MPO Pêches de l'Atlantique
Document de recherche 96 / 60

DFO Atlantic Fisheries
Research document 96 / 60

Estimation géostatistique de la production quotidienne d'oeufs chez le maquereau bleu (*Scomber scombrus* L.) du golfe du Saint-Laurent

Par

François Grégoire et Claude Bellehumeur[2]

Division des poissons et des mammifères marins
Ministère des Pêches et des Océans
Institut Maurice-Lamontagne
850, Route de la Mer
Mont-Joli, Québec,
G5H 3Z4

[2] Groupe STOPER, Université de Sherbrooke,
2 500 boul. Université, Sherbrooke
Québec, J1K 2R1

RÉSUMÉ

L'approche géostatistique a été utilisée à titre exploratoire pour calculer la production quotidienne d'oeufs du maquereau du golfe du Saint-Laurent. Les variogrammes des densités d'oeufs pour les directions Nord-Sud et Est-Ouest montrent clairement que les données sont autocorrélées spatialement. Les variogrammes mettent aussi en évidence la présence d'une structure spatiale bien définie et un effet de pépite relativement peu élevé comparativement à la variance des données. La production quotidienne d'oeufs a été évaluée à $2.2561 \bullet 10^{13}$ avec des limites inférieure et supérieure de $1.9438 \bullet 10^{13}$ et $2.5683 \bullet 10^{13}$ respectivement Cette estimation est améliorée par rapport aux estimations provenant des méthodes statistiques classiques.

ABSTRACT

The geostatistical approach was used on an exploratory basis to estimate the daily egg production for the Gulf of St. Lawrence mackerel. The variograms for the eggs densities in directions North-South and East-West clearly show that the data are spatially autocorrelated. The variograms also reveal a well defined spatial structure and a relatively low nugget effect in comparison with the variance of the data. The daily egg production was evaluated at $2.2561 \bullet 10^{13}$ with lower and upper limits of $1.9438 \bullet 10^{13}$ and $2.5683 \bullet 10^{13}$ respectively. This estimation is improved in comparison with the estimations given by the classic statistical methods.

INTRODUCTION

Les populations marines de vertébrés (Petitgas et Poulard 1989; Simard *et al.* 1993) ou d'invertébrés (Buestel *et al.* 1985; Armstrong *et al.* 1989; Simard *et al.* 1992) et les communautés planctoniques (Haury *et al.* 1978; Steele 1978; Mackas *et al.* 1985) ne sont pas distribuées dans l'environnement de façon aléatoire. On les retrouve plutôt agglomérées en taches ou distribuées en gradients selon la structure spatiale en présence. Dans de telles circonstances, la valeur d'une variable à un endroit donné est dépendante de la valeur de cette même variable à un autre endroit. Il y a alors autocorrélation spatiale (Margalef 1979; Isaaks et Srivastava 1989). La plupart des variables environnementales sont autocorrélées et leur description ou l'estimation de leur abondance nécessite une approche statistique différente de l'approche classique dont l'une des prémisses de base est la dispersion homogène et aléatoire de la variable à l'étude. L'approche géostatistique, qui est utilisée depuis peu dans le domaine des pêches, tient compte non seulement de la composante aléatoire mais aussi de la structure d'autocorrélation. La prise en considération de cette structure permet d'envisager un gain de précision sur l'estimation de l'abondance d'une variable. Ce gain sera d'autant plus important si la structure spatiale compte pour une grande part de la variance totale.

Dans le présent document, l'approche géostatistique est utilisée pour la première fois dans le calcul de la production quotidienne d'oeufs du maquereau (*Scomber scombrus* L.) du golfe du Saint-Laurent. Les résultats obtenus sont comparés à ceux provenant de la dernière évaluation d'abondance de cette espèce qui a été effectuée à partir des méthodes statistiques classiques. Dans le cas d'une amélioration de la précision de l'estimation, la géostatistique pourrait remplacer dans l'avenir l'approche classique pour le calcul de la production d'oeufs.

MATÉRIEL ET MÉTHODES

Mission d'évaluation par les oeufs

Les données d'abondance d'oeufs de maquereau pour l'année 1994 proviennent de l'étude de Grégoire *et al.* (1995). Elles ont été prélevées à 65 stations fixes, disposées sur une grille régulière et séparées en moyenne par environ 37 km. Les stations couvrent une région de 400 km dans la direction Est et de 350 km dans la direction Nord. Les oeufs ont été recueillis à l'aide d'un échantillonneur de type Bongo (Posgay et Marak 1980) dont les mailles des filets sont de 333 microns. Les traits, d'une durée minimale de 10 minutes, ont été effectués en dents de scie (Hempel 1973). À chacune des stations, un profil de la température en fonction de la profondeur a été réalisé de façon à calculer la température moyenne de l'eau dans les dix premiers mètres. C'est à l'intérieur de cette couche que se retrouve la plupart des oeufs, et la température de l'eau permet de déterminer leur temps d'incubation. Les différents stades de développement des oeufs ont été identifiés au retour de la mission. Le temps d'incubation et le décompte des oeufs du premier stade de développement ont permis de calculer les densités d'oeufs (nb/m^2) qui ont été pondus au moment de la fraie. La production quotidienne d'oeufs a par la suite été déterminée en utilisant ces valeurs et les équations d'estimation de la moyenne et de la variance associées à un plan d'échantillonnage aléatoire stratifié (Cochran 1977). L'utilisation de ce dernier par rapport à un plan d'échantillonnage simple a déjà été justifiée dans Grégoire (1992).

Variogramme

L'approche géostatistique consiste d'abord au calcul d'un variogramme. Ce dernier décrit les échelles de variations spatiales de la variable d'intérêt. En fait, le variogramme décrit la structure spatiale présente entre les points, i.e. les unités d'échantillonnage, qui sont séparés par différentes distances. Le variogramme s'exprime comme suit:

$$(1) \qquad \gamma^*(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} \left[z(x_i) - z(x_i + h)\right]^2$$

où:

- $\gamma^*(h)$ représente la semi-variation entre les couples de points (l'expression 1 est souvent appelée variogramme dans la littérature traitant de géostatistique même si en fait elle représente une semi-variation et non une variance entre les couples de points (Rossi *et al.* 1992);
- h est un vecteur de distance possédant une intensité et une direction;
- $N(h)$ représente le nombre de couples de points ayant servi au calcul de $\gamma^*(h)$;
- $Z(x_i)$ et $Z(x_i+h)$ sont les valeurs de la variable considérée aux positions (x_i) et $(x_i + h)$, séparées par le vecteur h.

Le calcul du variogramme est répété pour les différentes valeurs de h, ce qui permet de construire le diagramme des valeurs de $\gamma^*(h)$ contre h. Une fonction mathématique est ajustée aux points expérimentaux du variogramme pour calculer la covariance spatiale de toutes les distances géographiques considérées dans la région étudiée. L'ajustement des points expérimentaux n'est possible que si l'on assume certaines hypothèses de stationnarité comme l'homogénéité de la moyenne et de la covariance spatiale. Si la moyenne et la covariance spatiale d'une variable sont constantes dans la région étudiée, la variable est dite stationnaire d'ordre deux. L'hypothèse de stationnarité d'ordre deux implique:

- que l'espérance mathématique de la variable aléatoire $Z(x_i)$ existe et ne dépend pas de la position de l'échantillon:

$$(2) \qquad E\{Z(x_i)\} = m$$

- et que pour chaque paire de variables aléatoires $\{Z(x_i), Z(x_i+h)\}$, la covariance $C(h)$ existe et dépend seulement de la distance h et de son orientation:

$$C(h) = E\{Z(x_i+h) \bullet Z(x_i)\} - m^2$$

(3) $$C^*(h) = \frac{1}{N(h)} \sum_{i=1}^{N(h)} [Z(x_i) \bullet Z(x_i + h)] - m^2$$

où m est la moyenne.

Krigeage

La fonction de covariance décrite par l'équation 3 est utilisée pour procéder à une estimation optimale de la variable d'intérêt par la méthode du krigeage. L'estimation locale de la variable $Z(X_0)$ au point X_0 est effectuée en utilisant les points $Z(X_i)$ situés dans le voisinage de X_0:

(4) $$Z^*(X_0) = \sum w_i Z(X_i)$$

où les w_i sont des poids dont la somme est de un afin d'assurer une estimation non-biaisée. Les poids w_i sont estimés de façon à minimiser la variance des erreurs d'estimation ($Var[Z^*(X_0) - Z(X_0)]$ = minimum) sous la contrainte de non-biais. La solution minimisant la variance d'estimation du modèle s'exprime comme suit:

(5) $$\mathbf{C} \bullet \mathbf{W} = \mathbf{D}$$

$$\begin{bmatrix} C_{11} .. C_{1n} & 1 \\ . \quad . \quad . & \\ C_{n1} .. C_{nn} & 1 \\ 1 \; .. \; 1 & 0 \end{bmatrix} \bullet \begin{bmatrix} W_i \\ . \\ W_n \\ \mu \end{bmatrix} = \begin{bmatrix} C_{10} \\ . \\ C_{n0} \\ 1 \end{bmatrix}$$

où **C** est une matrice de covariance entre les n points participant à l'estimation et déduite du variogramme, **W** est le vecteur des poids recherchés et **D** le vecteur contenant les covariances entre les n points participant à l'estimation et le point à estimer (situé en X_0), et μ le paramètre de Lagrange qui est introduit pour minimiser la variance d'estimation sous la contrainte $\sum w_i = 1$. Les poids w_i sont trouvés en résolvant le système linéaire suivant:

(6) $$\mathbf{W} = \mathbf{C}^{-1} \bullet \mathbf{D}$$

Lorsque la superficie de chaque surface est la même, l'abondance totale des oeufs est donnée par l'équation suivante:

(7) $$\tau^* = \Sigma_k S_i \bullet 1/k \bullet \Sigma_k x_i$$

où *k* est le nombre de surfaces estimées, S_i la superficie de chaque surface, et X_i la moyenne obtenue par krigeage pour la i-ème surface.

À chaque estimation de surface est attachée une variance de krigeage qui s'exprime comme suit:

(8) $$\sigma_E^2 = 2\sum_{i=1}^{n} w_i \gamma(x_i, x_0) - \sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j \gamma(x_i, x_j)$$

Dans le cas présent, cette variance est difficilement utilisable pour procéder à une estimation globale étant donné que les variances obtenues ne sont pas indépendantes les unes des autres; les mêmes unités d'échantillonnage servant à l'estimation de plusieurs surfaces contiguës. Étant donné que l'échantillonnage est effectué de façon relativement uniforme, l'équation suggérée par Simard *et al.* (1992) est plutôt utilisée:

(9) $$\sigma^2_{xi} = \gamma(s,s) / n$$

où σ^2_{xi} est la variance d'estimation de la moyenne de la surface considérée, γ (s,s) la variance intra-surface d'une surface de dimension *s*, étant donné le modèle de variogramme retenu, et *n* est le nombre d'unités d'échantillonnage contenues dans la surface. La valeur de γ (s,s) peut être évaluée à l'aide des abaques donnés dans Journel et Huijbregts (1978), où l'on peut procéder à une évaluation numérique en calculant la valeur moyenne du variogramme, évaluée entre toutes les paires de points possibles retrouvées à l'intérieur d'une surface. La variance de l'estimation totale se calcule finalement par:

(10) $$\mathrm{Var}(\tau^*) = k \bullet S_i^2 \bullet \sigma^2_{xi}$$

RÉSULTATS

Les variogrammes de la densité d'oeufs de maquereau pour les directions Nord-Sud (NO°E) et Est-Ouest (N90°E) sont présentés à la Figure 1. Ces variogrammes montrent clairement que les données

sont autocorrélées spatialement. Ils mettent en évidence une structure spatiale bien définie et un effet de pépite (bruit de fond) relativement peu élevé comparativement à la variance des données (15 000 versus 288 429), laissant escompter un gain de précision appréciable en regard des méthodes statistiques classiques pour l'estimation de la moyenne globale du nombre d'oeufs. On remarque que le palier du variogramme de la direction Est-Ouest est plus élevé que celui de la direction Nord-Sud, indiquant une plus grande variabilité dans cette direction.

Les variations spatiales sont modélisées par un effet de pépite et un variogramme sphérique représenté par l'expression suivante (Journel et Huijbregts 1978):

$$\gamma(h) = C_0 + C_1 [\, 1.5 \bullet h/a - 0.5\, (h/a)^3] \qquad \text{pour } 0 < h \leq a$$

$$\gamma(h) = C_0 + C_1 \qquad \text{pour } h > a$$

$$\gamma(h) = 0 \qquad \text{pour } h = 0$$

où C_0 est l'effet de pépite, C_1 est la composante de variance structurée spatialement et *a* est la portée du modèle. Le Tableau 1 donne les paramètres des modèles de variogramme sphérique pour les directions Nord-Sud et Est-Ouest.

Les deux modèles de variogrammes ont servi à déterminer les valeurs w_i. La moyenne du nombre d'oeufs de maquereau a été estimée pour des surfaces de 35 km x 35 km lors du krigeage en discrétisant chaque surface par une grille régulière de deux points par deux points, c'est-à-dire que dans chaque surface, quatre points sont estimés par krigeage, et on attribue la moyenne de ces quatre points à la surface entière.

La variance intra-surface est évaluée à 31 520 (Annexe 1). La surface considérée dans cette estimation globale est légèrement plus élevée que celle donnée dans Grégoire *et al.* (1995) (58 • 35 km x 35 km = 71 050 km^2 versus 69 450 km^2). La raison de cette sur-évaluation est que la région étudiée est découpée en surfaces carrées régulières de 35 km x 35 km, ce qui ne permet pas une définition du contour des côtes. La production quotidienne totale est estimée à $2.2561 \bullet 10^{13}$ avec des limites inférieure et supérieure de $1.9438 \bullet 10^{13}$ et $2.5683 \bullet 10^{13}$ respectivement (Tableau 2). La variance de l'estimation de la production d'oeufs par géostatistique est inférieure aux variances reliées aux méthodes statistiques classiques (Tableaux 2 et 3).

CONCLUSIONS

Les variogrammes de la densité d'oeufs de maquereau ont permis de détecter des structures de corrélation spatiale. Ces structures sont d'une envergure de 180 km à 200 km et correspondent à la région située entre les Îles-de-la-Madeleine et le Nouveau-Brunswick. C'est dans cette région qu'on

observe toujours les plus importantes concentrations d'oeufs (Grégoire *et al.* 1995.). Un fort gradient Est-Ouest de la température de l'eau est aussi observé dans tout le Golfe.

L'estimation de la producion d'oeufs de maquereau pour la mission de 1994 a été améliorée en tenant compte de la structure spatiale observée dans le Golfe. Compte tenu de ces résultats, l'approche géostatistique devrait aussi être appliquée pour les missions antérieures.

RÉFÉRENCES

Armstrong, M., D. Renard., et P. Berthou. 1989. Applying geostatistics to the estimation of a population of bivalves. ICES C.M. / K37:22p.

Buestel, D., J. C. Dao., et F. Gohin. 1985. Estimation d'un stock naturel de coquilles Saint Jacques par une méthode combinant les dragages et la plongée. Traitement des résultats par une méthode géostatistique. ICEM. C. M. 1985/K:18.

Cochran, W. 1977. Sampling techniques. John Wiley and Sons, New York, NY. 413p.

Grégoire, F. 1992. Revue de la stratégie d'échantillonnage des oeufs utilisée lors des croisières d'évaluation de la biomasse reproductrice du maquereau bleu (*Scomber scombrus* L.) du golfe du Saint-Laurent. CSCPCA Document de recherche 92/52. 16p.

Grégoire, F., D. D'Amours, C. Lévesque., et D. Thibeault. 1995. Estimation de la biomasse reproductrice du stock de maquereau (*Scomber scombrus* L.) du golfe du Saint-Laurent. MPO Document de recherche 95/118. 81p.

Haury, L. R., J. A. McGowan., et P. H. Wiebe. 1978. Patterns and processes in the time-space scales of plankton distributions, p. 277-327. *In* J. H. Steele [ed.] Spatial pattern in plankton communities. Plenum Press, New York, NY.

Hempel, G. 1973. Fish egg and larval surveys. FAO Fisheries Technical Paper No. 122. 82p.

Isaaks, E. H., et R. M. Srivastava. 1989. Applied geostatistics. Oxford University Press, New York, NY. 561 p.

Journel, A. G., et Ch. J. Huijbregts. 1978. Mining geostatistics. Academic Press, New York, NY. 600p.

Mackas, D. L., K. L. Denman., et M. K. Abbott. 1985. Plankton patchiness: biology in the physical vernacular. Bull. Mar. Sci. 37: 652-674.

Margalef, R. 1979. The organization of space. Oikos 33: 152-159.

Petitgas, P., et J.-C. Poulard. 1989. Applying stationary geostatistics to fisheries: a study on hake in the bay of Biscay. ICES Demersal Fish. Comm. C. M. / G62: 21p.

Posgay, J. A., et R. R. Marak. 1980. The MARMAP Bongo zooplankton samplers. J. Northwest Atl. Fish. Sci. 1: 91-99.

Rossi, R. E., D. J. Mulla, A. G. Journel., et E. H. Franz. 1992. Geostatistical tools for modeling and interpreting ecological spatial dependence. Ecological Monographs. 62 (2): 277-314.

Simard, Y., P. Legendre, G. Lavoie., et D. Marcotte. 1992. Mapping, estimating biomass, and optimizing sampling programs for spatially autocorrelated data: case study of the northern shrimp (*Pandalus borealis*). Can. J. Fish. Aquat. Sci. 49: 32-45.

Simard, Y., D. Marcotte., et G. Bourgault. 1993. Exploration of geostatistical methods for mapping and estimating acoustic biomass of pelagic fish in the Gulf of St. Lawrence: size of echo-integration unit and auxiliary environmental variables. Aquat. Living Resour. 6: 185-199.

Steele, J. H. 1978. Spatial pattern in plankton communities. Plenum Press, New York, NY. 470 p.

Tableau 1. Paramètres des modèles de variogrammes de la densité d'oeufs de maquereau pour les directions Nord-Sud et Est-Ouest.

PARAMÈTRES	DIRECTION	
	Nord-Sud	Est-Ouest
C_0	15 000	15 000
Modèle	Sphérique	Sphérique
C_1	90 000	146 000
a_1	180 km	200 km

Tableau 2. Paramètres de l'estimation géostatistique du nombre d'oeufs de maquereau dans le golfe du Saint-Laurent.

k^{i}	58
x^{ii}	317.5358 oeufs/m^2
$\gamma(s, s)^{iii}$	31 520
τ^*	$2.2561 \bullet 10^{13}$
var $(\tau^*)^{iv}$	$2.5390 \bullet 10^{24}$
τ^* inf.^{v}	$1.9438 \bullet 10^{13}$
τ^* sup.^{v}	$2.5683 \bullet 10^{13}$
P^{vi}	13.84%

i: nombre de surfaces évaluées par krigeage;
ii: moyenne des 58 surfaces krigées;
iii: valeur moyenne du variogramme dans une surface de 35 km x 35 km;
iv: variance de l'estimation globale considérant le modèle de variogramme (éq. 10)
v: inf. et sup. désigne les limites inférieure et supérieure d'un intervalle de confiance à 95%;
vi: précision de l'estimation.

Tableau 3. Calcul de l'abondance totale des oeufs de maquereau (production journalière) par les relations statistiques classiques.

PARAMÈTRES	RELATIONS		
	Aléatoire Simple	Aléatoire Simple[i]	Aléatoire Stratifiée[ii]
n	65	63	65
x	260.7047		
s^2	288 429.2226	95 426.0993	
var (x)	4 437.3727	1514.7000	
S[iii]	$6.9450 \bullet 10^{10}\ m^2$		
τ^*	$1.8106 \bullet 10^{13}$		$2.2606 \bullet 10^{13}$
var (τ^*)	$2.1403 \bullet 10^{25}$	$7.3059 \bullet 10^{24}$	$3.2390 \bullet 10^{25}$
τ^* inf.[iv]	$0.9038 \bullet 10^{13}$	$1.2808 \bullet 10^{13}$	$1.1451 \bullet 10^{13}$
τ^* sup.[iv]	$2.7173 \bullet 10^{13}$	$2.3404 \bullet 10^{13}$	$3.3761 \bullet 10^{13}$
P[v]	50.08%	29.26%	49.34%

i: estimation de la variance où deux valeurs fortes sont enlevées;
ii: les résultats du plan aléatoire stratifié sont tirés de Grégoire *et al.* (1995);
iii: superficie de la région étudiée (Grégoire *et al.* 1995);
iv: inf. et sup. désigne les limites inférieure et supérieure d'un intervalle de confiance à 95%;
v: précision de l'estimation.

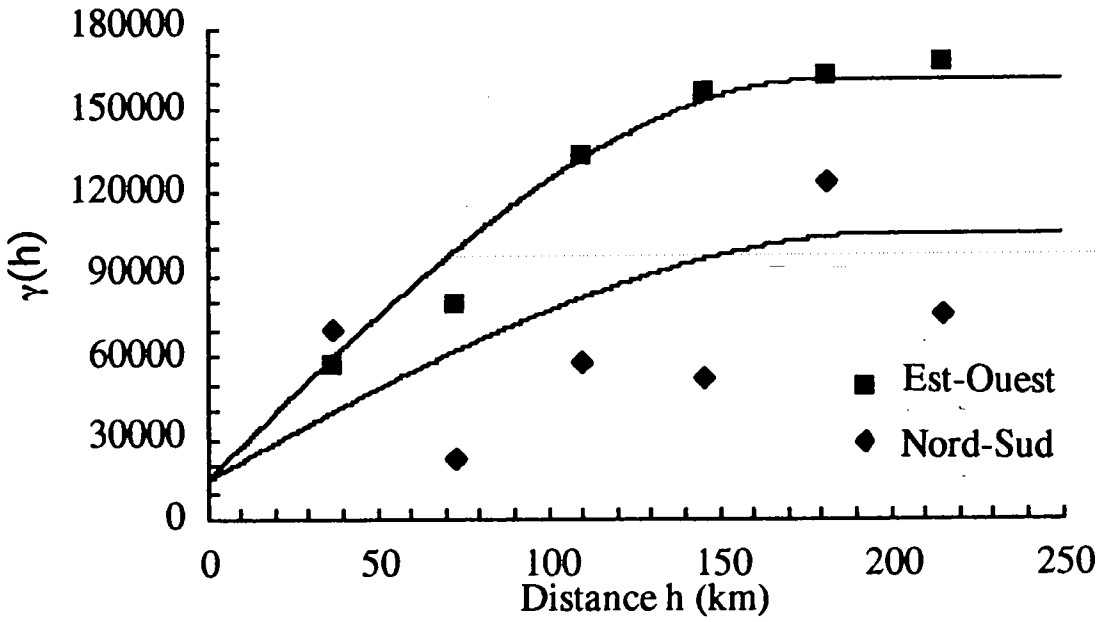

Figure 1. Variogrammes de la densité d'oeufs de maquereau pour les directions Est-Ouest et Nord-Sud.

Annexe 1. Calcul de la variance intra-surface selon un modèle de variogramme donné.

Le calcul de la variance d'estimation d'une moyenne requiert le calcul de la variance de la variable évaluée sur la surface pour laquelle est calculée cette moyenne. Cette variance est donnée par la valeur moyenne du variogramme, évaluée entre toutes les paires de points possibles retrouvées à l'intérieur d'une surface. On peut procéder au calcul numérique de cette variance en discrétisant la surface en un nombre fini de points, ou en utilisant des abaques qui donnent les paramètres appropriés.

Les abaques permettant de calculer la valeur moyenne du variogramme dans une surface (γ (s, s)) pour les modèles de variogrammes sphérique et exponentiel sont tirés de Journel et Huijbregts (1978). Ces abaques donnent une fonction de correction pour la partie structurée du variogramme en fonction des dimensions de la surface (L, 1) et de la portée du variogramme (a). Pour les directions NO°E et N90°E, les rapports L/a sont donnés par:

NO°E : L/a = 35/180 = 0.1944
N90°E : L/a = 35/200 = 0.175

Selon l'abaque du modèle sphérique, la valeur de F (L/a, 1/a) = F(0.1944, 0.175) = 0.14

Les paliers des modèles des directions NO°E et N90°E diffèrent légèrement. Une modélisation rigoureuse de cette anisotropie zonale requerrait l'expression de ces modèles selon une somme de modèles sphériques et linéaires valides pour toutes les directions, ce qui peut s'avérer assez fastidieux. Une approximation valable de la variance de cette échelle spatiale consiste à utiliser le palier moyen de ces deux paliers ((146 000 + 90 000) / 2). La valeur de (γ (s, s)) d'une surface de 35 km x 35 km est donnée par:

$$
\begin{aligned}
(\gamma(s, s)) &= C_0 + F(L/a, 1/a) \bullet C_1 \\
&= 15\,000 + 0.14 \bullet (146\,000 + 90\,000) / 2 \\
&= 31\,520
\end{aligned}
$$