

M. ROUX

Estimation des paléoclimats d'après l'écologie des foraminifères

Les cahiers de l'analyse des données, tome 4, n° 1 (1979),
p. 61-79

http://www.numdam.org/item?id=CAD_1979__4_1_61_0

© Les cahiers de l'analyse des données, Dunod, 1979, tous droits réservés.

L'accès aux archives de la revue « Les cahiers de l'analyse des données » implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
<http://www.numdam.org/>

ESTIMATION DES PALÉOCLIMATS D'APRÈS L'ÉCOLOGIE DES FORAMINIFÈRES [PALÉOCLIMATS]

*Critique et compléments à propos d'un travail
de J. Imbrie et N. G. Kipp
par M. Roux (1)*

Dans son principe, l'étude publiée en 1971 par J. Imbrie et N.G. Kipp (appelés I & K dans la suite) est assez simple : à partir de la composition faunistique actuelle (ou considérée comme telle) en foraminifères (*) des couches supérieures des sédiments marins et de la valeur connue des paramètres de l'environnement aux points de prélèvement, il s'agit d'estimer les valeurs de ces mêmes paramètres pour des époques reculées, à partir de la composition faunistique des sédiments correspondants. C'est typiquement un problème de régression et c'est bien aini qu'il a été traité par I & K.

Comme il est d'usage I & K basent leur régression, non sur les données brutes, mais sur les facteurs issus d'une analyse en composantes principales (en abrégé ACP) portant sur un tableau résultant d'une transformation du tableau initial des pourcentages d'espèces par échantillon. Nous considérerons ces aménagements des données au § 2 après avoir décrit au § 1 les données elles-mêmes. Nous proposerons ensuite (§ 3) notre propre méthodologie qui ne diffère de la précédente que par la substitution de l'analyse factorielle des correspondances (en abrégé AFC) sur les données brutes à l'ACP sur données transformées. On verra au § 4 les conséquences de cette substitution, tant du point de vue théorique que pratique.

Chemin faisant, l'AFC a clairement montré que les données actuelles présentent une variabilité écologique beaucoup plus grande que celle des paléochantillons à estimer : ceux-ci paraissant en fait s'insérer dans le domaine des échantillons tropicaux.

Si l'on apprécie comme il sied d'après ces seuls échantillons la qualité de la régression, les coefficients de corrélation estimés s'abaissent gravement. On a donc refait (§ 5) toute la série des calculs de régression eux-mêmes en se limitant au sous-ensemble des échantillons actuels le mieux assorti à l'ensemble des paléochantillons à estimer : d'où finalement, surtout par l'AFC, des coefficients de corrélation satisfaisants. La conclusion (§ 6), fait le bilan de ces travaux.

(*) *Etres unicellulaires pourvus d'une coquille calcaire percée d'un ou de plusieurs orifices (orifice = foramen), et intérieure à la cellule (endosquelette) ; les foraminifères ont d'innombrables espèces tant fossiles qu'actuelles, distinguées par la forme extrêmement variée de leurs coquilles.*

(1) *Chargé de recherches C.N.R.S.
Laboratoire de statistique de l'université Pierre et Marie Curie*

1 Présentation des données

1.1 Les tableaux et leur contenu

Les données publiées par I & K sont constituées de deux parties, la première concernant 61 observations de sédiments récents, provenant de stations assez bien réparties sur tout l'océan Atlantique et débordant même sur l'océan Indien, la deuxième portant sur 110 observations issues d'un seul carottage situé en mer des Caraïbes, au sud-ouest de l'île de Saint-Domingue (17°00 N et 74°24 W).

	30 espèces	3 var.
61 observations récentes	X	Y
110 observations anciennes	Z	

La première partie des données est constituée de deux blocs : un tableau faunistique (X) recensant les abondances de 30 espèces de foraminifères (dont la liste figure en annexe 1) et un tableau d'environnement (Y) comportant 3 variables : moyenne des températures d'été (en abrégé TE) moyenne des températures d'hiver (TH) et salinité moyenne (SA).

Ces variables sont mesurées sur les eaux de surface.

La seconde partie des données ne comprend qu'un tableau faunistique (Z) ; puisqu'il s'agit précisément d'estimer les valeurs des trois variables de l'environnement pour les échantillons anciens.

On conviendra, suivant l'usage, que les lignes des tableaux représentent les observations ou prélèvements, tandis que les colonnes représentent soit les espèces soit les variables de l'environnement. Les sous-tableaux faunistiques contiennent des valeurs en pourcentages calculés échantillon par échantillon. Ainsi à l'intersection de la ligne *i* et de la colonne *j* on a, pour l'espèce *j*, la valeur de son abondance relative dans l'échantillon *i*, la somme des termes de la ligne *i* valant 100.

Pour la partie environnement, les nombres portés dans le tableau Y sont les valeurs brutes des variables, en degrés Celsius pour les températures, en g/l pour les salinités.

1.2 Le choix des variables et des observations

Sur les 30 variables faunistiques on doit remarquer, tout d'abord, que l'une d'elles (n° 30) n'est pas une véritable espèce mais est constituée par les comptages cumulés de tous les individus qui n'ont pu être affectés à aucune des 29 espèces retenues. A noter également que deux espèces (n° 7 et n° 20) n'apparaissent jamais dans les sédiments anciens ; inversement trois espèces (n°s 27, 28 et 29) ne sont pas représentées dans les sédiments récents.

Les prélèvements sont constitués par une tranche de carotte ayant 1 cm d'épaisseur. En se basant sur diverses considérations I & K évaluent le taux de dépôt moyen des sédiments, à la station n° 61 où a eu lieu le carottage en profondeur, à 2,33 cm pour 1000ans ; compte-tenu du brassage dû aux animaux benthiques, brassage dont l'effet peut se

faire sentir jusqu'à 5 cm de profondeur, ils estiment que leurs échantillons couvrent une période variant de 2000 à 4000 ans ; les échantillons récents sont donc des témoins d'un climat entièrement post-glaciaire. Les prélèvements anciens ont été faits sur une profondeur de 11 m à raison d'un échantillon tous les 10 cm, ce qui couvre au total une période d'environ 450.000 ans. On doit encore noter que les températures comme les salinités actuelles résultent d'interpolations à partir de mesures faites en des points voisins des stations, et que les moyennes ne portent que sur quelques dizaines d'années tout au plus ce qui nuit évidemment à la précision des estimations.

2 Les calculs effectués par I & K

D'un point de vue pratique les calculs de I & K comportent trois étapes ; la première (programme CABFAC) réalise la transformation des données actuelles suivie de l'ACP ; la deuxième (REGRES) effectue le calcul des coefficients de régression applicables aux coordonnées, ou saturations ("*factor loadings*"), des prélèvements sur les axes factoriels retenus ; enfin la troisième étape (THREAD) calcule les coordonnées des prélèvements anciens sur ces mêmes axes, c'est-à-dire qu'elle évalue ce que nous appelons "les facteurs des éléments supplémentaires", puis applique à ces facteurs les coefficients de régression, trouvés à l'étape deux, pour obtenir une estimation des paramètres recherchés. Cette dernière étape étant sans difficulté nous n'examinerons en détail que les deux premières.

2.1 L'analyse en composantes principales

C'est de loin, la première étape qui pose le plus de problèmes, or c'est sur elle que repose le reste des calculs et en particulier la qualité des estimations finales. Il semble que, guidés par leur expérience et celle des chercheurs antérieurs (Blackman et Somayajulu, 1966, Imbrie et Van Andel, 1964), I & K aient rapidement fixé leur choix sur l'option "Q-mode", choix qui conditionne alors les transformations pré-alables des données. Rappelons, en effet que l'ACP privilégie nécessairement l'un des deux nuages de points en présence : le mode Q donne une bonne représentation du nuage $N(J)$ des espèces variables, tandis que le mode R, le plus couramment utilisé, donne une approximation précise du nuage $N(I)$ des prélèvements (individus). Dans les deux cas on peut obtenir une représentation approchée de l'autre nuage.

Si on appelle, suivant une notation devenue classique (Benzécri, 1973, TII A, n° 2, § 2.3), $F_\alpha(i)$ et $G_\alpha(j)$ la coordonnée de l'échantillon i et celle de l'espèce j , respectivement, sur l'axe factoriel de rang α , alors les ensembles $F_\alpha(I)$ et $G_\alpha(J)$, constitués par ces deux groupes de coordonnées, forment des vecteurs à n et p composantes, respectivement, n désignant le nombre d'échantillons et p le nombre d'espèces :

$$F_\alpha(I) = \{F_\alpha(i) \mid i \in I\} \quad ; \quad \text{Card } I = n \quad ;$$

$$G_\alpha(J) = \{G_\alpha(j) \mid j \in J\} \quad ; \quad \text{Card } J = p \quad .$$

Il se trouve que, dans l'analyse en mode Q les différents vecteurs $G_\alpha(J)$ sont orthogonaux deux à deux (au sens usuel) :

$$\alpha \neq \alpha' \Rightarrow \sum_j G_\alpha(j) G_{\alpha'}(j) = 0.$$

Par contre les vecteurs $F_\alpha(I)$ ne le sont pas, en général, et l'on verra (§ 2.2) que c'est un obstacle important à la réussite de la régression.

Mais revenons au tableau X, à n lignes et p colonnes (n = 61 et p = 30 pour I & K), soumis à l'analyse. Ce n'est pas le tableau brut des comptages, ni même celui des pourcentages par échantillon. En effet l'ACP est particulièrement sensible aux variations de taille, avantageant les espèces, ou les échantillons, pour lesquels les différences d'abondances sont particulièrement prononcées. C'est probablement pour pallier cet inconvénient que I et K effectuent consécutivement quatre transformations préliminaires :

1°) Calcul des pourcentages par échantillon (si ce n'est déjà fait)

2°) Elimination des espèces peu abondantes

3°) Uniformisation des variations par espèce

4°) Normalisation des échantillons

La première transformation uniformise les données ligne par ligne, chaque ligne obtenant ainsi un total égal à 100. Les espèces ne dépassant pas 2% dans au moins un des prélèvements superficiels et un des prélèvements de profondeur sont éliminés ensuite, en principe pour éviter les erreurs d'échantillonnage, mais I & K avouent immédiatement avoir mené la procédure jusqu'au bout de nombreuses fois avec des seuils différents avant de retenir celui de 2%...

La troisième transformation, appelée par I & K "*per cent range data*", consiste à appliquer, colonne par colonne, une "règle de trois" de façon à ce que le minimum de la colonne soit zéro et le maximum 100. Depuis ce travail I & K ont abandonné cette transformation, mais dans ce cas la transformation suivante (quatrième de notre liste) ne se justifie plus guère. Celle-ci constitue, en effet, une nouvelle uniformisation ligne par ligne, obtenue par multiplication, de façon à ce que la somme des carrés des termes de la ligne soit égale à l'unité ; cette modification des données réitère à peu près la première (pourcentages) mais n'en a pas le sens physique évident, qui est de rapporter tous les échantillons au même nombre d'individus recensés.

De même le calcul de la moyenne des abondances par échantillon et des écarts à cette moyenne n'auraient pas grand sens, c'est pourquoi I & K ne le font pas. Autrement dit leur ACP est une analyse Q-mode utilisant la matrice des produits scalaires entre observations (i. e. échantillons) normées mais non centrées. Les transformations préliminaires comme les options de calcul que l'on vient de voir font que les résultats sont assez difficiles à interpréter ce qui amène I & K à appliquer aux facteurs une rotation, suivant le critère "varimax", qui, tout en conservant globalement la quantité d'information apportée par les m premiers axes, fait passer ces axes "au voisinage" des variables initiales c'est-à-dire (dans leur cas) des échantillons ; mais cette rotation ne change rien à la non-orthogonalité, ou obliquité, des axes en question.

Ces manipulations, notamment l'élimination des espèces peu abondantes et la polarisation sur l'un des deux ensembles, ici les espèces, polarisation inhérente à la méthode statistique, *sont des choix arbitraires* qui peuvent sans doute être justifiés par la quantité des résultats obtenus, mais qui risquent d'être très contingents. En d'autres termes un changement dans l'échantillonnage des prélèvements superficiels, ou dans les subdivisions infraspécifiques, pourraient impliquer une révision de ces choix ce qui diminue considérablement la portée de la méthode, et sa reproductibilité avec des données d'autres domaines comme, par exemple, les pollens, les diatomées, les cocolithes, etc...

2.2 La régression multiple

Dans le problème actuel on a trois variables dites "à expliquer", savoir température d'été, température d'hiver et salinité, et quelques variables dites "explicatives" que sont les facteurs issus de l'ACP. On utilise généralement, comme critère d'évaluation de la régression, le coefficient de corrélation multiple entre la variable à expliquer et les variables explicatives, qui n'est autre que le coefficient de corrélation usuel entre la variable (à expliquer) réellement observée et la variable estimée par la régression. C'est également le critère utilisé par I & K, mais on sait que ce critère augmente automatiquement avec le nombre de variables explicatives, *même si celles-ci n'ont aucun rapport avec la variable à expliquer*. De plus, si les variables explicatives sont très corrélées entre elles, les coefficients de régression obtenus sont très instables et peuvent donc subir des variations importantes en cas de fluctuations, même légères, dans les données : par exemple adjonction d'une ou deux observations supplémentaires ou suppression d'une espèce. (Aussi parle-t-on à ce propos de : *régression illusoire* ; cf dans les *Cahiers Régr. Géom.*, Vol III n° 2 ; et la série *Méthodes de la régression*, I, II, III par P. Cazes ; où sont exposés les principes de protection de la régression).

Pour toutes ces raisons il convient de limiter au strict minimum le nombre des variables explicatives et de les choisir, si possible, non corrélées : c'est ce qui justifie l'emploi préalable de l'analyse factorielle, qui condense au mieux l'information initiale sur un petit nombre de facteurs non corrélés, ou orthogonaux. Précisons ces notions à l'aide de quelques formules. Si l'on appelle $T(i)$ la température relevée pour la i -ème observation, $E_j(i)$ le pourcentage de l'espèce j dans l'échantillon i , et, comme précédemment, $F_\alpha(i)$ la coordonnée sur l'axe factoriel α de l'observation i , alors les F_α jouent le rôle de nouvelles variables (explicatives) à la place des E_j . Ainsi, au lieu d'une formule liant directement la température et les p espèces retenues :

$$T(i) = a_1 E_1(i) + a_2 E_2(i) + \dots + a_p E_p(i) + a_0 + d_i \quad (1)$$

on cherche une formule analogue mais utilisant, par exemple, les quatre premiers facteurs de l'ACP :

$$T(i) = b_1 F_1(i) + b_2 F_2(i) + b_3 F_3(i) + b_4 F_4(i) + b_0 + e_i \quad (2)$$

Dans ces formules a_0 et b_0 désignent des constantes, e_i et d_i sont les résidus de la régression, résidus inévitables car les formules ne peuvent prétendre à un ajustement parfait aux données.

Ici nous devons exprimer quelques réserves sur la méthode employée par I & K. En effet, comme nous l'avons remarqué au § 2.1, avec l'analyse Q-mode ce sont les facteurs G_α sur J (ensemble des espèces) qui sont orthogonaux et non les facteurs F_α sur I (ensemble des échantillons). Or la formule (2) ci-dessus montre bien que ce sont ces derniers que l'on utilise dans la régression. En toute logique on devrait donc faire une analyse R-mode, ou bien une analyse factorielle des correspondances (cf § 3). Une autre critique que nous adressons à I & K concerne le nombre de facteurs. Après n'avoir retenu, comme liés au problème, que quatre facteurs, baptisés A, B, C, et D ils utilisent deux formes de régression qu'ils appellent linéaire et non-linéaire. La première n'utilise que les quatre facteurs en question, tandis que la seconde utilise, outre ces quatre facteurs, leurs combinaisons quadratiques à savoir : A^2 , B^2 , C^2 , D^2 , AB, AC, AD, BC, BD et CD ce qui porte à 14, au total, le nombre de variables explicatives !

Cette deuxième méthode est donc à proscrire, d'autant plus que, comme on le verra plus loin (§ 4.2), l'amélioration qu'elle apporte au coefficient de corrélation multiple est minime.

3 Amélioration de la méthode I & K : usage de l'AFC

Pour les raisons que l'on vient de voir nous suivrons maintenant les mêmes étapes que I & K mais en remplaçant l'ACP par l'AFC du tableau brut des pourcentages.

3.1 L'analyse des correspondances

L'ensemble des données disponibles, c'est-à-dire les trois tableaux du § 1.1 a été soumis à l'AFC, mais les 110 observations anciennes comme les 3 variables à expliquer ont été considérées comme éléments supplémentaires passifs ; c'est-à-dire que seul le tableau X a servi à la détermination des facteurs. Ce tableau, appelé tableau de base, contient donc tous les éléments actifs, les éléments supplémentaires sont simplement projetés sur les axes factoriels une fois ceux-ci déterminés par les éléments actifs.

Cette méthode a deux avantages : elle permet de voir si les observations de base recouvrent convenablement les observations anciennes ; elle autorise d'autre part une première appréhension, plutôt qualitative, des relations entre variables à expliquer et variables explicatives, comme on va le voir maintenant. L'examen du plan déterminé par les axes factoriels 1 et 2 pour les observations (Figure 1) montre l'existence d'un effet Guttman bien marqué : les stations se rangent suivant une courbe parabolique caractéristique, révélatrice d'un fort gradient au sein des données. D'un point de vue pratique, cela signifie que, si l'on réordonne dans le tableau X les observations ainsi que les variables suivant leurs coordonnées croissantes sur le premier axe, on obtient un tableau contenant de fortes valeurs au voisinage de la diagonale principale et des valeurs faibles ou nulles ailleurs (cf Benzécri, 1073, TII A n° 2, § 3.2).

Il est facile de voir que ce facteur largement prédominant (taux d'inertie expliquée 42%) n'est autre que le gradient de température : les observations qui proviennent des latitudes élevées sont du côté négatif du premier axe (avec de très fortes contributions) tandis que celles qui sont situées entre les tropiques sont du côté positif de l'axe. Le graphique relatif aux variables (Figure 2) est moins explicite. En effet, en vue de l'analyse, le tableau Y a été recodé en remplaçant chacune des trois variables à expliquer par quatre variables qualitatives, ou modalités, correspondant respectivement aux valeurs très faibles, faibles, fortes et très fortes de la variable originelle (Sur le découpage des variables continues en classes, cf *Cahiers* Vol II pp 115 sqq). Une observation reçoit la valeur 1 pour la modalité à laquelle elle appartient et possède des zéros pour toutes les autres modalités (on dit que les variables à expliquer sont mises sous forme disjonctive complète). Ainsi la variable "température d'été" est remplacée par quatre modalités, notées TE1, TE2, TE3 et TE4 ; une observation ayant une température d'été forte sera représentée par la suite des valeurs 0, 0, 1, 0 pour ces quatre modalités. On obtient de cette façon le tableau Y'.

Comme il a été dit ces variables supplémentaires n'ont pas été prises en compte pour la détermination des axes, mais leur position sur ces axes a été calculée. Grâce à l'artifice du découpage en classes (tableau Y'), chacune des trois variables est ainsi représentée par quatre points sur les graphiques, et l'on voit que les points relatifs aux températures suivent étroitement la courbe parabolique déjà examinée, les modalités étant rangées sur cette courbe dans leur ordre naturel. Il n'en est pas de même pour les modalités de la salinité, dont on voit sur le tableau des contributions qu'elles sont mieux

représentées sur le quatrième facteur (à l'exception de la première modalité - salinité très faible - bien liée, comme il est normal, aux faibles températures, donc située du côté de l'axe 1 négatif).

Mais revenons au graphique des observations, toujours pour les axes 1 et 2. Ce graphique appelle une remarque importante : on y voit tous les éléments supplémentaires (échantillons anciens) très resserrés, sans exception, dans la zone du graphique correspondant aux températures les plus élevées. Ce graphique nous permet donc d'apprécier d'un seul coup d'oeil la validité de l'ensemble de base (échantillons actuels) en vue de l'estimation des températures et salinités. On voit, en effet, que les observations de base, non seulement recouvrent bien les observations supplémentaires (ce qui est satisfaisant), mais qu'elles les débordent très largement, (ce qui est fâcheux, comme on le verra au § 5) : le nuage des observations de base est beaucoup plus dispersé que celui des observations supplémentaires. On peut se demander aussi si ces observations supplémentaires ne sont pas mieux représentées sur d'autres axes que les deux premiers, mais cette question est sans conséquence pour l'estimation des températures dans la mesure où l'on a reconnu que celles-ci étaient schématisées convenablement par ces deux premiers axes. Ce qui est confirmé par les calculs de corrélations que nous examinons maintenant.

3.2 Corrélations avec les variables de l'environnement

Quoique les résultats du paragraphe précédent soient assez clairs il nous a paru utile de les préciser numériquement à l'aide des coefficients de corrélation entre les facteurs et les trois variables à expliquer? Aux fins de comparaison, on a juxtaposé ci-dessous les tableaux relatifs à l'AFC, à l'ACP R-mode et à l'ACP Q-mode (avec rotations) soient les facteurs utilisés par I & K.

	AFC			ACP Q-mode			ACP R-mode		
	TE	TH	SA	TE	TH	SA	TE	TH	SA
F ₁	.8964	.8866	.7897	.8156	.8788	.5039	.7722	.8232	.4994
F ₂	-.3582	-.4224	-.0725	-.2533	-.1306	-.4540	.4013	.2744	.6263
F ₃	.1069	.0258	.1912	-.7965	-.7798	-.7496	.3142	.3499	.0647
F ₄	-.0148	.0917	-.2994	.5119	.5349	.2953	-.0956	-.0974	-.1907
F ₅	-.0298	-.0411	.1139	.3205	.2464	.5274	.0166	-.0300	-.1504

Tableau 1 : Corrélations entre facteurs et variables à expliquer

Les valeurs relativement élevées des corrélations entre les variables à expliquer et les facteurs issus de l'ACP Q-mode ne doivent pas faire illusion car ceux-ci sont souvent assez corrélés entre eux ; ainsi F₃ et F₄ sont corrélés à -.5064 et .6205, respectivement avec le facteur F₁, autrement dit l'information qu'ils apportent a déjà été largement apportée par le facteur F₁ (Voir au § 4.2 les coefficients de corrélation multiple).

4 Complément aux calculs de I & K pour l'estimation des paramètres de l'environnement

4.1 Fonctions de transfert explicites

Un autre avantage important de l'AFC est qu'elle permet, dans le cas particulier des données en pourcentages, d'exprimer les températures

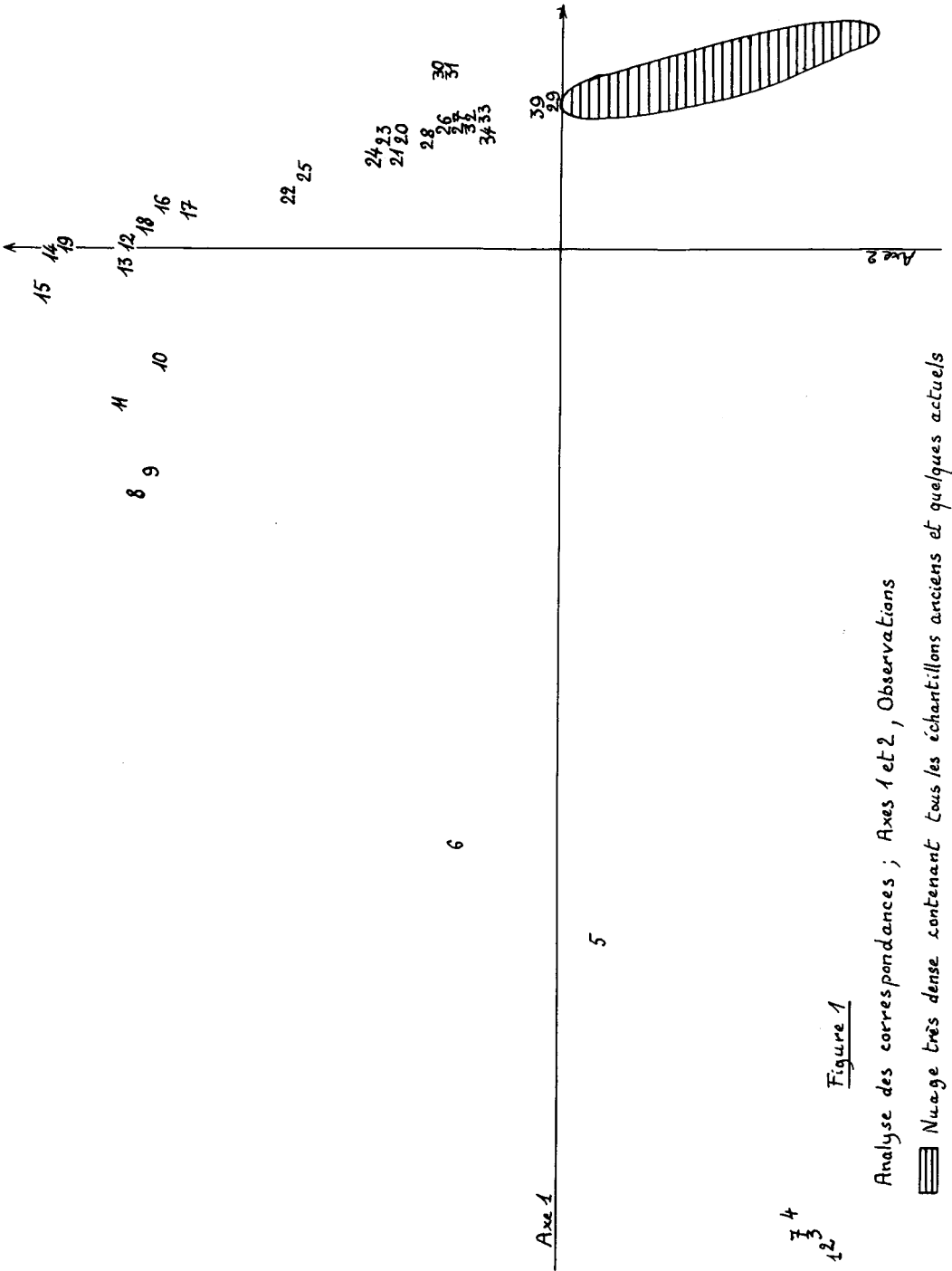


Figure 1

Analyse des correspondances ; Axes 1 et 2, Observations

☐ Nuage très dense contenant tous les échantillons anciens et quelques actuels

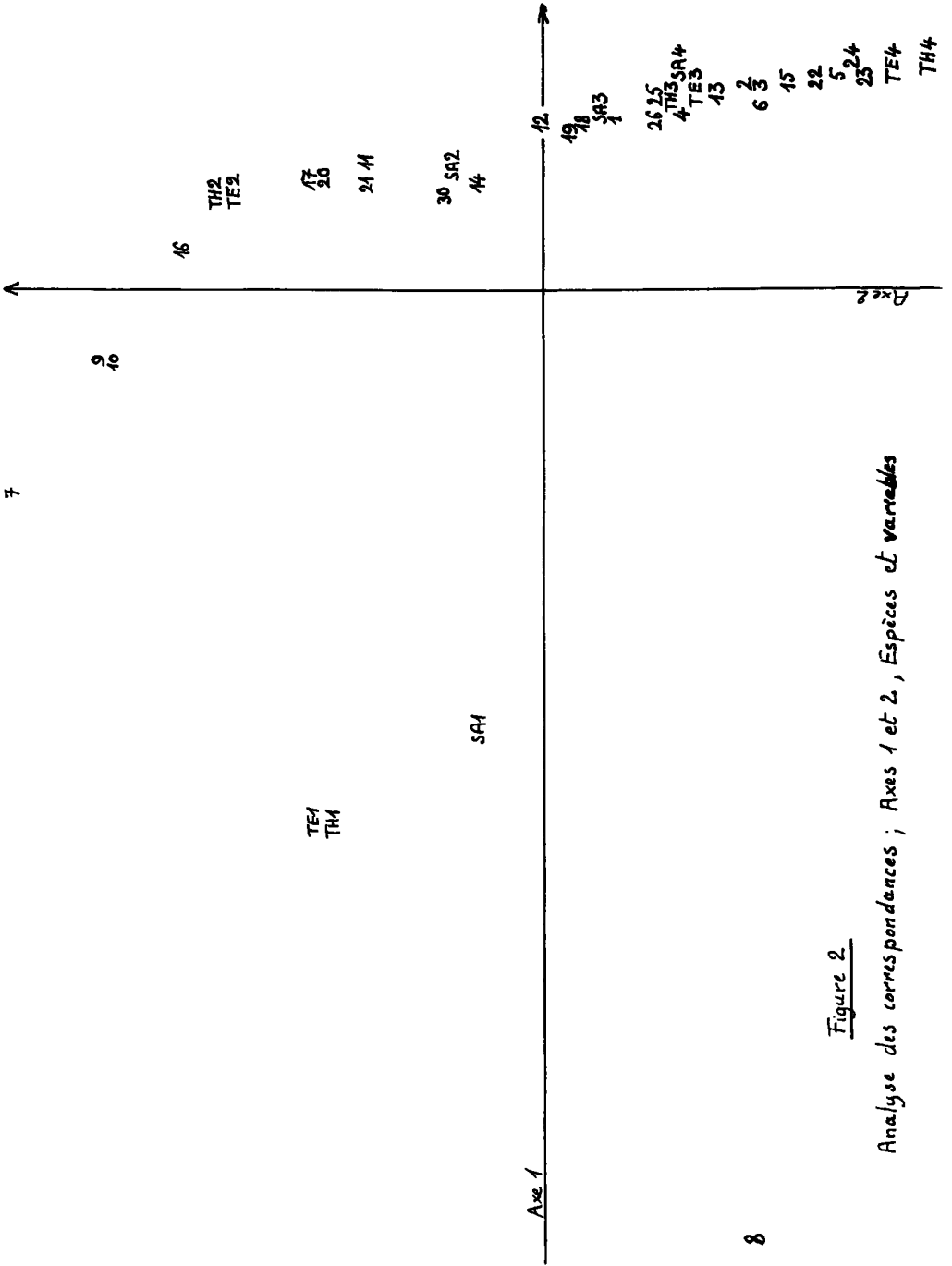


Figure 2

Analyse des cornes pondances ; Axes 1 et 2 , Espèces et variables

ou la salinité directement en fonction des espèces. On se souvient, en effet, que la méthode de I & K vise à exprimer les paramètres étudiés, non en fonction des données initiales, mais en fonction des coordonnées des échantillons sur les axes factoriels issus de l'analyse des prélèvements superficiels. Le retour aux données est pour eux impossible du fait des manipulations préliminaires qu'ils font sur elles, en particulier la normalisation ligne à ligne.

Ce retour est possible avec l'AFC grâce à la formule dite "de transition" qui relie les deux ensembles en présence, celui des espèces et celui des échantillons (cf Benzécri, 1973, TI A n° 2, § 3 et TII A n° 2).

$$F_{\alpha}(i) = (1/\sqrt{\lambda_{\alpha}}) \sum_j^P = 1 (E_j(i)/E_{*}(i)) G_{\alpha}(j)$$

où p désigne, comme précédemment le nombre total d'espèces, λ_{α} désigne la valeur propre ou inertie de l'axe α , $E_j(i)$ est, comme ci-dessus, le pourcentage de l'espèce j dans l'échantillon i et $E_{*}(i)$ est le poids, c'est-à-dire le total, de la ligne i ... or ce total est constant et égal à 100 puisqu'il s'agit de pourcentages! On peut donc mettre le coefficient $1/100$ devant le signe de sommation et supprimer ce poids :

$$F_{\alpha}(i) = (1/(100\sqrt{\lambda_{\alpha}})) \sum_j^P = 1 E_j(i) G_{\alpha}(j)$$

et en reportant cette expression dans la formule (2) de la régression sur facteurs (§ 2.2) on a pour la température de l'échantillon i :

$$T(i) = (b_1/(100\sqrt{\lambda_1})) \sum_j E_j(i) G_1(j) + \dots + \\ (b_4/(100\sqrt{\lambda_4})) \sum_j E_j(i) G_4(j) + b_0 + e_i$$

et, en regroupant les termes homologues de chaque sommation :

$$T(i) = (1/100) \sum_j E_j(i) [b_1 G_1(j)/\sqrt{\lambda_1} + \dots + b_4 G_4(j)/\sqrt{\lambda_4}] + b_0 + e_i$$

Si l'on appelle a_j la quantité entre crochets divisée par 100, alors la température s'exprime simplement au moyen d'une formule de type (1) (§ 2.2) :

$$T(i) = \sum_j a_j E_j(i) + b_0 + e_i \quad (3)$$

La constante b_0 n'est autre que la moyenne des températures calculée sur l'ensemble de base parce que les facteurs de l'AFC sont de moyenne nulle et que les poids $E_{*}(i)$ sont égaux. Le seul calcul restant à faire est donc celui des coefficients a_j des fonctions de transfert explicites :

$$a_j = \sum_{\alpha}^k = 1 b_{\alpha} G_{\alpha}(j)/\sqrt{\lambda_{\alpha}}$$

où k désigne le nombre de facteurs retenus (pour I & K, $k = 4$) et b_{α} le coefficient de la régression relatif au facteur α . Nous publions en annexe les valeurs de ces coefficients, obtenues en ne retenant que trois facteurs pour les températures et quatre facteurs pour la salinité. On voit immédiatement l'avantage de la formule (3) : étant donné un prélèvement quelconque dont les pourcentages d'espèces sont connus, on peut en calculer rapidement la température à la main ou à l'aide d'une petite machine à calculer de bureau ...

4.2 Qualité des fonctions de transfert

Avec l'évaluation de la qualité des résultats nous abordons une question particulièrement délicate. Le coefficient de corrélation multiple, couramment utilisé comme nous l'avons déjà dit, est un indicateur relatif à l'ajustement des *données de base*, c'est-à-dire les échantillons actuels ; or nous avons vu (§ 3.1) que ces données de base présentent une variabilité beaucoup plus grande que celle des éléments supplémentaires ce qui conduit à surestimer le coefficient de corrélation multiple et, par là même, nous donne une fausse assurance quand à la qualité des estimations obtenues pour les observations anciennes.

Il nous a paru plus raisonnable de calculer ce coefficient de corrélation multiple, non sur l'ensemble des échantillons de base, mais sur un sous-ensemble d'échantillons les plus "voisins" des échantillons anciens.

Nous nous sommes alors demandé s'il ne valait pas mieux refaire entièrement les calculs, en ne prenant pour données de base que ce sous-ensemble des plus proches voisins (appelé par la suite "ensemble PPV"). Nous avons donc été amené à conduire parallèlement deux séries de calculs l'une en prenant pour données de base l'ensemble des échantillons actuels, l'autre en prenant l'ensemble PPV.

Nous donnons d'abord ci-dessous (tableau 2) les valeurs du coefficient de corrélation multiple obtenues en utilisant tous les échantillons actuels et nous exposons au paragraphe suivant (§ 5) la méthode utilisée pour choisir les données PPV puis les résultats fournis par l'analyse de ces données.

AFC			ACP R - mode		
TE	TH	SA	TE	TH	SA
1 .8964	1 .8866	1 .7897	1 .7722	1 .8232	2 .6263
2 .9653	4 .9821	4 .8445	2 .8702	3 .8991	1 .8010
3 .9712	4 .9863	3 .8659	3 .9252	2 .9400	4 .8234
6 .9736	5 .9872	5 .8734	7 .9326	7 .9560	6 .8383
5 .9740	3 .9875	2 .8763	6 .9380	4 .9610	5 .8517

ACP Q-mode (I & K)		
TE	TH	SA
1 .8283	1 .8888	4 .5263
2 .9365	2 .9466	1 .6519
3 .9407	4 .9513	2 .8222
4 .9407	3 .9539	3 .8420

Tableau 2 : Coefficients de corrélation multiple quand on utilise successivement 1, 2, 3, etc... facteurs. Les colonnes impaires sont les numéros des facteurs les meilleurs à chaque pas.

Dans le tableau 2 on indique les valeurs du coefficient de corrélation multiple croissantes en fonction du nombre de facteurs retenus dans l'analyse factorielle ; on notera que ces facteurs ne sont pas forcément dans leur ordre naturel, celà tient à ce que l'on utilise de préférence les facteurs les plus liés aux variables à expliquer (cf Tableau du § 3.2).

Eu égard au peu d'amélioration obtenu par les facteurs ultérieurs

nous avons décidé de ne conserver que les trois meilleurs facteurs pour l'estimation des températures et les quatre meilleurs pour la salinité. Notons encore qu'avec 14 variables de la régression non-linéaire I & K obtiennent les coefficients de corrélation multiple suivants :

TE .9717, TH .9890, SA .8921

ce qui constitue une bien faible amélioration compte-tenu du nombre des variables explicatives introduites.

5 Calculs sur l'ensemble des plus proches voisins

5.1 Cloison des plus proches voisins (PPV)

A l'aide du programme POUBEL (cf Lebeaux, 1977) nous avons dressé pour chaque observation supplémentaire (paléoéchantillon) la liste des cinq échantillons actuels qui en sont les plus proches voisins au sens de la distance euclidienne usuelle (i.e. racine carrée de la somme des carrés des différences des coordonnées), mais dans l'espace des 7 premiers facteurs issus de l'AFC ; nous avons ensuite retenu tous les échantillons apparaissant au moins une fois comme voisin. Nous avons complété cette liste avec quelques échantillons qui nous sont apparus très proches des échantillons déjà retenus, soit sur les graphiques d'AFC, soit sur un arbre hiérarchique issu d'une classification ascendante basée sur le moment d'ordre deux d'une partition, classification calculée également sur les facteurs de l'AFC (cf Benzécri, 1973). Ces considérations nous ont permis d'arrêter la liste des PPV à 29 observations actuelles dont on trouvera les énumérations dans le tableau 4 en annexe (cette liste pouvant servir aux spécialistes qui reprendraient les recherches de I & K).

5.2 Les résultats des calculs de corrélations

Comme précédemment on n'a retenu dans chaque estimation que les quatre facteurs explicatifs ayant les plus fortes corrélations avec les variables à expliquer. Le tableau 2', ci-joint, donne les coefficients de corrélation calculés en distinguant trois cas suivant les échantillons utilisés.

1° cas (29 ; 29) : i.e. 29 observations de base (PPV) pour l'analyse factorielle et le calcul des coefficients de régression ; les coefficients de corrélation étant calculés sur les mêmes observations.

2° cas (61 ; 29) : 61 observations de base pour l'A.F. et les calculs de régression ; les coefficients de corrélation étant calculés sur les seules observations PPV (29 observations).

3° cas (61 ; 61) : 61 observations de base pour l'A.F. et la régression ; corrélations sur ces 61 observations (résultats extraits du tableau 2 ci-dessus).

Dans les deux cas on compare les résultats obtenus par notre stratégie, savoir AFC suivie d'une régression linéaire, et ceux obtenus par la méthode I & K, sans utiliser les combinaisons quadratiques des facteurs.

	29 observations de base (PPV); corrélations calculées sur ces 29 observations (PPV)			61 observations de base corrélations calculées sur 29 observations (PPV)			61 observations de base corrélations calculées sur ces 61 observations		
	TE	TH	SA	TE	TH	SA	TE	TH	SA
Notre méthode	.7924	.9136	.7674	.7337	.8567	.5634	.9740	.9875	.8763
Méthode I & K	.6560	.9081	.6480	.5161	.7827	.6406	.9407	.9539	.8420

Tableau 2': Comparaison des corrélations obtenues selon les échantillons et les méthodes

Pour interpréter les coefficients donnés dans le tableau 2', il faut se souvenir du but de l'étude : estimer d'après la paléoécologie des foraminifères les valeurs correspondantes des variables d'environnement. Donc quelques réserves que l'on puisse faire d'autre part quant à l'usage même des coefficients de corrélation (cf *infra*) il est clair en tout cas que la qualité de l'estimation obtenue pour le paléoclimat ne peut être appréciée que d'après la qualité des estimations contemporaines relatives aux 29 échantillons PPV. Les calculs de corrélation du 3° cas (61:61) effectués sur tous les 61 relevés contemporains comme l'ont fait I & K, sont donc illusoire ; et l'efficacité des formules de régression fondées sur 61 relevés de base doit être appréciée suivant le 2° cas (61:29) : et on constate alors combien les calculs du 3° cas surestiment cette efficacité ! Il est toutefois réconfortant de constater qu'en fondant la régression (comme nous le recommandons) sur les 29 relevés PPV on atteint finalement une efficacité meilleure (1° cas) : qui est véritablement celle que permet la méthode de I & K à condition de la corriger à la fois quant à l'analyse factorielle et à l'échantillonnage (*).

C'est donc ainsi que nous avons procédé pour calculer les fonctions de transfert explicites figurant au tableau 3 ainsi que pour les diverses estimations (tableaux 4 et 5).

Il faut cependant considérer les valeurs du tableau 2' avec prudence car le coefficient de corrélation multiple est un indicateur de variations concomitantes, mais il n'est pas sensible à l'amplitude de ces variations, ce qui veut dire que les courbes d'évolution des températures que l'on obtient sont sans doute valables quant au sens des variations mais que les valeurs absolues peuvent être assez éloignées de la réalité. D'ailleurs, si l'on examine la table 13 de I & K ou le présent tableau 4 (en annexe), en se limitant aux observations actuelles de climat tropical (n°s 30 à 61) on constate (colonne des résidus) des erreurs pouvant aller jusqu'à deux degrés... En somme un coefficient de corrélation multiple de valeur élevée n'est pas une condition suffisante pour avoir de bonnes estimations, mais c'en est cependant une condition indispensable.

5.3 Valeurs des variables estimées : Les valeurs trouvées pour les échantillons anciens sont listées dans le tableau 5 et illustrées par la figure 3 qui est une représentation de la première moitié de ces observations pour les températures. On constate, du point de vue des sens de variation, un excellent accord entre les estimations de I & K et les nôtres ; en revanche on peut observer des écarts notables entre les valeurs absolues de ces estimations, les nôtres étant généralement plus élevées. On remarque aussi que l'amplitude des oscillations est généralement plus grande chez I & K ; cela peut être causé par l'utilisation abusive des combinaisons quadratiques de facteurs qui, selon-nous, ne font qu'ajouter du "bruit" à l'information utile. On voit enfin que les oscillations des températures d'été sont beaucoup plus faibles que celles des températures d'hiver ; et il est d'ailleurs remarquable que ces oscillations soient concomitantes bien que les deux courbes soient calculées indépendamment l'une de l'autre.

5.4 Suggestions d'analyses complémentaires : Dans la présente étude, l'analyse factorielle a été faite indépendamment du problème de régression qu'on avait en vue : c'est-à-dire que les facteurs ont été définis d'après le seul tableau X des relevés de foraminifères : et il est apparu que ces facteurs, constituant un bon résumé des relevés, permettaient une régression satisfaisante des variables climatiques. Il est toutefois souvent nécessaire, dans des études analogues d'orienter la définition même de ces combinaisons des variables explicatives primaires (ici des abondances) que sont ces facteurs, d'après les variables à expliquer.

La méthode est exposée en détail par P. Cazes (Cahiers Vol III n° 4 pp 385 sqq) : bornons-nous ici à décrire ce qu'il faudrait pour estimer

(*) Voir note p. 76 après la conclusion.

e.g. la salinité. On construit un tableau F x S (30 x 4) : F = ensemble des 30 espèces de foraminifères ; S = ensemble des 4 modalités SA1 à SA4 de la variable de salinité (ici avec 29 relevés effectifs, il semble inutile de distinguer un plus grand nombre de modalités : mais 5 est possible ; en tout cas ces modalités doivent être définies non d'après l'histogramme des 61 relevés, mais d'après celui des 29 P.P.V.) ; la colonne SAx est constituée en faisant la somme des relevés (PPV) rentrant dans la classe de salinité x. A ce tableau les relevés, contemporains ou anciens peuvent être adjoints en éléments supplémentaires ; on peut alors soit effectuer une régression linéaire d'après les facteurs (calculés pour ces él. suppl.) ; soit (si cela apparaît indispensable) estimer la salinité afférente à chaque relevé comme une moyenne des salinités de ses plus proches voisins (recherchés dans l'espace des facteurs issus de la nouvelle analyse, parmi les 29 PPV)

6 Conclusion

Rappelons les principaux points acquis au cours de cette étude : l'analyse factorielle des correspondances, méthode à la fois Q - mode et R - mode, remplace avantageusement l'analyse en composantes principales puisqu'elle permet de travailler sur l'ensemble des espèces au complet et sur les données en pourcentages sans autre transformation, en donnant des facteurs mieux corrélés avec les variables de l'environnement. Elle permet également, d'obtenir des formules liant explicitement ces variables aux espèces.

L'examen des plans factoriels issus de l'AFC donne de plus des indications très utiles sur la validité de l'ensemble des données de base et permet d'extraire un sous-ensemble mieux assorti aux paléochantillons à estimer ; d'où une amélioration de la régression et une appréciation moins illusoire de la qualité de celle-ci.

Enfin il convient de limiter le nombre des facteurs entrant en jeu pour le calcul des coefficients des fonctions de transfert, en particulier la régression non-linéaire, ou quadratique, semble ici à proscrire. Les valeurs élevées du coefficient de corrélation multiple donnent une certaine sécurité quant aux sens des variations des variables estimées ; mais leurs valeurs absolues sont à considérer avec prudence.

Bien entendu ces conclusions demandent à être confirmées par le traitement d'autres données de foraminifères, nous pensons en particulier aux nouveaux échantillons de base publiés par N.G. Kipp elle-même (1976) ; il convient aussi d'analyser les données de pollens, diatomées, radiolaires, etc...

Note annoncée page 73

Dans le tableau 2' et dans l'analyse prenant les 29 observations PPV comme observations de base la "remontée" spectaculaire du coefficient de corrélation relatif à la salinité peut s'expliquer par une constatation intéressante que nous n'avons pu faire en plaçant, dans cette analyse relative aux eaux tropicales (comme précédemment dans l'analyse sur 61 échantillons), les modalités des variables écologiques (i.e. températures et salinité) en éléments supplémentaires : la salinité, comme il est bien connu, croît avec la température lorsqu'on passe des eaux polaires aux eaux tempérées et au contraire, coorélée négativement avec la température dans le domaine des eaux tropicales. Ce dernier fait n'avait pas, croyons-nous, été remarqué jusqu'ici.

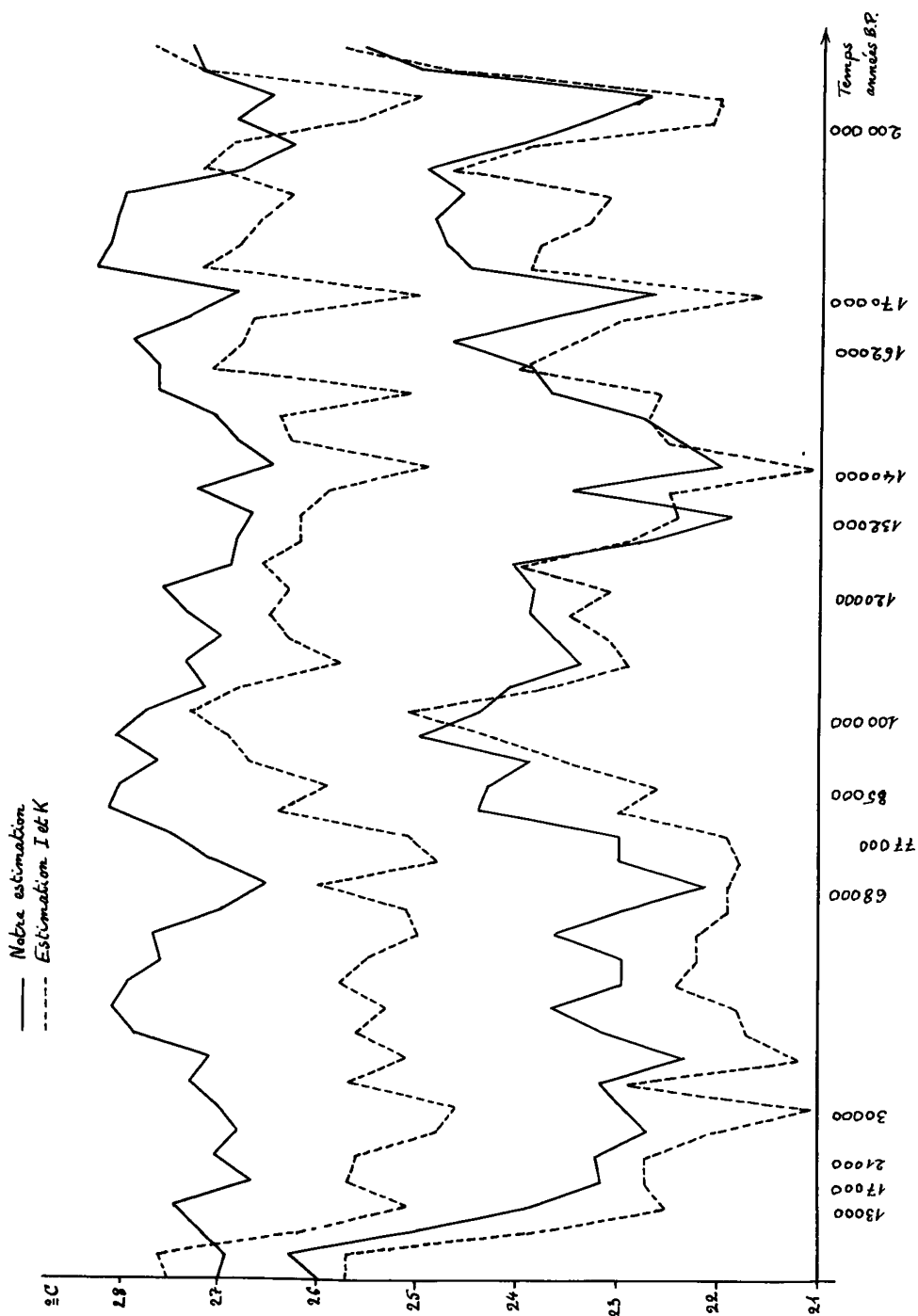
B I B L I O G R A P H I E

- Benzécri et col., 1973 :- L'analyse des données. TI La classification automatique, TII : l'analyse des correspondances. 1248 p, Dunod, Paris.
- Blackman, A. and B.L.K. Somayalu, 1966 :
Pacific pleistocene core : faunal analyses and geochronology. Science, 154, p 886.
- Imbrie, J. and N.G. Kipp, 1971 :
A new micropaleontological method for quantitative paleoclimatology : application to a late pleistocene caribbean core. In : "The late cenozoic glacial ages", K. Turekian ed., pp 71-181, Yale University Press.
- Imbrie, J. and T.H. Van Andel, 1964 :
Vector analysis of heavy mineral data. Bull. Geol. Soc. Am., 75, p 1131.
- Kipp, N.G., 1976 :
New transfer function for estimating post sea-surface conditions from sea-bed distribution of planktonic foraminiferal assemblages in the North-Atlantic. Geol. Soc. Am., Memoir 145. pp 3-41.
- Lebeaux, M.O., 1976 :
Notice sur l'utilisation du programme POUBEL. Cahiers de l'Analyse des données, Vol II, n° 4, pp 467-481.
- Benzécri, J.P., 1978 :
Problèmes statistiques et méthodes géométriques. Cahiers de l'Analyse des données, Vol III, n° 2, pp 131-146.
- Cazes, P., 1978 :
Méthodes de régression I - La régression sous contraintes. Cahiers de l'Analyse des données, Vol III, n° 2, pp 147-165.
- Cazes, P., 1978 :
Méthodes de régression II - Critères bayésiens. Cahiers de l'Analyse des données, Vol III, n° 3, pp 257-268.
- Cazes, P., 1978 :
Méthodes de régression III. L'Analyse des données. Cahiers de l'Analyse des données, Vol III, n° 4, pp 385-391.

A N N E X E

Tableau 3 : Fonctions de transfert explicites

	TE	TH	SA
1 <i>Orbulina universa</i>	- 3.4	- 2.2	- 0.9
2 <i>Globigerinoides conglobatus</i>	- 6.4	-16.0	6.0
3 <i>G. ruber</i>	1.4	1.3	- 0.2
4 <i>G. tenellus</i>	- 0.6	- 6.0	7.0
5 <i>G. sacculifer</i>	- 1.3	7.4	- 0.1
6 <i>Globigcrina rubescens</i>	5.0	- 3.9	1.2
7 <i>G. quinqueloba</i>	0.0	0.0	0.0
8 <i>G. pachyderma (senestre)</i>	8.3	- 3.1	- 3.8
9 <i>G. pachyderma (dextre)</i>	- 5.3	- 9.5	1.2
10 <i>G. bulloides</i>	0.9	-12.8	- 4.2
11 <i>G. falconensis</i>	- 6.0	- 7.8	2.9
12 <i>G. calida</i>	1.1	- 7.2	- 0.6
13 <i>Globigerinella aequilateralis</i>	0.3	- 1.0	1.1
14 <i>Globigerinita glutinata</i>	0.8	0.0	1.8
15 <i>Globoquadrina dutertrei</i>	9.6	6.5	-10.5
16 <i>Globorotalia inflata</i>	-12.0	-20.1	0.2
17 <i>G. truncatulinoides (senestre)</i>	-11.9	- 9.9	7.5
18 <i>G. truncatulinoides (dextre)</i>	- 0.3	-11.3	1.7
19 <i>G. crassaformis</i>	0.7	- 9.9	- 3.5
20 <i>G. hirsuta</i>	-13.3	-31.9	1.7
21 <i>G. scitula</i>	- 2.0	- 6.6	- 2.3
22 <i>G. menardii</i>	- 2.5	3.7	- 0.9
23 <i>Pulleniatina obliquiloculata</i>	4.6	4.9	- 1.6
24 <i>Candeina nitida</i>	- 2.9	- 1.1	4.1
25 <i>Sphaeroidinella dehiscens</i>	0.6	13.4	- 0.3
26 <i>Globigerina digitata</i>	- 2.5	-10.0	7.6
27 <i>Globoquadrina hexagona</i>	0.0	0.0	0.0
28 <i>G. conglomerata</i>	0.0	0.0	0.0
29 <i>Hastigerina</i>	0.0	0.0	0.0
30 Autres espèces	2.3	- 4.3	- 2.6
Constante à ajouter	21.4	16.9	35.7



ESTIMATION DE LA VARIABLE EH

ESTIMATION DE LA VARIABLE TH

ESTIMATION DE LA VARIABLE TE

	VALEUR OBSERVEE	VALEUR ESTIMEE	PESIDU	VALEUR OBSERVEE	VALEUR ESTIMEE	RESIDU	VALEUR OBSERVEE	VALEUR ESTIMEE	RESIDU
29	23.000	23.607	- 607	29	19.000	- 255	29	37.000	- 032
30	24.000	24.743	- 743	30	19.000	- 324	30	36.500	- 222
31	25.000	25.479	- 479	31	18.500	20.127	31	36.200	55.703
35	25.000	24.509	491	35	20.000	18.735	35	36.500	009
36	26.000	25.521	479	36	20.500	20.441	36	37.000	36.734
37	26.000	25.464	536	37	20.500	20.594	37	37.000	37.035
38	24.500	27.040	-2.540	38	22.000	23.453	38	36.700	782
40	26.200	25.396	804	40	21.000	20.130	40	37.200	36.810
41	25.000	24.974	026	41	22.500	22.924	41	37.000	36.430
42	26.500	26.637	- 137	42	21.500	23.173	42	36.000	35.933
43	26.200	26.933	- 733	43	22.000	23.973	43	35.200	56.263
44	26.000	26.805	- 805	44	22.500	22.647	44	37.200	36.668
45	26.000	26.988	- 988	45	23.000	24.233	45	35.700	35.761
46	27.000	26.345	655	46	22.500	21.950	46	35.200	35.345
47	27.000	27.292	- 292	47	23.500	23.717	47	35.000	35.752
48	27.500	26.865	834	48	24.000	25.173	48	35.700	35.948
49	27.000	27.155	- 155	49	24.500	24.464	49	36.500	36.330
50	27.000	27.078	- 778	50	25.000	25.477	50	35.700	36.065
51	27.000	27.153	- 153	51	25.000	24.460	51	37.000	36.226
52	27.000	26.631	369	52	25.500	24.971	52	35.700	36.272
53	27.000	26.864	136	53	26.000	25.952	53	36.000	35.973
54	29.000	27.363	1.637	54	24.500	24.103	54	36.200	35.659
55	28.500	27.048	1.452	55	25.000	22.424	55	36.500	37.043
56	27.500	27.803	- 303	56	26.000	24.505	56	35.000	35.210
57	27.500	27.557	- 557	57	26.000	25.092	57	36.500	35.823
58	27.500	27.656	- 156	58	26.000	26.076	58	35.000	35.108
59	27.000	26.926	074	59	26.500	25.926	59	35.200	35.943
60	27.000	27.037	- 037	60	26.500	25.867	60	35.700	35.731
61	28.000	26.980	1.020	61	26.500	26.027	61	36.000	35.758

Tableau 4 - Estimations pour les observations de base (P.P.V.)

Tableau 5 - Observations anciennes : valeurs estimées

	TE	TH	SA		TE	TH	SA
001	27.0	26.0	35.8	056	25.8	20.5	36.3
002	26.9	26.3	35.6	057	26.8	23.0	36.3
003	27.2	25.0	35.8	058	26.6	23.1	35.9
004	27.5	23.9	35.6	059	27.0	22.8	36.0
005	26.7	23.2	36.1	060	27.0	23.0	36.1
006	27.0	23.2	35.7	061	26.8	23.3	36.1
007	26.8	22.7	36.1	062	27.5	23.4	35.6
008	27.0	23.0	35.9	063	26.7	22.4	35.8
009	27.3	23.2	35.6	064	27.4	23.3	35.7
010	27.1	22.4	35.1	065	27.2	24.3	35.4
011	27.9	23.1	35.1	066	27.1	24.0	35.8
012	28.1	23.7	34.7	067	26.5	22.2	35.9
013	27.9	23.0	34.8	068	26.4	24.1	34.9
014	27.6	23.0	34.9	069	28.2	22.8	34.7
015	27.7	23.7	35.2	070	27.3	22.1	36.6
016	27.0	23.0	35.8	071	26.7	23.5	35.9
017	26.5	22.1	36.1	072	27.6	23.4	35.6
018	27.1	23.0	36.2	073	27.5	21.6	35.9
019	27.5	23.0	36.0	074	27.9	23.1	35.0
020	28.2	24.4	34.5	075	27.0	24.0	35.4
021	28.0	24.3	35.0	076	27.5	23.9	35.6
022	27.6	23.9	35.4	077	26.8	23.8	35.9
023	28.1	25.0	34.9	078	27.5	23.4	35.7
024	27.8	24.4	35.2	079	27.3	24.5	34.9
025	27.2	24.1	35.5	080	27.1	24.9	35.4
026	27.4	23.4	35.6	081	26.8	24.7	35.6
027	27.0	23.6	35.7	082	27.2	24.0	35.6
028	27.4	23.9	35.0	083	26.6	22.3	36.1
029	27.6	23.9	34.8	084	26.0	22.2	35.8
030	26.9	24.1	35.9	085	26.2	22.9	35.7
031	26.9	22.7	36.2	086	27.0	23.5	35.8
032	26.7	21.9	35.6	087	27.3	22.2	35.6
033	27.3	23.5	35.2	088	27.3	22.9	35.4
034	26.5	22.0	36.1	089	27.2	23.6	34.9
035	26.8	22.4	35.9	090	26.2	20.6	35.2
036	27.1	22.8	35.8	091	27.2	22.1	34.5
037	27.7	23.7	35.7	092	26.4	21.3	35.4
038	27.7	23.9	34.8	093	27.4	22.1	34.3
039	27.9	24.7	34.9	094	27.9	23.4	35.1
040	27.4	23.7	34.8	095	27.4	23.6	35.3
041	26.9	22.7	35.8	096	27.8	24.6	35.4
042	28.3	24.5	34.5	097	27.4	24.4	35.5
043	28.1	24.8	34.8	098	26.7	23.8	36.0
044	28.1	24.9	34.9	099	27.6	23.8	35.6
045	28.0	24.6	34.6	100	27.7	24.2	35.2
046	26.8	24.9	35.9	101	27.5	23.6	35.2
047	26.3	24.0	36.5	102	27.0	22.7	35.7
048	26.9	23.4	36.1	103	26.0	20.3	34.9
049	26.5	22.7	36.3	104	26.9	22.1	35.2
050	27.2	25.1	35.4	105	26.4	22.0	35.4
051	27.3	25.6	35.1	106	26.9	23.0	35.0
052	27.4	23.7	35.2	107	26.5	22.8	34.7
053	27.0	22.7	35.4	108	26.3	20.9	35.9
054	26.2	22.3	36.1	109	27.0	23.6	35.2
055	26.0	21.2	36.2	110	26.7	22.7	35.9