

A. ABDEL SHAHID

**Application de la régression d'après un tableau
de correspondance : l'estimation des paléoclimats
d'après l'écologie des foraminifères**

Les cahiers de l'analyse des données, tome 7, n° 1 (1982),
p. 93-111

http://www.numdam.org/item?id=CAD_1982__7_1_93_0

© Les cahiers de l'analyse des données, Dunod, 1982, 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/>

APPLICATION DE LA RÉGRESSION
D'APRÈS UN TABLEAU DE CORRESPONDANCE :
L'ESTIMATION DES PALÉOCLIMATS
D'APRÈS L'ÉCOLOGIE DES FORAMINIFÈRES
[RÉGR. CLIMATS]

par A. Abdel Shahid.⁽¹⁾

0 Rappel méthodologique : Les principes géométriques des calculs de régression sont connus depuis assez longtemps : un exposé synthétique en a été publié dans les *Cahiers* (cf. [REG. GEOM.] in *Cah.* Vol. III n° 2 pp 131-146 ; 1978) ; et dans une série d'articles P. Cazes a montré tout le parti que ces méthodes pouvaient tirer du calcul électronique et de l'analyse des données (cf. *Cahiers* Vol. III n° 2, 3, 4 ; et notamment [REGR. ANAL.] in *Cah* Vol. III n° 4 pp 385-391).

Placée dans un cadre théorique, où s'inscrivent de nombreuses applications pratiques, la régression n'en demeure pas moins un problème statistique où il est difficile de rechercher l'optimum en se gardant des solutions illusoires. L'objet du présent article est de montrer sur l'exemple de données concrètes, comment en progressant pas à pas on parvient à assurer et améliorer les résultats. Nous adressant à des lecteurs dont la plupart ne sont, pas plus que nous ne l'étions naguère nous-mêmes, initiés aux méthodes écologiques de la paléoclimatologie, il nous a paru utile de souligner les principes généraux du présent travail. L'article est rédigé de telle sorte qu'on en puisse saisir les conclusions (§ 4) si, après en avoir étudié l'introduction méthodologique (§ 0), on parcourt rapidement la présentation des données (§ 1) et la suite des étapes du calcul (§§ 2 et 3).

Une première difficulté des calculs de régression est qu'en utilisant autant (ou plus !) de variables explicatives qu'il y a d'individus de base on peut expliquer (reconstituer) parfaitement toute variable par une formule dont l'application à des individus nouveaux sera toutefois en général totalement illusoire. Contre une telle illusion on se protège d'une part, en réservant un échantillon d'épreuves, d'autre part en réduisant le nombre des variables primaires. Cependant quand est faible le nombre des individus pour lesquels sont connus à la fois les variables explicatives et les variables à reconstituer, réserver un échantillon d'épreuves (qui servira non au calcul des coefficients de la formule de régression, mais seulement, en bref, au choix du nombre des variables explicatives retenues) conduit à rétrécir dangereusement la base des calculs. Une judicieuse compression préalable de l'information en substituant aux nombreuses variables primaires un petit nombre de facteurs, fournit au contraire une base à la fois aussi large (tous les individus) et aussi stable (très peu de variables) qu'il est possible. De ce fait il devient possible d'avoir une saine estimation de l'erreur commise d'après les individus de base eux-mêmes.

(1) Docteur ingénieur. Maître de conférences - école polytechnique - Le Caire.

Mais ici deux difficultés attirent notre attention : d'une part comment réaliser au mieux, par l'analyse factorielle, ce qu'en termes imagés on a appelé une compression de l'information ; d'autre part par quels indicateurs estimer l'erreur, c'est-à-dire l'écart entre les véritables valeurs des variables à expliquer et leur estimation par une formule de régression en fonction des variables explicatives.

Supposons que les informations de base constituent un véritable tableau de correspondance (ou t. de contingence) croisant un ensemble I_1 d'individus de base et un ensemble J_1 de variables explicatives. Une première méthode de compression qui s'impose est de soumettre à l'a. des correspondances le tableau $I_1 \times J_1$. Mais cette analyse fournit des facteurs qui rendent compte au mieux de la diversité des individus de I_1 , quant aux variables de J_1 : or tel n'est pas exactement notre problème, la description de I_1 suivant J_1 ne nous intéresse qu'en tant qu'elle est susceptible d'expliquer la diversité de I_1 du point de vue des variables à expliquer. Plutôt que le tableau $I_1 \times J_1$ on analysera donc un tableau $J_2 \times J_1$, construit comme suit. Les variables à expliquer sont, d'après les histogrammes de leurs valeurs sur les individus de I_1 , découpées en classes : soit J_2 l'ensemble des modalités ainsi obtenues. Dans chacune de ces modalités j (e.g., si l'une des variables à expliquer est une température, la modalité : température comprise entre 20°C et 25°C ...) rentre un certain sous-ensemble $I(j)$ des individus de I_1 : la ligne j du tableau $J_2 \times J_1$ analysé sera tout simplement la somme des lignes i du tableau $I_1 \times J_1$ (tableau de base) pour lesquelles $i \in I(j)$. En adjoignant au tableau $J_2 \times J_1$ le tableau $I_1 \times J_1$ comme ensemble de lignes supplémentaires, on étend à l'ensemble I_1 (des individus de base) les facteurs issus du tableau $J_2 \times J_1$ (que nous avons construit) : les facteurs obtenus sur I_1 n'étant autres évidemment que des combinaisons linéaires particulières des données initiales (colonnes du tableau de base $I_1 \times J_1$). L'expérience montre que ces facteurs réalisent la compression cherchée de l'information pertinente.

Cette compression peut pourtant être améliorée, en considérant la deuxième des difficultés que nous annonçons ci-dessus : estimer l'erreur de la formule de régression. Cette estimation, on l'a dit, sera faite ici d'après l'ensemble I_1 tout entier, parce que mettre à part des échantillons d'épreuve rétrécirait trop la base des calculs. Mais il reste à mesurer globalement l'erreur commise (ou complémentaiement la qualité de l'ajustement). Une première idée classique est de calculer le coefficient de corrélation entre variable à estimer et résultat de la formule de régression linéaire ; une deuxième idée (non moins classique) est de calculer le maximum de l'erreur commise : on utilisera à la fois les deux méthodes : d'une part le calcul de corrélation servira à choisir les coefficients de la meilleure formule linéaire : d'autre part la qualité de la formule ainsi trouvée sera appréciée d'après l'erreur maxima.

Or ici une nouvelle question embarrassante se pose : sur quel ensemble d'individus calculer cette erreur maxima. L'idéal serait de prendre le maximum sur l'ensemble I_S des individus nouveaux pour lesquels sera utilisée la formule de régression établie : mais cela est impossible, car si se pose un problème de régression c'est précisément parce que pour les individus de I_S les variables à expliquer sont inconnues ! Pourtant à défaut de I_S lui-même on peut prendre pour base de l'estimation de l'erreur, le sous-ensemble des individus de I_1 qui recouvrent le domaine de I_S ; autrement dit on restreindra I_1 au sous-ensemble des individus qui sont parmi les plus proches voisins de ceux de I_S . On rejoint ici une autre technique qui sert grandement à la régression : la recherche des voisins : (cf. M.O. Lebeaux : programme POUBEL, in *Cahiers* Vol. IV n° 4 pp 467-481). Par voisins, il est clair qu'il faut entendre ici : voisins dans l'espace des premiers facteurs issus de l'analyse des données disponibles. Mais cette

restriction aux voisins (de IS dans II) ne vaut pas seulement pour le calcul de l'erreur maxima : elle vaut aussi en amont : pour le calcul des coefficients d'après la corrélation, et d'abord par le calcul des facteurs eux-mêmes. Telle est donc la voie que nous suivrons maintenant sur des données écologiques (ultérieurement paraîtra un travail de A. Hathout consacré au problème général du choix des voisins).

1 Origine du problème et des données : Nous reprenons ici un travail publié par Mrs Kipp, Bé et collab. (Geological Society of America Memoir 145 pp 3-41 et 165-300). Son principe 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 faudra estimer les valeurs de ces mêmes paramètres pour des époques reculées, à partir de la composition faunistique des sédiments correspondants.

L'estimation de Mrs Kipp et Bé est basée sur les facteurs issus d'une analyse en composantes principales (ACP) portant sur un tableau résultant d'une transformation du tableau initial des pourcentages d'espèces par échantillon.

Dans un travail récent, [PALEOCLIMATS] in *Cah.* Vol. IV n° 1 pp 61-79, 1979) M. Roux propose une autre méthodologie qui ne diffère de la précédente que par la substitution de l'analyse des correspondances sur les données brutes à l'ACP sur données transformées. Cette méthode, comme nous le montre M. Roux, aboutit à des coefficients de corrélations satisfaisants. Elle est utilisée dans notre travail et sera expliquée dans les §§ 2.1 et 3.2.

Nous proposons grâce aux conseils du professeur J.P. Benzécri et de M. Roux, une autre méthode qui consiste à appliquer l'analyse des correspondances, non sur les données brutes (tableau initial), mais sur un tableau que l'on formera d'une façon précise (cf. §§ 2.2 et 3.2.3) et que l'on appellera tableau cumulé par classe de variables climatiques des plus proches voisins (PPV).

Cette méthode comme nous le verrons dans les §§ 2.2 et 3.2 conduit à des coefficients de corrélation nettement plus satisfaisants.

Les données : Les données publiées par Mrs Kipp et Bé sont constituées de deux parties. La première partie se compose de 191 observations de sédiments récents, provenant de stations assez bien réparties sur l'Océan Atlantique du Nord (cf. fig. 1a), la deuxième partie se compose de 221 sédiments anciens (fossiles) issus de 7 carottes dont les localisations figurent dans la figure 1b. La première partie des données est constituée de 2 blocs (cf. fig. 1c) : un tableau (X) recensant les abondances de 42 espèces de foraminifères et un tableau (Y) d'environnement, comportant 5 variables : moyenne des températures d'hiver à la surface de l'eau (en abrégé THS), moyenne des températures d'été à la surface de l'eau (en abrégé TES), moyenne des températures d'été à une profondeur de 100 m (en abrégé TEP), moyenne de la salinité à la surface de l'eau (en abrégé SS) et moyenne de la salinité à une profondeur de 100 m. (en abrégé SP).

La seconde partie des données ne comprend qu'un tableau faunistique X_S , et il faudra estimer les valeurs des 5 variables de l'environnement correspondant à ce tableau X_S , c'est-à-dire estimer les valeurs du tableau Y_S .

La première remarque à noter à propos des espèces figurant dans ce tableau est que les espèces J11, J40, J41, J42, J44, J49, J50, J51 et J52 (cf. appendice 1) ne figurent plus dans les prélèvements

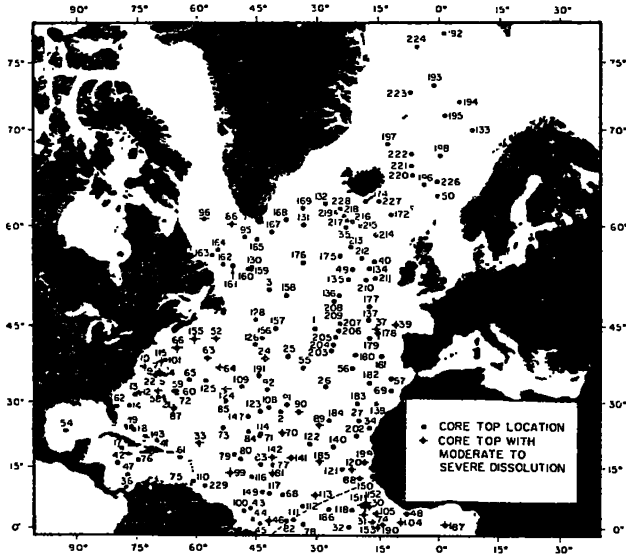


Figure 1a

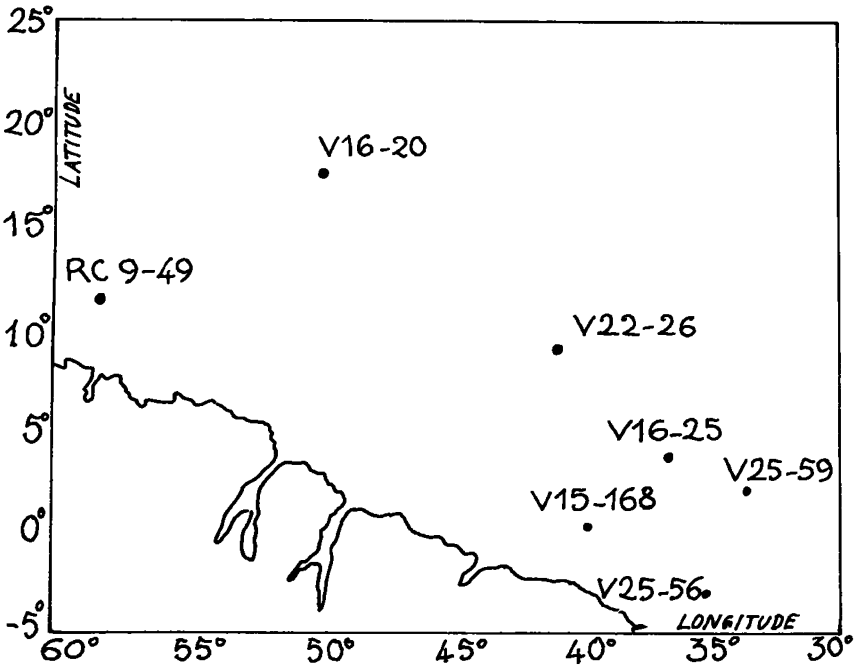


Figure 1b : Emplacement des 7 carottes élaborées pour l'obtention des 221 fossiles

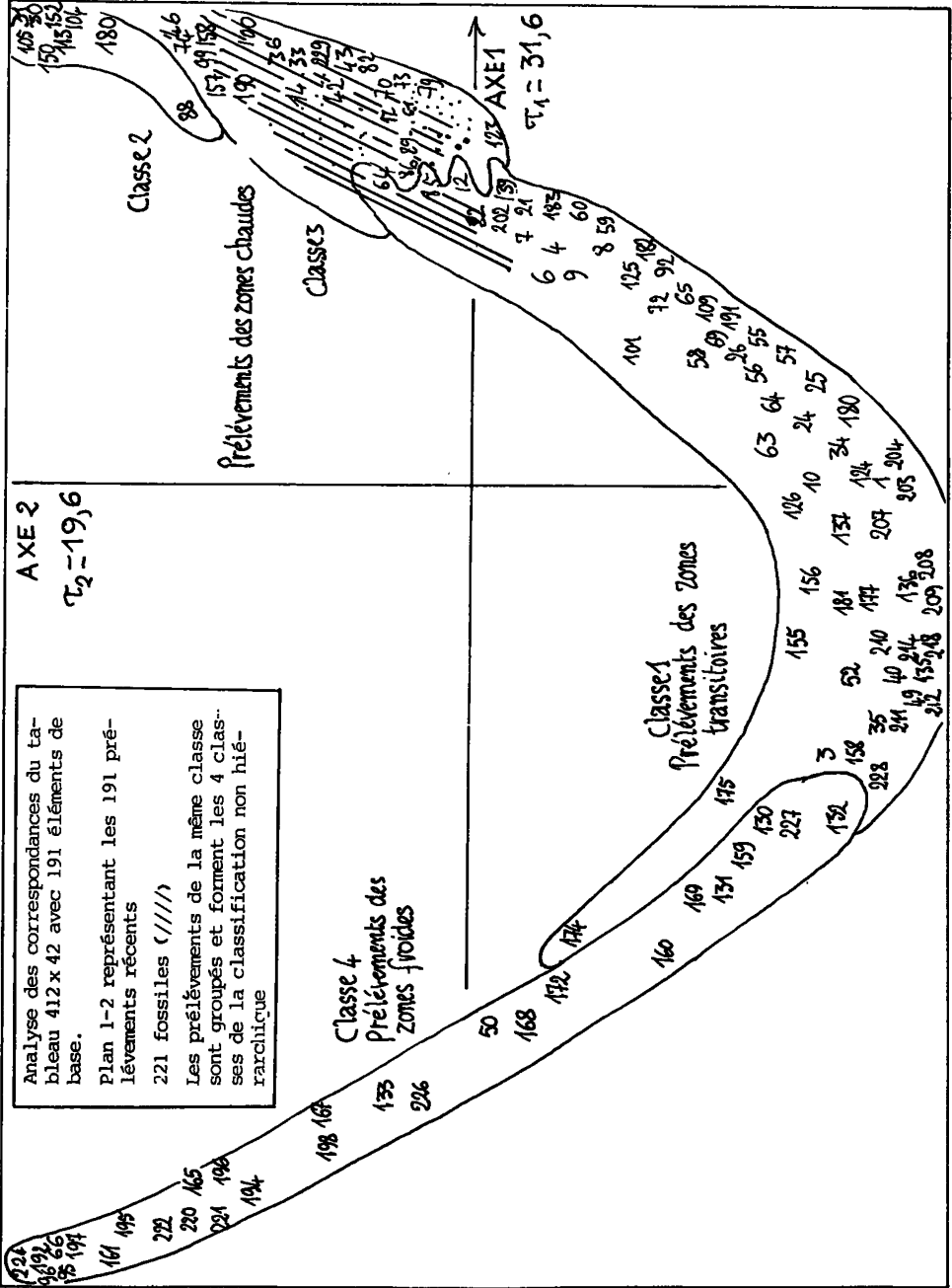


Figure 2.1a

		J1	J2
		42 espèces	5 paramètres
191 prélèvements récents	I ₁ {	X	Y
221 fossiles	I ₂ {	X _S	Y _S =?

Figure 1c

récents. Du reste de ces espèces, nous allons retenir, comme base de l'interprétation des résultats de l'analyse, les espèces ayant une abondance suffisante dans les prélèvements, et qui, d'après leurs distributions peuvent être affectées à une de ces 5 catégories.

- 1 Espèces polaires : J19, J20
- 2 Espèces circumpolaire : J14, J21, J39, J39
- 3 Espèces péritropicales : J15, J27, J31, J32
- 4 Espèces tropicales : J51 et J9
- 5 Espèces tropicales (Gyre Margin) : J22, J25, J37

2 Analyse des correspondances

A partir du tableau initial on forme un tableau $\{I_1 \cup I_2\} \times J_1$ de pourcentages, i.e. à l'intersection de la ligne i , $i \in I_1 \cup I_2$ et de la colonne j , $j \in J_1$ on a, pour l'espèce j , la valeur de son abondance relative dans l'échantillon i ; ainsi, la somme des termes de la ligne i , $i \in I_1 \cup I_2$, $j \in J_1$ est égale à 100.

2.1 Analyse du tableau des pourcentages 191 x 42 : Les 5 variables climatiques THS, TES, TEP, SS et SP sont découpées chacune en 8 modalités d'absence, présence par le programme STEAK (M. Roux, 1978). Les 40 modalités ainsi formées sont juxtaposées au tableau de pourcentages, formant ainsi le tableau 412 x 82. On soumet ce tableau à l'AFC en considérant les 191 prélèvements récents comme éléments de base, les 221 prélèvements anciens comme éléments supplémentaires et les 40 modalités de variables climatiques comme variables supplémentaires.

L'examen du plan factoriel 1-2 pour les observations (figure 2.2a) (51,23% de l'inertie) 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. Pratiquement 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 abondances au voisinage de la diagonale principale et des abondances faibles ou nulles ailleurs (cf. J.P.Benzécri 1973, TII A n° 2, § 3.2).

De même si l'on examine le plan 1-2, variables de base et variables supplémentaires (cf. fig. 2.1b) on peut voir que le premier axe est essentiellement créé par les variables J20, J5, J18, et J8, et que ces variables s'opposent deux à deux de part et d'autre de l'axe 1 : J5 et J8 à une extrémité, J18 et J20 à l'autre.

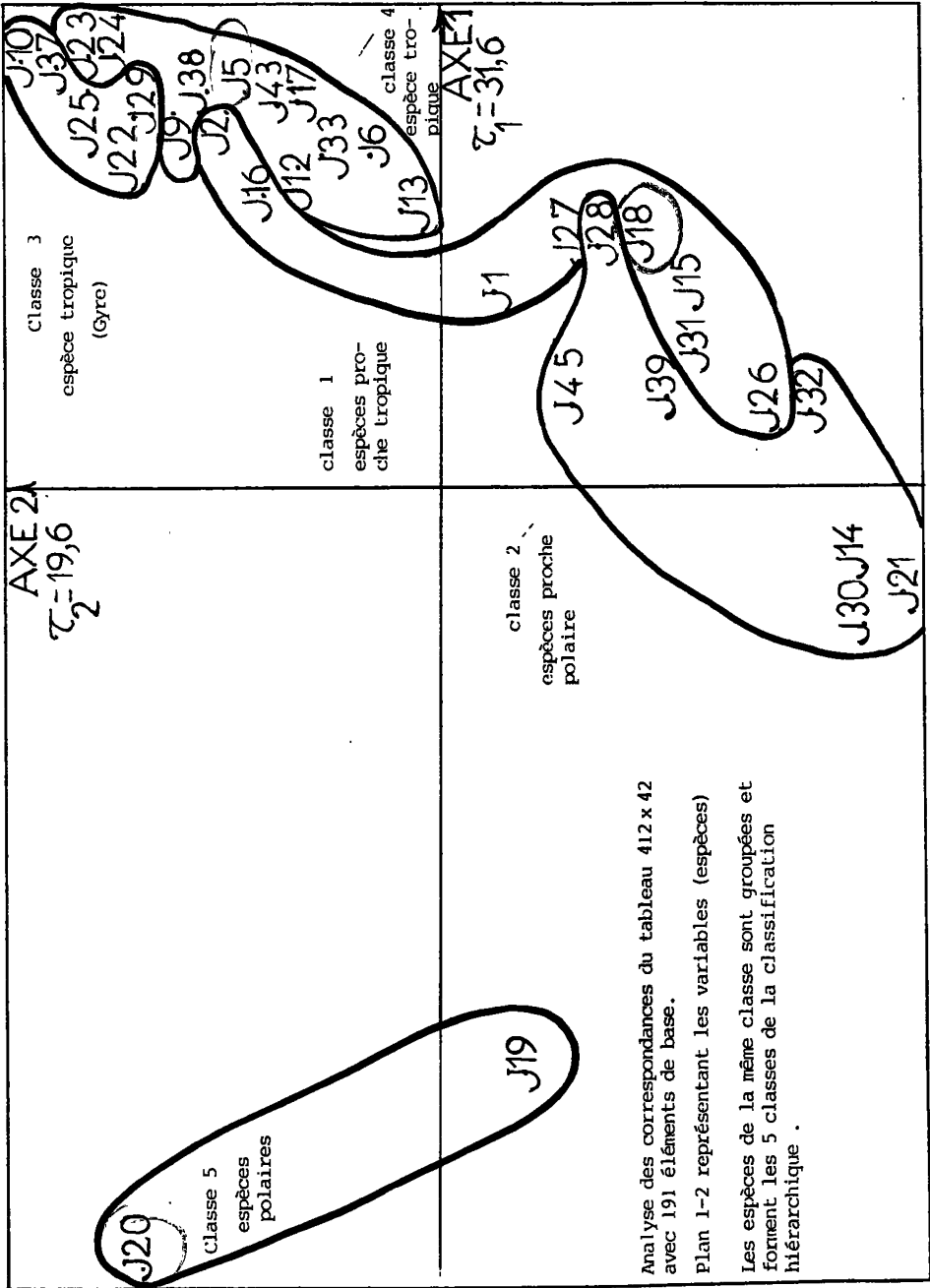


Figure 2.1b

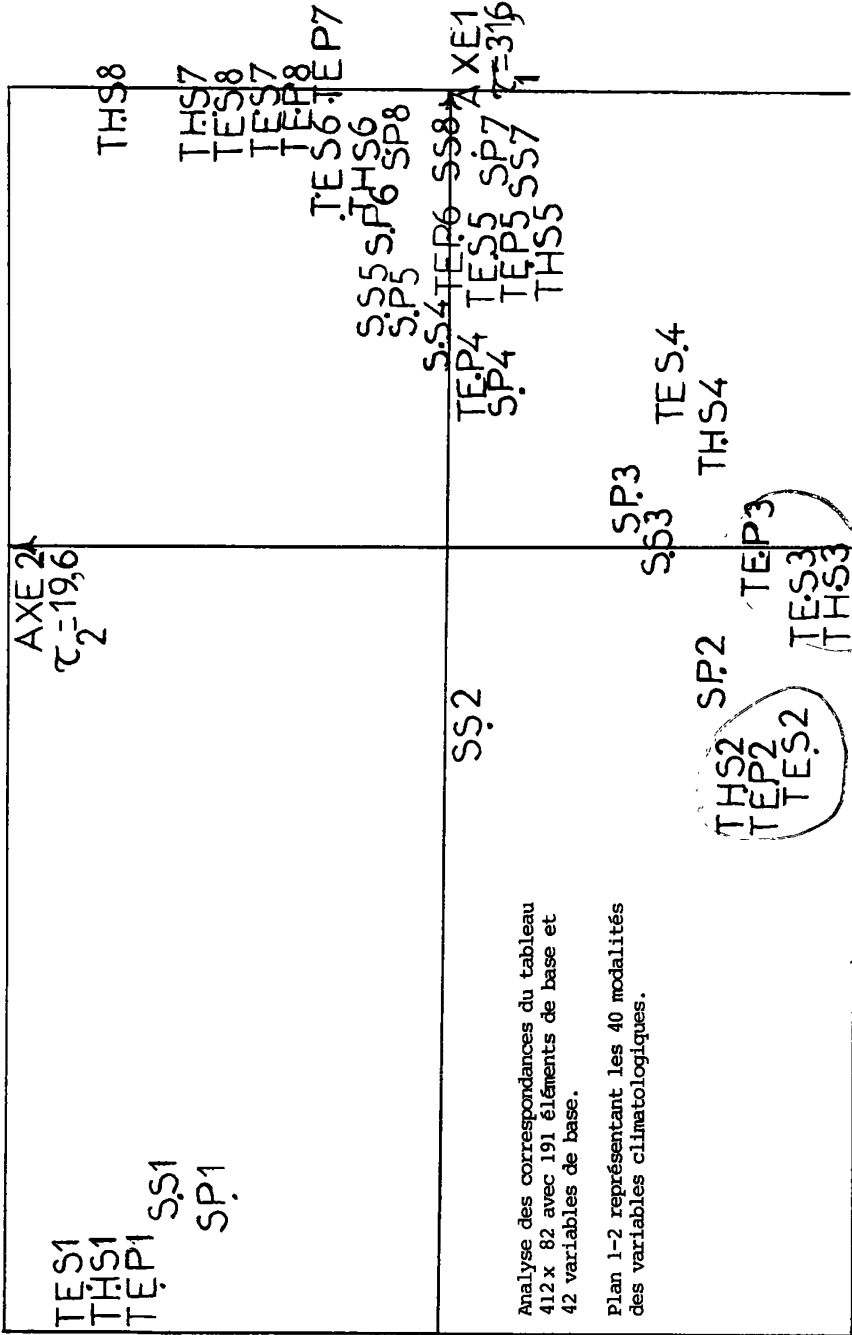


Figure 2.1c

De même les modalités des températures et salinités sont ordonnées dans leur ordre naturel, sur l'axe 1, tout en suivant la courbe parabolique.

De ces observations on peut conclure que le 1-er axe n'est autre que l'axe de température.

Si l'on retourne au plan 1-2 observations, on voit que tous les prélèvements anciens sont groupés, 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 (prélèvements récents) en vue de l'estimation des températures et salinités. On voit en plus que les prélèvements récents, non seulement recouvrent bien les observations supplémentaires (ce qui est satisfaisant) mais qu'elles les débordent très largement (ce qui est fâcheux).

N.B. Cette procédure n'est licite que parce que les lignes du tableau principal (lequel est un tableau de pourcentages) ont toutes un même total, de ce fait une colonne supplémentaire constante (modalité triviale, universellement vraie) se place au centre de gravité, comme il convient.

On obtient des résultats équivalents (à $\sqrt{\lambda}$ près) de façon plus sûre (car indépendamment du total des lignes) en créant pour chaque modalité des variables climatiques, une ligne supplémentaire obtenue par cumul des lignes des relevés rentrant dans cette modalité. Cette voie est d'ailleurs suivie au § suivant où le tableau ainsi construit est pris comme tableau principal.

2.2 Analyse du tableau cumulé par classes de variables climatiques : (40×42) formé d'après les différentes modalités de température et salinité.

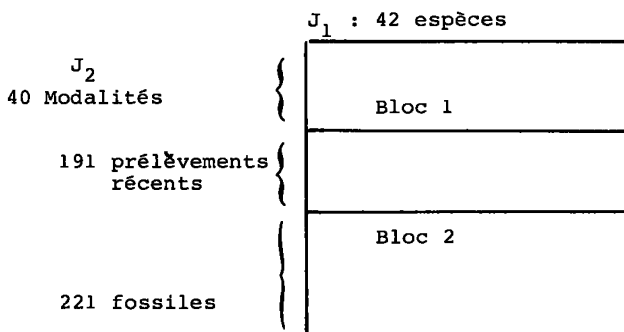


Figure : 2.2a

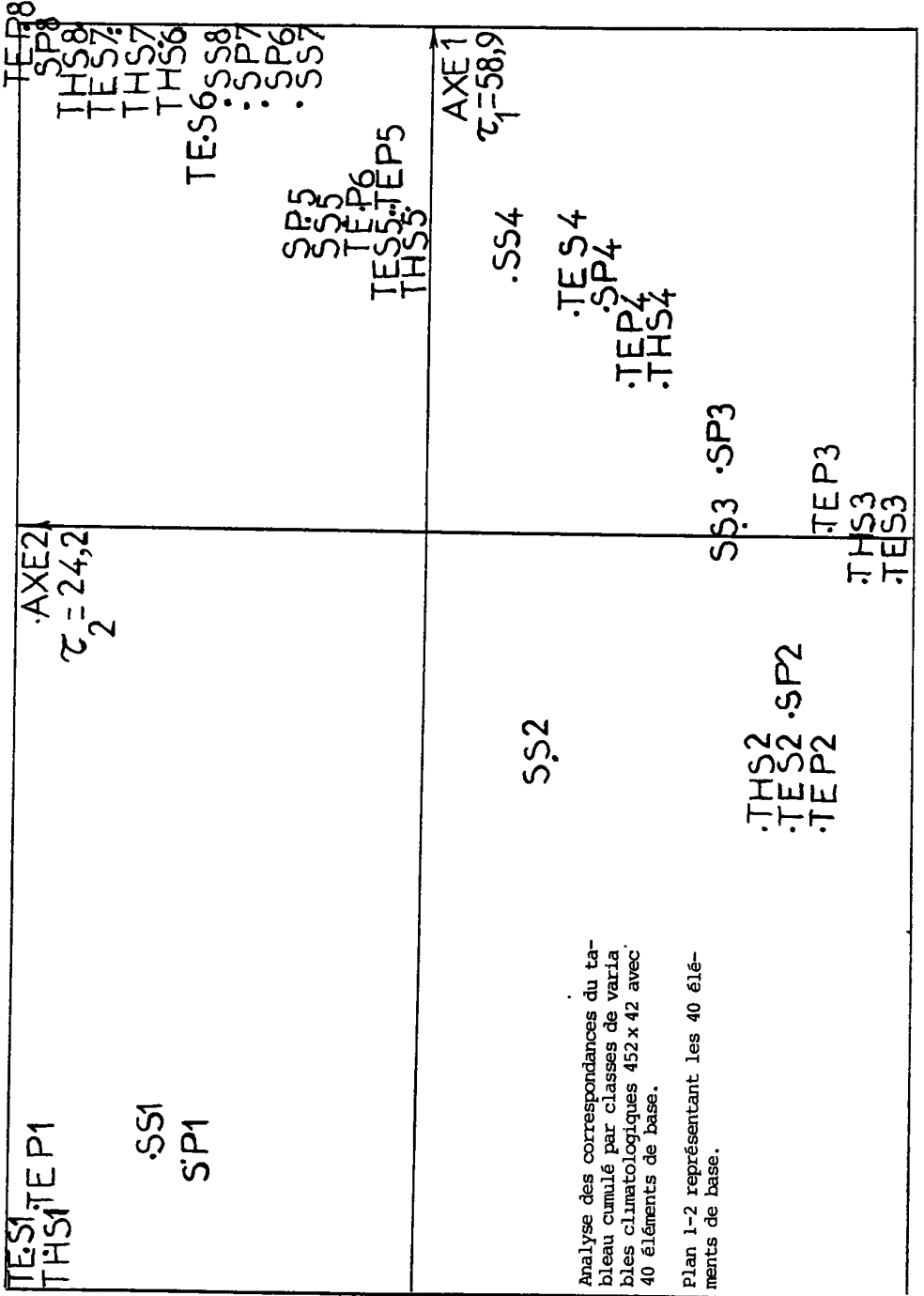


Figure 2. 2b

D'après le tableau 412 × 82 on forme un tableau 40 × 42 de la façon suivante :

Soit K_{IJ} le tableau 412 × 82

où $I = I_1 \cup I_2$, $I_1 = \{1, \dots, 191\}$, $I_2 = \{192, \dots, 412\}$

et $J = J_1 \cup J_2$, $J_1 = \{1, \dots, 42\}$, $J_2 = \{43, \dots, 82\}$

on forme le tableau $k'_{J_2 J_1}$ de la façon suivante :

$$k'(j_2 j_1) = \sum \{k(i, j_2) \cdot k(i, j_1) \mid i \in I_1\}$$

Autrement dit, la ligne j_2 du tableau cumulé $k'_{J_2 J_1}$ représente la somme de toutes les lignes du sous-tableau $K_{I_1 J_1}$ possédant la modalité j_2 . Au tableau k' on superpose les 412 prélèvements récents et fossiles (cf. fig. 2.2a) et on soumet le tableau obtenu à l'analyse des correspondances en considérant le bloc 1 comme tableau de base et le bloc 2 comme éléments supplémentaires.

Le résultat de cette analyse (voir fig. 2.2b) montre d'une façon plus sûre les résultats obtenus dans les §§ précédents.

- i) L'axe 1 est l'axe de température.
- ii) Les prélèvements anciens sont groupés dans l'extrémité "chaude" de l'axe 1.
- iii) Les prélèvements récents recouvrent et débordent plus largement les prélèvements anciens.

3 Estimation des paramètres climatiques d'après la régression multiple

3.0 La régression multiple : Le problème qui nous est posé est le suivant : Comment d'après la connaissance des tableaux X , Y et X_S on peut estimer les valeurs du tableau Y_S (cf. fig. 1c).

Pour résoudre le problème posé, on cherche une procédure ψ permettant de calculer les valeurs manquantes du tableau Y_S (cf. figure 1c) à partir des valeurs connues du tableau X_S :

$$Y_S = \psi(X_S)$$

Il est clair que la détermination de ψ ne peut être conçue que d'après les tableaux X et Y . Pour cela on cherchera d'abord ψ telle que le tableau Y_1 défini par :

$$Y_1 = \psi(X)$$

diffère le moins possible de Y .

Une fois que le résultat Y_1 est satisfaisant, on applique la même procédure ψ sur X_S pour donner une assez bonne estimation de Y_S .

Il est évident que cette méthode suppose que les données contenues dans les tableaux X et X_S décrivent des relevés véritablement analogues : ce qui géométriquement, veut dire que les nuages représentant I_1 et I_2 , dans l'espace des mesures sur J se recouvrent

mutuellement. C'est ici qu'intervient la remarque faite au § 2.1 : Le nuage $N(I_2)$ occupe l'extrémité de la parabole correspondant aux températures élevées. Il faudrait ainsi pour avoir une procédure ψ satisfaisante, se restreindre au sous-tableau de X correspondant aux prélèvements de I_1 qui embrassent, d'après l'analyse des correspondances, les prélèvements anciens (fossiles) I_2 , ce sera le thème du § 3.2.

Mathématiquement, la régression multiple consiste à avoir des variables dites explicatives, et des variables dites à expliquer. Dans notre cas les variables à expliquer sont les paramètres climatiques THS, TES, TEP, SS, SP et les variables explicatives seront les facteurs issus de l'analyse des correspondances.

Ainsi on cherche pour approcher e.g. TES, une formule du type :

$$TES'(i) = b_0 + \Sigma\{b_\alpha F_\alpha(i) \mid \alpha \in A\}$$

où les b_α sont des constantes (les mêmes pour tous les i). La détermination des b_α se fait aisément avec pour critère, sur l'ensemble I , que TES et TES' aient même moyenne et que leur coefficient de corrélation soit aussi élevé que possible.

En pratique il est plus favorable d'exprimer e.g. TES'(i), non en fonction des facteurs $F_\alpha(i)$, mais en fonction des espèces, et cela, dans le cas de l'analyse des correspondances, est aisé, grâce à la formule dite de transition.

$$F_\alpha(i) = \frac{1}{\sqrt{\lambda_\alpha}} \Sigma\{f_{j\alpha}^i G_\alpha(j) \mid j \in J_1\}$$

où $J_1 = \{1, 2, \dots, 42\}$ ensemble des espèces

$$\text{et } f_{j\alpha}^i = \frac{k(i, j)}{k(i)}$$

$$\text{or } k(i) = 100$$

on peut donc écrire

$$F_\alpha(i) = 1/(100 \sqrt{\lambda_\alpha}) \Sigma\{k(i, j) G_\alpha(j) \mid j \in J_1\}$$

et en remplaçant cette expression dans la formule de la régression sur facteurs on a :

$$TES(i) = b_0 + \Sigma\{b_\alpha / (100 \lambda_\alpha^{1/2}) k(i, j) G_\alpha(j) \mid j \in J_1, \alpha \in A\}$$

$$= b_0 + \Sigma\{a_j k(i, j) \mid j \in J_1\}$$

$$\text{où } a_j = \Sigma\{b_\alpha / (100 \lambda_\alpha^{1/2}) G_\alpha(j) \mid \alpha \in A\}$$

3.1 Première estimation des températures et salinités : La méthode de régression multiple citée au § précédent est utilisée 2 fois. La première on considère comme variables explicatives les facteurs issus de l'analyse du tableau 412×42 ($(I_1 \cup I_2) \times J_1$) avec 191 éléments principaux (cf. § 3.1) ; dans le deuxième, ce sont les facteurs issus de l'analyse du tableau cumulé par classe (cf. § 3.3). Le tableau qui figure ci-dessous nous montre les résultats de ces 2 estimations :

Variables à expliquer	Régression d'après l'analyse du tableau 412 x 42		Régression d'après l'a. du t. cumulé par c.(452x42)	
	Coef. de cor.	Résidu max.	Coef. de cor.	Résidu max.
THS	0,9901	5,12	0,9904	5,13
TES	0,9808	7,0	0,9800	5,43
TEP	0,9792	6,21	0,9799	4,09
SS	0,9238	1,67	0,9315	1,71
SP	0,9590	1,06	0,9675	1,1

Si l'on examine ce tableau nous remarquons que nous obtenons des coefficients de corrélation assez élevés, ce qui est une condition indispensable et non suffisante, mais en réalité ces valeurs sont illusoire car comme nous l'avons expliqué au § précédent la qualité de la régression est d'autant plus valable que les données de base et les données supplémentaires sont analogues ; ainsi, comme nous l'avons expliqué il faut se restreindre au sous-tableau de X correspondant aux prélèvements de I₁ qui embrassent d'après l'analyse des correspondances, les prélèvements anciens (fossiles) I₂ ; ceci reste le thème des §§ suivants.

3.2 Calcul sur l'ensemble des plus proches voisins

3.2.1. Détermination des plus proches voisins (PPV) : A l'aide du programme POUBEL nous avons déterminé pour chaque prélèvement ancien la liste des 5 plus proches prélèvements récents, dans l'espace des 7 premiers facteurs issus de l'analyse des correspondances, au sens de la distance euclidienne usuelle :

$$d^2(i, i') = \sum_{\alpha} (F_{\alpha}(i) - F_{\alpha}(i'))^2 \mid \alpha \in A.$$

D'après ces plus proches voisins on a dressé la liste des plus proches voisins de tous les prélèvements anciens, de la façon suivante : tous les prélèvements récents apparus au moins une fois comme PPV ont été retenus. La liste ainsi obtenue nous a permis d'arrêter la liste des PPV à 66 prélèvements récents.

3.2.2 Résultat des calculs de corrélation d'après les PPV : Si l'on considère seulement les 66 PPV, et que l'on réestime les paramètres inconnus, d'après la régression multiple, et d'après les facteurs de ces 66 PPV issus des deux analyses (2.1 et 2.2) nous obtenons les résultats ci-dessous.

Variables à expliquer	Régression (PPV) d'après l'a. du tableau 412 x 42		Régression (PPV) d'après l'a. du tableau cumulé (452 x 42)	
	Coef. de cor.	Résidu max.	Coef. de cor.	Résidu max.
THS	0,844	2,764	0,8793	3,92
TES	0,651	3,711	0,679	2,52
TEP	0,917	4,4	0,918	2,88
SS	0,880	0,536	0,887	0,48
SP	0,875	0,567	0,886	0,42

Comme nous l'avons dit ce sont ces résultats qui doivent être considérés et non ceux du § précédent, et comme on le voit les valeurs absolues des Résidus Maximum sont moins élevées que celles du § précédent.

3.2.3 Régression d'après les facteurs issus de l'analyse des 66 PPV seulement : Nous nous sommes demandé s'il ne valait pas mieux refaire entièrement les calculs, en ne prenant pour données de base que le sous-ensemble des PPV. Ainsi nous avons refait les analyses 2.1 et 2.2 mais en considérant seulement comme éléments de base ces 66 PPV. Deux régressions sur les facteurs issus de ces deux analyses ont donné les résultats ci-dessous.

Variables à expliquer	Régression d'après l'a. du tableau 287 × 42		Régression d'après l'a. du t. cumulé par classe (327 × 42)	
	Coef. de cor.	Résidu Max.	Coef. de cor.	Résidu Max.
THS	0,9440	2,006	0,9554	1,478
TES	0,6719	1,951	0,8069	1,271
TEP	0,9409	2,684	0,9577	1,97
SS	0,934	0,469	0,9431	0,448
SP	0,9456	0,429	0,9476	0,39

De ce tableau nous remarquons que les résultats de la régression basée sur les facteurs issus du tableau cumulé par classe, des PPV, sont les meilleurs, et les seuls non illusoire, du point de vue des coefficients de corrélation et des valeurs absolues de résidu maximum.

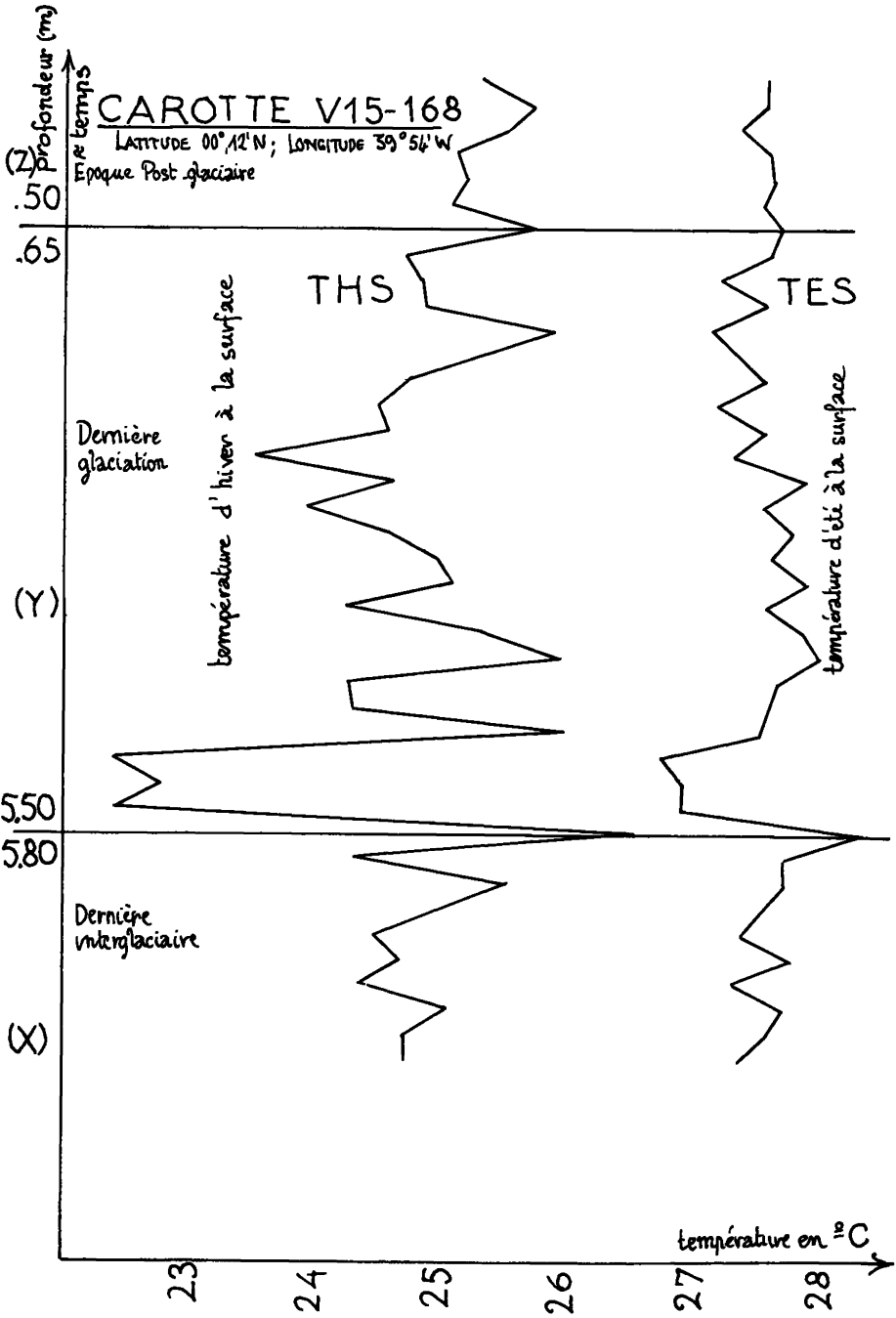
Comme exemple des résultats obtenus on dressera les courbes de température et salinité pour la carotte V15-168 (fig. 3.3a, 3.3b et 3.3c). Nous remarquons que les différentes époques sont dominées par des températures de nature tropicale et que les variations des températures d'hiver sont plus significatives que celles d'été.

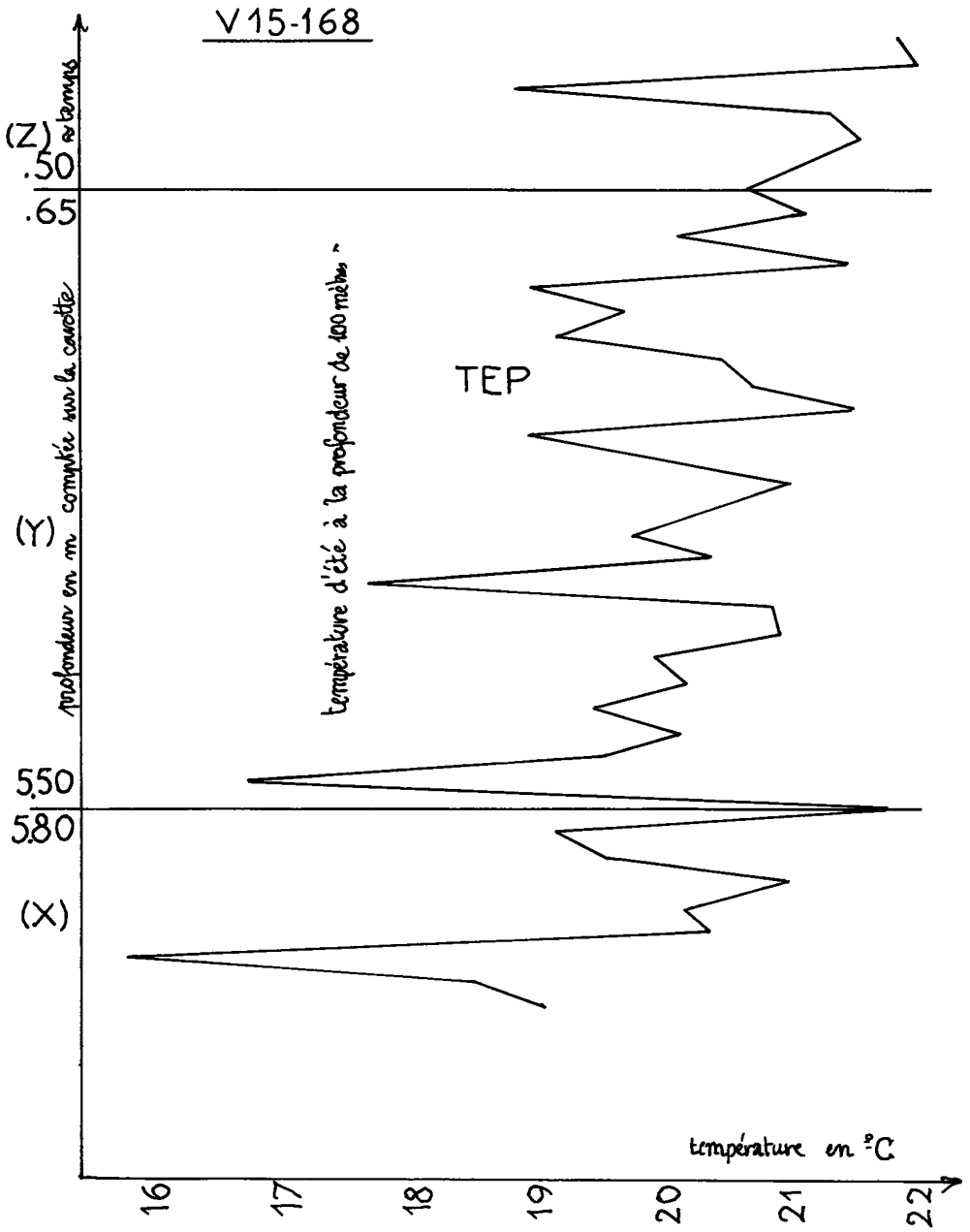
4 Conclusion

Pour conclure ce travail nous poserons la même question qui nous était posée au début : Comment 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 (stations) de prélèvements, on pourrait estimer les valeurs de ces mêmes prélèvements pour des époques reculées, à partir de la composition faunistique des sédiments correspondants ?

La réponse, que l'on a expliquée au fur et à mesure dans ce travail, est la suivante :

- i) Former le tableau de pourcentages des espèces dans chaque prélèvement à partir du tableau des données brutes.
- ii) Découper les variables climatologiques en différentes modalités et les juxtaposer au tableau de pourcentages.
- iii) Former le tableau cumulé par classes climatologiques et lui superposer le tableau de pourcentages.





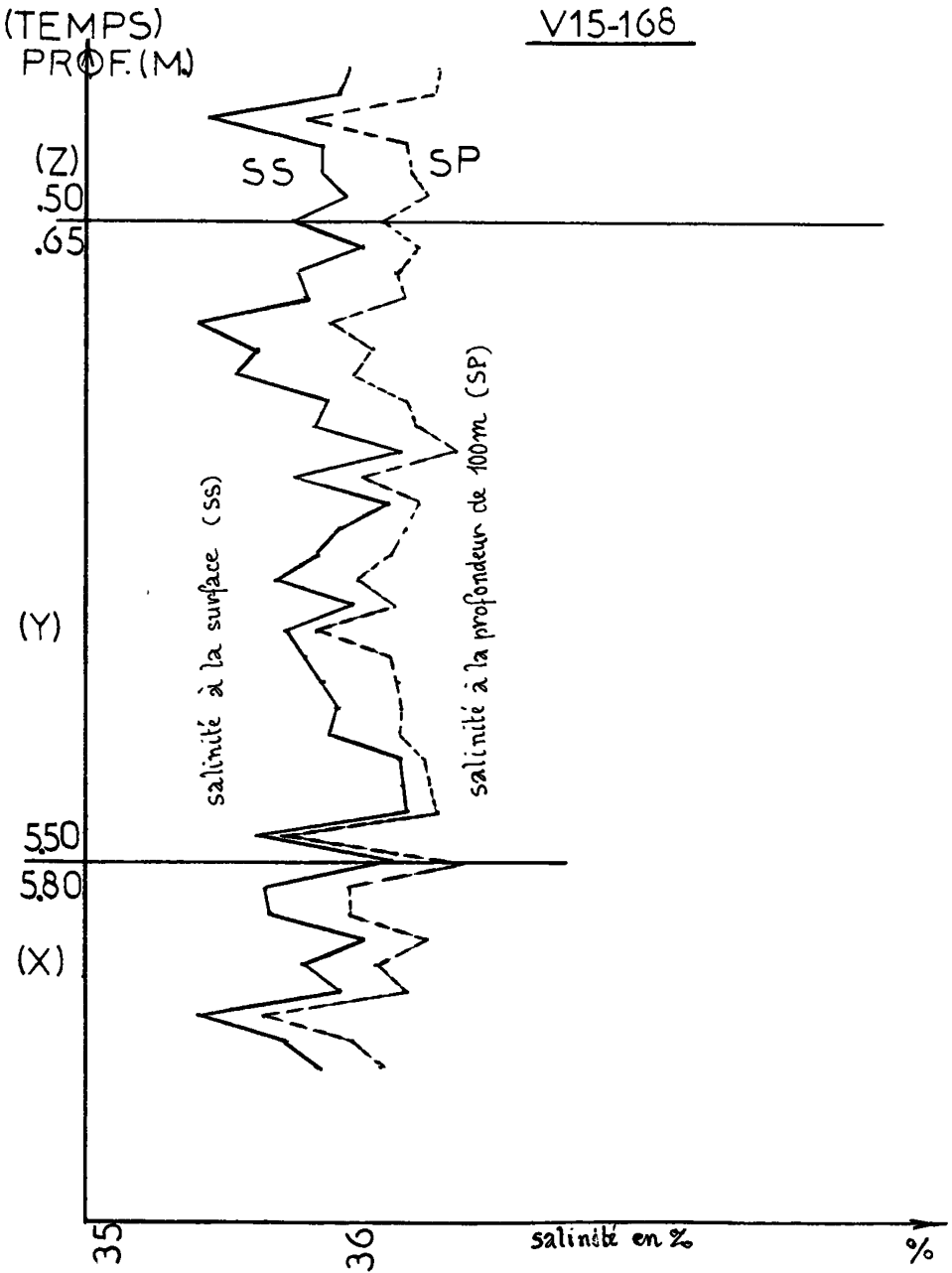


Figure 3.3c

J31	G. HIRSU	:	GLOBOROTALIA	HIRSUTA
J32	G. SCITU	:	"	SCITULA
J33	G. ANFRA	:	"	ANFRACTA
J37	G. M. COMP:	:	"	MENARDII COMPLEX (TOTAL)
J38	C. NITID	:	CANDEINA	NITIDA
J39	G. GLUTI	:	GLOBIGENITA	GLUTINATA
J40	G. IOTA	:	"	IOTA
J41	G. BRADY	:	"	BRADYI
J42	G. PUMIL	:	GLOBOROTALIA	PUMILIO
J43	G. PELAG	:	HASTIGERINA	PELAGICA
J44	H. DIGIT	:	"	DIGITATA
J45	OTHER	:	Autres espèces et non identifiées	
J49	G. UVULA	:	GLOBIGERINITA	UVULA
J50	G. CAVER	:	GLOBOROTALIA	CAVERNULA
J51	G. CRASS	:	"	CRASSULA
J52	U. JUVEN	:	UNIDENTIFIED	JUVENILES
J22	N. DUTER	:	"	DUTERTREI
J23	G. CGLOM	:	GLOBOQUADRINA	CONGLOMERATA
J24	G. HEXAG	:	GLOBOROTALOIDES	HEXAGONA
J25	P. OBLIQ	:	PULLE NIATINA	OBLIQUILOCULATA
J26	G. INFLA	:	GLOBOROTALIA	INFLATA
J27	G. TRNL	:	"	TRUNCATULINOIDES (LEFT COILING)
J28	G. TRNR	:	"	" (RIGHT COILING)
J29	G. CRASF	:	"	CRASSAFORMIS
J30	P. D. INT	:	"P.D. INTERERADE" (intermédiaire entre G. PAC. et G. DUTER)	

iiii) Faire l'analyse des correspondances sur le tableau ainsi obtenu en considérant comme éléments de base les données cumulées et comme éléments supplémentaires le tableau de pourcentages.

V i) Déterminer les plus proches voisins des fossiles, dans l'espace des 7 premiers facteurs issus de l'analyse et dresser la liste de ces PPV.

VI i) Former un autre tableau ayant comme éléments de base les PPV et comme éléments supplémentaires les fossiles.

VIII i) Découper les paramètres climatologiques correspondant à ces PPV et les juxtaposer au tableau.

VIII i) Former le tableau cumulé par classe de variables climatologiques.

IX i) Analyser ce tableau par l'AFC.

X i) Faire une régression multiple sur les facteurs issus de cette analyse.

Cette procédure nous mène à des résultats non illusoires et à des coefficients de corrélation très satisfaisants.

Liste des espèces - KIPP 1976

J 1	O. UNIVE	:	ORBULINA	UNIVERSA
J 2	G. CGLOB	:	GLOBIGERINOIDES	CONGLOBATUS
J 5	G. RUBT	:	"	RUBER (TOTAL)
J 6	G. TENEL	:	"	TENELLUS
J 9	G. SACT	:	"	SACCULIFER (TOTAL)
J10	S. DEHIS	:	SPHAEROIDINELLA	DEHISCENS
J11	O. ADAMS	:	GLOBIGERINELLA	ADAMSI
J12	G. AEQUI	:	"	AEQUILATERALIS
J13	G. CALID	:	GLOBIGERINA	CALIDA
J14	G. BULLO	:	"	BULLOIDES
J15	G. PALCO	:	"	PALCONENSIS
J16	G. DIGIT	:	"	DIGITATA
J17	G. RUBES	:	"	RUBESCENS
J18	G. HUMIL	:	"	HUMILIS
J19	G. QUINQ	:	"	QUINQUELOBA
J20	N. PACL	:	NEOGLOBOQUADRINA	PACHYDERMA (LEFT COILING)
J21	N. PACR	:	"	" (RIGHT COILING)