

ESTIMATION DES PARAMETRES DE CROISSANCE
APPORT DES TECHNIQUES DE REECHANTILLONNAGE
(Bootstrap - Jackknife)
ET DES STATISTIQUES ROBUSTES (Médianes et modes)

par

P. GROS*, A. LAUREC** et H. DUPOUY***

Résumé

La première partie est dévolue à une synthèse des problèmes statistiques posés par l'ajustement de la courbe de Von Bertalanffy. Un premier volet de cette synthèse est consacré à la régression classique, sous hypothèse de normalité et d'indépendance des résidus. Le deuxième volet est consacré aux techniques de rééchantillonnage, les points et formules essentiels du Jackknife et du Bootstrap étant présentés. La possibilité d'utiliser comme valeur centrale des longueurs pour un âge donné, les modes ou médianes, outre les moyennes arithmétiques, est discutée, en liaison avec les techniques de rééchantillonnage, les mieux à même de fournir dans un tel contexte des éléments d'inférences.

Un logiciel a été écrit, effectuant des ajustements par les moindres carrés (éventuellement pondérés) des valeurs centrales évoquées, et procédant aux rééchantillonnages classiques. Ce logiciel couvre, en outre, les possibilités de rétrocalcul. Un exemple est finalement traité.

Summary

This paper is devoted to statistical problems associated to the estimation of the parameters of a Von Bertalanffy growth curve. A first part corresponds to theoretical considerations, first within the classical regression point of view, corresponding to residuals normaly and independently distributed, then to resampling techniques (Jackknife/Bootstrap). The possibility of using not only the arithmetic mean of observed lengths for a given age, but also medians and modes is discussed.

A software has been written in Fortran, which makes it possible to fit by (Weighted) least squares either mean, median or modal length at age, and provides estimates of variances/covariances using the classical approximation within regression theory or resampling techniques. An example, corresponding real data is finally.

* IFREMER - B.P. 337 - BREST CEDEX

** IFREMER - B.P. 1049 - 44037 NANTES CEDEX 01

*** IFREMER - 8, rue François Toullec - 56100 LORIENT

INTRODUCTION

Le regain d'intérêt pour les méthodes reposant sur les distributions de longueur (JONES, 1979, 1981) a accru l'intérêt pour une estimation fiable des paramètres de l'équation de Von Bertalanffy.

Le développement des études de sensibilité, et notamment des méthodes delta (LAUREC et MESNIL, 1985 et LAUREC, 1986) a également mis l'accent sur la nécessité de disposer d'éléments d'inférences statistiques sur les estimations obtenues. Dans le même temps, les techniques de rééchantillonnage (Jackknife, Bootstrap) se sont développées, dotant les statisticiens d'outils nouveaux.

Le présent document se situe à la rencontre des besoins et des possibilités mentionnés. Il présente ainsi un rappel sur l'ajustement du modèle de Von Bertalanffy vu comme un problème de régression classique. Il aborde ensuite l'application des techniques de rééchantillonnage, en s'efforçant d'offrir aux halieutes un sommaire des points clés d'une approche statistique dont ils ne sont pas toujours familiers. Au passage, la possibilité d'utiliser comme valeur centrale des longueurs pour un âge donné, non la moyenne arithmétique, mais d'autres statistiques telles que la médiane ou le mode est évoqué. Le logiciel qui a été développé est ensuite présenté, avant qu'un exemple numérique soit traité.

1. Nature du modèle et notations

Le modèle postulé pour décrire la croissance linéaire est le classique modèle de Von Bertalanffy, selon lequel la longueur observée l_{ij} du j -ème poisson ($j = 1, 2, \dots, n_i$) appartenant au groupe d'âge i ($i = 1, 2, \dots, I$) s'exprime :

$$l_{ij} = \theta_1 (1 - \exp(-\theta_2 (t_i - \theta_3))) \quad (1)$$

où θ_1 , θ_2 et θ_3 sont habituellement notés L_∞ , K et t_0 respectivement. En notant θ le vecteur colonne des paramètres du modèle, ce dernier peut s'écrire sous forme condensée :

$$l_{ij} = f(t_i, \theta)$$

où t_i désigne l'âge du groupe i , considéré, en général, comme connu sans erreur.

Les développements qui vont suivre s'inscrivent dans une problématique générale, celle de l'ajustement de modèle. Il est considéré que le choix de la formulation analytique (1) n'est pas arbitraire, et qu'il est légitime de s'interroger sur la "vraie valeur" que possèdent les paramètres θ . Cela légitime la recherche d'une procédure qui non seulement résume au mieux les données, mais qui de surcroît fournisse des estimations les plus proches possible des vraies valeurs (inconnues) des paramètres, tout en demeurant stables d'un échantillon à l'autre. Cela nécessite d'introduire des notations ne prêtant pas à ambiguïté ; ainsi, seront adoptées les conventions d'écriture :

$\tilde{\theta}$: vrai vecteur des paramètres, inconnu et non aléatoire ;

$\hat{\theta}$: estimateur de $\tilde{\theta}$; cette notation désignera également les estimations considérant qu'il n'y a pas risque de confusion ;

θ : vecteur des valeurs courantes.

Dans ce contexte, la longueur observée l_{1j} (supposée exempte d'erreur de mesure) est considérée comme une réalisation particulière de la variable aléatoire L_{1j} (longueur à l'âge t_{1j}), dont l'espérance vaut précisément :

$$E(L_{1j}) = f(t_{1j}, \tilde{\theta}) \quad (2)$$

égalité qui suppose implicitement que le modèle f est convenablement choisi, i.e. qu'il est exact. La formulation théorique de l_{1j} s'écrit quant à elle :

$$l_{1j} = f(t_{1j}, \tilde{\theta}) + \epsilon_{1j} \quad (3)$$

2. Hypothèses usuelles sur la loi des erreurs et critères d'ajustement

Hypothèse a - La compatibilité des équations (2) et (3) implique que les résidus ϵ_{1j} soient d'espérance nulle, i.e. : $E(\epsilon_{1j}) = 0$, $i = 1, \dots, I$

Cette hypothèse ne suffit pas pour décider du choix d'un critère d'ajustement ; la dispersion des erreurs doit également être précisée. Le problème est alors de définir un corps d'hypothèses tel que les variances soient estimables à partir de l'unique échantillon (l_{1j} ; $i=1, \dots, I$; $j=1, \dots, n_1$). Ainsi :

Hypothèse b - Soit V la matrice de covariance des résidus ϵ_{1j} , ϵ_{2j} , ... ϵ_{Ij} . La deuxième hypothèse énonce que la répétition des mesures sur un autre échantillon (l_{ab} ; $a=1, \dots, I$; $b=1, \dots, n_a$) engendre des résidus dont la matrice de covariance est elle aussi V .

Hypothèse c - La troisième hypothèse habituellement formulée est celle d'absence de corrélation entre les résidus :

$$E(\epsilon_{1j}, \epsilon_{h1}) = 0 \text{ et } E(\epsilon_{1j}, \epsilon_{1j'}) = \sigma_1^2 \\ h \neq i \\ j \neq j'$$

Autrement dit, la matrice V est diagonale.

Dans le contexte linéaire, la vérification de a), b) et c) garantit que l'optimisation du critère des moindres carrés pondérés est la "meilleure" méthode d'identification des paramètres du modèle, en ce sens qu'elle fournit l'estimateur non biaisé de variance minimale. L'optimalité de l'estimateur des moindres carrés n'est cependant plus attestée lorsque le modèle est non linéaire. Aussi, l'estimateur du maximum de vraisemblance lui est-il généralement préférée du fait de ses qualités asymptotiques : convergence en probabilité, efficacité, et distribution d'échantillonnage normale (e.g., KENDALL et STUART, 1979).

Il apparaît donc nécessaire d'adjoindre aux hypothèses a), b) et c) une spécification de la loi des ϵ_i pour les deux raisons suivantes :

- pouvoir définir la vraisemblance, qui est entre autres déterminée par la densité jointe des résidus,
- donner le moyen de tenter des inférences statistiques dans un cadre paramétrique.

Hypothèse d - L'hypothèse de normalité de la loi des erreurs est alors généralement retenue et complétée de la sorte les conditions a), b), et c) deviennent :

$$\epsilon (\hat{\theta}) \approx N_I (0, V) \quad (4)$$

ou $\epsilon (\hat{\theta})$ désigne le $I \times 1$ vecteur des résidus. L'absence de corrélation c) implique désormais l'indépendance. La relation (4) signifie ainsi que ϵ suit une loi normale centrée, de variance-covariance V .

L'hypothèse de multinormalité n'est pas seulement une commodité de calcul : il peut être démontré (cf. par exemple BARD, 1974) que c'est également le seul moyen de n'introduire que le minimum d'information "exogène" pour définir la densité d'une variable aléatoire dont seules l'espérance et la variance sont connues. De manière qualitative, il ne faut toutefois pas perdre de vue quelles en sont les implications : les réalisations de ϵ_i se distribuent symétriquement autour de $f(t_i, \theta)$, la probabilité pour qu'elles s'en écartent de plus de $3\sigma_i$ étant quasi-nulle. Concrètement, cela impose un examen attentif des écarts à l'ajustement.

L'hypothèse (4) étant admise, il peut alors être aisément démontré que maximiser la vraisemblance équivaut à minimiser la forme quadratique :

$$S(\theta) = \frac{1}{2} e^T(\theta) V^{-1} e(\theta) \quad (5)$$

où le symbole \cdot^T désigne la transposition et $e(\theta)$ le vecteur colonne de composantes :

$$e_i(\theta) = \bar{y}_i - f(t_i, \theta) \quad , \quad \bar{y}_i = \sum_j y_{ij} / n_i \quad (6)$$

Sachant que les éléments diagonaux w_i de la matrice V^{-1} valent :

$$w_i = n_i / \sigma_i^2$$

alors l'estimateur $\hat{\theta}_{MV}$ du maximum de vraisemblance défini par :

$$S(\hat{\theta}_{MV}) = \min_{\theta} \{S(\theta)\}$$

est bien évidemment identique à l'estimateur $\hat{\theta}_{MC}$ des moindres carrés pondérés :

$$S(\hat{\theta}_{MV}) = \min_{\theta} \left\{ \sum_{i=1}^I w_i e_i^2(\theta) \right\} = S(\hat{\theta}_{MC})$$

En règle générale, les variances σ_1^2 sont inconnues et devront être estimées pour calculer le critère $S(\theta)$: c'est donc l'estimation \hat{V} de V , d'éléments diagonaux \hat{v}_1 qui apparaîtra ultérieurement dans les équations. Il doit par ailleurs être souligné que se sont les longueurs moyennes par classe d'âge l_1 qui figurent dans l'expression de $S(\theta)$; ce choix est justifié par le fait que le théorème de la limite centrale garantit la tendance vers la normalité de la distribution des l_1 , propriété qui conduit à préférer à l'équation (3) le modèle :

$$\bar{l}_1 = f(k, \theta) + \epsilon_1 \quad (7)$$

où ϵ_1 désigne dans ce cas un résidu aléatoire de dispersion plus faible, car associé aux moyennes empiriques et non aux observations individuelles l_{1j} .

Hypothèse e - Si aux hypothèses précédemment retenues est ajoutée celle d'homoscédasticité (4) devient :

$$\epsilon(\theta) \approx N_{1k}(0, \sigma^2 \mathcal{I}) \quad (8)$$

la matrice \mathcal{I} étant la matrice identité. L'estimateur θ_{MV} est alors obtenu en minimisant la somme des carrés des écarts :

$$S(\theta_{MV}) = \min_{\theta} \{e'(\theta)e(\theta)\}$$

et il équivaut par conséquent à l'estimateur $\hat{\theta}_{MV}$ des moindres carrés non pondérés. Cependant, si le nombre n_1 d'observations n'est pas le même dans chacune des k classes d'âge, le critère s'exprime :

$$S(\hat{\theta}) = \sum_i n_{1i} e_{1i}^2(\theta) \quad (9)$$

3. Approximation analytique de la matrice de covariance des paramètres sous l'hypothèse de normalité

Soit L la matrice des données, i.e. la matrice (l_{1j}) des valeurs observées. Elle peut être considérée comme un ensemble de réalisations particulières des variables aléatoires \mathcal{L} , observations à partir desquelles est obtenue une estimation $\hat{\theta}$ de θ . Un échantillon différent $L + \delta L$ fournirait une autre estimation $\hat{\theta} + \delta \hat{\theta}$. Si les deux échantillons sont saisis selon le même protocole sur le même stock de poissons, il est souhaité que la norme du vecteur $\delta \hat{\theta}$ soit faible, autrement dit que l'estimateur $\hat{\theta}$ soit stable. La méthode habituellement employée pour jauger la plus ou moins forte sensibilité de $\hat{\theta}$ aux fluctuations d'échantillonnage de \mathcal{L} consiste à quantifier la dispersion de la distribution d'échantillonnage de $\hat{\theta}$, i.e. à calculer une estimation approchée $\hat{V}(\hat{\theta})$ de la matrice $V(\hat{\theta})$ de covariance des paramètres.

La démarche qui va suivre est empruntée à BARD (1974). La meilleure estimation $\hat{\theta}$ est celle qui définit un extremum du critère S , fonction des données L ; donc :

$$S(\hat{\theta}, L) / \partial \theta = 0$$

Si les données sont légèrement perturbées (e.g., la matrice des données devient $L + \delta L$), la nouvelle estimation $\hat{\theta} + \delta \hat{\theta}$ satisfait de même la condition nécessaire :

$$\partial S(\hat{\theta} + \delta \hat{\theta}, L + \delta L) / \partial \theta = 0$$

Le développement de TAYLOR du premier membre au voisinage de $(\hat{\theta}, L)$ conduit à l'approximation, en ne retenant que les termes d'ordre 1 :

$$\partial S(\hat{\theta} + \delta \hat{\theta}, L + \delta L) / \partial \theta \approx \partial S(\hat{\theta}, L) / \partial \theta + (\partial^2 S(\hat{\theta}, L) / \partial \theta^2) \delta \hat{\theta} + (\partial^2 S(\theta, L) / \partial \theta \partial L) \delta L$$

Soit encore, après élimination des termes égaux au vecteur nul :

$$(\partial^2 S(\hat{\theta}, L) / \partial \theta^2) \delta \hat{\theta} + (\partial^2 S(\theta, L) / \partial \theta \partial L) \delta L = 0$$

$$\text{donc : } \delta \hat{\theta} = -\hat{H}^{-1} (\partial^2 S(\hat{\theta}, L) / \partial \theta \partial L) \delta L$$

où \hat{H} désigne le Hessien du critère S évalué à $\theta = \hat{\theta}$.

Par définition : $V(\hat{\theta}) = E(\delta \hat{\theta} \delta \hat{\theta}^T)$.

On en déduit :

$$V(\hat{\theta}) = \hat{H}^{-1} (\partial^2 \hat{S} / \partial \theta \partial L) E(\delta L \delta L^T) (\partial^2 \hat{S} / \partial \theta \partial L)^T \hat{H}^{-1} \quad (10)$$

où \hat{S} est l'abréviation de $S(\hat{\theta}, L)$. Cette formule approchée est d'application très générale, dès lors qu'est estimée, ou bien postulée, la matrice V de covariance des données (i.e., des erreurs) ; en effet :

$$E(\delta L \delta L^T) = E(\epsilon(\tilde{\theta}) \epsilon^T(\tilde{\theta})) = V$$

La condition de normalité, qui jusqu'ici n'a pas été prise en compte, peut désormais être introduite. Ainsi, sous l'hypothèse (4) et avec le critère (5) :

$$\partial S / \partial \theta = \hat{B}^T \hat{V}^{-1} e(\hat{\theta})$$

$$\partial^2 \hat{S} / \partial \theta \partial L = \hat{B}^T \hat{V}^{-1} \quad (11)$$

où \hat{B} est la 1×3 matrice d'éléments \hat{b}_{1p} :

$$\hat{b}_{1p} = -\partial e_1(\hat{\theta}) / \partial \theta_p = \partial f(t_1, \hat{\theta}) / \partial \theta_p, \quad p = 1, 2, 3$$

avec $e_1(\theta)$ et $f(t_1, \theta)$ définis par les équations (6). En introduisant (11) dans l'expression (10) de $V(\hat{\theta})$ est obtenue l'estimation approchée :

$$\hat{V}(\hat{\theta}) = \hat{H}^{-1} \hat{B}^T \hat{V}^{-1} \hat{B} \hat{H}^{-1} \quad (12)$$

La formule (12) nécessite le calcul des éléments h_{pq} du Hessien :

$$h_{pq} = \partial^2 S / \partial \theta_p \partial \theta_q = - \sum_i \frac{1}{V_{ii}} e_{1i} \partial^2 f_1 / \partial \theta_p \partial \theta_q + \sum_i \frac{1}{V_{ii}} (\partial f_1 / \partial \theta_p) (\partial f_1 / \partial \theta_q)$$

où e_{1i} et f_1 désignent respectivement $e_1(\theta)$ et $f(t_1, \theta)$.

méthode de Newton-Raphson peut être employée ici: elle consiste à ne retenir que le second terme de l'expression h_{pq} , considérant que les écarts e_i sont petits, spécialement au voisinage de l'optimum; d'où :

$$H = B'V^{-1} B$$

Avec cette simplification, (12) devient :

$$\hat{V}(\hat{\theta}) = \hat{H}^{-1} \approx (\hat{B}'\hat{V}^{-1}\hat{B})^{-1} \quad (13)$$

Et en ajoutant l'hypothèse d'homoscédasticité (cf. (8)), l'estimation $\hat{V}(\hat{\theta})$ s'exprime simplement :

$$\hat{V}(\hat{\theta}) = (\hat{\sigma}^2 \sum_i n_{i1} (\partial \hat{f}_{i1} / \partial \theta) (\partial \hat{f}_{i1} / \partial \theta)')^{-1} \quad (14)$$

avec :
$$\hat{\sigma}^2 = \sum n_{i1} e_{i1}^2(\hat{\theta}) / d \quad (15)$$

où $d = \sum n_{i1}$ ou $\sum n_{i1} - 3$ selon que l'on utilise l'estimateur du maximum de vraisemblance ou une estimation non biaisée pour σ^2 .

Le vecteur colonne $\partial \hat{f}_{i1} / \partial \theta$, évalué à l'optimum du critère S , a pour composantes (cf. équation (1)) :

$$\partial \hat{f}_{i1} / \partial \theta = \begin{bmatrix} 1 - \exp(-\hat{\theta}_2 (t_{i1} - \hat{\theta}_3)) \\ \hat{\theta}_1 (t_{i1} - \hat{\theta}_3) \exp(-\hat{\theta}_2 (t_{i1} - \hat{\theta}_3)) \\ -\hat{\theta}_1 \hat{\theta}_2 \exp(-\hat{\theta}_2 (t_{i1} - \hat{\theta}_3)) \end{bmatrix}$$

Les approximations (13) et (14) ne sont qu'un cas particulier d'un résultat établi pour une large classe d'estimateurs du maximum de vraisemblance (cf. KENDALL et STUART, 1979) : leur distribution d'échantillonnage est asymptotiquement normale, de matrice de covariance :

$$V(\hat{\theta}) = - (E \partial^2 LL / \partial \theta \partial \theta)^{-1}, \quad n \rightarrow \infty$$

où LL désigne la log vraisemblance des n observations. L'espérance n'est généralement pas calculée, et sous l'hypothèse de normalité, l'utilisation de la "meilleure valeur" $\theta = \hat{\theta}$ conduit à : $\hat{V}(\hat{\theta}) \approx \hat{H}^{-1}$.

approximation dont la qualité dépend de celle de l'ajustement. Néanmoins, même si l'ajustement est satisfaisant, l'estimation $\hat{V}(\hat{\theta})$ est elle-même sujette à fluctuation d'un échantillon à l'autre. Aussi faudra-t-il retenir, suivant en cela BARD (1974), qu'une estimation $\hat{V}(\hat{\theta})$ obtenue à partir d'un unique échantillon ne doit être regardée au mieux que comme une approximation rudimentaire, valide à un ordre de grandeur près.

4. Estimation non paramétrique de matrice de covariance de l'estimateur : application de techniques de rééchantillonnage

Les résultats qui viennent d'être présentés sont tributaires d'une formulation explicite de la nature de la loi des erreurs. Il peut cependant être considéré que les conditions (4) et (8) sont trop contraignantes, voire irréalistes ; en particulier, parce que la densité normale n'accorde pratiquement aucune chance de réalisation aux événements "extrêmes", i.e. éloignés de la tendance centrale. Plus généralement, le choix a priori d'un modèle de distribution des résidus peut s'avérer délicat, sinon totalement arbitraire, d'où l'idée de s'affranchir de cette étape, et de recourir à des procédures d'estimation non paramétriques (~~au sens de "distribution free"~~). Le modèle précédemment postulé pour la loi des ϵ_i sera donc progressivement allégé dans ce qui va suivre.

4.1. Abandon de l'hypothèse de normalité

Il est supposé que la moyenne de n_i longueurs $l_{i,j}$ du groupe d'âge i est exactement décrite par l'équation (7) :

$$\bar{l}_i = f(t_i, \tilde{\theta}) + \epsilon_i \quad i = 1, 2, \dots, I$$

et que :

$\epsilon_1, \epsilon_2, \dots, \epsilon_k$ correspondant à la fonction de répartition F

$E(\epsilon_i) = 0, E(\epsilon_i \epsilon_h) = \sigma^2 \delta_{ih}$ δ_{ih} correspondant au symbole de KRONECKER.

La nature de la fonction de répartition F n'est pas précisée : il est seulement supposé que la loi des erreurs est d'espérance nulle et de variance inconnue σ^2 . Sous ces hypothèses, l'estimation généralement retenue est celle qui réalise le minimum du critère (9) des moindres carrés (estimation qui désormais ne correspond plus au maximum de la vraisemblance).

L'estimateur $\hat{\theta}_{MC}$ demeure quoiqu'il en soit une fonction des erreurs ϵ_i , par conséquent il peut être noté $\hat{\theta}(\epsilon_1, \epsilon_2, \dots, \epsilon_I)$. La vraie matrice de covariance $V(\hat{\theta})$ peut donc s'écrire :

$$V(\hat{\theta}) = V(F, I, \hat{\theta}) \quad (= V(F) \text{ pour simplifier}).$$

Le problème est d'obtenir une estimation de $V(F)$. Pour résoudre cette question, il sera fait appel à la technique du bootstrap, proposée en 1977 par EFRON (EFRON, 1979). La base conceptuelle en est succinctement exposée ci-après.

A l'estimation $\hat{\theta}$ est associé un ensemble d'écarts à l'ajustement $e_i(\hat{\theta})$:

$$e_i(\hat{\theta}) = \bar{l}_i - f(t_i, \hat{\theta})$$

Ces écarts définissent une cumulée empirique \hat{F} , qui constitue une estimation de la vraie fonction de répartition F . L'estimation bootstrap de la matrice de covariance est simplement $V(\hat{\theta})$ évaluée à $F = \hat{F}$:

$$\hat{V}(\hat{\theta}) = V(\hat{F})$$

L'utilisation de l'estimateur naturel $V(\hat{F})$ est légitimée par un résultat fort (théorème de Glivenko-Cantelli) :

$$I \rightarrow \infty \Rightarrow \text{Proba} \left\{ \sup_u | \hat{F}_I(u) - F(u) | \rightarrow 0 \right\} = 1$$

L'expression analytique de $V(F)$ n'étant pas connue, il est nécessaire de faire appel à une méthode de Monte-Carlo pour calculer $\hat{V}(\hat{F})$. L'algorithme de bootstrap, qui procède par rééchantillonnage des écarts, se déroule en trois étapes :

Etape (i) - Construire la fonction de répartition empirique \hat{F} .
 \hat{F} : attribution de la masse $1/I$ à chaque $e_i(\hat{\theta})$
 $i = 1, \dots, I$

Etape (ii) - Extraire de \hat{F} un "échantillon bootstrap" de taille I ,
i.e. un $I \times 1$ vecteur $e^*(\hat{\theta})$, réalisation de $\epsilon^*(\hat{\theta})$,
où :
 $\epsilon^*_1, \epsilon^*_2, \dots, \epsilon^*_I$ suivent la fonction de répartition \hat{F}

Les I valeurs $e^*_i(\hat{\theta})$ sont simplement obtenues par échantillonnage aléatoire simple, et tirages avec remise, de l'ensemble des écarts $\{e_1(\hat{\theta}), \dots, e_I(\hat{\theta})\}$. Il peut être noté (EFRON, 1982) que la proportion attendue des éléments de cet ensemble qui seront absents de l'échantillon bootstrap vaut $(1-1/I)^I \approx e^{-1}$.

Sur l'échantillon bootstrap est calculée une estimation $\hat{\theta}^*$ définie par :

$$S(\hat{\theta}^*) = \min_{\theta} \{S(\theta)\} = \min_{\theta} \left\{ \sum_i w_i e^*_i(\theta) \right\} \quad \begin{array}{l} \text{les } w_i \text{ désignant ici} \\ \text{un système de} \\ \text{pondération quelconque.} \end{array}$$

Etape (iii) - Effectuer un grand nombre B de répétitions indépendantes de l'étape (ii), afin d'obtenir B "replicats bootstrap" :
 $\hat{\theta}^*_1, \dots, \hat{\theta}^*_B, \dots, \hat{\theta}^*_B$

dont la distribution constitue une représentation de la distribution d'échantillonnage de $\hat{\theta}$. Par conséquent, la matrice de covariance $V(\hat{\theta})$ est estimée par :

$$\hat{V}(\hat{\theta}) = \hat{V}_B(\hat{F}) = \left(\sum_{b=1}^B (\hat{\theta}^*_b - \hat{\theta}^*) (\hat{\theta}^*_b - \hat{\theta}^*)^T \right) / (B - 1) \quad (16)$$

avec :

$$\hat{\theta}^* = \sum_b \hat{\theta}^*_b / B$$

En pratique, pour des estimations de variance, le nombre B de replications doit être de 50 à 200 (EFRON, 1981 a,b). Cette seconde valeur étant communément adoptée. Le nombre d'échantillon bootstrap distincts qu'il est possible de créer vaut quant à lui $\binom{2I-1}{I}$ (EFRON, 1983), avec la notation usuelle de combinatoire.

Le bootstrap a été ici présenté dans un contexte relativement classique : critère des moindres carrés, hypothèse d'espérance nulle des ϵ_i . Il est important de remarquer que la méthode s'appliquerait tout aussi bien en dehors de ce cas particulier, par exemple, en supposant les résidus indépendants ϵ_i centrés au sens où :

$$\text{Proba}_F (\epsilon < 0) = 1/2$$

et en définissant un critère d'optimalité tel que : $S(\theta) = \sum w_i |e_i(\theta)|$

L'application de l'algorithme permet d'estimer $V(\hat{\theta})$ dans ce cas de figure réfractaire à une analyse standard (EFRON, 1979, 1982). A fortiori, le bootstrap peut fournir une estimation $\hat{V}(\hat{\theta})$ si l'hypothèse de normalité (8) est conservée : il suffit de remplacer \hat{F} par \hat{F}_{NORM} à l'étape (i) de l'algorithme,

$$\hat{F}_{\text{NORM}} = N_I (e, \sigma^2 I)$$

où $\hat{\sigma}^2$ est définie par (15), et e désigne le $I \times 1$ vecteur de composantes $e(\hat{\theta})$. \hat{F}_{NORM} est l'estimation paramétrique de F , au sens du maximum de vraisemblance, d'où le nom de "bootstrap paramétrique" attribué à cette approche (EFRON, 1982).

Par ailleurs, il n'a été jusqu'à présent question que de l'estimation de la matrice de covariance des paramètres, mais d'autres quantités peuvent être obtenues à l'aide de la distribution des replicats $\hat{\theta}^*_b$; par exemple le vrai biais β de l'estimateur $\hat{\theta}$:

$$\beta = E(\hat{\theta} - \tilde{\theta}) = E_F(\theta(\hat{F}) - \theta(F))$$

peut-être estimé par :

$$\hat{\beta} = \hat{\theta} - \sum_b \hat{\theta}^*_b / B = \hat{\theta} - \hat{\theta}^*$$

De sorte que les estimateurs $\hat{\theta}_p$ peuvent être comparés d'après leur erreur quadratique moyenne MSE :

$$\widehat{\text{MSE}}(\hat{\theta}_p) = \hat{v}_p + \hat{\beta}^2_p$$

où \hat{v}_p désigne le p -ème élément diagonal de $\hat{V}(\hat{\theta})$.

4.2. Abandon de l'hypothèse d'une loi commune aux I résidus

A ce stade ne sont plus conservées que les hypothèses a), b) et c) précédemment énoncées. Il est maintenant admis que la distribution des ϵ_i change avec l'âge t_i ; cela inclut en particulier le cas de l'hétéroscédasticité. Dans le groupe d'âge i , la répartition des $l_{i,j}$ autour de $f(t_i, \hat{\theta})$ définit donc n_i variables aléatoires $\epsilon_{i,j}$:

$\epsilon_{i,1}, \epsilon_{i,2}, \dots, \epsilon_{i,n_i}$ correspond à la fonction de répartition F_i

Les seules caractéristiques postulées de la loi F_i étant :

$$E(\epsilon_{i,j}) = 0, E(\epsilon_{i,j} \epsilon_{i,l}) = \sigma^2_{i,l}$$

Il est d'autre part supposé qu'il n'y a pas de corrélation entre les résidus associés aux différents âges t_i :

$$E(\epsilon_{i,j}, \epsilon_{h,j}) = 0 \quad i, h = 1, \dots, I \quad h \neq i$$

Sous ces hypothèses est choisi le critère d'optimalité :

$$S(\theta) = \sum_{i=1}^k \sum_{j=1}^{n_i} w_i (l_{i,j} - f(t_i, \theta))^2 = \sum_{i=1}^k \sum_{j=1}^{n_i} e_{i,j}^2(\theta) \quad (17)$$

où w_i peut être, par exemple, égal à 1 ou à $1/\hat{\sigma}_{t_i}^2$, $\hat{\sigma}_{t_i}^2$ étant estimé à partir des $l_{i,j}$. Le critère d'optimalité peut aussi se référer à une valeur centrale des $l_{i,j}$, pour i donné, et non aux valeurs individuelles.

De même que dans le modèle de corrélation, l'algorithme du bootstrap ne doit pas occulter la dépendance de F_i vis à vis de t_i (cf. FREEDMAN, 1981). En conséquence, les étapes (i) et (ii) sont modifiées comme suit :

Etape (i) - Construction des I estimations \hat{F}_i avec attribution de la masse $1/n_i$ à chaque $e_{i,j}(\hat{\theta})$ $j=1, 2, \dots, n_i$

Etape (ii) - La création de l'échantillon bootstrap procède par rééchantillonnage des I groupes d'âge :

ii a - dans le groupe d'âge t_i , échantillonnage aléatoire simple et tirages avec remise de n_i éléments dans l'ensemble des écarts $\{e_{i,1}(\hat{\theta}), \dots, e_{i,n_i}(\hat{\theta})\}$. Cette procédure livre n_i réalisations $e_{i,j}^*(\hat{\theta})$ des variables aléatoires :

$$e_{i,1}^*, \dots, e_{i,n_i}^* \text{ correspondant à } \hat{F}_i$$

ii b - puis effectuer $I-1$ répétitions indépendantes de l'étape (ii a) pour $i=2, 3, \dots, I$. Ainsi est obtenu l'échantillon bootstrap :

$$\{e_{i,j}^*(\hat{\theta}) ; j=1, \dots, n_i ; i=1, \dots, I\}$$

auquel correspond l'estimation $\hat{\theta}^*$ définie par :

$$S(\theta^*) = \min_{\theta} \left\{ \sum_i w_i \sum_j e_{i,j}^{*2}(\hat{\theta}) \right\}$$

les w_i étant définis selon une procédure quelconque.

L'étape (iii) de l'algorithme n'est pas modifiée, et l'estimation de la matrice de covariance des paramètres est donnée par la formule (16).

Ici encore, il peut être remarqué que lorsque l'hypothèse (4) de multinormalité est admise, l'approximation (13) obtenue par la voie analytique peut être confortée par un bootstrap paramétrique, avec :

$$\hat{F}_{NORM} = N_I(\bar{e}, \hat{V})$$

où \bar{e} est le vecteur de composantes $\bar{e}_i(\hat{\theta}) : e_i(\hat{\theta}) = \sum_j e_{i,j}(\hat{\theta})/n_i$ et \hat{V} la matrice diagonale d'éléments $\hat{\sigma}_{t_i}^2$.

4. 3. Deux cas particuliers typiques des études de croissance

4. 3. 1. Plusieurs couples âge-longueur extraits du même individu: abandon de l'hypothèse de non corrélation des erreurs

Cette configuration se présente lorsque sur chaque poisson d'âge t_1 est calculée à l'aide d'une relation d'allométrie ("rétrocalculée"), la longueur qu'il mesurait antérieurement aux âges t_{1-1}, t_{1-2}, \dots . Le j-ème poisson engendre ainsi plusieurs couples $(l_{1j}, t_1), (l_{1-1,j}, t_{1-1}), \dots$ etc, de sorte que l'hypothèse d'absence de corrélation entre les erreurs $\epsilon_1, \epsilon_{1-1}, \dots$ ne peut plus être acceptée.

Le problème est alors le suivant : une procédure de rééchantillonnage aléatoire des données doit opérer sur des unités d'échantillonnage indépendantes. Dans le cas présent, les seuls éléments saisis indépendamment les uns des autres sont les poissons. Ce sont donc ces derniers qui seront rééchantillonnés.

La technique du Jackknife peut aisément être appliquée à ce cas. Introduite par M. QUENOUILLE en 1949 (cf. MILLER, 1974), elle fut originellement conçue comme une méthode de réduction du biais. A la suite d'une conjecture de J. TUKEY, elle fut ensuite largement appliquée à l'estimation des variances. Elle a, en particulier, été employée pour estimer la matrice de covariance des paramètres dans les modèles de régression non linéaire (DUNCAN, 1978 ; FOX et al. 1980 ; SIMONOFF et TSAI, 1986). D'un point de vue théorique, il a été montré que le Jackknife est un bootstrap particulier, en ce sens qu'il ne rééchantillonne que n points au lieu de B parmi $\binom{2n-1}{n}$, et qu'en outre il ne nécessite pas de procédure de Monte-Carlo car il utilise une approximation par une fonctionnelle linéaire de la statistique dont on cherche à apprécier la variance (EFRON, 1982, 1983).

L'algorithme de calcul de $\hat{V}(\hat{\theta})$ est le suivant, soit : $\{J_1, J_2, \dots, J_n\}$ l'ensemble des n individus (poissons) qui sont servi à établir la base de données, plusieurs (âge, longueur) étant associés à chaque poisson. La minimisation du critère (17) permet d'obtenir l'estimation $\hat{\theta}_{MC}$, qui peut être notée :

$$\hat{\theta}_{MC} = \hat{\theta}(J_1, J_2, \dots, J_n)$$

soit :

$$\hat{\theta}_{-r} = \hat{\theta}(J_1, \dots, J_{r-1}, J_{r+1}, \dots, J_n)$$

c'est-à-dire l'estimation calculée avec le même critère (17), mais après suppression des couples (âge, longueur) obtenus sur le r -ème poisson. En répétant le calcul pour chacun des n individus J_r sont obtenus n vecteurs P_r :

$$P_r = n \hat{\theta}_{MC} - (n-1) \hat{\theta}_{-r}$$

et :

$$P. = \sum_{r=1}^n P_r/n \quad \begin{array}{c} \uparrow \\ \text{dites pseudo-valeurs définissant} \\ \text{l'estimateur Jackknife} \end{array}$$

La matrice de covariance des paramètres est estimée par :

$$\widehat{V}(\widehat{\Theta}) = \frac{n}{\sum_{r=1}^n (P_r - P.) (P_r - P.)^2} / (n(n-1)) \quad (18)$$

4.3.2. Erreurs d'âge

La lecture des pièces anatomiques qui enregistrent l'âge du poisson est souvent délicate. Il peut donc être considéré que la base de données inclut généralement une certaine proportion d'individus auxquels a été attribué un âge erroné : pour ceux-ci, le modèle postulé (3) est par conséquent faux.

Il sera supposé que la proportion des erreurs commises sur l'ensemble des lectures d'âge est faible : si tel n'était pas le cas, la question de l'ajustement d'un modèle aux données serait alors sans objet.

Le problème consiste à définir une procédure d'estimation qui soit la moins sensible possible à cette perturbation. Considérons le cas où il n'y aurait aucune erreur d'âge : dans le groupe i , les réalisations $l_{i,j}$ de la variable aléatoire L_i (de densité ϕ_i) se répartissent de manière équilibrée autour de $f(t_i, \Theta)$. Supposons au contraire qu'est introduite une erreur qui, par exemple, surestime de manière systématique l'âge de certains poissons : des poissons appartenant en réalité au groupe $i-1$ (voire $i-2$ ou pire) seront alors classés dans le groupe i . Cela aura pour effet d'épaissir artificiellement vers la gauche la densité ϕ_i de la variable "contaminée" L_i (si on admet que la forme de ϕ_i n'est pas affectée dans le même temps par l'erreur consistant à attribuer aléatoirement quelques individus du groupe i au groupe $i+1$, voire $i+2, \dots$). Un phénomène symétrique se produira si l'erreur tend à sous-estimer l'âge.

L'estimation de Θ par optimisation d'un critère tel que (9) repose sur une minimisation de la distance au modèle d'un indicateur de la tendance centrale des densités ϕ_i . Si les extrémités de celles-ci sont déformées par des erreurs d'âge, force est alors de définir un estimateur de position qui soit robuste (cf. HUBER, 1972 ; HAMPEL, 1973) face à ces queues de distribution épaissies.

Le choix de la moyenne arithmétique \bar{L}_i se révèle en l'occurrence fâcheux : cet estimateur est réputé particulièrement sensible aux extrémités de la distribution. Plusieurs estimateurs de position ont été proposés, pour lesquels l'influence des valeurs extrêmes est réduite : les propriétés des plus classiques d'entre eux ont été étudiées par ANDREWS et al. (1972). Par exemple, la moyenne α -tronquée (qui appartient à la classe des L -estimateurs, i.e. résultant d'une combinaison linéaire de statistiques d'ordre) est obtenue en supprimant les $[\alpha n_i]$ (avec $0 < \alpha < 1/2$) premières et dernière observations ordonnées $l_{i,j}$ du groupe d'âge i ; en supposant pour simplifier αn_i entier :

$$\bar{l}_{i,\alpha} = \frac{n_i - \alpha n_i}{\sum_{j=\alpha n_i + 1}^{n_i - \alpha n_i} l_{i,j}} / (n_i - 2\alpha n_i)$$

Aux bornes de α correspondent la moyenne arithmétique usuelle ($\alpha = 0$) et la médiane ($\alpha = 1/2$). L'idée d'utiliser la médiane pour calculer le critère d'écart à un modèle de régression n'est pas nouvelle : attribuée à EDGEWORTH, elle daterait d'environ un siècle. Dans cet esprit peuvent être définis :

$$e_i(\theta) = \bar{l}_{i, 1/2} - f(t_i, \theta) \text{ et } S(\theta) = \sum_i w_i |e_i(\theta)|$$

$$\text{ou } S(\theta) = \sum_i w_i e_i^2(\theta)$$

Les w_i définissant un système de pondération quelconque.

Il faut toutefois remarquer qu'un estimateur robuste tel que \bar{l}_i , va élaguer symétriquement les deux extrémités de la distribution empirique des $l_{i,j}$. Ce n'est pas la démarche la plus appropriée si l'erreur d'âgeage est systématique (vide supra) : seule la queue de droite est concernée s'il y a sous-estimation celle de gauche en cas de surestimation. Dans l'un ou l'autre cas, et pour les jeunes classe d'âge (i.e. pour lesquelles l'accroissement Δl entre deux âges consécutifs est élevé) un second mode tend à se dessiner dans la distribution de longueur des $l_{i,j}$. Etant admis que la proportion des âges erronés est faible, le mode majeur coïncide sensiblement avec celui qui serait observé en l'absence de lectures d'âge fausses (sous réserve que les effectifs n_i soient comparables). Au voisinage de l'asymptote de la courbe de croissance ("vieilles" classes, Δl négligeable), les erreurs d'âgeage deviennent pratiquement sans effet.

Soit donc $mode_i$ la position de la classe modale des $l_{i,j}$; s'il est suspecté que plusieurs poissons ont été classés à tort dans certains groupes d'âge, alors les écarts aux valeurs calculées par le modèle sont définies par :

$$e_i(\theta) = mode_i - f(t_i, \theta)$$

S'il est de surcroît souhaité n'accorder qu'une influence réduite aux écarts les plus forts, alors l'estimateur θ défini par minimisation de :

$$S(\theta) = \min \sum_i w_i |e_i(\theta)|$$

peut être préféré à l'estimateur θ_{MC} des moindres carrés. Dans l'un comme dans l'autre, l'estimation de la matrice de covariance de l'estimateur est calculée par l'algorithme du bootstrap (formule (16)), ou si le rétrocalcul est pratiqué, du Jackknife. Il faut d'autre part s'assurer que les valeurs estimées des paramètres ne sont pas affectées trop nettement par le choix de l'amplitude de classe de l'histogramme des $l_{i,j}$.

5. Intervalles de confiance - Région de confiance

5.1. Intervalles de confiance conditionnels

Les termes diagonaux v_p de la matrice de covariance $V(\hat{\theta})$ permettent d'associer à chaque estimation $\hat{\theta}_p$ ($p=1, \dots, m$, ici $m=3$) un intervalle de confiance, à condition que soit connue la distribution d'échantillonnage de l'estimateur. Lorsque $\hat{\theta}$ est un estimateur du maximum de vraisemblance (critère (5) ou (9) sous l'hypothèse (4) ou (8) respectivement), sa loi tend vers la normalité au fur et à mesure qu'augmente le nombre d'observations. Par conséquent :

$$\text{Proba } \{ \tilde{\theta}_p \in [\hat{\theta}_p - t_\alpha \hat{v}_k, \hat{\theta}_p + t_\alpha \hat{v}_k] \} \approx 1 - 2\alpha \quad (19)$$

où t_α désigne la borne de quantile supérieur de taille α de la loi de Student.

Lorsque la loi des erreurs n'est pas connue, la minimisation d'un critère d'écart définit un estimateur $\hat{\theta}_p$ dont la distribution d'échantillonnage est elle-même généralement inconnue. Il est néanmoins possible de construire un intervalle de confiance approché et non paramétrique à l'aide des replicats bootstrap $\hat{\theta}_{1,p}^*, \dots, \hat{\theta}_{b,p}^*, \dots, \hat{\theta}_{B,p}^*$, qui réalisent une estimation de la distribution d'échantillonnage de θ_p (EFRON, 1981 a). Soit donc \hat{C} la cumulée empirique des B replicats bootstrap :

$$\hat{C}(t) = \text{Proba}^* (\hat{\theta}_p^* < t) = (\# \hat{\theta}_{b,p}^* < t) / B$$

Le symbole $\#$ signifie "nombre de fois où", et la notation Proba^* rappelle que c'est le statisticien qui crée l'aléatoire en utilisant l'algorithme du bootstrap. L'intervalle de confiance à $100(1-2\alpha)\%$ de $\hat{\theta}_p$ est borné par les percentiles 100α et $100(1-\alpha)$ de \hat{C} .

$$\text{Soit : } \hat{\theta}_p(\alpha) = \hat{C}^{-1}(\alpha) \text{ et } \hat{\theta}_p(1-\alpha) = \hat{C}^{-1}(1-\alpha)$$

où \hat{C}^{-1} désigne la fonction réciproque de \hat{C} ; alors :

$$\text{Proba } \{ \tilde{\theta}_p \in [\hat{\theta}_p(\alpha), \hat{\theta}_p(1-\alpha)] \} \approx 1 - 2\alpha \quad (20)$$

Il convient de ne pas choisir un risque de première espèce trop faible, car il n'est pas assuré que les extrémités de la distribution empirique des $\hat{\theta}_{b,p}^*$ reproduisent correctement celles de la loi des $\hat{\theta}_p$ (voir la remarque de NASH dans la discussion de l'article de EFRON, 1981 a).

Par ailleurs, les intervalles (19) et (20) sont conditionnels : les limites de confiance approchées qui encadrent $\hat{\theta}_p$ sont en effet déterminées par les valeurs des autres composantes $\hat{\theta}_q$ ($q \neq p$) de l'estimation $\hat{\theta}$. Usuellement, les intervalles de confiance attachés à l'estimation de chacun des paramètres du modèle sont donnés pour tous les autres fixés à leur valeur optimale.

5.2. Région de confiance

Dans l'espace paramétrique R^m , la région de confiance est le domaine D_α limité par une hypersurface de dimension $m-1$, et défini par :

$\theta \in D_\alpha$ si le domaine de dispersion au seuil α attaché à l'hypothèse θ contient α , le domaine de dispersion étant limité par une isopente de la densité de probabilité et contenant $1 - \alpha$ de la probabilité totale. Si les domaines de dispersion sont de morphologie indépendante de θ , et symétrique par rapport à θ la région de confiance coïncide avec le domaine de dispersion attaché à $\hat{\theta}$.

A la différence des intervalles conditionnels, la région de confiance permet de détecter les liens qui existent entre les m composantes de $\hat{\theta}$. En règle générale, il ne doit y avoir entre les paramètres d'un modèle que la plus faible dépendance possible ; le modèle de Von Bertalanffy est à cet égard fort peu satisfaisant, compte tenu de la forte corrélation entre $\hat{\theta}_1$ et $\hat{\theta}_2$ (i.e., K et $L\infty$). Cela induit dans l'espace paramétrique un étirement de la région de confiance selon une direction de "mauvaise détermination". Des reparamétrages peuvent être utiles (GALLUCCI et QUINN, 1979), les considérations précédentes pouvant aisément s'adapter à ces nouvelles formulations des équations de croissance.

L'équation approchée qui définit la région D_α peut être obtenue par un développement de Taylor du critère $S(\theta)$ au voisinage de $\hat{\theta}$, jusqu'à l'ordre 2 :

$$S(\theta) \approx S(\hat{\theta}) + (\theta - \hat{\theta})^T (\partial S(\hat{\theta}) / \partial \theta) + 1/2 (\theta - \hat{\theta})^T \hat{H} (\theta - \hat{\theta})$$

Soit encore, compte tenu de la condition nécessaire d'extrémum :

$$S(\theta) - S(\hat{\theta}) \approx 1/2 (\theta - \hat{\theta})^T H (\theta - \hat{\theta})$$

Lorsque $\hat{\theta}$ est un estimateur du maximum de vraisemblance (critère (5) ou (9)), la tendance vers la normalité de la distribution permet d'écrire, avec l'approximation (13) :

$$D_\alpha \approx \{ \theta : (\theta - \hat{\theta})^T \hat{V}^{-1}(\theta) (\theta - \hat{\theta}) < x^2_{m, \alpha} \} \quad (21)$$

où $x^2_{m, \alpha}$ désigne la borne du quantile de taille α de la loi de χ^2 à m degrés de liberté. D_α est alors un m -ellipsoïde du volume :

$$\sqrt{x^2_{m, \alpha}} \Pi \text{ et } (V^{-1}) / \Gamma(m/2 + 1)$$

et dont le grand axe correspond à la direction de plus faible détermination des paramètres θ_p . Il est également possible de définir un contour exact pour D_α (cf. DRAPER et SMITH, 1966 ; BARD, 1974 ; KIMURA, 1980). La non linéarité du modèle (1) implique alors que le risque effectif diffère du risque nominal ; en outre, D_α n'est plus dans ce cas un ellipsoïde.

Quand la distribution d'échantillonnage de $\hat{\theta}$ n'est pas connue, il devient hautement spéculatif de recourir aux équations présentées dans les ouvrages mentionnés ci-dessus. En revanche, les relations entre paramètres peuvent être clairement mises en évidence par une analyse en composantes principales de la matrice $(\hat{\theta}^*_1, \dots, \hat{\theta}^*_B)$ des replicats bootstrap: les directions de plus forte et de plus faible détermination correspondent aux vecteurs propres associés respectivement à la plus petite et à la plus grande valeur propre de la $m \times m$ matrice des corrélations entre paramètres.

Pour un couple de paramètres, un histogramme bidimensionnel des estimations bootstrap sera d'autant plus utile qu'il permettra d'apprécier des distributions non normales et des domaines de dispersion ellipsoïdaux.

6. Le logiciel

6.1. Fonctions générales

Ecrit en Fortran IV pour ordinateur universel, il permet d'effectuer un ajustement par les moindres carrés, éventuellement pondérés, de longueurs centrales pour des âges régulièrement espacés d'un an. Les longueurs centrales considérées correspondent à la moyenne arithmétique, au mode principal ou à la médiane.

La pondération éventuelle peut correspondre aux effectifs, ou aux effectifs divisés par la variance estimée des longueurs pour un âge donné.

Les résultats de l'ajustement peuvent être systématiquement assortis de variances covariances calculées par la formule 12 et la formule approchée 13. En outre, des techniques de rééchantillonnage peuvent être appliquées dans tous les cas : Jackknife, Bootstrap non paramétrique et Bootstrap paramétrique. Les bootstraps peuvent être faits en opérant pour un âge avec une distribution spécifique de l'âge. Dans ce dernier cas on peut recentrer ou non les résidus pour garantir que leur espérance est ou non nulle, à chaque âge. Les éléments d'inférences fournis correspondent à une estimation des biais, variances et covariances.

Le programme prévoit en outre de traiter des données incluant des rétrocalculs, la technique non paramétrique suggérée étant alors le **Jackknife**.

6.2. Remarques techniques

L'ajustement par les moindres carrés est effectué selon la procédure de Tomlinson et Abramson (1961). Elle s'est en effet avérée très rapide, ce qui dans les techniques de rééchantillonnage est essentiel. Pour le Bootstrap la création de 200 échantillons a été choisie, sans créer de problème de temps de calcul.

Les données d'entrée sont "prétraitées" par le programme pour créer des histogrammes de distribution de longueur. Cela modifie un peu le calcul des valeurs centrales. Dans le calcul des moyennes arithmétiques pour chaque classe de taille, la moyenne de la classe est attribuée à tous les poissons concernés. Pour le calcul du mode, la classe de plus forte fréquence est d'abord identifiée. Si elle est multiple, la moyenne arithmétique de la valeur la plus forte et de la valeur la plus faible est considérée. En cas contraire, une seule classe, correspondant à un maximum strict apparaissant, on considérera d'une part le centre de la classe concernée, d'autre part le centre des deux classes adjacentes. La valeur retenue comme estimation du mode sera la moyenne pondérée de ces trois valeurs, la pondération étant faite selon les fréquences respectives.

Si la fréquence maximale observée est à une extrémité de l'histogramme, seule la valeur adjacente interne est considérée.

Pour le calcul de la médiane, du fait de la création d'un histogramme, la fonction de répartition est discrétisée. La valeur correspondant à une probabilité de 0.5, définissant la médiane est obtenue par interpolation linéaire.

6.3. Développement

La possibilité de construire des histogrammes uni et bidimensionnels des pseudo-valeurs (Jackknife) ou des valeurs obtenues par les échantillons Bootstraps sera introduite au plus tôt.

Dans un deuxième temps, la possibilité d'effectuer un ajustement au sens des moindres valeurs absolues, et non des moindres carrés sera incorporée.

Une transcription en Fortran 77 sera enfin effectuée.

Le logiciel ainsi banalisé sera disponible auprès des auteurs, sans garantie, ni maintenance.

7. Une illustration

L'impossibilité de disposer suffisamment tôt d'une version opérationnelle du logiciel a réduit son application à un exemple, dont les conclusions ne sauraient être généralisées.

7.1. Données traitées

Les données correspondantes ont gracieusement été mises à notre disposition par notre collègue P. BERTHOU (1). Elles concernent des spicules (Spisula ovalis) provenant du golfe de Gascogne, pour un gisement et pratiquement une cohorte. Ces précautions sont liées au fait que la croissance de la spicule varie notablement d'un banc, et d'une cohorte à l'autre. Les "longueurs" analysées correspondent à des hauteurs aux anneaux sur un ensemble de 35 individus.

Les données ont délibérément été tronquées pour créer un échantillon où les âges maximaux considérés se répartissent de façon équilibrée de un à sept ans.

Les données de base ainsi définies apparaissent dans le tableau I.

7.2. Comparaison des différents résultats

Les ajustements ont été faits soit en excluant les données de rétrocalcul (tableau II), soit en les incorporant (tableau III). Les valeurs centrales et considérées correspondent aux trois possibilités évoquées (moyenne - mode principal - médiane).

(1) P. BERTHOU - IFREMER (RH/BPA) - B. P. 337 - 29273 Brest Cédex

Les ajustements ont été faits par les moindres carrés, soit équipondérés, soit pondérés par les effectifs, soit enfin pondérés par le rapport effectif/variance. Cette dernière variante n'a été appliquée qu'aux moyennes. Les résultats reportés donnent d'une part valeurs centrales observées et prédites, d'autre part les estimations et variances estimées pour K et L_{∞} , à l'exclusion de t_0 et des covariances, pour ne pas alourdir les tableaux.

Les techniques de Bootstrap n'ont pas été appliquées à la situation autorisant le rétrocalcul, car les hypothèses d'indépendance implicitement requises ne paraissent pas satisfaites. Les calculs correspondant à des bootstraps où la distribution des résidus varie d'un âge à l'autre n'ont pas été reportés. Il apparaît en effet que les données ne sont pas suffisamment nombreuses pour estimer un histogramme de résidus par âge (bootstrap non paramétrique) ou une variance par âge (bootstrap paramétrique).

Les résultats obtenus donnaient des variances systématiquement plus faibles. On peut à cet égard noter que jusqu'à présent le problème du défaut d'ajustement (lack of fit) n'a pas été couvert. Le problème est néanmoins réel et sur l'exemple traité un examen des variances résiduelles comparé à l'erreur pure nous montre qu'il est même important. Les calculs de l'erreur pure, sous hypothèse d'homoscédasticité et avec rétrocalcul, conduisent à une variance estimée à 1.6. Le calcul des variances résiduelles, après regression, par la formule (15) mène, par exemple pour l'ajustement pondéré par les n_i des moyennes arithmétiques, donne une variance estimée à 7.3. Les autres techniques mènent à des variances analogues. Un facteur de 4.5 apparaît donc. On peut voir sur les tableaux II et III que la plus grave difficulté est liée aux âges 2 et 3. Les hauteurs prédites à deux ans sont toujours inférieures aux observations, cette situation l'inversant à trois ans. Lorsque l'on crée une distribution globale des résidus, la dispersion correspondante confond en réalité erreur pure et défaut d'ajustement. Lorsque l'on garde une distribution spécifique par âge, seule l'erreur pure sera prise en compte dans les rééchantillonnages. Les variances correspondantes seront naturellement plus faibles. Cette remarque joue aussi quand on effectue des comparaisons avec le Jackknife, car en l'occurrence les rééchantillonnages ne viennent pas mêler défaut d'ajustement et erreur pure. Dans le même ordre d'idée on notera que les calculs théoriques correspondant aux formules 12 et 13, lorsqu'ils calculent les variances par la formule 15, confondent erreur pure et défaut d'ajustement. Si, en revanche, on estime directement σ_i^2 et $w_i = n_i/\sigma_i^2$ d'après la distribution des longueurs pour l'âge i , seule l'erreur pure sera prise en compte.

L'ajustement par les n_i/σ_i^2 ne correspond pas à des calculs de Bootstrap dans les tableaux II et III. En fait, les calculs ont été faits, mais conduisent à des aberrations. Dans les échantillons Bootstrap il apparaît systématiquement un ou quelques cas où une variance σ_i^2 est très faible, voire nulle pour le Bootstrap non paramétrique, créant ainsi des pondérations déraisonnables.

Au delà de ces remarques préliminaires on peut d'abord noter la faiblesse des biais estimés, à quelques exceptions où les variances correspondantes sont suffisamment fortes pour que l'estimation du biais paraisse peu fiable. Cela recoupe les simulations de VAUGHAN et KANCIRUK (1982). On peut aussi noter la concordance d'ensemble pour l'ajustement sur

moyennes arithmétiques, entre variances calculées par les formules théoriques, et estimations obtenues par bootstrap paramétrique. Les hypothèses utilisées par les deux approches sont en fait les mêmes, les formules théoriques impliquant des approximations. La seule exception correspond à l'ajustement avec pondération par n_i/σ_i . La divergence correspond alors au fait que le bootstrap opère avec la variance résiduelle calculée après régression, les formules théoriques utilisant des estimations directes des σ_i^2 . On a déjà noté la forte différence, liée au défaut d'ajustement. Hors ce cas particulier, les formules théoriques paraissent respecter, pour les moyennes arithmétiques, les ordres de grandeur. L'approximation correspondant à la formule (13) marque cependant un différence non négligeable avec la formule (12), sans remettre en cause les ordres de grandeur. Elle est systématiquement plus faible. On notera que l'application aux ajustements sur modes et médianes des formules (12) et (13) mène à de fortes divergences avec le bootstrap paramétrique. Rien ne garantit en effet, surtout avec de faibles effectifs, la normalité des résidus obtenus avec modes et médianes.

La comparaison du bootstrap paramétrique et du bootstrap non paramétrique met en évidence, pour l'ajustement sur moyennes arithmétiques, une bonne concordance. Cela suggérerait que la distribution des valeurs individuelles $l_{i,j}$ ne joue pas un rôle majeur. Cela n'est plus vrai toutefois pour modes et médianes.

La comparaison de l'ensemble des autres techniques avec le Jackknife montre cette fois de fortes divergences. Elles peuvent être rapprochées du fait précédemment signalé : le Jackknife ne confond pas erreur pure et défaut d'ajustement. Les autres techniques, à l'exception de celle opérant une pondération par n_i/σ_i^2 le font. Elle pèchent ainsi par pessimisme et les variances obtenues par Jackknife sont systématiquement plus faibles. Toutefois, les résultats correspondant à la pondération par n_i/σ_i^2 sont eux "optimistes" par comparaison au Jackknife.

Une comparaison de colonne à colonne montre une plus faible variance des résultats obtenus en utilisant les moyennes arithmétiques si l'on considère les variances Jackknife. C'est particulièrement vrai lorsque les données de rétrocalcul sont incorporées. Un examen attentif des pseudo-valeurs, et la possibilité d'un tel examen constitue un réel atout du Jackknife, montre que celles-ci sont particulièrement variables lorsque l'on ôte un individu parmi ceux d'âge 6 ou 7 ans. L'ajustement est très sensible aux valeurs centrales utilisées à ces âges. Modes et médianes sont plus sensibles que la moyenne et ce d'autant plus que les effectifs à 6 et 7 ans sont faibles. L'histogramme à 6 ans obtenu avec rétrocalcul, s'il est d'effectif le plus élevé qu'en l'absence de rétrocalcul, ne présente pas de pic net. Il est donc très sensible à la perturbation créée par la suppression d'un seul poisson.

Au delà de l'exemple traité, il apparaît que modes et médianes sont des statistiques fragiles quand les effectifs sont faibles.

On peut encore noter que les divergences entre résultats, globalement faibles sur L_∞ , sont plus fortes sur K , mais que les divergences sont plus grandes selon que l'on inclut ou non le rétrocalcul que d'une technique d'ajustement à l'autre.

Les comparaisons d'un système de pondération à l'autre, en gardant par exemple le Jackknife comme indicateur de la stabilité, ne font pas apparaître le bénéfice net à utiliser une technique plutôt qu'une autre. Tout au plus remarquera-t-on que la pondération par n_i/σ_i^2 ne paraît pas donner de très bons résultats. Cela peut être rapproché du fait que les estimations de σ_i^2 obtenus à partir des dispersions des $l_{i,j}$ peuvent être médiocres si les n_i sont faibles.

Equipondération et pondération par les n_i ne peuvent en outre être comparées uniquement en termes de variance. Du fait du défaut d'ajustement, ce n'est pas exactement la même courbe qui est estimée lorsque l'on passe d'un système d'ajustement à l'autre. C'est particulièrement vrai en cas de rétrocalcul, où les effectifs aux jeunes âges sont largement plus forts. Dans le cas traité les divergences sont modérées, mais le tableau III fait néanmoins apparaître des valeurs toujours plus fortes pour K en cas d'équipondération.

7.3. Perturbation des lectures d'âge

A partir des données correspondant au tableau I, un jeu perturbé a été créé en introduisant un "faux" anneau, à mi-distance entre les "vrais" premiers et seconds anneaux, pour les individus 1, 6, 13, 18 et 23. Les résultats obtenus par les différentes techniques apparaissent dans le tableau IV. Les perturbations sur L_∞ sont globalement faibles. K est plus sensible, et globalement ce sont les méthodes utilisant le mode qui résistent le mieux. Ce résultat est strictement conforme aux espérances, à l'exception des performances médiocres des résultats liés à la médiane.

Les illustrations devront être multipliées, mais il n'est pas étonnant que l'utilisation des modes protège mieux contre les erreurs d'âgeage, tout en étant plus sensible aux variabilités individuelles de croissance, car la détermination du mode principal n'utilise qu'une fraction de l'information apportée par la distribution des longueurs pour un âge donné.

CONCLUSION

Le logiciel développé permet d'appliquer à l'ajustement de la courbe de Von Bertalanffy des procédures désormais classiques en statistiques, mais qui à notre connaissance n'avaient pas fait l'objet d'une application systématique à ce problème particulier.

Seule, l'expérience accumulée sur un ensemble de cas réels, et de données simulées pour couvrir des difficultés précises permettra d'apprécier le mérite relatif des différentes procédures. La question des défauts d'ajustement mériterait dans ses simulations une attention particulière.

REFERENCES CITEES

- ANDREWS, D. L., BICKEL P. J., HAMPEL F. R., HUBER P. J., ROGERS W. H. & TUKEY J. W., 1972. - Robust estimates of location. Survey and advances. Princeton University press, Princeton, N. J., 373 p.
- BARD, Y., 1974. - Non linear parameter estimation. Academic Pres, New-York, San Francisco, London, 341 p.
- DRAPER N. R., & SMITH H., 1966. - Applied Regression Analysis. John Wiley & Sons, INC. New-York, London, Sydney, 407 p.
- DUCAN G. T., 1978. - An empirical study of Jackknife - Constructed confidence regions in non linear regression. Technometrics 20 (2) : 123-129.
- EFRON B., 1979. - The 1977 Rietz lecture. Bootstrap methods : another look at the Jackknife. Ann. Statist 7 (1) : 1-26.
- EFRON B., 1981 a. - Non parameters standard errors and confidence intervals. Can. J. Stat. 9 (2) : 139-172.
- EFRON B., 1981 b. - Non parametric estimates of standard error : the Jackknife, the bootstrap, and other methods. Biometrika 68 (3) : 589-599.
- EFRON B., 1982. - The jackknife, the bootstrap and other resampling plans. CMBS-NSF, Regional conference series in Applied Mathematics n° 38, SIAM ed., Philadelphia, Pennsylvania, 92 p.
- EFRON B. & G. GONG, 1983. - A leisurely look at the bootstrap, the jackknife and cross-validation. Amer.stat. 37 (1) : 36-48.
- FOX T., HINKLEY D. & LARNTZ, 1980. - Jackknifing in non linear regress. Technometric, 22 (1) : 29-33.
- GALLUCI V. F. & QUINN T. J., 1979. - Reparameterizing, fitting and testing a simple growth model. Trans. Am. Fish. Soc., 108 : 14-25.
- FREEDMAN, D. A., 1981. - Bootstrapping regression models. Ann. Statist. 9 (6) 1218-1228.
- HAMPEL F. R., 1973. - Robust estimation : a condensed partial survey. Z. Wahrscheinlichkeitsstheorie Werw. Geb., 27 : 87-104.
- HUBER P. J., 1972. 6 The 1972 Wald lecture. Robust statistics : a review. Ann. Math. Statist. 43 (4) : 1041-1067.
- JONES R., 1974. - Assessing the long term effects of changes in fishing effort and mesh size from legth composition data. Cons.int. Explor. Mer, C. M. 1974/F : 33, 13 p. (mimeo).

- JONES R., 1979. - An analysis of a Nephrops stock using length composition data. Rapp. P.V. Reun. Cons.int. Explor. Mer, 175, pp. 259-269.
- KENDALL M. & STUART A., 1979. - The advanced theory of Statistics. Vol. II : inference and relationship. Charles Griffin & Co., Ltd., London & High Wycombe, 4th edition, 748 p.
- KIMURA, D. K., 1980. - Likelihood methods for the Von Bertalanffy growth curve. Fish. Bull, 77 (4) : 765-776.
- LAUREC A., 1986. - Les méthodes delta en halieutique. Evaluation des sensibilités, approximation des biais et variances à l'aide de développements limités. Rapp. techn. IFREMER DRV/-86. 02/RH - Centre IFREMER de Nantes, 64 p.
- LAUREC A. et MESNIL B., 1985 a. - Analytical investigations of errors in mortality rate estimated from length distribution of catches. Document présenté à "International conference on the theory and application of length based methods of stock assessment". Mazzara del Vallo, Sicily, Italy (février, 1985).
- LAUREC A. et MESNIL B., 1985 b. - Rendement par recrue et analyse des cohortes de JONES. Etude de la sensibilité. Cons.int. Explor. Mer, C. M. 1985/G : 23 réf. K.
- MILLER R. G., 1974. - The Jackknife : a review. Biometrika 61 (1) : 1-15.
- SIMONOFF J. S. & TSAI C. L., 1986. - Jackknife - based estimators and confidence regions in non linear regression. Technometrics 28 (2) : 103-112.
- TOMLINSON P. K. and ABRAMSON N. J., 1961. - Fitting a Von Bertalanffy growth curve by least squares. Calif. Dpt. Fish and Game. Fish. Bull. 116, 19 p.
- VAUGHAN D. S. and KANCIRUK P., 1982. - An empirical comparison of estimation procedures for the Von Bertalanffy growth equation. J. Cons. int. Explor. Mer, 40 : 211-219.

Individus	Age	1	2	3	4	5	6	7
1		10.0	17.0	22.0	23.0	24.0	24.3	25.0
2		08.5	16.0	21.4	23.6	24.3	24.5	25.0
3		08.0	15.0	20.0	22.0	22.5	23.0	24.0
4		05.0	13.0	20.0	23.0	24.0	25.0	26.0
5		08.0	15.0	20.0	22.0	23.0	24.2	25.0
6		09.0	16.0	22.0	24.5	25.5	26.0	
7		08.0	15.0	22.0	24.5	26.0	27.0	
8		07.0	16.0	22.0	24.0	25.5	26.0	
9		08.0	16.0	22.0	24.2	25.4	26.0	
10		08.5	16.0	22.0	24.0	26.0	26.4	
11		08.0	16.0	21.4	23.0	24.0	24.8	
12		08.0	15.0	22.0	24.0	24.5	25.0	
13		08.5	17.0	23.0	25.0	26.0		
14		09.0	16.5	23.0	24.0	24.8		
15		10.0	18.0	24.0	25.0	25.4		
16		09.0	14.0	21.0	23.0	24.2		
17		08.5	17.5	23.5	26.0	27.0		
18		08.0	16.0	22.5	24.0			
19		09.5	17.8	23.0	25.0			
20		06.0	12.8	18.0	20.0			
21		07.5	16.0	23.0	25.0			
22		08.0	17.0	22.5	24.2			
23		09.0	18.0	23.0				
24		07.0	15.9	20.5				
25		07.5	15.0	21.0				
26		08.0	16.0	22.0				
27		07.8	12.0					
28		09.0	16.0					
29		09.0	14.0					
30		08.0	15.6					
31		09.0	17.0					
32		09.0						
33		06.8						
34		08.4						
35		07.0						

Tableau I : Hauteur aux âges.

Statistique centrale Pondération		Moyenne			Mode		Médiane	
Hauteur anneau		1	n_1	n_1/σ_1^2	1	n_1	1	n_1
1	Observée	8.05	8.05	8.05	7.63	7.63	7.8	7.8
	Prédite	7.67	7.59	7.86	7.18	7.08	7.55	7.49
2	Observée	15.1	15.1	15.1	14.8	14.8	16.3	16.3
	Prédite	16.33	16.26	16.9	16.28	16.21	16.96	16.96
3	Observée	21.8	21.8	21.8	21.8	21.8	21.8	21.8
	Prédite	21.08	21.03	21.52	21.16	21.12	21.99	22.01
4	Observée	23.9	23.9	23.9	24.8	24.8	25.55	25.55
	Prédite	23.67	23.65	23.90	23.77	23.75	24.68	24.71
5	Observée	25.7	25.7	25.7	25.3	25.3	26.55	26.55
	Prédite	25.09	25.10	25.11	25.17	25.17	26.12	26.15
6	Observée	26.16	26.16	26.16	26.16	26.16	27.17	27.17
	Prédite	25.87	25.90	25.73	25.92	25.94	26.88	26.92
7	Observée	25.3	25.3	25.3	25.3	25.3	26.3	26.3
	Prédite	26.30	26.33	26.05	26.32	26.37	27.29	27.32
L _b	Estimation	26.81	26.87	26.39	26.78	36.82	27.76	27.79
	V ₁	.81	.46	.10	1.01	.57	.52	.30
	V ₂	.67	.37	.08	.81	.45	.44	.25
L _J	Estimation	26.69	26.77	26.55	26.41	26.50	26.95	26.73
	Variance	.19	.24	.38	.22	.28	.98	.69
L _{NP}	Estimation	26.98	27.03		26.68	26.68	27.57	27.66
	Variance	.48	.45		.79	.72	.66	.55
L _P	Estimation	26.80	26.66		26.74	26.75	26.54	26.70
	Variance	.46	.54		1.01	2.07	6.9	5.9
K _b	Estimation	.60	.60	.67	.62	.62	.63	.63
	V ₁	.009	.006	.004	0.13	.008	.006	.004
	V ₂	.007	.004	.003	.009	.006	.005	.003
K _J	Estimation	.595	.585	.578	.64	.63	.86	.90
	Variance	.0037	.004	.014	.005	.006	.027	.017
K _{NP}	Estimation	.60	.60		.63	.63	.64	.63
	Variance	.0045	.0044		.009	.007	.007	.007
K _P	Estimation	.596	.61		.63	.62	.64	.63
	Variance	.0060	.005		.015	.009	.013	.012

Tableau II : Synthèse des résultats obtenus en excluant le rétro-calcul.

L_b = valeur de base obtenue par ajustement.

L_J = estimations par Jackknife

L_{NP} = estimation corrigée du biais estimé par Bootstrap non paramétrique

L_P = idem supra avec Bootstrap paramétrique

Les mêmes conventions s'appliquent à K pour définir par exemple K_b et K_J.

V₁ est la variance estimée par la formule 13.

V₂ correspond à la formule 12.

Statistique centrale Pondération		Moyenne			Mode		Médiane	
Hauteur anneau		1	n_1	n_1/σ_1^2	1	n_1	1	n_1
1	Observée	8.5	8.5	8.5	8.6	8.6	9.57	9.57
	Prédite	8.25	8.32	8.38	8.37	8.42	9.38	9.39
2	Observée	16.11	16.11	16.11	16.46	16.46	17.32	17.32
	Prédite	17.03	16.87	17.04	17.90	17.25	18.23	18.08
3	Observée	22.19	22.19	22.19	22.60	22.60	23.40	23.40
	Prédite	21.52	21.43	21.51	21.81	21.73	22.68	22.60
4	Observée	24.18	24.18	24.18	24.44	24.44	25.33	25.33
	Prédite	23.81	23.87	23.83	23.96	24.02	24.90	24.97
5	Observée	25.09	25.09	25.09	24.75	24.75	25.93	25.93
	Prédite	24.98	25.17	25.03	25.01	25.18	26.01	26.18
6	Observée	25.5	25.5	25.5	25.50	25.50	26.50	26.50
	Prédite	25.58	25.86	25.65	25.53	25.77	26.56	26.82
7	Observée	25.5	25.5	25.5	25.50	25.50	26.50	26.50
	Prédite	25.89	26.24	25.97	25.78	26.07	26.83	27.15
L _b	Estimation	26.21	26.67	26.31	26.02	26.38	27.11	27.50
	V ₁	.29	.49	.07	.29	.50	.28	.48
	V ₂	.25	.42	.06	.26	.44	.25	.41
L _J	Estimation	26.20	26.60	26.38	25.97	26.21	29.50	31.47
	Variance	.11	.17	.21	.98	1.3	.32	.79
L _{np}	Estimation							
	Variance							
L _p	Estimation							
	Variance							
K _b	Estimation	.67	.63	.66	.72	.68	.69	.65
	V ₁	.0055	.0044	.0009	.007	.006	.006	.005
	V ₂	.0045	.0036	.0008	.006	.005	.005	.004
K _J	Estimation	.67	.63	.2	.71	.68	.60	.43
	Variance	.0010	.0009	.0016	.006	.007	.0013	.003
K _{np}	Estimation							
	Variance							
K _p	Estimation							
	Variance							

Tableau III : Résultats obtenus en incorporant les données issues du rétrocalcul.

L_b = valeur de base obtenue par ajustement.

L_J = estimations par Jackknife

L_{np} = estimation corrigée du biais estimé par Bootstrap non paramétrique

L_p = idem supra avec Bootstrap paramétrique

Les mêmes conventions s'appliquent à K pour définir par exemple K_b et K_J.

V₁ est la variance estimée par la formule 13.

V₂ correspond à la formule 12.

Statistique centrale Pondération		Moyenne			Mode		Médiane	
		1	n_i	n_i/σ_i^2	1	n_i	1	n_i
Sans rétrocalcul								
L_{∞}	Non perturbé	26.8	26.9	26.6	26.4	26.5	26.9	26.7
	Perturbé	26.7	26.80	26.6	26.7	26.7	27.5	25.8
K	Non perturbé	.60	.60	.67	.62	.62	.62	.63
	Perturbé	.57	.57	.62	.61	.61	.65	1.22
Avec rétrocalcul								
L_{∞}	Non perturbé	26.2	26.7	26.3	26.0	26.4	27.1	27.5
	Perturbé	26.5	26.8	26.7	26.00	26.3	27.2	27.5
K	Non perturbé	.67	.63	.66	.72	.68	.69	.65
	Perturbé	.58	.54	.56	.70	.67	.63	.60

Tableau IV : Influence d'une erreur de lecture de l'âge.