

IBM-PC Windows

EVA

logORSTOM

EVA

Estimation

VARiance

Logiciel de géostatistique
pour l'analyse structurale
et les calculs de variance

ORSTOM
éditions

Pierre PETITGAS
Arnaud PRAMPART

EVA

Estimation VARIance

Logiciel de géostatistique

pour l'analyse structurale
et les calculs de variance

Pierre PETIGAS
Géostatisticien marin

Arnaud PRAMPART
Informaticien

Éditions de l'Orstom

INSTITUT FRANÇAIS DE RECHERCHE SCIENTIFIQUE
POUR LE DÉVELOPPEMENT EN COOPÉRATION

Collection logORSTOM

Paris, 1995

*La collection LogOrstom est
consacrée aux programmes
et logiciels expérimentaux
élaborés dans le cadre
des recherches de l'Orstom.*

*Les critères importants
de leur sélection sont
la pertinence scientifique,
le multilinguisme, la fiabilité de
fonctionnement et la qualité
de la documentation.*

C o m i t é d e l e c t u r e

Nicolas Bez	Jean-Olivier Job
Pascal Boivin	Francis Laloë
Daniel Cardon	Jean-Jacques Lechaue
Gérard Cochonneau	Jean-Yves Loyer
Daniel Corre	Jean-François Nouvelot
Bernard de Mérona	Jacques Rivoirard
Jean Déjardin	Dominique Rémy
Laurent Delache	Patrick Séchet*
Laurent Drapeau	Francis Sondag
Jean-Marc Écoutin	Marc Souris
Jocelyne Ferraris	Michel Voltz
Marc Fournier	
Francis Gerlotto	* <i>Directeur</i>
Jacques Guiscafré	<i>de la collection Logorstom</i>

Fabrication-coordination : Catherine Plasse

Maquette de couverture : Michèle Saint-léger

« Par dérogation au 2° de l'article L. 122-5 lorsque l'œuvre est un logiciel, toute reproduction autre que l'établissement d'une copie de sauvegarde par l'utilisateur ainsi que toute utilisation d'un logiciel non expressément autorisée par l'auteur ou ses ayants droit, ou ayants cause, est illicite ». Art. L. 122-6 de la loi du 1^{er} juillet 1992 (code de la propriété intellectuelle, partie législative).

Le code de la propriété intellectuelle n'autorisant, aux termes des alinéas 2 et 3 de l'article L. 122-5, d'une part, que les « copies ou reproductions strictement réservées à l'usage privé du copiste et non destinées à une utilisation collective », et, d'autre part, que les analyses et les courtes citations dans un but d'exemple et d'illustration, « toute représentation ou reproduction intégrale, ou partielle, faite sans le consentement de l'auteur ou de ses ayants droit ou ayants cause, est illicite » (alinéa 1^{er} de l'article L. 122-4).

Cette représentation ou reproduction, par quelque procédé que ce soit, constituerait donc une contrefaçon passible des peines prévues au titre III de la loi précitée.

E

P. Petitgas, A. Prampart

V

A
© Orstom, 1995

Rappels
méthodologiques :

ce que fait EVA

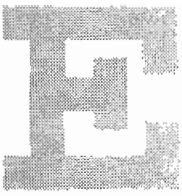
I

On rassemble ici les principes généraux de deux grandes catégories de méthodes qui permettent d'estimer les quantités totales à partir de campagnes d'échantillonnage. L'une relève de la théorie de l'échantillonnage aléatoire et de l'utilisation des calculs sur les variables aléatoires, l'autre est l'application géostatistique de la théorie des fonctions aléatoires. On replace dans ce cadre les calculs que permet d'effectuer EVA. L'estimation par d'autres approches, en particulier par simulation de processus, n'est pas abordée.

Toute estimation s'effectue explicitement ou implicitement sur une zone donnée, définie avant ou après l'échantillonnage. La quantité totale estimée et la précision de cette estimation dépendent de la définition de cette zone.

EVA traite de l'estimation globale sur une zone définie par l'utilisateur, lorsque l'effort d'échantillonnage est réparti de façon homogène sur cette zone. On entend par là que les échantillons peuvent être agrégés mais qu'il n'y a pas d'orientation systématique de l'effort dans une sous-zone donnée.

Tout suréchantillonnage dans certaines sous-zones de la zone à estimer introduira un biais dans l'estimation que EVA peut fournir, car ces sous-zones seraient surreprésentées dans le jeu de données. Dans un tel cas, on pourra utiliser EVA de deux manières. Soit, on veut faire l'estimation à partir du jeu global et alors il est nécessaire de réduire cette surreprésentation en travaillant sur des moyennes locales des données. Soit, on veut séparer les sous-zones et faire une estimation distincte dans chacune. On pourra alors spécifier les variations des paramètres structuraux d'une sous-zone à l'autre, lors de la modélisation du variogramme (effets proportionnels...).



ESTIMATEURS DE LA MOYENNE SUR ZONE

On cherche à estimer une Intégrale d'espace sur la zone V , à partir d'échantillons répartis de façon aléatoire ou non. La moyenne sur zone à estimer est :

$$M_v = \frac{1}{V} \int_V f(x) dx$$

où $f(x)$ est la densité au point x . On pourra trouver dans SIMMONDS *et al.* (1992) une revue des différents estimateurs.

Estimateurs linéaires

La moyenne arithmétique des données est un estimateur sans biais de la moyenne sur la zone, lorsque l'effort est réparti de façon homogène sur toute la zone à estimer. Cet estimateur ne suppose aucune hypothèse sur la distribution statistique des données (= histogramme = *probability distribution function* = pdf). Reprenant la classification proposée par SMITH (1990), la moyenne arithmétique est un *design based estimator*, toujours calculable et toujours valide sans hypothèse statistique pourvu que l'effort soit homogène, par opposition aux *model based estimators* qui supposent que le phénomène sous-jacent générant les données soit compatible avec le modèle inféré.

Le krigeage de la moyenne sur la zone est une moyenne pondérée des données. À chaque échantillon est attribué un poids qui est fonction de sa position et du modèle de structure spatiale. Le krigeage de

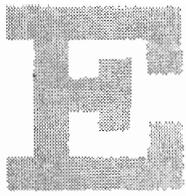
Rappels méthodologiques :
ce que fait EVA

la moyenne sur la zone peut être qualifié de *spatial model based estimator*. Il est surtout intéressant lorsque les échantillons sont agrégés et peu nombreux (disons inférieurs à la centaine) et que les valeurs sont corrélées. Intuitivement, on conçoit bien que des échantillons proches formant un paquet apporteront des informations redondantes. Le krigeage leur donnera à chacun un poids plus faible dans la moyenne globale qu'à une information isolée. Quand le nombre d'échantillons est grand (disons supérieur à la centaine), seuls quelques échantillons ont des poids de krigeage différents de $1/n$ et la pondération ne joue plus beaucoup de rôle dans l'estimation (PETITGAS et RIVOIRARD, 1991).

La moyenne arithmétique et le krigeage sont des estimateurs linéaires : ils estiment la moyenne sur zone par une moyenne pondérée des données où les poids ne dépendent pas de la valeur des échantillons.

Estimateurs non linéaires

Un estimateur non linéaire est une moyenne pondérée des données où les poids dépendent de la valeur des échantillons (des quantiles auxquels appartiennent les échantillons). Un estimateur non linéaire dépend du modèle d'histogramme choisi ; c'est un *pdf model based estimator*. Il peut être intéressant lorsque l'histogramme des échantillons est dissymétrique car il est alors plus précis que la moyenne arithmétique, mais cette précision est conditionnelle au modèle utilisé. Intuitivement, on conçoit bien que la moyenne arithmétique est très influencée par les fortes valeurs alors que celles-ci auront un poids plus faible que $1/n$ dans l'estimation non linéaire, car elles ont une faible fréquence d'apparition sur la loi de probabilité (le modèle d'histogramme). Toutefois, l'estimation non linéaire peut être fortement biaisée dans la mesure où il n'est pas facile en général d'ajuster



P. Petitgas, A. Prampart



© Orstom, 1995

I-5

convenablement, à la fois dans les faibles et les fortes valeurs, un modèle d'histogramme.

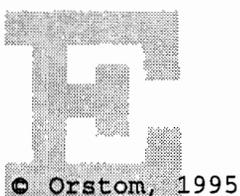
PENNINGTON (1983) propose un estimateur de la loi lognormale pour les données de chalutage et d'ichtyoplancton qui présentent un certain nombre de zéros. SMITH (1988) discute des biais et de la non-robustesse de cet estimateur par rapport aux variations du modèle. MACLENNAN et MACKENZIE (1988), pour les données d'acoustique, proposent d'utiliser les transformations puissance de Box-Cox afin de choisir un modèle d'histogramme adapté aux données, puis d'en déduire l'estimateur de sa moyenne. Théoriquement, ces estimateurs ne sont valides que pour des données non corrélées spatialement.

Dans la catégorie des estimateurs linéaires, EVA ne fournit que la moyenne arithmétique. Dans la catégorie des estimateurs non linéaires, EVA fournit un estimateur lognormal géostatistique (MATHERON, 1974). C'est l'estimateur de la moyenne sur la zone V pour des données lognormales et corrélées. On pourrait le qualifier de *structural and pdf model based estimator*. Bien qu'intéressante, son utilisation reste dangereuse à cause des écarts à la lognormalité.

ESTIMATEURS DE LA VARIANCE D'ESTIMATION

On s'intéresse maintenant à calculer la précision des estimations précédentes. Celle-ci dépend du plan d'échantillonnage effectué et de l'autocorrélation entre les échantillons. Les variances calculées ne rendent pas compte des erreurs de mesure ni des biais lors de l'échan-

Rappels méthodologiques :
ce que fait EVA

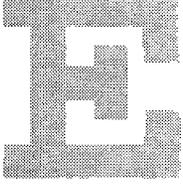


tillonnage. Ces « variabilités » restent dans les données et sont comptabilisées dans la structure. Le problème abordé est celui de l'erreur commise lorsque l'on estime une intégrale d'espace (la quantité totale) à partir d'informations fragmentaires (les échantillons).

Plan aléatoire : statistique classique et géostatistique

Dans la théorie de l'échantillonnage aléatoire, on mesure une variable en des points dont la position est aléatoire sur une zone V . Cette zone joue le rôle d'une urne. Chaque échantillon est interprété comme le résultat d'un tirage au hasard dans l'urne, et toute la distribution spatiale se trouve résumée par la loi de probabilité des valeurs dans V (histogramme régional). On dispose, quelle que soit la distribution spatiale, de n informations indépendantes non localisées. Chaque échantillon est un estimateur de la moyenne de la loi de probabilité des valeurs. La variance associée à un tel plan est un *design based estimator*. Toutefois, géostatistique et statistique classique donnent nécessairement les mêmes valeurs numériques de variance. En effet, les formules géostatistiques correctes (MATHERON, 1971) ne font intervenir que la variance de dispersion sur toute la zone. Le modèle étant bien calé, celle-ci doit égaler la variance de la population, c'est-à-dire la variance calculée en $1/(n-1)$.

La variance d'estimation pour le plan aléatoire n'est pas fournie dans EVA mais les éléments pour la calculer sont fournis. Dans ce logiciel, les positions des échantillons sont considérées fixes et non aléatoires. EVA fournit des variances pour les plans à maille régulière, c'est-à-dire plans systématiques, et les plans non orthodoxes (selon la théorie de l'échantillonnage aléatoire). On divisera par n le G_{VV} fourni par EVA



(voir plus loin), afin d'avoir la variance d'estimation que donnerait un plan aléatoire pour un même nombre d'échantillons.

Autres plans : géostatistique transitive et intrinsèque

Lorsque les échantillons ne sont pas prélevés indépendamment les uns des autres et que la distribution spatiale échantillonnée est structurée, le calcul de toute variance, et en particulier de la variance d'estimation, requiert la prise en compte de la structure spatiale au travers d'un modèle. Quel que soit le type d'échantillonnage, dès que l'on a un modèle structural, on peut calculer des variances. On pourra trouver les justifications théoriques dans MATHERON (1971), COCHRAN (1977) ou JOURNEL et HUIJBREGTS (1978). Les estimateurs de variance proposés ici sont donc des *spatial model based estimators*.

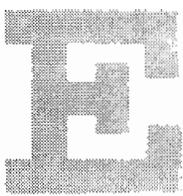
Pour effectuer ces calculs, deux méthodes géostatistiques sont à disposition : la méthode intrinsèque et la méthode transitive. Cette dernière est la théorie de l'estimation globale à l'aide de mailles régulières dont l'origine est aléatoire, c'est-à-dire plans systématiques de la terminologie de COCHRAN (1977). La méthode transitive ne fait appel à aucune hypothèse statistique sur la distribution spatiale échantillonnée. Le covariogramme transitif en est l'outil structural. Dans la méthode transitive, seules sont prises en compte les valeurs strictement positives. Dans cette méthode, c'est la quantité totale qui est estimée, non la moyenne. La variance d'estimation de la méthode transitive a la signification d'une variance de l'estimateur (pour toutes les implantations possibles de la grille). La méthode transitive peut se comprendre comme une extension de la théorie de l'échantillonnage aléatoire au cas des mailles régulières.

Rappels méthodologiques :
ce que fait EVA

La méthode intrinsèque est une application de la théorie des fonctions aléatoires à l'estimation locale (cartographie) et globale. Dans cette approche, on interprète la réalité comme l'une des réalisations possibles d'une fonction aléatoire que l'on observerait à l'intérieur d'un polygone donné, c'est-à-dire la zone à estimer. Les modèles intrinsèques ont une double propriété : leurs distributions spatiales possèdent un certain taux d'homogénéité (stationnarité d'ordre k), et elles ne sont pas influencées par les bordures de la zone (pas d'effet de bord). Le variogramme est la fonction structurale du modèle intrinsèque d'ordre zéro. Dans les modèles intrinsèques, la moyenne sur la zone à estimer (intégrale d'espace) est une variable aléatoire : elle varie en effet à chaque réalisation. La variance d'estimation a la signification d'une variance d'interpolation : c'est la variance de l'erreur commise à chaque réalisation, entre la moyenne sur zone et son estimation. La définition des limites de la zone à estimer est indispensable au calcul de la variance d'estimation.

Quand il n'y a pas d'effet de bord, les méthodes intrinsèque et transitive sont équivalentes et fournissent les mêmes valeurs numériques de variance (MATHERON, 1965).

EVA effectue les calculs des variances d'estimation associées aux estimateurs précédents de la moyenne, par la méthode intrinsèque d'ordre zéro, pour toute configuration d'échantillonnage à effort homogène et points non aléatoires. Dans le cas des campagnes acoustiques effectuées le long de transects parallèles et équidistants, EVA permet d'utiliser la méthode transitive sur les valeurs cumulées le long des transects. On ne travaille alors plus qu'à une dimension de façon très simple. Cette méthode aboutit aux mêmes résultats que si l'estimation était effectuée à deux dimensions avec un variogramme (MATHERON, 1965 ; PETITGAS, 1993).



P. Petitgas, A. Prampart



© Orstom, 1995

I-9

DIFFÉRENTES VARIANCES : VARIANCE DE DISPERSION, VARIANCE DE L'ESTIMATEUR, VARIANCE D'ESTIMATION

■ Méthode intrinsèque

La variance sur une zone est une fonction de l'autocorrélation entre toutes les valeurs sur cette zone. On appelle variance de dispersion sur la zone V , la variance de la population contenue dans V . On appelle variance d'estimation, la variance de l'erreur entre la moyenne sur V et son estimateur. Les variances sont calculées avec le modèle de variogramme Inféré.

La variance de dispersion sur V sera notée G_{VV} . C'est la moyenne du variogramme pour toutes les distances dans V . Elle équivaut à la variance des données calculée en $1/(n - 1)$, pour un plan aléatoire dans V . Quand les données sont corrélées, cette variance n'est pas un bon estimateur de la variance de la population dans V .

MATHERON (1971) suggère de se servir du G_{VV} pour caler le modèle de variogramme à la variance dans V . Le palier du variogramme (s'il y en a un) n'est pas nécessairement égal à la variance dans V : si la portée est petite, relativement aux dimensions de V , le palier sera très proche de la variance dans V ; dans le cas contraire, le palier sera logiquement supérieur à la variance dans V .

La moyenne du variogramme pour toutes les distances entre les échantillons sera notée G_{ab} . La variance de l'estimateur est : $G_{VV} - G_{ab}$. Elle a la signification physique d'une mesure de la variabi-

Rappels méthodologiques :
ce que fait EVA



lité entre estimateurs obtenus pour le même réseau rigide d'échantillons mais implantés en différents endroits de la zone V.

Pour une estimation linéaire par la moyenne arithmétique d'échantillons non aléatoires, la variance d'estimation sur V se note :

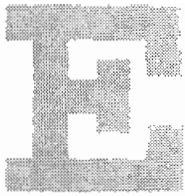
$$\sigma_E^2 = 2G_{aV} - G_{VV} - G_{ab}$$

La variance d'estimation dépend de trois termes qui sont des moyennes du variogramme. Le terme G_{VV} (variance sur V) prend en compte la forme de la zone V. Le terme G_{ab} prend en compte la forme du réseau rigide d'échantillonnage (position relative des échantillons entre eux). Le terme G_{aV} est la moyenne du variogramme pour toutes les distances entre successivement chaque échantillon et tous les points de V. Le terme G_{aV} prend en compte la disposition du réseau des échantillons dans la zone V à estimer.

Dans une estimation non linéaire, la variance d'estimation présente aussi trois termes qui ont les mêmes significations physiques.

EVA permet de calculer les trois types de variance : variance sur V, variance de l'estimateur, variance de l'erreur d'estimation.

EVA permet de tester les performances de différents plans d'échantillonnage. Dans les formules de variance, seuls interviennent la zone sur laquelle on veut faire l'estimation, les positions des échantillons, le modèle de variogramme. Les valeurs des échantillons n'interviennent pas dans le calcul de variance puisque la variance est une fonction de l'autocorrélation entre les points. Par contre les valeurs interviennent dans l'inférence du modèle d'autocorrélation spatial. Pour tester un plan d'échantillonnage, il suffira de générer la position des échantillons pour ce nouveau plan, et de calculer la variance d'estimation.



Pour mesurer l'influence sur l'estimation de l'indétermination des limites de la zone, on pourra calculer la variance d'estimation pour différentes hypothèses de limites de zone.

I Méthode transitive : cas des grilles régulières

Le calcul de la variance d'estimation par la méthode transitive est proposé dans EVA à une dimension, pour les valeurs cumulées le long des transects des campagnes acoustiques.

Dans cette méthode à une dimension, la quantité à estimer est :

$$Q = \int_{-\infty}^{+\infty} q(x) dx$$

où $q(x)$ représente le cumul sur le transect de coordonnée x . En dehors des limites de la zone, les transects étant nuls, l'estimateur de Q est une somme discrète finie :

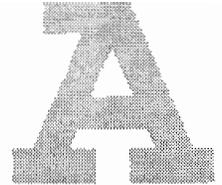
$$\hat{Q} = a \sum_{k=1}^n q(x_0 + ka)$$

où a est la distance intertransects, où x_0 est la coordonnée origine du premier transect et n le nombre de transects.

La variance d'estimation s'écrit :

$$\sigma_{\hat{Q}}^2 = a \sum_{k=-\infty}^{+\infty} g(ka) - \int_{-\infty}^{+\infty} g(h) dh$$

où a est la distance intertransects et g est le modèle structural (covariogramme transitive). La variance d'estimation est la différence entre l'intégrale de la fonction structurale et son approximation grâce à la



grille d'échantillonnage. Plus la maille sera fine et plus la distribution spatiale sera régulière, plus la variance d'estimation sera donc faible. La variance d'estimation ne dépend que de la maille de la grille et du modèle de covariogramme transitif.

L'hypothèse implicite de la méthode est que toute la zone est couverte par la grille : au-delà de la grille, les valeurs sont nulles. Les limites de la zone tombent donc entre deux points de la grille. Les limites de la zone sont donc indéterminées à la maille près. Cette indétermination est comptabilisée dans la formule ci-dessus.

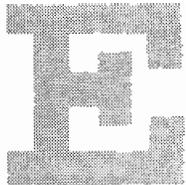
EVA permet facilement de construire la courbe de l'évolution de la variance d'estimation en fonction de la distance intertransects. Il suffit de faire varier ce paramètre et de recalculer la variance.

FONCTIONS STRUCTURALES EXPÉRIMENTALES

On trouvera dans MATHERON (1971), JOURNEL et HUIJBREGTS (1978), ARMSTRONG *et al.* (1992) tous les éléments relatifs aux calculs et aux modèles de covariance, variogramme, et covariogramme transitif.

▮ Variogramme expérimental

Le variogramme mesure l'écart quadratique moyen entre deux points séparés d'une distance vectorielle h . C'est une mesure moyenne sur toute la zone, ce qui suppose une certaine homogénéité. Le variogramme a la signification d'une mesure de variance entre points dis-



tants de h , quelle que soit la position de ces points dans la zone. Si le variogramme croît avec la distance, c'est qu'il y a une structure (corrélation spatiale) : deux points proches se ressemblent plus que deux points éloignés.

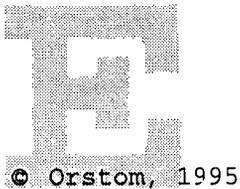
Le calcul expérimental s'effectue ainsi. On considère tous les couples entre les échantillons de la zone (ou de la sous-zone s'il a fallu stratifier). Pour chacun, on calcule la direction et la norme du vecteur que les deux échantillons définissent. On calcule le carré de la différence entre leurs valeurs, c^2 . On regroupe ensuite les angles par classes d'angles. Dans chaque classe d'angle, on regroupe les distances par classes de distance. Dans chaque classe de distance, on calcule la demi-moyenne des c^2 : cette demi-moyenne, c'est le variogramme. Le calcul expérimental requiert donc le choix de paramètres. On choisit les classes de façon à calculer la moyenne des c^2 sur un nombre suffisant de couples.

$$\gamma^*(u, h) = \frac{1}{2n(u, h)} \sum_i (f(x_i) - f(x_i + h))^2$$

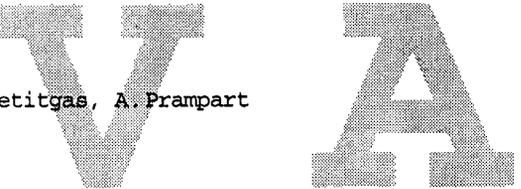
avec $f(x_i)$ la valeur mesurée en x_i et $n(u, h)$ le nombre de couples d'échantillons pour la direction u et la distance h .

On définit les classes d'angle en définissant des portions de plan, par le choix d'un angle et d'une tolérance, $\pm \text{tolu}$, autour de cette angle. Dans EVA, les angles sont comptés dans le sens trigonométrique, par rapport à l'horizontale. Soit u l'angle choisi et tolu la tolérance. À partir de chaque échantillon, x , tout autre échantillon, y , sera utilisé pour le calcul du variogramme dans la direction u si le point, y , se trouve dans la portion de plan centré sur x et défini par $[u - \text{tolu}, u + \text{tolu}]$ et $[u + \pi - \text{tolu}, u + \pi + \text{tolu}]$.

Rappels méthodologiques :
ce que fait EVA



P. Petitgas, A. Prampart



EVA propose un calcul isotrope (aucune direction n'est différenciée), un calcul anisotrope dans deux directions (u et $u + \pi/2$) et un calcul anisotrope dans quatre directions (u , $u + \pi/4$, $u + \pi/2$, $u + 3\pi/4$).

On définit les classes de distance en choisissant un pas, p , et une tolérance, $\pm \text{tolp}$, autour. Le variogramme est calculé pour n pas successifs. La classe i contient toutes les distances appartenant à l'intervalle $]i p - \text{tolp}, i p + \text{tolp}]$. La valeur du variogramme est centrée sur la distance $i p$. La classe $]0, \text{tolp}]$ est individualisée dans le calcul effectué par EVA, ce qui permet d'utiliser les réplicats pour estimer le variogramme à l'origine. Dans cette classe, la valeur du variogramme est centrée sur la moyenne des distances appartenant à la classe.

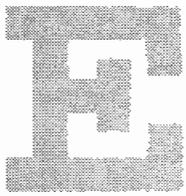
Covariogramme transitif expérimental

Le covariogramme transitif est une sorte de covariance non centrée dans laquelle la géométrie de la zone intervient. Il n'est facilement calculable que pour les grilles régulières. À une dimension sur les cumuls par transect, il se calcule en faisant glisser la grille sur elle-même d'une distance h , multiple de la maille, et en sommant les produits $q(x_i)q(x_i + h)$.

$$g^*(ka) = a \sum_i q(x_i)q(x_i + ka)$$

où a est la maille (distance intertransects) et où les $q(x)$ représentent les cumuls par transect.

Le covariogramme transitif est fonction des valeurs $q(x)$, mais aussi du nombre de produits sommés pour chaque distance ka . Il s'annule pour les distances supérieures à la longueur de la grille.



P. Petitgas, A. Prampart



© Orstom, 1995

I-15

MODÈLES STRUCTURAUX

Les variances sont calculées en utilisant des modèles structuraux. Ceux-ci doivent satisfaire certaines conditions pour des raisons de cohérence (MATHERON, 1971). La gamme des modèles d'usage courant (JOURNEL et HUIJBREGTS, 1978) permet de rendre compte de propriétés physiques des distributions spatiales. Ces propriétés sont caractérisées par le comportement du variogramme aux petites distances (i.e. à l'origine pour $h = 0$) et aux grandes distances (i.e. existence ou non d'un palier).

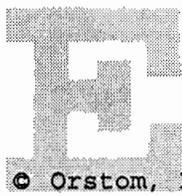
Il n'y a pas de différence fondamentale dans les propriétés des modèles de variogramme et de covariogramme transitifs. Toutefois la méthode transitive n'explicitant pas d'hypothèse probabiliste, les covariogrammes présentent toujours un palier puisqu'ils s'annulent pour les distances supérieures aux limites de la grille d'échantillonnage.

Modèles à palier, modèles sans palier

Les modèles à palier montrent une saturation (palier) de la variance quand la distance croît. L'existence d'un palier veut dire qu'il existe une variance dans le modèle : c'est le palier. Quand on choisit un modèle à palier, on interprète les données comme issues d'un modèle sous-jacent stationnaire (une fonction aléatoire stationnaire) dont la variance est finie. Le palier égale la variance de la fonction aléatoire. Il n'est pas toujours égal à la variance que montre cette fonction sur une zone donnée.

Le G_{VV} est la variance totalisée par la fonction aléatoire sur une zone V . Le palier est la variance de la fonction. Quand il n'y a pas de corrélations spatiales (variogramme plat), le G_{VV} égale le palier.

Rappels méthodologiques :
ce que fait EVA



Le palier est associé à un autre paramètre, la portée. C'est la distance à partir de laquelle le palier est atteint. Deux points séparés par une distance supérieure à la portée sont indépendants, alors que deux points séparés d'une distance inférieure à la portée sont corrélés. La portée donne une mesure quantitative de la notion de zone d'influence. C'est en gros le diamètre moyen des agrégats de la régionalisation.

Les modèles sans palier sont relatifs aux fonctions aléatoires qui ne possèdent pas de variance (variance infinie) : plus la distance croît, plus la variance entre les points croît, sans limite. Toutefois, sur une zone donnée, on peut toujours calculer la variance : c'est la moyenne du variogramme pour les distances mises en jeu sur cette surface.

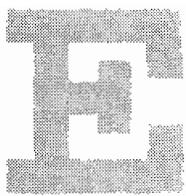
Comportement aux petites distances

Considérons que la distribution spatiale échantillonnée est une surface. Près de l'origine $h = 0$, le comportement du modèle de variogramme rend compte de sa rugosité.

Les comportements à tangente verticale rendent compte d'une très grande rugosité ; au contraire, ceux à tangente horizontale rendent compte d'une très grande régularité. Entre les deux, les comportements linéaires rendent compte d'une rugosité « moyenne ».

Pépité

La pépité est une discontinuité d'une amplitude C_0 à l'origine sur le variogramme ($\gamma(0) = 0$; $\gamma(h) = C_0 > 0$ dès que $h > 0$), comme sur le covariogramme ($g(0) = C_1$; $g(h) = C_1 - C_0$ dès que $h > 0$).



P. Petitgas, A. Prampart



C'est une composante totalement aléatoire de la distribution spatiale dont la variance est C_0 . La pépite a trois interprétations physiques indissociables dans une étude pratique. Elle prend en charge une composante aléatoire pure de la régionalisation, et/ou elle prend en charge une erreur de mesure additive, indépendante des valeurs et aléatoire pure, et/ou elle prend en charge des structures dont les portées sont plus faibles que la maille d'échantillonnage.

Plus d'éléments sur les modèles peuvent être trouvés dans MATHERON (1971), JOURNEL et HUIJBREGTS (1978), ARMSTRONG *et al.* (1992). Le choix d'un modèle est une caractérisation : on met dans le modèle ce que l'on pense du phénomène sous-jacent. On choisit un type de modèle pour ses propriétés aux grandes et aux petites distances. C'est pourquoi dans EVA, il n'y a pas de modélisation automatique : l'utilisateur doit choisir son type de modèle puis les paramètres de ce modèle. L'ajustement se fait ensuite de visu.

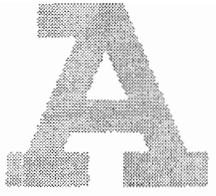
La variance d'estimation dépend surtout du comportement à l'origine du modèle de variogramme et du comportement aux grandes distances (MATHERON, 1971, 1978). Des modèles un peu différents mais ayant des comportements similaires donneront des variances d'estimation similaires (MATHERON, 1978) : la variance d'estimation est sur ce point assez robuste.

Modèles

Un modèle peut être la somme de différents modèles emboîtés, si les données permettent de déceler des structures à différentes échelles. EVA autorise au maximum un effet de pépite et trois modèles emboîtés.

EVA propose les modèles du tabl. 2. Par extension, pour tous ces modèles, le paramètre C est appelé palier et le paramètre r est appelé portée.

Rappels méthodologiques :
ce que fait EVA



Les modèles exponentiels et gaussiens n'atteignent leur palier que asymptotiquement. JOURNEL et HUIJBREGTS (1978) ont défini pour ces modèles une « portée pratique », r' : c'est la distance à partir de laquelle 95 % du palier C est atteint. On a donc : pour l'exponentiel, $r = r' / 3$ et pour le gaussien, $r = r' / \sqrt{3}$. À partir de r' , les courbes marquent une inflexion très nette. Dans EVA, c'est la valeur de r' qui est demandée.

Le covariogramme transitif triangle est le modèle de covariogramme à une dimension pour un segment.

TABLEAU 1 -
Propriétés
des modèles
de variogramme.

Type de modèle	Palier	Comportement à l'origine	Rugosité modélisée
Sphérique	Oui	Linéaire	Moyenne
Exponentiel	Asymptotique	Linéaire	Moyenne
Gaussien	Asymptotique	Parabolique (tangente horizontale)	Très lisse
Triangle (1D uniquement)	Oui	Linéaire	Moyenne
Puissance (h^α)	$0 < \alpha < 1$ Non	Tangente de plus en plus verticale quand α tend vers 0	Très rugueux
	$\alpha = 1$ Non	Linéaire	Moyenne
	$1 < \alpha < 2$ Non	Tangente de plus en plus horizontale quand α tend vers 2	Lisse

Rappels méthodologiques :
ce que fait EVA

Type de modèle	Variogramme (2 D) : $\gamma(h) =$	Variogramme transitif (1 D) : $g(h) =$
Sphérique	$C(1.5(h /r) - 0.5(h /r)^3)$ <i>si</i> $ h \leq r$ C <i>si</i> $ h \geq r$	$C(1 - 1.5(h /r) + 0.5(h /r)^3)$ <i>si</i> $ h \leq r$ 0 <i>si</i> $ h \geq r$
Exponentiel	$C(1 - \exp(- h /r))$	$C \exp(- h /r)$
Gaussien	$C(1 - \exp(-h^2/r^2))$	$C \exp(-h^2/r^2)$
Triangle	Non disponible	$C(1 - h /r)$ <i>si</i> $ h \leq r$ 0 <i>si</i> $ h \geq r$
Puissance	Ch^a	Non disponible

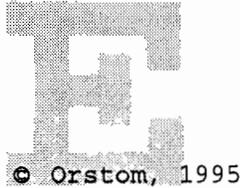
TABLEAU 2 -
Formules
des modèles
de variogramme
et covariogramme
transitif.

Anisotropies

Anisotropie signifie que toutes les directions de l'espace ne sont pas équivalentes. Un variogramme anisotrope modélise le fait que la variabilité a des caractéristiques différentes dans différentes directions. JOURNEL et HUIJBREGTS (1978) reconnaissent deux types d'anisotropie, géométrique et zonale, par référence aux modèles à paller.

Dans l'anisotropie géométrique, le paller est le même dans toutes les directions, mais la portée varie de façon elliptique en fonction de la direction. Physiquement, cela veut dire que toutes les directions sont autant hétérogènes les unes que les autres, mais que la forme des agrégats est elliptique plutôt que d'être circulaire comme dans le cas isotrope. Dans l'anisotropie zonale, la variance de la fonction (i.e. le paller) est une fonction de la direction : le phénomène est plus hétérogène dans certaines directions. Les deux anisotropies peuvent se combiner.

Rappels méthodologiques :
ce que fait EVA

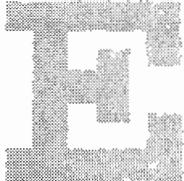


L'anisotropie géométrique est paramétrée par un angle de rotation, un coefficient d'anisotropie X pour cette direction et un coefficient Y pour la direction perpendiculaire. L'angle définit la rotation qu'il faut effectuer pour amener l'axe des abscisses sur le grand axe (ou le petit) de l'ellipse des portées. Les coefficients marquent la dilatation ou la contraction des distances sur les deux axes principaux de l'ellipse qu'il faut effectuer pour transformer l'ellipse en cercle. Par exemple, si on suppose que dans la direction 45°, la portée est deux fois plus longue que dans la direction perpendiculaire, la rotation est de 45° et les coefficients sont : 1 pour X et 2 pour Y ou bien 0.5 pour X et 1 pour Y.

L'anisotropie zonale est modélisée par un modèle à structures emboîtées. Dans une direction, on a un modèle de moins. Pour ce faire, il suffit d'annuler le coefficient d'anisotropie dans cette direction. Pour éviter les sauts brusques, il est préférable d'éviter un modèle constitué de deux modèles emboîtés où chacun ne vaudrait que dans une portion du plan. JOURNEL et HUIJBREGTS (1978) recommandent de prendre un modèle où ni X ni Y ne sont nuls et d'y ajouter d'autres modèles où soit X soit Y est nul.

INTERPRÉTATION PROBABILISTE D'UNE RÉGIONALISATION : MOYENNE ET VARIANCE SUR ZONE, MOYENNE ET VARIANCE DE LOI

Quand on interprète la distribution spatiale échantillonnée comme l'une des réalisations d'une fonction aléatoire que l'on observerait dans

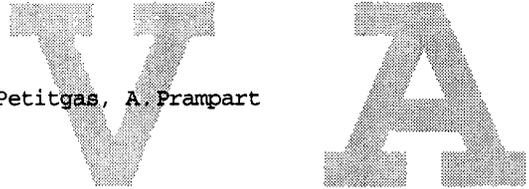


une fenêtre V , la moyenne sur V pour une réalisation particulière n'est pas nécessairement égale à la moyenne de la fonction aléatoire.

La moyenne arithmétique des données est un estimateur sans biais des deux quantités. Par contre, la variance d'estimation de cet estimateur pour l'une ou l'autre des deux quantités n'est pas la même. Tout autre estimateur de la moyenne estimera, dans le modèle de fonction aléatoire, soit l'une soit l'autre des quantités. Pour l'estimation lognormale, l'option prise dans EVA est d'estimer la moyenne sur zone pour la réalisation échantillonnée. Les éléments pour calculer la moyenne de la loi sont néanmoins fournis. Mais, la variance calculée dans EVA ne correspond pas à cet estimateur. La variance d'estimation correspond à la moyenne, sur toutes les réalisations, de l'erreur entre la moyenne sur V à chaque réalisation et son estimation. La variance d'estimation est un indicateur de précision de l'estimateur de la moyenne sur V . C'est la variance d'estimation qui est calculée dans EVA.

Le variogramme expérimental est relatif à la réalisation échantillonnée sur V (variogramme régional). L'opération de modélisation consiste à choisir un modèle moyen qui a la signification de la moyenne des variogrammes expérimentaux que d'autres échantillonnages sur cette zone nous permettraient de calculer. La variance sur la zone V déduite du variogramme n'est donc pas relative dans le modèle à une réalisation particulière, mais à la variabilité moyenne de la fonction aléatoire sur la zone V , pour toutes les réalisations. La variance sur V est égale au G_V alors que la variance de la loi (pdf) est égale au palier du variogramme. Considérons un modèle dont le variogramme n'a pas de palier. La loi (pdf) de ce modèle n'a pas de variance, elle est infinie. Le théorème central limite ne s'applique plus : on ne peut pas définir d'intervalle autour de la moyenne qui contienne un pourcentage donné, $t\%$, des valeurs. Mais, on peut toujours calculer la variance sur une zone donnée grâce au variogramme.

Rappels méthodologiques :
ce que fait EVA



Dans la théorie de l'échantillonnage aléatoire, la zone V n'est pas une fenêtre dans laquelle on observe une fonction : la zone V est identifiée à tout l'univers et on ne fait pas la distinction entre les paramètres statistiques de la loi et sur la zone.

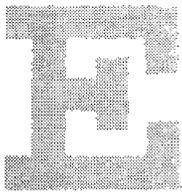
Le paramètre clé est la taille relative de la portée dans la zone V : c'est lui qui détermine la qualité statistique de l'inférence du modèle (MATHERON, 1978) comme les écarts numériques entre les paramètres sur zone et ceux du modèle. Plus la portée est petite relativement aux dimensions de V , meilleure est l'inférence statistiquement et plus ces écarts se réduisent. Toutefois, au terme d'une réflexion épistémologique, MATHERON (1978) montre que passer par un modèle de fonction aléatoire est un moyen méthodologique qui ne sert qu'à expliciter l'opération de modélisation du variogramme. Les paramètres de moyenne et variance de la fonction aléatoire restent conventionnels, ce sont les grandeurs régionales (i.e. paramètres sur V) qui ont un sens pratique et qui nous intéressent dans les calculs (MATHERON, 1978). Ainsi, pour la modélisation du variogramme, le comportement à l'origine et la valeur du G_{VV} sont des paramètres plus importants que la valeur explicite du prier, car c'est bien la régionalisation dans V qu'il appartient de caractériser.

ALGORITHMES DE CALCUL DES VARIANCES

■ Méthode transitive

La formule à une dimension de la variance d'estimation :

Rappels méthodologiques :
ce que fait EVA



$$\sigma_{\rho}^2 = a \sum_{k=-\infty}^{+\infty} g(ka) - \int_{-\infty}^{+\infty} g(h)dh, \text{ est calculée directement.}$$

Les modèles de covariogramme utilisés ont des primitives évidentes et leurs intégrales ont donc été calculées formellement. Pour les modèles à palier, la portée constitue les bornes de la sommation et de l'intégration : le calcul est direct et sans approximation. Pour les modèles à palier asymptotique, la sommation et l'intégration infinies sont tronquées grâce à l'approximation suivante. On a considéré que ces modèles présentaient une portée fictive égale à la distance pour laquelle 99,999 % du palier est atteint.

Méthode intrinsèque

VARIANCE D'ESTIMATION LORSQUE LA MOYENNE SUR ZONE EST ESTIMÉE PAR LA MOYENNE ARITHMÉTIQUE D'ÉCHANTILLONS CORRÉLÉS

La variance d'estimation s'écrit (MATHERON, 1971) :

$$\sigma_E^2 = 2G_{av} - G_{VV} - G_{ab}$$

Dans EVA, cette formule est calculée numériquement.

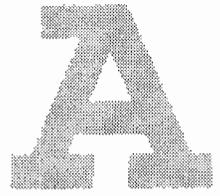
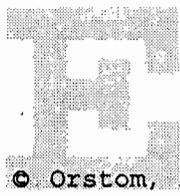
$$G_{ab} = \frac{1}{n^2} \sum_a \sum_b \gamma(x_a - x_b)$$

(a et b sont les indices sur les échantillons et x désigne une position)

$$G_{av} = \frac{1}{nV} \sum_a \int_V \gamma(x_a - y) dy$$

(x_a désigne chacun des échantillons, y désigne chacun des points de V)

Rappels méthodologiques :
ce que fait EVA



$$G_{VV} = \frac{1}{V^2} \int_V dx \int_V \gamma(x-y) dy$$

(x et y prennent toutes les positions de V indépendamment l'un de l'autre)

Pour calculer les sommations intégrales sur la zone V, celle-ci est discrétisée finement. Dans EVA, l'utilisateur définit lui-même la discrétisation. Le calcul de G_{VV} est effectué selon l'algorithme de la maille aléatoire stratifiée proposé par JOURNAL et HUIJBREGTS (1978) : x décrit chaque nœud de la grille de discrétisation de V alors que y a une position aléatoire dans chaque cellule de la grille.

MATHERON (1971) a développé des calculs approchés de la formule de σ_E^2 , dans le cas des grilles régulières et pour des structures isotropes. Les valeurs de la variance d'estimation peuvent être lues directement sur des abaques et le calcul fait à la main. Pour ces approximations, MATHERON (1971) montre qu'un modèle de variogramme pour les petites distances suffit (quasi-stationarité) : la variance d'estimation approximée ne dépend que du comportement à l'origine du variogramme. Dans le calcul numérique, la variance dépend aussi du comportement aux grandes distances et on devra inférer un modèle réaliste pour toutes les distances, pas uniquement pour les petites.

Le calcul des intégrales G_{VV} et G_{AV} est basé sur la discrétisation de la zone V. Comme la variance est une différence entre ces termes et le G_{Ab} , si la discrétisation n'est pas assez fine par rapport à la maille d'échantillonnage, on risque de trouver une variance négative. Il faut discrétiser la zone V de façon suffisamment fine pour obtenir une bonne estimation du G_{VV} , du G_{AV} , donc de la variance d'estimation.

EVA fournit la variance d'estimation σ_E^2 , le G_{VV} , le G_{Ab} . Comme la variance d'estimation est une fonction linéaire du variogramme, EVA



fournit aussi les contributions de chacun des modèles emboîtés à la variance σ_E^2 , en pourcentage.

La variance d'estimation que donnerait un plan aléatoire avec le même nombre d'échantillons peut être calculée : il suffit de diviser le G_{VV} par le nombre d'échantillons. Quand les échantillons ne sont pas aléatoires et que les données sont corrélées, la variance calculée en $1/(n-1)$ n'est pas un bon estimateur de la variance dans V. On peut tout aussi bien lui préférer le G_{VV} . Par contre, pour un plan aléatoire, le G_{VV} doit égaler la variance calculée en $1/(n-1)$.

MOYENNE SUR ZONE ET VARIANCE D'ESTIMATION SOUS L'HYPOTHÈSE DE LOGNORMALITÉ POUR DES DONNÉES CORRÉLÉES

EVA fournit les estimateurs proposés par MATHERON (1974). Le variogramme utilisé pour cette estimation est celui des Logarithmes des densités échantillonnées. Il n'y a pas de garde-fou dans EVA : l'estimation lognormale pourra se faire avec un tout autre variogramme. L'utilisateur devra être prudent.

La moyenne de la fonction aléatoire lognormale sur V est :

$M_V = \exp(m^a + G_{VV} / 2)$, où m^a est la moyenne sur V du Log de la fonction aléatoire et où G_{VV} est calculé avec le variogramme des Logs.

L'estimateur de M_V pour la réalisation échantillonnée et l'échantillonnage effectué est : $M_V^* = \exp(m^{*a} + G_{ab} / 2)$, où m^a est la moyenne arithmétique du Log des échantillons et où G_{ab} est calculé avec le variogramme des Logs.

Les erreurs d'estimation sont les différences, à chaque réalisation, entre la moyenne sur zone M_V et l'estimateur M_V^* . La variance d'estimation est :

$$\sigma_E^2 = M_V^2 \exp(G_{VV}) (T_{VV} - 2T_{aV} - T_{ab})$$

Rappels méthodologiques :
ce que fait EVA



avec $T_{VV} = \int_V dx \int_V \exp(-\gamma(x-y)) dy$; $T_{aV} = \frac{1}{V} \int_V \exp\left(-\frac{1}{n} \sum_a \gamma(x_a - y)\right) dy$

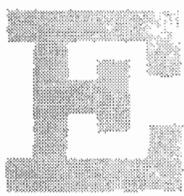
$T_{ab} = \exp(-G_{ab})$

Le calcul des termes T_{VV} et T_{aV} s'effectue par discrétisation de la zone V, comme précédemment. Et, de même, pour des raisons numériques, il est important de prendre une discrétisation suffisamment fine.

Dans l'expression de la variance d'estimation, σ_E^2 , intervient la valeur moyenne de la fonction aléatoire sur V, que l'on ne connaît pas. EVA fournit l'erreur relative d'estimation par rapport à M_V , σ_E / M_V , et les valeurs de G_{VV} et de G_{ab} .

La variance d'estimation pour un plan aléatoire ayant le même nombre n d'échantillons est : $(M_V^2 \exp(G_{VV}) - 1) / n$.

Elle peut être approchée en utilisant la formule : $(M_V^2 \exp(G_{VV}) - 1) / n$.



P. Petitgas, A. Prampart



© Orstom, 1995

I-27

BIBLIOGRAPHIE

ICES, 1991 — *Report of the Workshop on the applicability of spatial statistical techniques to acoustic survey data*. Copenhagen, ICES, CM91/D: 40, 20 p.

ARMSTRONG (M.), RENARD (D.), RIVOIRARD (J.), PETITGAS (P.), 1992 — *Geostatistics for fish survey data*. Fontainebleau, ENSMP, centre de Géostatistique, 90 p.

COCHRAN (W.), 1977 — *Sampling techniques*. New York, John Wiley and Sons, 3^e édition.

JOURNEL (A.), HUIJBREGTS (C.), 1978 — *Mining Geostatistics*. London, Academic Press, 600 p.

MACLENNAN (D.), MACKENZIE (I.), 1988 — Precision of acoustic abundance estimates. *Can. J. Fish. Aquat. Sci.*, 45 : 605-616.

MATHERON (G.), 1965 — *Les variables régionalisées et leur estimation*. Paris, Masson, 306 p.

MATHERON (G.), 1971 — *The theory of regionalised variables and their applications*. Fontainebleau, ENSMP, centre de Géostatistique, coll. Les cahiers du centre de Morphologie mathématique, fasc. 5, 212 p.

MATHERON (G.), 1974 — *Effet proportionnel et lognormalité ou le retour du serpent de mer*. Fontainebleau, centre de Géostatistique, note N-374.

MATHERON (G.), 1978 — *Estimer et choisir*. Fontainebleau, centre de Géostatistique, coll. Les cahiers du centre de Morphologie mathématique, fasc. 7, 175 p.



- MATHERON (G.), 1989 — *Estimating and choosing. An essay on probability in practice*. Berlin, Springer-Verlag, 141 p.
- PENNINGTON (M.), 1983 — Efficient estimators of abundance, for fish and plankton surveys. *Biometrics*, 39 : 281-286.
- PETITGAS (P.), 1991 — *Contributions géostatistiques à la biologie des pêches maritimes*. Thèse de doctorat, centre de Géostatistique, Fontainebleau, 211 p.
- PETITGAS (P.), RIVOIRARD (J.), 1991 — « Global estimation : s^2/n and the geostatistical estimation variance, Appendix B ». In ICES 1991 : appendix B.
- PETITGAS (P.), 1993 — Geostatistics for fish stock assessments: a review and an acoustic application. *Ices J. Mar. Sci.*, 50 : 285-298.
- SIMMONDS (E. J.), WILLIAMSON (N.), GERLOTTO (F.), AGLÉN (A.), 1992 — *Acoustic survey design and analysis procedures: a comprehensive review of current practice*. Copenhagen, ICES, Coop. Res. Report, n° 187.
- SMITH (J.), 1990 — Use of statistical models for the estimation of abundance from groundfish trawl survey data. *Can. J. Fish. Aquat. Sci.*, 47 : 894-903.
- SMITH (J.), 1988 — Evaluating the efficiency of the delta-distribution mean estimator. *Biometrics*, 44 : 485-493.

E

P. Petitgas, A. Prampart

V

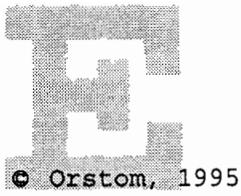
© Orstom, 1995

A

Installation,
mise en route :

principes d'utilisation

II



P. Petitgas, A. Prampart



II-2

INSTALLATION ET MISE EN ROUTE

Le logiciel EVA est écrit en *visual basic* 1.00. Il fonctionne sous Windows 3.1 et nécessite l'utilisation d'un tableur pour Windows, par exemple Excel 3.0 ou 4.0. L'exécutable fourni, EVA.exe, comme tous les exécutables de *visual basic*, utilise une librairie pour Windows nommée VBRUN100.dll. Cette librairie est indispensable au fonctionnement d'EVA.

La disquette fournie contient deux répertoires : EVA et DATA. Le répertoire DATA contient trois fichiers de données servant d'exemples de traitement qui sont détaillés dans les chapitres III, IV et V.

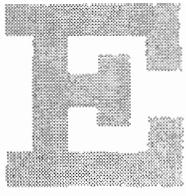
Le répertoire EVA contient EVA.exe et VBRUN100.dll. On conseille de garder cette architecture, donc de créer sur C:\ un répertoire EVA puis d'y copier le contenu du répertoire A:\EVA. Ces opérations s'effectuent facilement avec le gestionnaire de fichiers de Windows : il suffit de prendre avec la souris le répertoire EVA sur A:\ et d'aller le déposer dans C:\. Si VBRUN100.dll n'est pas installé dans le même répertoire que EVA.exe, il faut modifier en conséquence la variable PATH dans l'AUTOEXEC.BAT.

Créer une icône

Une fois installé sur C:\, on peut créer une icône pour le programme EVA, en utilisant les commandes du gestionnaire de programmes de Windows :

- faire Fichier, puis Nouveau dans la barre de menu du gestionnaire de programmes ;
- une fenêtre apparaît : choisir Groupe de programmes, puis donner

Installation, mise en route :
principes d'utilisation



P. Petitgas, A. Prampart



© Orstom, 1995

un nom au groupe de programmes, par exemple Geostat, un groupe de programmes est créé :

— refaire Fichier, puis Nouveau dans la barre de menu du gestionnaire de programmes, répondre cette fois Programme ;

— une autre fenêtre apparaît : remplir comme on le désire les cases, par exemple : nom = EVA, ligne de commande = C:\EVA\EVA.EXE, répertoire de travail = C:\DATA, cliquer sur le bouton Changer d'icônes et choisir le poisson ;

— faire OK.

On peut désormais lancer EVA à partir du gestionnaire de fichiers ou à partir de son icône dans le groupe Géostat.

PRINCIPES D'UTILISATION

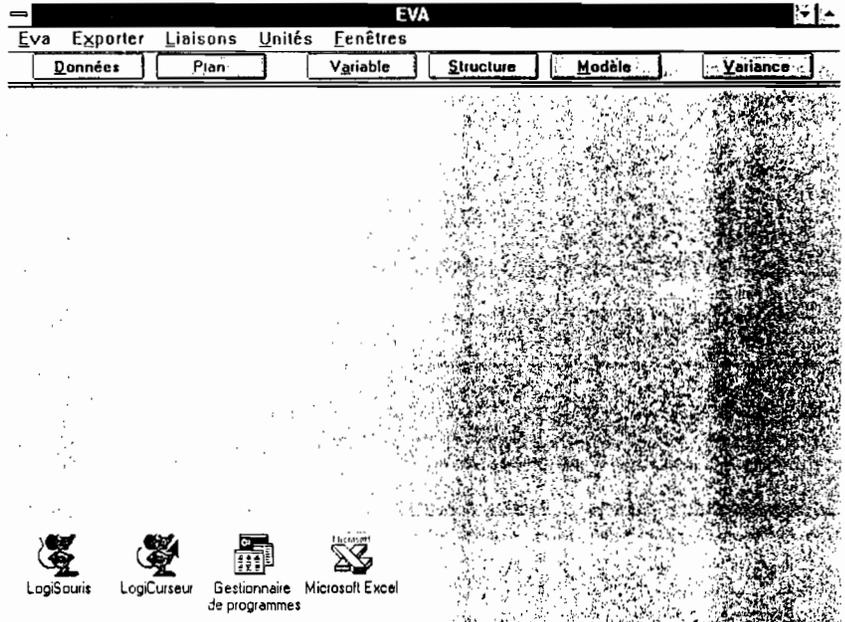
Une fois lancé, le logiciel EVA se présente par deux barres en haut de l'écran : une barre de menus (Eva, Exporter, Liaisons, Unités, Fenêtres) et, au-dessous, une barre d'exécution constituée d'une succession de boutons (Données, Plan, Variable, Structure, Modèle, Variance) (fig. 1).

Pour aller des données à gauche à la variance à droite, il faut passer par une succession d'étapes matérialisées par les boutons. À chaque bouton sont associées différentes fenêtres qui sont utiles à l'accomplissement de l'étape. Mais les fenêtres sont indépendantes les unes des autres. Chaque fenêtre qui a été ouverte reste en mémoire et pourra être rappelée à tout moment. Ce sont ces propriétés de l'interface qui nous permettent de naviguer librement entre les étapes du traitement.

Installation, mise en route :
principes d'utilisation

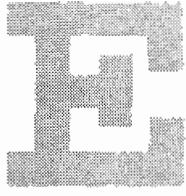
II-4

FIGURE 1 -
Le logiciel EVA
ouvert après
le double clic
sur l'icône.
Barre des menus
et barre
des boutons
d'exécution.



La morphologie de chaque fenêtre peut être modifiée grâce aux flèches en haut à droite de son cadre : elle peut être agrandie ou réduite. La réduction n'est pas une fermeture : la fenêtre réduite garde ses paramètres ; elle reste en mémoire et peut être rappelée.

Cliquer sur un nouveau bouton d'exécution ouvre les fenêtres qui lui sont associées. Cliquer à nouveau sur un bouton d'exécution d'une étape déjà effectuée réactive les fenêtres déjà ouvertes dans cette étape. Les fenêtres ouvertes dans une étude ne sont jamais refermées sauf dans le cas suivant. Si au cours de l'étude, on revient dans la



P. Petitgas, A. Prampart



© Orstom, 1995

fenêtre Plan et que le choix du plan est modifié, toutes les fenêtres se fermeront. Ceci est logique dans la mesure où changer de plan consiste à entreprendre une autre étude.

Passer par une succession de boutons revient à parcourir un chemin. En fonction du type de données, du type de plan d'échantillonnage et du type de méthode géostatistique (transitive ou intrinsèque), on aboutit à l'étape du calcul de variance, en ayant parcouru différents chemins.

À chaque instant, tous les chemins ne sont pas possibles : certains boutons sont grisés ou activables en fonction des étapes précédemment achevées.

Les boutons Variable, Structure et Modèle sont spécifiques à l'analyse structurale. Les boutons Données et Plan permettent de choisir un chemin pour aboutir à la variance. EVA permet de croiser la méthode de calcul de variance, transitive ou intrinsèque, avec le plan effectué. Toutefois, c'est la même formule qui est utilisée pour tout plan dans la méthode intrinsèque.

EVA permet deux utilisations :

- l'une consiste à faire une étude complète des données expérimentales en passant successivement par toutes les étapes, c'est-à-dire par tous les boutons de la barre d'exécution ;
- l'autre consiste à calculer la variance avec la méthode voulue, connaissant à priori le modèle structural, par exemple celui des campagnes précédentes. On utilisera alors uniquement les boutons Données, Plan puis Variance. Ceci permet de tester des efforts et des plans d'échantillonnage, connaissant une structure.

Outre l'utilisation d'EVA uniquement pour les boutons de l'analyse structurale, quatre chemins sont possibles (tabl. 3).



II-6

TABLEAU 3 -
Utilisation
de EVA.
Quatre chemins
possibles.

Connaissance du modèle de structure spatiale	Méthode géostatistique du calcul de variance d'estimation	Boutons de la barre d'exécution à utiliser	Numéro de chemin
Oui	Transitif 1D	Plan, Variance	1
Oui	Intrinsèque	Données, Plan, Variance	2
Non	Intrinsèque	Données, Plan, Variable, Structure, Modèle, Variance	3
Non	Transitif 1D	Données, Plan, Variable, Structure, Modèle, Variance	4

IMPORTATION, MISE À JOUR, EXPORTATION

EVA fonctionne comme une « application » de Windows au même titre que Word, Excel ou Paintbrush. La fonction Copier-Coller sous Windows est mise à profit, d'une part, pour importer les données et, d'autre part, pour exporter les graphes. La fonction Copier-Coller fait transiter l'information par le presse-papiers et permet ainsi de passer cette information d'une application à l'autre de Windows. Pour effectuer les Copier-Coller, il faut que soient ouvertes en même temps les applications source et arrivée : EVA (= arrivée) et un tableur (= source), Excel par exemple, pour importer des données ; EVA (= source) et un éditeur de graphiques (= arrivée), Paintbrush par exemple, pour exporter les graphes.

Le raccourci clavier Alt + Tab (on garde la touche Alt appuyée en faisant Tab) permet sous Windows de basculer d'une application à l'autre, très facilement. En gardant la touche Alt appuyée et en faisant Tab plusieurs fois, on fait défiler successivement toutes les applications qui sont ouvertes. Pour activer ce raccourci clavier, il faut modifier la configuration de Windows : aller dans Groupe principal, puis dans Panneau de configuration, puis dans Bureau et cocher la case correspondant au déplacement rapide avec Alt + Tab.

Importations

LECTURE DES DONNÉES, COLONNE À COLONNE

La suite des étapes à effectuer est la suivante :

- lancer EVA. Cliquer sur le bouton d'exécution Données. La fenêtre associée s'ouvre ;
- Alt + Tab : on revient au gestionnaire de programmes. Lancer Excel ou un autre tableur sous Windows ;
- dans Excel, lire ses données. Elles sont désormais dans une feuille de calcul où chaque colonne correspond à une variable : x, y, valeur échantillonnée, etc. Sélectionner une colonne. Faire Édition puis Copier dans la barre de menu. Cette colonne est désormais dans le presse-papiers ;
- Alt + Tab : on revient dans la fenêtre Données de EVA ;
- dans la fenêtre Données de EVA, cliquer dans la colonne désirée. Un double cadre marque alors cette sélection. Cliquer sur le bouton Coller ou sur le bouton Coller avec liaison. Un signal sonore (bip) indique que l'opération est finie. On voit les chiffres s'afficher.

Installation, mise en route :
principes d'utilisation



Répéter cette suite pour chacune des colonnes d'intérêt. Cliquer sur le bouton Fermer valide la lecture.

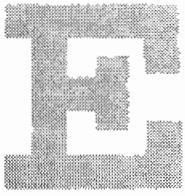
Les colonnes de la fenêtre Données de EVA ne sont pas unifiées et ne sont pas des éditeurs. Elles servent uniquement à voir ses données pour vérification. Toute modification des valeurs s'effectue dans Excel. Les colonnes modifiées doivent être réimportées dans EVA.

Dans la fenêtre Données de EVA, on peut lire trois types de fichiers : Échantillons, Polygone, Côte. Chacun est Importé depuis Excel, colonne à colonne. On conseille d'organiser une feuille de calcul Excel qui contienne les trois fichiers de façon à se repérer plus facilement. Dans une feuille Excel, le fichier Données échantillons est un champ qui comporte autant de lignes que d'échantillons et trois colonnes, x, y et des valeurs. Les fichiers Côte et Polygone ne comportent que deux colonnes, x et y.

EVA permet d'utiliser différents chemins comme spécifié dans le tableau 3. Le tableau 4 donne les colonnes qu'il est nécessaire de remplir dans la fenêtre Données de EVA.

TABLEAU 4 -
Les Données
entrées pour
chacun des
quatre chemins
de EVA.

Numéro du chemin	Fichier données échantillons			Fichier polygone
	x	y	Valeurs	
1	-	-	-	-
2	Oui	Oui	-	Oui
3	Oui	Oui	Oui	Oui
4	-	-	Oui	-



P. Petitgas, A. Prampart



© Orstom, 1995

Le fichier Polygone est une suite de points. Le polygone doit être fermé dans le fichier : la dernière ligne égale la première. Le polygone ne peut pas contenir de « trous » à l'intérieur.

Le fichier Côte est aussi une suite de points. Il contient la ligne de côte qui n'est pas forcément fermée, suivie éventuellement d'îles. Chaque île est un polygone fermé dans le fichier. Chaque entité de côte, ligne de côte non fermée ou île, est une suite de points. Ces entités sont à la suite les unes des autres dans le fichier. La répétition de la dernière ligne de chaque entité marque la séparation entre entités dont EVA a besoin. Pour construire ce fichier sous Excel, il suffit de répéter la dernière ligne de chacune des suites de points, puis de coller ces petites entités à la suite les unes des autres. Le fichier Côte est utilisé uniquement lors de la représentation des données sur la carte : il ne sert pas dans les calculs.

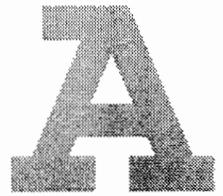
LIAISON ET MISE À JOUR

Si on choisit Coller avec liaison dans la fenêtre Données de EVA, une liaison est établie avec la colonne de la feuille Excel correspondante. Tout changement des valeurs dans cette colonne Excel sera passé directement à EVA grâce à la fonction Liaison-mise à jour de la barre de menu de EVA.

Toute modification des valeurs est possible à tout moment, en revenant aux données dans Excel. La mise à jour équivaut à la lecture automatique de nouvelles valeurs. Cette étape effectuée, il faudra refaire les calculs.

Une bonne organisation de la feuille de calcul d'Excel permet de mettre à profit le lien qui existe entre une colonne Excel et une colonne EVA : on peut facilement changer de polygone, changer de variable, ou modifier la variable sur laquelle on travaille. Il suffit d'échanger le

Installation, mise en route :
principes d'utilisation



contenu de la colonne liée avec celui d'une autre colonne de la feuille de calcul, ou de modifier le contenu de la colonne liée.

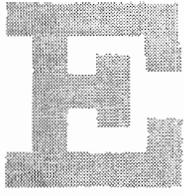
RÈGLES DE COPIE DES CARACTÈRES DEPUIS EXCEL

La fonction Copier d'Excel a certaines propriétés qui en limitent l'utilisation. Il faut faire attention à la taille de la colonne Excel, au séparateur des décimales, au caractère alphanumérique. La fonction Copier-Coller travaille en format texte. Dans EVA, après collage, il y a reconversion du texte en numérique quand on clique sur le bouton Fermer dans la fenêtre Données.

La fonction Copier d'Excel ne copie que ce qui est visible dans la colonne. Si la colonne n'est pas assez large, les chiffres transmis à EVA seront tronqués à leur partie visible. Il faut agrandir les colonnes d'Excel si besoin est.

Dans EVA, le séparateur des décimales est le point (.). Si on colle des chiffres dont le séparateur des décimales est une virgule, seule la partie entière sera prise en compte numériquement dans EVA. Il faut changer si besoin est le format des nombres dans la configuration de Windows. Le chemin d'accès à cette configuration est le suivant : aller dans le gestionnaire de programmes, puis dans le Groupe principal, puis dans le Panneau de configuration, puis dans International, puis faire Modifier dans le cadre Format nombre et mettre un point comme séparateur des décimales. Pour prendre en compte le changement de configuration, il n'y a pas besoin de relancer Windows, par contre, il faudra relancer Excel.

Les lettres sont interprétées par EVA comme des valeurs nulles. Avant de copier, il faut soit enlever soit ne pas sélectionner les mots et lettres qui figurent dans la colonne Excel.



P. Petitgas, A. Prampart



© Orstom, 1995

Exportations : impressions et sauvegardes

Les impressions et sauvegardes s'effectuent par exportation dans d'autres applications de Windows. On peut exporter les figures, les valeurs du variogramme et du covariogramme grâce à la fonction Exporter de la barre de menu de EVA. On peut exporter tout ou une partie de l'écran par la touche clavier Copie d'écran.

FONCTION EXPORTER DE LA BARRE DE MENU

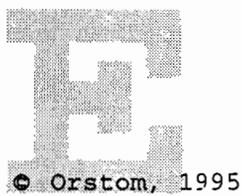
On peut exporter des figures et les valeurs des variogrammes et des covariogrammes expérimentaux. Les figures sont exportées au format bitmap, les valeurs au format texte. Les figures peuvent être collées dans un éditeur de dessins comme Paintbrush, les valeurs du variogramme dans une feuille Excel.

La succession des étapes est la suivante :

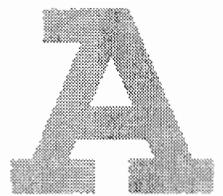
- dans EVA, aller dans Exporter de la barre de menu et y choisir le graphe à exporter dans la liste. Un signal sonore (bip) indique que le graphe se trouve dans le presse-papiers ;
- Alt + Tab : on bascule dans l'application arrivée. Ce peut être Word, Excel, Write, Paintbrush ;
- dans l'application arrivée, faire Édition puis Coller. Le graphe s'affiche. On peut alors l'éditer, le sauver, l'imprimer ;
- Alt + Tab : on revient dans EVA.

On peut copier une figure agrandie. On agrandit une fenêtre en cliquant sur la flèche qui pointe vers le haut, en haut à droite de la

Installation, mise en route :
principes d'utilisation



P. Petitgas, A. Prampart



fenêtre. Pour exporter la figure agrandie, il faut réduire la fenêtre (cliquer sur la flèche qui pointe vers le bas en haut à droite de la fenêtre). On a à nouveau accès à la barre de menu. En effectuant la suite des étapes précédentes, c'est la figure agrandie qui est exportée.

Le collage dans une feuille Excel des valeurs du varlogramme ou du covariogramme donne automatiquement un en-tête spécifiant la direction et le nom de chaque colonne. Les distances, les valeurs et le nombre de couples sont automatiquement collés dans trois colonnes différentes.

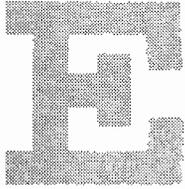
TOUCHE CLAVIER DE COPIE D'ÉCRAN

La touche Copie d'écran permet d'exporter dans le presse-papiers tout l'écran. On peut ensuite le coller dans toute application de Windows.

La touche Alt + Copie d'écran permet d'exporter dans le presse-papiers toute la fenêtre active, son contenu et ses bordures. On peut ensuite la coller dans une application Windows. On peut ainsi sauver et imprimer facilement les fenêtres des paramètres des modèles choisis. Une fenêtre active a sa bordure supérieure qui est plus foncée.

BARRE DE MENU

La barre de menu contient les fonctions Eva, Exporter, Liaisons, Unités, Fenêtres. Chacune ouvre une liste déroulante. La fonction Eva permet d'accéder à l'ordre Quitter qui ferme le logiciel. Les fonctions Exporter et Liaisons ont été décrites dans le chapitre précédent : « Importation, mise à jour, exportation ».



P. Petitgas, A. Prampart



© Orstom, 1995

Unités

La fonction Unités permet de définir les unités des coordonnées et des distances qui seront utilisées dans les calculs. À l'issue de la saisie lors de l'échantillonnage, les coordonnées des échantillons sont en général en degrés, minutes et dixièmes de minutes. Or la distance que représente un mille de longitude varie en fonction de la latitude. Une transformation des coordonnées peut s'avérer nécessaire, afin de disposer d'un repère dans lequel on puisse calculer des distances cartésiennes. Toute transformation des coordonnées doit se faire sous Excel.

Deux situations sont permises dans EVA : les coordonnées sont en degrés et centièmes de degrés et les distances sont en miles nautiques (c'est l'option par défaut) ; ou bien coordonnées et distances sont exprimées dans la même unité. L'option active est celle devant laquelle figure une marque dans la liste déroulante.

Fenêtres

La fonction Fenêtre permet de ranger les fenêtres. Chaque fenêtre est associée à un bouton de la barre d'exécution : quand on clique dans le bouton, ces fenêtres s'ouvrent. Le même bouton permet d'accéder à différentes fenêtres en fonction du chemin dans lequel on se trouve. Chaque fenêtre ouverte reste en mémoire et peut être rappelée à tout moment. La fonction Fenêtre-Afficher/Mettre à jour donne accès à la liste des fenêtres. Les noms des fenêtres déjà ouvertes sont en gras, les autres sont grisés. La sélection d'une fenêtre la rappelle.

Chaque fenêtre est indépendante : elle est accessible à tout moment après qu'elle ait été ouverte ; elle peut être déplacée en cliquant

dans sa bordure supérieure et en déplaçant la souris ; elle peut être réduite à l'aide de la flèche qui pointe vers le bas en haut à droite de son cadre. Très rapidement, l'écran peut devenir un fouillis que l'on peut vouloir ranger. L'option Rangement automatique permet d'effectuer cette opération. La marque située devant l'option montre qu'elle est active. Si les fenêtres sont éparpillées et/ou réduites et si l'option Rangement automatique est active, quand on clique sur le bouton d'exécution, les fenêtres associées se replacent à leur endroit d'origine.

FICHIERS DE DONNÉES FOURNIS POUR EXEMPLES D'UTILISATION

Le répertoire DATA sur la disquette fournie contient trois jeux de données : SPAWHER.TXT, HEREGG.TXT et REGEGG.TXT. Ils sont fournis dans le format texte-ASCII, avec une tabulation comme séparateur des colonnes. Ces données sont donc directement transférables dans un tableur sous Windows.

Les chapitres III, IV et V suivants décrivent pas à pas la suite des étapes à effectuer pour réaliser des études types avec EVA. Ces chapitres permettent de se familiariser avec les méthodes tout en manipulant le logiciel : on peut les comprendre comme des séances de travaux dirigés. Le tableur utilisé est Excel 4.

E

P. Petitgas, A. Prampart

V

© Orstom, 1995

A

Estimation par cumul
le long de transects
acoustiques :

exemple d'utilisation

III



III-2

Le jeu de données qui sert d'exemple ici a été fourni par le Marine Research Institute de Bergen (Norvège) comme jeu test (test data 4) pour un workshop du CIEM sur les statistiques spatiales (ICES, 1991). Il s'agit d'une prospection acoustique. On trouvera une description et un traitement plus approfondis de ces données dans ICES (1991) et dans PETITGAS (1991, 1993). La partie régulière du plan d'échantillonnage est fournie sur la disquette dans le répertoire DATA. Le nom du fichier est SPAWHER.TXT.

La prospection acoustique a eu lieu le long de transects parallèles et équidistants sur l'une des plus importantes frayères du hareng norvégien (*norwegian spring spawning herring*) au large de More, en février 1989. La prospection a eu lieu pendant la période de reproduction et n'a duré que quelques jours alors que le poisson reste sur zone plusieurs semaines. La distance intertransects est de 4.54 milles nautiques (M.n.). Chaque valeur d'échointégration est une valeur moyenne des Index ponctuels de réverbération acoustique des poissons : moyenne sur la colonne d'eau, puis moyenne de ces valeurs sur des segments de 1 M.n. le long de la route du navire. L'unité acoustique des index est arbitraire. Selon les propriétés de l'échointégration, les valeurs sont proportionnelles à la quantité de poisson qui était présente sous le navire. Les valeurs sont rapportées à l'unité de surface de $1(\text{M.n.})^2$ et représentent donc une densité en poisson. Le jeu de données est visualisé sur la figure 2. Le rayon de chaque cercle est proportionnel à la densité pour ce mille.

Comme ces valeurs sont des moyennes le long de la route du navire, PETITGAS (1993) a proposé de travailler sur la quantité de poisson par transect (cumul des valeurs par transect) et d'utiliser la méthode transitive à une dimension sur les cumuls pour calculer la précision de l'estimation de la quantité totale présente sur zone.

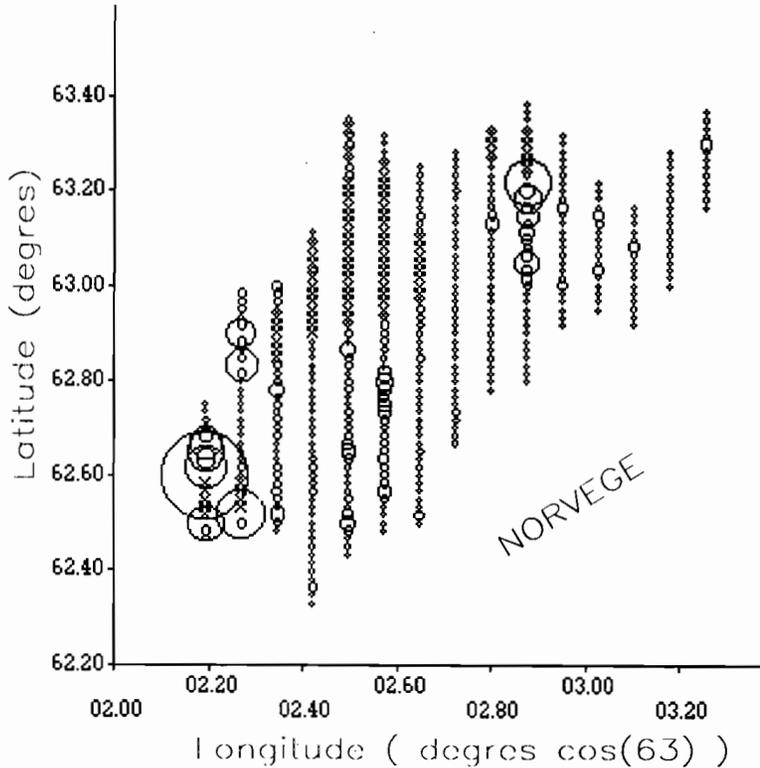


FIGURE 2 -
Représentation
proportionnelle
des données
du fichier
SPAWHER.TXT.

ORGANISATION DE LA FEUILLE EXCEL

La figure 3 est une copie d'écran de la feuille de calcul Excel contenant le fichier qui va être importé dans EVA. On a effectué au préalable avec Excel les calculs suivants. On a d'abord transformé les coordonnées : les latitudes, y, ont été exprimées en degrés et centièmes,

Estimation par cumul le long
de transects acoustiques

III-4

FIGURE 3 -

Données
du fichier
SPAWHER.TXT
dans une feuille
de calcul Excel.
B = longitude,
C = latitude
(en milles
nautiques
et centièmes),
D = densité
acoustique
de poisson,
F = numéro
des transects,
G = biomasse
par transect
(densité moyenne
du transect
multipliée
par sa longueur),
H = longueur
des transects
(milles nautiques).

	B	C	D	E	F	G	H
1	2 271	62 5	3543		1	196149	37
2	2 271	62 5166667	36400		2	103920	42
3	2 271	62 5333333	0		3	49427	59
4	2 271	62 55	0		4	13883	55
5	2 271	62 5666667	0		5	63489	63
6	2 271	62 5633333	0		6	74406	61
7	2 271	62 6	80		7	21587	57
8	2 271	62 5166667	652		8	12627	47
9	2 271	62 6333333	100		9	16061	44
10	2 271	62 65	52		10	128681	43
11	2 271	62 5666667	66		11	16712	36
12	2 271	62 6833333	171		12	12054	26
13	2 271	62 7	65		13	8768	31
14	2 271	62 7166667	65		14	4058	31
15	2 271	62 7333333	84		15	13260	21
16	2 271	62 75	52				
17	2 271	62 7666667	50				
18	2 271	62 7833333	10				

Prêt

les longitudes, x , ont été exprimées en degrés et centièmes et ont été multipliées par le cosinus de la latitude moyenne. Une telle transformation des longitudes est nécessaire à 63° de latitude nord afin qu'un mille représente une même distance sur les x et sur les y . Les colonnes B et C sont respectivement celles des x et des y . La colonne D est celle des valeurs acoustiques de la densité en poisson. La colonne A est celle du loch. On a ensuite repéré chaque radiale et on a fait le cumul des valeurs pour chaque radiale. Les radiales sont nord-sud ce qui facilite les notations. Le cumul s'opère le long des y Indicés par j :

$$q_i = a \sum_{j=1}^n f_{i,j}$$

Estimation par cumul le long
de transects acoustiques

où $f_{i,j}$ est la densité en poisson sur le j° segment de longueur a du i° transect, et où n_j est le nombre d'échantillons sur le transect i ($a = 1$ M.n. lcl).

La colonne F contient le numéro des radiales, du sud vers le nord de la zone.

La colonne G contient les valeurs cumulées le long de chaque transect.

La colonne H contient la longueur des segments positifs le long de chaque transect (i.e. nombre de valeurs strictement positives le long du transect).

On importe dans EVA, la colonne G.

On sélectionne la colonne G ou le champ des 15 premières lignes de la colonne G. On clique dans la fonction Édition de la barre de menu et on choisit Copier. Les 15 valeurs sont maintenant dans le presse-papiers (fig. 4).

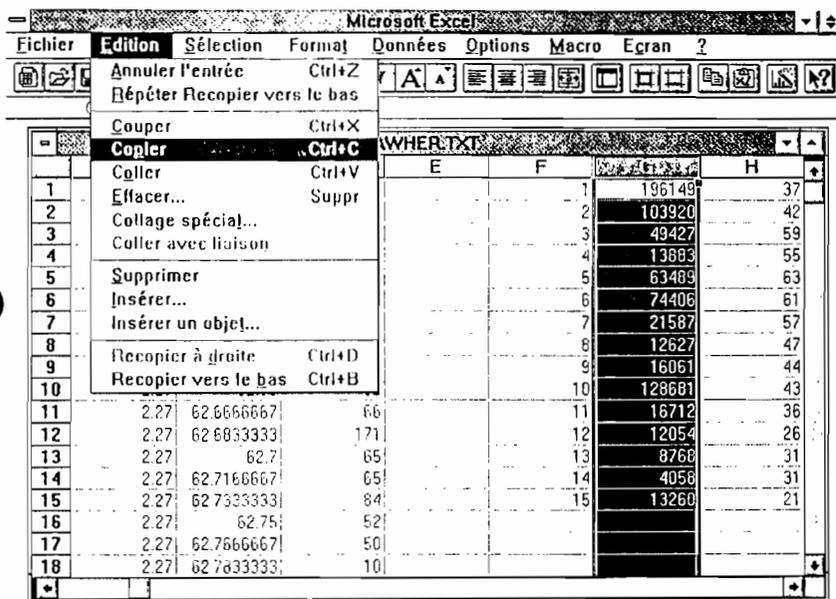


FIGURE 4 -
Copier
la colonne G
dans le
presse-papier.

Copier les cellules sélectionnées dans le Presse-papiers

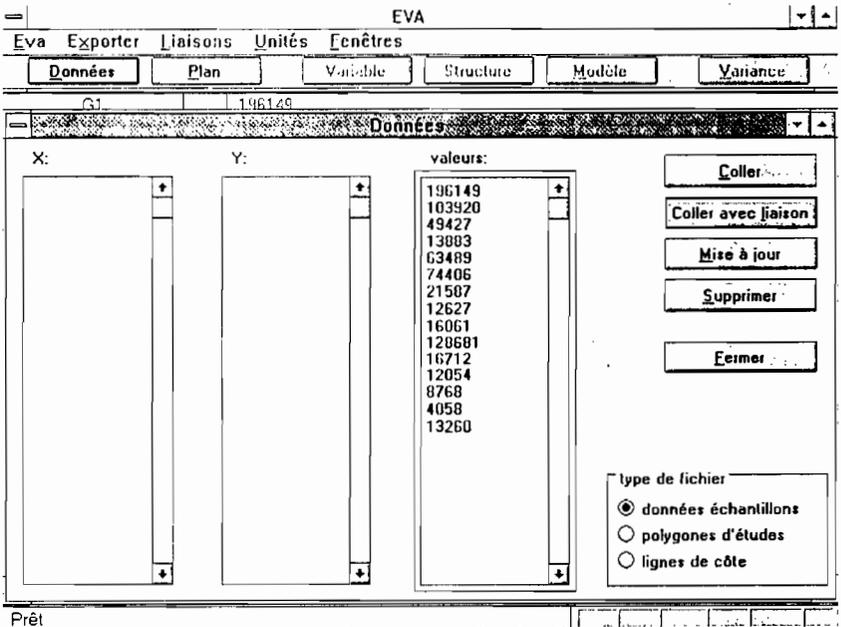
Estimation par cumul le long
de transects acoustiques

III-6

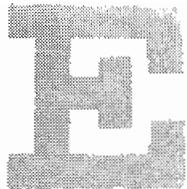
Alt + Tab : on revient au gestionnaire de programmes de Windows et on lance EVA en cliquant deux fois sur son icône. Les deux applications Excel et EVA sont ouvertes. On est désormais dans EVA.

IMPORTATION DES DONNÉES ET CHOIX DE LA MÉTHODE GÉOSTATISTIQUE

FIGURE 5 -
Bouton d'exécution Données.
Importer la colonne G dans EVA.
Cliquer dans la colonne valeurs, puis sur Coller, puis sur Fermer.



Estimation par cumul le long de transects acoustiques



P. Petitgas, A. Prampart



Bouton Données

Le bouton Données de la barre d'exécution permet d'importer les données. On clique sur le bouton Données : la fenêtre associée s'ouvre (fig. 5). Le type de fichier Données échantillons est la sélection par défaut. Cela nous convient. On sélectionne la colonne Valeurs en cliquant dedans. On importe les 15 cumuls en cliquant sur l'un des deux boutons Coller ou Coller avec liaison. On choisit par exemple Coller avec liaison. Après le signal sonore, on clique sur le bouton Fermer. Les 15 valeurs sont importées.

Bouton Plan

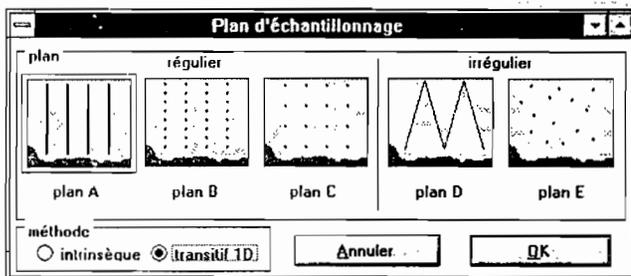
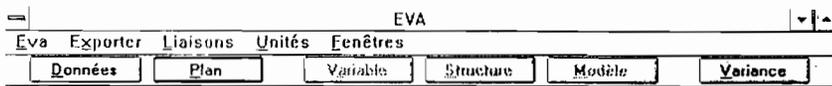
Le bouton Plan permet de croiser le plan d'échantillonnage effectué avec la méthode géostatistique utilisée pour le calcul de la variance. On clique sur le bouton Plan : la fenêtre associée s'ouvre (fig. 6).

On choisit le plan A en cliquant sur l'icône correspondante. Un cadre l'entoure qui montre que la sélection est maintenant active. On choisit la méthode transitive en cliquant sur le rond ou sur le nom. Le rond devient noir, indiquant par là que cette sélection est active. Pour fermer la fenêtre, cliquer sur OK.

Pendant tout le temps de l'importation des données et du choix de la méthode, les autres boutons de la barre d'exécution restent grisés : les opérations qu'ils permettent de faire ne sont pas accessibles. Une fois les données importées et la méthode choisie, les boutons se dégrisent. L'étude commence.

III-8

FIGURE 6 -
Bouton
d'exécution Plan.
Croiser le plan
d'échantillonnage
avec la méthode
d'analyse.
Cliquer sur
Plan A,
puis sur
transitif 1D,
puis sur OK.



LogiSouris



LogiCursur



Gestionnaire
de programmes



Microsoft Excel

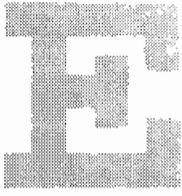
ANALYSE STRUCTURALE

Cette étape s'effectue grâce aux trois boutons d'exécution : Variable, Structure, Modèle.

■ Bouton Variable

Deux fenêtres y sont associées. La fenêtre Cumuls fournit une représentation des valeurs sous forme de barres. La fenêtre Variable fournit les

Estimation par cumul le long
de transects acoustiques



statistiques élémentaires des valeurs. La variance fournie est la variance des données, calculée en $1/(n - 1)$ (fig. 7).

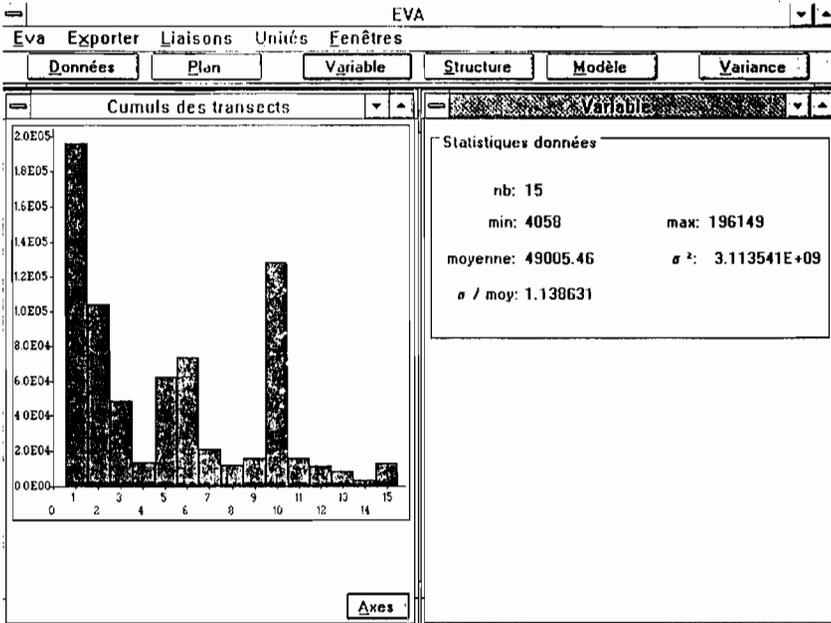


FIGURE 7 - Bouton d'exécution Variable. Représentation spatiale et statistiques élémentaires des données.

Le bouton Axes de la fenêtre Cumuls ouvre une fenêtre dans laquelle il est possible de modifier les échelles sur les axes : T signifie Transects et V signifie Valeurs.

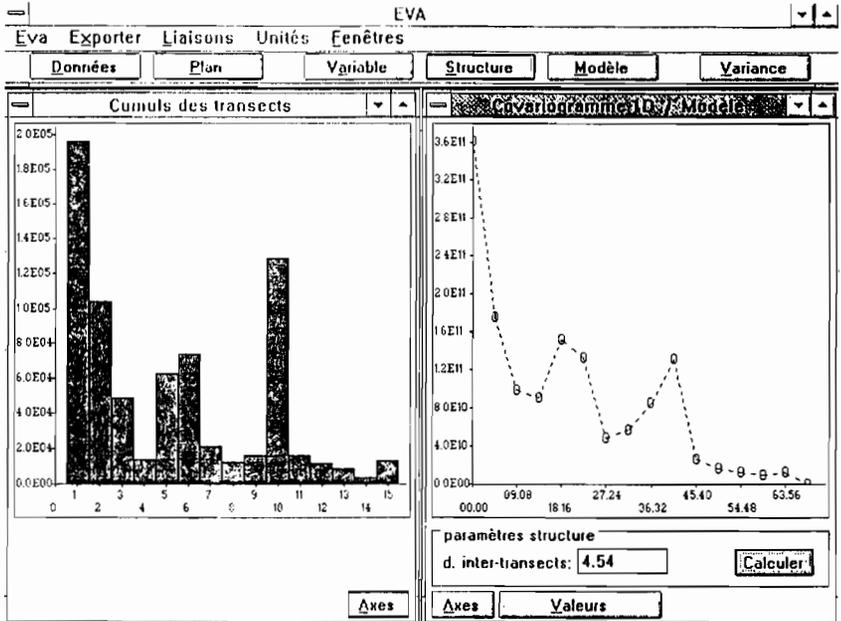
Bouton Structure

Deux fenêtres y sont associées : la fenêtre Cumuls qui est la même que précédemment et la fenêtre Covarlogramme 1D. Ceci permet de voir en

Estimation par cumul le long de transects acoustiques

même temps et les données et leur structure. La fenêtre Covariogramme permet d'effectuer le calcul du covariogramme expérimental (fig. 8).

FIGURE 8 -
Bouton
d'exécution
Structure.
Entrer
la valeur
de la distance
intertransect
et cliquer
sur Calculer.
Le covariogramme
transitif
expérimental
s'affiche
à l'écran.



On écrit 4.54 dans la case de la distance intertransects, et on clique sur le bouton Calculer de la fenêtre Covariogramme.

Le bouton Axes de la fenêtre Covariogramme ouvre une fenêtre dans laquelle il est possible de modifier les échelles sur les axes du covariogramme : D signifie distance et V signifie valeurs du covariogramme.

Le bouton Valeurs de la fenêtre Covariogramme ouvre une fenêtre dans laquelle on lit les valeurs du covariogramme.

Bouton Modèle

Deux fenêtres y sont associées : la fenêtre Covariogramme qui est la même que précédemment et la fenêtre Paramètres modèle. Ceci permet d'effectuer l'ajustement du modèle interactivement, de visu (fig. 9).

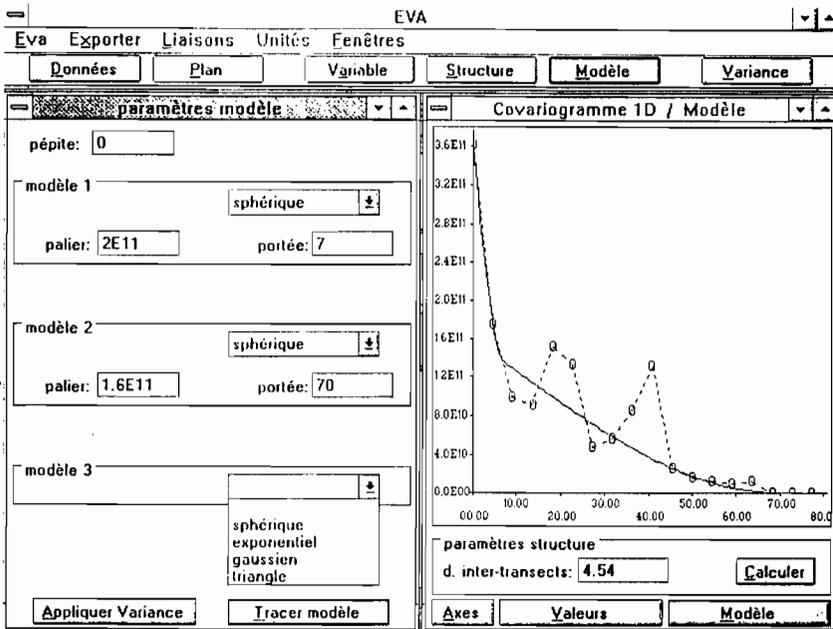


FIGURE 9 -
Bouton d'exécution Modèle. Choisir les types de modèle, entrer les valeurs des paramètres. Cliquer sur Tracer modèle. Changer les paramètres interactivement jusqu'à ce que l'ajustement soit visuellement satisfaisant. Ensuite, cliquer sur Appliquer variance.

Dans la fenêtre Covariogramme, un nouveau bouton est apparu, le bouton Modèle. Ce bouton et le bouton Valeurs permettent de basculer à loisir entre la fenêtre des valeurs du covariogramme expérimental et la fenêtre des paramètres du modèle. On peut ainsi choisir les paramètres du modèle de manière à le faire passer exactement par certains points expérimentaux.

Estimation par cumul le long
de transects acoustiques

taux. Dans la fenêtre Paramètres modèle, le type de modèle est choisi dans chacune des trois listes déroulantes accessibles par les flèches. Le bouton Appliquer variance de la fenêtre Paramètres modèle permet de faire passer dans la fenêtre Variance le modèle ajusté sans avoir à le redonner plus loin.

VARIANCE D'ESTIMATION

Le bouton d'exécution Variance conduit au calcul de la variance d'estimation (fig. 10).

FIGURE 10 -
Bouton
d'exécution
Variance.
Cliquer sur
Calculer
variance.

EVA

Eva Exporter Liaisons Unités Fenêtres

Données Plan Variable Structure Modèle Variance

Variance pour plan A et transectif D

paramètres modèle

pépite: 0

modèle 1 sphérique

palier: 2E+11 portée: 7

modèle 2 sphérique

palier: 1.6E+11 portée: 70

modèle 3

paramètres variance

distance inter-transects: 4.54

Calculer variance

variance

$\sigma^2_{est.}$: 1.667316E+11

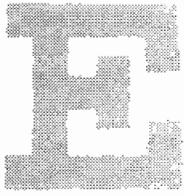
pépite: 0.0%

modèle 1: 93.0%

modèle 2: 7.0%

modèle 3: -

Estimation par cumul le long
de transects acoustiques



Les paramètres précédents ont été transmis. Il suffit de cliquer dans le bouton Calculer variance pour la calculer.

On calcule facilement la variance d'estimation pour d'autres distances intertransects. On obtient alors facilement la courbe variance d'estimation-effort d'échantillonnage.

Dans la méthode transitive, la variance d'estimation est relative à la quantité totale et non à la moyenne. Il y a 15 transects, la distance intertransects est de 4.54 M.n. et la valeur moyenne des cumuls est 49 005.46. La quantité totale est estimée à :

$$\dot{Q} = 15 * 4.54 * 49005.46 = 3.3 * 10^6$$

L'erreur relative d'estimation est donc :

$$\sigma_{est} / \dot{Q} = 12.3\%$$

Si on considérait un modèle sans corrélation spatiale entre les cumuls par transect, on calculerait une précision relative par la formule classique :

$$\frac{\sqrt{\sigma^2 / n}}{m}$$

où σ^2 et m sont la variance et la moyenne des cumuls, et où n est le nombre de transects. Ces paramètres sont fournis dans la fenêtre Variable (on peut la rappeler avec la fonction Fenêtre-Afficher/mise à jour de la barre de menu). On trouverait une précision relative de 29.4 %. L'estimation est bien meilleure que ne le laisse supposer ce calcul. Ce calcul est erroné car les cumuls par transect montrent nécessairement une structure spatiale. On trouvera plus de détails et de discussion dans PETITGAS (1993).



EXPORTATIONS

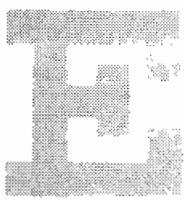
La fonction Exporter de la barre de menu déroule une liste dans laquelle seuls les noms, Cumuls, Covariogramme et Valeurs covariogramme ne sont pas grisés. Cumuls correspond au diagramme en barres des cumuls. Covariogramme correspond à la figure sur laquelle le modèle est superposé au covariogramme expérimental. On exporte facilement ses graphes un à un comme indiqué au chapitre II. On peut sauvegarder les valeurs du covariogramme en les collant dans une feuille Excel.

La fenêtre Variance contient les paramètres du modèle et la variance d'estimation. La fenêtre Variable contient les statistiques élémentaires des cumuls. On peut imprimer ces fenêtres comme indiqué dans le chapitre II. Il suffit de rappeler la fenêtre désirée grâce à la fonction Fenêtre de la barre de menu. Une fois active, l'ordre Alt + Cople d'écran l'envoie dans le presse-papiers (se reporter au chapitre II).

UTILISATION DE LA LIAISON

Supposons que l'on veuille maintenant refaire une partie de l'étude, non plus sur les cumuls des densités par transect mais sur la longueur positive des transects de façon à mieux caractériser géométriquement la frayère. La colonne G de la feuille d'Excel est liée avec la colonne Valeurs de EVA. Il suffit de mettre en G le contenu de la colonne H.

Alt + Tab : on retourne dans Excel. On sélectionne la colonne G. Grâce aux fonctions Couper et Coller, on déplace le contenu de la colonne G dans la colonne E. La colonne G est libre. On sélectionne maintenant la colonne H et, par le même procédé, on l'amène en G.



P. Petitgas, A. Prampart



© Orstom, 1995



Alt + Tab : on revient dans EVA. On utilise la fonction Liaison de la barre de menu. Le signal sonore (bip) indique que les nouvelles valeurs sont importées. Il n'est pas nécessaire de retourner dans Données. La sélection dans Plan n'a pas lieu d'être modifiée. On recommence l'étude à partir de Variable. On choisira un modèle gaussien, car il montre un comportement à tangente horizontale à l'origine qui est théoriquement celui d'une variable indicatrice cumulée le long de lignes (MATHERON, 1965).



E

P. Petitgas, A. Prampart

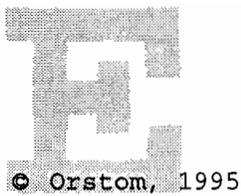
V

A
© Orstom, 1995

Estimation
de la moyenne sur zone
à partir d'échantillons
non aléatoires :

exemple d'utilisation

IV



P. Petitgas, A. Prampart



© Orstom, 1995

IV-2

Le jeu de données qui sert d'exemple ici a été fourni par le Marine Laboratory de Aberdeen (Écosse) comme exemple d'essai lors du cours de géostatistique qui a eu lieu à Fontainebleau pour les biologistes et statisticiens de la communauté du CIEM (ARMSTRONG *et al.*, 1992). Il s'agit d'une récolte d'œufs de hareng, effectuée à la benne sur un sédiment constituant un banc de ponte (*spawning bed*) dans une ria ouest-écossaise. Le fichier est fourni sur la disquette dans le répertoire DATA sous le nom HEREGG.TXT.

On sait avant l'échantillonnage où se trouve en gros le (ou les) banc(s) de ponte, mais comme les sédiments bougent d'une année à l'autre et que les géométries des bancs ne sont pas fixes à cause principalement des courants et du comportement des géniteurs, on ne peut pas définir à priori la géométrie du (ou des) banc(s) de ponte. L'échantillonnage doit apporter deux informations : l'une géométrique, relative aux limites du (ou des) banc(s), et l'autre énergétique, relative aux densités d'œufs sur le (ou les) banc(s). Les contraintes, tant du milieu que des manipulations à bord, sont telles que les échantillons ne peuvent pas être prélevés selon un plan bien défini au départ : le plan effectué n'est ni aléatoire ni régulier. Par nature, ce type d'échantillonnage est non orthodoxe. Avec des degrés divers, tous nos échantillonnages pour l'estimation des stocks ont ces caractéristiques. L'intérêt de ce jeu de données est qu'il les concentre.

Chaque valeur représente une densité : c'est le nombre d'œufs comptés par unité de surface du sédiment. Les coordonnées des points sont exprimées en kilomètres par rapport à une origine dans la ria.

Estimation de la moyenne sur zone
à partir d'échantillons non aléatoires

ORGANISATION DE LA FEUILLE EXCEL

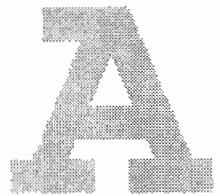
Le fichier contient 8 colonnes. Sur la feuille Excel, la description des colonnes est la suivante (fig. 11) :

	B	C	D	E	F	G	H
1	26.1	17.14	1286			25.9	17.05
2	26.03	17.18	1352			26	17
3	26.03	17.14	1588			26.12	17.08
4	26.05	17.09	1039			26.2	17.08
5	26.19	17.18	2064			26.26	17.1
6	26.21	17.11	1216			26.25	17.16
7	26.13	17.13	1353			26.34	17.2
8	26.15	17.12	1529			26.4	17.17
9	26.05	17.21	723			26.51	17.26
10	26.08	17.11	798			26.34	17.35
11	26.19	17.32	966			26.24	17.35
12	26.22	17.26	1165			26.24	17.54
13	26.01	17.05	631			26.14	17.54
14	26.21	17.39	925			26.14	17.23
15	26.25	17.29	1103			26.09	17.19
16	26.27	17.22	1357			26.05	17.29
17	26.34	17.3	1786			25.94	17.2
18	26.35	17.25	1355			25.95	17.12
19	26.38	17.24	1342			25.8	17.05
20	26.41	17.23	1399				

FIGURE 11 -
Données
du fichier
HEREGG.TXT
dans une feuille
de calcul Excel.
B et C =
abscisses
et ordonnées (km)
des points
échantillons,
D = densité
en œufs
(nb par surface
élémentaire),
G et H =
abscisses
et ordonnées
des sommets
du polygone
délimitant
le champ d'étude.

- colonne A : numéro de la station ;
- colonnes B et C : coordonnées des stations (en kilomètres), respectivement x et y ;
- colonne D : densités des œufs ;

Estimation de la moyenne sur zone
à partir d'échantillons non aléatoires



- colonnes E et F : vides ;
 - colonnes G et H : coordonnées x et y d'un premier polygone définissant le banc de ponté ;
 - colonnes I et J : coordonnées x et y d'un second polygone qui correspond à une seconde hypothèse géométrique définissant le banc de ponté.
- Les polygones ont été définis avec EVA.

IMPORTATION DES DONNÉES ET CHOIX DE LA MÉTHODE GÉOSTATISTIQUE

On importe successivement les colonnes B, C et D d'Excel dans les colonnes X, Y, Valeurs de la fenêtre Données d'EVA pour le type de fichier Données échantillons. On résume rapidement la suite des opérations à effectuer qui a été détaillée dans les chapitres II et III.

Dans Excel, sélectionner la colonne B puis faire Édition-Copier. La colonne est dans le presse-papiers. Faire Alt + Tab pour revenir dans le gestionnaire de programmes de Windows et lancer EVA. Dans EVA, cliquer le bouton d'exécution Données et sélectionner le type de fichier Données échantillons. Cliquer dans la colonne X puis dans le bouton Coller ou Coller avec liaison, par exemple ici, choisir Coller. Le signal sonore indique que le collage est terminé. Faire Alt + Tab et basculer dans Excel. Sélectionner alors la colonne C puis faire Édition-Copier. Faire à nouveau Alt + Tab pour revenir dans EVA. On est directement dans la fenêtre Données. Cliquer alors dans la colonne Y puis dans le

bouton Coller. Retourner une troisième fois dans Excel et recommencer ces opérations pour importer la colonne D que l'on viendra coller dans la colonne Valeurs de EVA. Cliquer ensuite dans le bouton Fermer. Les données sont importées (fig. 12).

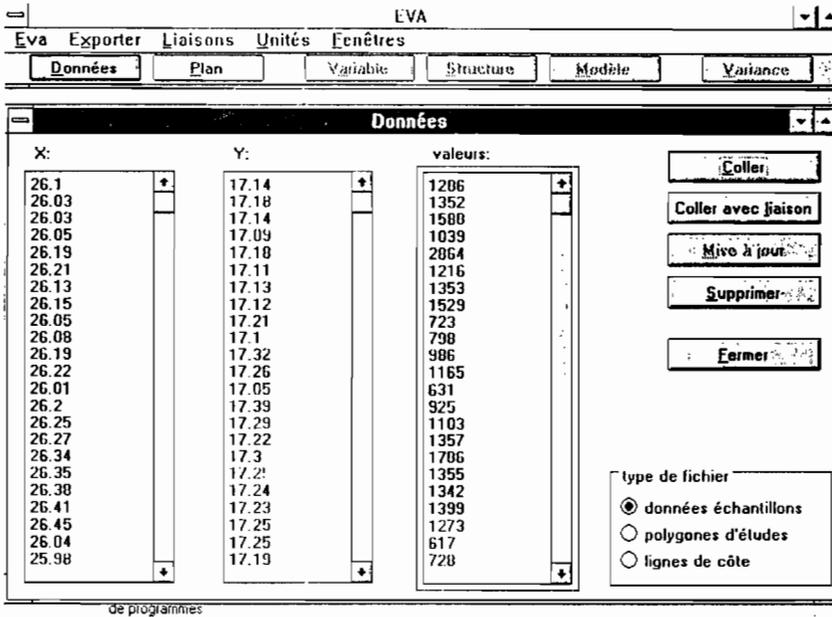


FIGURE 12 - Bouton d'exécution Données. Fenêtre Données après avoir importé chaque colonne. Cliquer sur polygone d'étude et importer les colonnes G et H, puis cliquer sur Fermer.

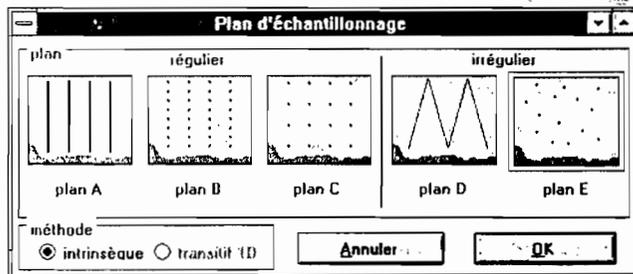
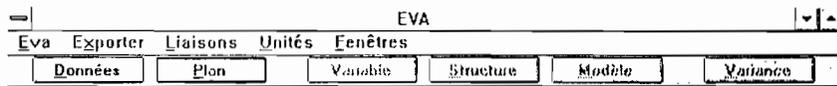
Cliquer dans le bouton Plan. Choisir le plan E : seule la méthode Intrinsèque est accessible (fig. 13). Cliquer dans OK. Les autres boutons d'exécution sont maintenant dégrisés et l'étude commence.

Si, au cours de l'étude, on revient dans la fenêtre Plan et que le choix du plan est modifié, toutes les fenêtres se fermeront. Ceci est logique dans la mesure où changer de plan consiste à entreprendre une autre étude.

Estimation de la moyenne sur zone
à partir d'échantillons non aléatoires

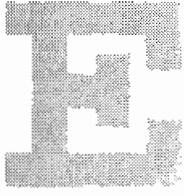
IV-6

FIGURE 13 -
Bouton
d'exécution
Plan. Cliquer
sur Plan E,
puis sur OK.
Seule la
méthode
intrinsèque
est possible.



Le bouton Plan permet de choisir entre les méthodes intrinsèque ou transitive. Il ne donne accès à la méthode transitive que pour le plan régulier à transects échantillonnés en continu. Il permet donc de croiser le plan A avec l'une ou l'autre méthode. Pour tous les autres plans, il ne donne accès qu'à la méthode intrinsèque. Dans la méthode intrinsèque, quelle que soit l'icône de plan choisie, ce sera toujours la même formule de la variance d'estimation qui sera utilisée (chapitre I). Pour un jeu de données avec la méthode Intrinsèque, l'icône de plan choisie n'aura pas d'importance dans les calculs. Les icônes rappellent seulement que l'on travaille avec des plans à effort d'échantillonnage homogène. Si on ne peut choisir une icône correspondant au plan d'échantillonnage de

Estimation de la moyenne sur zone
à partir d'échantillons non aléatoires



ses données, on doit se demander s'il ne serait pas judicieux d'effectuer un prétraitement, comme expliqué au début du chapitre I.

ANALYSE STRUCTURALE

Bouton Variable

L'étape du bouton Variable sert à disséquer les données. À la suite de cette étape, on doit pouvoir décrire la régionalisation avec des mots, donc savoir ce que l'on veut modéliser au travers du variogramme ou pourquoi le variogramme est un outil peu adapté à la modélisation que l'on souhaite entreprendre.

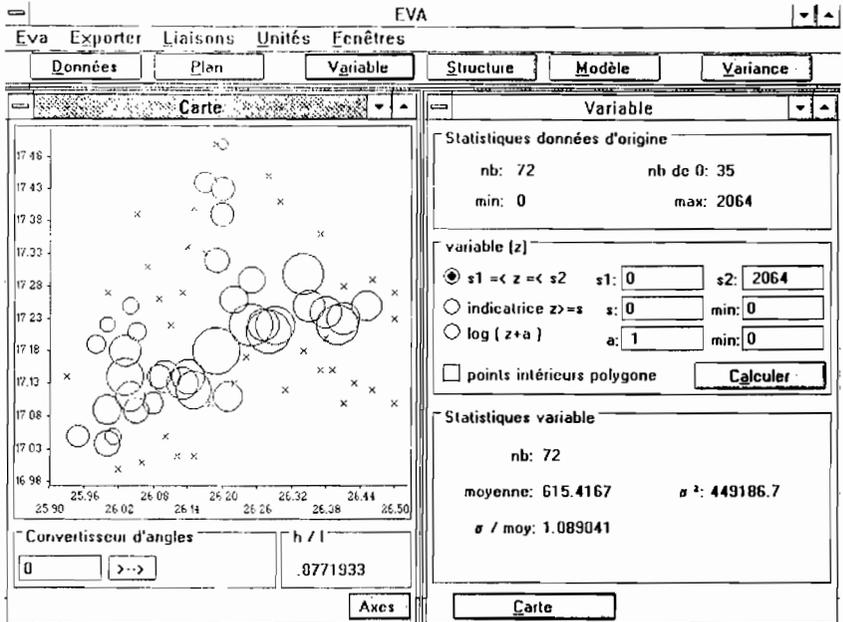
Deux fenêtres sont associées au bouton d'exécution Variable (fig. 14). La fenêtre Variable à droite offre la possibilité de travailler sur les données en définissant des variables de la variable. Sont en même temps fournies les statistiques élémentaires relatives aux variables de la variable choisies. La fenêtre Carte à gauche fournit une représentation spatiale des variables de la variable choisies. Les valeurs sont représentées par des cercles dont le rayon est proportionnel à la valeur. La représentation est relative à la valeur maximale : un rayon unité est attribué à la valeur maximale et tous les autres s'en déduisent par proportionnalité. Si toutes les valeurs du jeu de données sont identiques, la représentation cartographique n'est pas fournie.

Les calculs qui suivront dans les étapes successives de l'étude (boutons Structure, Modèle, Variance) s'effectueront sur la variable qui aura été retenue dans la fenêtre Variable.

Estimation de la moyenne sur zone
à partir d'échantillons non aléatoires

IV-8

FIGURE 14 -
 Bouton
 d'exécution
 Variable.
 Cliquer
 sur Calculer :
 les statistiques
 relatives
 au choix effectué
 dans le champ
 variable (z)
 sont affichées.
 Cliquer alors
 sur Carte pour
 la représentation
 spatiale.



La fenêtre Variable est découpée en trois zones. La première appelée Statistiques données d'origine donne des caractéristiques élémentaires du jeu de données. La deuxième appelée Variable (z) permet de définir une variable des données. Le bouton Calculer donne accès à la troisième zone appelée Statistiques variable, où sont calculées les statistiques élémentaires de la variable choisie. Le bouton Carte ouvre la fenêtre Carte où l'on visualise une représentation de la distribution spatiale de la variable choisie. Dans la zone Variable (z), on peut effectuer une sélection sur la valeurs des échantillons et/ou sur leur position.

Estimation de la moyenne sur zone
 à partir d'échantillons non aléatoires

OPTION POINTS INTÉRIEURS POLYGONE DE LA FENÊTRE VARIABLE

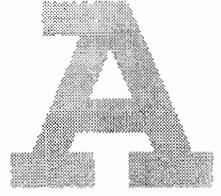
Elle permet de ne retenir que les points intérieurs au polygone fourni (i.e. zone sur laquelle on fait l'estimation). Les statistiques et l'analyse structurale s'effectueront sur le jeu de données réduit aux échantillons intérieurs. Par contre, le calcul de la variance d'estimation s'effectue toujours avec les échantillons intérieurs au polygone, même si l'option n'a pas été sélectionnée. La raison en est que l'on peut accéder au calcul de la variance d'estimation sans passer par l'analyse structurale si l'on connaît déjà le modèle et que l'on désire tester des plans d'échantillonnage.

L'option permet de distinguer les zéros intérieurs à la zone de présence (trous de densité) aux zéros extérieurs, qui sont hors de la zone de présence et qui marquent les limites de celle-ci. Cette option peut être croisée avec les options suivantes.

OPTION $s1 \leq z \leq s2$ DE LA FENÊTRE VARIABLE

Elle permet de tronquer les données et de ne retenir dans les calculs que les données entre les deux seuils, $s1$ et $s2$. La représentation proportionnelle des valeurs est relative à $s2$. Ceci permet de garder l'information quantitative entre les quantiles et de voir à partir de quelle valeur de $s2$ l'aspect de la régionalisation change. On peut ainsi mieux comprendre la régionalisation de la toile de fond des valeurs moyennes que peuvent masquer les fortes valeurs. On peut ainsi donner un sens biologique au terme « fortes valeurs » si elles se comportent différemment des autres dans l'espace. Sur la carte, les valeurs en dehors des bornes sont représentées par les symboles supérieurs ou inférieurs. Les zéros du fichier d'origine sont symbolisés par des croix.

Estimation de la moyenne sur zone
à partir d'échantillons non aléatoires



OPTION INDICATRICE $z \geq s$ DE LA FENÊTRE VARIABLE

Elle permet de travailler sur la variable indicatrice d'un quantile. L'indicatrice vaut 1 quand la valeur de l'échantillon est supérieure à s , et vaut 0 sinon.

Les statistiques fournies sur cette variable sont des pourcentages. Soit n le nombre de données et p le nombre de valeurs supérieures à s . Si on range les valeurs par ordre croissant, on a :

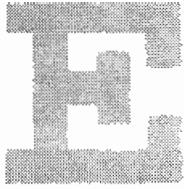
$$P(z \geq s) = p / n \quad ; \quad \%Q = \sum_{i=p}^n z(i) / \sum_{i=1}^n z(i)$$

$$\% \text{ var} = \sum_{i=p}^n (z(i) - m)^2 / \sum_{i=1}^n (z(i) - m)^2$$

avec m la moyenne arithmétique des n données d'origine z .

La quantité $\% Q$ caractérise la contribution des valeurs supérieures à s à la quantité totale. La quantité $\% \text{ var}$ caractérise la contribution des valeurs supérieures à s dans la variance des données. Ces paramètres décrivent l'histogramme. La case « min » permet de définir un seuil en dessous duquel les valeurs ne sont pas comptabilisées.

La carte de l'indicatrice comporte trois symboles. Les cercles représentent les 1 et les points représentent les 0 de l'indicatrice. Les symboles inférieurs marquent les valeurs inférieures au minimum choisi. Les croix marquent les zéros des données d'origine. La représentation spatiale d'une indicatrice permet d'une part de visualiser l'organisation spatiale du quantile concerné. Le variogramme de l'indicatrice décrira géométriquement les ensembles aléatoires que définissent les 1. D'autre part, lorsque l'on fait varier le seuil, on visualise alors l'agencement spatial des différents quan-



tiles et on comprend mieux leurs liens, comment ils s'agencent les uns par rapport aux autres. La zone Statistiques variable nous fournit en même temps les contributions de chaque quantile. En particulier, en diminuant le seuil s progressivement, on voit où sont les graines de fortes valeurs et si les valeurs plus faibles s'agencent autour (modèle de diffusion) ou non (modèle sans transition). Cette dissection quantile à quantile peut donner des arguments naturalistes pour une stratification des valeurs et pour s'orienter vers un certain type de modèle géostatistique non linéaire, si besoin était. Cet outil permet de se forger une opinion sur ses données mais pas d'effectuer une étude non linéaire complète.

OPTION $\text{Log}(z+a)$ DE LA FENÊTRE VARIABLE

Elle permet de travailler sur les logarithmes népériens. On peut effectuer une translation, a , avant de prendre les Logs. On peut définir un minimum : les valeurs en dessous de ce seuil ne seront pas prises en compte dans les calculs. En particulier, on peut ainsi garder les zéros ou les enlever.

FENÊTRE CARTE

Quatre types de symboles sont utilisés pour représenter les valeurs sur la carte. Les cercles sont une visualisation proportionnelle aux valeurs positives sélectionnées, et relative à la valeur maximale sélectionnée dans la fenêtre Variable. Les croix représentent les zéros du fichier d'origine, quelles que soient les sélections effectuées. Les symboles inférieurs et supérieurs marquent les échantillons en dehors des bornes choisies dans la fenêtre Variable.

Dans la fenêtre, la carte est un objet carré : le rapport entre les dimensions réelles des axes n'est pas respecté. Une direction apparente sur

Estimation de la moyenne sur zone
à partir d'échantillons non aléatoires

IV-12

la carte n'est pas nécessairement égale à la direction réelle. Le convertisseur d'angle permet de transformer une direction apparente à l'écran en direction réelle. Le rapport h/l est un rapport d'anisotropie entre la longueur de l'axe y , en distance réelle, et la longueur de l'axe x , en distance réelle.

Le bouton Axes permet de modifier les échelles de graduation sur les axes. De plus, il donne accès à la possibilité de représenter la côte et le fichier Polygone.

Outre le fait qu'il permet de mieux la voir, l'agrandissement de la carte, à l'aide de la flèche qui pointe vers le haut en haut à droite de la fenêtre Carte, permet d'accéder à une nouvelle fonction. Le champ Position souris donne la position de la souris là où on clique dans la carte. Cette fonction permet de définir rapidement un polygone et de faire plusieurs hypothèses géométriques (séparation des zéros en zéros intérieurs et extérieurs). La procédure est manuelle : il faut recopier les positions à la main sur une feuille, puis aller dans Excel et les entrer dans la feuille de calcul. Puis on les importe dans EVA, par la procédure d'importation déjà décrite (chapitres II, III, IV). Il faudra choisir le type de fichier Polygone d'étude dans la fenêtre Données. Le bouton Variable dans la fenêtre Carte agrandie permet de basculer dans la fenêtre Variable. Le bouton Axes donne accès à la même fenêtre Axes que précédemment. La double flèche qui pointe vers le haut et vers le bas en haut à droite de la fenêtre Carte agrandie permet de faire revenir la fenêtre à sa taille normale (fig. 15).

EXEMPLE

Dans un premier temps, on visualise toutes les données : s_1 et s_2 sont les valeurs par défaut qui correspondent au minimum et maximum des

Estimation de la moyenne sur zone
à partir d'échantillons non aléatoires

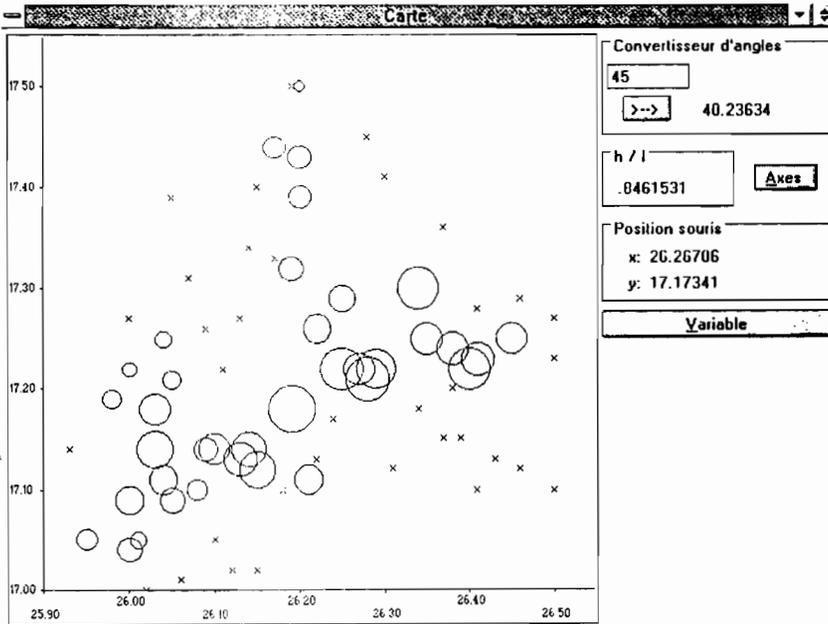


FIGURE 15 -
Agrandissement
plein écran
de la fenêtre
Carte obtenue
en cliquant
sur le bouton
possédant
la flèche
vers le haut,
en haut à droite
de la fenêtre
Carte. Cliquer
sur la souris
n'importe où
dans la carte,
sa position
est marquée
dans le champ
Position souris.

valeurs du fichier. On clique sur le bouton Calculer : les statistiques du jeu total nous sont fournies. On clique sur le bouton Carte : la fenêtre Carte nous fournit une représentation proportionnelle des valeurs (fig. 14). On clique dans le bouton Axes, la fenêtre Axes s'ouvre, on y modifie comme suit les graduations de façon à avoir une figure plus propre : rentrer les valeurs suivantes $X_{\min} = 25.9$, $Y_{\min} = 17$, $X_{\max} = 26.55$, $Y_{\max} = 17.55$, pas sur $X = 0.1$, pas sur $Y = 0.1$, puis faire OK. On décrit la régionalisation.

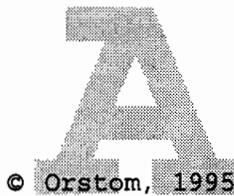
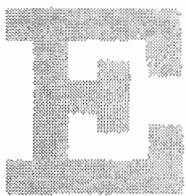
Il y a 49 % d'échantillons nuls. Les zéros entourent géographiquement une zone constituée essentiellement de valeurs positives. On interprète

Estimation de la moyenne sur zone
à partir d'échantillons non aléatoires



donc les zéros comme des zéros extérieurs au banc de ponte qui est la zone aux valeurs positives. Celles-ci s'alignent sur un axe principal d'orientation apparente nord-est-sud-ouest. Le convertisseur d'angle donne un angle de 40° pour un angle apparent de 45° (les angles sont comptés dans le sens trigonométrique à partir de l'axe des X). De cet axe principal partent deux langues de ponte vers le nord-nord-ouest. Les valeurs positives sont entourées de zéros qui parfois les touchent de très près sur les bordures. Mais l'intérieur de l'axe ne contient pas de zéros. On est confronté à un problème géométrique : il y a une anisotropie marquée du banc de ponte (sans doute à cause de la répartition particulière du sédiment, mais nous n'avons pas d'information sur cette variable), ce qui influence un peu la régionalisation des œufs sur le banc et, d'autre part, la limite du banc n'est pas toujours bien claire d'après l'échantillonnage effectué.

Si on effectue l'analyse structurale sur le jeu total, on mélangera l'information structurale due à la géométrie propre du banc (structure de l'indicatrice des valeurs positives dans l'ensemble du jeu de données) avec l'information structurale des œufs sur le banc (structure des valeurs à l'intérieur du banc). Si on définit un polygone traçant les contours du banc, en ne travaillant que sur les données intérieures, on ne caractérisera que la structuration sur le banc de ponte. C'est cette approche qui est suivie. On fera l'estimation dans le polygone définissant les contours du banc. On pourra ensuite modifier ce polygone pour mesurer l'influence, sur l'estimation et sa précision, de l'indétermination des limites du banc. Pour les plans réguliers, on peut calculer l'influence de l'erreur géométrique sur l'estimation de la moyenne, en utilisant la formule approchée donnée par MATHERON (1965, 1971). On la trouvera aussi dans JOURNEL et HUIJBREGTS (1978), ARMSTRONG *et al.* (1992) ou PETITGAS et RIVOIRARD (1991). EVA ne permet pas de calculer ce terme d'erreur.



Dans un deuxième temps, on s'attache à définir les limites du banc qui constitueront les limites du polygone d'estimation. On procède de la façon suivante :

— dans la fenêtre Carte de EVA, on agrandit la fenêtre en cliquant sur la flèche qui pointe vers le haut en haut à droite de la fenêtre Carte. On définit un polygone grâce à la fonction Position souris (fig. 15). On clique dans la carte aux endroits qui marquent les sommets du polygone que l'on veut définir. On note les positions affichées sur ses tablettes. Par exemple, on choisit un polygone qui suive les valeurs positives et qui laisse quelques valeurs nulles à l'intérieur ;

— Alt + Tab : on revient dans la feuille Excel. On choisit une colonne pour y écrire les valeurs des X et une colonne pour y écrire les valeurs des Y. Le fichier EGGHER.TXT contient déjà deux polygones (voir chapitre IV, « Organisation de la feuille Excel ») en colonnes G et H et en colonnes I et J. Ils ont été définis selon la procédure précédente. On sélectionne la colonne G et on la copie dans le presse-papiers à l'aide de la fonction Édition-Copier ;

— Alt + Tab : on retourne dans EVA. On redimensionne la taille de la fenêtre Carte de façon à pouvoir avoir accès à la barre d'exécution. On ouvre la fenêtre Données en cliquant sur le bouton Données. On y sélectionne le type de fichier Polygone d'étude en cliquant dessus ou dans le rond qui lui correspond : le rond devient noir, la sélection est active. Deux colonnes apparaissent, l'une pour les X et l'autre pour les Y. On sélectionne X en cliquant dedans, puis on clique dans le bouton Coller avec liaison car plus tard on voudra sans doute faire l'estimation avec un autre polygone ;

— on recommence pour la colonne Y : Alt + Tab, retour à Excel, sélection de la colonne H, copie de cette colonne dans le presse-papiers,



Alt + Tab, retour à EVA, sélection de la colonne Y du fichier Polygone d'étude, et Copie avec liaison ;

— on clique dans le bouton Fermer de la fenêtre Données de EVA. Le polygone est importé.

On rappelle les fenêtres Carte et Variable en cliquant, par exemple, sur le bouton d'exécution Variable. La variable n'est plus à jour : Il faut recliquer sur le bouton Calculer. Le bouton Axes de la fenêtre Carte permet d'accéder à l'option de visualisation des limites du polygone. Dans la fenêtre Axes, l'option Afficher polygone est maintenant dégrisée et accessible. On la sélectionne en cliquant dessus ou dans la case correspondante : une croix apparaît dans la case indiquant que la sélection est active. Faire alors OK. La carte est retracée avec les limites du polygone.

Dans un troisième temps, on peut sélectionner uniquement les points intérieurs au polygone grâce à l'option correspondante dans la fenêtre Variable : on sélectionne cette option en cliquant dedans, une croix apparaît dans la case correspondante. On travaillera désormais sur ce fichier réduit : les zéros extérieurs en sont éliminés. En cliquant sur le bouton Calculer, on obtient les statistiques pour ce fichier réduit : $n = 46$, $m = 963$, $\text{var} = 3.66 \text{ E5}$. La carte est retracée et les zéros extérieurs y sont figurés par des croix grisées (fig. 16).

Dans un quatrième temps, on peut disséquer statistiquement et spatialement le fichier réduit en utilisant la sélection « indicatrice » de la fenêtre Variable. On clique sur le mot « indicatrice » ou dans le rond qui lui correspond, celui-ci devient noir. Laisser le minimum à zéro et prendre un seuil de 2 000. Faire Calculer : les chiffres sont relatifs au fichier réduit. La carte montre qu'une seule valeur est supérieure à ce seuil et qu'elle est située au centre de l'axe du banc. Cette forte valeur

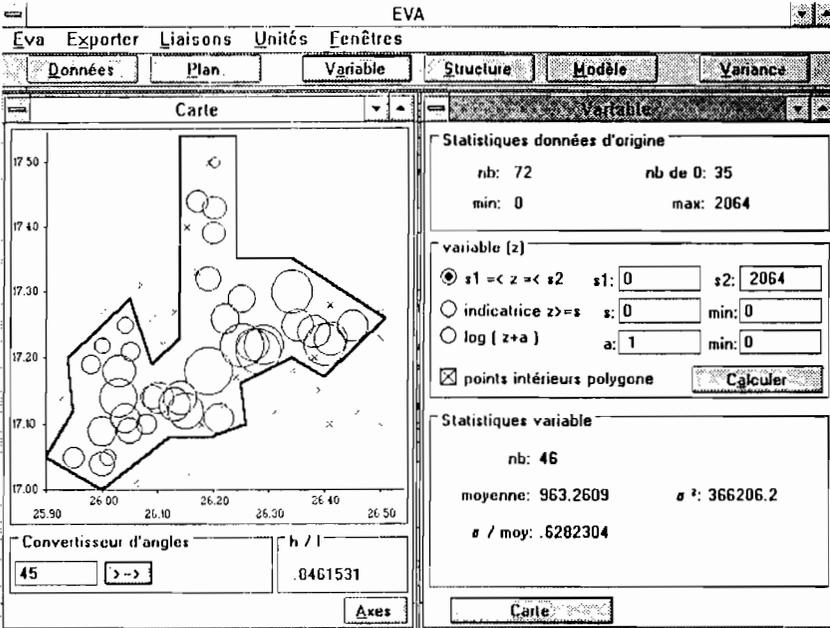


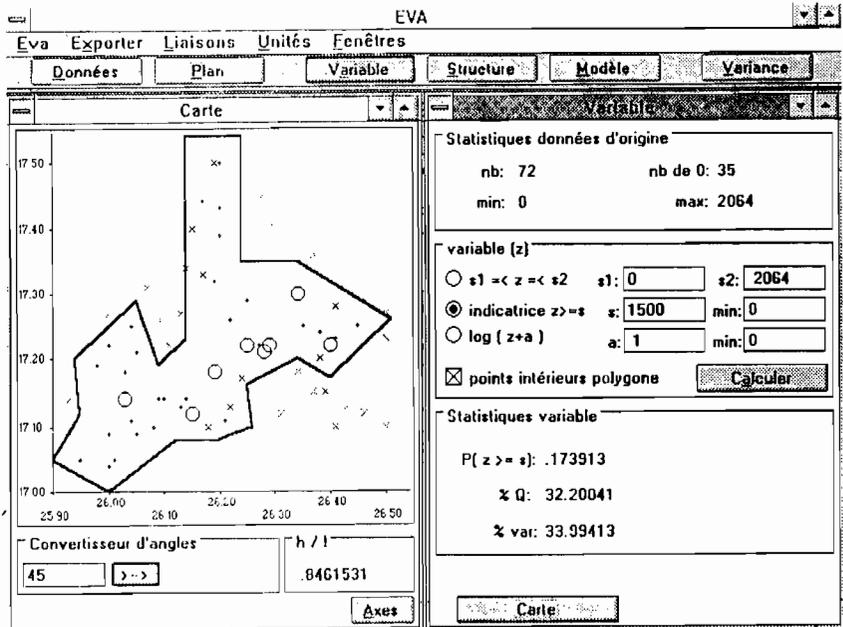
FIGURE 16 -
Bouton
d'exécution
Variable.
Modifier
vos sélections
dans le champ
Variable (z).
Cliquez
sur Points
intérieurs
polygone,
puis sur
Calculer.
Cliquez
sur Axes
et choisissez
l'option
Afficher
polygone,
puis OK.

représente 5 % de la quantité totale et 7 % de la variance dans le polygone défini (i.e. sur le banc). Choisir un seuil plus bas, par exemple 1 500. Faire Calculer. Ces huit fortes valeurs représentent le tiers de la moyenne et de la variance sur le banc. Les valeurs immédiatement inférieures à la plus forte ne sont pas implantées dans son voisinage immédiat. Elles se répartissent plutôt homogènement sur tout l'axe principal du banc. Les axes secondaires ne montrent pas de valeurs fortes. La bordure sud du banc est très abrupte car des valeurs fortes sont très proches de la limite et côtoient des zéros. La bordure nord est plus douce. L'information spatiale principale est là (fig. 17).

Estimation de la moyenne sur zone
à partir d'échantillons non aléatoires

IV-18

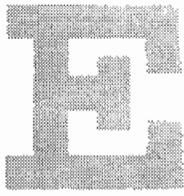
FIGURE 17 -
 Bouton
 d'exécution
 Variable.
 Modifier
 vos sélections
 dans le champ
 Variable (z).
 Choisir
 Indicatrice
 $Z > = z$,
 et entrer
 la valeur 1 500,
 puis cliquer
 sur Calculer.



Les densités d'œufs forment un dôme elliptique dont le centre est la plus forte valeur, dont le grand axe est l'axe principal, dont la bordure sud est un peu tronquée et dont la bordure nord est plus douce et qui possède deux ramifications. Cette régionalisation peut être modélisée de différentes façons.

Avant de passer à la suite, sélectionner à nouveau l'option $s_1 \leq z \leq s_2$, car c'est sur cette variable que l'on désire travailler. L'option Points Intérieurs polygone est toujours active. Faire Calculer à nouveau. Le jeu réduit est en mémoire.

Estimation de la moyenne sur zone
 à partir d'échantillons non aléatoires



P. Petitgas, A. Prampart



© Orstom, 1995

Bouton Structure

Cliquer le bouton d'exécution Structure. Deux fenêtres y sont associées : la fenêtre Carte à droite qui est la même que précédemment et la fenêtre Variogramme à gauche. Ceci permet de visualiser la régionalisation tout en la caractérisant à l'aide d'un variogramme.

Le variogramme est une variance entre les points en fonction de la distance qui les sépare. Les coordonnées des points sont en kilomètres. On calculera les distances dans la même unité. On active donc ce choix dans la barre de menu : faire Unités-coordonnées et distances, puis cliquer sur l'option Identiques. Une marque montre que ce choix est désormais actif.

Les paramètres pour le calcul du variogramme expérimental sont à fournir dans la zone Paramètre structure. Ils ont été définis au chapitre I. Les valeurs des paramètres choisis sont les suivantes : distances : nb. pas = 6 ; pas = 0.05 ; tolérance pas = 0.025. La distance maximale calculée sera de 3 km correspondant à peu près à la moitié du grand axe. Angle : isotrope.

Le bouton Valeurs ouvre une fenêtre donnant les valeurs du variogramme expérimental. Le bouton Axes ouvre une fenêtre permettant de modifier les graduations de la figure Variogramme : D signifie distance et V signifie valeurs du variogramme. On clique sur ce bouton et on modifie les graduations sur les distances. Prendre $D_{max} = 0.32$ et pas sur les distances = 0.05 puis faire OK (fig. 18).

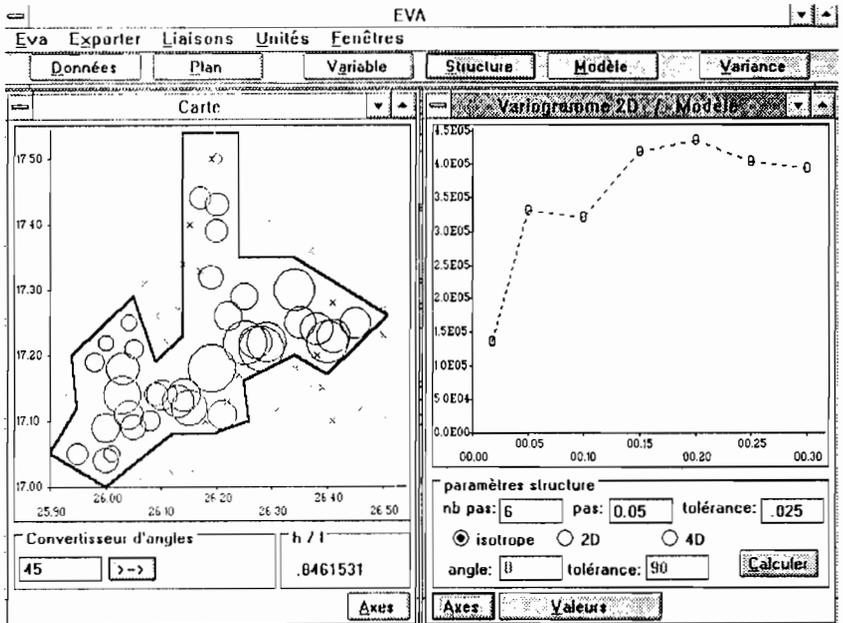
Bouton Modèle

Deux fenêtres y sont associées : la fenêtre Variogramme précédente à gauche et la fenêtre Paramètres modèle à droite. Dans la fenêtre

Estimation de la moyenne sur zone
à partir d'échantillons non aléatoires

Variogramme, un nouveau bouton est apparu : le bouton Modèle. Il permet de basculer entre la fenêtre des valeurs du variogramme et celle des paramètres du modèle : on peut ainsi régler les paramètres pour faire passer le modèle par certains points expérimentaux, exactement.

FIGURE 18 -
 Bouton d'exécution Structure.
 Entrer les paramètres choisis pour le calcul du variogramme expérimental et cliquer sur Calculer. Le variogramme est calculé sur la sélection choisie dans la fenêtre Variable (ici, les sélections actives sont : $s1 \leq Z \leq s2$, $s1 = \min$, $s2 = \max$ et points intérieurs au polygone).



La signification des paramètres des modèles a été exposée au chapitre I. On choisit les types de modèle dans les listes déroulantes qui sont accessibles en cliquant sur les flèches pointant vers le bas dans les zones modèle 1, 2 ou 3. Pour sélectionner un type, il suffit de cliquer sur son nom. La première ligne de la liste est vide. Lorsqu'elle est active, aucun modèle n'est défini. Le type de modèle choisi, la zone modèle

Estimation de la moyenne sur zone à partir d'échantillons non aléatoires

s'ouvre et on peut alors spécifier ses paramètres. On choisit par exemple : pépite = $5 E4$ + modèle 1 = exponentiel (palier = $3.7 E5$, portée pratique = 0.15). Le modèle est isotrope ; les paramètres d'anisotropie sont : coeff. X = coeff. Y = 1, rotation = 0. Le bouton Tracer modèle permet de visualiser l'ajustement. On modifie interactivement les paramètres du modèle jusqu'à ce que l'ajustement soit jugé satisfaisant par l'utilisateur (fig. 19).

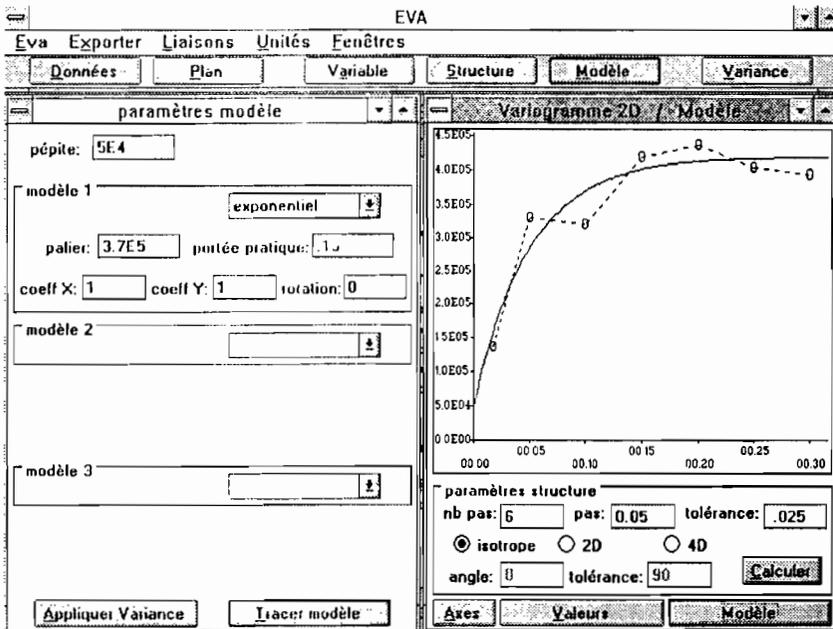


FIGURE 19 -
Bouton
d'exécution
Modèle.
Ajuster
le modèle
visuellement
de façon
interactive.
Lorsque
vous avez
retenu
un modèle,
cliquer
sur Appliquer
variance.

Le bouton Appliquer variance permet de transmettre à l'étape du calcul de la variance d'estimation les paramètres du modèle choisi : il ne sera pas besoin de les recopier plus tard.

Estimation de la moyenne sur zone
à partir d'échantillons non aléatoires

CALCUL DE LA VARIANCE D'ESTIMATION

Le bouton d'exécution Variance conduit au calcul de la variance d'estimation. Ce calcul n'est possible que si un polygone et des points échantillons ont été importés (voir chapitres I et II).

Dans le calcul de la variance d'estimation, seuls sont pris en compte les points intérieurs au polygone, même si cette option n'a pas été sélectionnée dans la fenêtre Variable. La raison en est que, pour tester un plan d'échantillonnage, on peut accéder à la variance d'estimation par le chemin 2 (voir chapitres I et II), c'est-à-dire simplement avec un fichier de points, un polygone et un modèle de variogramme. Si l'analyse structurale est effectuée sur le fichier total des données et si un polygone n'englobant pas toutes les valeurs est défini, la variance d'estimation sera calculée avec les données Intérieures au polygone et avec un modèle relatif à l'ensemble des valeurs.

Deux fenêtres sont associées au bouton d'exécution Variance.

I Fenêtre Type d'estimation

Le bouton Variance ouvre une première fenêtre qui s'appelle Type d'estimation. Elle se présente sous forme d'un tableau qui résume les hypothèses des différents types d'estimation afin de replacer les méthodes utilisées dans un ensemble plus large (fig. 20).

On doit spécifier le type d'estimation que l'on effectue : estimation linéaire par la moyenne arithmétique ou estimation lognormale, dans le cas de données corrélées. Les autres cas sont grisés et inaccessibles. On clique dans la colonne Choix, dans l'un des deux ronds correspondant à l'un des deux cas dégrisés. Le rond devient noir : la sélection est

Estimation de la moyenne sur zone
à partir d'échantillons non aléatoires

active. On peut aussi bien cliquer sur les mots « oui », « indifférent », « lognormal », pour sélectionner la ligne. On clique ensuite sur OK. On accède alors au calcul de la variance d'estimation pour le cas sélectionné. Choisir linéaire puis faire OK. Une seconde fenêtre s'ouvre, appelée Variance structure spatiale.

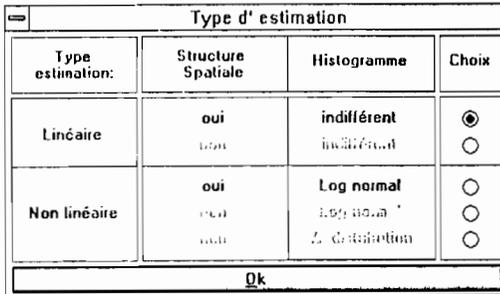
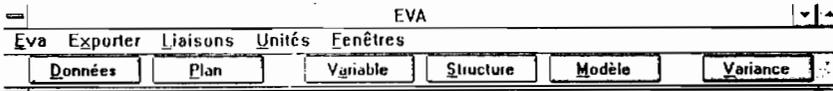


FIGURE 20 - Bouton d'exécution Variance. Première fenêtre permettant de choisir le type d'estimation à effectuer. Le variogramme a été calculé sur la variable brute. Cliquer sur la ligne Estimation linéaire et Histogramme indifférent, puis sur OK.



Fenêtre Variance structure spatiale

La fenêtre Variance structure spatiale, dans le cas de l'estimation linéaire, est composée de trois champs. Le champ à droite nommé Paramètres modèle est identique au contenu de la fenêtre Paramètres modèle.

Estimation de la moyenne sur zone
à partir d'échantillons non aléatoires

Comme on avait cliqué sur le bouton Appliquer variance, on note que les paramètres du modèle ont bien été transmis. La partie droite est composée de deux champs. Les résultats sont donnés dans le champ inférieur nommé Variance d'estimation. Pour la signification des résultats fournis, on se reportera au chapitre I. Le champ supérieur nommé Discrétisation permet de définir les paramètres de la discrétisation du polygone. Elle sert au calcul numérique des termes G_{VV} et G_{GV} de la formule de la variance d'estimation, comme expliqué au chapitre I (fig. 21).

FIGURE 21 -
Bouton
d'exécution
Variance.
Seconde fenêtre
permettant
le calcul
de la variance
d'estimation.
Entrer
les paramètres
de la discrétisation
et cliquer sur
 G_{VV} normé pour
contrôler si la
discrétisation
est suffisante.
Puis, cliquer
sur Calcul
variance.

EVA

Eva Exporter Liaisons Unités Fenêtres

Données Plan Variable Structure Modèle Variance

Variance structure spatiale

paramètres modèle pépite: 50000

modèle 1 exponentiel

palier: 370000 portée pratique: .15

coeff X: 1 coeff Y: 1 rotation: 0

modèle 2

modèle 3

discrétisation, sélection points intérieurs polygone

longueur sur x: 0.61 nb cases x: 20

longueur sur y: 0.54 nb cases y: 20

Gvv normé -> .9254123

Calcul variance

variance d'estimation

σ^2_{est} : 6202.769

pépite: 17.5% surface: 00.12

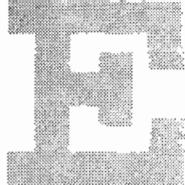
modèle 1: 82.5% Gvv: 392402.6

modèle 2: 0.0% Gab: 385207.1

modèle 3: 0.0%

Cliquer à nouveau sur le bouton d'exécution Variance : la fenêtre Type d'estimation s'ouvre à nouveau. Choisir maintenant l'estimation non

Estimation de la moyenne sur zone
à partir d'échantillons non aléatoires



linéaire lognormale et faire OK. Une fenêtre Variance structure spatiale s'ouvre qui correspond au calcul de la variance d'estimation pour des données lognormales et corrélées. Cette fenêtre comprend les trois champs précédents plus un quatrième, en haut à gauche, qui donne les paramètres des Logs transmis automatiquement depuis la fenêtre Variable quand la sélection Log est utilisée. Comme on n'a pas utilisé cette sélection, aucune valeur n'a été transmise : les valeurs par défaut sont nulles. Pour la signification des résultats fournis dans le champ Variance d'estimation, on se reportera au chapitre I. Revenir dans l'estimation linéaire : cliquer le bouton d'exécution Variance, puis choisir l'estimation linéaire, puis faire OK.

CHAMP DISCRÉTISATION

La longueur maximale du polygone sur X est donnée dans la case Longueur sur x, et de même pour Y. Ces longueurs sont données dans les unités qui ont été choisies avec la fonction Unités de la barre de menu. On définit une grille de calcul en donnant le nombre de cases sur X et sur Y. Seules les cases dont les centres sont à l'intérieur du polygone sont utilisées dans les calculs. En définissant la grille de calcul par une maille non carrée, on peut vouloir tenir compte de l'anisotropie de la forme du polygone et/ou de l'anisotropie du modèle de variogramme.

Le bouton G_{VV} normé lance le calcul du G_{VV} dont le résultat est fourni pour des modèles de palier unité. C'est un résultat intermédiaire qui permet d'optimiser la grille de calcul. Si la grille définie est très fine, le calcul des G_{VV} et G_{AV} prendra beaucoup de temps. Par contre, si la grille n'est pas assez fine, les calculs numériques ne seront pas assez précis. En augmentant progressivement le nombre de cases de la grille, le G_{VV} normé se modifie mais ses variations deviennent de plus en plus petites. La précision requise sur le G_{VV} normé dépend de la

Estimation de la moyenne sur zone
à partir d'échantillons non aléatoires

valeur du palier du variogramme. Supposons un palier dont l'unité est 10^6 . On demandera une précision de l'ordre de 10^{-4} ou 10^{-5} sur le G_{VV} normé. Il faudra progressivement augmenter le nombre de cases afin que la précision requise sur le G_{VV} normé se stabilise à ce niveau là.

Dans l'algorithme de calcul du G_{VV} par la maille aléatoire stratifiée (voir chapitre I), on fait appel à un générateur de nombres au hasard. Pour une même grille, en cliquant deux fois sur le bouton G_{VV} normé, on aura donc des valeurs différentes. Cette variabilité est un second paramètre de contrôle.

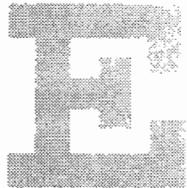
L'étape intermédiaire du calcul du G_{VV} normé n'est pas obligatoire. Si on clique directement sur le bouton Calcul variance après avoir défini une grille de calcul, la variance d'estimation sera calculée avec cette grille. La valeur du G_{VV} normé correspondante sera fournie. Dans ce cas, il est conseillé de prendre une maille fine, 50×50 ou 100×100 . Le G_{VV} normé ayant déjà été calculé, si le contenu de l'une des cases du modèle de variogramme ou de la discrétisation est modifié, le bouton Calculer variance relancera automatiquement le calcul du G_{VV} normé.

Dans l'exemple, on a jugé par essais successifs qu'une grille 20×20 était satisfaisante (fig. 21).

BOUTON CALCUL VARIANCE

Ce bouton lance le calcul de la variance d'estimation, pour la grille de calcul, le modèle défini et le G_{VV} normé calculé. Si le G_{VV} normé n'a pas déjà été calculé ou si le contenu des cases Paramètres modèle ou Discrétisation a été modifié, le bouton Calcul variance relance automatiquement le calcul du G_{VV} normé. La surface est celle du polygone : elle est calculée à partir du nombre de cases Intérieures au

Estimation de la moyenne sur zone
à partir d'échantillons non aléatoires



polygone et est exprimée dans l'unité de distance choisie avec la fonction Unités de la barre de menu.

Soit m^a , la moyenne arithmétique des valeurs Intérieures au polygone. Elle est fournie dans la fenêtre Variable. L'erreur relative d'estimation est :

$$\frac{\sigma_{est}}{m^a} = \frac{\sqrt{6203}}{963} = 8.2\%$$

On note que la valeur du G_{VV} est un peu plus forte que celle de la variance des données mais reste du même ordre de grandeur. Un plan aléatoire avec 46 stations sur le banc de ponton comme défini par le polygone donnerait une précision relative de :

$$\frac{\sqrt{G_{VV}/n}}{m^a} = \frac{\sqrt{3.92F5/46}}{963} = 9.6\%$$

EXPORTATION DES RÉSULTATS

Comme expliqué au chapitre II, on utilise la fonction Exporter ou la touche de Copie d'écran. La fonction Exporter de la barre de menu déroule une liste dans laquelle les noms, Carte, Variogramme et Valeurs variogramme sont dégrisés.

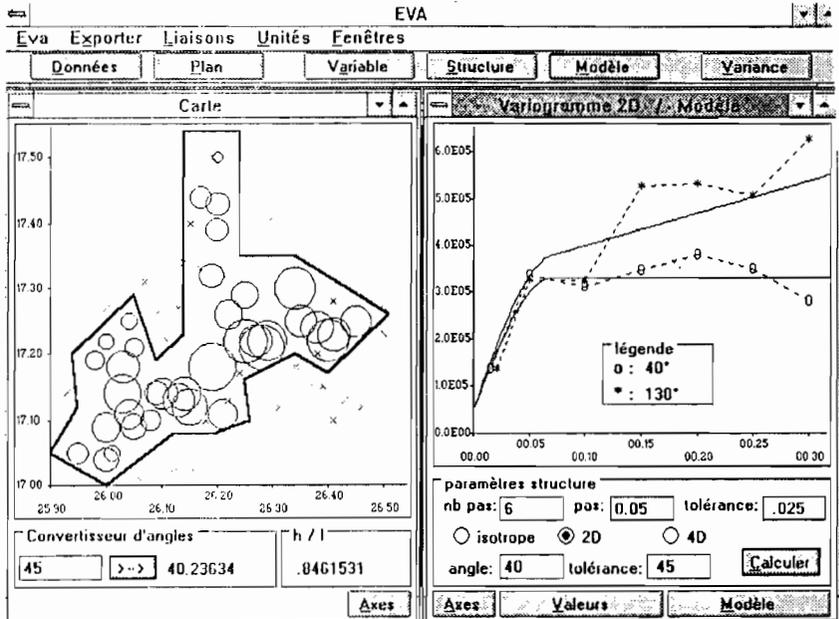
AUTRES MODÈLES, MODÉLISATION ANISOTROPE

À tout moment, on peut revenir à une étape antérieure de l'étude, en rappelant la fenêtre voulue. Revenons par exemple sur le choix du modèle.

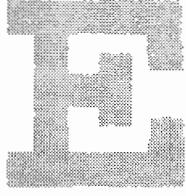
On se propose d'effectuer un calcul de variogramme anisotrope. On rappelle les fenêtres Structure et Carte grâce à la fonction Fenêtre-afficher/mise à jour de la barre de menu, ou tout simplement en cliquant à nouveau sur le bouton d'exécution Structure.

Dans la fenêtre Carte, le convertisseur d'angle nous indique qu'un angle apparent de 45° est un angle réel de 40° . Dans la fenêtre Variogramme, on sélectionne le choix 2 D : le rond devient noir. On inscrit 40° dans la case Angle, et on laisse la tolérance angulaire par défaut de $\pm 45^\circ$. On clique sur le bouton Calculer. Deux courbes apparaissent, une pour chaque direction, à savoir 40 et sa perpendiculaire 130 (fig. 22).

FIGURE 22 -
Modèle
anisotrope.
Les fenêtres
Cartes et
Variogramme/2D
sont apposées.
Cliquer
dans la figure
Variogramme,
la légende
apparaît.



Estimation de la moyenne sur zone
à partir d'échantillons non aléatoires



Cliquer avec la souris dans la figure du variogramme, un cadre de légende apparaît qui indique quelle courbe correspond à quelle direction. Le cadre de légende reste à l'écran tant que l'utilisateur garde le bouton de la souris appuyé. Cette information se trouve aussi dans la fenêtre Axes du variogramme.

Cliquer sur le bouton Valeurs. Dans la fenêtre Valeurs figurent maintenant deux boutons en bas correspondant aux directions du variogramme : sur chacun est marquée la valeur de l'angle. Cliquer sur l'un des boutons : les valeurs du variogramme pour cette direction s'affichent dans la fenêtre et la valeur de l'angle sur le bouton est encadrée par des crochets, indiquant ainsi la direction affichée.

Les trois premiers points du variogramme expérimental sont identiques dans les deux directions : la structure est isotrope jusqu'à 0.1 km. Ensuite, dans la direction 40, on observe un palier alors que le variogramme continue à grimper dans la direction 130. Le variogramme expérimental s'interprète de manière naturaliste : les œufs sont structurés en paquets isotropes dont le diamètre moyen est de l'ordre de 0.05 à 1 km. Dans la direction 40, les paquets ont en moyenne des densités similaires et le variogramme atteint un palier. Dans la direction 130, au-delà de 0.1 km, la distance augmentant, les paquets ont des densités de plus en plus différentes et le variogramme continue à croître.

Ceci quantifie la description faite de la morphologie elliptique du banc de ponte : on avait décrit un axe principal de direction 40 dont la bordure nord était plus douce que la bordure sud.

On propose donc un modèle à anisotropie zonale prenant ces deux aspects en compte. À un modèle isotrope à palier prenant en charge la structuration en paquet sur le banc, on ajoute un modèle linéaire, uniquement dans la direction 130, qui prend en charge dans cette

direction la structuration en pente douce. Le modèle dans la direction 130 signifie bien que les œufs sont répartis en paquets et que, la distance augmentant, ces paquets ont des densités moyennes de plus en plus différentes. Le modèle choisi est (pour le choix des paramètres d'anisotropie, se reporter au chapitre II) :

pépite = 5 E4

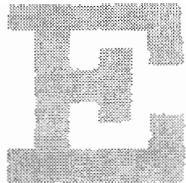
+ sphérique isotrope (palier = 2.8 E5, portée = 0.07) ; coeff. X = coeff. Y = 1, direction = 40

+ puissance (palier = 7 E5, exposant = 1) ; coeff. X = 0, coeff. Y = 1, direction = 40.

La figure 22 montre la fenêtre Variogramme après modification des axes et modélisation du variogramme, à côté de laquelle a été rappelée la fenêtre Carte grâce à la fonction Fenêtre-mise à jour de la barre de menu.

On peut exporter chacune des directions dans une feuille Excel. Faire Exporter (dans la barre de menu), Valeurs variogramme, puis choisir une direction. Le signal sonore indique que les valeurs sont dans le presse-papiers. Basculer dans Excel, ouvrir une nouvelle feuille, cliquer dans une cellule puis faire Édition-Coller. Recommencer pour l'autre direction.

Revenir dans la fenêtre Paramètres modèle de EVA. Cliquer le bouton Appliquer variance. Cliquer dans le bouton d'exécution Variance pour calculer la variance avec ce nouveau modèle. Choisir la ligne de l'estimation linéaire dans la fenêtre Type d'estimation. Dans la fenêtre Variance structure spatiale, comme on garde la grille de calcul de 20 x 20 cases et que le modèle est importé, il suffit de cliquer dans le bouton Calculer variance. Le G_{VV} est un peu plus fort : 4.2 E6.



Le modèle linéaire dans la direction 130 contribue 15 % à la variance d'estimation. L'erreur relative d'estimation est du même ordre que précédemment : 8.4 %.

Le modèle isotrope précédent n'est pas si éloigné du modèle anisotrope choisi ici. Les portées sont similaires et rendent compte de la même structuration isotrope des œufs. Dans le modèle isotrope précédent, l'hétérogénéité plus grande dans la direction 130 était prise en charge par un palier isotrope plus important. Dans ces modèles, l'emplacement des paquets dans le polygone n'est pas spécifié.

AUTRE POLYGONE, AUTRE HYPOTHÈSE GÉOMÉTRIQUE

À tout moment, on peut revenir à une étape antérieure de l'étude. Redéfinissons par exemple un autre polygone.

On retourne dans la fenêtre Carte pour redéfinir un polygone. Il suffit de rappeler cette fenêtre, soit en cliquant sur le bouton d'exécution Variable, soit en utilisant la fonction Fenêtre-Carte de la barre de menu. On procède comme expliqué plus haut en agrandissant la fenêtre. Un second polygone a déjà été défini et il se trouve sur la feuille Excel dans les colonnes I et J.

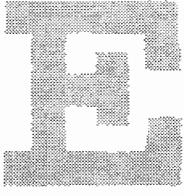
On met à profit le lien établi entre les colonnes G et H de la feuille Excel et les colonnes X et Y du fichier Polygone d'EVA.

Alt + Tab : on bascule dans Excel. Grâce à la fonction Couper-Coller, on déplace le contenu des colonnes G et H dans les colonnes E et F.

IV-32

On déplace le contenu des colonnes I et J dans les colonnes G et H.

Alt + Tab : on revient dans EVA. On utilise la fonction Liaison-mise à jour de la barre de menu. Seules les colonnes du fichier Polygone ont été importées avec liaison : seul Polygone n'est pas grisé dans la liste déroulante. On choisit Polygone. Deux signaux sonores (bip bip) indiquent que les colonnes X et Y sont mises à jour. La variable n'est plus à jour, il faut recommencer la suite des étapes, Variable, Structure, Modèle, Variance : c'est une nouvelle estimation qui est effectuée.

A large, bold, stylized letter 'E' composed of a grid of small squares, positioned in the top left corner.

P. Petitgas, A. Prampart

A large, bold, stylized letter 'V' composed of a grid of small squares, positioned in the top center.A large, bold, stylized letter 'A' composed of a grid of small squares, positioned in the top right corner.

© Orstom, 1995

Test d'un plan
d'échantillonnage
connaissant la structure:
exemple d'utilisation

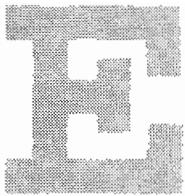
A large, bold, stylized letter 'V' composed of a grid of small squares, positioned in the center of the page.

Cette étude fait suite à l'étude précédente. On a mis en évidence une structuration sur le banc de ponte, caractérisée par le variogramme isotrope exponentiel suivant : $g(h) = \text{pépité} (5 \text{ E}4) + \text{exponentiel}$ (pallier = $3.7 \text{ E}5$, portée pratique = 0.15 km). Le polygone du banc de ponte a été défini. L'échantillonnage était constitué de stations disséminées sur le banc et en dehors du banc, de façon ni aléatoire ni régulière. Quelle serait la variance d'estimation s'il était possible d'effectuer un plan d'échantillonnage régulier du banc, pour la même structure spatiale ?

Pour répondre à cette question, il nous faut le fichier Polygone dont on dispose déjà, le modèle de variogramme dont on dispose déjà et le fichier des coordonnées des nouvelles stations qu'il faut générer. Il n'est pas possible de prévoir combien exactement de stations tomberont dans le polygone en définissant une grille donnée. On choisit une maille carrée de 0.06 km de côté dont l'origine est le point de coordonnées ($x = 25.9$, $y = 17$) et qui est composée de 11 cases sur X et de 11 cases sur Y.

GÉNÉRATION D'UNE GRILLE D'ÉCHANTILLONNAGE DANS UNE FEUILLE EXCEL 4

On organise donc une nouvelle feuille Excel. On génère facilement cette grille régulière avec Excel 4, en utilisant les propriétés récursives des formules, la fonction Édition-Copier vers le bas et la fonction Édition-Copier/Coller. Dans la case A1, on écrit 25.9. On écrit dans la case A2 la



P. Petitgas, A. Prampart



© Orstom, 1995

V-3

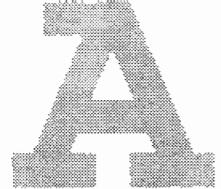
formule : = A1 + 0.06. On sélectionne ensuite le champ A2-A11, et on fait Édition-Copier vers le bas. Comme on doit importer des données dans EVA et non des formules, il faut sélectionner le champ A2-A11 et transformer les formules en valeurs. Pour ce faire, le champ A2-A11 restant sélectionné, on fait Édition-Copier, puis Édition-Collage spécial, et on sélectionne Valeurs dans la fenêtre, puis on fait OK. On écrit 17 dans la case B1 et on la recopie vers le bas jusqu'à la case B11. Voilà générée la première ligne de points de la grille d'échantillonnage. On répète cette opération onze fois, en incrémentant à chaque fois de 0.06 les valeurs de Y. La colonne A représente les X et la B les Y. Afin de pouvoir visualiser sur la carte le plan d'échantillonnage dans EVA, il faut fournir des valeurs fictives. On recopie dans la colonne C une colonne de 1. On modifie par exemple la première valeur que l'on égale à 2 car si toutes les valeurs sont égales, la carte ne pourra pas être fournie dans EVA. On peut maintenant lancer EVA et importer ce fichier à trois colonnes.

Le fichier construit précédemment est fourni sur la disquette dans le répertoire DATA sous le nom de REGEGG.TXT. Dans ce fichier, les colonnes D et E contiennent les valeurs des X et des Y du polygone précédent définissant le banc de ponté.

IMPORTATION DE LA GRILLE D'ÉCHANTILLONNAGE ET DU POLYGONE DANS EVA

Dans EVA, on ouvre la fenêtre Données en cliquant sur le bouton d'exécution Données. On choisit le type de fichier Données échan-

Test d'un plan d'échantillonnage
connaissant la structure



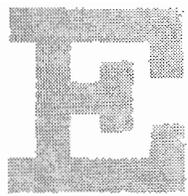
tilons. On importe les colonnes A, B et C de la feuille Excel REGEGG dans les colonnes X, Y et Valeurs de la fenêtre Données d'EVA, selon la procédure déjà décrite. On choisira de préférence le bouton Coller avec liaison dans la fenêtre Données d'EVA, au cas où l'on voudrait par la suite modifier la maille de la grille d'échantillonnage : une fois les modifications faites dans Excel, il suffira alors de mettre à jour dans EVA. On sélectionne ensuite le type de fichier Polygone d'étude. On importe les colonnes D et E de la feuille Excel REGEGG dans les colonnes X et Y du fichier Polygone d'étude. On clique dans le bouton Fermer. Les données et le polygone sont importés.

On choisit dans la fenêtre Plan n'importe quel plan du moment qu'il est croisé avec la méthode intrinsèque. L'étude commence.

REPRÉSENTATION DE LA GRILLE D'ÉCHANTILLONNAGE

Le bouton d'exécution Variable nous permet de visualiser le plan d'échantillonnage. La sélection Points intérieurs polygone n'aura d'influence que sur la représentation cartographique et non sur le calcul de la variance d'estimation. Si on choisit cette option, les points extérieurs seront grisés et les points intérieurs seront dénombrés.

Il peut être pratique de choisir la sélection « indicatrice » et de prendre un seuil supérieur à la valeur maximale : seuls de petits points représenteront les stations.



Choisir la sélection $s1 \leq z \leq s2$ et cliquer sur le bouton Calculer : on lit qu'il y a 34 stations intérieures au polygone, soit moins que dans l'échantillonnage précédent. Si on voulait plus de points, il faudrait aller modifier la maille de la grille et mettre à jour les données en utilisant la fonction Liaison de la barre de menu. On se contente de ce plan-ci.

Choisir maintenant la sélection « indicatrice » et prendre comme seuil 3 (le maximum des valeurs est 2), laisser la sélection Points intérieurs polygone active. Cliquer sur le bouton Calculer, puis sur le bouton Carte. Cliquer sur le bouton Axes de la fenêtre Carte. Activer l'option Afficher polygone dans la fenêtre Axes de la carte puis faire OK. On obtient la figure 23.

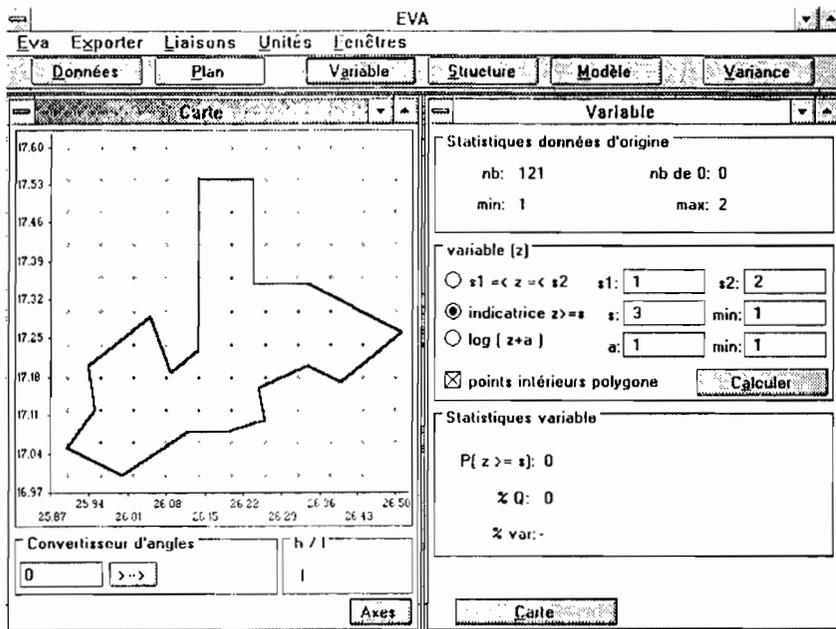


FIGURE 23 - Test d'un nouveau plan d'échantillonnage dans le polygone d'étude. La figure montre les deux fenêtres du bouton d'exécution Variable. La variance d'estimation pour ce nouveau plan sera calculée.



VARIANCE D'ESTIMATION

On aura besoin de calculer des distances. Les coordonnées des points échantillons sont exprimées en kilomètres. Aller dans la fonction Unités de la barre de menus et choisir coordonnées et distances « Identiques ».

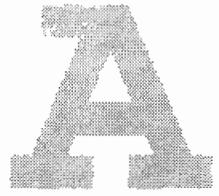
Cliquer dans le bouton d'exécution Variance, choisir l'estimation linéaire dans la fenêtre Type d'estimation. La fenêtre Variance structure spatiale s'ouvre. Entrer à la main les paramètres du modèle isotrope obtenu dans l'étude précédente, à savoir : péplte = 5 E4 + exponentiel (palier = 3.7 E5 ; portée pratique = 0.15). Choisir la même grille de calcul 20 x 20 que dans l'étude précédente. Cliquer directement dans le bouton Calculer variance. En considérant que cet échantillonnage nous donnerait la même moyenne que précédemment, à savoir 963, on trouve une erreur relative d'estimation de 8.2 %.

Le plan régulier donne la même précision que précédemment, avec 10 stations de moins.

Table des matières

I Rappels méthodologiques : ce que fait EVA

ESTIMATEURS DE LA MOYENNE SUR ZONE	I-3
Estimateurs linéaires	I-3
Estimateurs non linéaires	I-4
ESTIMATEURS DE LA VARIANCE D'ESTIMATION	I-5
Plan aléatoire : statistique classique et géostatistique	I-6
Autres plans : géostatistique transitive et intrinsèque	I-7
DIFFÉRENTES VARIANCES : VARIANCE DE DISPERSION, VARIANCE DE L'ESTIMATEUR, VARIANCE D'ESTIMATION	I-9
Méthode intrinsèque	I-9
Méthode transitive : cas des grilles régulières	I-11
FONCTIONS STRUCTURALES EXPÉRIMENTALES ..	I-12
Variogramme expérimental	I-12
Covariogramme transitif expérimental	I-14



MODÈLES STRUCTURAUX I-15
 Modèles à palier, modèles sans palier I-15
 Comportement aux petites distances I-16
 Pépite I-16
 Modèles I-17
 Anisotropies I-19

 INTERPRÉTATION PROBABILISTE
 D'UNE RÉGIONALISATION :
 MOYENNE ET VARIANCE SUR ZONE,
 MOYENNE ET VARIANCE DE LOI I-20

 ALGORITHMES DE CALCUL DES VARIANCES I-22
 Méthode transitive I-22
 Méthode intrinsèque I-23
 VARIANCE D'ESTIMATION LORSQUE LA MOYENNE SUR ZONE EST ESTIMÉE
 PAR LA MOYENNE ARITHMÉTIQUE D'ÉCHANTILLONS CORRÉLÉS I-23
 MOYENNE SUR ZONE ET VARIANCE D'ESTIMATION SOUS L'HYPOTHÈSE
 DE LOGNORMALITÉ POUR DES DONNÉES CORRÉLÉES I-25

 BIBLIOGRAPHIE I-27

II Installation, mise en route, principes d'utilisation

INSTALLATION ET MISE EN ROUTE II-2
 Créer une icône II-2

PRINCIPES D'UTILISATION	II-3
IMPORTATION, MISE À JOUR, EXPORTATION ..	II-6
Importations	II-7
LECTURE DES DONNÉES, COLONNE À COLONNE	II-7
LIAISON ET MISE À JOUR	II-9
RÈGLES DE COPIE DES CARACTÈRES DEPUIS EXCEL	II-10
Exportations : impressions et sauvegardes	II-11
FONCTION EXPORTER DE LA BARRE DE MENU	II-11
TOUCHE CLAVIER DE COPIE D'ÉCRAN	II-12
BARRE DE MENU	II-12
Unités	II-13
Fenêtres	II-13
FICHIERS DE DONNÉES FOURNIS POUR EXEMPLES D'UTILISATION	II-14

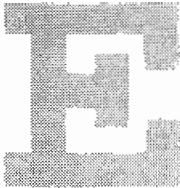
Estimation par cumul le long
de transects acoustiques :
exemple d'utilisation

ORGANISATION DE LA FEUILLE EXCEL	III-3
IMPORTATION DES DONNÉES ET CHOIX DE LA MÉTHODE GÉOSTATISTIQUE	III-6
Bouton Données	III-7
Bouton Plan	III-7

ANALYSE STRUCTURALE	III-8
Bouton Variable	III-8
Bouton Structure	III-9
Bouton Modèle	III-11
VARIANCE D'ESTIMATION	III-12
EXPORTATIONS	III-14
UTILISATION DE LA LIAISON	III-14

IV Estimation de la moyenne sur zone à partir d'échantillons non aléatoires : Exemple d'utilisation

ORGANISATION DE LA FEUILLE EXCEL	IV-3
IMPORTATION DES DONNÉES ET CHOIX DE LA MÉTHODE GÉOSTATISTIQUE	IV-4
ANALYSE STRUCTURALE	IV-7
Bouton Variable	IV-7
OPTION POINTS INTÉRIEURS POLYGONE DE LA FENÊTRE VARIABLE	IV-9
OPTION $s1 \leq z \leq s2$ DE LA FENÊTRE VARIABLE	IV-9
OPTION INDICATRICE $z \geq s$ DE LA FENÊTRE VARIABLE	IV-10
OPTION $\text{Log}(z+a)$ DE LA FENÊTRE VARIABLE	IV-11
FENÊTRE CARTE	IV-11
EXEMPLE	IV-12



Bouton Structure IV-19
 Bouton Modèle IV-19
 CALCUL DE LA VARIANCE D'ESTIMATION IV-22
 Fenêtre Type d'estimation IV-22
 Fenêtre Variance structure spatiale IV-23
 CHAMP DISCRÉTISATION IV-25
 BOUTON CALCUL VARIANCE IV-26
 EXPORTATION DES RÉSULTATS IV-27
 AUTRES MODÈLES, MODÉLISATION ANISOTROPE . IV-27
 AUTRE POLYGONE, AUTRE HYPOTHÈSE GÉOMÉTRIQUE . IV-31

V Test d'un plan d'échantillonnage
 connaissant la structure :
 Exemple d'utilisation

GÉNÉRATION D'UNE GRILLE D'ÉCHANTILLONNAGE
 DANS UNE FEUILLE EXCEL 4 v-2
 IMPORTATION DE LA GRILLE D'ÉCHANTILLONNAGE
 ET DU POLYGONE DANS EVA v-3
 REPRÉSENTATION DE LA GRILLE
 D'ÉCHANTILLONNAGE v-4
 VARIANCE D'ESTIMATION v-6



Reference : ICES CM1993/D:65

EVA

Estimation VARIance

A geostatistical software

*for structure characterization
and variance computation*

Pierre PETITGA
Marine geostatistica

Arnaud PRAMPAR
Computer scienti

Éditions de l'Orstom

INSTITUT FRANÇAIS DE RECHERCHE SCIENTIFIQUE
POUR LE DÉVELOPPEMENT EN COOPÉRATION

Collection logORSTOM

Paris, 1995



E

P. Petitgas, A. Prampart

A
© Orstom, 1995

Estimation
theory:

EVA Computations

I



Here we gather the general principles of two great categories of methods which allow to estimate the total quantities based on sampling surveys. The first one depends on the random sampling theory and the use of random variables. The second one is the geostatistical application of the theory of random functions. The computations performed by EVA belong to this context. The estimation by other approaches, particularly by process simulation will not be tackled here.

Any estimation is performed explicitly or implicitly in a given area defined before or after sampling. The total quantity estimated and the accuracy of this estimation depend on the defined area.

EVA deals with estimating the total quantity in a given area when the sampling effort is distributed homogeneously in this area. This means that samples can be aggregated but that the effort shall not systematically be focused on a given sub-area. Any over-sampling in some sub-areas of the area to be estimated will lead to a bias in the estimation likely to be provided by EVA for these sub-areas would be over-represented in the data set. In such a case, EVA could be used in both of the following ways. Either one wants to perform the estimation on the basis of the general data set and then it is necessary to reduce this over-representation by averaging the data on small blocks. Or one wants to separate sub-areas and to perform a different estimation in each sub-area. Then, one will be able to specify the variations in the structural parameters from one sub-area to the other thus defining a particular variogram model (proportional effects...).

ESTIMATORS OF THE MEAN IN AN AREA

One tries to estimate the integral of the fish density (space integral) over a given area, V , on the basis of randomly or non randomly located punctual

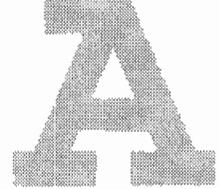
samples. The mean to be estimated in the V area is: $M_V = \frac{1}{V} \int_V f(x) dx$

where $f(x)$ is the density at point x . A description of the different estimators can be found in SIMMONDS *et al.* (1992). They can be grouped in two categories, linear and non linear.

Linear estimators

The arithmetical mean of the data (simple average) is an unbiased estimator of the mean in an area when the effort is distributed homogeneously over the whole area to be estimated. This estimator does not assume any hypothesis on the statistical distribution of the data (= histogram = probability distribution function = pdf). Using the classification put forward by SMITH (1990), the arithmetical mean is a "design based estimator" which can always be computed and is always valid without any statistical hypothesis provided that the effort should be homogeneous. As opposed to this, "model based estimators" require the inference of a model. Their use implies that the underlying phenomenon generating the data is consistent with the inferred model.

Estimating the mean in an area by kriging is equivalent to computing a weighted data mean. In the kriging procedure the weight assigned to each sample depends on its position and on the model of spatial structure. The kriging of the mean in an area can be qualified as a "spatial model based estimator". It is interesting mainly when samples are aggregated and not very numerous (say lower than one hundred) and when values are correlated. Intuitively, it is conceived that close samples which form a patch will bring redundant information. Kriging will give to each of them a smaller weight in the general mean than to isolated individual data. When the number of samples is large (say



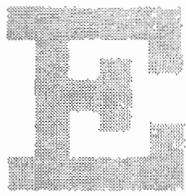
higher than one hundred), only a few samples get kriging weights different from $1/n$ and weighting no longer plays a great role in the estimation (PETITGAS and RIVOIRARD, 1991).

The arithmetical mean and the kriged mean are linear estimators: they estimate the mean in an area using a weighted average of the data where the weights do not depend on the sample values.

Non linear estimators

A non linear estimator is a weighted data mean where the weights depend on the sample values (quantiles to which samples belong). A non linear estimator depends on the model of histogram selected. It is a "pdf model based estimator". It can be interesting when the sample histogram is dissymmetrical for it is then more accurate than the arithmetical mean, but this accuracy depends on the model used. Intuitively, it is conceived that the arithmetical mean is highly influenced by high values as these contribute to the mean estimate with the same weight ($1/n$) as any other value. In a non linear estimation procedure high values are given weights lower than $1/n$ for they get a low frequency of occurrence on the probability distribution (the histogram model). However the non linear estimation can be highly biased in so far as generally it is not easy to fit adequately a model of histogram both in the low and high values.

PENNINGTON (1983) suggested an estimator of the lognormal distribution concerning trawl and ichthyoplankton data which get a number of zeros. SMITH (1988) studied the bias and the non robustness of this estimator in relation to the variations of the model. MACLENNAN and MACKENZIE (1988) suggested to use the Box-Cox power transformation functions in the case of acoustics data in order to choose a model of



P. Petitgas, A. Prampart



© Orstom, 1995

histogram better adapted to the data. Theoretically, these estimators are valid only for data which are not spatially correlated.

In the category of linear estimators, EVA supplies only the arithmetical mean. In the category of non linear estimators, EVA supplies a geostatistical lognormal estimator (MATHERON, 1974). It is the estimator of the mean in the V area concerning lognormal and correlated data. It could be qualified as a "structural and pdf model based estimator". Although it is interesting, its use remains dangerous because of the deviations from the lognormal distribution.

ESTIMATORS OF THE ESTIMATION VARIANCE

Now one is concerned with calculating the accuracy of the previous estimations. It depends on the sampling scheme performed and on the autocorrelation between samples. The variances computed account neither for measurement errors nor for biases in the course of the sampling. These "variabilities" remain in the data and are counted in the structure. The problem dealt with here concerns the error made in the estimation of a spatial integral (the total quantity) based on fragmentary information (samples).

Random sampling scheme: classical statistics and geostatistics

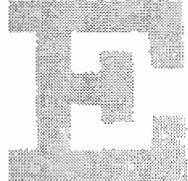
In the theory of random sampling, a variable is estimated at points whose positions are random in a V area. This area plays the role of a

ballot box. Each sample is interpreted as the result of a random selection in the ballot box and the whole spatial distribution is summarized in the probability distribution of the values in V (regional histogram). Whatever the spatial distribution may be, the samples are n independent outcomes. Each sample is an estimator of the mean of the probability distribution of values. The variance associated with such a scheme is a "design based estimator". However, geostatistics and classical statistics give necessarily the same numerical values of variance. The appropriate geostatistical formula (MATHERON, 1971) only use the dispersion variance in the whole area. When the model is well adjusted, the latter must be equal to the population variance, that is to say the variance computed in terms of $1/(n - 1)$.

The estimation variance of a random set of n samples in V is supplied in EVA only indirectly. In EVA, the positions of samples are considered fixed thus non random. EVA supplies variances concerning regular mesh sampling schemes, that is to say systematic schemes and non conventional schemes according to the theory of random sampling. When the data are correlated, the estimation variance corresponding to a random sampling scheme of n samples is obtained by dividing by n the G_{VV} supplied by EVA (see further).

Other sampling schemes: transitive and intrinsic geostatistics

When samples are not taken independently from one another and when the sampled spatial distribution is structured, the computation of any variance and particularly of the estimation variance requires to take account of the spatial structure through a model. Whatever the sampling scheme, it is possible to compute variances as soon as a



structural model is available. Theoretical corroborations can be found in MATHERON (1971), COCHRAN (1977) or JOURNEL et HUIJBREGTS (1978). Therefore, the variance estimators put forward here are "spatial model based estimators".

In order to perform these computations, two geostatistical methods are available: the intrinsic method and the transitive method. The transitive method is the theory of the global estimation (abundance estimation) using regular meshes whose origin is random, that is to say systematic schemes as defined by COCHRAN (1977). The transitive method does not make use of any statistical hypothesis on the sampled spatial distribution. The transitive covariogram is its structural tool. In the transitive method, only the strictly positive values are considered. In this method, the total quantity is estimated and not the mean. The estimation variance computed in the transitive method is equivalent to the variance of the estimator (for all the possible origins of the grid). The transitive method can be interpreted as an extension of the random sampling theory to systematic sampling schemes.

The intrinsic method is an application of the theory of random functions to the local estimation (mapping) and to the global estimation (abundance estimation). In this approach, reality is interpreted as one of the possible outcomes of a random function which would be observed within a given polygon, that is to say the area to be estimated. The intrinsic models get a double characteristic: their spatial distribution show a certain rate of homogeneity (stationarity of order k) and they are not influenced by the edges of the area (no edge effect). The variogram is the structural tool of the intrinsic model of order zero. In intrinsic models, the mean in the area to be estimated (integral over space) is a random variable: it varies with each outcome. The estimation variance is equivalent to an interpolation variance: it is the



P. Petitgas, A. Prampart



variance of the error made at each outcome between the mean in the area and its estimation. It is necessary to define the limits of the area to be estimated in order to compute the estimation variance.

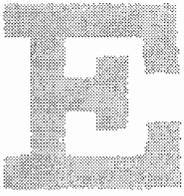
When there is no edge effect, the intrinsic and transitive methods are equivalent and supply the same numerical values of variance (MATHERON, 1965).

EVA performs the computations of estimation variances associated with the previous estimators of the mean using the intrinsic method of order zero, for any sampling scheme with homogeneous effort and non random points. In the case of acoustic surveys performed along parallel equidistant transects, EVA allows to use the transitive method on the values summed up along the transects. In this procedure one works only in one dimension very easily. This method leads to the same results as if the estimation was performed in two dimensions with a variogram (MATHERON, 1965; PETITGAS, 1993).

DIFFERENT VARIANCES: DISPERSION VARIANCE, VARIANCE OF THE ESTIMATOR, ESTIMATION VARIANCE

Intrinsic method

The variance in an area is dependent on the autocorrelation between all the values in this area. The population variance contained in V is called dispersion variance in the V area. The error variance between



the mean in V and its estimator is called estimation variance. Variances are computed using the model of variogram inferred.

The dispersion variance in V will be designated by G_{VV} . It is the mean of the variogram for all the distances in V. It is equivalent to the data variance computed in $1/(n - 1)$ concerning a random scheme in V. When the sample values are correlated and when the sample points are not randomly located, this variance is not a good estimator of the population variance in V.

MATHERON (1971) suggested to use the G_{VV} in order to scale the variogram model to the variance in V. The sill of the variogram (if there is one) is not necessarily equal to the variance in V: if the range is small as related to the dimensions of V, the sill will be very close to the variance in V. In the opposite case, the sill will be logically higher than the variance in V.

The mean of the variogram concerning all the distances between samples will be designated by G_{ab} . The variance of the estimator is : $G_{VV} - G_{ab}$. It gets the physical meaning of a measure of the variability between estimators obtained for the same network of samples but when it is set up in different places on the V area.

Now consider the mean in a V area is estimated by the simple average of non randomly located samples the values of which are correlated. The corresponding estimation variance is: $\sigma_E^2 = 2G_{aV} - G_{VV} - G_{ab}$. The estimation variance depends on three terms which are variogram averages. The G_{VV} term (variance in V) takes account of the shape of the V area. The G_{ab} term: takes account of the shape of the strict sampling network (relative position of samples as related to one another). The G_{aV} term is the mean of the variogram concerning all the distances between each sample and all the V points. The G_{aV} term

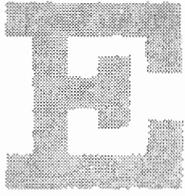
takes account of the position of the sample network in the V area to be estimated.

In a non linear estimation, the estimation variance also shows three terms which get the same physical meanings.

EVA allows to compute the three types of variance: dispersion variance in V (G_{VV}), variance of the estimator ($G_{VV} - G_{ab}$), variance of the estimation error (estimation variance, σ_E^2).

EVA allows to test the performances of different sampling schemes. In the variance formulas, the area on which one wants to make the estimation, the positions of samples and the model of variogram are involved only. The sample values are not involved in the variance computation since the variance is a function of the autocorrelation between points. In order to test a sampling scheme, it will be enough to generate the position of samples for this new scheme and to compute the estimation variance.

In general the limits of the fish population are not known exactly before the survey. These must be estimated from the values of the samples. The undetermination on the limits of the area contributes an error term on the mean estimate. In the previous formula of σ_E^2 the limits are considered exactly known: the formula does not account for a geometric error variance term. In order to evaluate the geometric error variance term, one may compute the estimation variance for different hypotheses put forward concerning the area limits. When the number of samples is sufficient, the geometric error variance is small. For regular sampling schemes, the influence of the geometric error on the estimation of the mean can be computed using an approximate formula given by MATHERON (1965, 1971). It will also be found in JOURNEL et HUIJBREGTS (1978), ARMSTRONG *et al.* (1992) or PETITGAS (1991). EVA does not allow to compute this error term.



Transitive method: case of regular grids

The computation of the estimation variance through the transitive method is proposed in EVA for the one dimensional case obtained by summing up the values along transects continuously sampled as in acoustic surveys.

In this one-dimensional method, the quantity to be estimated is:

$$Q = \int_{-\infty}^{+\infty} q(x) dx$$

where $q(x)$ represents the cumulated value on the transect of x coordinate. Outside the limits of the area, the transects are equal to zero and the Q estimator is a finite discrete sum:

$$Q^* = a \sum_{k=1}^n q(x_0 + ka)$$

where a is the inter-transect distance, where x_0 is the random coordinate of the first transect and n is the number of transects.

The estimation variance writes as follows:

$$\sigma_Q^2 = a \sum_{k=-\infty}^{+\infty} g(ka) - \int_{-\infty}^{+\infty} g(h) dh$$

where a is the inter-transect distance and g is the structural model (transitive covariogram). The estimation variance is the difference between the integral of the structural function and its approximation with the sampling grid. The finer the mesh and the more regular the spatial distribution, the lower the estimation variance. The estimation variance only depends on the mesh of the grid and on the model of transitive covariogram.

The implicit hypothesis of the method is that the whole area is covered by the grid: beyond the grid, values are equal to zero. Therefore, the



P. Petitgas, A. Prampart



limits of the area lie between two points of the grid. The impact of this geometric indetermination on the estimate is accounted for in the above formula.

EVA allows to build easily the curve of the estimation variance as a function of the inter-transect distance. It is enough to have this parameter vary and to compute again the variance.

EXPERIMENTAL STRUCTURAL FUNCTIONS

All the elements concerning computations and the models of covariance, variogram and transitive covariogram will be found in MATHERON (1971), JOURNEL et HUIJBREGTS (1978) and ARMSTRONG *et al.* (1992).

Experimental variogram

The variogram measures the mean quadratic deviation between two points separated by a vectorial distance h . It is an average measure in the whole area and therefore implies that the data show some homogeneity. The variogram is equivalent to the variance between points distant from h . If the variogram increases with distance, there is a structure (spatial correlation): two close points are more similar than two remote points.

The variogram does not take account of the location of the samples. It works only on the distances between them. Therefore it does not specify the position of dense areas. It characterizes their average structure as if these were randomly distributed in the area.

The experimental computation is made as follows. All the couples between the samples of the area are considered (or of the sub-area if it was necessary to stratify). For each of them, the direction and the norm of the vector defined by the two samples are computed. The square of the difference between their values, c^2 is computed. Then angles are gathered into classes of angle. In each class of angle, distances are gathered into classes of distance. In each class of distance, the half-mean of c^2 is computed: this half-mean is the variogram. Therefore, the experimental computation requires to choose parameters. Classes are selected in order to compute the mean of c^2 with a sufficient number of couples. The experimental variogram writes:

$$\gamma^*(u, h) = \frac{1}{2n(u, h)} \sum_i (f(x_i) - f(x_i + h))^2$$

$f(x_i)$ being the value at point x_i and $n(u, h)$ being the number of couples of samples for the direction u and the distance h .

Classes of angle are defined by defining portions of the plane. These are defined by using an angle and a tolerance $\pm \text{tolu}$ round this angle. In EVA, angles are counted trigonometrically starting from the horizontal line. Let u be the selected angle and tolu the tolerance. On the basis of each sample, x , any other sample, y , will be used to compute the variogram in the u direction if the point, y , is found in the portion of plane focused on x and defined by $]u - \text{tolu}, u + \text{tolu}[$ and $]u + \pi - \text{tolu}, u + \pi + \text{tolu}[$. EVA provides an isotropic computation (no direction is distinguished), an anisotropic computation in two directions $]u$ and $u + \pi/2[$ and an anisotropic computation in four directions $(u, u + \pi/4, u + \pi/2, u + 3\pi/4)$.

Classes of distance are defined by selecting a lag, p , and a tolerance $\pm \text{tolp}$ round it. The variogram is computed for n successive lags. Class i



includes all the distances belonging to the interval $] | p - \text{tolp}, | p + \text{tolp}]$. The value of the variogram is centered on the distance $| p$. The class $] 0, \text{tolp}]$ is individualized in the computation performed by EVA. This allows to use replicated samples in order to estimate the variogram near $h = 0$. In this class, the value of the variogram is centered on the mean of distances belonging to the class.

Experimental transitive covariogram

The transitive covariogram is a type of non centered covariance in which the geometry of the area is involved. It can be easily computed only for the regular grids that have no missing samples. In one dimension on transect cumulated values, it is computed by superposing the grid on itself and shifting it by a distance h which is a multiple of the mesh and then by summing up the products $q(x_i)q(x_i + h)$. The experimental transitive covariogram writes: $g^*(ka) = a \sum_i q(x_i)q(x_i + ka)$ where a is the mesh (inter-transect distance) and where the $q(x)$ represent the transect cumulated values.

The transitive covariogram depends on the values $q(x)$ but also on the number of products summed for each distance ka . It is reduced to zero for distances higher than the length of the grid.

STRUCTURAL MODELS

Variances are computed using structural models. These must meet some conditions for reasons of coherence (MATHERON, 1971). The models gene-

rally used (JOURNEL et HUIJBREGTS, 1978) allow to account for physical properties of spatial distributions. These properties are characterized by the behaviour of the variogram at small distances (that is to say at the origin for $h = 0$) and at great distances (that is to say existence or not of a sill).

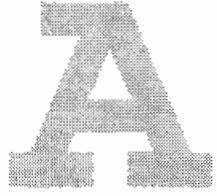
There is no basic difference in the properties of the models of variogram and transitive covariogram. However, covariograms always show a sill since they are equal to zero for distances higher than the limits of the sampling grid.

Models with a sill, models without a sill

The models with a sill show a saturation (sill) of the variance when the distance increases. The existence of a sill means that there is a variance in the model: it is the sill. When a model with sill is selected, data are interpreted as being derived from a stationary underlying model (a stationary random function) whose variance is finite. The sill is equal to the variance of the random function. It is not always equal to the variance shown by this function in a given area.

G_{VV} is the total variance reached by the random function in a V area. The sill is the variance of the function. When there is no spatial correlation (flat variogram), G_{VV} is equal to the sill.

The sill is associated with another parameter, the range. It is the distance beyond which the sill is reached. Two points separated by a distance higher than the range are independent, while two points separated by a distance lower than the range are correlated. The range gives a quantitative evaluation of the notion of influence area. In practice, it is roughly the average diameter of the aggregates in the spatial distribution.



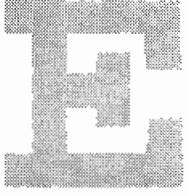
The models without sill concern random functions which do not get any variance (infinite variance): the more the distance increases, the more the variance between points increases without any limit. However, in a given area, the variance can always be computed, it is the mean of the variogram concerning distances involved in this area.

Behaviour at small distances

Let us consider that the sampled spatial distribution is a density surface. Close to the origin $h = 0$, the behaviour of the variogram model accounts for the roughness of the density surface. The model behaviours with vertical tangent account for a very high roughness; on the contrary, those with horizontal tangent account for a very high regularity. Between both, the linear behaviours account for a "mean" roughness.

Nugget

The nugget is a discontinuity of an amplitude C_0 at the beginning of the variogram ($\gamma(0) = 0$; $\gamma(h) = C_0 > 0$ as soon as $h > 0$) as in the covariogram ($g(0) = C_1$; $g(h) = C_1 - C_0$ as soon as $h > 0$). It is a totally random component of the spatial distribution whose variance is C_0 . The nugget gets three physical interpretations which cannot be dissociated in a practical study. It can be generated by a pure random component of regionalization and/or it can be generated by an additive measurement error which is independent from the values and is purely random and/or it can be generated by spatial structures whose ranges are lower than the sampling grid.



More information concerning models can be found in MATHERON (1971), JOURNEL et HUIJBREGTS (1978), ARMSTRONG *et al.* (1992). The choice of a model is a characterization of the regionalization of the values: the model contains what one thinks of the underlying phenomenon. A type of model is selected for its properties at great and small distances. Therefore, in EVA, there is no automatic modelling: the user must choose his type of model and the parameters of this model.

The estimation variance depends mainly on the behaviour of the variogram model near $h = 0$ and on the behaviour at great distances (MATHERON, 1971, 1978). Models which are somewhat different but display similar behaviours will give similar estimation variances (MATHERON, 1978).

Models

A model can be the sum of different nested models, if the data allow to reveal structures at different scales. EVA allows to fit a model made at the most of a nugget effect and three nested models.

EVA enables to use the models of the table 2. For all these models, the parameter C is called sill and the parameter r is called range.

The exponential and Gaussian models reach their sill only asymptotically. JOURNEL et HUIJBREGTS (1978) defined for these models a "practical range", r' : it is the distance at which 95 % of the sill C is reached.

Therefore, we get for the exponential model, $r = r' / 3$ and for the Gaussian one, $r = r' / \sqrt{3}$. In EVA, the user is asked to enter the values of r' .

The triangular model of transitive covariogram is the one-dimensional covariogram model for a segment.

TABLE 1 -
Characteristics
of variogram
models.

Type of model	Sill	Behaviour near h = 0	Modelized roughness
Spherical	Yes	Linear	Medium
Exponential	Asymptotic	Linear	Medium
Gaussian	Asymptotic	Parabolic (horizontal tangent)	Very smooth
Triangular (only in 1 D)	Yes	Linear	Medium
Power (h^a)	$0 < a < 1$ No	Tangent more and more vertical when a tends to 0	Very rough
	$a = 1$ No	Linear	Medium
	$1 < a < 2$ No	Tangent more and more vertical horizontal when a tends to 2	Smooth

TABLE 2 -
Model equations
for variogram
and transitive
covariogram.

Type of model	Variogram (2 D) : $\gamma(h) =$	Transitive covariogram (1 D) : $g(h) =$
Spherical	$C(1.5(h /r) - 0.5(h /r)^3)$ C si $ h \geq r$	$C(1 - 1.5(h /r) + 0.5(h /r)^3)$ si $ h \leq r$ 0 si $ h \geq r$
Exponential	$C(1 - \exp(- h /r))$	$C \exp(- h /r)$
Gaussian	$C(1 - \exp(-h^2/r^2))$	$C \exp(-h^2/r^2)$
Triangular	Non available	$C(1 - h /r)$ si $ h \leq r$ 0 si $ h \geq r$
Power	Ch^a	Non available

Anisotropies

Anisotropy means that all the spatial directions are not equivalent. An anisotropic variogram models the fact that variability gets different characteristics in different directions. JOURNEL et HUIJBREGTS (1978) defined two types of anisotropy, a geometric and a zonal one with reference to models with sill.

In the geometric anisotropy, the sill is the same in all directions but the range varies elliptically according to the direction. Physically, it means that all the directions show similar heterogeneity but that the shape of aggregates is elliptical rather than circular as in the isotropic case. In the zonal anisotropy, the variance of the random function (say the sill) varies with the direction: the phenomenon is more heterogeneous in some directions. Both anisotropies can be combined.

The parameters of the geometric anisotropy are an angle of rotation, a coefficient of anisotropy X on the x -axis and a coefficient Y on the perpendicular direction. The angle defines the rotation which must be made in order to bring the x -axis on one of the axes of the range ellipse (the major or the minor one). Coefficients denote the expansion or the contraction of distances on the two main axes of the ellipse which must be performed in order to transform the ellipse into a circle. For instance, if one assumes that in the direction 45° , the range is twice as large as in the perpendicular direction, the rotation amounts to 45° and coefficients are: 1 for X and 2 for Y or 0.5 for X and 1 for Y .

The zonal anisotropy is modelled using a model with nested structures. In one direction, we get one model less. For this purpose, it is sufficient to equal the anisotropy coefficient to zero in the required direction. In order to avoid discontinuities in the model, it would be better to avoid

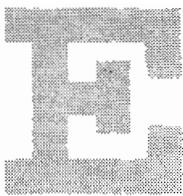


a model composed of two nested models where each would be applied only in a portion of the plane. JOURNAL et HUIJBREGTS (1978) recommend to take a model where neither the coefficient X nor Y are equal to zero and to add to it other models where either X or Y is equal to zero.

PROBABILISTIC INTERPRETATION OF A REGIONALIZATION: MEAN AND VARIANCE IN AN AREA, MEAN AND VARIANCE OF A DISTRIBUTION

Consider the sampled spatial distribution is one outcome from a random function which would be observed in a V area window. The mean in V for a specific outcome is not necessarily equal to the mean of the random function.

The simple average of the data is an unbiased estimator of both quantities, mean of the random functions distribution and mean of the function in V for the outcome sampled. But the estimation variance is not the same when this estimator estimates either quantity. Any other estimator than the simple average will estimate either one quantity or the other. For the lognormal estimation, the option selected in EVA consists in estimating the mean in the V area for the functions outcome sampled. The elements necessary to compute the distribution mean are nevertheless given. But, the variance computed in EVA does not correspond to the estimator of this quantity as the variance computed in EVA is the estimation variance. The estimation variance is the variance



P. Petitgas, A. Prampart



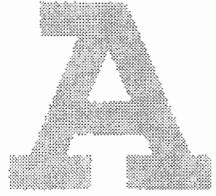
© Orstom, 1995

of the estimation error made for each outcome between the functions mean in V and its estimation.

The experimental varlogram is computed on the values of one outcome of the random function, the one sampled in V (regional varlogram). The modeling operation consists in selecting an average model which gets the meaning of the average of experimental varlograms which would be computed by other samplings in this area. The variance in the V area deduced from the varlogram does not therefore correspond in the model to a specific outcome but to the mean variability of the random function in the V area for all the outcomes. The variance in V is equal to the G_{VV} , while the variance of the functions distribution (pdf) is equal to the sill of the varlogram. Let us consider a model whose varlogram does not get any sill. The distribution (pdf) of this model does not get any variance, it is infinite. The central limit theorem is no longer applicable: no interval can be defined round the mean which contains a given percentage of values. But it is always possible to compute the variance in a given area using the variogram.

In the theory of random sampling, the V area is not a window where a function is observed: the V area is identified to the whole universe and no distinction is made between the statistical parameters of the pdf and the parameters in the area.

The key parameter is the size of the V area relatively to the range: it determines the statistical quality of the inference of the model (MATHERON, 1978) as well as the numerical deviations between the parameters in the area and those of the model. The smaller the range in comparison to the size of V , the better the statistical inference and the smaller the deviations. However, at the end of an epistemological reflection, MATHERON (1978) points out that a model of random function



is a methodological method which is used only to clarify the modeling of the variogram. The mean and the variance of the random function's pdf remain conventional, whereas the regional parameters (mean, variance and variogram $\ln V$) have a practical meaning and are of interest for us in the computations (MATHERON, 1978). Thus, concerning the modeling of the variogram, the behaviour near $h = 0$ and the value of G_{VV} are parameters which are more significant than the explicit value of the sill.

ALGORITHMS FOR THE COMPUTATION OF VARIANCES

Transitive method

In EVA, the one-dimensional formula for the estimation variance,

$$\sigma_Q^2 = a \sum_{k=-\infty}^{+\infty} g(ka) - \int_{-\infty}^{+\infty} g(h)dh \text{ is computed directly.}$$

The models of covariogram used get obvious primitives and their integrals have therefore been computed formally. Concerning the models with sill, the range represents the limits of summation and integration: the computation is direct and without any approximation. Concerning the models with asymptotic sill, infinite summation and integration are truncated and the following approximation is used. It is considered that these models showed a range equal to the distance for which 99.999 % of the sill is reached.

Intrinsic method

ESTIMATION VARIANCE WHEN THE MEAN IN AN AREA IS ESTIMATED BY THE SIMPLE AVERAGE OF CORRELATED SAMPLES

The estimation variance is written as follows (MATHERON, 1971):
 $\sigma_E^2 = 2G_{av} - G_{VV} - G_{ab}$. In EVA, this formula is computed numerically.

$$G_{ab} = \frac{1}{n^2} \sum_a \sum_b \gamma(x_a - x_b)$$

(a and b are the indexes on samples and x denotes a position)

$$G_{av} = \frac{1}{nV} \sum_a \int_V \gamma(x_a - y) dy$$

(x_a denotes each of the samples, y denotes each of the points of V)

$$G_{VV} = \frac{1}{V^2} \int_V dx \int_V \gamma(x - y) dy$$

(x and y take all the positions in V independently from one another)

In order to compute the integral summations in the V area, the latter is discretized with a fine mesh grid. In EVA, the user defines himself the discretization of V. The computation of G_{VV} is made according to the algorithm of the stratified random mesh proposed by JOURNEL et HUIJBREGTS (1978): x gets the position of each node of the discretization grid of V, while y gets a random position in each cell of the grid.

MATHERON (1971) developed approximate computations for the σ_E^2 formula in the case of regular grids and for isotropic spatial structures. The

values of the estimation variance can be read directly on charts and computation can be performed manually. Concerning these approximations, MATHERON (1971) showed that a variogram model for small distances is sufficient (quasi-stationarity): the approximate estimation variance only depends on the behaviour of the variogram near $h = 0$. In the numerical computation performed in EVA, the variance also depends on the behaviour of the model at great distances and a realistic model will have to be inferred for all the distances, not only for small ones.

The computation of the G_{VV} and G_{AV} integrals is based on the discretization of the V area. As the variance is a difference between these terms and G_{ab} , if the discretization is not fine enough in comparison to the sampling mesh, a negative variance may be found. The V area must be discretized finely enough to get a good estimation of G_{VV} , of G_{AV} and therefore of the estimation variance.

EVA provides the following variance results: the estimation variance σ_E^2 , the G_{VV} and the G_{ab} .

The estimation variance is a linear function of the variogram. Now consider the variogram model is the sum of several nested models. Each of the nested structures contributes to the estimation variance σ_E^2 its own variance term which can be express as a percentage of the total estimation variance.

The estimation variance which would be given by a random sampling scheme with the same number of samples can be computed: it is sufficient to divide the G_{VV} by the number of samples. When samples are not randomly located and data are correlated, the variance computed in $1/(n - 1)$ is not a good estimator of the dispersion variance in V. One may prefer the G_{VV} to it. On the contrary, for a random scheme, G_{VV} must be equal to the variance calculated in $1/(n - 1)$.

**MEAN OVER THE AREA AND ESTIMATION VARIANCE
FOR CORRELATED DATA UNDER LOGNORMAL HYPOTHESIS**

EVA supplies the estimators proposed by MATHERON (1974). The variogram used for this estimation is the variogram of the Logs of the data. In EVA the lognormal estimation can be performed using any variogram. So the user must be cautious in using the variogram model for the Logarithms of the data.

The mean of the lognormal random function in V is:

$M_V = \exp(m^a + G_{VV} / 2)$, where m^a is the mean in V of the Logarithm of the random function and where G_{VV} is computed with the variogram of the Logs. The estimator of M_V for the sampled outcome of the random function and the sampling performed is:

$M_V^* = \exp(\bar{m}^a + G_{ab} / 2)$ where \bar{m}^a is the arithmetical mean of the Logs of the data and where G_{ab} is computed with the variogram of the Logs.

The estimation variance is: $\sigma_E^2 = M_V^2 \exp(G_{VV}) (T_{VV} - 2T_{aV} - T_{ab})$

with $T_{VV} = \int_V dx \int_V dy \exp(-\gamma(x-y))$; $T_{aV} = \int_V \exp\left(-\frac{1}{n} \sum_a \gamma(x_a - y)\right) dy$

$T_{ab} = \exp(-G_{ab})$

The computation of the T_{VV} and T_{aV} terms is performed with the discretization of the V area as previously. Similarly, for numerical reasons, it is important to take a fine enough discretization.

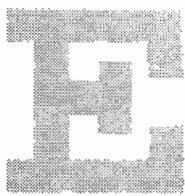
In the expression of the estimation variance σ_E^2 , the mean of the random function in V is involved though its value is unknown in practice. EVA gives the relative estimation error, σ_E / M_V , and the values of G_{VV} and of G_{ab} .



The estimation variance for a random sampling scheme with the same number n of samples is: $(M_V^2 \exp(G_W) - 1) / n$. It can be approximated by using the formula:

$$(M_V^2 \exp(G_W) - 1) / n$$

All the terms of this formula are given in EVA, but the formula is not explicitly computed.



REFERENCES

ICES, 1991 — *Report of the Workshop on the applicability of spatial statistical techniques to acoustic survey data*. Copenhagen, ICES, CM91/D: 40, 20 p.

ARMSTRONG (M.), RENARD (D.), RIVOIRARD (J.), PETITGAS (P.), 1992 — *Geostatistics for fish survey data*. Fontainebleau, ENSMP, centre de Géostatistique, 90 p.

COCHRAN (W.), 1977 — *Sampling techniques*. New York, John Wiley and Sons, 3rd édition.

JOURNEL (A.), HUIJBREGTS (C.), 1978 — *Mining Geostatistics*. London, Academic Press, 600 p.

MACLENNAN (D.), MACKENZIE (I.), 1988 — Precision of acoustic abundance estimates. *Can. J. Fish. Aquat. Sci.*, 45 : 605-616.

MATHERON (G.), 1965 — *Les variables régionalisées et leur estimation*. Paris, Masson, 306 p.

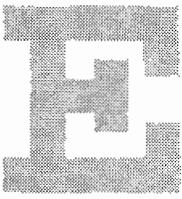
MATHERON (G.), 1971 — *The theory of regionalised variables and their applications*. Fontainebleau, ENSMP, centre de Géostatistique, coll. Les cahiers du centre de Morphologie mathématique, fasc. 5, 212 p.

MATHERON (G.), 1974 — *Effet proportionnel et lognormalité ou le retour du serpent de mer*. Fontainebleau, centre de Géostatistique, note N-374.

MATHERON (G.), 1978 — *Estimer et choisir*. Fontainebleau, centre de Géostatistique, coll. Les cahiers du centre de Morphologie mathématique, fasc. 7, 175 p.



- MATHERON (G.), 1989 — *Estimating and choosing. An essay on probability in practice*. Berlin, Springer-Verlag, 141 p.
- PENNINGTON (M.), 1983 — Efficient estimators of abundance, for fish and plankton surveys. *Biometrics*, 39 : 281-286.
- PETITGAS (P.), 1991 — *Contributions géostatistiques à la biologie des pêches maritimes*. Thèse de doctorat, centre de Géostatistique, Fontainebleau, 211 p.
- PETITGAS (P.), RIVOIRARD (J.), 1991 — « Global estimation : s^2/n and the geostatistical estimation variance, Appendix B ». In ICES 1991: appendix B.
- PETITGAS (P.), 1993 — Geostatistics for fish stock assessments: a review and an acoustic application. *Ices J. Mar. Sci.*, 50 : 285-298.
- SIMMONDS (E. J.), WILLIAMSON (N.), GERLOTTO (F.), AGLÉN (A.), 1992 — *Acoustic survey design and analysis procedures: a comprehensive review of current practice*. Copenhagen, ICES, Coop. Res. Report, n° 187.
- SMITH (J.), 1990 — Use of statistical models for the estimation of abundance from groundfish trawl survey data. *Can. J. Fish. Aquat. Sci.*, 47 : 894-903.
- SMITH (J.), 1988 — Evaluating the efficiency of the delta-distribution mean estimator. *Biometrics*, 44 : 485-493.



P. Petitgas, A. Prampart

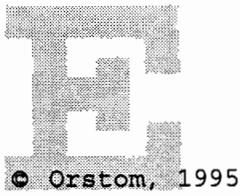


© Orstom, 1995

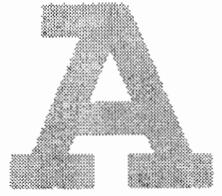
Installing,
getting started:

principles of use

II



P. Petitgas, A. Prampart



INSTALLING AND GETTING STARTED

The EVA software is written in Visual Basic 1.00. It works under Windows 3.1 and requires the use of a spreadsheet program for Windows, for instance Excel 3 or 4. The program supplied, EVA.exe, like all the applications of Visual Basic makes use of a library for Windows called VBRUN100.dll. This library is necessary for EVA to work.

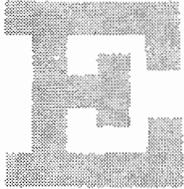
The disk supplied contains two directories: EVA and DATA. The DATA directory contains three data files which are used as processing examples and are described in detail in sections III, IV and V.

The EVA directory contains EVA.exe and VBRUN100.dll. It is advisable to keep this architecture and therefore to create on C:\ an EVA directory and to copy the contents of the A:\EVA directory. These operations are performed easily with the Windows File Manager: it is enough to click with the mouse on the EVA directory in A:\ and to put it in C:\. If VBRUN100.dll is not installed in the same directory as EVA.exe, the PATH variable must be modified consequently in the AUTOEXEC.BAT.

Creating an icon

Once the installation on C:\ is finished, it is possible to create an icon for the EVA program by using the commands of the Windows Program Manager:

- click File, then New from the menu bar of the Program Manager;
- a window appears: choose Group of Programs, then give a name to it, for instance Geostat, a group of programs is created;



P. Petitgas, A. Prampart



© Orstom, 1995

- click again File, then New from the menu bar of the Program Manager, and click Program;
- another window appears: fill up the cells as desired, for instance: name = EVA, command line = C:\EVA\EVA.EXE, working directory = C:\DATA, click the Change icon button and choose the fish;
- click OK.

Now, one can start EVA from the File Manager or from its icon in the Geostat group.

PRINCIPLES OF USE

Once it is started, the EVA software displays two bars in the upper part of the screen: a menu bar (EVA, Export, Links, Units, Windows) and below an execution bar composed of successive buttons (Data, Sampling, Variable, Structure, Model, Variance) (fig. 1).

In order to go from Data at the left to Variance at the right, one must go through successive steps symbolized by buttons. Each button opens different windows which are useful for the achievement of the step. But windows are independent from one another. Each window which has been opened is kept in memory and can be called back at any time. The interface allows us to go freely through the processing steps.

The morphology of each window can be modified with the arrows positioned in the right upper part of its frame: it can be enlarged to the full screen dimensions or reduced to an icon. Reducing the window does not mean closing it: the reduced window keeps its parameters; It

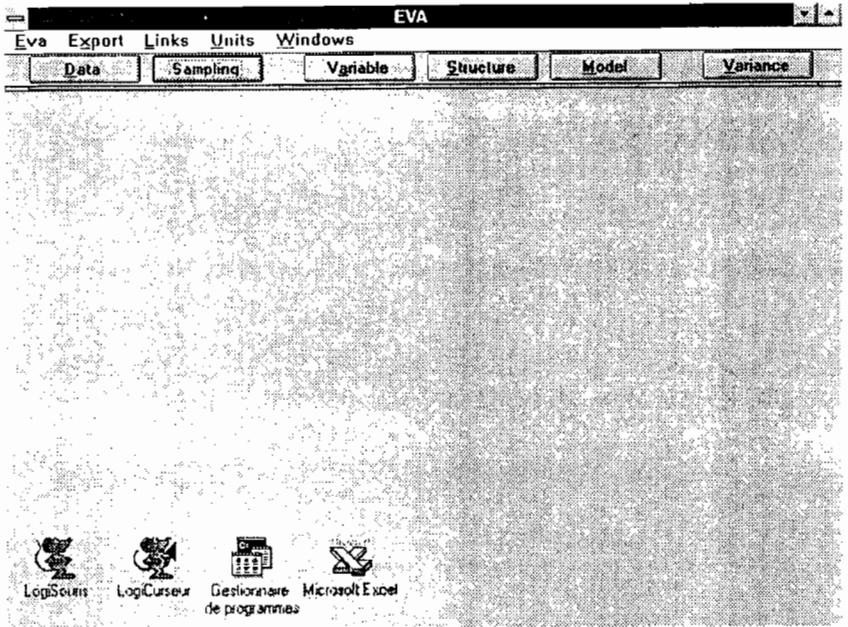
EVA

P. Petitgas, A. Prampart

© Orstom, 1995

II-4

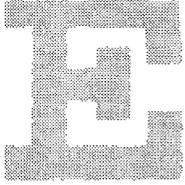
FIGURE 1 -
EVA software
as it appears
when double
clicking
in the
EVA icon.
Bar of menus
and bar of
executions
boutons.



Is kept in memory and can be called back at any time. Clicking a new execution button opens the related windows. Clicking again an execution button corresponding to an already performed step calls back the related windows. The windows opened in a study are never closed except in the following case. If in the course of the study, one returns to the Sampling window and if a new sampling scheme is selected, all the windows will be closed. This is understandable in so far as changing a scheme consists in carrying out a new study.

Going through a series of buttons amounts to following a path. According to the type of data, the type of sampling scheme and the

Installing, getting started
and principles of use



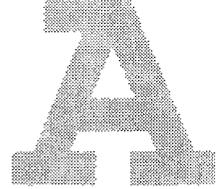
type of geostatistical method (transitive or Intrinsic), one is led to the variance computation after having followed different paths. At each moment, all the paths are not possible: some buttons are inactive or can be activated according to the steps previously achieved.

The Variable, Structure and Model buttons are specific to the structural analysis. The Data and Sampling buttons allow to choose a path in order to get to the variance computation. EVA allows to cross the geostatistical method (transitive or intrinsic) with the sampling scheme performed. However, the intrinsic method uses the same variance formula whatever the sampling scheme.

EVA allows to perform two types of studies. One consists in performing a complete study of experimental data. One will go through all the

Knowledge of the model of spatial structure	Geostatistical method for computing the estimation variance	Buttons of the execution bar to be used	Path code
Yes	One-dimensional transitive	Sampling Variance	1
Yes	Intrinsic	Data, Sampling, Variance	2
No	Intrinsic	Data, Sampling, Variable, Structure, Model, Variance	3
No	One-dimensional transitive	Data, Sampling, Variable, Structure, Model, Variance	4

TABLE 3 - Four different paths when using EVA.



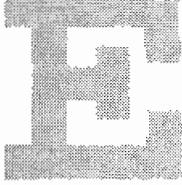
successive buttons of the execution bar. Eva also enables to compute the estimation variance directly with the positions of the samples but without the data values. This can be used for testing sampling schemes when the structure is known a priori, for instance that of previous surveys. In this case, only the buttons Data, Sampling and Varlance need be used. Four paths are possible (table 3).

Eva can also be used only for the structural analysis.

IMPORTATION, UPDATING, EXPORTATION

EVA works as an application program of Windows like Word, Excel or Paintbrush. The copy-paste function under Windows is used on the one hand to import data and on the other hand, to export graphics and numerical results. The copy-paste function places the Information temporarily into the clipboard, thus allowing this information to be transferred from one application of Windows to the other. In order to operate the copy-paste function, the source and arrival applications must be opened simultaneously. When importing data, EVA is the arrival application and the spreadsheet (Excel) is the source. When exporting results, EVA is the source and a graph editor for Windows (Paintbrush or other) is the arrival application.

The Alt + Tab keyboard shortcut (the Alt key is held down while pressing the tabulation key) allows under Windows to toggle very easily from one application to another. By holding down the Alt key and by pressing the Tab several times, all the successively opened applications are



accessed. In order to activate this keyboard shortcut, the Windows configuration must be modified if necessary: enter Main Group then Configuration Panel, then Desk and click in the cell corresponding to the quick toggle with Alt + Tab.

Importations

DATA READING BY COLUMN

The operations to be performed are as follows:

- start EVA. Click the Data button. The related window opens;
- Alt + Tab: return to the Program Manager of Windows. Start Excel or another spreadsheet under Windows;
- In Excel, read the data file you want to work on. The file is now in a worksheet. Each column corresponds to a variable: x, y, sampled value,... Select a column. Click in Edit from the menu bar then select Copy. This column is now in the clipboard;
- Alt + Tab: return to the EVA Data window;
- in the EVA Data window, click in the desired column. The selected column is marked by a double frame. Click the Paste button or the Paste and link button. A sound signal (bip) reveals that the operation is completed. The values imported from the clipboard are displayed.

Repeat this series of operations for each column concerned. Clicking in the Close button enables reading.

The columns of the Eva Data window are no editor. They are used only to display the imported data in order to check the success of the data importation. If the values need be modified, modifications

have to be made in Excel. The modified columns must be Imported again in EVA.

In the EVA Data window, three types of files can be read: Data samples, Polygon, Coast. Each one is imported from Excel column by column. It is advisable to organise an Excel worksheet which includes the three files in order to work more easily. In an Excel worksheet, the Data sample file is a field which includes so many lines as samples and three columns, x, y and values. The Coast and Polygon files include only 2 columns, x and y.

EVA allows to use different paths as specified in the table 3. The table 4 gives the columns which must be filled up in the EVA Data window.

TABLE 4 -
Input Data
for each
of the four
paths of EVA.

Path code	Sample data file			Polygon file
	x	y	Values	
1	-	-	-	-
2	Yes	Yes	-	Yes
3	Yes	Yes	Yes	Yes
4	-	-	Yes	-

The Polygon file is made of a series of points. The polygon must be closed in the file: the last line is equal to the first one. The polygon cannot contain "holes" within it.

The Coast file is also a series of points. The Coast file contains the coast line which is not necessarily closed and can be followed by islands.

Each island is a closed polygon in the file. Each coast unit, non closed coast line or island, is a series of points. These units follow each other in the file. When two successive lines are the same in the file, EVA understands that the repeated line marks the separation between coast units. In order to build this file under Excel, it is enough to append these small units one after another. The Coast file is used only when data are represented on the map: it is not used in the computations.

LINKS AND UPDATING

If Paste and link is selected in the EVA Data window, a link is established with the corresponding column of the Excel worksheet. Any change in values in this Excel column will be directly passed on to EVA using the Links-update function from the EVA menu bar.

Any modification in values can be made at any moment by returning to the data in Excel. The updating amounts to the automatic reading of new values. There is no updating of the series of computations: once the variable is updated, computations must be performed again.

An adequate configuration of the Excel worksheet allows to use easily the link between an Excel column and an EVA column. One can exchange the polygon, exchange the variable or modify the working variable. It is enough to exchange in Excel the contents of the linked column for that of another column of the worksheet or to modify in Excel the contents of the linked column.

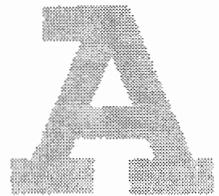
PRINCIPLES OF COPYING CHARACTERS FROM EXCEL

The Excel Copy function gets some properties which limit its use. One must be careful with the size of the Excel column, the decimal separa-

Installing, getting started
and principles of use



P. Petitgas, A. Prampart



© Orstom, 1995

tor and the alphanumerical character. The Copy-Paste function works with the text format. In EVA, after pasting, the text is reconverted in the numerical form when the Close button is pressed in the Data window.

The Excel Copy function copies only what is visible in the column. If the column is not wide enough, the figures transmitted to EVA will be truncated in their visible part. The Excel columns must be enlarged if necessary.

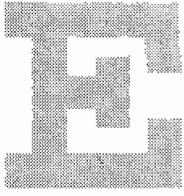
In EVA, the decimal separator is the point (.). If numbers are pasted whose decimal separator is a coma, only the integer part of the value will be considered numerically in EVA. The number format must be changed if necessary in the configuration of Windows. The path to this configuration is as follows: enter Program Manager, then Main Group, then Control Panel, then International and click modify in the Number Format cell and insert the point as a decimal separator instead of the coma. In order to take account of the change in configuration, it is not necessary to restart Windows, but Excel will have to be restarted.

Letters are interpreted as zero values by EVA. Before copying, words and letters which are found in the Excel column must be either removed or not selected.

Missing lines in the Excel columns will mark the end of the file during the importation in EVA. Only the lines before the missing one will be imported.

■ Exportations: printings and savings

Printings and savings are performed by exporting graphics or files to other applications of Windows. This is done with the Export function from the EVA menu bar. The whole screen or EVA windows can also be exported using the Screen Copy key.



P. Petitgas, A. Prampart



© Orstom, 1995

EXPORT FUNCTION FROM THE MENU BAR

Figures as well as the values of experimental varlograms and covarlograms can be exported. Figures are exported with the Bitmap format and values with the text format. Figures can be pasted in a graphics editor like Paintbrush and the values of the varlogram in an Excel worksheet. The successive operations are as follows:

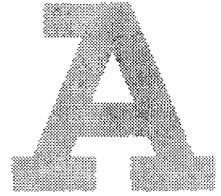
- In EVA, click Export from the menu bar and choose the graph to be exported in the scrolling list. A sound signal (blp) reveals that the graph is in the clipboard;
- Alt + Tab: toggle into the arrival application. It can be Word, Excel, Write, Paintbrush or other;
- In the arrival application, click Edit and Paste. The graph is displayed. It can be edited, saved and printed;
- Alt + Tab: return to EVA.

An enlarged figure can be copied. An EVA window is enlarged to the full screen dimension by clicking the up arrow in the top right corner of the window. In order to export the full screen figure, the window must be reduced (click the down arrow in the top right corner of the window). The menu bar is again displayed. By performing the series of previous operations, the enlarged figure is exported.

The pasting of the values of the varlogram and covarlogram in an Excel worksheet gives automatically a headline which specifies the direction and the name of each column. Distances, values and the number of couples are automatically pasted in three different columns.



P. Petitgas, A. Prampart



II-12

SCREEN COPY KEY

The screen Copy key allows to export the whole screen in the clipboard. Then it is possible to paste it in any application of Windows.

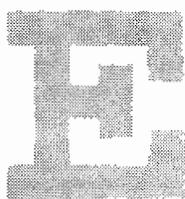
The Alt + Screen Copy key allows to export the whole active window in the clipboard as well as its contents and its edges. Then it is possible to paste it in an application of Windows. Therefore, the windows concerning the parameters of the models selected can be saved and printed easily. The upper edge of an active window is darker.

MENU BAR

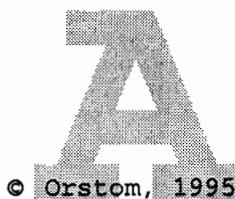
The menu bar includes the Eva, Export, Links, Units and Windows functions. Each one opens a scrolling list. The Eva menu bar function allows to access the command Quit which closes the software EVA. The Export and Links functions have been described in the previous section called Importation, updating, exportation.

Units

The computations of the variogram and of the variance require that distances be calculated. The Units function allows to define the units of the coordinates and distances which will be used in the computations. At the end of the sampling, the coordinates of the samples are generally expressed in terms of degrees, minutes and tenths of minutes. However, the distance represented by one mile of longitude varies with the latitude. A transformation of the coordinates can be neces-



P. Petitgas, A. Prampart



© Orstom, 1995

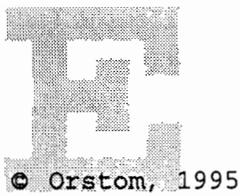
sary to assure that one mile in latitude represents the same distance as one mile in longitude over the studied area. Any transformation of coordinates must be performed under Excel.

EVA allows two situations: coordinates are expressed in terms of degrees and hundredths of degrees and distances are expressed in terms of nautical miles (it is the default option); or coordinates and distances are expressed with the same units. The active option is labelled in the scrolling list of the Units function.

Windows

The Windows function allows to catalog the windows opened. Each window is associated with one or two execution buttons; when the button is pressed, the related windows open. The same button allows to access different windows according to the path. Each open window is kept in memory and can be called back at any moment. The Windows-Show/Update function leads to the list of windows. The names of the already open windows are in bold-faced type, the others are shaded. Selecting a window calls it.

Each window is independent: it can be called back open at any moment once it has been opened; it can be shifted by clicking in its higher edge and by moving the mouse; it can be reduced using the down arrow which is found in the top right corner of its frame. Very quickly, the screen can become a mess likely to be reorganized. The option Arrange automatically in the scrolling list allows to perform this operation. The label found in front of the option means that it is active. If the EVA windows are scattered and/or reduced and if the Arrange automatically option is active, clicking on an execution button will automatically replace the related windows in their original position.



P. Petitgas, A. Prampart



© Orstom, 1995

II-14

DATA FILES SUPPLIED AS EXAMPLES OF USE

The DATA directory in the disk supplied includes three data sets: SPAWHER.TXT, HEREGG.TXT and REGEGG.TXT. They are supplied in Text-ASCII format with a tabulation as column separator. Therefore, these data can be transferred directly into a spreadsheet under Windows.

The following sections III, IV and V describe step by step the series of operations to be performed in order to carry out typical studies with EVA. These sections allow to become familiar with the methods while working with the software. Performing the following case studies is thus recommended. The spreadsheet used is Excel 4.

E

P. Petitgas, A. Prampart

V

A
© Orstom, 1995

Abundance estimation
performed by summing up
biomass along acoustic
transects:

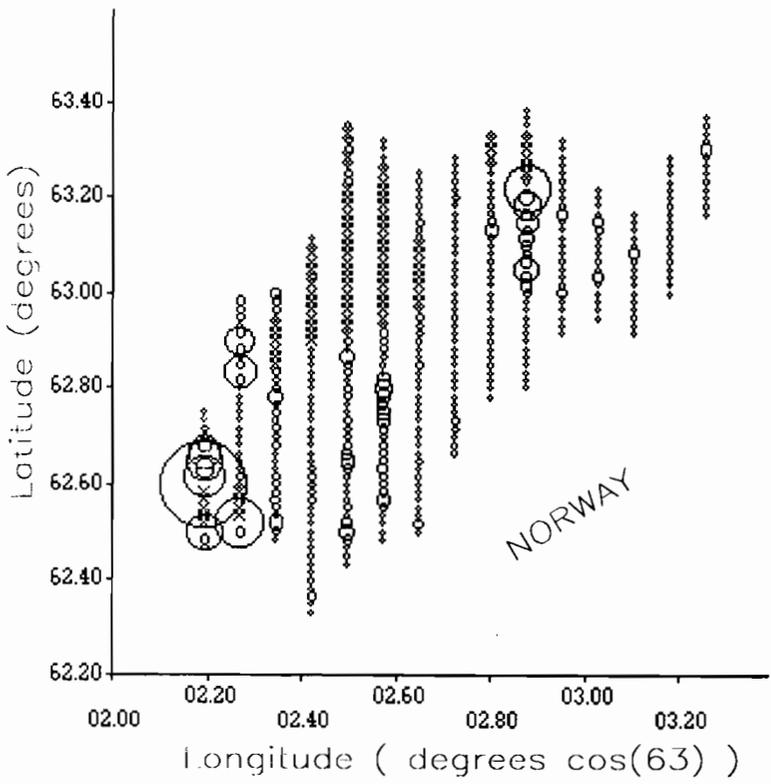
exemple of use

I I I

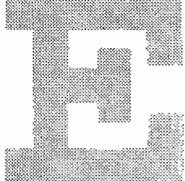
III-2

The data set which is used here as an example has been supplied by the Marine Research Institute of Bergen (Norway) as test data set number 4 at the ICES workshop on the applicability of spatial statistical techniques to acoustic survey data (ICES, 1991). The data correspond to an acoustic survey. A more detailed description and processing of these data will be found in ICES (1991) and PETITGAS (1991, 1993). The regular part of the sampling scheme is supplied on the disk in the DATA directory. The file name is SPAWHER.TXT.

FIGURE 2 -
Proportional
representation
of the date
in the file
SPAWHER.TXT.



Abundance estimation performed by
summing up biomass along acoustic transects



The acoustic survey covered a major spawning ground of Norwegian spring spawning herring situated off Moore. It was performed along parallel and equidistant transects and was designed to assess the spawning biomass amongst other objectives. The survey took place in february during the reproduction period and lasted only a few days, while herring remain in the area for several weeks. The inter-transect distance amounts to 4.54 nautical miles (N.m.). Each value of echo-Integration is an average value of punctual indexes of fish acoustic reverberation: average on the water column and average of these values on segments of 1 N.m. along the ships sailing track. According to the properties of the echo-integration method, each value is proportional to the fish density present under the ship. Each value is proportional to a fish density and is expressed as an average acoustic reverberation index over one nautical mile (N.m.) square of sea surface. The data set is displayed in figure 2. The radius of each circle is proportional to the density for this mile.

As the acoustic surveying is continuous along the survey track, PETITGAS (1993) suggested to work on the amount of fish per transect (transect cumulated values) and to use the one-dimensional transitive method on these cumulated values in order to compute the precision of the abundance estimation.

CONFIGURATION OF THE EXCEL WORKSHEET

Figure 3 is a screen copy of the Excel worksheet which includes the data to be imported in EVA. The following computations were previously performed with Excel.

Abondance estimation performed by
summing up biomass along acoustic transects

III-4

FIGURE 3 -

Data file
SPAWHER.TXT
in its Excel
worksheet.

B = longitude,
C = latitude
(in nautical
miles
and hundreds),
D = fish
acoustic
density,
F = transect
code,
G = biomass
per transect
(mean transect
density
multiplied
transect length),
H = length
of transect.

	B	C	D	E	F	G	H
1	2.27	62.5	3543		1	196149	37
2	2.27	62.5166667	36400		2	103920	42
3	2.27	62.5333333	0		3	49427	59
4	2.27	62.55	0		4	13883	55
5	2.27	62.5666667	0		5	63489	63
6	2.27	62.5833333	0		6	74406	61
7	2.27	62.6	80		7	21587	57
8	2.27	62.6166667	652		8	12627	47
9	2.27	62.6333333	100		9	16061	44
10	2.27	62.65	52		10	128681	43
11	2.27	62.6666667	66		11	16712	36
12	2.27	62.6833333	171		12	12054	26
13	2.27	62.7	65		13	8768	31
14	2.27	62.7166667	65		14	4058	31
15	2.27	62.7333333	04		15	13260	21
16	2.27	62.75	52				
17	2.27	62.7666667	50				
18	2.27	62.7833333	10				

Prêt

First, coordinates were transformed: latitudes, y, have been expressed in degrees and hundredths, longitudes, x, have been expressed in degrees and hundredths and have been multiplied by the cosine of the mean latitude. Such a transformation in longitudes is necessary at latitude 63° North so that one mile should represent the same distance on x and y. Columns B and C are respectively those of x and y. Column D is that of the acoustic values of the fish density. Column A is that of the loch.

Then each transect has been identified and the values corresponding to each transect have been cumulated. Transects are oriented north-

Abundance estimation performed by
summing up biomass along acoustic transects

south. The transect biomass sum is performed as follows along the y axis Indexed by j:

$$q_i = l \sum_{j=1}^{n_i} f_{i,j}$$

where $f_{i,j}$ is the fish density for the j^{th} segment of length l on the i^{th} transect and where n_i is the number of samples on the transect i ($i = 1$ M.n. here).

Column F includes the number of transects from the south to the north of the area. Column G includes the values cumulated along each transect. Column H includes the length of the positive segments along each transect (that is to say the number of strictly positive values along the transect).

	E	F	G	H
1		1	196149	37
2		2	103920	42
3		3	49427	59
4		4	13803	55
5		5	63489	63
6		6	74406	61
7		7	21587	57
8		8	12627	47
9		9	16061	44
10		10	128681	43
11	2.27	62.6666667	16712	36
12	2.27	62.6833333	12054	26
13	2.27	62.7	8768	31
14	2.27	62.7166667	4056	31
15	2.27	62.7333333	13260	21
16	2.27	62.75		
17	2.27	62.7666667		
18	2.27	62.7833333		

Copier les cellules sélectionnées dans le Presse-papiers

FIGURE 4 -
Copying
the G
column
in the
clipboard.

Abundance estimation performed by summing up biomass along acoustic transects



P. Petitgas, A. Prampart



© Orstom, 1995

III-6

Now import column G in EVA. Select column G or the field G1-G15. Click in the Edit function from the menu bar and choose Copy. The fifteen values are now in the clipboard (fig. 4).

Alt + Tab: return to the Program Manager of Windows and start EVA by double-clicking on its icon. We are now in EVA and both Excel and EVA are open.

DATA IMPORTATION AND SELECTION OF THE GEOSTATISTICAL METHOD

Data button

The Data button from the execution bar allows to import data. Click the Data button: the related window opens (fig. 5). The type of file Data samples is the default selection. This is OK. Select the column Values by clicking in it. The fifteen cumulated values are imported by clicking on one of the two Paste or Paste and link buttons. For instance, choose Paste and link. After the sound signal, click on the Close button. The fifteen values have been imported.

Sampling button

The Sampling button allows to cross the sampling scheme performed with the geostatistical method used for computing the variance. Click on the Sampling button: the related window opens (fig. 6). Choose Scheme A by clicking on the corresponding icon. The frame of the

Abundance estimation performed by summing up biomass along acoustic transects

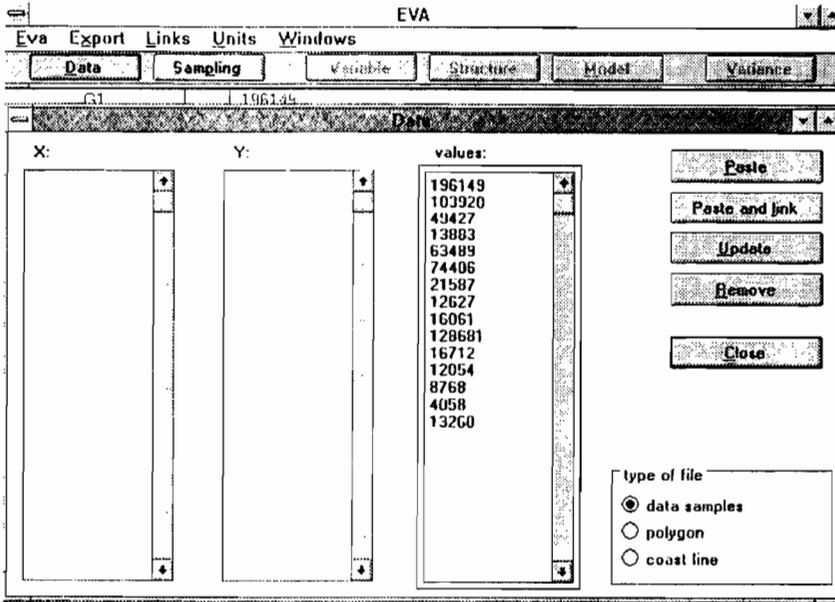
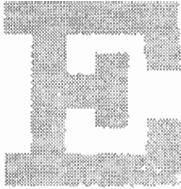


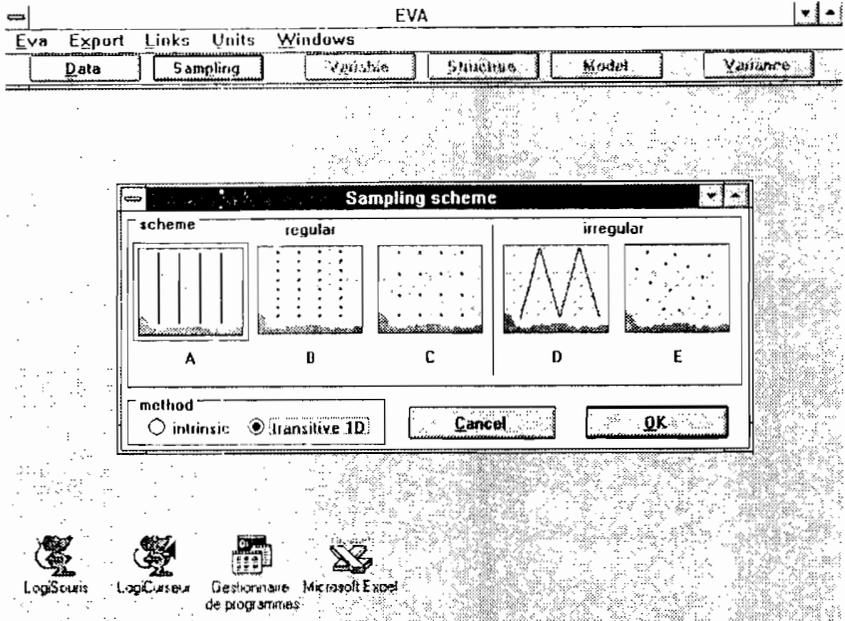
FIGURE 5 - Execution bouton Data. Import the G column in EVA. Click in column Values, then Paste, then Close.

icon is doubled. This shows that the selection is now active. Choose the transitive method by clicking on the circle or on the name. The circle becomes black inside, thus showing that this selection is active. In order to close the Sampling window, click OK.

In the course of the data importation and selection of the method, the other buttons from the execution bar remain inactive. Their names remain shaded. The operations likely to be performed by them are not available. Once the data have been imported and the method selected, buttons become active. The study begins.

Abundance estimation performed by summing up biomass along acoustic transects

FIGURE 6 - Execution bouton Sampling. Cross survey design with analysis method. Click Scheme A, then transitive 1D, then OK.



STRUCTURAL ANALYSIS

This step is performed through the three execution buttons, Variable, Structure and Model.

Variable button

Two windows are related to it. The window named Transect cumulated values gives a representation of the values with a bar diagram. The window

Abundance estimation performed by summing up biomass along acoustic transects

named Variable supplies the elementary statistics of these values. The variance supplied is the data variance which is computed in $1/(n-1)$ (fig. 7).

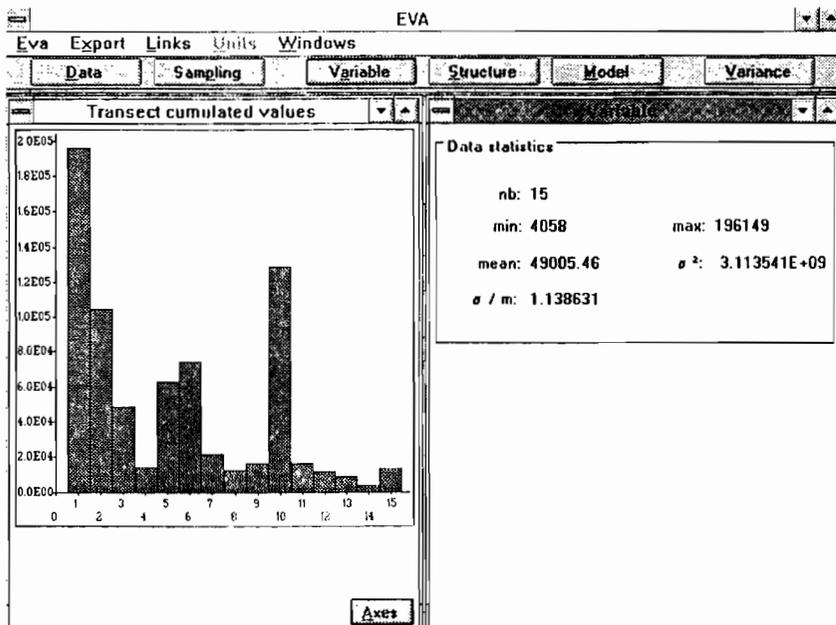


FIGURE 7 - Execution bouton Variable. Spatial representation and basic statistics of the data.

The Axis button from the window Transect cumulated Values opens a window where it is possible to modify the scales on the axes: T means Transects and V means Values.

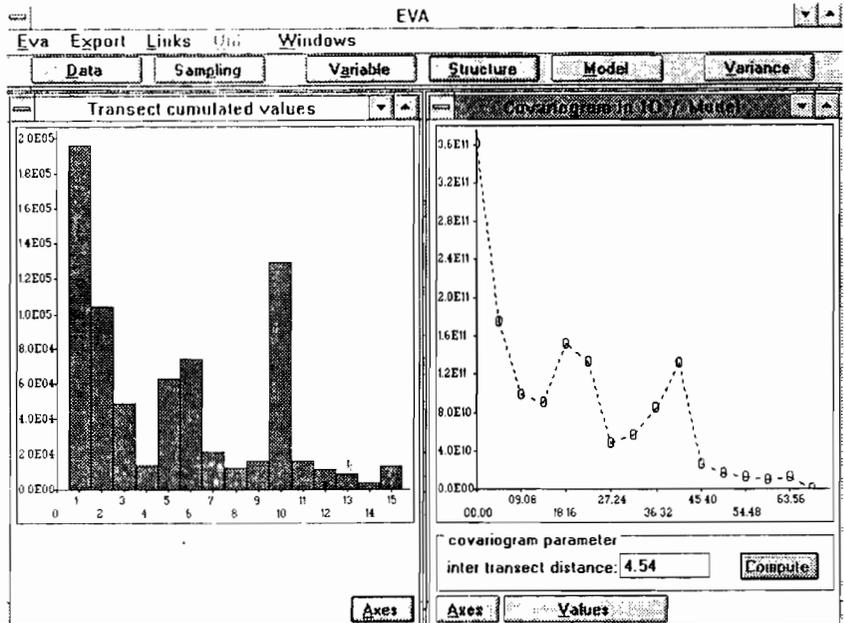
Structure button

Two windows are related to it: the window Transect cumulated values which is the same as previously and the window called Covariogram 1D. This allows

Abundance estimation performed by summing up biomass along acoustic transects

to see at the same time the data and their structure. The Covariogram window allows to compute the experimental covariogram (fig. 8).

FIGURE 8 -
Execution bouton
Structure.
Enter the
inter-transect
distance
and click
Compute.
The experimental
transitive
covariogram
appears
on the screen.



Write 4.54 in the cell of the inter-transect distance and click on the Compute button from the Covariogram window. The curve of the experimental covariogram is displayed.

The Axes button from the Covariogram window opens a window where it is possible to modify the scales on the axes of the covariogram graph: D means distance and V means values of the covariogram.

The Values button from the Covariogram window opens a window where the values of the covariogram are displayed.

Abundance estimation performed by
summing up biomass along acoustic transects

Model button

Two windows are related to it: the Covariogram window which is the same as previously and the Model Parameters window. This allows to fit the model in an interactively (fig. 9).

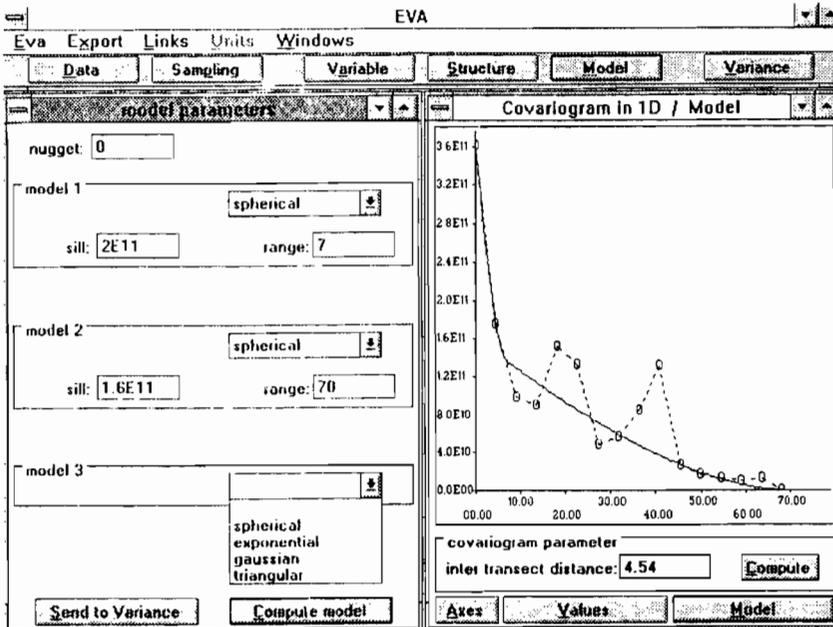


FIGURE 9 - Execution bouton Model. Choose the model types and enter parameters values. Click Compute model. Change parameters interactively until visual fit is satisfactory. Then, click Send to variance

In the Covariogram window, a new button has appeared, the Model button. This button and the Values button allow to toggle at the users convenience from the window of the experimental covariogram values to the window of the model parameters. Thus, one can choose the model parameters so that it goes exactly through required experimental points.

Abundance estimation performed by summing up biomass along acoustic transects



In the Model Parameters window, the type of model is chosen in each of the three scrolling lists which can be accessed by clicking on the arrows. Enter the model parameters as shown on fig. 9. Click on the button called Compute model. The model curve is displayed on the covariogram graph. The model parameters can be changed until the model fit is judged satisfactory by the user.

The Send to variance button from the Model Parameters window allows to pass on automatically to the Variance window the parameters of the fitted model. There will be no need to enter them again. Click this button. The sound signal (bip) indicates that the parameters have been passed on.

ESTIMATION VARIANCE

The Variance execution button enables to compute the estimation variance (fig. 10). The previous parameters have been transferred. It is enough to click in the Compute Variance button in order to start its computation.

The estimation variance is easily computed for other inter-transect distances. It is enough to enter another inter transect distance and do the computation again. Then the values of the curve estimation variance versus/sampling effort are easily obtained. This curve must be drawn with other means than EVA.

In the transitive method, the estimation variance refers to the total quantity and not to the mean. There are fifteen transects, the inter-transect distance is 4.54 N.m. and the simple average of the cumulated values is 49,005.46. The total quantity is estimated as:

$$Q^* = 15 * 4.54 * 49005.6 = 3.3 * 10^6$$

Therefore the relative estimation error is: $\sigma_{est} / \dot{Q} = 12.3\%$

If a model without any spatial correlation between the transect cumulated values was considered, a relative error would be computed using the classical formula,

$$\frac{\sqrt{\sigma^2 / n}}{m}$$

where σ^2 and m are the variance and the mean of the cumulated values and where n is the number of transects. These parameters are supplied in the Variable window. This window can be called back with the Window-show/update function from the menu bar. A relative accuracy of 29.4 % would be found. The estimation is more precise than what is suggested by this result. This computation is erroneous because the transect cumulated values necessarily show a spatial structure. More details and discussion can be found in PETITGAS (1993).

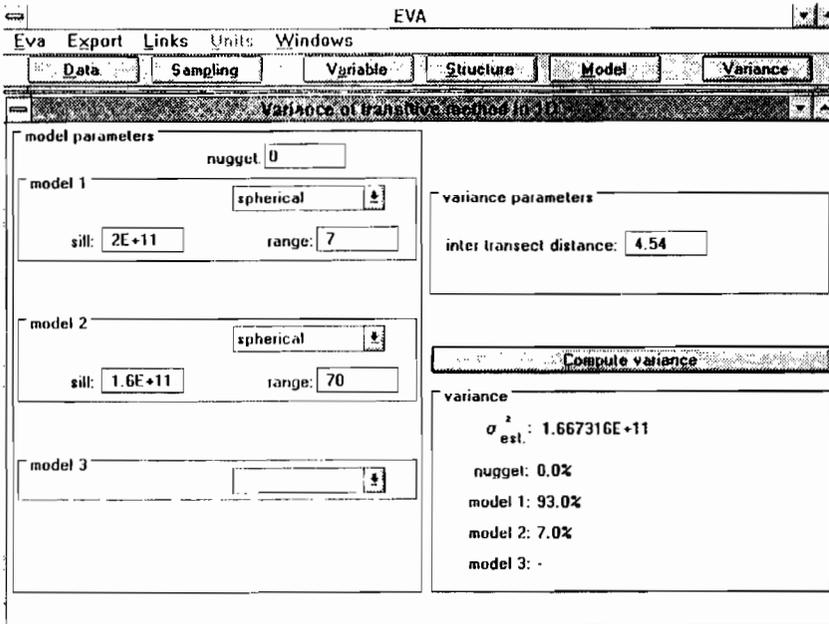


FIGURE 10 - Execution bouton Variance. Click Compute variance.

Abundance estimation performed by summing up biomass along acoustic transects

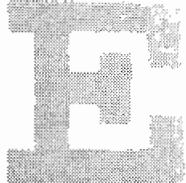
RESULTS EXPORTATION

The Export function from the menu bar unrolls a scrolling list where only the names, Cumulated Values, Covariogram and Covariogram values are in bold characters. These functions are active. The ones with shaded names are inactive and cannot be accessed. The Cumulated Values correspond to the bar diagram of the cumulated values. Covariogram corresponds to the graph where the experimental covariogram and its model are plotted. These figures are easily exported one by one as shown in section II. The values of the covariogram can be saved by pasting them in an Excel worksheet (see section II).

The Variance window contains the model parameters and the estimation variance. The Variable window contains the elementary statistics of the cumulated values. These windows can be printed as shown in section II. The desired window can be called back using the Window/show/update function from the menu bar. Once the required window is active, the command Alt + Screen copy sends it into the clipboard (section II).

USING THE LINK

Let us assume that we now want to make again part of the study, not on the transect cumulated values but on the positive length of transects in order to characterize the geometry of the spawning ground. Column G of the Excel worksheet is linked to the Values column of EVA from the EVA Data window. It is enough to put in G the contents of column H.



P. Petitgas, A. Prampart



© Orstom, 1995

III-15

Alt + Tab: return to Excel. Select Column G. Move the contents of column G to column E using the Cut and Paste functions. Column G is free. Now select column H and move it to column G using the same method.

Alt + Tab: return to EVA. Use the Link function from the menu bar. The sound signal (bip) indicates that the new values are imported. It is not necessary to return to the Data execution button. The selection of the sampling scheme need not be modified. The study starts again from the Variable execution button. A Gaussian model will be selected for it shows a behaviour with horizontal tangent near $h = 0$. This is theoretically that of an indicator variable which has been cumulated along lines (MATHERON, 1965).

Abundance estimation performed by
summing up biomass along acoustic transects



E

P. Petitgas, A. Prampart

V

© Orstom, 1995

A

Estimation of the mean
in an area
from non randomly
located samples:

exemple of use

IV



IV-2

The data set used here as an example has been supplied by the Marine Laboratory of Aberdeen (Scotland) as a test example during the geostatistics course given at Centre de Géostatistique, Fontainebleau, for biologists and statisticians of the ICES community (ARMSTRONG *et al.*, 1992). The data are herring egg densities on a spawning bed in a fjord on the west coast of Scotland. The herring eggs lie in specific sediments which form the beds. The sampling is performed by dredging the sediment with an appropriate gear. The file is supplied on the disk in the DATA directory. It is called HEREGG.TXT.

Before sampling, the spawning area is roughly known. But since sediments move from one year to the other and since the shapes of beds are not fixed due mainly to the currents and to the spawning herring behaviour, it is not possible to define the exact shape of the bed(s) prior to the survey. Thus the sampling must bring two pieces of information: a geometrical one concerning the limits of the bed(s) and a numerical one concerning the egg densities on the bed(s). The constraints linked to the environment and to the work carried out on board are such that samples cannot be taken accordingly to a well defined scheme: the scheme performed is neither random nor regular. By nature, this type of sampling is unorthodox. Many surveys for marine resource assessment show these characteristics with various degrees. This data set is interesting in that it gathers them.

Each value represents a density: it is the number of eggs counted per unit area of sediment. The coordinates of the sample points are expressed in kilometers. They measure longitudinal and latitudinal distances separating each sample from a given origin in the fjord.

CONFIGURATION OF THE EXCEL WORKSHEET

The file includes 8 columns. In the Excel worksheet, the columns are the following (fig. 11):

	B	C	D	E	F	G	H
1	26.1	17.14	1266			25.9	17.05
2	26.03	17.18	1352			26	17
3	26.03	17.14	1568			26.12	17.08
4	26.05	17.09	1039			26.2	17.08
5	26.19	17.18	2064			26.26	17.1
6	26.21	17.11	1216			26.25	17.16
7	26.13	17.13	1353			26.34	17.2
8	26.15	17.12	1529			26.4	17.17
9	26.05	17.21	723			26.51	17.26
10	26.08	17.1	798			26.34	17.35
11	26.19	17.32	986			26.24	17.35
12	26.22	17.26	1165			26.24	17.54
13	26.01	17.05	631			26.14	17.54
14	26.2	17.39	925			26.14	17.23
15	26.25	17.29	1103			26.09	17.19
16	26.27	17.22	1357			26.05	17.29
17	26.34	17.3	1786			25.94	17.2
18	26.35	17.25	1355			25.95	17.12
19	26.38	17.24	1342			25.9	17.05
20	26.41	17.23	1399				

FIGURE 11 -
Data file
HEREGG.TXT
in its Excel
worksheet.
B et C =
abscissa
and ordinate
of sample
points (km),
D = egg density
(nb per unit
surface),
G et H =
abscissa
and ordinate
of the polygon
vertexes
delimiting
the area
studied.

- column A: number of the station;
- columns B and C: coordinates of stations (in km), respectively x and y;
- column D: egg densities;

Estimation of the mean in an area
from non randomly located samples



P. Petitgas, A. Prampart



IV-4

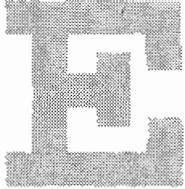
- columns E and F: empty;
 - columns G and H: x and y coordinates of a first polygon defining the spawning bed;
 - columns I and J: x and y coordinates of a second polygon corresponding to a second geometric hypothesis defining the spawning bed.
- Polygons have been defined using EVA.

DATA IMPORTATION AND SELECTION OF THE GEOSTATISTICAL METHOD

Import successively the Excel columns B, C and D into the X, Y and Values columns of the EVA Data window. The series of operations to be performed is quickly summed up followingly and has been detailed in sections II and III.

In Excel, select column B, click in Edit then Copy. The column is in the clipboard. Press Alt + Tab to return to the Program Manager of Windows and start EVA. In EVA, click the Data execution button and select the type of file Data samples. Click in column X, then in the Paste button or in the Paste and link button. For instance here, choose Paste. The sound signal (bip) indicates that pasting is completed. Press Alt + Tab and toggle into Excel. Then, select column C, click in Edit then Copy. Press again Alt + Tab to return to EVA. One returns directly in the Data window. Now, click in column Y and in the Paste button. Return a third time to Excel and repeat these operations to import column D which will be pasted in the EVA Values

Estimation of the mean in an area
from non randomly located samples



column. Then, click in the Close button. The data have been Imported (fig. 12).

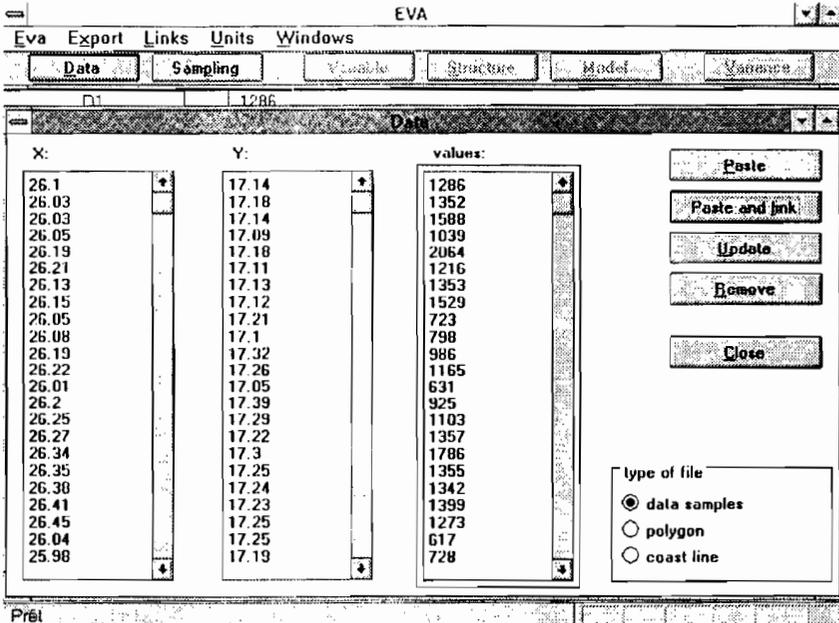


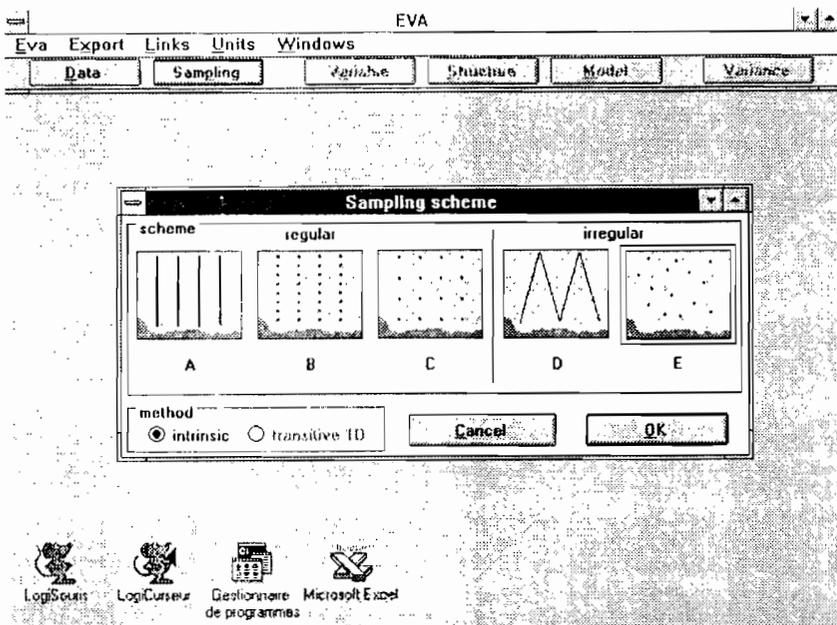
FIGURE 12 - Execution bouton Data. Data window after having imported each data column. Also click on polygon and import columns G and H. Then, click Close.

Click in the Sampling button. Choose scheme E: only the Intrinsic method can be accessed (fig. 13). Click OK. The other execution buttons are now active and the study begins.

If in the course of the study, the Sampling window is called back and the selection of the scheme is modified, all the windows will be closed. This is understandable in so far as changing the scheme consists in carrying out another study.

Estimation of the mean in an area from non randomly located samples

FIGURE 13 - Execution bouton Sampling. Click in Scheme E, then OK. Just the intrinsic method is only possible.



The Sampling button allows to choose between the intrinsic method or the transitive one. It gives access to the transitive method only for the regular scheme with continuously sampled transects. Therefore, it allows to cross scheme A with either method. For all the other schemes, it only gives access to the intrinsic method. In the intrinsic method, whatever the scheme icon selected, the same formula of the estimation variance will be used (section I). For a given data set analysed with the intrinsic method, the scheme icon selected will have no influence in the computations.

STRUCTURAL ANALYSIS

Variable button

The Variable button enables to dissect the data. Once this step has been performed, one should be able to describe the spatial distribution with words, therefore one should be able to know what one wants to model through the variogram or why the variogram is not very adapted to the modeling desired.

Two windows are related to the Variable execution button (fig. 14):

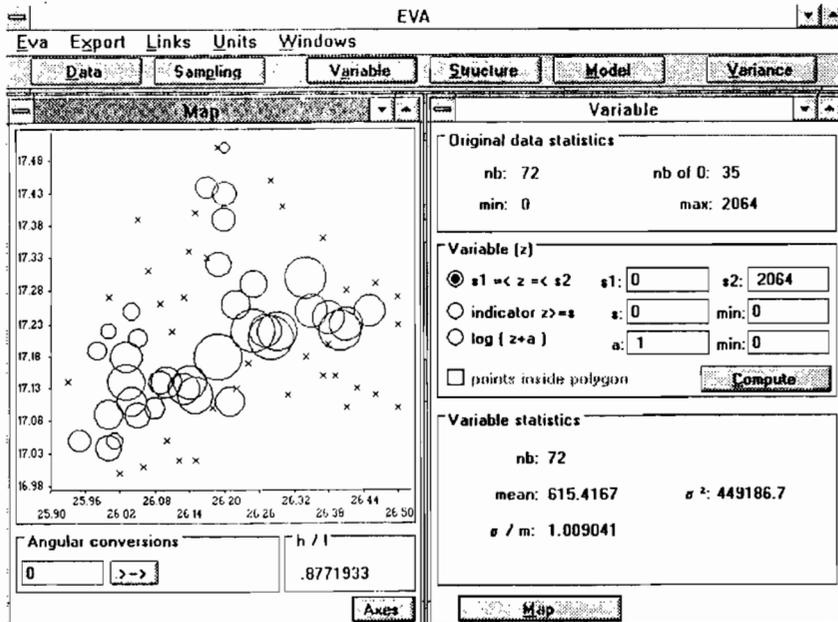


FIGURE 14 - Execution bouton Variable. Click Compute: statistics relative to the choice made in the field variable (z) appear. Click then Map for the spatial representation.

Estimation of the mean in an area from non randomly located samples



the Variable window on the right handside of the screen and the Map window on the left. The Variable window enables to work on the data by defining new variables which are functions of the original data values. At the same time, elementary statistics concerning the new variables are supplied. The Map window gives a spatial representation of the variable chosen. Each value is represented by a circle whose radius is proportional to the value. The representation is displayed relatively to the maximum value: a unit radius is ascribed to the maximum value and all the others are deduced from it by proportionality. If all the values of the data set are identical, the map cannot be supplied.

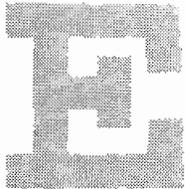
The following computations in the successive steps of the study (Structure, Model, Variance buttons) will be performed on the variable which will have been selected in the Variable window.

The Variable window is divided into three areas. The first one called Original data statistics gives elementary characteristics of the data set. The second one called Variable (z) allows to define a new variable which is a function of the data values. The Compute button leads to the third area called Variable statistics where the elementary characteristics of the variable chosen are computed. The Map button opens the Map window where a representation of the spatial distribution of the variable chosen is displayed. In the Variable (z) area, a selection can be made on sample values and/or on their position.

OPTION POINTS INSIDE POLYGON FROM THE VARIABLE WINDOW

It allows to select only the sample points standing inside a polygon. The polygon defines the area on which the estimation will be performed. When the option is selected, the statistics and the structural analysis will be computed on the polygon inner samples. But the

Estimation of the mean in an area
from non randomly located samples



estimation variance will always be computed with the inner samples only, even though the option has not been selected. This allows the estimation variance to be calculated without going through the structural analysis if the model is already known and if one wishes to test sampling schemes.

The option allows to distinguish the zeroes inside the habitat (density holes) from the zeroes outside it which delimit it. This option can be crossed with the following options.

OPTION $s1 \leq z \leq s2$ FROM THE VARIABLE WINDOW

It allows to truncate the data set by selecting only the data values between both thresholds, $s1$ and $s2$. Variogram and variance computations will then be performed on this reduced data set. On the map displayed the proportional representation of values will be relative to $s2$.

The option allows to keep the quantitative information between quantiles and so to see from which value of $s2$, the spatial distribution changes aspect. Therefore, one can understand better the spatial distribution of the medium values which can be concealed by the high values. Thus, one can give a biological meaning to the term "high values" if their behaviour is different from the behaviour of the other values. On the map, values outside the threshold limits are represented by the symbol higher than ($>$) or the symbol lower than ($<$). The zeroes of the original file are symbolized by crosses.

OPTION "INDICATOR $z \geq s$ " FROM THE VARIABLE WINDOW

It allows to work on the indicator variable which indicates a quantile. The indicator equals 1 when the sample value is higher than s and it equals 0 otherwise.

Estimation of the mean in an area
from non randomly located samples



P. Petitgas, A. Prampart



IV-10

The statistics supplied for this variable are percentages. Let n be the number of data samples and let p be the number of values higher than s . If the values are ranked in ascending order, we get:

$$P(z \geq s) = p/n ; \%Q = \frac{\sum_{i=p}^n z(i)}{\sum_{i=1}^n z(i)}$$

$$\% \text{ var} = \frac{\sum_{i=p}^n (z(i) - m)^2}{\sum_{i=1}^n (z(i) - m)^2}$$

with m being the simple average of the n data values.

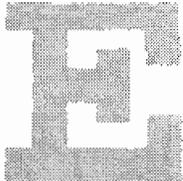
The quantity $\%Q$ characterizes the contribution made by the values higher than s to the total quantity. The quantity $\% \text{ var}$ characterizes the contribution made by the values higher than s to the data variance. These parameters describe the dispersion of values on the histogram.

The number of samples n is the number of the polygon inner samples if the option Points inside polygon is active.

The cell "min" allows to define a threshold below which values are not considered.

The map of the indicator includes three symbols. Circles represent the value 1 and points represent the value zero of the indicator. The symbols lower than ($<$) denote the values lower than the minimum value chosen. Crosses denote the zeroes of the original data set.

The spatial representation of an indicator allows to display the spatial organization of the quantile concerned. The indicator variogram will describe the geometry of the random sets defined by the 1s. More over, when the user modifies the threshold s , the spatial setting of the different quantiles will be displayed. Their links can so be better understood. At the same time, in the Variable statistics area the contributions made by each quantile to the total mean and variance are supplied.



In particular, by decreasing gradually the threshold s , one can see where are the high seeds and whether the lower values are arranged round them (diffusive model) or not (model without transition). Such quantile by quantile dissection may give biological arguments for a stratification of values. It may lead to developing a certain type of non linear geostatistical model, if necessary. This option allows to understand better ones data but it does not allow to carry out a complete non linear study.

OPTION $\text{Log}(z + a)$ FROM THE VARIABLE WINDOW

It allows to work on the Neperian logarithms of the data values. A shift, a , can be made before taking the Logs. A minimum value can be defined: values below this threshold will not be considered in the computations. In particular, this makes possible to keep the zero values or to remove them.

MAP WINDOW

Four types of symbols are used to represent values on the map. The values represented are the ones of the variable chosen: the data values between s_1 and s_2 , an Indicator or the Logs. Each value is represented by a circle. The display is linearly proportional to a maximum value. This maximum is the value of s_2 when the option $s_1 \leq s \leq s_2$ is selected. It is the maximum of the data Logs when the option $\text{Log}(z + a)$ is selected. Crosses represent the zeroes of the original data set. The symbols "higher than" ($>$) and "lower than" ($<$) denote samples outside the thresholds selected in the Variable window. The symbols are shaded when the samples are outside the polygon limits when the option Points inside polygon is active.

Estimation of the mean in an area
from non randomly located samples



In the window, the map is a square object: the ratio between the actual dimensions of the axes is not respected. An apparent direction in the map is not necessarily equal to the actual direction. The angle converter calculates the value of the actual direction when an apparent direction is given. The angle values are given in degrees and hundredths and are counted trigonometrically starting from the x-axis. The h/l ratio is a ratio of anisotropy between the length of the y-axis in terms of actual distance and the length of the x-axis in terms of actual distance.

The Axes button allows to modify the graduation values and the scales on the axes. Moreover, it gives access to the options Show polygon and Show coast line which allow to display on the map the polygon and/or the coast. When the corresponding files have not been imported, these options are inactive.

The up arrow in the top right corner of the Map window allows to enlarge it full screen. This leads to a new function: the mouse positioner. The Mouse localisation cell gives the position of where one clicks in the map with the mouse. This function allows to define quickly a polygon and to make several geometric hypotheses (separation of zeroes into inner and outer zeroes). Coordinates of vertexes will have to be copied manually on a sheet, then entered and saved in an Excel worksheet. Then the file can be imported into EVA (see sections II, III, IV).

The Variable button in the enlarged Map window allows to toggle back into the Variable window. The Axes button leads to the same Axes window as previously. The double up and down arrow in the top right corner of the full screen Map window allows to return the window to its normal size (fig. 15).

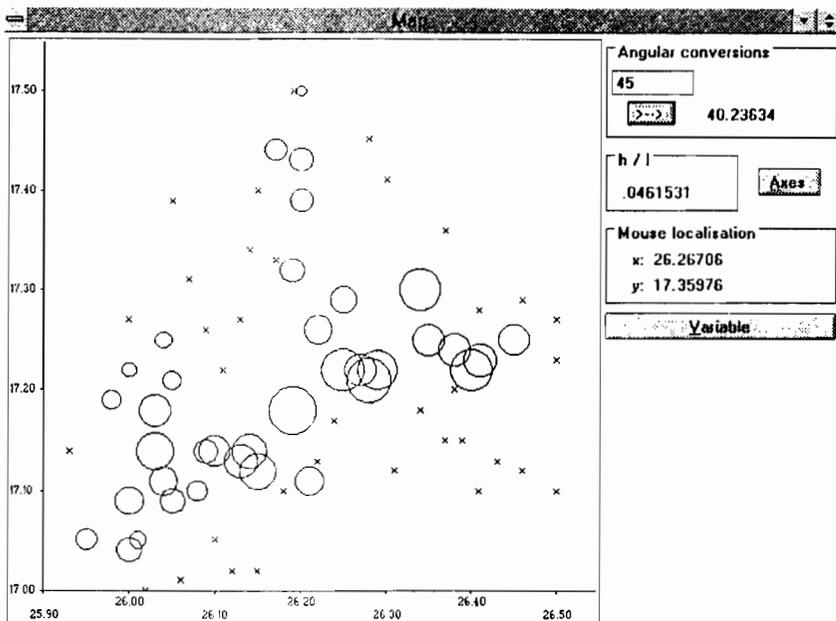


FIGURE 15 - Full screen view of the Map window obtained by clicking the top right bouton of the window Map which shows an upwards arrow. Click anywhere in the map, mouse position is given in the field Mouse localisation.

EXAMPLE

Firstly in the default option, all the data are displayed: s1 and s2 are the minimum and maximum values of the file. Click on the Compute button: the statistics of the data set are supplied. Click on the Map button: the Map window supplies us with a proportional representation of values (fig. 14). Click on the Axes button, the Axes window opens. Modify the graduations as follows in order to get a clearer figure. Insert the following values: Xmin = 25.9, Ymin = 17, Xmax = 26.55, Ymax = 17.55, lag on X = 0.1, lag on Y = 0.1, and click OK. The spatial distribution shall now be described.

Estimation of the mean in an area
from non randomly located samples



There are 49 % of zero samples. Zeroes surround geographically an area composed mainly of positive values. Therefore, zeroes are interpreted as zeroes outside the spawning bed which is the area with positive values. These are aligned on a main axis whose apparent direction is north-east south-west. The angle converter gives an angle of 40° for an apparent angle of 45° (angles are counted trigonometrically from the X axis). From this main axis, there are two secondary axes orientated north-north-west. The zeroes surrounding the positive values can be very close to them particularly near edges. But the inner part of the axis does not contain any zero. We are faced with a geometric problem: there is a pronounced anisotropy of the spawning bed (probably due to the particular distribution of the sediment, but no information is available on this variable). This exerts a little influence on the spatial distribution of the egg densities on the bed. The bed limits cannot be well defined in the areas where the sampling performed lacks sample points.

If the structural analysis is made on the total data set, the structural information due to the proper shape of the bed (structure of the indicator of positive values in the whole data set) will be mixed with the structural information concerning the egg densities on the bed (structure of values within the bed). If a polygon drawing the outlines of the bed is defined then the inner data can be selected and the specific structuration on the spawning bed can be characterized. This approach is adopted. The estimation will be made in the polygon which defines the outlines of the bed. Then this polygon can be modified to measure the influence of the undetermined limits of the bed on the estimation and on its precision.

Particular attention is paid to defining the limits of the bed which will constitute the limits of the estimation polygon:

Estimation of the mean in an area
from non randomly located samples

— In the EVA Map window, enlarge full screen the window by clicking on the up arrow in the top right corner. Define a polygon by using the Mouse localisation function (fig. 15). Click in the map on the summits of the polygon to be defined. The positions displayed must be copied manually on a sheet. For instance, a polygon can be selected which follows the positive values and leaves some zero values within it;

— Alt + Tab: return to the Excel worksheet. Choose a column where to type the X values and another one for the Y values. The EGGHER.TXT file already contains two polygons (see section IV, "Configuration of the Excel worksheet") in columns G and H and in columns I and J. They have been defined according to the previous method. Now select column G and copy it into the clipboard using the Edit-Copy function;

— Alt + Tab: return to EVA. The Map window is resized in order to be able to have access to the execution bar. Open the Data window by clicking on the Data button. Select the Polygon type of file by clicking on the name or in the corresponding circle: the circle becomes black, the selection is active. Two columns appear, one for X and the other for Y. Select X by clicking inside the X column, then click in the Paste and link button. Later on one may want to make the estimation with another polygon so linking is preferred;

— the same operation is repeated for column Y: Alt + Tab, return to Excel, select the column H, copy this column in the clipboard, Alt + Tab, return to EVA, select the column Y of the Polygon file and Paste with link;

— click in the Close button from the EVA Data window. The polygon has been Imported.

Call back the Map and Variable windows by clicking, for instance, on the Variable execution button. The variable is no longer updated: click

Estimation of the mean in an area
from non randomly located samples



IV-16

again on the Compute button. The Axes button from the Map window allows to lead to the option that enables to display the limits of the polygon. In the Axes window, the Show polygon option can now be accessed. Click in the Axes button: the Axes window opens. Select the Show polygon option by clicking on the name or in the corresponding cell: a cross appears in the cell showing that the selection is active. Then click OK. The map is displayed with the limits of the polygon.

Select the option Points inside polygon by clicking on the name or in the corresponding square cell: a cross appears in the cell indicating the selection is active. Henceforth, one will work on the file reduced to the inner samples: outer zeroes have been removed. Click on the Compute button: statistics are obtained for this reduced file. These are: $n = 46$, $m = 963$, $\text{var} = 3.66 \text{ E}5$. The map is drawn again and outer zeroes are represented by shaded crosses (fig. 16).

The reduced file can be dissected statistically and spatially by using the "indicator" selection from the Variable window. Click on the "indicator" word or in the corresponding circle, the latter becomes black. Leave the minimum value to zero and take a threshold of 2,000. Click on Compute: statistics are computed on the polygon inner data. The map shows that a single value is higher than this threshold and that it is situated in the middle of the axis of the bed. This high value represents 5 % of the total quantity and 7 % of the variance in the polygon defined (that is to say on the bed). Select a lower threshold, for instance 1,500. Click Compute. Eight values are higher than 1,500. These values represent a third of the mean and of the variance on the bed. The values which are immediately lower than the highest one are not positioned in its immediate vicinity. They are distributed rather homogeneously on the whole main axis of the bed. The secondary axes do not show any high values. The southern edge of the bed is very steep, for high values

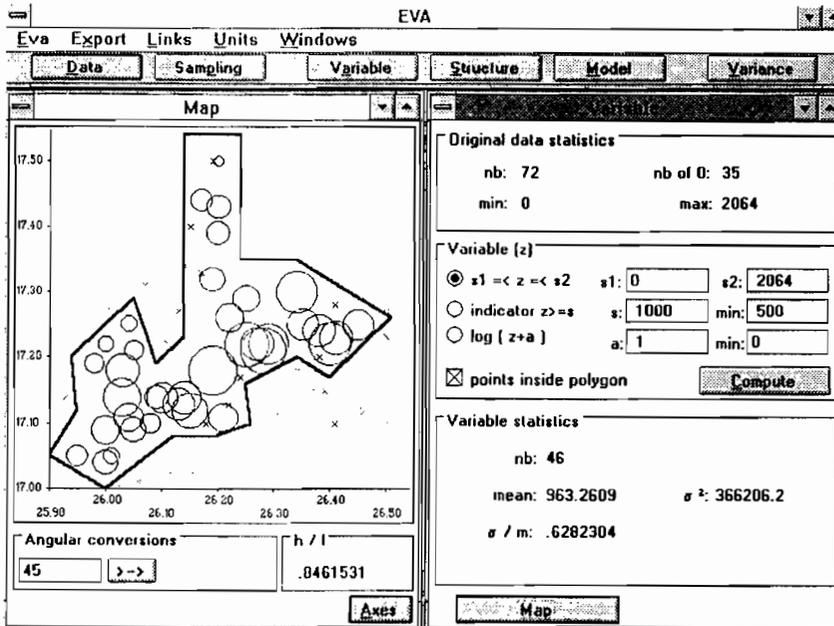


FIGURE 16 - Execution bouton Variable. Modify your selections in the field variable (z). Click Points inside polygon. Click Axes and choose option Show polygon, then OK.

are very close to the border of zeroes. The northern edge is smoother. These are the main features of the spatial distribution (fig. 17).

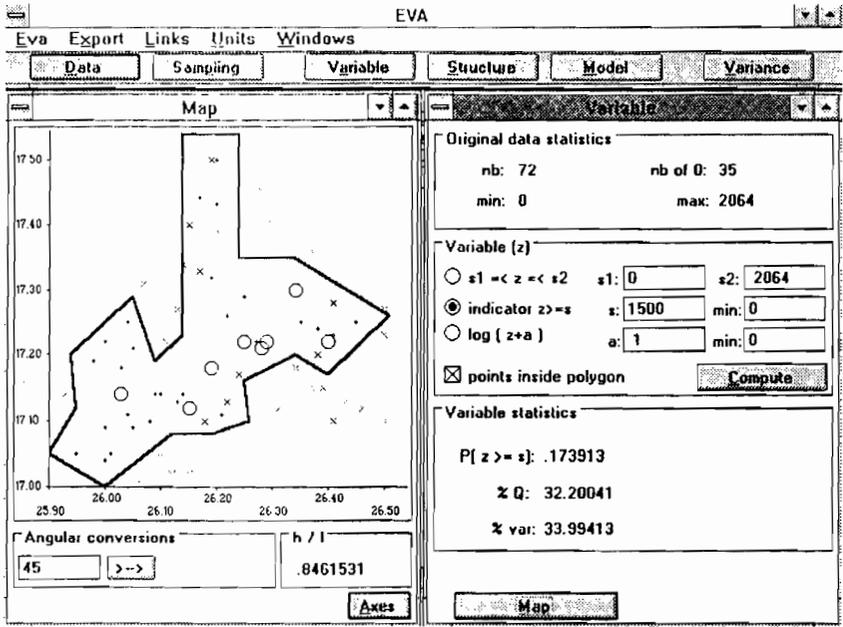
The egg densities has an elliptical bell shape whose center is the highest value, whose great axis is the main axis, whose southern edge is somewhat truncated and northern edge is smoother and which gets two ramifications. This regionalization can be modelled in different ways.

Before going on, select again the $s_1 \leq z \leq s_2$ option for it defines the variable we shall work on. Keep the Points inside polygon option selected. Click again Compute. The data set formed by the polygon inner samples is in memory and the analysis can be carried on.

Estimation of the mean in an area
from non randomly located samples

IV-18

FIGURE 17 -
 Execution
 bouton
 Variable.
 Modify your
 selections
 in the field
 variable (z).
 Choose
 Indicator
 Z > = z,
 enter
 value 1500,
 then click
 Compute.



Structure button

Click the Structure execution button. Two windows are related to it: the Map window which is the same as previously and the Variogram window on the lefthand side. This allows to display the spatial distribution while characterizing it using a variogram.

The variogram is a variance between points as related to the distance separating them. Coordinates of the points are expressed in kilometers. Distances will be computed in the same unit. Therefore, this choice

Estimation of the mean in an area from non randomly located samples

must be activated in the menu bar: click Units from the menu bar then choose the option Coordinates and distances in identical units. A label shows that this selection is now active.

The parameters concerning the computation of the experimental variogram must be supplied by the user. This is done in the lower part of the Variogram window called Variogram parameters. The parameters necessary for computing the experimental variogram have been defined in section I. For instance, type the following parameters. Distances: nb. lag = 6; lag = 0.05; lag tolerance = 0.025. The maximum distance computed will be 3 km corresponding roughly to

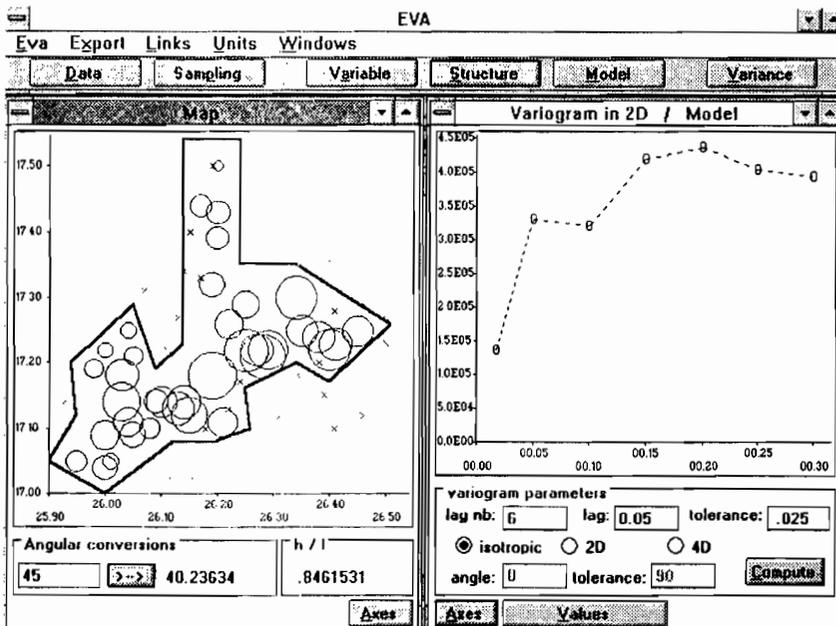


FIGURE 18 - Execution bouton Structure. Enter your parameters for computation of experimental variogram. Variogram is computed on the selection chosen in the Variable window (here active selections are: $s1 \leq Z \leq s2$, $s1 = \min$, $s2 = \max$ and Points inside polygon).

Estimation of the mean in an area from non randomly located samples



P. Petitgas, A. Prampart



half the great axis. Angle: Isotropic. Click on the Compute button: the experimental variogram graph is displayed after some computational time.

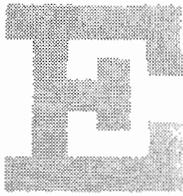
The Values button opens a window which gives the values of the experimental variogram.

The Axes button opens a window which allows to modify the graduations of the Variogram figure: D means distance and V means values of the variogram. Click on this button to modify the graduations concerning the distances. Take $D_{max} = 0.32$ and distance lag = 0.05 and click OK. The figure is redrawn with the new axes (fig. 18).

Model button

Two windows are related to it: the previous Variogram window on the left handside and the Model parameters window on the right. In the Variogram window, a new button appears: the Model button. It allows to toggle from the Variogram values window to that of the Model parameters window. Thus parameters can be adjusted to make the model go through required experimental points.

Model parameters have been detailed in section I. The types of models are chosen in the scrolling lists which can be accessed by clicking on the down arrows in the Model 1,2,3 areas. In order to select a type, it is enough to click on its name. The first line of the list is empty. When it is active, no model is defined. Once the type of model has been chosen, the model area opens and its parameters can then be specified. Now click in the arrow of the Model 1 area. Choose the exponential type of model. Fill in the parameter cells as follows: nugget = 5 E4 + model 1 = exponential (sill = 3.7 E5, practical range = 0.15).



The model is isotropic. So leave the anisotropy parameters as: coeff. X = coeff. Y = 1, rotation = 0.

The Compute model button allows to display the model curve on the experimental variogram graph. The parameters of the model can be modified interactively until the model fit is considered as satisfactory by the user (fig. 19).

The Send to variance button allows to pass on the model parameters to the Variance step. It will not be necessary to type them again for the variance computation. Click on the Send to Variance button. A sound signal (bip) indicates that the parameters have been passed on.

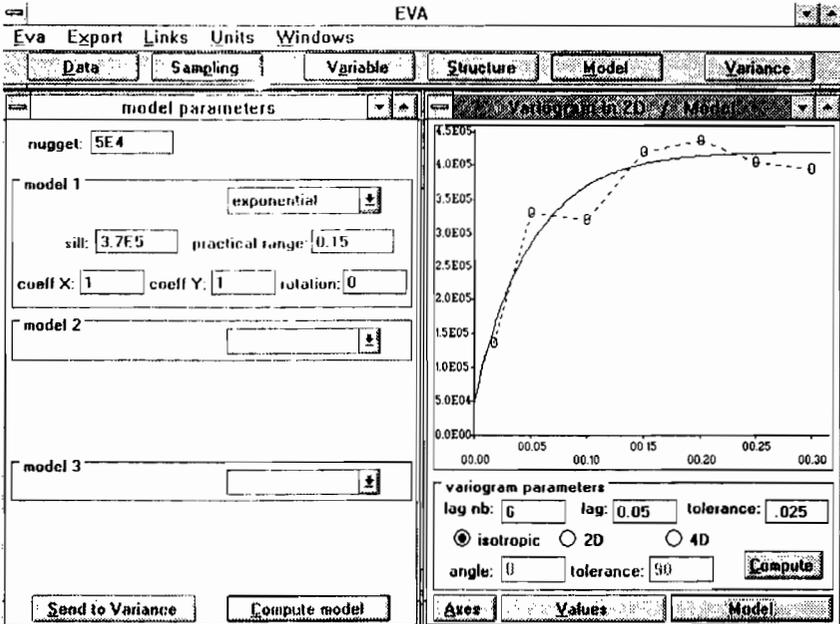


FIGURE 19 - Execution bouton Model. Fit model interactively and visually. When finished, click Send to variance.

ESTIMATION VARIANCE

The Variance execution button leads to the computation of the estimation variance. This computation is possible only if a Polygon file and a file of sample points have been imported in EVA (see sections I and II).

In the computation of the estimation variance, only the points within the polygon are considered, even though this option has not been selected in the Variable window. The reason for this is that, in order to test a sampling scheme, it is possible to access to the estimation variance through path 2 (see sections I and II), that is to say merely with a file of sample coordinates, a Polygon file and a Varlogram model. If the structural analysis is made on the total data file and if a polygon which does not embrace all the values is defined, the estimation variance will be computed with the data within the polygon but using a model concerning the whole set of values.

Two windows are related to the variance execution button. The Variance button opens first a window called Estimation method where the type of estimation is chosen. Then for the required type, the estimation variance is computed in a second window called Variance with spatial structure which opens later.

Estimation method window

The Variance button opens a first window called Type of estimation. It is in the form of a table which sums up the hypotheses of the different estimation methods in order to position the methods used into a larger context (fig. 20).

Estimation of the mean in an area
from non randomly located samples

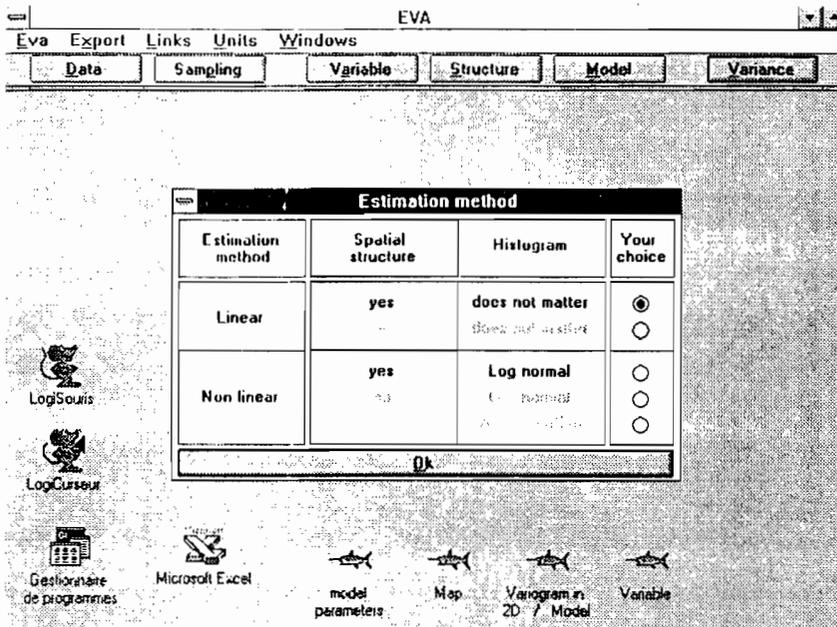
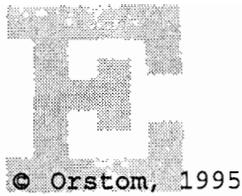


FIGURE 20 - Execution bouton Variance. First window enabling specification of estimation procedure. Variogram was computed for the untransformed variable. Click on the line Linear estimation and Histogram does not matter, then click OK.

The type of estimation performed must be specified: linear estimation using the arithmetical mean or lognormal estimation in the case of correlated data. The other cases are written in shaded characters, they are inactive and cannot be accessed. In the column Your choice, click in one of the two circles corresponding to one of the two possible cases. The circle becomes black: the case is selected. It is also possible to click on the words "yes", "does not matter", "lognormal" in order to select the line. Choose linear estimation with spatial structure, then click OK. A second window opens called Variance with spatial structure. The estimation variance for this required case shall be computed.

Estimation of the mean in an area from non randomly located samples



Variance with spatial structure window

The Variance with spatial structure window in the case of the linear estimation is composed of three fields. The field on the left hand-side called Model parameters is identical to the contents of the Model Parameters window. The previous Click on the Send to variance button has transmitted the model parameters here. On the right handside there are two fields. Results are given in the lower field called Estimation variance. The results supplied are detailed in section I. The higher field called Discretisation of polygon allows to define the parameters concerning the polygon discretisation. This is used for the numerical computation of the G_{VV} and G_{AV} terms of the estimation variance formula as explained in section I (fig. 21).

Click again on the Variance execution button: the Estimation method window opens again. Now choose the non linear lognormal estimation and click OK. A new Variance with spatial structure window opens which now corresponds to the computation of the estimation variance concerning lognormal and correlated data. This window includes the three previous fields in addition to a fourth one in the top left corner which gives the Logs parameters transmitted automatically from the Variable window when the Log selection is used. As this selection has not been used, no value has been transmitted: the default values are equal to zero. The results supplied in the Estimation variance field are detailed in section I. Now return to the linear estimation: click on the Variance execution button, choose the linear estimation in the table and click OK.

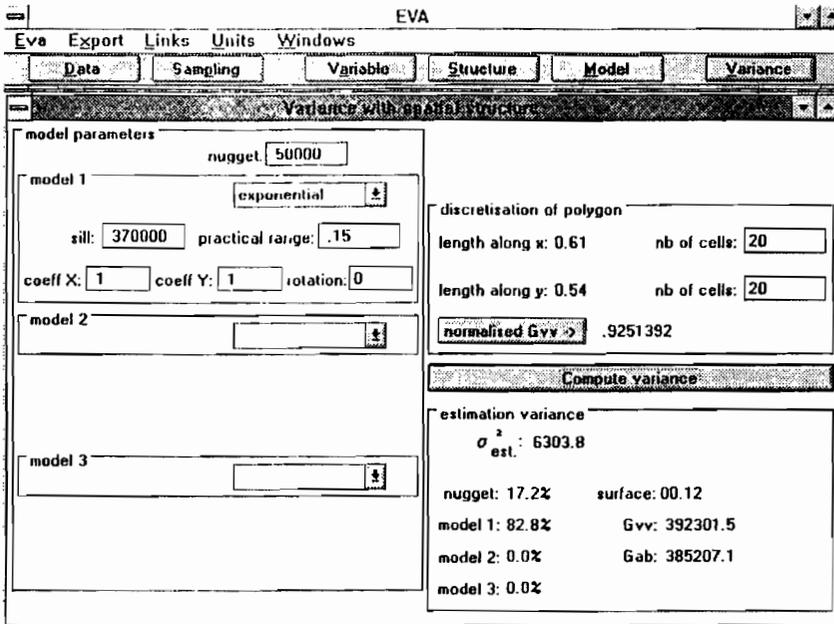
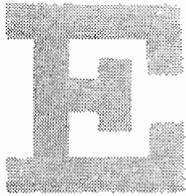


FIGURE 21 - Execution bouton Variance. Second window enabling computation of estimation variance. Enter parameters of discretisation and click G_{VV} for controlling if discretisation is sufficient. When ready, click on Compute variance.

SCREEN FIELD DISCRETIZATION OF POLYGON

The maximum length of the polygon on X is given in the "length along x" cell and the same holds true for Y. These lengths are expressed with the units which have been chosen with the Unit function from the menu bar. A computation grid is defined by giving the number of grid cells on X and Y. Grid cells whose centres are within the polygon will only be used in the computations. Grids with a non square mesh can be defined.

The button called Normalized G_{VV} starts a standardized G_{VV} computation. The result is given for the models defined on the left handside of

Estimation of the mean in an area from non randomly located samples

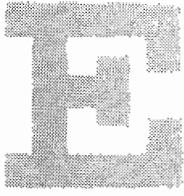
the screen but in this computation there sill is equalled to 1. It is an intermediate result which allows to optimize the computation grid. If the grid defined is very fine, the computation of G_{VV} and G_{GV} will take much time. On the contrary, if the grid is not fine enough, the numerical computations will not be precise enough. The mesh of the computation grid should not be too big in comparison to the mesh of the sampling grid.

By increasing gradually the number of grid cells, the standardized G_{VV} is modified, but this variation gets gradually smaller and smaller. The accuracy required on the standardized G_{VV} depends on the value of the variogram sill. Let us assume a sill whose unit is 10^6 . An accuracy of about 10^{-4} or 10^{-5} will be required for the standardized G_{VV} . The number of cells will have to be gradually increased so that the accuracy required on the standardized G_{VV} should be stabilized at that level.

The computation of the G_{VV} is performed using the algorithm of the stratified random mesh as detailed in section I. For the same grid, different values for the normalized G_{VV} will be obtained by running the computation several times. This second source of variability also helps define an adequate mesh size for the computation grid.

The computation of the standardised G_{VV} is an intermediate step which is not necessary. Clicking directly on the Variance computation button after having defined a computation grid will start the computation of the estimation variance with this grid. The corresponding value of the standardized G_{VV} will be supplied. In this case, it is advisable to take a fine enough mesh, say a grid of 50 x 50 cells or 100 x 100.

Suppose that the standardised G_{VV} has already been computed. If one of the parameters of the variogram model or of the computation grid has been modified, the Compute variance button will automatically restart the computation of the standardised G_{VV} .



In the example, a 20 x 20 grid has been considered as satisfactory through successive trials (fig. 21).

COMPUTE VARIANCE BUTTON

This button starts the computation of the estimation variance for the computation grid chosen, the model defined and the standardized G_{VV} computed. The button will restart automatically the standardised G_{VV} computation if a parameter value of the model or of the computation grid has been modified.

The area inside the polygon limits is that over which the estimation variance is computed. The surface is calculated by counting the number of computation grid cells whose centers stand inside the polygon limits. The units for the surface are the square of the distance units that have been selected with the Units function from the menu bar.

Let m^a be the simple average of the data inside the polygon. This parameter is given in the Variable window. Thus the relative estimation error is:

$$\frac{\sigma_{est}}{m^a} = \frac{\sqrt{6304}}{963} = 8.2\%$$

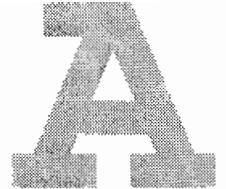
Note that the G_{VV} value is a little higher than that of the data variance (computed on the inner samples) but it remains of the same order of magnitude. A sampling scheme made of the same number of 46 stations randomly located in the polygon would give a relative estimation error of:

$$\frac{\sqrt{G_{VV} / n}}{m^a} = \frac{\sqrt{3.92E5 / 46}}{963} = 9.6\%$$

Estimation of the mean in an area
from non randomly located samples



P. Petitgas, A. Prampart



IV-28

RESULTS EXPORTATION

As explained in section II, the Export function from the menu bar or the screen copy key are used for exporting results. The Export function from the menu bar unrolls a scrolling list where the names, Map, Variogram and variogram Values are active.

OTHER MODEL, ANISOTROPIC MODELLING

At any time, it is possible to return to a previous step of the study by calling back the window desired. Let us return, for instance, to the selection of the model.

Let us compute an anisotropic variogram. Call back the Structure and Map windows by using the Window-Show/update function from the menu bar or merely by clicking again on the Structure execution button. In the Map window, use the angle converter: write 45 in the left handside cell and press the button, the result is given in the right handside cell. An apparent angle of 45° is an actual angle of 40° . In the Variogram window, the variogram parameters concerning the distance lags need not be changed. Change the variogram parameters concerning the directions. Select the option 2D: the circle becomes black. Write 40 in the angle parameter cell and leave the default angular tolerance at $\pm 45^\circ$. Click on the Compute button. After some computational time, two curves appear, one for each direction, say 40 and its perpendicular 130 (fig. 22).

Estimation of the mean in an area
from non randomly located samples

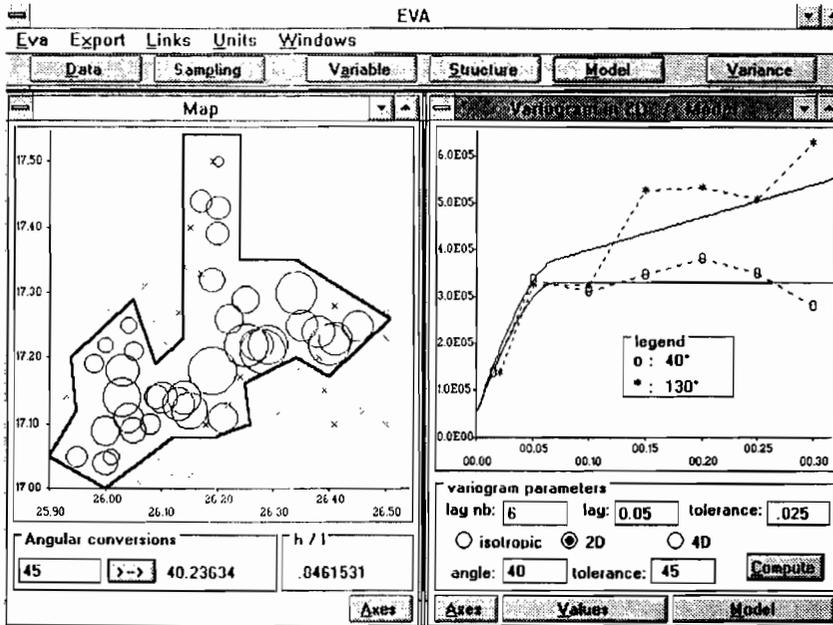


FIGURE 22 - Anisotropic model. Windows Map and Variogram/2D were put together. Click anywhere in the variogram figure: the legend appears.

Click with the mouse in the variogram figure, a legend box appears. This box indicates which curve symbol corresponds to which direction. The legend box is displayed as long as the user holds the mouse click down. This information is also found in the variogram Axes window.

Click on the Values button. In the Values window, there are two buttons in the bottom of the window. An angle value is written on each of them. These buttons allow the display of the variogram values for each direction. Click on one of them; the variogram values for this direction are displayed in the window and the angle value on the but-



ton is surrounded with brackets. The brackets indicate which direction is being displayed.

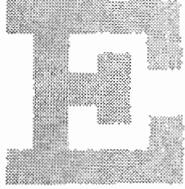
The first three points of the experimental variogram are similar in both directions: the structure is isotropic up to 0.1 km. Then, in direction 40, a sill is observed, while the variogram goes on climbing in direction 130. The experimental variogram is interpreted biologically: Eggs are structured into isotropic aggregates whose mean diameter ranges from 0.05 to 1 km. In direction 40, aggregates show similar average densities and the variogram reaches a sill. In direction 130, beyond 0.1 km, while distance is increasing, aggregates show increasingly different average densities and the variogram goes on increasing. This quantifies the description made with words of the elliptical morphology of the spawning bed: a main axis had been described with direction 40 whose northern edge was smoother than the southern edge (section 4.3.1.6).

Therefore, a model with zonal anisotropy is suggested which would consider these two aspects. We suggest to take a sum of two models: an isotropic model with a sill and a linear model only in direction 130. The isotropic model with a sill models the structuration on the bed. The linear model in direction 130 only models the gentle gradient that the spatial distribution shows in this direction. In direction 130, the sum of the two models means that the eggs are structured in isotropic aggregates and that, while distance increases, these aggregates get increasingly different mean densities.

The model selected is: nugget = 5 E4

+ isotropic spherical (sill = 2.8 E5, range = 0.07); coeff. X = coeff. Y = 1, direction = 40

+ power (sill = 7 E5, exponent = 1); coeff. X = 0, coeff. Y = 1, direction = 40.



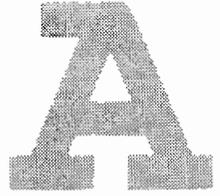
Fill in the Model parameter window with these parameters. Click in the Compute model button. For the choice of the anisotropy parameters, refer to section II.

Figure 22 shows the Variogram window once the axes have been modified and the variogram has been modelled. On the right hand-side, the Map window has been called back by using the Window-Show/update function from the menu bar.

The variogram values for each direction can be exported into an Excel worksheet. Click on Export (in the menu bar) then select variogram values, then select a direction. The sound signal (bip) indicates that the values are in the clipboard. Toggle Into Excel, open a new worksheet, click in a cell and Click Edit-Paste. Repeat the same operation for the other direction.

Now let us compute the estimation variance with this new model. Return to the EVA Model parameters window. Click in the Send to variance button. Click in the Variance execution button. Select the line of the linear estimation in the Estimation method window. In the Variance with spatial structure window, the computation grid parameters stay unchanged (20 x 20 grid). The new model has been imported. It is thus enough to click in the Compute variance button. G_{VV} is somewhat higher: 4.2 E6. The linear model in direction 130 contributes by 15 % to the estimation variance. The relative estimation error is of the same order of magnitude as previously: 8.4 %.

The previous Isotropic model is not very different from the anisotropic model fitted here. Ranges are similar and account for the same Isotropic egg structuration. In the previous isotropic model, the bigger heterogeneity in direction 130 was accounted for by a higher isotropic sill. In these models, the positions of the aggregates in the polygon are not specified.



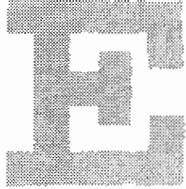
OTHER POLYGON, OTHER GEOMETRIC HYPOTHESIS

At any time, it is possible to return to a previous step of the study. Let us now define another polygon. Return to the Map window in order to define a new polygon. Call back this window either by clicking on the Variable execution button or by selecting Map in the scrolling list of the Windows - Show/update function from the menu bar. The method used is the same as explained above (section IV): enlarge the window full screen and use the Mouse localisation function. A second polygon has already been defined and it is found in the Excel worksheet in columns I and J.

The link established between columns G and H from the Excel worksheet and columns X and Y from the EVA Polygon file is used.

Alt + Tab: toggle into Excel. Using the Cut and Paste functions, move the contents of columns G and H into columns E and F. Now move the contents of columns I and J into columns G and H.

Alt + Tab: return to EVA. Go into the Links-update function from the menu bar. In the scrolling list, only polygon is written in bold characters and is active: only the columns of the Polygon file have been imported with a link. Select Polygon. Two sound signals (bip bip) show that columns X and Y have been updated. But the variable must be updated. One will have to perform the study with the new polygon going through the Variable, Structure, Model, Variance execution buttons.



P. Petitgas, A. Prampart



© Orstom, 1995

test
of a sampling scheme
when the spatial
structure is known:

exemple of use



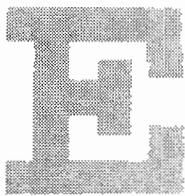
This study is a continuation of the previous study (section IV). A structuration on the spawning bed has been characterized by the following exponential isotropic variogram: $\gamma(h) = \text{nugget } (5 \text{ E}4) + \text{exponential } (\text{sill} = 3.7 \text{ E}5, \text{ practical range} = 0.15 \text{ km})$. The polygon of the spawning bed has been defined. The sampling was composed of stations scattered over the bed and outside it, neither randomly nor regularly. What would the estimation variance be if it was possible to sample the bed with a regular grid of samples, given the same spatial structure?

In order to answer this question, already available information is necessary as well as new one. The existing information is the Polygon file and the Variogram model parameters. The new information is the coordinate file of the new stations. Data values are not required. The new station coordinate file must be generated. It is not possible to define exactly a sampling effort on the bed because it is not possible to foresee exactly how many stations will be found in the polygon by defining a given mesh size. A grid with a mesh 0.06 km square whose origin is the coordinate point (x = 25.9, y = 17) has been chosen. The new sampling grid is made of 11 cells on X and 11 cells on Y.

MAKING A REGULAR SAMPLING GRID WITH AN EXCEL 4 WORKSHEET

The new regular sampling grid is easily generated with Excel 4 by using the recursive properties of formulas, the Edit-Copy down function and the Edit-Copy/Paste function. The following sequence of operations is suggested.

Test of sampling scheme when
the spatial structure is known



Create a new Excel worksheet. Column A will represent the Xs and column B the Ys. Write 25.9 in cell A1. In cell A2, write the formula: = A1 + 0.06. Then select the rectangle A2-A11. In Edit, click the function Copy down. As data and not formulas must be imported into EVA, the rectangle A2-A11 must be selected and formulas must be transformed into values. For this purpose, as the rectangle A2-A11 is still selected, in Edit, click Copy, then in Edit click Special pasting, then select the option Values in the window, then click OK. Go to the B column. Write 17 in cell B1. Copy this value down up to cell B11. The first row of points of the sampling grid has been generated. Repeat this operation 11 times. In column A the rectangle A2-A11 shall be copied each time without modification, from A12 to A23, from A24 to A35, and so forth. In column B the value Y copied for generating the row must be increased each time by 0.06.

In order to display the map of the sampling scheme in the EVA Map window, fictitious values must be given to the grid samples. Copy a column of 1 in column C. But if all sample values are equal EVA will not be able to display a map. So, modify the first value by writing for instance 2 in cell C1. Now, EVA can be started and the three column file can be imported.

This file is supplied on disk in the DATA directory and is called REGEGG.TXT. In this file, the D and E columns contain the values of X and Y of the previous polygon which defines the spawning bed.

IMPORTATION IN EVA

In EVA, click on the Data execution button. The Data window opens. Select the type of file Data samples. Columns A, B and C are imported

Test of sampling scheme when
the spatial structure is known



P. Petitgas, A. Prampart



V-4

from the Excel REGEGG worksheet into columns X, Y and Values using the already described method. The Paste and link button will be preferably chosen in case one would like to modify the mesh of the sampling grid: if modifications are made in Excel, it will be enough to update in EVA. Now select the type of file Polygon and import columns D and E from the Excel REGEGG worksheet into columns X and Y. Click in the Close button. The new sampling grid and the polygon have been imported.

In the Sampling window, any scheme can be selected in so far as it is crossed with the intrinsic method. Choose one. The study begins.

REPRESENTATION OF THE SAMPLING GRID

The Variable execution button allows us to display the sampling scheme. Selecting the option Points inside polygon will have an influence only on the map aspect but not on the computation of the estimation variance. If this option is selected, the outer points will be shaded inactive on the map, the inner points will be counted and the statistics in the Variable window will be computed only with the inner points.

It can be useful to choose the "indicator" selection and to take a sill higher than the maximum value: only small points will represent the stations.

Select the options $s1 \leq z \leq s2$ and Points inside polygon then click on the Compute button: one reads that there are 34 stations within the

Test of sampling scheme when
the spatial structure is known

polygon; say less than in the previous sampling. If more points were required, it would be necessary to modify the mesh of the grid in Excel and to update data using the Link function from the menu bar. This scheme is satisfactory.

Now select the "indicator $z \geq s$ " option and take 3 for the threshold value s (the maximum value is 2). Leave the Points inside polygon selection active. Click on the Compute button, then on the Map button. Click on the Axes button from the Map window. Activate the Show polygon option in the map Axes window and click OK. Figure 23 is obtained.

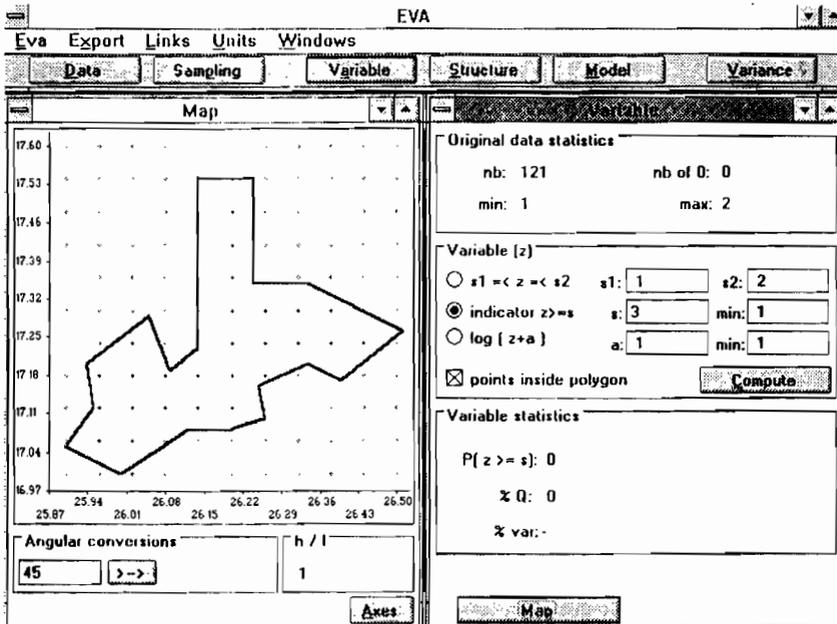
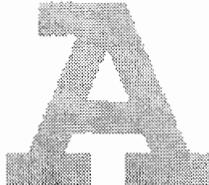


FIGURE 23 - Testing another sampling scheme inside the polygon. The figure shows the window of the Execution bouton Variable. Estimation variance for this new scheme will be computed

Test of sampling scheme when the spatial structure is known

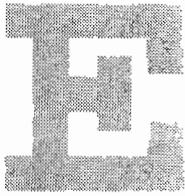


ESTIMATION VARIANCE

Distances will have to be calculated. Coordinates of sample points are expressed in kilometers. Click on the Units function from the menu bar and select the option Coordinates and distances in identical units.

Click on the Variance execution button, choose the linear estimation in the Estimation method window. The Variance with spatial structure window opens. Insert manually the parameters of the isotropic model obtained in the previous study, namely : nugget = $5 \text{ E}4$ + exponential (sill = $3.7 \text{ E}5$; practical range = 0.15). Choose the same 20×20 computation grid than in the previous study. Click directly in the Compute variance button. Considering that this sampling would give us the same mean as previously, namely 963, a relative estimation error of 8.2 % is found.

The regular scheme gives the same accuracy as previously but with ten stations less.



P. Petitgas, A. Prampart

© Orstom, 1995

Contents

I Estimation theory, EVA computations

ESTIMATORS OF THE MEAN IN AN AREA I-2

Linear estimators I-3

Non linear estimators I-4

ESTIMATORS
OF THE ESTIMATION VARIANCE I-5

Random sampling scheme :

classical statistics and geostatistics I-5

Other sampling schemes :

transitive and intrinsic geostatistics I-6

DIFFERENT VARIANCES :

DISPERSION VARIANCE,
VARIANCE OF THE ESTIMATOR,
ESTIMATION VARIANCE I-8

Intrinsic method I-8

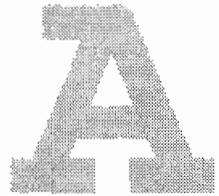
Transitive method :

case of regular grids I-11

EXPERIMENTAL STRUCTURAL FUNCTIONS I-12

Experimental variogram I-12

Experimental transitive covariogram I-14



STRUCTURAL MODELS I-14

Models with a sill, models without a sill I-15

Behaviour at small distances I-16

Nugget I-16

Models I-17

Anisotropies I-19

PROBABILISTIC INTERPRETATION
OF A REGIONALIZATION :
MEAN AND VARIANCE IN AREA,
MEAN AND VARIANCE OF A DISTRIBUTION I-20

ALGORITHMS FOR THE COMPUTATION OF VARIANCES . I-22

Transitive method I-22

Intrinsic method I-23

ESTIMATION VARIANCE WHEN THE MEAN IN AN AREA IS ESTIMATED
BY THE SIMPLE AVERAGE OF CORRELATED SAMPLES I-23

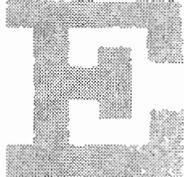
MEAN OTHER THE AREA AND ESTIMATION VARIANCE FOR CORRELATED
DATA UNDER LOGNOMAL HYPOTHESIS I-25

REFERENCES I-27

II Installing, getting started
and principles of use

INSTALLING AND GETTING STARTED II-2

Creating an icon II-2



PRINCIPLES OF USE II-3

IMPORTATION, UPDATING, EXPORTATION II-6

 Importations II-7

 DATA READING BY COLUMN II-7

 LINKS AND UPDATING II-9

 PRINCIPLES OF COPYING CHARACTERS FROM EXCEL II-9

 Exportations : printings and savings II-10

 EXPORT FUNCTION FROM THE MENU BAR II-11

 SCREEN COPY KEY II-12

MENU BAR II-12

 Units II-12

 Windows II-13

DATA FILES SUPPLIED
AS EXAMPLES OF USE II-14

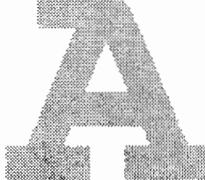
III Example of use : abundance estimation
performed by summing up biomass
along acoustic transects

CONFIGURATION OF THE EXCEL WORKSHEET .. III-3

DATA IMPORTATION AND SELECTION
OF THE GEOSTATISTICAL METHOD III-6

 Data button III-6

 Sampling button III-6



STRUCTURAL ANALYSIS III-8

Variable button III-8

Structure button III-9

Model button III-11

ESTIMATION VARIANCE III-12

RESULTS EXPORTATION III-14

USING THE LINK III-14

IV Example of use :
 estimation of the mean in an area
 from non randomly located samples

CONFIGURATION OF THE EXCEL WORKSHEET IV-3

DATA IMPORTATION AND SELECTION
 OF THE GEOSTATISTICAL METHOD IV-4

STRUCTURAL ANALYSIS IV-7

Variable button IV-7

OPTION POINTS INSIDE POLYGON FROM THE VARIABLE WINDOW IV-8

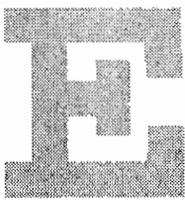
OPTION $s1 \leq z \leq s2$ FROM THE VARIABLE WINDOW IV-9

OPTION INDICATOR $z \geq s$ FROM THE VARIABLE WINDOW IV-9

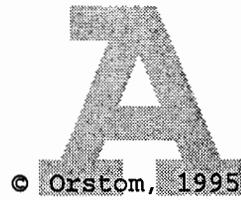
OPTION $Log(z+a)$ FROM THE VARIABLE WINDOW IV-11

MAP WINDOW IV-11

EXAMPLE IV-13



P. Petitgas, A. Prampart



© Orstom, 1995

Structure button IV-18

Model button IV-20

ESTIMATION VARIANCE IV-22

Estimation method window IV-22

Variance with spatial structure window IV-24

SCREEN FIELD DISCRETIZATION OF POLYGON IV-25

COMPUTE VARIANCE BUTTON IV-27

RESULTS EXPORTATION IV-28

OTHER MODEL, ANISOTROPIC MODELLING IV-28

OTHER POLYGON,
OTHER GEOMETRIC HYPOTHESIS IV-32

V Example of use :
test of a sampling scheme
when the spatial structure is known

MAKING A REGULAR SAMPLING GRID
WITH AN EXCEL 4 WORKSHEET V-2

IMPORTATION IN EVA V-3

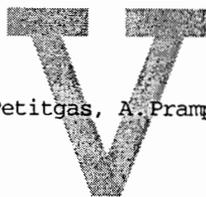
REPRESENTATION OF A SAMPLING GRID V-4

ESTIMATION VARIANCE V-6

A large, bold, stylized letter 'E' composed of a grid of small squares, serving as a logo.

© Orstom, 1995

P. Petitgas, A. Prampart

A large, bold, stylized letter 'V' composed of a grid of small squares, serving as a logo.A large, bold, stylized letter 'A' composed of a grid of small squares, serving as a logo.

Mise en page : Post-Factory (Montpellier)

Photocopies : Inf'arT (Montpellier)

Façonnage des classeurs : Luxord (Lunel)

Duplication des disquettes : TSI (Gagny)

EVA est un logiciel de géostatistique pour IBM-PC, fonctionnant sous Windows, qui fournit les paramètres statistiques de l'estimation d'abondance des stocks marins. Il permet d'effectuer l'analyse structurale et le calcul de la variance d'estimation globale de façon visuelle et interactive : dialogue avec les données, retour en arrière, modification des options, mesure de leurs conséquences. EVA permet de tester différents plans et efforts d'échantillonnages. Le logiciel n'effectue pas le krigeage. Le manuel utilisateur est à la fois un cours succinct sur les principes de géostatistique et un manuel d'utilisation. Trois exemples de traitement type — dont les jeux de données sont fournis — sont détaillés, pas à pas, avec copies d'écran.

EVA is a geostatistics software for IBM-PC working under Windows. It computes the estimation statistics for fish stock assessments. It performs the structural analysis and computes the global estimation variance. It works in a visual and interactive manner: the user dialogues with his data, goes back and forth in his analysis, modifies options, measures their consequences. EVA enables testing of different survey designs and sampling efforts. No kriging is performed. The user manual contains a review of basic geostatistics, the user guide and case studies. Three examples of typical data analysis — for which data are supplied — are detailed step by step with screen copies.

Contact technique / Technical adviser : Pierre Petitgas
Orstom, HEA, 911 av. Agropolis, BP 5045
34032 Montpellier cedex 1.
e-mail : petitgas@orstom.orstom.fr

Géostatistique, échantillonnage, variogramme, variance, abondance
Geostatistics, surveys, variograms, variance, abundance

150 FF t.c.