

# ECHANTILLONNAGE DES DÉBARQUEMENTS EN CRIÉE POUR L'ÉTUDE DES STRUCTURES DÉMOGRAPHIQUES

## Sources de variabilité et optimisation du plan de sondage

**Robert CHEVALIER**

IFREMER Centre de Nantes.  
Départ. Evaluation des Ressources halieutiques.  
B.P. 1049 - 44037 Nantes Cedex 01, France.

### *Abstract*

SAMPLING OF COMMERCIAL LANDINGS FOR ESTIMATION OF THE AGE STRUCTURES. SOURCES OF VARIABILITY AND OPTIMIZATION OF SAMPLING DESIGN.

In a previous report we have analyzed problems encountered to select a sampling design for estimating the age structures of landings. In this one we study the effects of various factors of variability on the optimal values of the sampling rates. The method used for this analysis could be employed to define an optimal sampling strategy to assess the age structures of landings from fish markets.

### *Résumé*

Dans un précédent rapport nous avons répertorié les contraintes auxquelles sont soumises les opérations d'échantillonnage en criée et analysé les diverses options qui se présentent lors de la mise au point du plan d'échantillonnage et du choix des estimateurs. Prolongeant cette analyse nous présentons une étude sur l'optimisation du plan type qui met en évidence l'impact de différentes sources de variabilité sur les valeurs optimales des taux d'échantillonnage. Moyennant quelques apports complémentaires, la démarche pourrait être utilisée pour préparer les plans de sondage en criée.

La mise au point d'un plan de sondage exige des informations sur au moins trois points essentiels (COCHRAN, 1981) : Quelle précision peut-on espérer pour une certaine quantité d'effort (coût) disponible ? Quelle est la meilleure voie pour améliorer la précision des résultats ? Si la précision qu'il est permis d'espérer est trop faible, de combien doit-on augmenter l'effort pour obtenir une précision jugée satisfaisante ?

C'est le problème de l'optimisation du plan de sondage qui est ainsi posé, voire son utilité dans la mesure où la précision souhaitée entraîne des coûts jugés prohibitifs. Toutefois, une réponse précise à ces questions ne pourra être fournie que par rapport au produit final obtenu qui a logiquement motivé l'étude du plan. Dans le domaine qui nous intéresse, les mensurations réalisées sur les espèces mises à terre et les déterminations d'âge subséquentes permettent une estimation des structures démographiques des débarquements. L'information potentielle ainsi recueillie sera pleinement exploitée par l'utilisation de modèles. L'expression mathématique de ces modèles, que nous qualifierons pour clarifier notre exposé de grandeur objectif, permettra d'obtenir des résultats chiffrés dont la valeur dépendra essentiellement de la qualité des données recueillies dans la mesure où l'on suppose une adéquation parfaite entre le choix du modèle et la réponse attendue. Le plus souvent, ce modèle fera intervenir plusieurs paramètres dont il conviendra d'estimer la valeur. La qualité du produit final dépendra des incertitudes liées à toutes ces estimations.

L'étude que nous présentons porte essentiellement sur l'estimation des structures démographiques. Nous nous proposons d'analyser dans quelle mesure le choix d'une grandeur objectif peut modifier le plan d'échantillonnage optimal. Plus concrètement, il s'agit d'évaluer comment, pour un plan d'échantillonnage optimal donné et une précision souhaitée, varient les valeurs optimales des taux d'échantillonnage entraînant un coût minimal de l'opération. Cela nous amènera à préciser quelques points déjà évoqués dans une précédente étude (CHEVALIER et MESNIL, 1986) sur le plan de sondage. Par ailleurs, les grandeurs objectives que l'on pourra définir seront très différentes selon le but poursuivi et le nombre de paramètres retenus. Toutefois, par rapport à la seule variable prise en compte ici, la structure démographique, elles peuvent être ramenées à deux types : fonctions linéaires ou non linéaires. Pour des grandeurs objectives appartenant à ces deux types, nous évaluerons sur des exemples les fluctuations des valeurs optimales d'échantillonnage et les coûts relatifs du sondage.

### **Présentation du problème.**

#### **Le plan de sondage.**

Dans le rapport cité plus haut, une première analyse a permis de préciser les contraintes à prendre en compte pour mettre au point une stratégie d'échantillonnage. Ceci nous a conduit à proposer un plan de sondage qui tienne compte d'une subdivision spatio-temporelle imposée par la nature du problème : dispersion des débarquements et observations continues dans le temps. A l'intérieur d'une subdivision, le schéma classique qui paraît s'imposer est un plan à deux degrés avec stratification des unités primaires (jours de vente ou bateaux) par catégories commerciales. Il sera défini par les caractéristiques suivantes : tirages équiprobables, sans remise, des unités primaires et des unités secondaires (poissons).

Des variantes peuvent intervenir suivant que les proportions d'âge dans les captures sont évaluées directement sur un échantillonnage aléatoire simple (E.A.S.) des unités du deuxième niveau ou que l'on utilise l'information supplémentaire fournie par les distributions de longueur, en appliquant les techniques du double échantillonnage.

D'autre part, dans la mesure où l'on est confronté à une technique de sous-échantillonnage, on peut envisager la mise en œuvre, pour l'estimation des fréquences absolues  $C_a$ , ou relatives  $P_a$ , d'un des trois estimateurs suivants :

- estimateur rapport  $\hat{P}_a = \frac{\sum T_i \hat{P}_{ai}}{\sum T_i}$
- estimateur simple  $\hat{P}_a = \frac{\sum \hat{P}_{ai}}{n}$
- estimateur pondéré  $\hat{P}_a = \frac{1}{n} \frac{\sum T_i \hat{P}_{ai}}{\bar{T}}$

$T_i$  et  $\bar{T}$  représentent l'effectif de l'unité  $i$  et l'effectif moyen pour une période déterminée.

Selon la structure de la population, l'un ou l'autre de ces estimateurs se révèlera plus précis ; l'estimateur simple correspond à un certain type d'utilisation de la clé taille-âge.

L'analyse de la sensibilité du plan de sondage au choix de la fonction objectif sera faite par strate spatio-temporelle afin de limiter le nombre de paramètres à estimer dans une même opération. Eventuellement, la répétition de la procédure pour chacune des strates spatio-temporelles fournira des informations supplémentaires quant à l'influence de la structure des populations échantillonnées. Si l'on considère l'optimalité du plan de sondage dans son ensemble, ce fractionnement doit théoriquement conduire à une solution sub-optimale dont on peut supposer qu'elle devrait être suffisante compte tenu des adaptations nécessaires des résultats théoriques à la réalité.

#### **Choix d'une grandeur objectif.**

La précision de l'estimation d'une structure démographique n'est pas un concept simple à définir dans la mesure où elle est la résultante de la précision de ses diverses composantes. De plus, elle n'a qu'une valeur relative. Nous sommes, en effet, plus intéressés à connaître quels vont être les effets des erreurs d'échantillonnage des captures aux âges sur les évaluations subséquentes qui pourront être réalisées.

Pour atteindre l'efficacité maximale, en terme de coût, du plan de sondage, nous ne disposons pas d'une estimation unique, moyenne ou total des réalisations d'une variable donnée, mais d'un vecteur aléa-

toire dont le poids relatif des composantes pourra fluctuer suivant le type d'évaluation ou de prévision effectué. Or, il s'avère que pour un niveau d'échantillonnage donné, la précision des estimations des effectifs diffère sensiblement suivant les âges. Il s'ensuit que pour une précision globale donnée, les coûts pourront varier assez fortement selon l'objectif. Autrement dit, le coût optimal variera en fonction de l'objectif.

Des solutions de compromis peuvent être envisagées.

YATES (1981) a proposé des méthodes d'optimisation pour des plans de sondage ayant pour objectif l'estimation de plusieurs caractères. Le principe de ces méthodes repose sur la notion de pertes. En effet, les estimations qui sont faites sont entachées d'erreurs d'échantillonnage. Les décisions pouvant être prises au vu de ces estimations peuvent donc entraîner des pertes plus ou moins importantes liées notamment au poids de l'estimation dans la décision et à l'importance de l'erreur qui y est attachée. Si l'on désigne par  $P_i$  et  $\hat{P}_i$ , respectivement la valeur vraie, inconnue, de la variable  $i$  et son estimation, on définira une fonction de pertes  $\pi$  des grandeurs  $(\hat{P}_i - P_i)$ , l'estimation  $\hat{P}_i$  étant issue d'un plan de sondage quelconque. Dans les procédures d'optimisation, on utilisera la fonction de pertes moyennes, prise sur tous les échantillons possibles :

$$\Phi(P_1, \dots, P_i, \dots) = E \{ \Pi [(\hat{P}_1 - P_1), \dots, (\hat{P}_i - P_i), \dots] \}$$

Le choix de la fonction de pertes est plus ou moins arbitraire. On utilise fréquemment la fonction quadratique :

$$\Pi(P_1, \dots, P_i, \dots) = \sum a_i (\hat{P}_i - P_i)^2$$

La fonction moyenne de pertes est dans ce cas une fonction linéaire des variances des diverses estimations  $\hat{P}_i$  :

$$\Phi = \sum_i a_i \text{Var}(\hat{P}_i)$$

Les valeurs des taux d'échantillonnage sont explicités dans l'expression de  $\text{Var}(\hat{P}_i)$ . Les coefficients  $a_i$  peuvent être choisis arbitrairement ou découler de la structure de la grandeur objectif choisie. La recherche des valeurs optimales des taux d'échantillonnage consistera à minimiser la fonction moyenne de pertes pour un coût donné.

Si les coefficients  $a_i$  sont déterminés, on retrouve un problème classique d'optimisation. Un exemple est fourni par KIMURA (1977) qui étudie les mérites comparés, dans l'hypothèse d'un échantillonnage en deux phases, des taux de sous-échantillonnage fixe ou proportionnel pour la détermination de l'âge. Il utilise, en effet, comme fonction moyenne de pertes la relation :

$$\Phi = \sum_i a_i \text{Var}(\hat{P}_i) \quad a_i = 1 \text{ pour tout } i$$

Ce système de coefficients, ni plus ni moins arbitraire qu'un autre, lui a permis d'apporter une réponse claire à un problème précis. Mais, il n'est pas assuré que cet ensemble de valeurs soit le meilleur possible pour une série de grandeurs objectifs fréquemment utilisées en dynamique de populations.

Si les coefficients  $a_i$  ne sont pas connus, des conditions supplémentaires doivent être fournies. Une méthode proposée par YATES (1981) consiste à fixer les valeurs de  $\text{Var}(\hat{P}_i)$ . La procédure d'optimisation n'est pas immédiate. En effet, dans une première étape, on sera amené à déterminer les valeurs des  $a_i$  avec un système d'équations pas toujours facile à résoudre. Ce peut-être un handicap. Mais l'inadéquation de cette approche à nos problèmes paraît d'ordre théorique. En effet, la solution précédente, et celles qui ont été développées à la suite de YATES par d'autres auteurs s'appliquent dans le cas où plusieurs critères indépendants ont été observés au cours du sondage. Dans la mesure où les variables représentant ces critères seront utilisées pour des études différentes, il n'est pas illogique de fixer pour chacune d'elles un niveau d'erreur selon l'importance qu'on lui accorde.

Dans notre étude, nous sommes intéressés par un seul caractère de la population, la structure démographique, représentée par une variable vectorielle dont les éléments, les captures aux âges, interviennent avec probablement des poids différents pour une même application. Fixer la valeur des variances de ces différents composants reviendrait à choisir une fonction moyenne de pertes plus ou moins arbitraire et donc finalement à se donner une grandeur objectif elle-même arbitraire.

En effet, si l'on se donne une grandeur objectif quelconque  $f = f(\hat{C})$ , la procédure d'optimisation consistera à minimiser la variance de  $f$  pour un coût global fixé. Ceci revient à choisir une fonction moyenne de pertes :

$$\Phi(\hat{C}) = \sum_i \sum_j b_i b_j \text{cov}(\hat{C}_i, \hat{C}_j) \quad (1)$$

relation exacte si  $f$  est une fonction linéaire du vecteur capture  $\hat{C}$  ou approchée (développement en série de Taylor) dans le cas contraire.

Nous retiendrons une fonction de chaque type pour apprécier l'influence du choix de la grandeur objectif sur les valeurs optimales des taux d'échantillonnage et sur les coûts globaux.

**Grandeur objectif liée à l'analyse des cohortes : estimation d'un F moyen.**

Les valeurs de F, taux instantané de mortalité par pêche, sont estimées par les techniques d'analyse de cohortes. Nous avons choisi le modèle simplifié de POPE (1972) pour lequel les estimations en début d'année de l'effectif d'une cohorte, dépendent uniquement de l'estimation des captures subséquentes. Les coefficients du modèle sont le taux instantané de mortalité naturelle M, et le taux d'exploitation dans le dernier groupe d'âge. Rappelons les deux formules de base du modèle :

$$N_j = N_{j+1} \exp(M) + C_j \exp(M/2)$$

$$\exp(F_j) = N_j \exp(-M)/N_{j+1}$$

L'estimation des effectifs est une fonction linéaire des captures. Par contre, l'estimateur des F et donc d'un F moyen, ne jouit pas de cette propriété.

La non linéarité entraîne deux conséquences.

- L'existence de biais sur les estimateurs de F. En effet,  $\hat{\theta}$  étant l'estimateur d'une variable aléatoire quelconque, on a la relation :

$$E[f(\hat{\theta})] \neq f[E(\hat{\theta})]$$

E symbolisant la valeur moyenne de la fonction ou de la variable. Si f est une fonction convexe :

$$E[f(\hat{\theta})] \geq f[E(\hat{\theta})]$$

relation connue sous le nom d'inégalité de Jensen (RAO, 1973).  $E[f(\hat{\theta})]$  sera d'autant plus proche de la valeur exacte  $f[E(\hat{\theta})]$  que l'estimation  $\hat{\theta}$  sera plus précise.

- La relation (1) utilisée pour calculer la variance de  $f(\hat{C})$  ne donne qu'une estimation approchée de la vraie variance. Dans les deux cas, il conviendra d'évaluer les biais résultants.

**Fonction linéaire des captures.**

Un objectif qui présente un intérêt certain est la prévision des captures à court terme. Désignant par  $\hat{C}$ , la prévision des captures en nombre, on aura la fonction d'estimation :

$$\hat{Y} = \sum_{i=1}^l \hat{C}_i \omega_i \quad \omega_i \text{ désignant le poids moyen à l'âge } i.$$

Si on utilise le modèle simplifié de Pope, sous l'hypothèse d'équilibre de la pêcherie, on pourra écrire :

$$\hat{C}_i = \hat{N}_i \exp(-M/2) - \hat{N}_{i+1} \exp(M/2)$$

Les  $\hat{C}_i$  apparaissent comme combinaisons linéaires des  $\hat{C}_j$  ( $j = 1, l$ ).

En fait, une estimation des captures en période d'équilibre peut être approchée par la relation :

$$\hat{Y} = \sum_{i=1}^l \hat{C}_i \omega_i$$

qui permet une évaluation sélective de l'impact des erreurs d'échantillonnage des captures aux âges sur la prévision des captures.

Remarquons que ce qui précède revient à écrire qu'en situation d'équilibre une estimation raisonnable de prévision des captures est fournie par les captures observées les années précédentes.

**Hypothèses et méthodes.**

**Principes généraux.**

Nous considérons une pêcherie en situation d'équilibre pour laquelle les débarquements sont assimilés aux captures. La population cible, supposée connue, est créée à partir d'une distribution moyenne de longueur évaluée pour une période donnée et d'une clé taille-âge déterminée sur la même période. Nous avons rappelé plus haut le plan de sondage retenu et supposé appliqué dans une strate spatio-temporelle. Le contenu des différentes grappes ou unités primaires de la strate est obtenu à partir de tirages aléatoires effectués sur la distribution moyenne prise comme référence. Le plan pourra être simplifié pour illustrer certains points particuliers et entre autres pour justifier certains choix concernant la stratégie ou les estimateurs retenus pour cette analyse.

Les données utilisées portent sur des débarquements de lieu noir (*Pollachius virens*) en criée de Lorient au cours de l'année 1986. Les poissons sont répartis en 5 catégories commerciales dont les tailles limites peuvent fluctuer d'un trimestre à l'autre. Tenant compte de ces fluctuations, nous avons fixé les tailles inférieures de chaque catégorie à : 37 cm - 55 cm - 67 cm - 82 cm - 97 cm. D'autre part, les effectifs initiaux

sont reportés en classes de taille de 1 cm d'amplitude. Nous les avons regroupés en classes de 3 cm. En effet, à l'encontre de ce qui a pu être vérifié lors de traitements (analyse de Jones) appliqués sur des distributions de longueur (LAUREC et MESNIL, 1985a), il ne semble pas nécessaire de conserver un découpage aussi fin pour des données destinées à être redistribuées en classes d'âge.

Pour l'évaluation des coûts, nous avons utilisé les fonctions qui sont reprises dans tous les ouvrages classiques. A partir d'informations fournies par le laboratoire IFREMER, nous avons évalué des coûts moyens pour l'accès à une unité primaire ( $C_p$ ) et les coûts moyens de mesurage ( $C_1$ ) et d'ageage ( $C_2$ ) supposés constants quelle que soit la taille (âge) du poisson.

La précision des résultats est exprimée en valeur relative (coefficient de variation) et nous avons utilisé les fréquences relatives  $\hat{P}_i$  des captures. Pour le calcul de la variance des  $\hat{P}_i$  et des taux d'échantillonnage optimaux, nous avons utilisé les formules classiques du sous-échantillonnage avec stratification des unités primaires (JESSEN, 1978). Mais dans l'expression de la variance de  $f(\hat{p})$  peuvent intervenir des termes de covariance entre les composantes  $\hat{P}_i$  du vecteur  $\hat{P}$ . L'estimation d'un F moyen sur une cohorte isolée fait intervenir des estimations  $P_i$  réalisées au cours d'années successives. Ces estimations peuvent être considérées comme indépendantes et donc  $cov(\hat{P}_i, \hat{P}_j) = 0$ . Il n'en est pas de même pour la deuxième grandeur objectif proposée. Toutefois, pour assurer la cohérence de l'analyse, nous négligerons les termes de covariance. Nous verrons cependant que leur inclusion peut sensiblement améliorer la précision de la fonction  $f(\hat{p})$ , les termes de covariance étant tous négatifs.

**Choix particuliers.**

*Echantillonnage en deux phases et catégories commerciales.*

Un échantillon relativement important de  $n_1$  poissons est sélectionné au cours d'une opération. Les individus sont reportés en K strates de longueur de poids relatif  $w_k$ . Un échantillon aléatoire de  $n_{2k}$  éléments est tiré dans chaque strate pour détermination de l'âge. Pour un âge donné, une estimation non biaisée de la proportion d'individus de cet âge dans la population échantillonnée est donnée par la relation :

$$\hat{p} = \sum_k w_k \hat{p}_k \tag{2}$$

dans laquelle  $\hat{p}_k$  représente la proportion d'individus d'âge dans la strate k.

Deux stratégies d'allocation de  $n_{2k}$  peuvent être envisagées :

- l'allocation optimale, dite de Neyman avec

$$n_{2k} = n_2 \frac{W_k S_k}{\sum W_k S_k} \tag{3}$$

$n_2 = \sum n_{2k}$ , conduisant à l'expression suivante pour la variance, les termes de correction pour population finie étant négligés :

$$\text{Var (opt)} = \frac{(\sum W_k S_k)^2}{n_2} + \frac{\sum W_k (P_k - P)^2}{n_1} \tag{4}$$

- l'allocation proportionnelle pour laquelle  $n_{2k} = \lambda n_{1k}$  et

$$\text{Var (p)} = \frac{\sum W_k P_k Q_k}{n_2} + \frac{\sum W_k (P_k - P)^2}{n_1} \tag{COCHRAN, 1977} \tag{5}$$

$$n_2 = \lambda n_1$$

On vérifie que :

$$\text{Var (prop)} - \text{Var (opt)} = \frac{\sum W_k (S_k - \bar{S})^2}{n_2}$$

avec  $\bar{S} = \sum W_k S_k$  et  $S_k^2 = P_k (1 - P_k)$ .

Sauf exception la stratégie optimale pour une précision donnée doit théoriquement se révéler plus efficace. Un exemple est fourni tableau 2 pour les deux populations témoins que nous avons retenues et l'estimateur  $f = \sum \omega_i P_i$ . On trouvera également dans ce tableau les résultats relatifs à un échantillonnage aléatoire simple. Dans la mesure où il n'existe pas de stratification par catégorie commerciale, on peut donc s'attendre à des différences sensibles entre les coûts des trois stratégies.

Toutefois, l'allocation de Neyman n'est pas sans présenter quelque faiblesse quand les calculs sont faits à partir d'échantillons, à l'imprécision due aux erreurs d'échantillonnage s'ajoute le fait que l'allocation optimale dépend du poids relatif des classes de taille qui peut évoluer dans le temps. Les gains d'efficacité peuvent être partiellement illusoire. Aussi, on peut s'interroger sur le réel intérêt de cette stratégie quand les débarquements sont stratifiés en catégories commerciales et donc qu'il existe déjà une partition de la

population échantillonnée par classes de longueur. En effet une part des gains d'efficacité imputables à l'allocation de Neyman est déjà prise en compte par la stratification.

Age	Poids moyens	Distribution 1			Distribution 2		
		Fréquence des captures	Effectifs relatifs	F	Fréquences des captures	Effectifs relatifs	F
1	0,480	18,790	0,265 93.10	0,081	1,932	0,169 82.10	0,013
2	0,921	11,771	0,200 72.10	0,067	45,580	0,137 29.10	0,457
3	1,330	16,768	0,153 69.10	0,128	27,342	0,711 59.10	0,553
4	1,880	23,008	0,110 66.10	0,261	11,910	0,335 19	0,499
5	2,775	14,350	0,697 79	0,258	8,578	0,166 67	0,841
6	3,885	5,036	0,441 46	0,135	3,419	0,588 40.10 <sup>-1</sup>	1,027
7	5,000	3,543	0,315 86	0,132	0,781	0,172 42.10 <sup>-1</sup>	0,694
8	5,943	1,121	0,226 55	0,056	0,334	0,704 95.10 <sup>-2</sup>	0,743
9	6,950	0,962	0,175 34	0,063	0,050	0,274 59.10 <sup>-2</sup>	0,227
10	7,650	0,084	0,134 85	0,007	0,024	0,179 17.10 <sup>-2</sup>	0,163
11	8,270	0,653	0,109 65	0,068	0,004	0,124 58.10 <sup>-2</sup>	0,033
12	9,361	0,877	0,838 64.10 <sup>-1</sup>	0,123	0,009	0,986 86.10 <sup>-3</sup>	0,112
13	10,194	3,036 <sup>o</sup>	0,607 25.10 <sup>-1</sup>	0,200	0,012	0,722 19.10 <sup>-3</sup>	0,196
14*	11,329				0,024*	0,485 84.10 <sup>-3</sup>	0,200

TABL. 1. — Principales caractéristiques des deux populations témoins utilisées dans l'étude et résultats de l'analyse de cohortes. (\* ) 14 ans et + ; (°) 13 ans et +.

Dans le tableau 3 sont reportés pour les deux populations témoins préalablement stratifiées en catégories commerciales, les efforts et coûts afférents qu'il serait nécessaire d'engager pour obtenir la même précision de la grandeur objectif  $f = \sum \omega_i P_i$ , selon le plan de sondage mis en œuvre. Les différents plans résultent de combinaisons de stratégies d'échantillonnage intra et inter-strates.

Le tableau 4 contient des résultats intermédiaires : poids relatifs des strates, proportions d'individus âgés par strate. La prise en compte de la stratification en catégories commerciales dans le plan de sondage réduit sensiblement le niveau des coûts même si les écarts restent importants entre les différents plans testés.

On peut noter qu'à partir d'une stratégie combinant une allocation proportionnelle par strate pour la détermination de l'âge et une répartition optimale de l'effort entre les strates, on obtient des résultats équivalents à ceux affichés dans le tableau 2 pour une allocation optimale du nombre d'individus âgés. Ce résultat est sans doute fortuit. Mais, si on tient compte également de la sensibilité des résultats à la structure des populations échantillonnées, les observations faites plus haut sur les gains réels d'efficacité induits par l'allocation de Neyman paraissent justifiées dans le cas d'une population répartie en catégories commerciales.

Type d'échantillonnage	Taux d'échantillonnage		$\lambda = n_2/n_1$	Coût
	$n_1$	$n_2$		
Distribution 1. — Statistiques de la fonction objectif : variance = 0,048 87    moyenne = 2,210 72 coefficient de variation = 0,10				
E.A.S.	178	178	1,000	3 243
Proportionnel	839	102	0,122	2 052
Optimal	641	58	0,090	1 190
Distribution 2. — Statistiques de la fonction objectif : variance = 0,021 23    moyenne = 1,456 90 coefficient de variation = 0,10				
E.A.S.	115	115	1,000	2 101
Proportionnel	524	76	0,145	1 496
Optimal	437	52	0,119	1 043

TABL. 2. — Efforts et coût comparés, pour deux populations différentes et les plans de sondage : aléatoire simple ; 2 phases, allocation proportionnelle ; 2 phases, allocation optimale.  
Grandeur objectif :  $f = \sum \omega_i P_i$   
Coûts unitaires :  $C_1 = 0,25$  ;  $C_2 = 18$  ;  $C_p = 120$ .

Aussi, nous avons utilisé dans notre analyse l'allocation proportionnelle des individus âgés, stratégie qui se révèle en général plus efficace qu'une allocation fixe non optimale (KIMURA, 1977). Par contre, la répartition des efforts entre les catégories commerciales est théoriquement optimale.

Clé taille-âge et sous-échantillonnage.

Comme nous l'avons mentionné plus haut, l'échantillonnage des débarquements fera intervenir le plus souvent des plans de sondage à deux degrés, les jours de vente représentant les unités primaires.

Plusieurs estimateurs de la fréquence des individus d'un âge donné  $t$  peuvent être envisagés. Aucun d'entre eux ne s'impose *a priori*. Le choix dépendra de la nature des informations disponibles et des caractéristiques de la population échantillonnée. Toutefois, nous nous sommes servi pour notre étude de l'estima-

Plans de sondage		Taux d'échantillonnage			
Stratégie à l'intérieur d'une strate	Type d'allocations entre les strates	$n_1$	$n_2$	$\lambda = n_2/n_1$	Coût
		E.A.S. 2 phases	proportionnelle optimale	130 73	
Allocation proportionnelle	proportionnelle optimale	445 275	99 59	0,222 0,214	1 887 1 128
E.A.S. 2 phases	proportionnelle optimale	86 60	86 60	1 1	1 573 1 100
	Allocation proportionnelle	254 202	73 52	0,287 0,257	1 381 987

*Distribution 1*  
 statistiques de la grandeur objectif  
 variance = 0,048 87  
 moyenne = 2,210.72  
 coefficient de variation = 0,10

*Distribution 2*  
 statistiques de la grandeur objectif  
 variance = 0,021 23  
 moyenne = 1,456 89  
 coefficient de variation = 0,10

TABL. 3. — Efforts moyens et coûts comparés pour deux distributions témoins, stratifiées en catégories commerciales et les plans de sondage par strate : E.A.S. ; 2 phases, allocations proportionnelles. Estimateur :  $f = \sum \omega_i P_i$ .

ECHANTILLONNAGE INTRA-STRATE							
E.A.S.					2 phases allocation proportionnelle		
Catégories	$\lambda = n_2/n_1$	Poids relatif des catégories		$\lambda = n_2/n_1$	Poids relatif des catégories		
		Proportionnel	Optimal		Proportionnel	Optimal	
Distribution 1	1	1	0,23	0,05	0,173	0,23	0,05
	2	1	0,26	0,17	0,250	0,26	0,17
	3	1	0,33	0,31	0,273	0,33	0,32
	4	1	0,12	0,22	0,233	0,12	0,22
	5	1	0,07	0,26	0,149	0,07	0,23
Distribution 2	1	1	0,53	0,28	0,405	0,53	0,29
	2	1	0,35	0,43	0,205	0,35	0,41
	3	1	0,11	0,23	0,262	0,11	0,23
	4	1	0,02	0,06	0,255	0,02	0,06
	5	1	0,00	0,01	0,148	0,00	0,01

TABL. 4. — Poids relatifs des différentes strates et valeur du rapport : nombre d'individus âgés sur nombre d'individus mesurés.

teur simple, bien qu'il puisse fournir des résultats biaisés. Ce choix découle de deux observations. En premier lieu, c'est celui qui correspond le mieux à la méthode d'estimation des structures démographiques largement utilisée actuellement, basée sur l'utilisation d'une clé taille-âge (CHEVALIER et MESNIL, 1986). Cette dernière

provient du cumul non pondéré d'observations faites sur différentes unités primaires. Si la structure démographique varie fortement entre les unités primaires, les estimations seront entachées de biais. C'est du même défaut que souffre l'estimateur simple. Mais le biais sera faible et l'estimateur performant si la proportion  $P_i$  d'individus d'âge  $t$  pour l'unité  $i$  est indépendante de l'effectif  $T_i$  de l'unité  $i$ , et donc si la structure démographique est stable. Dans le cas contraire, l'estimateur simple sera biaisée et en ce qui concerne la clé taille-âge, on se trouve dans une situation dénoncée par KIMURA (1977) où la clé est utilisée pour estimer la structure démographique d'une population qui serait en réalité différente de celle dont elle est issue.

Or, compte tenu de ce qui précède et si l'on admet que pour un intervalle de temps assez court les proportions d'âge par classes de longueur restent stables dans les débarquements, les fluctuations proviendront essentiellement des variations des poids relatifs des classes de longueur. La stratification en catégories commerciales est de nature à réduire sensiblement l'amplitude de ces variations.

Par ailleurs, la méthode que nous avons utilisée pour reconstituer les unités primaires n'introduit que des variations aléatoires et donc conserve la structure sous-jacente de la population de référence. Dans ces conditions, les biais induits par l'estimateur simple devraient être peu importants. C'est ce que nous avons pu vérifier sur les simulations réalisées. Ils restent inférieurs à 4 % sur les estimations des proportions d'âge, ce qui entraîne des écarts négligeables sur les grandeurs objectifs.

### Analyse des résultats.

#### Grandeur objectif non linéaire.

Avant d'examiner les résultats obtenus sur les diverses situations testées, il nous paraît utile d'évaluer les erreurs induites sur l'estimation du  $F$  moyen, fonction non linéaire des captures, et de sa variance quand on utilise une approximation linéaire par les développements en séries de Taylor.

Le modèle d'analyse de cohortes de Pope fournit les relations :  $\bar{F} = M + [L_n(\hat{N}_s) - L_n(\hat{N}_{s+r})] \frac{1}{r}$

$$\text{Var}(\hat{\bar{F}}) = \frac{1}{r^2} \left[ \frac{\text{Var}(\hat{N}_s)}{N_s^2} + \frac{\text{Var}(\hat{N}_{s+r})}{N_{s+r}^2} - \frac{2 \text{Cov}(\hat{N}_s, \hat{N}_{s+r})}{N_s \times N_{s+r}} \right]$$

$r$  représentant le nombre de composantes  $F_i$  incluses dans la moyenne et  $s$  le rang de la première valeur retenue.

Distribution	1		2	
	$n = 300$	$n = 500$	$n = 300$	$n = 500$
Valeur vraie $\bar{F}$ $E(\hat{\bar{F}}) = \bar{F}$	0,155 1	0,157 1	0,564 9	0,564 9
	0,158 1	0,157 6	0,595 2	0,587 7
Biais	0,003 0	0,002 5	0,030 3	0,022 8
S.C.E.M. vraie S.C.E.M. calculée	0,000 635	0,000 388	0,012 03	0,007 76
	0,000 547	0,000 329	0,010 69	0,006 41
Biais	0,000 088	0,000 059	0,001 34	0,001 35

TABLEAU 5. — Valeurs des biais sur les estimateurs de  $\bar{F}_{16}$ , moyennes sur les groupes d'âges de 1 à 6 ans.

Soit  $\bar{F}$  la valeur exacte dont on cherche une estimation, il s'agit de calculer les écarts :

$$E(\hat{\bar{F}}) - \bar{F} \qquad \text{Var}(\hat{\bar{F}}) - E(\hat{\bar{F}} - \bar{F})^2$$

Les calculs ont été faits sous l'hypothèse d'un sondage aléatoire simple. Les valeurs exactes, supposées connues, sont extraites du tableau 1. Mille échantillons de 300 et 500 individus ont été simulés à partir

de chacune des deux distributions de référence. Pour chacun des échantillons, on obtient une estimation du  $\bar{F}$  moyen dont la moyenne  $\hat{\bar{F}}$  sur le total des échantillons d'une même série peut être considérée comme une mesure de  $E(\hat{\bar{F}})$  :  $\hat{\bar{F}} = E(\hat{\bar{F}})$ .

Pour évaluer les biais sur la variance, on utilise la relation :

$$E(\hat{\bar{F}} - \bar{F})^2 = \frac{\sum (\hat{\bar{F}} - \bar{F})^2}{N}, \text{ pour } N = 1000$$

Les résultats sont reportés dans le tableau 5. Pour les 4 séries simulées, les biais sur les estimateurs surestiment les vraies valeurs de 1 à 5 %. Pour les variances, les écarts sont plus importants. On obtient une sous-estimation de 11 à 17 %. Précisons que ces écarts représentent bien la valeur de l'erreur induite quand on se limite au terme d'ordre 1 du développement en série de Taylor de la fonction  $\bar{F}$  car nous avons calculé  $\text{Var}(\hat{\bar{F}})$  par rapport aux valeurs exactes  $\bar{F}$ . Il s'agit en fait, en terminologie statistique, de l'écart quadratique moyen calculé à partir de la somme des carrés des écarts à la moyenne (S.C.E.M.).

Les valeurs des coefficients de variation, que nous avons utilisés comme critère de précision, sont sous-estimées, l'écart moyen étant légèrement supérieur à 1 % :

valeurs exactes :	0,163	0,127	0,194	0,156
valeurs calculées :	0,151	0,117	0,183	0,142

## Résultats-discussion.

### Commentaires généraux.

Les résultats des simulations sont rassemblés dans les tableaux 6 et 7. Les facteurs ayant une influence marquée sur la précision des estimations et donc sur les coûts ont été individualisés. La valeur du coefficient de variation de la grandeur objectif est fixée à 0,10. Les symboles  $q$  et  $q_h$  représentent respectivement le nombre total d'unités primaires à échantillonner pour l'ensemble des catégories commerciales et le poids relatif de la catégorie commerciale  $h$ . Les quatre facteurs mis en évidence ne jouent pas des rôles symétriques dans le processus de choix du plan optimal.

- *Structure de la population cible.* — Représentée par le vecteur capture dont on cherche une estimation, objectif immédiat du plan de sondage, on peut noter que cette structure a une influence non négligeable sur le coût du plan optimal. La structure évoluant dans le temps, il y a lieu d'être attentif à la définition des strates temporelles en tenant compte également des quantités débarquées par saison.
- *Catégories commerciales.* — Avec la présentation en catégories commerciales, l'enquêteur bénéficie de tous les avantages potentiels de la stratification sans avoir à en payer le prix. Mesurés en termes de coût, ces avantages sont importants. A précision égale, l'efficacité de la stratification peut se mesurer par le rapport des coûts optimaux d'échantillonnage. Pour les exemples traités, ce rapport varie entre 1,4 et 15,5. Nous n'insisterons pas davantage sur l'intérêt de la stratification en catégories commerciales qui est d'ailleurs incluse dans le plan de sondage type car elle s'impose à l'enquêteur dans de nombreux cas.
- *Choix de la grandeur objectif.* — L'efficacité du plan de sondage, en terme de coût relatif, va dépendre du choix de la grandeur objectif. Mais, le choix ne se fera pas sur un critère d'efficacité. Seule interviendra l'utilité de la fonction objectif. C'est dire que le statisticien n'a aucun contrôle sur ce point.
- *Echantillonnage aléatoire ou en deux phases pour l'estimation des proportions par âge.* — La décision sur ce point appartient à l'enquêteur. Le choix peut s'appuyer sur un critère d'efficacité. En termes de coût et pour les exemples traités, l'efficacité relative du double échantillonnage varie entre 1,04 et 1,27 pour des populations préalablement stratifiées. Les gains sont beaucoup plus importants en l'absence de stratification. Il est d'ailleurs normal que les gains d'efficacité diminuent au fur et à mesure que le plan de sondage s'améliore.

Pratiquement et pour le schéma type que nous avons retenu, les gains d'efficacité par l'intermédiaire du plan de sondage pourront être acquis par une bonne définition des strates temporelles et l'utilisation de la taille des individus comme variable auxiliaire.

### La grandeur objectif.

L'estimation du vecteur capture sera souvent utilisée à plusieurs fins et nous pouvons donc être en présence de fonctions des observations de types très différents. Une hiérarchie pourra être établie sur un

critère d'utilité. Cependant, il n'est pas inintéressant d'évaluer quelles peuvent être les conséquences sur la précision de l'estimation d'une fonction donnée, des écarts aux solutions optimales des taux d'échantillonnage. Considérant les deux fonctions retenues,  $F$  moyen et poids moyen des captures et pour le plan d'échantillonnage adopté, les rapports des coûts à l'optimum sont respectivement, pour les populations 1 et 2, 1,56 et 1,21. Pour une même précision, l'estimation du poids moyen des captures est plus économique. L'examen du tableau 7 permet de constater que les rapports ageage-mesurage ( $\lambda_h$ ) varient peu selon la grandeur objectif considérée. Les différences sont par contre très marquées entre les deux populations tests qui ont, il est vrai, des structures très différentes. Ceci souligne l'intérêt d'une bonne stratification temporelle. Pour ce qui est des poids relatifs des catégories commerciales ( $q_h$ ), on peut noter qu'ils sont pratiquement insensibles à la méthode d'échantillonnage des âges. Par contre, leur distribution dépend fortement de la structure de la population et de la fonction considérée.

Grandeur objectif		$f = \sum \omega_i P_i$				$f = \bar{F}_{1,6}$			
		Présence		Absence		Présence		Absence	
Catégories commerciales									
Type d'échantillonnage		aléatoire	2 phases	aléatoire	2 phases	aléatoire	2 phases	aléatoire	2 phases
Distribution 1	$\lambda$	.	.	1	0,122	.	.	1	0,080
	$n_1$	84	318	189	913	140	602	717	3 482
	$n_2$	84	69	189	112	140	100	717	280
	$q$	2,2	2,0	2,6	2,1	4,5	4	11,1	7,7
	coût	1 811	1 580	3 773	2 498	3 146	2 470	14 563	6 903
Distribution 2	$\lambda$	.	.	1	0,143	.	.	1	0,094
	$n_1$	68	226	125	579	64	180	1 250	6 120
	$n_2$	68	60	125	83	64	60	1 250	573
	$q$	1,6	1,5	1,6	1,4	3,8	3,8	25,1	18,4
	coût	1 452	1 329	2 482	1 809	1 677	1 612	25 910	14 315

TABLE. 6. — Taux d'échantillonnage optimaux et coûts afférents pour un coefficient de variation de 0,10 de la grandeur objectif (\* détails par strate tableau 7).

En utilisant pour chacune des fonctions considérées, les solutions optimales obtenues pour l'autre et toujours pour une précision relative de 10 %, on obtient comme efficacité relative, en terme de coût :

	Distribution 1	Distribution 2
Poids moyen des captures	0,86	0,36
$F_{1,6}$	0,83	0,26

Les pertes d'efficacité sur l'estimation des grandeurs objectifs non prioritaires peuvent être importantes et, dans ce domaine, la différence de structure entre les populations est aussi prépondérante. Mais sur ce point précis, la stratification temporelle ne pourra apporter aucune amélioration.

L'action combinée de la structure de la population et de la nature de la fonction des observations apparaît clairement si l'on considère la formule 1. Faisant abstraction des termes de covariance, on peut écrire :

$$\text{Var} [f(\hat{p})] = \sum a_i \text{Var}(\hat{P}_i)$$

les variances traduisent l'influence de la structure de la population ; les coefficients  $a_i$  étant liés à la grandeur objectif avec :

$$a_i = \omega_i^2 \text{ pour } f = \sum_i P_i$$

$$a_i = \left[ \frac{\partial \bar{F}}{\partial P_i} \right]^2 \text{ pour } f = \bar{F}$$

Les formules des dérivées partielles  $\frac{\partial \bar{F}}{\partial P_i}$  et des coefficients de sensibilité ont été explicitées par LAUREC et MESNIL (1985b) et LAUREC (1986) et peuvent être calculées par un programme développé par ces auteurs pour des valeurs individuelles de F. A titre d'exemple, nous avons pris  $F = F_4$ . On trouvera dans le tableau 8 la valeur des coefficients  $a_i$  pour 3 grandeurs objectifs :

- $f = \sum \omega_i P_i$
- $f = F_4$ , taux de mortalité par pêche à l'âge 4
- $f = N_1$ , effectif de la cohorte à l'âge 1.

Pour cette dernière fonction, les coefficients ont été calculés à partir du modèle de Pope.

Distribution		1				2			
		$f = \sum \omega_i P_i$		$f = \bar{F}_{1,6}$		$f = \sum \omega_i P_i$		$f = \bar{F}_{1,6}$	
Type d'échantillonnage		aléatoire	2 phases	aléatoire	2 phases	aléatoire	2 phases	aléatoire	2 phases
$\lambda_h$	$h = 1$	1	0,177	1	0,159	1	0,406	1	0,392
	$h = 2$	1	0,252	1	0,253	1	0,206	1	0,218
	$h = 3$	1	0,276	1	0,285	1	0,268	1	0,352
	$h = 4$	1	0,240	1	0,231	1	0,367	1	0,331
	$h = 5$	1	0,147	1	0,110	1	1	1	1
$q_h$	$h = 1$	0,05	0,05	0,05	0,05	0,11	0,11	0,02	0,02
	$h = 2$	0,13	0,13	0,09	0,09	0,30	0,30	0,03	0,03
	$h = 3$	0,17	0,17	0,09	0,09	0,27	0,27	0,12	0,120
	$h = 4$	0,26	0,26	0,15	0,15	0,26	0,26	0,50	0,50
	$h = 5$	0,39	0,39	0,63	0,63	0,05	0,05	0,33	0,33

TABL. 7. — Détail des proportions d'individus âgés par strate ( $\lambda_h$ ) et du poids relatif de chaque strate ( $q_h$ ).

Une première remarque s'impose : le niveau moyen des coefficients peut être très différent. D'où l'intérêt d'utiliser des précisions relatives (coefficients de variation) dans une analyse de ce type. Par ailleurs, la valeur des coefficients  $a_i$  pour des fonctions des observations issues de l'analyse des cohortes sera fortement influencée par le niveau du taux de mortalité naturelle  $M$  et à un moindre degré du taux d'exploitation  $E$  de la dernière classe d'âge ( $M = 0,20$ ,  $E = 0,50$  dans nos exemples). Les coûts optimaux pour les trois fonctions s'élèvent respectivement à 1 580, 6 240 et 1 434, pour une même précision. Les pertes d'efficacité entre la première et la troisième seront faibles par contre, on peut s'attendre à ce qu'elles soient plus sévères avec la deuxième. Nous avons reporté sur la figure 1 les coefficients de variation des proportions estimées des différents âges correspondant au plan de sondage optimal obtenu pour chacune des fonctions retenues. Les profils sont comparables pour les trois fonctions et nettement liés à la structure démographique de la population. Les niveaux moyens de précision dépendent de la grandeur objectif.

Age	$\sum \omega_i P_i$	$N_i$	$F_4$
1	0,230	1,221	0
2	0,848	1,823	0
3	1,769	2,719	0
4	3,538	4,056	0,987
5	7,701	6,052	0,129
6	15,093	9,024	0,193
7	25,000	13,462	0,288
8	35,319	20,088	0,424
9	48,303	29,965	0,661
10	58,523	44,703	0,868
11	68,443	66,684	1,433
12	87,628	99,481	2,122
13	103,897	486,042	10,583
	456,292	785,320	17,688

TABLE. 8. — Valeurs des coefficients de la formule de variance pour quelques types de grandeur objectif.

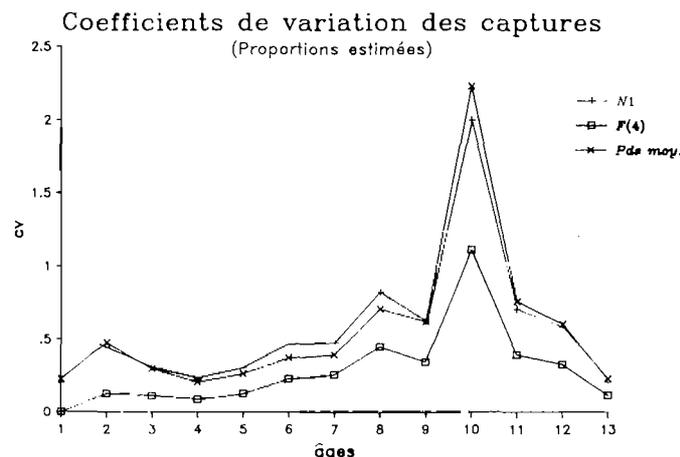
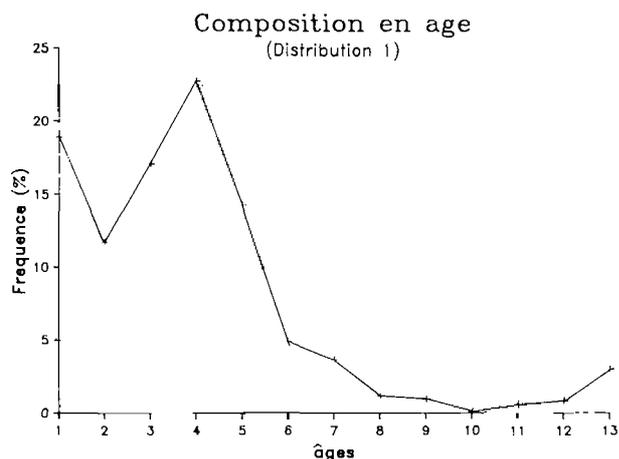


FIG. 1. — Covariation entre les proportions d'âge dans les débarquements (distribution 1) et les coefficients de variation afférents pour trois types de grandeur objectif.

#### Problèmes pratiques d'optimisation.

Le plan de sondage défini et les estimateurs choisis, il sera possible de faire des estimations du vecteur des captures et d'évaluer la variance des résultats obtenus en utilisant des taux de sondage plus ou moins arbitraires. Il n'est pas nécessaire de connaître, *a priori*, les effectifs des unités primaires et les poids relatifs des catégories commerciales. Les informations indispensables pour calculer les estimations et leur variance seront fournies à la fin des opérations d'échantillonnage. Cependant, un choix arbitraire des taux d'échantillonnage peut entraîner des pertes d'efficacité, en termes de coût ou de précision, qui peuvent être importantes. C'est pourquoi, avant de démarrer des sondages intensifs, des enquêtes pilotes sont parfois réalisées pour obtenir les informations manquantes, notamment sur les variances des variables observées et sur les coûts. Mais, ceci suppose une certaine stabilité de la population échantillonnée, ce qui n'est pas le cas dans le problème qui nous préoccupe. Rappelons que les populations tests que nous avons utilisées représentent la composition moyenne des captures de lieu noir pour deux trimestres consécutifs.

Les plans de sondage optimaux correspondants sont tout à fait différents. Toutefois, on dispose de données historiques sur les captures ; dans l'hypothèse d'une stabilité relative des pêcheries, les données acquises les années antérieures peuvent être utilisées dans le processus d'optimisation par strates spatio-temporelles.

Les coûts exacts des différentes phases de l'échantillonnage ne sont pas toujours faciles à déterminer notamment si les coûts marginaux des unités échantillonnées doivent être utilisés ou s'il existe des problèmes d'imputations. Toutefois, nous disposons de suffisamment d'expérience pour estimer des coûts moyens. Il semblerait par ailleurs que les valeurs optimales des taux d'échantillonnage soient relativement peu sensibles à des variations modérées des coûts relatifs.

En ce qui concerne les composantes des variances des variables observées, nous ne disposons pas, en général, de données antérieures qui permettraient de les estimer. C'est pourquoi nous avons utilisé des techniques de simulations pour créer des débarquements fictifs, quantités et poids relatifs des classes de longueur. On introduit ainsi une variabilité par rapport aux débarquements réels. Cependant, il faut noter que pour engager un processus d'optimisation, nous sommes obligés de faire référence au passé. Or même si l'on admet une stabilité moyenne de la structure des débarquements d'une année sur l'autre, on peut imaginer que dans le détail ils ne se reproduiront pas identiquement et donc une certaine variabilité existe naturellement. Variabilité qui dans l'un et l'autre cas se fait sentir principalement sur la composante inter unités primaires des variances. Ceci permet des estimations des valeurs optimales de  $\lambda_h$  et  $q_h$ , proportion d'individus âgés et poids relatif des strates, relativement stables. Pour la détermination du nombre total d'unités primaires à échantillonner, phase finale du processus d'optimisation, il serait peut-être préférable de travailler à coût constant dans la mesure où l'on dispose d'une bonne estimation des coûts moyens. Quoi qu'il en soit, les calculs classiques de variances effectués en fin de sondage devraient permettre de juger la fiabilité du plan optimal proposé.

### **Discussion générale.**

Le concept de précision est un critère de base pour juger de la valeur d'une estimation. Comme nous l'avons écrit au début de cette note, c'est une valeur centrale de l'analyse du statisticien. Aussi, nous voudrions dans ce paragraphe revenir sur les problèmes de précision qui apparaissent à différents niveaux de ce rapport.

En premier lieu, il faut souligner qu'il arrive fréquemment que la détermination des valeurs optimales des taux d'échantillonnage se fasse à partir de données non précises : effectifs relatifs des strates, évaluation grossière des coûts et estimation des variances des variables observées. L'imprécision attachée à toutes ces informations entraîne des erreurs sur les solutions optimales et donc sur la précision espérée. Il n'en reste pas moins qu'une amélioration de l'efficacité du plan de sondage est attendue de ces résultats et justifie le lancement d'enquêtes pilotes. Or, dans la mesure où une certaine stabilité peut être observée d'une année sur l'autre dans la structure des débarquements, les informations recueillies dans le passé peuvent nous servir à planifier les sondages futurs avec l'espoir raisonnable d'améliorer l'efficacité de nos interventions.

Toutefois, nous ne disposons pas toujours, actuellement, de tous les éléments qui permettraient cette analyse. C'est pourquoi nous avons utilisé les techniques de simulation afin de situer les principales sources de variabilité des estimations. Réduite à cet objectif, la méthode utilisée donne des informations utiles. Par contre, son utilisation pour la prévision et la planification des opérations d'échantillonnage pose un problème quant à la précision des résultats attendus. En effet, et nous l'avons déjà souligné, même si d'une année sur l'autre et pour une même strate spatio-temporelle, les structures moyennes des populations débarquées sont relativement comparables, il est à peu près certain que dans le détail la composition des débarquements n'est pas reproduite identiquement. En conséquence, la part de la variance due à ces fluctuations entre débarquements (unités primaires) va pouvoir fluctuer avec les années. On n'ajoute donc pas une nouvelle source de variabilité en utilisant des valeurs simulées des unités primaires. Par contre, il ne sera pas toujours aisé d'apprécier le degré de variabilité qu'il faut introduire. Or, selon le nombre de tirages aléatoires effectué sur la population témoin on obtiendra une image plus ou moins précise de sa structure et une variabilité plus ou moins forte entre les effectifs simulés. Dans nos analyses, pour les deux populations considérées comprenant 24 et 30 classes de taille, nous avons réalisé 300 tirages pour chaque simulation.

Afin d'apprécier les conséquences induites par la précision des simulations, nous avons aussi analysé des séries de 150 tirages des deux distributions. Les valeurs optimales des taux d'échantillonnage sont reportées dans le tableau 9 avec les résultats analogues pour  $N = 300$ . Les taux d'échantillonnage ( $\lambda_h$ ) des individus âgés restent remarquablement stables. En fait, ils ne dépendent pas de la variabilité entre unités primaires. Les poids relatifs optimaux des catégories commerciales sont relativement stables pour les trois premières catégories commerciales, mais varient plus ou moins fortement pour les catégories contenant les individus les plus âgés et les effectifs les plus faibles.

Par ailleurs, et ceci est théoriquement prévisible, le nombre d'unités primaires à échantillonner augmente quand on augmente la variabilité entre les effectifs simulés. Au total et pour la même précision, le coût minimal est en moyenne, sur nos exemples, de 25 % plus élevé pour  $f = \bar{F}_{1,6}$  et de 15 % environ quand on s'intéresse au poids moyen des captures. Il faut noter qu'à l'égard de ce facteur, la stratification en catégories commerciales s'avère très efficace. Les pertes sont en effet plus importantes sur des populations non stratifiées.

Distribution		1				2			
		$f = \sum \omega_i P_i$		$f = \bar{F}_{1,6}$		$f = \sum \omega_i P_i$		$f = \bar{F}_{1,6}$	
Simulation N		150	300	150	300	150	300	150	300
$\lambda_h$	h = 1	0,183	0,177	0,164	0,159	0,422	0,406	0,413	0,392
	h = 2	0,253	0,252	0,253	0,253	0,205	0,206	0,217	0,218
	h = 3	0,275	0,276	0,284	0,285	0,280	0,268	0,367	0,352
	h = 4	0,242	0,240	0,235	0,231	0,393	0,367	0,416	0,331
	h = 5	0,161	0,147	0,123	0,110	1	1	1	1
$q_h$	h = 1	0,05	0,05	0,04	0,05	0,12	0,11	0,02	0,02
	h = 2	0,10	0,13	0,06	0,09	0,34	0,30	0,04	0,03
	h = 3	0,16	0,17	0,07	0,09	0,28	0,27	0,23	0,120
	h = 4	0,19	0,26	0,10	0,15	0,25	0,26	0,70	0,50
	h = 5	0,51	0,39	0,73	0,63	0,00	0,05	0,02	0,33
q		3,9	2	8,9	4	2,1	1,5	5,0	3,8

TABL. 9. — *Détail des proportions d'individus âgés par strate ( $\lambda_h$ ) et du poids relatif de chaque strate ( $q_h$ ). q = total d'unités primaires à échantillonner pour l'ensemble des strates.*

Pour un coût donné, la précision attendue pourra être sur ou sous-évaluée selon l'appréciation qui aura été faite sur la variabilité entre les divers débarquements intervenant au cours d'une période donnée. Cette erreur est théoriquement et pratiquement inévitable. Toutefois, elle doit être appréciée en tenant compte de toutes les autres sources d'erreurs qui vont affecter les résultats finaux. Or, pour un grand nombre d'espèces échantillonnées en criée, les compositions en âge des débarquements annuels sont utilisées dans des commissions internationales et cumulées, avant traitements, avec des données similaires estimées par nos partenaires scientifiques étudiant les mêmes stocks à partir de leurs propres pêcheries. La plupart des modèles utilisés, pour ne pas dire la totalité, utilisent les captures plutôt que les débarquements. Les données sont alors corrigées d'une estimation des rejets qui peuvent ne pas être négligeables selon les cas. Toutes ces manipulations font qu'il est très difficile d'apprécier et de maîtriser la précision des produits finaux.

En supposant que les statistiques qui nous sont fournies pour faire une estimation sont fiables, il n'est pas interdit de définir une précision des données que nous fournissons, tenant compte de l'importance de nos captures par rapport au total des prises sur le stock. Il y a, dans ce cas, tout intérêt à obtenir le résultat désiré au coût minimal, c'est-à-dire à optimiser le plan de sondage. Des solutions empiriques, plus ou moins intuitives, sont utilisées. Une approximation du plan de sondage optimal doit pouvoir être fournie par les méthodes de Monte-Carlo, sous réserve du choix d'une grandeur objectif. Ce choix peut d'ailleurs impliquer l'existence de covariance entre les fréquences estimées des différents âges. Si c'est le cas, les gains de précision peuvent être importants. Si on s'intéresse, par exemple, au poids moyen des captures, le coût minimal peut être réduit par le facteur 10 et à coût égal, le coefficient de variation est réduit d'environ un tiers.

Précisons que nous avons implicitement admis dans notre analyse que l'effectif total des poissons débarqués était connu. En fait, il devra être estimé et l'erreur attachée à cette estimation prise en compte pour l'évaluation du plan optimal.

Nous avons, jusqu'ici, considéré des sources de variabilité liées au plan d'échantillonnage et aux techniques de simulation. Mais un sondage implique la mesure de certains caractères sur les individus sélectionnés. La précision de l'estimation faite à partir de ces mesures dépendra de l'erreur d'échantillonnage, due au fait qu'on obtient qu'une partie de l'information, celle fournie par l'échantillon, et de l'exactitude de la mesure. Nous avons admis que la détermination de l'âge était réalisée sans erreur. S'il n'en est pas ainsi l'erreur de mesure vient s'ajouter à l'erreur d'échantillonnage. L'erreur de mesure peut être génératrice de biais si elle intervient toujours dans le même sens ou accroître la variabilité des observations si elle est purement aléatoire. Une solution satisfaisante à ce problème ne peut venir que d'une amélioration de la mesure des caractères étudiés.

## Conclusion.

Considérant les différentes sources de variabilité qui interviennent au cours du processus d'échantillonnage, nous avons essayé d'évaluer l'impact de chacune d'entre elles sur la précision du résultat final et donc pour une précision donnée sur les coûts de l'opération. Cette analyse repose sur la recherche du plan optimal correspondant à une situation précise. N'ayant pas toutes les informations pour réaliser les calculs, nous avons utilisé des techniques simples de simulation pour estimer les données manquantes. La méthode pourrait être utilisée, le plan de sondage étant défini et la grandeur objectif précisée, pour évaluer des taux d'échantillonnage approximativement optimaux en tenant compte des résultats exposés dans ce document.

## BIBLIOGRAPHIE

- CHEVALIER (R.) et MESNIL (B.), 1986. — Echantillonnage en criée pour l'estimation des structures démographiques. — Document technique. IFREMER, DRV-86.006 RH/Nantes.
- COCHRAN (W.G.), 1977. — Sampling techniques. — New York : 3rd ed. J. Wiley and Sons.
- COCHRAN (W.G.), 1981. — The Information supplied by the sampling Results. Contributions to Statistics. Contribution n° 12. New York : J. Wiley and Sons.
- JENSEN (R.J.), 1978. — Statistical Survey Techniques : J. Wiley.
- KIMURA (D.K.), 1977. — Statistical Assessment of the age length key. — *J. Fish. Res. Board Can.*, **34** : 317-324.
- LAUREC (A.), 1986. — Les méthodes Delta en halieutique. Evaluation des sensibilités, approximation des biais et des variances à l'aide de développements limités. — Document technique IFREMER, DRV-86.002 RH/Nantes.
- LAUREC (A.) et MESNIL (B.), 1985 (a). — Analytical investigations of errors in mortality rates 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, Italie (février 1985).
- 1985 (b). — Sensibilité des analyses de cohortes et des projections déduites. — Document présenté au Groupe de Travail « méthode » du CIEM. Copenhague (novembre 1985).
- POPE (J.G.), 1972. — An Investigation of the Accuracy of Virtual Population Analysis Using Cohort Analysis. — *Res. Bull. ICNAF*, **9**.
- POPE (J.G.) et GRAY (D.), 1983. — An Investigation of the Relationship between the Precision of Assessment Data and the Precision of total Allowable Catches dans Doubleday W.G. and Rivard D. ed. — L'échantillonnage des prises commerciales de poissons et d'invertébrés marins. — Publ. spéc. Can. Sci. halieut. aquat., **66** : 290.
- RAO (C.R.), 1973. — Linear Statistical Inference and its applications. — New York : J. Wiley.
- YATES (F.), 1981. — Sampling Methods for Censuses and Surveys. — London : 4 Ed. C. Griffin and Co.

Manuscrit soumis le 7-9-1987, accepté le 22-9-1987.

ECHANTILLONNAGE DES DÉBARQUEMENTS EN CRIÉE POUR L'ÉTUDE DES STRUCTURES DÉMOGRAPHIQUES

Liste des symboles utilisés

- $i$  = indice des unités primaires ou des âges  
 $h$  = indice des catégories  
 $k$  = indice des strates de longueur  
 $n_{th}$  = nombre d'individus mesurés dans la catégorie  $h$   
 $n_1$  = nombre total d'individus mesurés (1<sup>re</sup> phase)  
 $n_{2k}$  = nombre d'individus âgés dans la strate de longueur  $K$   
 $n_2$  = nombre total d'individus âgés  
 $P$  = proportion d'individus d'âge  $t$  donné dans la population  
 $\underline{P}$  = vecteur des proportions par âge  
 $\underline{P}_k$  = proportion d'individus d'âge  $t$  donné dans la strate de longueur  $K$   
 $\hat{p}$  = estimation du vecteur  $\underline{P}$   
 $q$  = nombre total d'unités primaires échantillonnées  
 $q_h$  = poids « statistique » de la catégorie commerciale  $h$   
 $T_i$  = effectif total de l'unité primaire  $i$   
 $\bar{T}$  = effectif moyen des débarquements au cours d'une période donnée  
 $W_k$  = proportion de la population dans la strate de longueur  $K$   
 $w_k$  = proportion de l'échantillon dans la strate de longueur  $K$   
 $\lambda$  =  $n_2/n_1$ , proportion d'individus âgés  
 $\omega_i$  = poids moyen à l'âge  $i$   
 $\Phi$  = fonction de pertes quadratique moyenne  
 $\Lambda$  = signe d'identification des estimations
-