

J.-P. BENZÉCRI

T. K. GOPALAN

**Sur l'application des méthodes
multidimensionnelles à une anthologie de
données. (1) Morphométrie**

Les cahiers de l'analyse des données, tome 18, n° 4 (1993),
p. 427-446

http://www.numdam.org/item?id=CAD_1993__18_4_427_0

© Les cahiers de l'analyse des données, Dunod, 1993, 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/>

SUR L'APPLICATION DES MÉTHODES MULTIDIMENSIONNELLES À UNE ANTHOLOGIE DE DONNÉES

(1) MORPHOMÉTRIE

[MÉTH. ANTH. DONNÉES (1)]

J.-P. BENZÉCRI
T. K. GOPALAN

0 Le Projet d'étude comparative

Bryan F.J. MANLY: *Multivariate Statistical Methods: a Primer*; Chapman & Hall; Londres, New York; pp.xii+159; (1985).

0.1 Sources de l'anthologie des données

Il y a quelque huit ans, qu'est parue, sous la signature d'un professeur de l'Université d'Otago, en Nouvelle Zélande, une introduction à l'analyse multidimensionnelle, dont son auteur décrit, modestement, l'objet en ces termes:

"Ce livre entend être une introduction aux méthodes multidimensionnelles, à l'usage de non mathématiciens. Il n'a pas la prétension d'être complet le but est, au contraire, de donner, avec un minimum de détails, une vue exacte de ce qu'il est possible de faire. Autrement dit, le livre vise à vous 'lancer sur les voies' (get you going) d'un domaine particulier des méthodes statistiques

En effet, tout en donnant, des bases mathématiques, un exposé adapté au niveau de ceux qui ont reçu dans une université de langue anglaise une formation statistique élémentaire, l'auteur ne propose rien qu'il n'illustre par des exemples; publiant des tableaux de données complets; lesquels, tout en étant bien en deçà de ce que peuvent traiter les micro-ordinateurs d'aujourd'hui, n'en sont pas moins assez étendus pour poser au praticien de véritables problèmes dont la solution, dans ses grandes lignes et, *a fortiori*, dans ses nuances, ne se découvre que par un calcul multidimensionnel.

0.2 Multivariate Analysis et Analyse des Données

Ces exemples nous paraissent offrir une anthologie de ce que l'auteur, et d'autres praticiens avec lui, considèrent comme le domaine de la 'Multivariate Analysis'; et il nous a paru que, sans entreprendre aucunement d'exposer les méthodes que nous utilisons, il valait la peine de parcourir ce domaine du point de vue de l'Analyse des Données.

On déplorera seulement que, l'auteur n'ayant pas proposé de données hétérogènes étendues, ou de dépouillemnt de questionnaires, l'occasion ne nous ait pas été offerte de découper utilement des variables en classes; sinon pour adjoindre le climat en supplément à une anlyse biologique (cf.§.).

Dans huit cas sur neuf, les données de base constituent un tableau croisant un ensemble I d'individus avec un ensemble J de variables; et il est apparu que le tableau, après une élaboration minimale, pouvait être soumis à l'analyse des correspondances; les résultats ainsi obtenus étant éclairés par la CAH, l'analyse discriminante, voire l'introduction de groupes aléatoires.

Alors que l'auteur du *Primer*, se propose de proportionner le choix des méthodes à la diversité des problèmes, nous tentons, au contraire, de ne pas sortir du cadre géométrique adopté au départ. De la statistique élémentaire elle-même, nous ne retenons, pour seule épreuve de validité, que la considération de la place d'un sous-groupe réel relativement à la distribution de groupes aléatoires.

0.3 Enchaînement des exemples

Sans prétendre qu'aucun plan s'impose, nous avons voulu rapprocher les études particulières qui ont entre elles des analogies, notamment dans le caractère des données; tout en progressant du simple au complexe.

Le §1 comprend trois études morphométriques. Au §1.1, 25 gobelets préhistoriques, de forme régulière, sont décrits par six mesures. Au §1.2, on considère 77 mandibules de canidés (chiens, chacals,...) décrites chacune, encore qu'imparfaitement, par 9 mesures: on s'interroge sur la discrimination des espèces. Le §1.3, porte sur 150 crânes humains, relevés en Égypte, dans des sépultures antiques de la région de Thèbes. Avec 4 mesures seulement, la description crâniométrique est très sommaire. Mais nous intéresse la succession temporelle: depuis -4000 jusqu'à ≈150 après l'Incarnation.

Au §2, comme dans deux des études du §1, il s'agit de données biologiques; mais se pose, de plus, la question de l'adaptation des formes vivantes.

La répartition, en six pics consécutifs, des formes variantes d'une enzyme a été déterminée, par électrophorèse, dans seize colonies d'une espèce de papillons: au §2.1, la diversité des 16 profils est mise en rapport avec les modalités de quatre variables de l'environnement.

De 49 moineaux, est donnée une description morphométrique sommaire, d'après laquelle on se proposerait de caractériser les individus ayant survécu à une tempête. Nous croyons avoir montré, au §2.2, que ce projet est vain: mais afin d'écarter l'hypothèse du biologiste, pour qui la sous-population des survivants a une concentration exceptionnelle autour de leur centre de gravité, on effectue des calculs de corrélation et d'inertie, qui pourraient servir ailleurs.

Avec la répartition des emplois, par secteurs, dans les pays d'Europe, le §3.1 aborde l'économie; au §3.2, on retrouve une image globale analogue dans la répartition des sources des protéines consommées en Europe.

Deux des exemples du *Primer*, sont proposés dans le cadre de ce que nous appelons: la représentation euclidienne d'un ensemble muni de masses et distances. Au §4, tout en rappelant des solutions de ce problème, nous en restreignons la portée pratique. En particulier, l'un des exemples n'est autre que l'interprétation, en terme de matrice de similitude, des attitudes de votes dans une assemblée parlementaire: or il est classique de soumettre à l'analyse des correspondances le tableau même des votes; avec l'avantage d'obtenir une représentation simultanée des attitudes et des personnes.

§0.4 Les programmes utilisés

Le *Primer* s'applique à suggérer au lecteur comment programmer, par lui-même, les algorithmes du *Multidimensional Scaling*. L'intérêt pédagogique de cette voie est manifeste. Mais notre but est, au contraire, de montrer que l'ensemble de l'Analyse des Données peut être intégré en un système de programmes communiquant entre eux par des fichiers de format uniforme. Ne pouvant entrer partout dans le détail de l'enchaînement des appels de programme, nous donnons ici un tableau des programmes du logiciel MacSAIF utilisés pour traiter les divers exemples. Le tableau ci-après récapitule les emplois des programmes dans les diverses applications.

Programmes utilisés dans les exemples

	11	qori	plnF	cums	CAH	zran	qsup	cor1	juxt	disc	cumh	cpli
\$1.1	1	1	1	1	1	0	0	0	0	0	0	0
\$1.2	1	1	0	1	1	1	1	0	0	1	0	0
\$1.3	1	1	1	0	0	0	1	0	0	1	1	0
\$2.1	1	1	0	1	1	1	1	0	0	0	0	0
\$2.2	1	1	1	0	0	1	1	1	1	0	1	1
\$3.1	1	1	1	1	0	0	0	0	0	0	0	0
\$3.2	1	1	1	0	1	0	1	1	1	0	0	0
\$4.1	1	1	0	0	0	0	0	0	0	0	0	0
\$4.2	1	1	1	0	0	0	0	0	0	0	0	0

Les programmes 'qori' et 'plnF', d'analyse de correspondances et d'affichage des graphiques, servent partout; par 'qorelsup', sont souvent adjoints des tableaux supplémentaires extérieurs au tableau de base, notamment en vue d'analyses discriminantes, par 'discr'. On a recours à la CAH dans la moitié des applications. Des programmes de création de tableaux, le plus fréquemment utilisé est le programme 'cums', de création de lignes et colonnes par cumul; les échantillons aléatoires étant définis par 'cumh'; 'zrang' sert au découpage de variables en classes, ou à l'affichage de croisements. On a parfois recours à 'juxtab' ou 'coupli' pour extraire et combiner des blocs. Enfin, pour se placer dans l'optique du *Primer*, on effectue, par 'corel', des calculs de corrélation.

1 Trois études morphologiques

1.1 Gobelets préhistoriques Thais

1.1.1 Description des gobelets

Données du Pr. C.F.W. HIGHAM, Université d'Otago.

Le mot français de gobelet est passé en anglais pour désigner une pièce de vaisselle pouvant servir à boire, sans anse, mais avec un pied: par exemple une coupe ou un verre à pied.

La forme des objets préhistoriques, étudiés par le Pr. HIGHAM, est décrite par six mesures, {Lo Lg Ht Lp Lc Hp}, que nous expliquerons en renvoyant aux schémas, disposés dans le plan (1, 2) issu de l'analyse factorielle. Dans ces schémas, le contour du pied est assimilé à deux quarts d'ellipse, posés sur un segment; et la vue frontale de la coupe à une ellipse, tronquée en haut par une corde horizontale.

Le pied est décrit par 3 mesures {Hp Lp Lc}: Hp = hauteur du pied; Lp = largeur, ou diamètre, du pied, à sa base; Lc = largeur du pied à son extrémité supérieure, ou col, par laquelle il se raccorde à la coupe.

De même, on a, pour la coupe, ou globe, les 3 mesures {Hg Lg Lo}: Hg = hauteur, du sommet du pied au plan horizontal de l'ouverture supérieure; cette hauteur est la différence, (Ht-Hp), de la hauteur totale et de la hauteur du pied; Lo = largeur, ou diamètre, de l'ouverture supérieure; Lg = largeur maxima du globe; pour quatre spécimens (notés: B V W X), on a Lg=Lo, ce qui implique qu'à l'ouverture, la forme est celle d'un compotier ou d'une coupe plate, non d'un verre ballon.

Le tableau de base croise un ensemble I de 25 gobelets (dont chacun a pour sigle une lettre capitale) avec l'ensemble des six mesures {Lo Lg Ht Lp Lc Hp Hg}. Dans le tableau analysé, figurent en principal Hp et la différence Hg; tandis que Ht est en supplément. Dans un tableau de mensuration, il est en effet souhaitable de mettre en principal les segments successifs que des repères manifestes déterminent sur une ligne (ou plusieurs): ainsi, est créée une sorte de bilan; qui est la structure même pour laquelle l'analyse des correspondances a été conçue.

gobelets Thai							
	Lo	Lg	Ht	Lp	Lc	Hp	Hg
7							
C	19	23	24	20	6	12	12
D	17	18	16	16	11	8	8
H	12	22	25	15	7	7	18
V	9	9	10	7	4	3	7
X	9	9	8	4	2	2	6
Z	12	19	27	18	5	12	15

Nous publions les données relatives à 6 gobelets dont 5 figurent sur le graphique du plan (1,2).

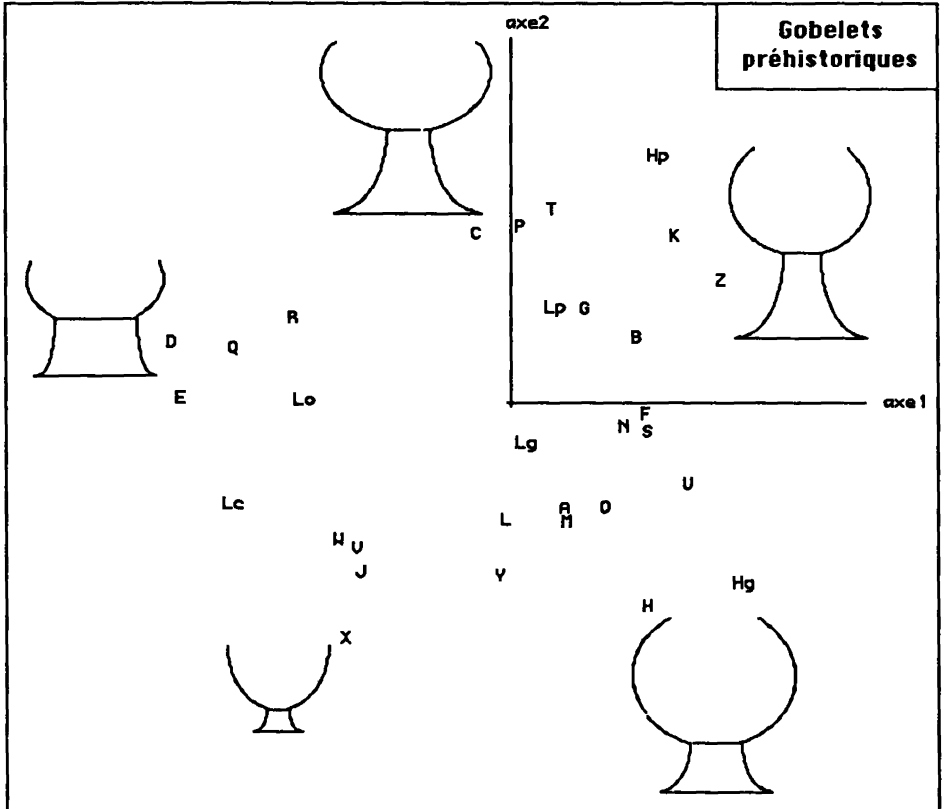
1.1.2 Analyse factorielle du tableau des mensurations

gobelets Thai							
trace :	2.775e-2						
rang :	1	2	3	4	5		
lambda :	133	83	37	19	7	e-4	
taux :	4776	2980	1328	680	236	e-4	

L'axe 1 rend compte de près de la moitié de l'inertie; on a représenté le plan (1,2); mais, dans l'interprétation, nous citerons aussi l'axe 3.

La hauteur totale, Ht, est parfaitement corrélée à l'axe 1: ceci implique que comme Ht, barycentre de {Hp Hg}, le barycentre des 4 autres mesures principales, les largeurs { Lo Lg Lp Lc }, est également approximativement sur l'axe 1, mais du côté (F1<0).

Plus précisément, {Lg Lp}, largeurs maxima du globe et du pied, sont près de l'origine sur le demi-axe (F1>0): l'opposition est donc entre {Hp Hg} et {Lo Lc}. Du côté (F1>0) des formes élancées avec un globe refermé (Lo petit)



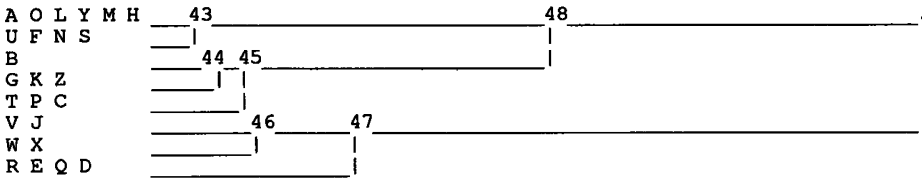
posé sur un pied éfilé (Lc petit); du côté (F1<0) des coupes évasées, sur un pied dont la forme confine au cylindre.

SIGJ	QLT	PDS	INR	F 1	CO2	CTR	F 2	CO2	CTR	F 3	CO2	CTR	F 4	CO2	CTR
Lo	996	183	209	-150	715	313	4	0	0	-91	262	412	24	18	54
Lg	903	246	49	6	7	1	-27	142	23	-11	24	9	-63	730	522
Lp	908	201	78	27	66	11	71	474	123	34	106	62	53	262	299
Lc	991	88	213	-200	596	266	-72	77	55	146	317	509	-4	0	1
Hp	963	112	185	100	216	84	182	722	448	12	3	4	-31	22	59
Hg	992	170	267	159	584	326	-130	391	350	-8	2	3	27	17	65
ci dessous élément(s) supplémentaire(s)															
Ht	961	282	195	136	958	392	-6	2	1	0	0	0	4	1	2

L'axe 2 est créé par l'opposition entre Hp et Hg: du côté (F2>0), où se trouve Hp, la hauteur du pied approche, ou même égale, celle du globe; du côté (F2<0), la hauteur du pied peut n'être que le tiers de la profondeur de la coupe.

Le gobelet 'X', figuré sur le plan (1,2), nous pose une énigme: certes, avec ($F1 < 0$), c'est une véritable coupe, nullement un verre ballon; et son pied, de faible hauteur relativement au tout, s'accorde avec ($F2 < 0$). Mais, relativement à l'ensemble, la largeur d'ouverture Lo est grande, tandis que Lc , col du pied, est faible: or, dans le plan (1,2), X est plus proche de Lc que de Lo . L'énigme se résoud si l'on considère l'axe 3, créé exclusivement par l'opposition entre Lc et Lo : sur cet axe, X, associé à Lc , s'oppose à Lo . La CAH précise cet aspect des formes.

1.1.3 Classification des gobelets



Le listage FACOR montre que la dichotomie supérieure est parfaitement corrélée à l'axe 1: $i48$, ($F1 > 0$), formes élancées s'oppose à $i47$, ($F1 < 0$). De même, la classe 48, se partage exactement suivant l'axe 2: $F2(i43) < 0$, faible hauteur du pied; $F2(i45) > 0$, hauteur égale du pied et du ballon (comme pour Z et C, qui sont dessinés).

Le partage de $i47$ en $i46$ et {R E Q D} se fait dans le plan (2,3), principalement suivant l'axe 2: {R E Q D} (dont D est représenté) est une classe concentrée autour de son centre, proche de l'axe 1; tandis que pour $i46$, $F1$, $F2$ et $F3$ sont négatifs. Cette dernière classe, se subdivise en {V J} et {W X}; la dichotomie est parfaitement corrélée à l'axe 3. Pour les gobelets {W X}, qui sont quasi identiques entre eux, on a ($F3 < 0$); Fc est faible vis-à-vis de Fo , le pied est négligeable dans toutes ses dimensions. Pour {V J}, le pied se place sous la coupe comme une sorte de pastille tronconique de largeur non négligeable relativement à celle de l'ouverture, Lo .

Dans sa présentation des données, Br. F.J. M. s'interroge sur les rapports entre taille et forme; et suggère de prendre pour indice de la taille soit la hauteur totale, Ht , soit la somme de nos mesures. Il apparaît que les gobelets sont tous de taille équivalente, à l'exception de {V W X}, voire de J; ce qui nous ramène à la classe $i46$. L'analyse des formes suffit donc pour acquérir une vue d'ensemble de la diversité des gobelets.

1.2 Mandibule et dents chez les canidés

C.F.W. HIGHAM, A. KIJNGAM, B.F.J. MANLY: An analysis on prehistoric canid remains from Thailand; *Journal of Archeological Science*, Vol. 7, pp.149-165; (1980).

1.2.0 Analyse préliminaire des descriptions de 7 espèces

Dans le corps du livre de Br. F.J. M., on trouve un tableau 7x6 croisant 7 espèces ou groupes de canidés avec un ensemble de 6 mesures prises sur la mandibule (machoire inférieure) et les dents. La crâniométrie offrirait certainement une meilleure base pour la taxinomie: mais le choix des mesures est justifié par le but de l'étude: comparer un groupe préhistorique à 6 groupes contemporains. En effet la mandibule et les dents se conservent mieux que la boîte crânienne.

À notre avis, il y a peu à attendre de l'analyse de telles données: la proximité mutuelle des espèces et leurs empiètements éventuels ne se révèle que par l'analyse de populations, non de moyennes.

Tcan	8	10	12
Slup			
π can			
Hlup	9	11	
Chcl			
Ding			
Quon			

On se borne donc ici à publier la CAH des espèces, issue de l'analyse du tableau 7x6. Les sigles des espèces sont { π can: chien Thaï préhistorique; Tcan: chien Thaï actuel; Slup: loup de Chine; Hlup: loup de l'Inde; Chcl: chacal; Ding: dingo; Quon: cuon}. Nous citons sans explication les sigles des 6 mesures: {MndL MndH molH molL 3mol 4prm}; qui se retrouvent dans la suite.

En appendice au livre, est publié un tableau 77 x 9 croisant, avec 9 mesures, un ensemble de 77 individus, appartenant au groupe préhistorique ou à 4 des groupes contemporains: c'est sur ce tableau que nous concentrerons notre attention. Même si l'étude de ces données elles-mêmes, tout en montrant l'efficacité des méthodes géométriques mises en œuvre grâce à l'outil de calcul électronique, suscite, chemin faisant, diverses critiques; lesquelles seront reprises avec les conclusions au §1.2.6.

1.2.1 Le tableau principal: 77 individus de 5 espèces

Ainsi qu'on l'a dit, le tableau principal croise un ensemble I de 77 canidés des deux sexes, avec un ensemble J de 9 mesures prises sur la mandibule et les dents. On signalera que la rangée molaire est de trois dents, sauf chez les cuons, où l'on n'en compte que deux.

les loups et les chiens
 Mnd...:longueur de la mandibule; MndL:Largeur de la Mnd;
 Cndl:Larg du condyle articulaire; MndH:Hauteur Mnd;
 molH:longueur de la pr. molaire; molL:Largeur d'icelle;
 tmol:longueur de la rangée molaire; qprm: id prémolaire;
 canL:largeur d'une canine inférieure; tout en mm/dix;

9	Mnd...	MndL	Cndl	MndH	molH	molL	3mol	4prm	canL
Ta	1234	101	231	228	192	78	322	330	56
Tb	1274	96	190	219	192	78	322	404	58
Tc	1211	102	178	210	206	79	345	375	62

De façon précise, les individus sont répartis en un ensemble Is de 9 groupes. Les 10 spécimens préhistoriques de chiens Thaïs sont désignés par π suivi d'une minuscule de a à j: le sexe des individus dont proviennent les mandibules est inconnu.

Pour les quatre espèces actuelles, on a un groupe divisé en 2 sous-groupes suivant le sexe. À chaque espèce est affectée une lettre: {T = chien Thaï; C = Chacal; Q = cuon; H = loup indien}: Un individu mâle est désigné par la lettre de son espèce, en capitale, suivi d'une minuscule, en guise de numéro; on fait le contraire pour les femelles. Ainsi Tc désigne le 3-ème individu de la série des chiens thaïs; cD, le 4-ème de la série des chacals femelle.

Au tableau de base, on adjoint des lignes, afférentes aux espèces ou demi-espèces, créées par cumul de lignes décrivant des individus: e.g., aux chiens mâles contemporains est associé une ligne §H cumul des lignes de 8 individus; aux chiens préhistoriques, § π , cumul de 10 individus; etc.

1.2.2 Analyse factorielle du tableau de base

les loups et les chiens

trace :	2.732e-3								
rang :	1	2	3	4	5	6	7	8	
lambda :	14	6	3	2	1	1	1	0	e-4
taux :	4995	2103	1011	712	514	264	221	181	e-4
cumul :	4995	7097	8108	8820	9334	9598	9819	10000	e-4

Sur l'axe 1, la longueur des rangées molaires et prémolaires {3mol 4prm}, ($F1 > 0$), s'oppose à la hauteur de la mandibule et à la largeur de la mandibule et du condyle articulaire: {MndH MndL Mnd...}.

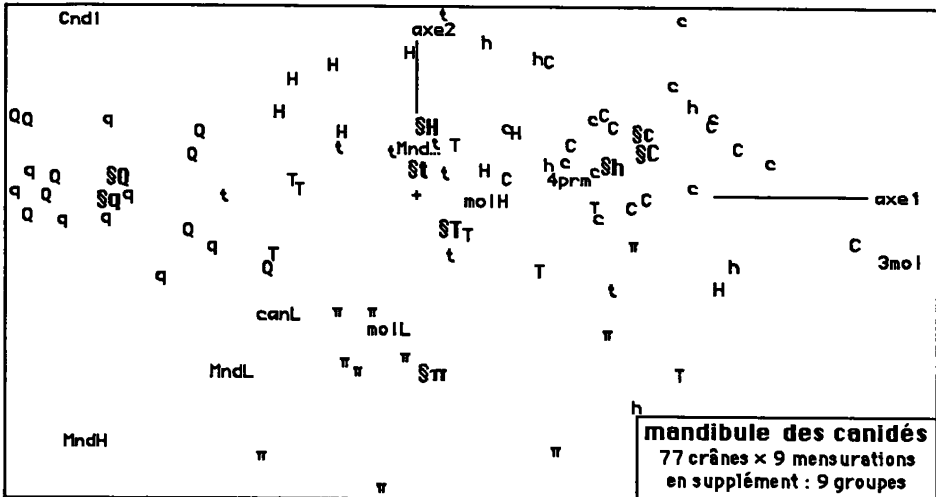
SIGI	QLT	PDS	INR	F 1	CO2	CTR	F 2	CO2	CTR	F 3	CO2	CTR	F 4	CO2	CTR
ci-dessous élément (s) supplémentaire (s)															
\$T	846	104	7	4	102	2	-8	350	12	0	0	0	9	394	41
\$t	679	99	5	0	3	0	6	244	6	5	193	10	6	240	18
\$C	972	115	69	38	875	121	10	56	18	6	19	13	-6	23	22
\$c	871	110	76	38	750	114	14	103	37	6	16	12	2	1	2
\$Q	997	121	131	-51	916	240	4	4	3	-11	43	55	-9	34	62
\$q	984	106	120	-53	922	222	-1	1	1	-10	33	39	-9	28	48
\$H	662	129	21	1	1	0	16	583	58	6	78	16	0	0	0
\$h	925	90	39	32	844	66	7	36	7	-7	42	16	2	3	2
\$π	974	126	87	1	1	0	-41	931	383	4	9	8	8	33	40
SIGJ	QLT	PDS	INR	F 1	CO2	CTR	F 2	CO2	CTR	F 3	CO2	CTR	F 4	CO2	CTR
Mnd...	732	450	49	-3	37	4	12	475	110	-6	143	69	5	76	52
MndL	724	35	62	-34	238	30	-41	353	105	12	29	18	-22	104	91
Cndl	973	77	188	-60	540	203	41	253	226	35	180	334	-1	0	1
MndH	981	75	204	-58	467	190	-58	453	439	3	1	3	21	60	172
molH	685	71	46	9	44	4	-1	1	0	1	1	0	-33	639	415
molL	498	28	37	-7	16	1	-30	255	45	-2	2	1	-28	225	118
3mol	998	113	294	78	861	507	-14	31	43	27	101	293	6	5	22
4prm	687	130	76	23	330	50	4	10	3	-23	341	256	3	6	6
canL	519	21	44	-26	123	11	-27	137	29	-18	60	26	-33	199	123

Quant aux espèces, mises ici en supplément, les cuons mâles et femelles {\$C \$c} s'opposent, sur l'axe 1, aux chacals des deux sexes {\$C \$c}. Au centre, sont les chiens thaïs, {\$T \$t}, les loups mâles, \$H, et les chiens préhistoriques, \$π. Les louves, \$h, sont, sur l'axe 1, proches des chacals; mais la dispersion des individus loups interdit de rien affirmer quant aux espèces en général.

Il importe de noter que les cuons, chez qui la rangée molaire est réduite à deux dents se trouvent opposés à la longueur de cette rangée: cette opposition peut être regardée comme un artéfact lié au choix des mesures.

Sur l'axe 2, \$π, parfaitement corrélé (COR2=931), s'oppose à toutes les espèces contemporaines. Plus précisément, sur l'axe 2, \$π, lié à MndH, s'oppose à Cndl: il se peut que la mauvaise conservation du condyle articulaire fasse sousestimer Cndl sur des spécimens préhistoriques.

Sur le graphique du plan (1, 2), on voit les trois ensembles J (caractères), Is (groupes spécifiques, mis en supplément) et I (individus, figurés par la première lettre du sigle). L'amas des cuons (Q) est dans le demi-plan (F1<0); celui des chacals (C) dans (F1>0). Les mandibules préhistoriques sont toutes dans (F2<0). Chiens Thaïs et loups de l'Inde s'étalent sur une bande allant d'une zone centrale (décalée vers F2>0) jusque dans le quart de plan (F1>0; F2<0). Il y a, pour chacune de ces espèces, un écart entre centre des mâles et



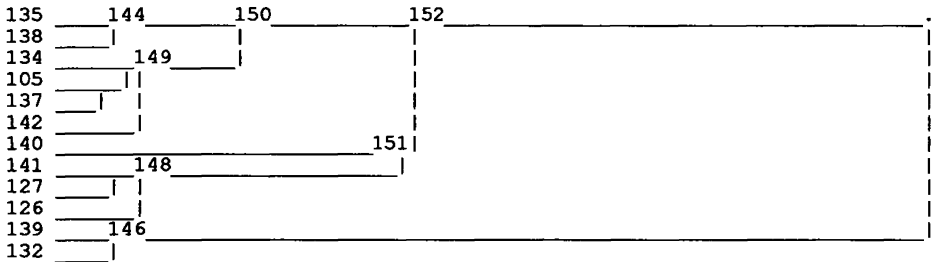
centre des femelles; sans discrimination nette entre les sexes présents sur toute la bande; remarque que confirmeront les analyses ultérieures.

1.2.3 Classification des 77 mandibules

La CAH est faite sans tenir compte de l'espèce ou du sexe, mais seulement de la forme, décrite par 9 mesures. Si des groupes spécifiques se reconnaissent dans les classes, cela prouvera non seulement qu'ils peuvent être discriminés; mais encore qu'ils s'imposent, sans information a priori, de par la seule étude inductive des mandibules.

c	Partition en 12 classes : Sigles des individus de la classe numéro c																					
135	Ta	tH	Hd	Td	tA	Hf	2T	2t	2H													
138	Hh	Ha	He	tF	tE	Te	1T	2t	3H													
134	Tg	hF	tB	Hb	Tb	2T	1t	1H	1h													
105	Hg	hE							1H	1h												
137	Tf	Ci	cF	Ch	Cf	cD	Cj	tD	cH	cG	Cd	cB	1T	1t			5C	5c				
142	hB	tG	cJ	cE	cC			1t	1h								3c					
140	πe	πi	πj	πh	πc	πd	πb	πf												8π		
141	Th	Hc	πg	Ca	hC	Cg	cY	hD	1T		1H	2h	2C	1c						1π		
127	Ce	cA	Cc	Cb														3C	1c			
126	tC	πa	Tc	hA										1T	1t	1h				1π		
139	Qb	qB	qG	Qg	Qf	qC														3Q	3q	
132	qE	qF	qA	Qc	qD	Qa	qH	Qi	Qe	Qh	Qd										6Q	5q

Le bilan des classes fait suivant les groupes montre d'abord que



l'ensemble des 17 cuons - mâles et femelles mêlés, constitue une classe i146 qui se sépare du reste des individus au niveau supérieur de la hiérarchie. Mais on a déjà rappelé que les cuons se distinguent *a priori* par la forme de leur rangée molaire.

Au sein de la classe i152, on distingue trois subdivisions {150 140 148}; dont une seule s'interprète clairement: i140 contient 8 des 10 mandibules π réhistoriques.

Au niveau le plus bas de la hiérarchie retenue, certaines classes sont à peu près confinées dans une espèce: i137 comprend 10 chacals, avec 2 chiens T; et i127 est formée de 4 chacals; mais aucune classe ne s'identifie à une espèce.

1.2.4 Analyse discriminante

Dans l'espace euclidien des profils, rapporté aux 8 axes issus de l'analyse factorielle, les 77 mandibules se projettent comme éléments principaux; les 9 groupes, comme supplémentaires. Le programme 'discr' affecte chaque individu au groupe qui en est le plus proche.

nombre de facteurs utilisés = 8											
affectation des 77 i aux 9 iS (4x2 actuelles + 1 fossile)											
Tb	Te	Tg	tD							-> \$T	
Ta	Td	Tf	tA	tB	tE	tF	tG	tH	[C1]	-> \$t	
[tC]	Cb	Cc	Cd	Ce	Cf	Ch	cA	cY		-> \$C	
Cj	cB	cC	cD	cE	cF	cG	cH	cJ		-> \$c	
Qa	Qb	Qd	Qf	Qh	Qi	qD				-> \$Q	
Qc	Qe	Qg	qA	qB	qC	qE	qF	qG	qH	-> \$q	
Ha	Hb	Hd	He	Hf	Hg	Hh	hB	hE		-> \$H	
[Th	Ca	Cg]	Hc	hC	hD	hF	[π g]			-> \$h	
[Tc	hA]	π a	π b	π c	π d	π e	π f	π h	π i	π j	-> \$ π

Nous donnons la liste des individus affectés à chaque centre; en signalant, entre crochets [], les rares discordances relatives à l'espèce (8/77). Quant au sexe, l'accord est moins bon; toutefois dans la plupart des espèces le partage n'est pas aléatoire: ainsi, à \$Q sont affectés 6 mâles et une femelle; et à \$q, 7 femelles et 3 mâles.

tableau des affectations : in\$*x* à ->\$*y*

9	in\$T	in\$t	in\$C	in\$c	in\$Q	in\$q	in\$H	in\$h	in\$π
->\$T	3	1	0	0	0	0	0	0	0
->\$t	3	6	1	0	0	0	0	0	0
->\$C	0	1	6	2	0	0	0	0	0
->\$c	0	0	1	8	0	0	0	0	0
->\$Q	0	0	0	0	6	1	0	0	0
->\$q	0	0	0	0	3	7	0	0	0
->\$H	0	0	0	0	0	0	7	2	0
->\$h	1	0	2	0	0	0	1	3	1
->\$π	1	0	0	0	0	0	0	1	9

Le tableau d'affectation, construit d'après la liste, confirme la stabilité des espèces, représentées par des blocs carrés le long de la diagonale.

1.2.5 Analyse discriminante avec échantillon d'épreuve

Le succès de l'analyse discriminante dont on vient de rendre compte incite à reprendre la discrimination en réservant un échantillon d'épreuve. On sait que, pour cela, l'ensemble I des individus est partagé entre Ib, échantillon de base, et Ie, échantillon d'épreuve. Le centre de gravité de chaque espèce (ou, ici, demi-espèce) est calculé d'après les individus de Ib.

les loups et les chiens Lupu\$
 en principal: I\$, dé ni d'après les 52 individus de base de Ib.
 trace : 1.528e-3
 rang : 1 2 3 4 5 6 7 8
 lambda : 11 3 1 0 0 0 0 0 e-4
 taux : 7237 1847 431 269 142 52 21 3 e-4
 cumul : 7237 9083 9514 9783 9925 9976 9997 10000 e-4

Plus précisément, on analyse le tableau principal croisant I\$ avec J; où la ligne afférente à chaque espèce \$s est le cumul de ses individus compris dans Ib; et chaque individu, de Ib ou de Ie, est affecté au centre de I\$ qui en est le plus proche.

matrice de confusion pour l'échantillon d'épreuve

9	in\$T	in\$t	in\$C	in\$c	in\$Q	in\$q	in\$H	in\$h	in\$π
->\$T	0	0	0	0	1	0	0	0	0
->\$t	1	2	0	0	0	0	2	0	0
->\$C	0	0	2	3	0	0	0	0	0
->\$c	0	0	0	1	0	0	0	0	0
->\$Q	0	0	0	0	1	1	0	0	0
->\$q	0	0	0	0	1	1	0	0	0
->\$H	0	1	0	0	0	0	1	2	0
->\$h	1	0	1	0	0	0	0	0	0
->\$π	0	0	0	0	0	0	0	0	3

Les centres étant déterminés d'après Ib, on attend que l'affectation soit plus exacte pour cet ensemble que pour Ie; mais, d'autre part, c'est l'affectation de Ie qui seule représente ce que vaudrait notre système de centres pour affecter des individus quelconques, extérieurs aux données de la présente étude. Ici, l'échantillon d'épreuve comprend 25 individus (ceux dont le rang - de 1 à 77 - se trouve divisible par 3: rangs 3, 6, ... 75).

En dehors de l'échantillon d'épreuve, i.e. pour les 52 individus dont le rang n'est pas divisible par 3, il n'y a, quant à l'espèce, que 8 erreurs d'affectation.

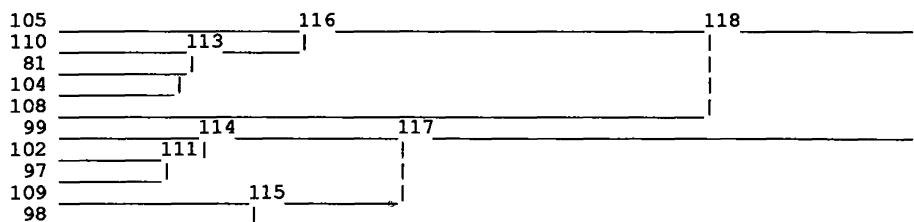
Pour l'échantillon d'épreuve, les affectations exactes quant à l'espèce sont au nombre de 19/25; soit plus des trois quarts; ce qui ne peut être dû au hasard. En revanche, sur 16 affectations exactes quant à l'espèce et concernant des individus non préhistoriques, dont le sexe est donné, 8 sont erronées quant au sexe: le sexe n'est donc aucunement reconnu dans l'échantillon d'épreuve.

1.2.6 Conclusion: discrimination des espèces de canidés

L'analyse multidimensionnelle a montré que l'espèce des cuons, dont la rangée molaire mesurée est réduite à deux dents, ne peut utilement être traitée avec les autres espèces: on doit donc chercher les conclusions de l'étude en se restreignant aux 60 individus des 7 autres groupes.

Voici d'abord une CAH effectuée d'après l'analyse du tableau 60x9, comme au §1.2.3 d'après l'analyse 77x9.

c	Partition en 10 classes : Sigles des mandibules de la classe c									
105	Ta	tH	Hd	Td	tA	Hf				
							1T	1H	2h	2C 1c 1π
110	Hh	Ha	He	tF	tE	Te	hB	tG		
81	Hg	hE							1H	1h
104	Tg	hF	tB	Hb	Tb				2T	1t 1H 1h
108	πb	πf	πe	πi	πj	πh	πc	πd		8π
99	Tf	Ci	cF	tD	cH	cG	Cd	cB	1T	1H 2C 4c
102	cC	cJ	cE	Ch	Cf	cD	Cj			3C 4c
97	Ce	cA	Cc	Cb						3C 1c
109	Th	Hc	πg	Cg	cY	hD	Ca	hC	1T	1H 1h 3C 1c 1π
98	tC	πa	Tc	hA					1T 1t	1h 1π



Pour les espèces retenues, la CAH diffère peu de celle faite en présence des cuons. Les mandibules préhistoriques sont bien reconnues dans 108; ainsi

que les chacals dans 114. Aucune séparation n'apparaît entre chiens Thaïs et loups, qui sont intimement mêlés dans 116 comme dans 115.

On effectue ensuite une analyse discriminante avec échantillon d'épreuve: comme au §1.2.5, le tableau principal croise avec l'ensemble J des mensurations, des groupes, ici 7, dont les profils sont définis par cumul sur Ib, en réservant Ié comme échantillon d'épreuve. Comme au §1.2.5, on a, sur l'échantillon d'épreuve, un taux élevé de reconnaissance des espèces: le taux des erreurs n'est que de 5/20.

De façon précise, voici la liste de ces erreurs:

(Tc->X §h) (tG->X §H) (Cb->X §h) (Ha->X §t) (Hd->X §t) ;

mis à part le chacal mâle Cb affecté au groupe des louves, §h, il s'agit de confusions entre chiens et loups.

Puisqu'il est apparu, au §1.2.4, que le sexe ne peut être reconnu d'après la forme, on a tenté de faire une analyse discriminante avec échantillon d'épreuve, non avec 7 centres de groupes, mais avec 4 centres d'espèce: {&Tt &Cc &Hh §π}. Nous avons eu la surprise de trouver des résultats nettement moins bons qu'avec 7 centres. Cette expérience apporte une conclusion importante, qu'on interprétera a posteriori. Dans l'espace des formes, ainsi qu'on l'a vu dès l'analyse factorielle du §§2, une espèce n'est pas nécessairement concentrée autour d'un point: elle peut s'étaler sur une bande. La discrimination sera meilleure si, pour une telle espèce, on dispose de deux centres qui rendent compte de son extension spatiale.

Reste la confusion entre chiens et loups. Il se trouve que cette confusion, manifeste quant à la forme, se résoud par la considération de la taille. Chez les chiens Thaïs des deux sexes, la longueur Mnd... de la mandibule n'est supérieure à 130,1 mm que dans un cas; chez les loups de l'Inde, au contraire, Mnd... dépasse 140,9 mm, excepté dans un cas.

Il est connu que relativement au crâne, la mandibule est beaucoup plus développée chez le loup que chez le chien: ainsi, la différence de taille que l'on voit ici sur un échantillon de mandibules, procède d'une différence de forme: critère bien plus satisfaisant, compte tenu de la grande variabilité de taille des chiens domestiques.

Il semble donc que le problème de discrimination posé soit résolu. Mais on

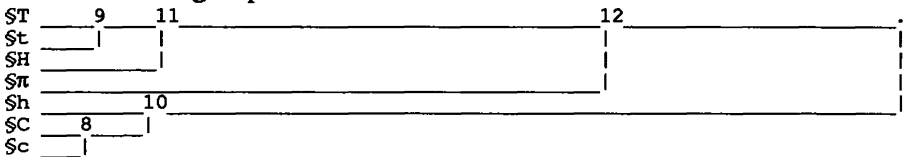
doit rappeler qu'on a exclu les cuons, avec leur rangée à deux molaires; et que l'opposition à CndI, propre aux restes préhistoriques, peut dépendre d'une usure des condyles.

les loups et les chiens
 trace : 8.243e-4
 rang : 1 2 3 4 5 6
 lambda : 5 2 1 0 0 0 e-4
 taux : 6521 2554 608 211 83 24 e-4
 cumul : 6521 9074 9683 9893 9976 10000 e-4

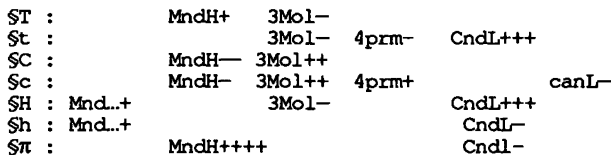
SIGI	QLT	PDS	INR	F 1 CO2 CTR	F 2 CO2 CTR	F 3 CO2 CTR	F 4 CO2 CTR
ST	966	134	41	14 787 49	3 48 8	1 3 2	6 129 251
\$t	924	128	56	5 83 7	15 655 142	8 183 167	1 2 5
{C	993	148	134	-23 735 151	-12 199 104	0 1 2	-6 58 368
{c	991	143	168	-27 769 198	-10 117 77	10 95 261	3 10 76
{H	987	167	129	-2 11 2	24 938 475	-3 22 47	-3 16 96
{h	992	117	72	-14 416 46	-6 92 26	-14 438 521	5 46 157
{π	999	163	400	42 890 547	-14 107 167	0 0 0	-2 2 47

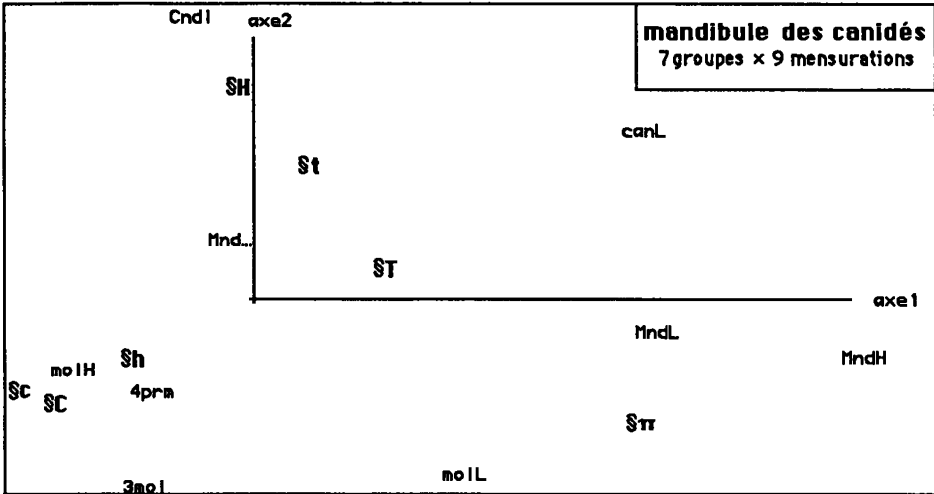
SIGJ	QLT	PDS	INR	F 1 CO2 CTR	F 2 CO2 CTR	F 3 CO2 CTR	F 4 CO2 CTR
MndL	999	449	54	-4 217 18	7 475 101	-5 275 246	2 31 81
MndL	975	34	84	44 945 121	-3 7 2	4 9 12	-5 13 51
CndI	999	75	129	-8 57 11	33 749 379	16 175 372	-5 18 114
MndH	999	73	399	66 978 599	-6 10 16	6 8 54	4 3 59
molH	974	71	53	-22 830 68	-8 104 22	1 1 1	-4 39 97
molL	884	28	42	22 380 25	-22 384 64	-4 16 11	-11 104 209
3mol	995	118	110	-14 261 44	-23 717 308	1 2 3	-3 15 80
4prm	987	131	61	-13 484 45	-10 291 70	7 134 135	6 79 231
canL	995	21	67	42 673 69	20 145 38	-19 152 166	-8 25 78

Enfin, pour faire pendant à l'analyse préliminaire du §1.2.0, nous donnons, avec des CAH, l'analyse du tableau 7x9 dont chaque ligne est le cumul de tous les individus d'un groupe.

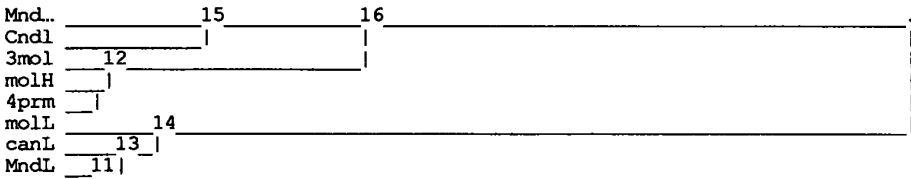


Dans la CAH des espèces (ou demi-espèces), le groupe préhistorique, {π, est isolé; les deux sexes des chacals {{C {c} sont agrégés; comme le sont ceux des chiens {{T \$t}, bien qu'on voie, dans le plan (1,2), un décalage; mais {H





va avec les chiens, §h avec les chacals: ainsi, Loups et louves sont séparés: l'étalement de l'espèce a déjà été commenté.



Les variables {MndH CndI 3Mol}, chacune présente dans l'une des trois classes principales de variables, {j15 j12 j14}, apportent, à la CAH des espèces, les principales contributions.

1.3 Données crâniométriques issues de sépultures égyptiennes antiques

A. THOMSON & R. RANDALL-MACIVER: *Ancient Races of the Thebaid*; Oxford University Press; (1905).

1.3.1 Description des données

Les données concernent des crânes humains, de sujets présumés de sexe mâle, relevés en Égypte, dans des sépultures antiques de la région de Thèbes.

L'ensemble I des crânes se répartit, suivant la période, en un ensemble Is de 5 séries de 30: Is = {<1 (2 =3)4 +5}:

$$<1 \approx -4000 ; (2 \approx -3300 ; =3 \approx -1850 ;)4 \approx -200 ; +5 \approx 150$$

La première série, dont les individus seront notés sur le graphique par le sigle '<', remonte au prédynastique ancien: soit quelque 4000 ans avant l'Incarnation. La série 2, '(', est de la fin du prédynastique: -3300. La série 3 se situe vers -1850, sous les douzième et treizième dynasties. La série 4, '=', est de l'époque des Lagides. La dernière, 5, '+', +150, remonte à la domination impériale romaine.

Sur chaque crâne, est relevé un ensemble J de 4 mesures. Dès 1973, des données crâniométriques beaucoup plus précises ont été soumises à l'analyse des correspondances; et l'on trouve dans les travaux de Louis BELLIER, (in *Traité*, TI Cn°5, n°6,) une critique pertinente du choix des mesures. Ici, nous disposons de $J = \{Lr Hh Ba Nz\}$; soit: Largeur maxima; Hauteur de la base du crâne au bregma; longueur Basi-alvéolaire; hauteur du Nez. Sur les quatre dernières mesures, nous n'avons qu'un schéma, que nous avons tenté de reproduire sur le plan (1,2).

Ainsi, on a un tableau de mensurations $I \times J$, $(5 \times 30) \times 4$; auquel s'ajoute un ensemble I_s de 5 lignes, dont chacune est le cumul des lignes descriptives afférentes aux 30 individus d'une série.

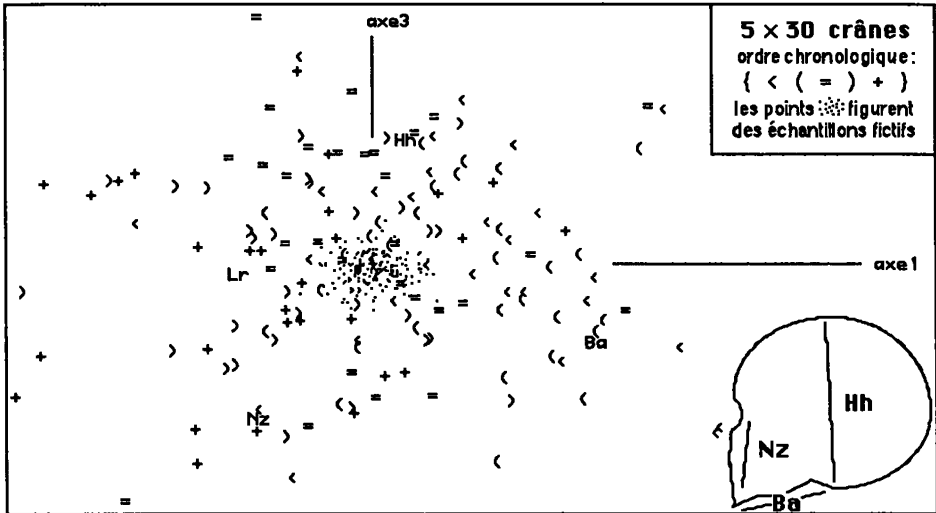
1.3.2 Variation diachronique de la forme des crânes

La structure même de l'échantillon I, atteste que les données ont été rassemblées pour déceler une éventuelle variation diachronique dans la forme des crânes. Or cette variation est très nette: elle apparaît dès l'analyse du tableau $I \times J$, pris en principal, d'après la place de l'ensemble I_s des éléments supplémentaires.

crânes, Égypte :	5 périodes;			
trace :	1.481e-3			
rang :	1	2	3	
lambda :	7	4	4	e-4
taux :	4835	2730	2436	e-4
cumul :	4835	7564	10000	e-4

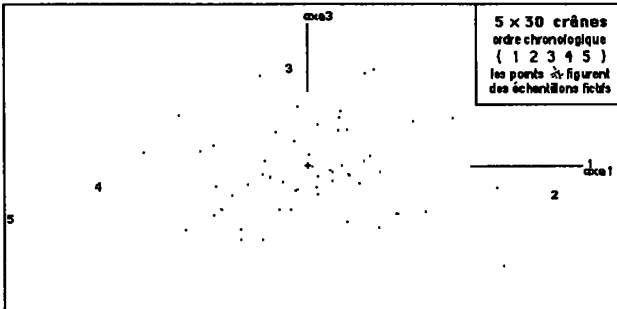
L'opposition entre prédynastique, $\{<1 (2)\}$, et postdynastique, $\{)4 +5\}$, se lit exactement sur l'axe 1, avec des CO2 (contributions relatives, i.e. coefficients de corrélation au carré) supérieurs à 929. La série du Moyen Empire, =3, sort sur l'axe 3; en sorte que l'image du plan (1,3), issu de $I \times J$, ne diffère pas de celle qu'on obtiendrait dans le plan (1,2) issu de $I_s \times J$. Au cours des millénaires, Ba tend à diminuer relativement à la largeur Lr.

Les différences entre séries, sont hautement significatives, comme on peut s'en assurer en projetant un nuage Iha de 200 échantillons fictifs, représentés



chacun par le cumul des lignes de 30 individus tirés au hasard. Sur l'axe 1, tous ces échantillons, sauf un (projeté entre 4 et 5), se projettent dans l'intervalle compris entre {5 4} et {2 1}. De plus, sur l'axe 3, deux échantillons seulement ont une abscisse supérieure à celle de la série 3.

SIGJ	QLT	PDS	INR	F 1	CO2	CTR	F 2	CO2	CTR	F 3	CO2	CTR
ci-dessous élément(s) supplémentaire(s)												
<1	1000	200	42	17	956	84	-3	44	7	0	0	0
(2	1000	200	32	15	975	64	2	11	1	-1	13	2
=3	1000	200	5	-1	35	0	1	17	0	6	948	21
)4	1000	200	24	-13	954	47	-2	36	3	-1	9	1
+5	1000	199	49	-18	929	94	4	41	7	-3	30	6
SIGJ	QLT	PDS	INR	F 1	CO2	CTR	F 2	CO2	CTR	F 3	CO2	CTR
Lr	1000	324	253	-26	619	324	21	378	350	-1	3	3
Hh	1000	320	173	6	39	14	-12	207	131	25	755	535
Ba	1000	233	325	42	847	570	9	43	51	-15	110	147
Nz	1000	123	250	-23	180	93	-39	513	469	-30	307	315



1.3.3 Essai de reconnaissance de la période d'un crâne d'après sa forme

Il n'y a pas lieu d'attendre que les différences de forme entre séries soient d'une telle amplitude que la période dont provient un crâne puisse être connue avec certitude d'après quatre mesures. Il vaut cependant la peine de considérer les résultats d'une analyse discriminante subordonnée à l'analyse factorielle dont on vient de rendre compte.

Chaque élément i de I est affecté à la série is du centre de laquelle il est le plus proche (dans l'espace des profils, muni de la distance du χ^2 ; i.e., dans l'espace euclidien où F_1, F_2, F_3 font un système de coordonnées orthonormées).

affectations de I à Is par A_f
en lignes, les centres de Is ;
en col. les classes d'individus

	5	<<<	(((=))	+++
c1	13	9	5	2
c2	8	8	4	4
c3	3	7	13	5
c4	4	5	3	7
c5	2	1	5	12

Affectation selon la
distance de Mahalanobis
(mêmes notations qu'en A_f)

	5	<<<	(((=))	+++
c1	12	10	4	3
c2	8	8	4	4
c3	4	5	15	5
c4	4	4	2	5
c5	2	3	5	11

Sur le tableau cumulatif, on voit, e.g., que 8 des individus de la série 1, '<', ont été attachés au centre c2 de la série 2; 12 individus, '+', de la série 5, ont été correctement affectés au centre c5...

D'après la distance de Mahalanobis, on obtient une matrice d'affectations peu différente. Le total des termes diagonaux (affectations exactes) est 51, contre 53 ci-dessus. Les erreurs graves, affectations du prédynastique au postdynastique ou réciproquement, (échanges entre {1 2} et {4 5},) sont de 26, au lieu de 24 ci-dessus.

Ces différences numériques sont insignifiantes; il importe davantage que l'ensemble des résultats intéressants s'ordonne dans l'espace des profils, muni de la distance du χ^2 et rapporté aux axes que détermine l'analyse des correspondances.