

ESTIMATEURS D'INDICES D'ABONDANCE
DANS LE CAS D'ECHANTILLONNAGES STRATIFIES

J. BERTRAND*, R. CHEVALIER**



Abstract :

Abundance indice estimators based on stratified random trawl surveys.

One of the primary uses of trawl survey data is the estimation of abundance of the species involved. But these direct estimates give results with generally a high variability. Reducing this latter is one of the main problems for the scientists involved in such a research. Methods to achieve a better efficiency have been implemented. Nevertheless estimations remain highly variable, so the information on the changes in population, enclosed in the observed series of survey indices is mixed with a noise component result of both the survey sampling variability and effect of uncontrolled factors.

PENNINGTON (1985) proposed time series models to filter measurement error from the signal. However the length of the series does not always allow using time series methodology. So for a short series we have tried smoothing the observed indices using the variance analysis model to calculate parametric estimators. It is assumed that the data are measured with multiplicative random errors. The values so estimated are more efficient than the classical estimates. The indices so obtained for age group 3 (recruitment) are more closely related to the VPA estimates than the current ones.

Résumé :

Le premier objectif des campagnes expérimentales de chalutage est l'estimation d'indices d'abondance des espèces suivies qui serviront notamment à la calibration des modèles de dynamique.

Malgré l'amélioration des stratégies d'échantillonnage les estimations obtenues restent hautement variables. Il s'ensuit que les informations sur l'évolution des populations, contenues dans une série historique le plus souvent relativement courte, sont masquées par une multitude de signaux parasites provenant de variables non contrôlées et bien souvent imprévisibles.

Pour filtrer les informations utiles PENNINGTON(1985) préconise l'étude des séries observées par les techniques classiques d'analyse des séries temporelles. Pour des séries très courtes la technique peut ne pas être très performante. Par ailleurs, il est quand même dommage de perdre l'information accumulée aussi brève soit-elle. C'est pourquoi nous avons essayé sur une série de données du Banc St-Pierre (subdivision 3Ps de la NAFO) un lissage des données par un modèle d'analyse de variance à deux facteurs contrôlés. Les principaux résultats sont reportés dans cette note.

Introduction

Les estimations directes d'abondance par l'intermédiaire de chalutages expérimentaux donnent des résultats qui sont le plus souvent entachés d'une forte variabilité. Pour améliorer la précision des estimations et donc l'efficacité de la technique plusieurs solutions ont été adoptées.

a) Choix d'une stratégie d'échantillonnage

Les acquis de la théorie de l'échantillonnage permettent de rechercher la stratégie la mieux adaptée. Depuis que GROSSLEIN (1969) a mis en évidence les avantages des plans aléatoires stratifiés, la technique est largement utilisée pour programmer les campagnes expérimentales. Il faut noter toutefois que les gains en précision sur une stratégie aléatoire simple ne sont pas automatiques et dépendent aussi bien du schéma de stratification que de la répartition de l'effort d'échantillonnage entre les strates (GAVARIS et SMITH, 1987).

b) Choix d'estimateurs efficaces

La technique proposée par PENNINGTON (1983) est basée sur le fait que la distribution des organismes marins est en général surdispersée. Il peut être utile de "normaliser" la série des observations pour utiliser des tests paramétriques ou calculer des intervalles de confiance.

Si l'hypothèse d'une distribution log-normale peut être retenue pour la série des observations Y , une transformation logarithmique des données entraînera une normalisation de la série transformée. De plus si on désigne par θ_1 le moment d'ordre 1 de la loi log-normale on dispose de deux estimateurs non biaisés de ce paramètre :

1) $\bar{y} = (\Sigma y)/n$ estimé sur les données non transformées,

2) $\theta_1 = \exp(\bar{x}) \Phi_1(s^2_{\bar{x}})$, \bar{x} et $s^2_{\bar{x}}$ représentant les estimateurs classiques de la moyenne et de la variance des données transformées et Φ_1 une fonction de $s^2_{\bar{x}}$ connue par son développement en série.

On dispose également de deux estimateurs non biaisés pour la variance θ_2 de la loi log-normale qui conduisent à des estimations différentes des variances des deux estimateurs précédents. Des résultats dus à FINNEY (1941) (KENDALL et Al. 1983) montrent que l'estimateur θ_1 est plus efficace que y . Il y a donc un intérêt théorique à choisir le premier.

c) Modélisation des données

Une série d'indices d'abondance estimés à partir de chalutages expérimentaux est fortement contaminée par du bruit provenant de l'imprécision des mesures ou d'action de facteurs non contrôlés et souvent imprévisibles. Pour dégager l'information utile du bruit environnant et obtenir une meilleure efficacité dans l'analyse,

notamment en terme de réduction de variance, il peut être intéressant d'appliquer des techniques de lissage.

PENNINGTON (1985) utilise les techniques décrites par BOX et JENKINS pour modéliser une série d'indices d'abondance et obtenir une méthode de lissage fournissant une série corrigée d'indices affectés chacun d'une variance nettement plus faible que la variance de l'indice brut. Théoriquement la structure du modèle se déduit des données empiriques. Cependant ce résultat ne peut être obtenu qu'avec de longues séries ($>=30$), éventualité rarement rencontrée à ce jour dans le cas d'indices d'abondance évalués à partir de chalutages expérimentaux. Faute de disposer de séries suffisamment longues FOGARTY et al. (1986), à la suite de PENNINGTON, proposent une structure définie a priori. Si les résultats obtenus paraissent encourageants, il faut quand même souligner que les séries traitées comportent une vingtaine d'années dans les deux cas.

Avec une série beaucoup plus courte il peut quand même être intéressant d'utiliser les informations accumulées au cours des années pour obtenir un lissage de la série et une réduction de la variance résiduelle. Nous avons utilisé pour cela le modèle classique d'analyse de variance.

Méthode

Des campagnes de chalutages expérimentaux sont effectuées selon un protocole identique chaque année sur la subdivision 3Ps de la NAFO par le laboratoire de St-Pierre et Miquelon depuis 1978. L'échantillonnage est réalisé suivant un schéma aléatoire stratifié. On dispose donc d'une série d'observations structurées suivant un plan expérimental à deux facteurs croisés: strates-années, modèle classique d'analyse de variance. Il s'agit d'un modèle fixe qui se présente comme un cas particulier du modèle général de régression linéaire.

Compte tenu de la nature des données l'analyse est conduite sur une échelle logarithmique. Il est attendu de cette transformation qu'elle normalise les erreurs ou tout au moins qu'elle stabilise leur variance.

Le modèle classique complet peut s'écrire :

$$x_{i,j,k} = \mu + \alpha_i + \beta_j + \pi_{i,j} + \epsilon_{i,j,k}$$

$$i=1, \dots, nL, \quad j=1, \dots, nC, \quad k=1, \dots, n_{i,j}$$

Il est utilisé pour tester la présence éventuelle d'interaction et estimer la somme des carrés résiduelle.

Des analyses préliminaires n'ayant pas nettement mis en lumière l'existence d'interaction, le modèle simplifié suivant a été utilisé pour estimer les paramètres μ , α et β :

$$x_{i,j,k} = \mu + \alpha_i + \beta_j + \epsilon_{i,j,k}$$

$$\sum \alpha_i = 0 \quad \text{et} \quad \sum \beta_j = 0$$

Si l'on désigne par m, a_1, b_j les estimateurs de μ, α_1 et β_j l'estimation retenue comme indice d'abondance pour l'année j et la strate i sera donnée par la relation :

$$u_{ij} = m + a_1 + b_j$$

Compte tenu des hypothèses retenues, u_{ij} a une distribution gaussienne et :

$$\text{Var}(u_{ij}) = c * S/n$$

S représentant la somme des carrés résiduelle, n le nombre de degrés de liberté du plan analysé et c un coefficient calculé à partir des éléments de la matrice inverse des équations normales. La moyenne et la variance des données brutes s'obtiennent à partir des valeurs précédentes en utilisant les formulations de HOYLE (1968).

Les indices d'abondance corrigés sont calculés à partir des estimations ainsi obtenues, par strate et par année.

Résultats

a) analyse de variance et résidus

L'étude porte sur la morue, principale espèce recherchée dans la région considérée. Les analyses ont été réalisées sur les observations faites au cours de 10 années (1978-1987) et sur 18 strates couvrant l'essentiel de l'aire de répartition du stock de morue de la subdivision 3Ps. Cette concordance est d'ailleurs indispensable si l'on veut utiliser les résultats des campagnes pour le calibrage des VPA.

Nous avons reporté sur la figure 1 le graphique des points : logarithme de la variance versus logarithme de la moyenne, valeurs calculées sur toutes les strates informées. L'ajustement sur ces points de la fonction de Taylor:

$$\text{Log}(\sigma^2) = \text{Log}(a) + b \text{Log}(\mu)$$

fournit les estimations suivantes :

$$b = 2.17, \quad a = -0.786, \quad R^2 = 0.93,$$

résultats qui paraissent justifier la transformation logarithmique adoptée.

Le tableau récapitulatif de l'analyse de variance (tab.1) effectuée sur les données transformées montre un effet global significatif du au temps et à la stratification. Cependant les gains de précision sur l'estimation des indices d'abondance ne peuvent se déduire directement des résultats de cette analyse de variance. Plus important, l'hypothèse nulle en ce qui concerne l'interaction ne peut être rejetée ce qui nous autorise à utiliser le modèle additif, plus simple.

Un examen attentif des résidus autorise deux observations. La première découle de la figure 3 qui représente la distribution cumulée des résidus après transformation des fréquences cumulées en probit. L'alignement des points justifie l'hypothèse de normalité et le bien fondé de la transformation logarithmique. La deuxième repose sur la figure 2 qui donne la distribution des résidus en fonction des valeurs prédites. Le nuage obtenu traduit assez bien l'absence de corrélation entre les deux variables et confirme l'adéquation du modèle sans interaction (modèle additif) aux données étudiées.

b) Indices d'abondance et VPA

Nous reportons ici les résultats obtenus sur les estimations d'abondance en nombre.

Le modèle étant défini, on dispose de deux estimations pour l'indice d'abondance annuel :

- une estimation qu'on peut considérer comme classique utilisant directement les valeurs numériques observées et les valeurs moyennes subséquentes,
- une estimation paramétrique calculée à partir des paramètres du modèle linéaire.

Les graphiques des figures 4 à 6 donnent les estimations des valeurs moyennes et des coefficients de variation afférents pour les groupes d'âge 3+ et les estimations des indices d'abondance pour le groupe d'âge 3 pour les deux types d'estimateurs. Il apparaît nettement que les estimations paramétriques conservent les grandes tendances avec cependant un effet de lissage qui écrête les pics. Les coefficients de variation sont sensiblement réduits.

Les séries normalisées des estimations d'abondance issues des VPA ont été représentées sur les mêmes graphiques 4 à 6. L'adéquation entre les estimations fournies par la VPA d'une part et celles issues des campagnes expérimentales d'autre part a été mesurée par la somme des carrés des écarts (SCE) entre les indices normalisés de ces séries. Pour les âges cumulés les valeurs obtenues sont identiques. Il faut toutefois noter que la série des estimations directes a été utilisée pour le calibrage des VPA privilégiant ainsi la relation entre ces deux séries.

De plus, dans une précédente analyse utilisant la série des données directes, MABEAU et al (1986) ont montré que, contrairement à ce qui est observé pour les autres groupes d'âge, les corrélations entre les indices d'abondance des groupes 5-6 et les données de VPA correspondantes étaient très mauvaises. Ces résultats suggéraient une relative inadéquation du plan d'échantillonnage au regard du comportement de ces jeunes adultes. Le lissage n'améliore vraisemblablement pas significativement les résultats pour ces classes d'âge.

La même analyse faite sur le groupe 3, âge du recrutement, montre que la coïncidence entre la série VPA et celle des estimations paramétriques ($SCE = 1,10$) est meilleure que celle obtenue entre estimations directes et VPA ($SCE = 2,38$). Pour ce groupe d'âge la convergence de la VPA est largement assurée ce qui replace les deux types d'estimateurs à un même niveau d'indépendance vis-à-vis des données issues de la VPA.

Un des objectifs importants des campagnes expérimentales est de pouvoir corréler, aux fins de prédiction du recrutement, les résultats fournis par la VPA et ceux obtenus au cours de ces campagnes. Les résultats précédents laissent présager que la série paramétrique sera plus performante que la série des estimations directes. Les régressions VPA versus indices d'abondance confirment cette analyse.

En effet, si on retient la variance résiduelle comme critère de l'ajustement, l'efficacité relative de la série paramétrique par rapport à la série directe est égale à $3,62/2,52$ soit $1,44$, valeur qui reste sensiblement identique si on ne considère que les classes d'âge pleinement recrutées.

Discussion

Deux points importants sont à considérer en ce qui concerne la validité du modèle. Le premier porte sur l'hypothèse d'indépendance des erreurs sous-jacente au modèle d'analyse de variance. La répartition spatio-temporelle des observations peut faire naître quelque crainte à cet égard. Rappelons ici que le non-respect de cette hypothèse n'affecte pas les estimations des paramètres mais réduit l'efficacité du plan et modifie la variance de ces estimations. La répartition aléatoire des points de prélèvements dans les strates devrait donner du poids à l'hypothèse de non-corrélation en ce qui concerne les observations d'une même campagne. L'existence de corrélations entre les observations de campagnes différentes n'est pas à rejeter a priori. Le critère de Durbin-Watson n'est pas très performant, surtout pour des séries aussi courtes. Cependant pour l'ensemble des strates analysées le critère fournit une majorité de valeurs négatives. Ce qui semblerait indiquer une absence de corrélations positives entre les observations d'années successives.

Le deuxième point important quant à la valeur des estimations obtenues est relatif au type du modèle utilisé. Le modèle simplifié, sans interaction, a été retenu dans la mesure où les investigations effectuées n'ont pas donné de fortes évidences sur la réalité de ces interactions. Les risques de biais existent par contre l'estimation des variances reste correcte. De plus, l'exclusion des termes d'interaction réduit sensiblement les dimensions du modèle et induit en parallèle un gain de précision sur les paramètres.

Conclusion

Le modèle d'analyse de variance est fréquemment utilisé pour faire des estimations de données manquantes pour des strates accidentellement non échantillonnées. En utilisant dans tous les cas les estimateurs que permet de calculer cette technique on obtient une série d'indices d'abondance qui conserve les tendances d'une série "classique" tout en écrétant les pics. Ce lissage s'accompagne d'une réduction sensible de la variance des estimations. Les écarts entre les séries observées et la série calculée par la VPA sont du même ordre de grandeur en ce qui concerne les groupes d'âge 3+. Par contre on observe une meilleure coïncidence entre la série VPA et la série paramétrique pour le groupe d'âge 3 ce qui se traduit par un meilleur ajustement entre ces deux séries et probablement une meilleure performance des capacités prédictrices du recrutement.

REFERENCES

- FOGARTY (M. J), IDOINE (J. S), ALMEIDA (F. P) and PENNINGTON (M), 1986. - Modelling trends in abundance based on research vessel surveys. Cons. Int. Explor. Mer, C. M. 1986/G: 2
- GAVARIS (S) and SMITH (S. J), 1987. -Effect of allocation and stratification strategies on precision of survey abundance estimates for atlantic cod (Gadus morhua) on the eastern scotian shelf. J. Northw. Atl. Fish Sci. 7, : 137-144.
- GROSSLEIN (M. D), 1969. -Groundfish survey programm of BCF Woods Hole. - U. S. Fish Wildl. Serv., Comm. Fish. Rev., 31(8-9), 22-35.
- HOYLE (M. H), 1968. -The estimation of variances after using a gaussianating transformation. The Annals of Mathematical Statistics, 39, n° 4, : 1125-1143.
- KENDALL (M), STUART (A) and ORD (J. K), 1983. - The advanced theory of statistics. Design and analysis and time-series 3. C. GRIFFIN and Co Ltd.
- MABEAU (S), POULARD (J. C) & MAHE (J. C), 1986. -Research survey abundance indices for the cod stock in NAFO subdivision 3Ps, their reliability and compatibility with results of a cohort analysis. Cons. Inter. Explor. Mer, C. M. 1986/C : 38.
- PENNINGTON (M), 1983. -Efficient estimation of abundance for fish and plankton surveys. Biometrics, 39, : 281-286.
- PENNINGTON (M), 1985. -Estimating the relative abundance of fish from a series of trawl surveys. Biometrics, 41, : 197-202.

Tableau d'analyse de variances

<u>Source</u>	<u>somme des carrés</u>	<u>DL</u>	<u>carrés moyens</u>	<u>F</u>
Total	1397.1070	527	2.6511	
Effet strate	300.5469	17	17.6792	9.4127
Effet année	76.9727	9	8.5525	4.5535
Interaction	361.0288	153	2.3597	1.2563
Résiduel	653.6240	348	1.8782	

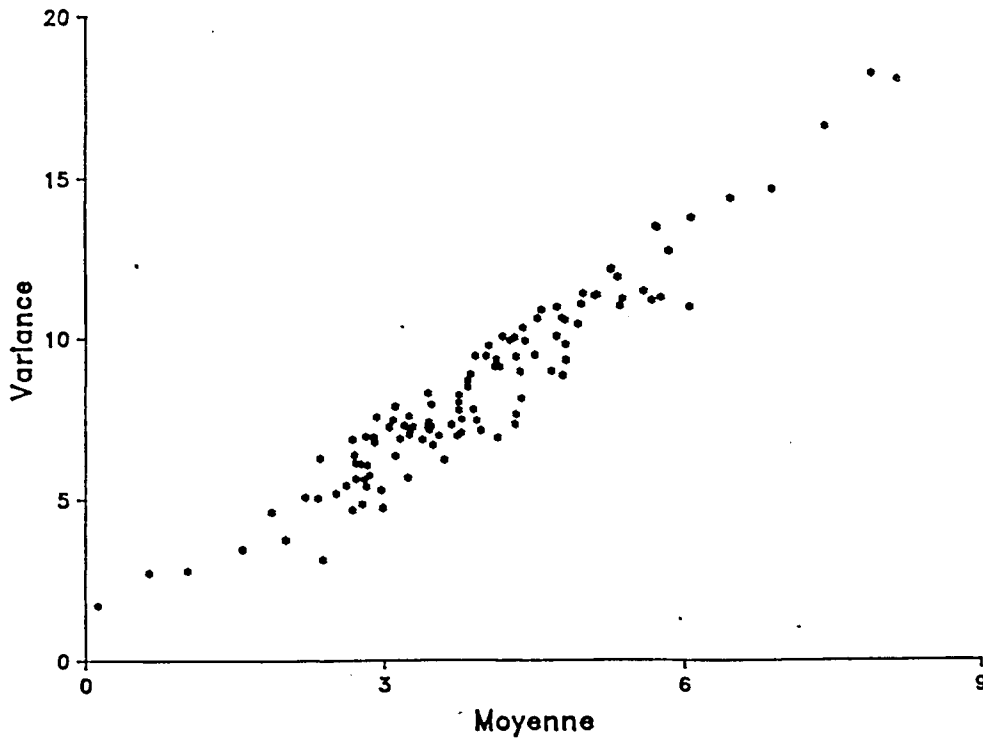


Figure 1 - Ajustement de la formule de Taylor : $\sigma^2 = a\mu^b$ sur échelle logarithmique.

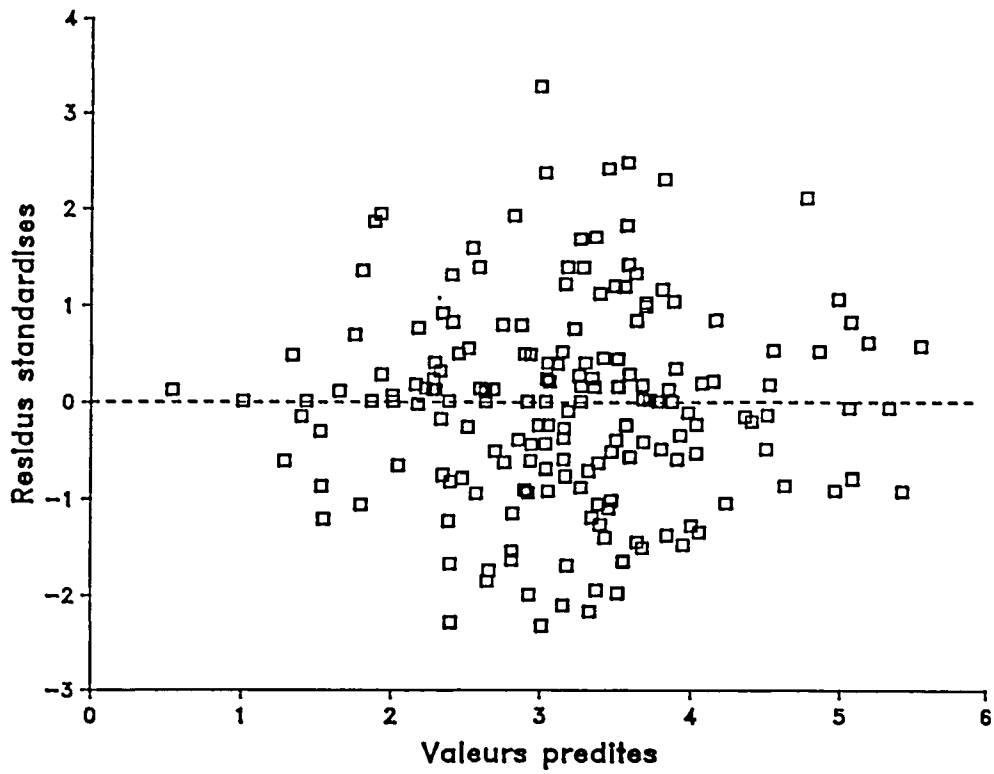


Figure 2 - Relation entre résidus standardisés et valeurs prédites.

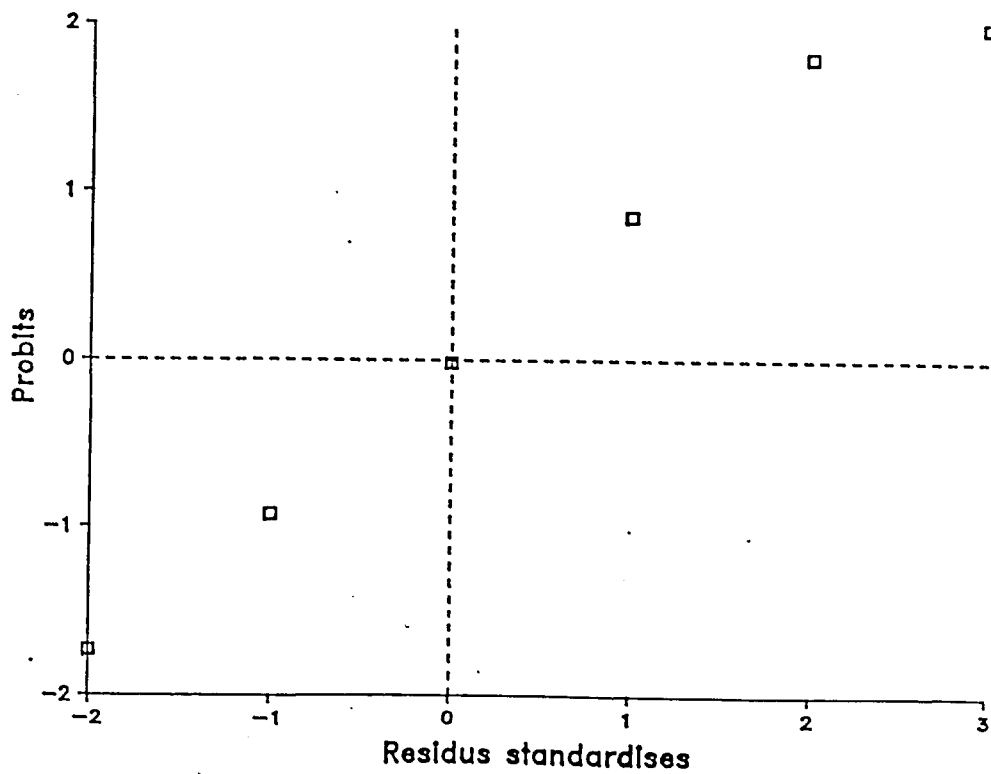


Figure 3 - Distribution comulée des résidus après transformation des pourcentages en probits.

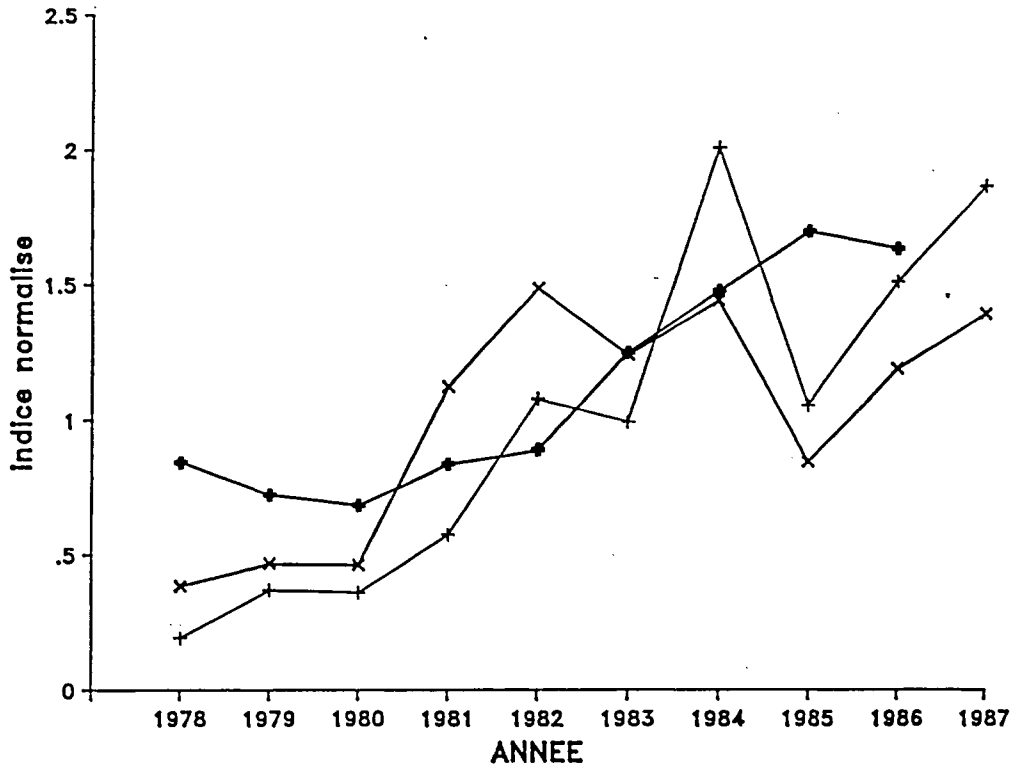


Figure 4 - Indices d'abondance des morues de 3 ans et plus.
 Campagnes d'évaluation + estimation directe
 x estimation paramétrique
 analyse des cohortes +

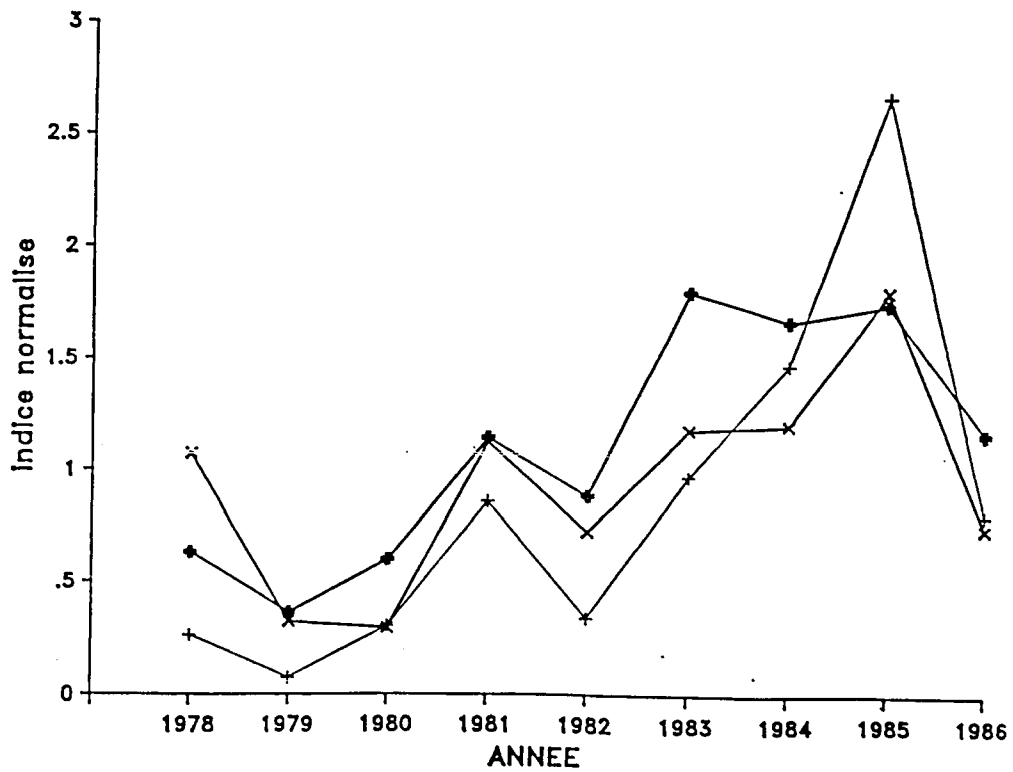


Figure 5 - Indices d'abondance à l'âge du recrutement (3 ans).
 Campagnes d'évaluation + estimation directe
 x estimation paramétrique
 analyse des cohortes +

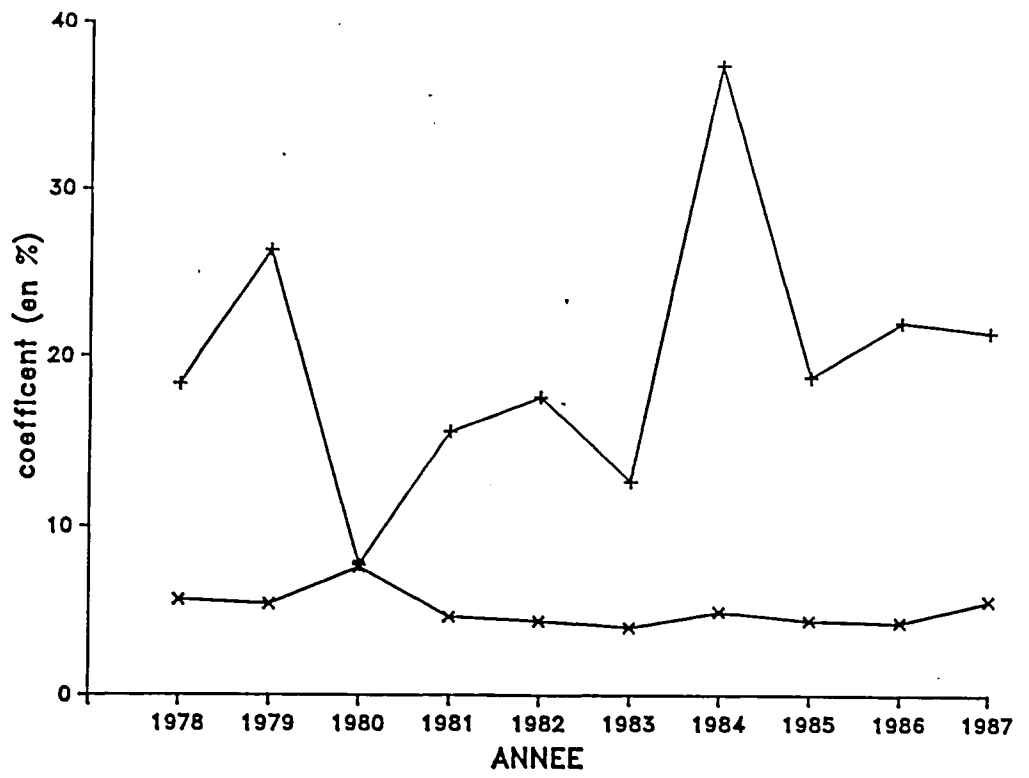


Figure 6 - Coefficients de variation des estimations d'abondance présentées sur la figure 4.