptable au cas où l'abondance apparente se décompose en deux effets multiplicatifs, le plus souvent une composante annuelle et une composante liée à un effet croisé de secteur et saison. Après transformation, cela correspond à un modèle linéaire à trois facteurs. Il a notamment été utilisé par Laurec et Fonteneau (1979), Gavaris (1980), discuté dans Anon (1984). Le système à résoudre aura pour dimension la somme des deux plus petits nombres N_a , N_b et N_s , N_b étant le nombre d'années, N_s le nombre de strates spatio-temporelles à l'intérieur de l'année.

Le premier paragraphe du présent article est consacré à un rappel sur le problème fondamental d'ajustement. Le second applique la technique de substitution à la résolution des problèmes linéaires, sans rechercher des inférences sur les résultats obtenus. Le troisième paragraphe est consacré aux problèmes de variance.

Une première annexe présente quelques remarques sur la correction dite de LAURENT (1963). Une seconde traite d'un problème connexe, non linéarisable où sont connus pour chaque navire les captures par marées et les temps de pêche par secteur, mais non les captures par secteur.

I - AJUSTEMENT DU MODELE LINEAIRE

I-1 - Notations

Un ensemble de Nb flottilles ou bateaux indicés par i sera considéré. La puissance de pêche du bateau i est P_i , son logarithme est p_i . Les bateaux opèrent selon le modèle de ROBSON dans N_s strates spatio temporelles indicées par j . L'abondance apparente est notée A_j , son logarithme a_j .

Dans le modèle à trois facteurs, où plusieurs années sont considérées, avec séparation d'un effet annuel et d'un effet "spatio-saisonnier", les années sont indicées par k , $k = 1, \dots, N_y$. Les strates restent indicées par j ($j = 1, \dots, N_s$). L'abondance logarithmique apparente sera ainsi $q_j + d_k$, d_k étant l'indice annuel, q_j le facteur spécifique de la strate.

Chaque observation correspond à une prise par unité d'effort (p.u.e). Les observations sont indicées par ℓ ($\ell = 1, \dots, N_o$). Après passage au logarithme la p.u.e est u_ℓ . Chaque observation se rapporte à un bateau $b(\ell)$ à une strate $s(\ell)$, éventuellement à une année $y(\ell)$. Un poids éventuel m_ℓ peut être accordé à l'observation ℓ . Enfin il sera fait un large usage du symbole de Kronecker :

$$\delta_{m,n} = 1 \text{ si } m = n, = 0 \text{ si } m \neq n$$

Avec ces notations, le modèle de Robson (1966) s'écrit

$$(1) \quad u_{\ell} = p_b(\ell) + a_s(\ell) + \varepsilon_{\ell}$$

où ε_{ℓ} est un résidu. Dans le modèle simple, les ε_{ℓ} sont centrés, normaux, mutuellement indépendants.

On sait que les p_i sont définis à une constante additive près. Pour simplifier, on posera $p_1 = 0$. Dans la suite les sommations sur i le feront toujours courir de 2 à Nb.

Le modèle à trois facteurs utilisé par Laurec et Fonteneau (1979), GAVARIS (1980), Anon (1984) s'écrit quant à lui

$$(2) \quad u_{\ell} = p_b(\ell) + d_y(\ell) + q_s(\ell) + \varepsilon_{\ell}$$

Les ε_{ℓ} font l'objet dans le modèle simple des mêmes hypothèses que précédemment.

Pour lever l'indétermination, on posera $p_1 = 0$ et $d_1 = 0$.

Les équations et systèmes les plus importants sont encadrés et numérotés en chiffres arabes pour les équations, en chiffres romains pour les systèmes.

I-2 - Ajustement du modèle de ROBSON

L'ajustement par les moindres carrés pondérés conduira à minimiser :

$$\phi = \sum_{\ell=1}^{N_0} (u_{\ell} - p_{b(\ell)} - a_{s(\ell)})^2$$

ϕ est fonction des valeurs attribuées au $N_b - 1$ puissances logarithmiques de pêche p_i , ($i = 2, \dots, N_b$) et aux N_s abondances logarithmiques. Pour minimiser ϕ , on annulera ses dérivées qu'il convient donc d'expliquer.

$$\text{Ainsi } \frac{d\phi}{dp_i} = -2 \sum_{\ell} (w_{\ell} \delta_{i,b(\ell)} (u_{\ell} - p_i - a_{s(\ell)}))$$

L'annuler équivaut à satisfaire l'égalité

$$(3) \quad w_i \cdot p_i + \sum_{j=1}^{N_s} w_{i,j} a_j = V_i$$

$$\text{où } w_{i,j} = \sum_{\ell} \delta_{i,b(\ell)} \delta_{j,s(\ell)} m_{\ell}$$

$$w_i = \sum_{j=1}^{N_s} w_{i,j}$$

$$V_i = \sum_{\ell} \delta_{i,b(\ell)} m_{\ell} u_{\ell}$$

De même annuler $\frac{d\phi}{da_j}$ mène à :

$$(4) \quad \sum_{i=2}^{N_b} w_{i,j} p_i + w_{.j} a_j = V_{.j}$$

$$\text{avec } w_{.j} = \sum_{i=2}^{N_b} w_{i,j}$$

$$V_{.j} = \sum_{\ell} \delta_{j,s(\ell)} m_{\ell} u_{\ell}$$

I-3 Ajustement du modèle à trois facteurs

On minimisera cette fois

$$\Psi = \sum_{\ell} m_{\ell} (u_{\ell} - p_b(\ell) - q_s(\ell) - d_y(\ell))^2$$

Les inconnues sont les $(Nb - 1)$ p_i ($i = 2, \dots, Nb$), les $(Ns - 1)$ q_j ($j = 2, \dots, Ns$), et les Ny d_k ($k = 1, \dots, Ny$)

La même procédure que précédemment conduit au système défini par les équations (5), (6) et (7)

$$(5) \quad w_{i..} p_i + \sum_{j=2}^{Ns} w_{i,j} q_j + \sum_{k=1}^{Ny} w_{i,k} d_k = v_{i..}$$

pour $i = 2, \dots, Nb$

$$(6) \quad \sum_{i=2}^{Nb} p_i w_{i,j} + w_{.j} d_j + \sum_{k=1}^{Ny} w_{.j,k} d_k = v_{.j}$$

pour $j = 2, \dots, Ns$

$$(7) \quad \sum_{i=2}^{Nb} p_i w_{i,k} + \sum_{j=2}^{Ns} w_{.j,k} q_j + w_{..k} d_k = v_{..k}$$

pour $k = 1, \dots, Ny$

$$\text{Avec } w_{i,j,k} = \sum_{\ell} \delta_{i,b(\ell)} \delta_{j,s(\ell)} \delta_{k,y(\ell)} m_{\ell}$$

$$w_{i,j} = \sum_{k=1}^{Ny} w_{i,j,k}$$

$$w_{i..} = \sum_{j=2}^{Ns} w_{i,j}$$

$$v_{i..} = \sum_{\ell} \delta_{i,b(\ell)} u_{\ell} m_{\ell}$$

Les autres termes se construisent de même.

II - RESOLUTION APRES REDUCTION DES DIMENSIONS

II-1 Partition des matrices

Les paragraphes I-2 et I-3 conduisent à des systèmes linéaires de dimension respective $N_b + N_s - 1$ et $N_b + N_s + N_y - 2$.

Considérons de façon générale le système (o) $E X = Y$, de dimension N .

Décomposons N en $N_1 + N_2$ et partitionnons ainsi les matrices E , X et Y pour aboutir au système :

$$I \quad \left(\begin{array}{c|c} E_{11} & E_{12} \\ \hline E_{21} & E_{22} \end{array} \right) \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}$$

X_1 et Y_1 sont des vecteurs de dimension N_1 , X_2 et Y_2 sont de dimension N_2 . E_{11} est une matrice carrée de dimension N_1 , E_{22} est de dimension N_2 . E_{12} possède N_1 lignes et N_2 colonnes, ces dimensions étant échangées pour E_{21} . Si, comme ce sera le cas dans les problèmes traités, E est symétrique, E_{11} et E_{22} le sont aussi, tandis que $E_{21} = {}^t E_{12}$, la présence du (t) en position supérieure haute notant la transposition.

II-2 - Résolution partielle et substitution

Le système $E X = Y$ correspond à l'ensemble de sous-systèmes.

$$I(a) : \boxed{E_{11} X_1 + E_{12} X_2 = Y_1}$$

et

$$I(b) : \boxed{E_{21} X_1 + E_{22} X_2 = Y_2}$$

Le premier système conduit à :

$$E_{11} X_1 = Y_1 - E_{12} X_2$$

$$\text{d'où (II)} \quad \boxed{X_1 = E_{11}^{-1} (Y_1 - E_{12} X_2)}$$

E_{11}^{-1} étant la matrice inverse de E_{11}

Une substitution dans le sous système $I(b)$ conduit à :

$$E_{21} E_{11}^{-1} (Y_1 - E_{12} X_2) + E_{22} X_2 = Y_2$$

ou

$$(E_{22} - E_{21} E_{11}^{-1} E_{12}) X_2 = Y_2 - E_{21} E_{11}^{-1} Y_1$$

En posant $H = E_{22} - E_{21} E_{11}^{-1} E_{12}$

et

$$Z_2 = Y_2 - E_{21} E_{11}^{-1} Y_1$$

Cela conduit à

$$(III) \quad \boxed{H X_2 = Z_2}$$

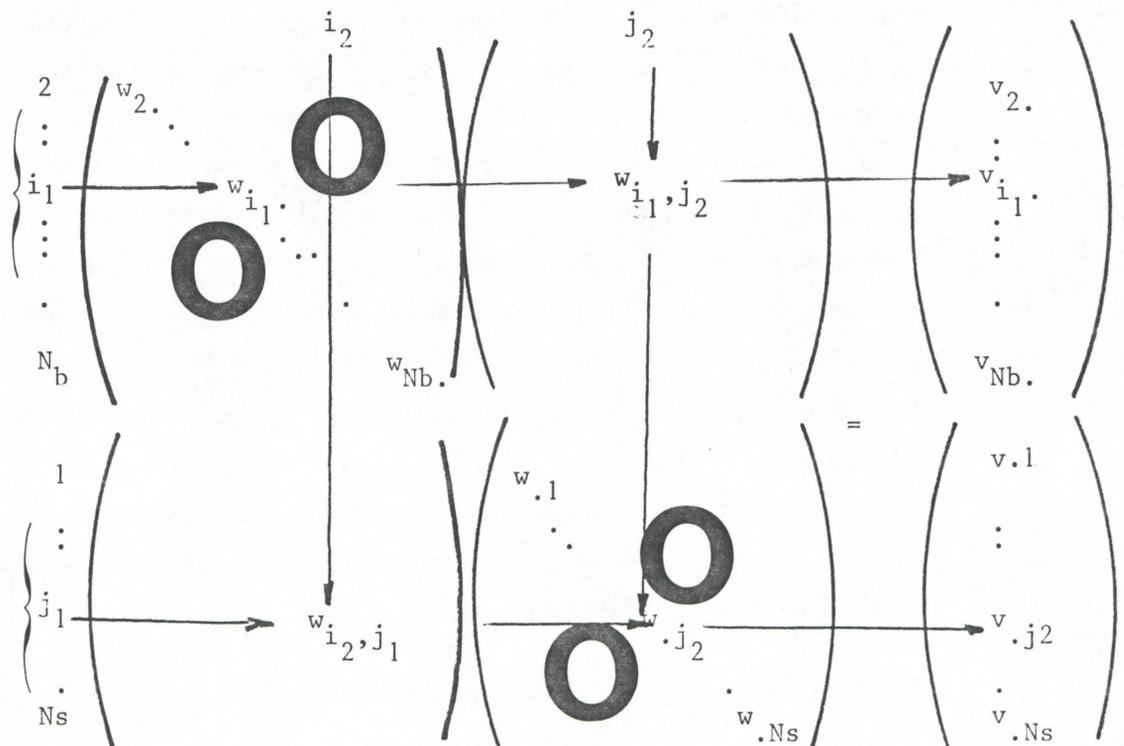
Il convient donc, pour suivre cette procédure, d'inverser la matrice E_{11} , puis de résoudre le système III. Cela donnera X_2 . On en déduira X_1 par l'équation II. Le système III est de plus faible dimension que le système I. En revanche, il faut de façon préliminaire inverser la matrice E_{11} . La procédure suggérée sera donc particulièrement pertinente quand cette inversion est triviale. C'est bien le cas pour les problèmes traités ici, puisque l'on fera en sorte que E_{11} soit une matrice diagonale.

II-3 - Application aux problèmes du paragraphe I

A) Modèle à deux facteurs

Le système de dimension $N_b + N_s - 1$, donné par les équations (3) et (5), peut s'écrire sous forme d'un système partitionné avec $N_1 = N_b - 1$ et $N_2 = N_s$.

Cette partition peut être résumée par le schéma suivant :



Le système II se compose ainsi des équations :

$$(8) \quad p_i = (v_{i,j} - w_{i,j} a_j) / w_i \quad i = 2, \dots, Nb$$

Le système III devient quant à lui :

$$(9) \quad \sum_{j_2=1}^{Ns} (\delta_{j_1, j_2} w_{j_1} - \sum_{i=2}^{Nb} \frac{w_{i, j_1} w_{i, j_2}}{w_i}) a_{j_2} = v_{j_1} - \sum_{i=2}^{Nb} \frac{w_{i, j_1}}{w_i}$$

pour $j_1 = 1, \dots, Ns$

Le terme générique de la matrice H du paragraphe II-2 est ainsi

$$h_{j_1, j_2} = \delta_{j_1, j_2} w_{j_1} - \sum_{i=2}^{Nb} \frac{w_{i, j_1} w_{i, j_2}}{w_i}$$

$$\text{tandis que } z_{j_1} = v_{j_1} - \sum_{i=2}^{Nb} \frac{w_{i, j_1} v_i}{w_i}$$

Les gains de temps permis par cette procédure sont considérables. Si l'on traite d'un exemple avec 400 strates et 100 bateaux, la réduction de la dimension du système par rapport à celui traité par le programme de gales (in ABRAMSON, 1971) est de $125 = (5)^3$. Si l'on traite d'un exemple où le nombre de strates excède celui du nombre de navires, il convient d'échanger navires et strates dans les raisonnements et formules précédents. Dans toutes les situations, la dimension du système à résoudre sera divisée par un facteur supérieur ou égal à 2, le temps de calcul par un facteur supérieur ou égal à $8 = 2^3$.

En terme d'optimisation informatique, le problème de l'espace mémoire requis doit aussi être évoqué. La matrice H est ainsi symétrique, et seul un triangle supérieur ou inférieur sera conservé. On peut en outre éviter d'avoir à stocker en mémoire centrale le tableau E_{12} , ou de façon équivalente le tableau de $w_{i,j}$ ($i = 2, \dots, Nb$; $j = 1, \dots, Ns$)

Pour ce faire, on calculera préalablement les w_i par la formule :

$$w_i = \sum_{\ell} m_{\ell} \delta_{i, b(\ell)}$$

On remarquera ensuite, se référant à la formule (9), qu'une boucle sur les i (bateaux), amènera à calculer les $w_{i,j}$, puis à les effacer à chaque fois qu'un nouveau bateau est concerné. On incrémentera à chaque fois les sommes :

$$\frac{w_{i,j_1} w_{i,j_2}}{w_i} \text{ et } \frac{w_{i,j_1} v_i}{w_i}$$

Les performances du logiciel FORTRAN que nous avons écrit, baptisé PUIS2, ont été comparées avec celles de FPOW, sur un exemple lié à la pêche française de Mer Celtique. Il contient 6719 C.P.U.E. de baudroie effectuées par 295 navires dans 122 mois-secteurs. Les calculs ont été menés sur un ordinateur DPS7/80. Les ressources requises apparaissent dans le tableau ci-dessous.

	temps mn.	memoire centrale 1u. = 2084 octets	memoire auxiliaire 1u. = 8092 octets
PUIS2	2.43	107	0
FPOW	518.19	345	2800

B) Modèle à trois facteurs

La procédure reste la même. On réaménagera le système de départ pour faire apparaître d'abord les paramètres correspondants à la catégorie la plus nombreuse. Selon les cas, ce sera N_{b-1} , N_{s-1} , ou de façon moins vraisemblable N_y . La première option mène à un système de dimension $N_{s-1} + N_y$, la seconde à $N_{b-1} + N_y$. Nous ignorons si une seconde procédure de substitution serait utile. Comme N_y est généralement faible, les réductions permises par la simple substitution seront substantielles. Pour traiter le cas des p.u.e. de navires de recherche, ceux-ci étant généralement peu nombreux, les séries historiques courtes, les calculs seront généralement rapides, même si le nombre de strates est élevé.

Dans le cas où $N_b - 1 \geq N_s - 1$ et $N_s - 1 \geq N_y$, les formules équivalentes aux formules (8) seront.

$$(10) \quad P_i = (v_{i..} - \sum_{j=2}^{N_s} (w_{i,j} q_j - \sum_{k=1}^{N_y} w_{i,k} d_k)) / w_{i..} \quad (i = 2, \dots, N_b)$$

La formule (9) sera remplacée par des formules de deux types :

$$(11) \quad \sum_{j_2} \left(\delta_{j_1, j_2} w_{\cdot j_1 \cdot} - \sum_i \frac{w_{i, j_2} \cdot w_{i, j_1} \cdot}{w_{i \cdot \cdot}} \right) q_{j_2} + \sum_k \left(w_{\cdot j_1, k} - \sum_i \frac{w_{i, j_1} \cdot w_{i, k}}{w_{i \cdot \cdot}} \right) d_k$$

$$= v_{\cdot j_1 \cdot} - \left(\sum_i (w_{i, j_1} \cdot v_{i \cdot \cdot} / w_{i \cdot \cdot}) \right)$$

pour $j_1 = 2, \dots, N_s$

et

$$(12) \quad \sum_j \left(w_{\cdot j, k_1} - \sum_i \frac{w_{i, k_1} \cdot w_{i, j} \cdot}{w_{i \cdot \cdot}} \right) q_j + \sum_{k_2} \left(\delta_{k_1, k_2} w_{\cdot \cdot k_1} - \sum_i \frac{w_{i, k_1} \cdot w_{i, k_2}}{w_{i \cdot \cdot}} \right) d_{k_2}$$

$$= v_{\cdot \cdot k_1} - \left(\sum_i (w_{i, k_1} \cdot v_{i \cdot \cdot} / w_{i \cdot \cdot}) \right)$$

pour $k_1 = 1, \dots, N_y$

III - INFERENCE STATISTIQUES

III-1 - Généralités

La résolution du système III peut se faire sans inverser la matrice H. Cette inversion sera en revanche nécessaire pour effectuer des inférences. Si Czz est la matrice de variance - covariance du vecteur Z, la matrice de variance - covariance de X₂ sera alors égale à H⁻¹ Czz H⁻¹. Pour obtenir la matrice de variance - covariance de X₁, et la matrice des covariances de X₁ et X₂ les formules seront plus complexes.

Pour résoudre ce problème on peut réaménager le système des équations II et III.

Le système III peut en effet s'écrire

$$X_2 = H^{-1} (Y_2 - E_{21} E_{11}^{-1} Y_1)$$

Soit

$$(IV) \quad X_2 = (- H^{-1} E_{21} E_{11}^{-1}) Y_1 + H^{-1} Y_2$$

Le système II devient quant à lui :

$$X_1 = E_{11}^{-1} Y_1 - E_{11}^{-1} E_{12} (- H^{-1} E_{21} E_{11}^{-1}) Y_1 - E_{11}^{-1} E_{12} H^{-1} Y_2$$

ou

$$(V) \quad X_1 = (E_{11}^{-1} + E_{11}^{-1} E_{12} H^{-1} E_{21} E_{11}^{-1}) Y_1 - (E_{11}^{-1} E_{12} H^{-1}) Y_2$$

Globalement, on en revient à

$X = G Y$ avec une partition de la matrice G comme suit

$$N_1 \quad G_{11} \quad G_{12}$$

$$N_2 \quad G_{21} \quad G_{22}$$

$$N_1 \quad N_2$$

$$\text{où } G_{11} = (E_{11}^{-1} + E_{11}^{-1} E_{12} H^{-1} E_{21} E_{11}^{-1})$$

$$G_{12} = -E_{11}^{-1} E_{12} H^{-1} = {}^t G_{21}$$

$$G_{22} = H^{-1}$$

En fait la matrice G qui vient d'être définie est l'inverse de la matrice E du système (o). On est donc ramené à la procédure classique. Si C_{xx} est la matrice de variance - covariance de X , C_{yy} celle de Y :

$$(VI) \quad C_{xx} = G C_{yy} G = E^{-1} C_{yy} E^{-1}$$

La procédure de partition-substitution permet d'opérer un retour au cas général.

Disposant de Cxx, il est possible d'en déduire les variances des p_i, a_j ou p_i, d_j et q_k selon que l'on traite du problème à un ou deux facteurs. Plus généralement la variance de toute combinaison linéaire pourra être obtenue. Ce fait est utile si l'on souhaite revenir sur la convention ayant levé l'indétermination sur les p_i . On a en effet choisi $p_1 = 0$, mais toute autre convention du type

$$\text{Nb} \quad \sum_{i=1} \lambda_i p_i = 0.$$

avec

$$\text{Nb} \quad \sum_{i=1} \lambda_i = 1.$$

est concevable. On passera en fait de l'estimation p_i obtenue avec $p_1 = 0$ aux p_i^* correspondant au cas où $\sum_i \lambda_i p_i^* = 0$, par les relations :

$$p_i^* = p_i + \mu \quad \text{avec} \quad \mu = - \sum_i \lambda_i p_i$$

La variance de $p_{i_0}^*$ est alors égale à :

$$\sum_{i_1, i_2} r_{i_0, i_1} r_{i_0, i_2} c_{i_1, i_2}$$

$$\text{si } r_{i_0, i_1} = \delta_{i_0, i_1} - \lambda_{i_1} \quad \text{et} \quad r_{i_0, i_2} = \delta_{i_0, i_2} - \lambda_{i_2}, \quad c_{i_1, i_2}$$

désignant la covariance de p_{i_1} et p_{i_2} , donné par Cxx. La même procédure s'appliquerait aux calculs de covariances comme à la résolution du système à trois facteurs.

III-2 Particularité du modèle stochastique simple

Par modèle simple, on entend un modèle où les différents résidus centrés, indépendants, et d'une variance égale à $2/m_\ell$. La pondération par les m_ℓ a alors pour but de rééquilibrer les variances. Dans ce cadre on vérifie que la ^{co}variance de v_{i_1} et v_{i_2} est égale à :

$$\sum_{\ell} \delta_{i_1, b(\ell)} \delta_{i_2, b(\ell)} \frac{m_{\ell}^2}{m_{\ell}} = \delta_{i_1, i_2} \left(\sum_{\ell} \delta_{i_1, b(\ell)} m_{\ell} \sigma^2 \right)$$

soit encore $\delta_{i_1, i_2} w_{i_1} \sigma^2$
 La covariance de $v_{.j_1}$ et $v_{.j_2}$ est de même égale à :

$$\delta_{j_1, j_2} w_{j_2} \sigma^2$$

Enfin la covariance de $v_{i.}$ et $v_{.j}$ est égale à :

$$\sum_{\ell} \delta_{i, b(\ell)} \delta_{j, s(\ell)} m_{\ell} \frac{\sigma^2}{m_{\ell}} = \sigma^2 w_{i, j}$$

Il apparaît ainsi que $C_{yy} = E \sigma^2$

Le système (VI) se simplifie alors en

$$C_{xx} = \sigma^2 E^{-1} E E^{-1} = \sigma^2 E^{-1}$$

(VII) $\boxed{C^{xx} = \sigma^2 E^{-1}}$

Le calcul de $G = E^{-1}$ donne ainsi : la matrice de variance - covariance recherchée.

On pourrait d'ailleurs vérifier que le système VII ne correspond qu'à un cas particulier de relations liant variance-covariance des estimations et Hessian de la fonction de Log-vraisemblance.

Si l'on s'en tient aux variances et covariances des puissances de pêche, la matrice recherchée sera égale à :

$$(VIII) \quad G_{11} = \begin{pmatrix} E_{11}^{-1} & & & & \\ & E_{11}^{-1} & & & \\ & & E_{12} & & \\ & & & E_{21} & \\ & & & & E_{11}^{-1} \end{pmatrix}$$

Reste que cet élégant résultat, exploité par le logiciel de Berude (in, Abramson, 1971) repose sur des hypothèses d'indépendance des résidus et d'un retour possible à l'homoscédasticité. Les logiciels conçus s'attachent ainsi à étudier de près, la structure des résidus, ainsi qu'il est de rigueur pour toute régression bien conçue. Aux histogrammes habituels s'ajoutant ainsi l'examen de diverses sources possibles d'hétéroscédasticité, (variances liées aux bateaux, et/ou aux strates), sous le système à trois facteurs les liaisons entre résidus seront étudiées au travers de corrélations spatiales et temporelles. Le premier intérêt d'un logiciel spécifique, par opposition à l'utilisation d'un logiciel standard d'ajustement du modèle linéaire correspond d'ailleurs à la possibilité d'effectuer un examen des corrélations entre résidus, adaptés à la nature du problème.

III-3 Considérations pratiques

Le système VIII tel qu'il est écrit laisse peu d'espoir d'éviter le stockage du tableau E_{12} , correspondant aux $w_{i,j} = \sum_l \delta_{i,b(l)} \delta_{j,s(l)} m_l$. C'est ce que l'on avait cherché à éviter précédemment. Le souci d'effectuer des inférences amène donc à accroître le temps de calcul requis, puisqu'il faudra inverser une matrice, et la place mémoire nécessaire.

Si la possibilité d'effectuer des inférences est à payer d'un prix, qui peut être élevé, le surcroît d'informations doit être apprécié. On sait que pour obtenir une estimation non biaisée des puissances et des abondances, et non de leurs logarithmes, une correction est nécessaire.

Elle n'est possible que si la variance du logarithme est connue ou estimée. L'importance de cette correction ne doit pas être surestimée en pratique (voir annexe A). Il est aussi souhaitable de pouvoir apprécier la fiabilité des estimations, d'effectuer des inférences. Il est ainsi souhaitable de disposer d'un logiciel calculant les matrices de variance - covariance. Il paraît toutefois sage de disposer en parallèle d'une version simplifiée, n'opérant pas les inférences. Elle est utile pour effectuer des passages préliminaires, destinés à épurer les données de bases. Par ailleurs, les hypothèses requises pour effectuer les inférences selon les formules présentées au paragraphe III-2 peuvent être mises en doute. Des techniques de rééchantillonnage (Jackknife, Bootstrae) peuvent être utilisées. Disposer d'un logiciel rapide est alors particulièrement précieux.

L'expérience nous a aussi démontré que lorsque l'on opère sur des bateaux individuels, pour effectuer ensuite une régression des puissances selon les caractéristiques connues des navires, l'erreur d'estimation des puissances était une composante faible des variances résiduelles, sauf exception. Ces exceptions correspondent aux navires ayant fourni peu de données. C'est pourquoi, le logiciel simplifié fournit les w_i , qui offrent une première base pour effectuer une régression pondérée.

A terme, un logiciel permettant ou non des inférences sera disponible. Actuellement, seule la version simplifiée, sans inférence, est opérationnelle.

CONCLUSION

La disponibilité de données fines, issues notamment de livres de bords, laisse envisager le calcul systématique des puissances de pêche individuelles. Faire appel à des algorithmes adaptés à la nature du problème traité permet une réduction considérable des temps de calculs, et accessoirement des espaces requis en mémoire centrale. Les gains de temps et d'espace sont particulièrement importants si l'on opère une estimation de paramètres sans opérer d'inférence. Cette résolution simplifiée est suffisante dans beaucoup d'applications. Comme il est également indispensable pour d'autres applications de pouvoir effectuer des inférences, il est souhaitable de disposer de logiciels couvrant les deux approches. Le logiciel du à BERUDE (in ABRAMSON, 1971, SIMS, 1986) remarquable à l'origine du point de vue informatique doit lui être considéré comme désuet. D'une part il ne permet pas de choisir entre ajustement simple ou, après inversion, avec inférence, d'autre part il n'exploite pas les particularités du système linéaire créé.

REFERENCES CITEES

Abramson, N.J. (1971) : Computer programmes for Fish Stock Assessment.
F.A.O. Fish. Biol. techn. paper. 101 : 4 pp + unpag.

Anon, (1984) : Reports of the ad hoc working group on the use of Effort
Data in Assessments and of the working group on methods of Fish Stock
Assessments ICES, Coop, Res. Rep. No. 129.

Charuau et collaborateurs (1985) : Etude d'une gestion optimale des
pêcheries de langoustines et de poissons démersaux en Mer Celtique.
Rapport final de la deuxième phase
Direction Générale de la C.E.E. pour les pêcheries Bruxelles.

Charuau et collaborateurs (1986) : Etude d'une gestion optimale des
pêcheries de langoustines et de poissons démersaux en Mer Celtique.
Rapport final de la deuxième phase.
Rapport final de la troisième phase Tome I
Direction générale de la C.E.E. pour les pêcheries Bruxelles.

Gavaris S. (1980) : Use of multiplicative model to estimate catch rate
and effort from commercial data. Cdn. J. Fish. Aquat. Sci. 37 :
2272-2275

Gulland, J.A. (1956) : On the fishing effort in English Demersal
Fisheries
Fischery Invest., Lond., Ser 2, SO : (5), 41 pp.

Laurec A et Fonteneau A. (1979) : Estimation de l'abondance d'une classe
d'âge : utilisation des c.p.u.e. de plusieurs engins, en différentes
zones et saisons. Recueil de Doc. scientifique. CICTA S.C.R.S./78/86
CICTA Madrid

Laurent A.G. (1963) : Lognormal distribution and the translation method : description and estimation problems. J. Ameri. Statist. Assoc., 58 : 231-5

Morrison D.F. (1976) : Multivariate statistical methods. Mc Graw Hill Series in probability and statistics. Mc Graw-Hill New York

Robson, D.S. (1966) : Estimation of the relative fishing power of individual ships Res. Bull. Int. Comm. N.W. Atlant. Fish. (3) : 5-14

Sims S.E. (1985) Selected computer programmes in FORTRAN for fish stock assessment. F.A.O. Fish tech. pap. N° 259 183 pp.

REMARQUE SUR L'IMPORTANCE DU FACTEUR DE CORRECTION DU BIAS

On sait que si l'on tente d'estimer \hat{y} par $e^{\hat{x}}$, \hat{x} étant un estimateur non biaisé d'espérance μ et de variance σ^2 , l'espérance de $e^{\hat{x}}$ n'est pas égale à e^{μ} mais à $e^{(\mu + \sigma^2/2)}$, si \hat{x} est normalement distribué. Il est donc nécessaire en théorie d'introduire un facteur de correction $e^{-\sigma^2/2}$. Lorsque l'on dispose d'une estimation s^2 de σ^2 , il convient d'en déduire une estimation de $e^{-\sigma^2/2}$. A l'expérience les coefficients de correction estimés sont rarement cruciaux. Lorsque s^2 est faible, ils sont vite négligeables. Ce n'est plus vrai quand s^2 croît, mais la variance croît encore plus vite que le biais. Si l'on considère l'erreur quadratique moyenne, somme de la variance et du carré du biais, le premier terme est largement prépondérant.

Cela peut se comprendre si l'on considère la variance de $\hat{y} = e^{\hat{x}}$. Elle est égale à $e^{2\mu + \sigma^2} (e^{\sigma^2} - 1)$. Dans le même temps, le carré du biais est égale à $e^{\mu} (e^{\sigma^2/2} - 1)^2$.

Le rapport carré du biais/variance est ainsi égal à

$$\frac{e^{\sigma^2/2} (e^{\sigma^2/2} - 1)^2}{e^{\sigma^2} (e^{\sigma^2} - 1)} = \frac{e^{\sigma^2/2} (e^{\sigma^2/2} - 1)^2}{e^{\sigma^2} (e^{\sigma^2/2} - 1) (e^{\sigma^2/2} + 1)}$$

Soit donc
$$\frac{e^{\sigma^2/2} - 1}{e^{\sigma^2} (e^{\sigma^2/2} + 1)}$$

Dans le même temps, le carré du coefficient de variation de $e^{\hat{x}}$ est égal à $e^{\sigma^2} - 1$.

Il apparaît ainsi que soit σ^2 est fort et le coefficient de variation est grand et toute estimation très imprécise, soit σ^2 est faible, et les problèmes de biais négligeables.

Lorsque, faute des σ^2 , on dispose d'estimations $\hat{\sigma}^2$, les formules se compliquent, mais l'expérience nous a confirmé que pour l'essentiel le diagnostic ci-dessus restait fondé. Il convient donc de ne pas dramatiser l'impact de la correction de LAURENT (1963)

Ce fait est illustré par les figures 1 et 2. Elles correspondent à l'exemple évoqué dans le paragraphe II-3 A. L'utilisation du programme FPOW a permis le calcul du biais relatif tel que calculé par la méthode de LAURENT (1963), et des écarts-types. Pour l'ensemble des navires, les biais relatifs (figure 1), et écarts-types (figure 2) ont été reportés sur un graphique où apparaît en abscisse le nombre d'observations données pour un navire. Du fait de l'équipondération ce nombre n'est autre que w_i . On notera ainsi l'excellente relation des biais et écarts-types avec w_i , et la confirmation du fait que lorsque la dispersion est faible, le biais est négligeable.

UNE METHODE NON LINEAIRE

Il se peut que les données de base ne correspondent pas directement à des captures par strate, mais à des captures par marée, cumulées éventuellement sur plusieurs strates, assorties d'évaluation du temps de pêche dévolu aux différentes strates. Ce problème va être traité dans le cas du modèle à deux facteurs, mais pourrait également l'être dans un contexte à trois facteurs.

1 - Notations

On notera alors C_ℓ la ℓ^e capture (pour une espèce et/ou une catégorie commerciale donnée). Le temps passé dans la strate j est $T_{j,\ell}$. Le bateau correspondant est $b(\ell)$. Sa puissance de pêche logarithmique est $p_{b(\ell)}$ correspondant à la puissance non transformée $P_{b(\ell)}$. La strate $s(\ell)$ correspond à une abondance logarithmique $a_{s(\ell)}$ soit sans transformation $A_{s(\ell)}$.

Le modèle de base s'écrira alors :

$$(1) \quad C_\ell = P_{b(\ell)} \left(\sum_j A_j T_{j,\ell} \right)$$

ou après transformation logarithmique :

$$(2) \quad c_\ell = p_{b(\ell)} + \text{Log} \left(\sum_j A_j T_{j,\ell} \right)$$

Comme précédemment les inconnues sont les p_i et les a_j . La même indétermination jouant, on posera : $p_1 = 0$.

Le modèle (1) ne peut être linéarisé. Cela enlève un premier attrait à une transformation logarithmique. De plus, les C_ℓ correspondants à un cumul sur plusieurs journées, il n'est pas déraisonnable de considérer que les résidus puissent être gaussiens. L'ajustement par les moindres carrés, éventuellement pondérés, du modèle non transformé (équation (1)) trouve alors un surcroît d'intérêt.

Les deux points de vue, correspondant respectivement à l'ajustement par les moindres carrés des équations (1) et (2), vont être développés. Seule l'expérience pratique permettra à l'avenir de trancher entre les deux voies.

II - AJUSTEMENT DU MODELE NON TRANSFORME

Dans ce cas on minimisera

$$\phi = \sum_{\ell} m_{\ell} (C_{\ell} - P_{b(\ell)} (\sum_j A_j T_{j,\ell}))^2$$

Annuler la dérivée par rapport à P_i mène à :

$$(3) \quad P_i = \frac{\sum_{\ell} \beta_{i,\ell} (\sum_j A_j T_{j,\ell})}{\sum_{\ell} \beta_{i,\ell} (\sum_j A_j T_{j,\ell})^2}$$

Si l'on pose $\beta_{i,\ell} = m_{\ell} \delta_{i,b(\ell)}$

Annuler la dérivée par rapport à A_j mène à :

$$(4) \quad \sum_j \sum_{\ell} m_{\ell} T_{j,\ell} T_{j_0,\ell} P_{b(\ell)}^2 = \sum_{\ell} m_{\ell} P_{b(\ell)} T_{j_0,\ell} C_{\ell}$$

Par substitution, on peut éliminer les inconnues correspondantes aux P_i dans l'équation (4). Le système ainsi créé est cependant non linéaire.

III - AJUSTEMENT APRES TRANSFORMATION LOGARITHMIQUE

Gardant l'éventualité d'une pondération, on sera conduit à minimiser :

$$\phi = \sum_{\ell} m_{\ell} (c_{\ell} - (p_{b(\ell)} + \text{Log}(\sum_j A_j T_{j,\ell})))^2$$

Annuler $\frac{d\phi}{dp_i}$ mène à l'équation :

$$\sum_{\ell} \beta_{i,\ell} (c_{\ell} - p_i - \text{Log}(\sum_j A_j T_{j,\ell})) = 0.$$

soit :

$$(6) \quad p_i = \frac{1}{(\sum_{\ell} \beta_{i,\ell})} \sum_{\ell} \beta_{i,\ell} (c_{\ell} - \text{Log}(\sum_j A_j T_{j,\ell}))$$

Annuler $\frac{d\phi}{dA_{j_0}}$ conduit de même à :

$$(7) \quad \sum_{\ell} \beta_{j_0,\ell} (p_{b(\ell)} + \frac{T_{j_0,\ell}}{\sum_j A_j T_{j,\ell}}) = \sum_{\ell} \beta_{j,\ell} c_{\ell}$$

Là encore l'équation (6) peut servir de base à une substitution, pour créer un système ne contenant que les A_j comme inconnues. Ce système ne sera pas linéaire.

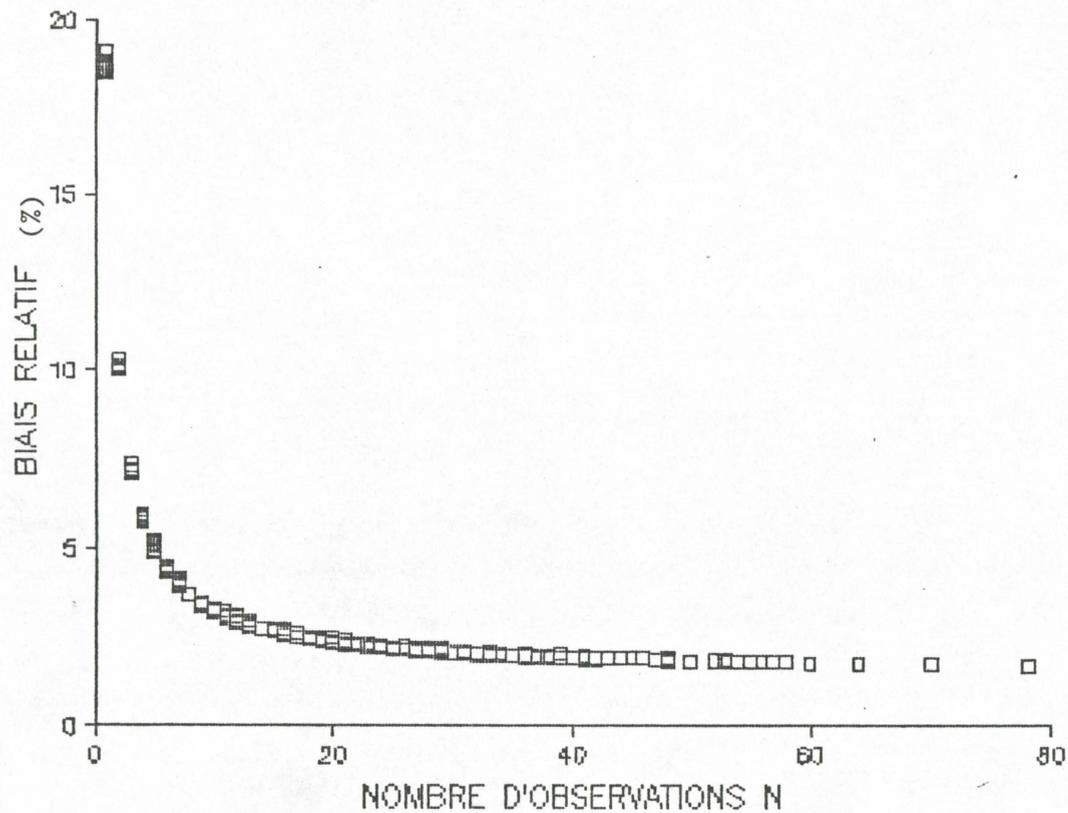


Figure 1: Relation entre le biais relatif (%) et le nombre d'observations par bateau

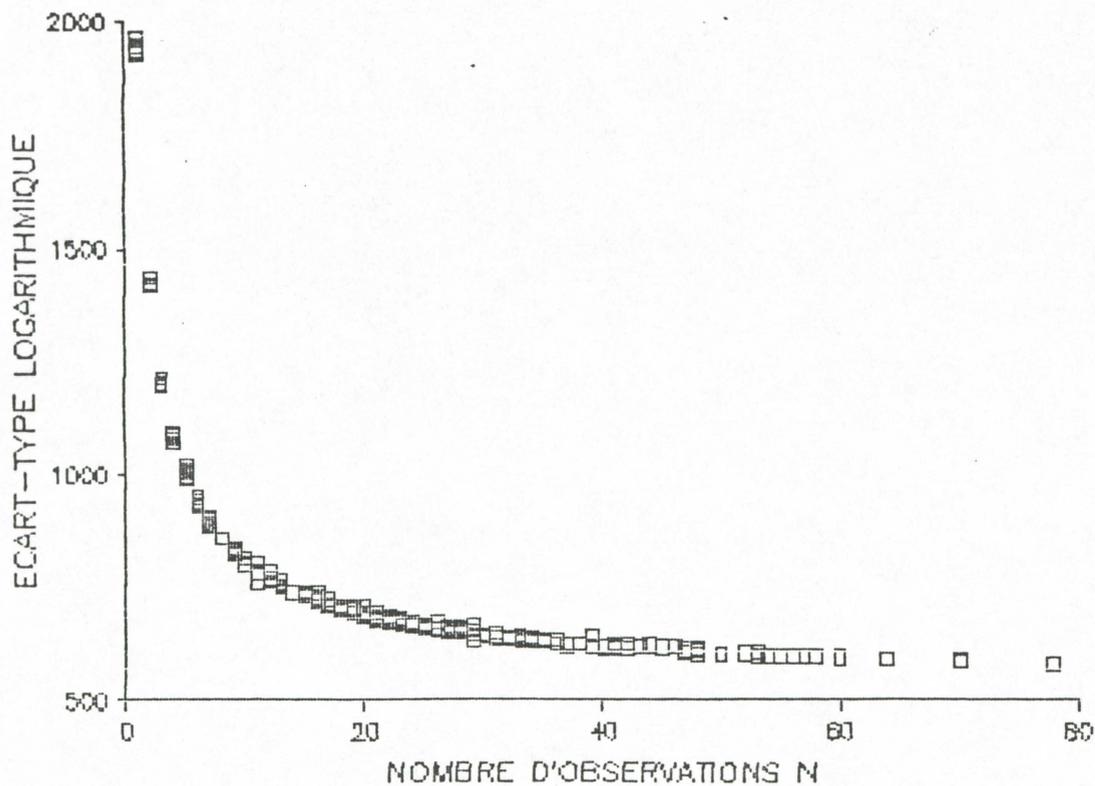


Figure 2 : Relation entre l'écart-type des puissances de pêche logarithmiques et nombre d'observations.