

B. ABI-BOUTROS

L. BELLIER

**Contribution à la taxinomie des micromammifères.
Application au genre *Crocidura***

Les cahiers de l'analyse des données, tome 2, n° 4 (1977),
p. 435-450

http://www.numdam.org/item?id=CAD_1977__2_4_435_0

© Les cahiers de l'analyse des données, Dunod, 1977, 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/>

CONTRIBUTION A LA TAXINOMIE DES MICROMAMMIFÈRES.

APPLICATION AU GENRE *CROCIDURA*

[CROCIDURES]

par B. Abi-Boutros ⁽¹⁾
et L. Bellier ⁽²⁾

1 Données et buts de l'étude

L'Office de la Recherche Scientifique et Technique Outre-Mer (O.R.S.T.O.M.) se consacre, entre autres, aux recherches d'écologie animale, qui peuvent aider au progrès de l'agriculture et de l'élevage.

Les données concernant plusieurs centaines de petits insectivores SORICIDES appartenant au groupe "*Crocidura*" ou Musaraignes, ont été recueillies sous la direction de M. le Professeur L. Bellier dans les stations de l'O.R.S.T.O.M. réparties dans les différentes régions de la Côte d'Ivoire (Adiopodoumé, Abengourou, Lamto...).

Les individus étudiés se répartissent en deux groupes : un groupe de base au nombre de (238) individus dont le nom de l'espèce est déterminé, et un second groupe de (350) individus anonymes. Dans le premier groupe le nombre d'animaux est très différent d'une espèce à l'autre, ainsi nous avons

L'espèce	Nombre d'individus	Sigle
<i>Crocidura occidentalis</i>	128	O
<i>Crocidura theresae</i>	58	T
<i>Crocidura juvenetae</i>	21	J
<i>Crocidura poensis</i>	15	P
<i>Crocidura lamottei</i>	5	L
<i>Crocidura giffardi</i>	2	G
<i>Crocidura bottegi</i>	3	B
<i>Crocidura lusitania</i>	1	U
<i>Crocidura wimmeri</i>	1	W
<i>Sylvisorex megalura</i>	4	S

(*) Le présent article expose les résultats de la thèse de M. B. Abi-Boutros (3^e cycle Paris 1977) "Contribution à la taxinomie mathématique des micromammifères : application au genre *Crocidura*". Les lecteurs intéressés par des résultats numériques complets et des graphiques détaillés se reporteront à cette thèse. Quant aux données elles sont en possession de L. Bellier, Docteur ès Sciences, Directeur à la Société Euréquip, (19, rue Yves du Manoir, 92420 Vaucresson) qui en a assuré la collecte alors qu'il travaillait au centre d'Adiopodoumé (Côte d'Ivoire) et en a dirigé les analyses publiées ici.

C'est le groupe de base que nous allons étudier d'abord.

Nous disposons de (15) mesures prises sur le crâne et de (5) sur la mandibule ; les informations complémentaires dont on dispose généralement sont (outre le nom de l'espèce si l'animal a été déterminé) : la localité ou la région où il a été capturé, le sexe, et les trois mesures prises sur le corps de l'animal fraîchement tué ; pour un certain nombre de spécimens ces dernières données manquent. Chaque individu a un numéro de terrain et un numéro d'analyse. (Le schéma explicatif des mesures est placé à la fin du § 2).

Les régions de la Côte d'Ivoire où les Crocidures ont été capturées en nombre assez important sont :

Région	Sigle	Région	Sigle
Adiopodoumé	1	Abengourou	9
Lamto	2	Matonguiné	D
Divo	5	Port Bouet	Q
Mopoyem	6	Seampli	R
Dabou	7		

On a (32) régions.

Sigles ou codages : On a noté les individus en codes de (4) cases, car il faut qu'un codage rassemble le plus d'informations possibles sur ce qu'on va analyser :

Dans la 1^o case : - le sigle de l'espèce si elle était déterminée (groupe de base).

- blanc sinon

La seconde case : - le sigle de la région où l'animal a été capturé.

La 3^o case : - Féminin F, ou Masculin M.

La 4^o case : - un n^o d'ordre pour différencier les individus de chaque série formée d'animaux d'une même espèce, capturés dans une même région et du même sexe.

Les mensurations et les variables : ont été prises sur le squelette de la tête entre des points notés sur le schéma ci-contre.

Des données biométriques nous amènent naturellement au domaine de la taxinomie ; c'est-à-dire :

Au sujet des espèces connues :

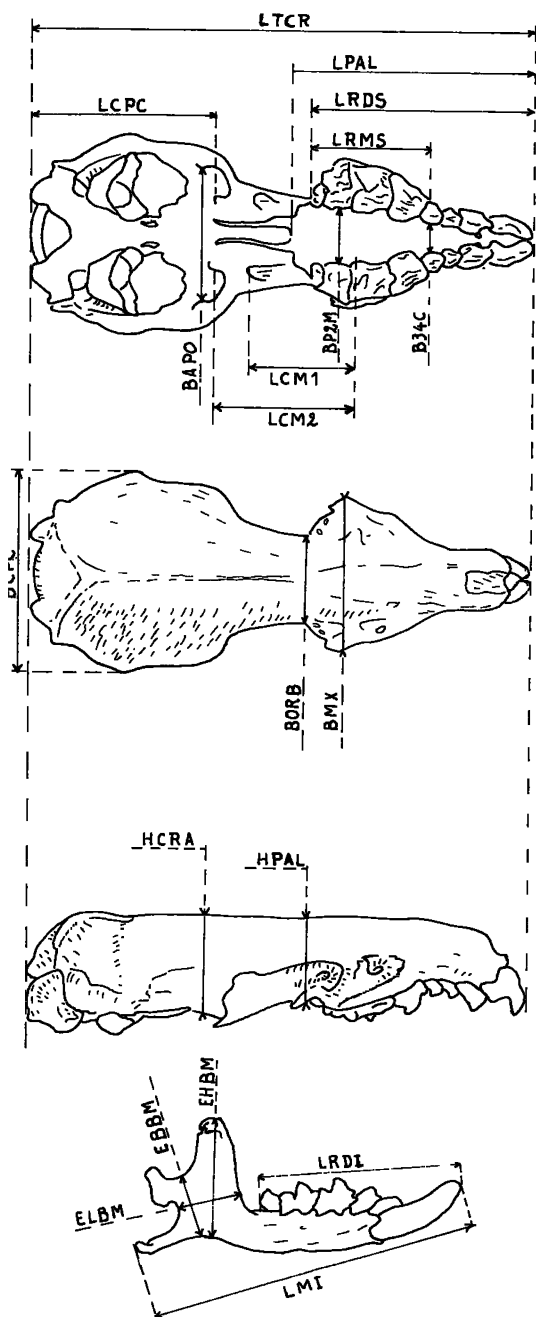
- A la différenciation des espèces
- A la mise en évidence des groupes
- A la caractérisation de chaque espèce par rapport aux autres.
- Aux caractéristiques de chaque espèce, et à ses différenciations géographiques et sexuelles.

Au sujet d'espèces ne figurant pas sur la liste des déterminations:

- A la recherche de groupes assez bien séparés des autres pour relever de nouvelles espèces.

- Enfin à la détermination d'un individu anonyme, c'est-à-dire l'identification du groupe dont les caractéristiques sont les plus proches des siennes.

En plus de tout ceci, il y a des individus parmi le groupe de base qui peuvent être d'une autre espèce que celle qui leur est attribuée (inattention ou erreur d'identification)



LTCR : longueur totale du crâne .

LCPC : longueur de la capsule crânienne .

LPAL : longueur du palais .

LRDS : longueur de la rangée dentaire supérieure .

LRMS : longueur de la rangée molaire supérieure .

LCM1 : petite longueur entre la capsule et le maxillaire .

LCM2 : deuxième longueur entre la capsule et le maxillaire .

BAPC : largeur entre les apophyses .

BP2M : petite largeur du palais entre les deux molaires .

B34C : larg. du maxillaire au niv. des dents entre les 3^e et 4^e canines .

BCPC : largeur de la capsule crânienne .

BORB : largeur entre les orbites (cons. triction interorbitaire) .

BMX : largeur du maxillaire (hors tout) .

HCRA : hauteur du crâne .

HPAL : hauteur du palais .

LRDI : longueur de la rangée dentaire inférieure .

LMI : longueur de la mandibule (hors tout) .

ELBM : longueur de la branche montante .

EBBM : largeur de la branche montante .

EHBM : hauteur de la branche montante .

L'analyse des correspondances et la classification hiérarchique nous ont permis de bien répondre à toutes ces questions ; et d'aboutir (§8) à une clé de détermination permettant de classer tout crâne d'après des calculs simples effectués sur ses mensurations (non sans avoir corrigé au passage de multiples erreurs de mesures, ou d'identification : § 3 ; § 5).

2 L'analyse du tableau initial K(I,J) où :

I = le groupe de base des (238) individus,

J = les (20) mensurations sur le crâne.

Les valeurs propres :

λ_i	τ_i	cumul
78.5×10^{-5}	50.9%	
20.1×10^{-5}	13.0%	63.9%
9.4×10^{-5}	6.1%	70.2%

Les valeurs propres sont petites, c'est dû à ce que les profils des mensurations sont très proches les uns des autres.

Le nuage des individus : les individus de l'espèce *C. occidentalis* ont tous un premier facteur positif, à quelques exceptions près (on verra au § 4 à quoi sont dues ces anomalies!).

Le premier axe discrimine entre l'espèce *occidentalis* qui a le plus grand effectif, les *giffardi* et les *lamottei* d'un côté ; de l'autre côté s'échelonnent, du bout au centre, les *bottegi*, les *Sylvisorex*, qui sont assez différentes des autres espèces pour former leur nuage du côté négatif du 1^o axe, ensuite les *theresae* qui se confondent avec les *poensis*, les *wimmeri*, *lusitania* et les *jouvenetae*.

On a donc trois sous-nuages dans le plan (1,2).

Cette analyse nous a signalé les choses suivantes

- Trois des (5) *lamottei* se groupent ensemble du côté des *occidentalis*, alors que les deux autres notés (L2M2, et L2M3) sont du côté des *theresae* et bien distants des premiers.
- Un des *jouvenetae* noté (J5F1) se groupe avec les *bottegi*.
- Un autre noté *jouvenetae* (J2M1) se groupe avec les *Sylvisorex*.

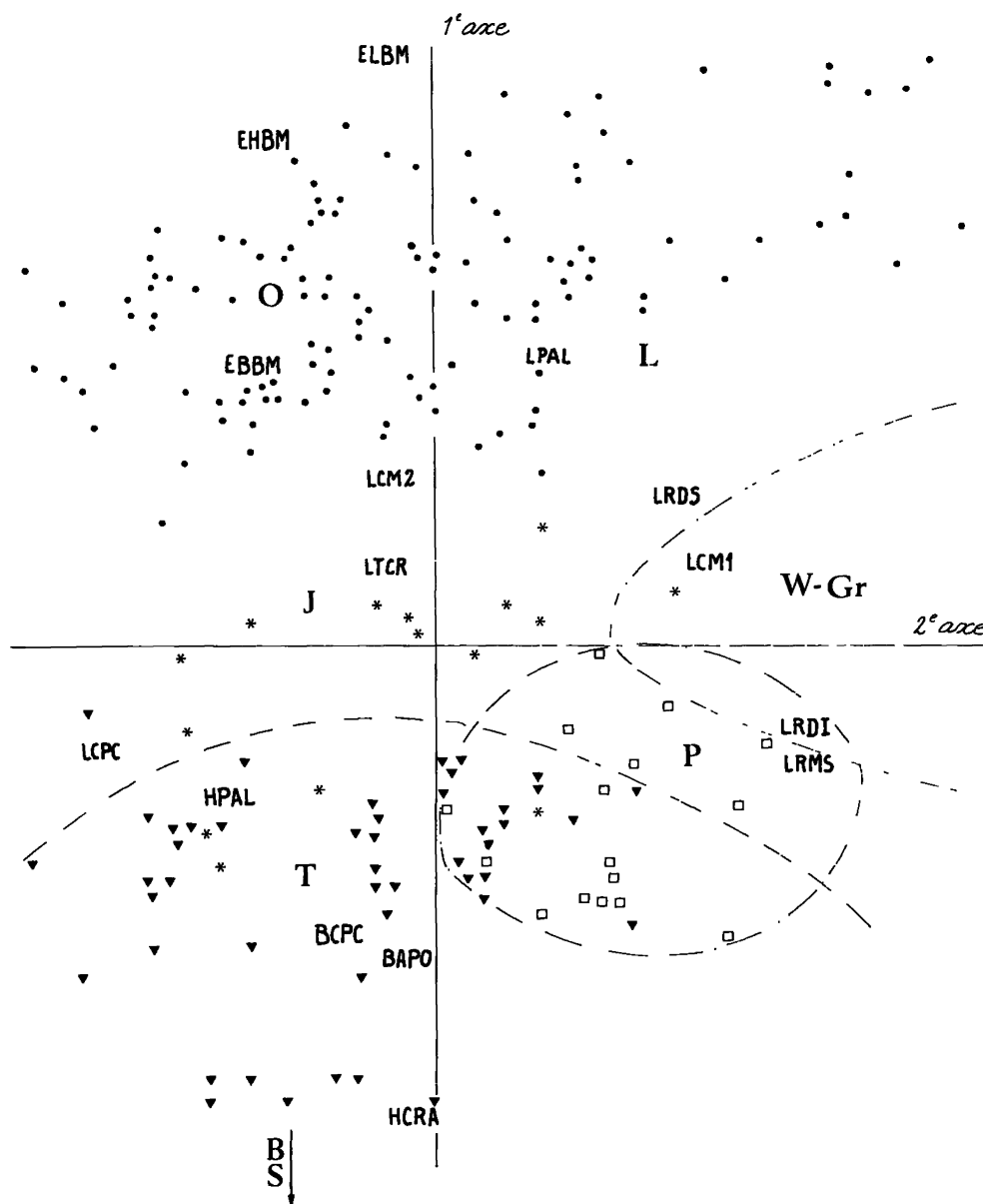
Alors que rien pour le moment ne permet de distinguer entre *theresae*, *poensis*, et *jouvenetae*, *wimmeri* et *lusitania*.

On doit porter les squelettes de la tête et les peaux de ces exemplaires au spécialiste mondial des Crocidures M. Heim de Balzac pour l'identification authentique (cf § 3 *in fine*).

On peut expliquer le premier facteur de deux façons complémentaires : la première est que c'est un facteur de forme lié à la taille : En effet les *bottegi* sont les plus petits, suivent les *Sylvisorex*, ensuite le groupe formé des *poensis*, *theresae* et *jouvenetae*, ensuite les *lamottei* et les *occidentalis* qui sont les plus grands.

Toujours est-il qu'il y a quelques exceptions :

- Les *jouvenetae* dans l'ensemble sont plus petits que les *theresae* et les *poensis*, alors que ce sont les plus proches des *occidentalis* sur le premier axe.



Plan 1-2 de l'analyse pondérée : les individus des quatre espèces O, T, J, P (O: *C. occidentalis*; T: *C. theresae*; J: *C. jouvenetae*; P: *C. poensis*) sont pondérés par l'inverse de l'effectif de leur espèce ; on les a représentés par les sigles : • (O) ; ▼ (T) ; * (J) ; □ (P) . la lettre L indique la place des *C. lamottei* ; on a esquissé par leurs contours les domaines (empruntés sur ce plan) des espèces T et P, ainsi que le domaine approximatif W-Gr de l'espèce *C. wimmeri* et groupes adjacents très peu représentés dans l'échantillon ; une flèche signale les espèces *C. bottegii* et *Sylvosorex m.* très écartées à l'extrémité inférieure de l'axe .

Les *giffardi*, qui sont les plus grands en taille et dépassent de beaucoup les *occidentalis*, se confondent avec ceux-ci et n'ont pas un premier facteur discriminatoire.

La seconde façon : c'est un facteur qui discrimine entre les deux espèces de plus grands effectifs (poids) : les *occidentalis* et les *theresae* ; ensuite entrent en jeu les espèces de plus petits effectifs par leurs différenciations ou leurs similitudes aux premiers.

Le second axe : On remarque que les individus de l'espèce *poensis* sont d'un seul côté de cet axe, les *Sylvisorex* de l'autre, alors que les individus des autres espèces se mettent indifféremment sur cet axe.

Pour les espèces qui se confondent dans cette analyse il vaudra mieux les soumettre à part à l'analyse factorielle, deux à deux ou ensemble (§4).

Les variables et les caractérisations des espèces les unes par rapport aux autres :

Le premier axe : D'un côté, du bout au centre, on a : la longueur et la hauteur de la branche montante et la longueur du palais (ELBM, EHBM, LPAL), dont les profils sont supérieurs chez les *occidentalis*, *giffardi* et *lamottei*, par rapport aux autres espèces.

De l'autre côté de cet axe, de l'extrémité au centre, on a : la constriction interorbitaire, la hauteur du crâne (BORB, HCRA) qui vont avec les *bottegi* *Sylvisorex*, et qui vont aussi avec les *theresae* ; la largeur du crâne entre les apophyses, et la hauteur du palais (BCPC, BAPO, HPAL) dont les profils sont plus grands chez les *jouvenetae*, *poensis* et *theresae* que chez les *occidentalis*.

Dans le plan (1,2), on peut dire que les *poensis* se caractérisent par les profils relativement forts pour les longueurs de la rangée dentaire inférieure et de la rangée molaire supérieure (LRDI, LRMS) et par des profils faibles pour la longueur de la capsule crânienne (LCPC).

Les *lamottei* sont bien représentés dans le plan (2,3) et sont groupés avec (LCM1, LCM2) les deux longueurs prises entre des points fixes sur la capsule et le maxillaire.

Le 4° axe oppose, à peu près, les *theresae* d'un côté aux *jouvenetae* et *poensis* de l'autre, qui sont caractérisés ainsi par des profils faibles pour la petite largeur du palais entre les molaires et la largeur du maxillaire (BP2M, BMX) qui forment le 4° axe (51% de son inertie).

Note : S'il y a plus que 3 espèces (dans notre cas) avec des effectifs très différents, cela fait tourner les individus d'une espèce de faible effectif autour du nuage de l'espèce la plus proche (similaire) de grand effectif ; et ceci élimine la discrimination entre ces deux espèces, comme dans le cas des *jouvenetae* (20) et des *theresae* (60).

Pour y remédier (note de M. J.P. Benzécri) il faut soumettre à l'analyse factorielle les individus pondérés par l'inverse de l'effectif de leur espèce ($1/Ne(i)$), et ceci pour les espèces ayant au moins une vingtaine d'individus (puisque il faut que l'échantillon de chaque groupe soit représentatif) ; les individus des autres espèces ayant moins d'une vingtaine d'individus et les individus anonymes pourront être mis en éléments supplémentaires.

En effet ceci aide à équilibrer les poids, et les contributions à l'inertie du nuage, des espèces les unes par rapport aux autres ; et le principe de l'équivalence distributionnelle permet de faire cette démarche.

3 La classification automatique sur les (238) individus du groupe de base muni de leurs (20) mensurations crâniennes.

Nous allons utiliser la classification ascendante hiérarchique, avec pour distance celle du χ^2 , et pour critère d'agrégation la maximisation du moment centré d'ordre deux d'une partition.

Les niveaux d'agrégation des noeuds sont très petits de l'ordre de (10^{-4}) comme les valeurs propres de l'analyse factorielle (on sait que les v. p. majorent les niveaux des noeuds) ; puisque les données sont de nature biométrique : les profils des individus (resp. des variables) sont très proches les uns des autres.

Nous désignons par A(I) (ainé de I) et B(I) (benjamin de I) les deux classes supérieures de l'arbre en lesquelles se divise I ; et de même AA(I) (ainé de l'ainé) BA(I) (benjamin de l'ainé) ; etc ; les subdivisions successives de ces classes.

L'ainé A(I) contient les *occidentalis*, les 2 *giffardi* et les 3 *lamottei* qui forment leur noeud de n° d'ordre 459 (le n° d'ordre du noeud (I) des 238 individus est : $(2 \text{ Card}(I) - 1) = (475)$).

Le benjamin B(I) contient toutes les autres espèces :

- AB(I) : Les *theresae* forment son ainé AAB(I), d'effectif (61), qui contient avec les *theresae* les deux éléments notés *lamottei* (L2M3, L2M2), et des éléments notés *poensis* (P11M, P13F, PW1F) et un élément noté *jouvenetae* (JW2) ; de même que l'exemplaire des *lusitania* (UJM1).

- BAB(I) contient les *bottegi* qui en forment l'ainé et se joint à eux l'élément noté *jouvenetae* (J51F), et les *Sylvisorex* et se joint à leur classe l'élément (J21M) confirmant ainsi les résultats de l'analyse factorielle.

- BB(I) : contient d'une part : les *poensis* et avec eux 3 éléments notés *theresae* (T6CM), T25F, T64F) et un élément noté *jouvenetae* (J42F) ainsi que l'exemplaire de l'espèce *C. wimmeri* (W8F1) ; et d'autre part les *jouvenetae* et se joignent à eux les éléments (P12M) noté *poensis* et (T24F) noté *theresae*, et un groupe bizarre (P14F, T21M, OV1M).

Résultats : On a apporté ces individus qui se classent avec d'autres groupes que les leurs au spécialiste mondial M. Heim de Balzac, et l'on a eu les résultats suivants :

- les deux *lamottei* (L2M2, L2M3) qui se mettent avec les *theresae* sont identifiés comme tels et nous les noterons : (T2M2, et T2M3).

- (J2M1), qui se met avec les *Sylvisorex*, est identifié comme tel et renomé (S2M4).

- (J5M1), qui se joint aux *bottegi*, est identifié comme étant un *bottegi* et renomé (B5F1).

Ces individus ont été détectés par la classification automatique et par l'analyse factorielle.

- (T2F5) se groupant avec les *poensis* sera identifié comme tel et renomé : (P2F1).

- (P1M2) se groupant avec les *jouvenetae* est un *jouvenetae* et renomé (J1M5).

Les deux derniers individus ont été détectés par la classification automatique.

Alors que les autres individus qui se groupent par la classification avec d'autres espèces que celles qu'on leur avait attachées sont :

- (T2M1) non retrouvé ; (P1M4) dont le crâne est abîmé ; (P1M1) dont on a retrouvé la peau mais non le crâne ; et (T6F4) non retrouvé.

Il doit s'agir soit de mesures douteuses (erreur de mensuration) soit d'erreurs de perforation :

- dans le chiffre des centaines, qu'on peut rectifier facilement.
- dans le chiffre des dizaines, que l'on peut arriver à rectifier après analyse des données.
- dans le chiffre des unités, qu'on ne peut pas détecter car : elles entraînent pour le profil des variations de faible amplitude qui peuvent transformer le profil réel en un profil vraisemblable.

Pour y remédier, il faut remesurer les individus dont on a retrouvé le crâne, c'est ce qui est fait pour : (J42F, P14M, T24F et OVM1).

4 Les analyses partielles :

Nous avons fait des analyses factorielles sur les espèces, ensemble ou deux à deux, pour celles ayant un effectif supérieur à une vingtaine d'individus que l'analyse globale avait superposés. Celles des analyses sur les tableaux de mesures brutes des individus des espèces prises deux à deux, nous ont donné des résultats assez bons, avec des pourcentages d'erreurs de classement descendant jusqu'à (2/20) soit (10%) ; tandis que quand on soumet à l'analyse trois espèces (ou plus), leurs individus se confondent, et surtout ceux des trois espèces les plus proches : *jouvenetae*, *theresae* et *poensis* ; alors que les analyses factorielles sur les individus pondérés par l'inverse de l'effectif de leur espèce, se sont avérées très utiles dans la discrimination entre les espèces, avec des pourcentages d'erreur de classement descendant jusqu'à (5%), soit (95% de pourcentage de bien classés).

- *theresae* et *poensis* : $\lambda_1 = 20.1 \times 10^{-5}$; $\tau_1 = 30.0\%$.

Une bonne séparation se fait sur le premier axe, l'hyperplan discriminant est orthogonal au 1^o axe au centre du nuage, (T2F5) renoté (P2F1) se met avec les *poensis* et un des *poensis* (PWF1) se met à la frontière des *theresae*.

- *theresae* et *jouvenetae* : $\lambda_1 = 33.0 \times 10^{-5}$; $\tau_1 = 37.1\%$

se discriminent de la même manière.

- *poensis* et *jouvenetae* : on a eu une séparation sur le 1^o axe, (P1M2) renoté (J1M5) se met avec les *jouvenetae*, et (J4F2) se met à la frontière des *poensis*.

- *theresae*, *jouvenetae* et *poensis* :

$$\lambda_1 = 23.6 \times 10^{-5} \quad ; \quad \tau_1 = 27.7\%$$

$$\lambda_2 = 15.7 \times 10^{-5} \quad ; \quad \tau_2 = 18.5\%$$

Dans le plan (1,2), d'un côté autour du 1^o axe : les *jouvenetae* ; de l'autre côté les *theresae*, alors que les *poensis* forment leur groupe d'un côté du second axe, à côté des premières espèces ; la séparation entre les trois espèces est beaucoup plus nette que dans l'analyse sur le tableau des mesures brutes de leurs individus.

- *theresae*, *jouvenetae*, *poensis* et *occidentalis* :

$$\lambda_1 = 60 \times 10^{-5} \quad ; \quad \lambda_2 = 17.5 \times 10^{-5} \quad ; \quad \lambda_3 = 13.3 \times 10^{-5}$$

$$\tau_1 = 45.3\% \quad ; \quad \tau_2 = 13.2\% \quad ; \quad \tau_3 = 10\%.$$

Les *occidentalis* se séparent, sur le premier axe, des autres espèces, qui gardent elles la même disposition décrite plus haut. Et en mettant les individus des autres espèces que celles-ci, elles auront les

mêmes dispositions que dans l'analyse principale sur le groupe de base (i.e. les *Sylvisorox* et les *bottegi* se placent à l'extrémité du premier axe, du côté opposé aux *occidentalis*).

Nous ferons une classification sur l'ensemble des éléments dont nous disposons : les (238) identifiés et les (350) anonymes, soit (588) individus, pour délimiter les groupes des espèces ; ensuite nous soumettrons à l'analyse factorielle les groupes obtenus par la classification, deux à deux ou ensemble, pour avoir des discriminations plus ou moins parfaites entre les espèces, et pour essayer de détecter, s'il y a lieu, d'autres espèces s'il en existe.

Etant donné que la classification ascendante hiérarchique donne des limites aux groupes, et que les analyses partielles (avec pondération ou non) discriminent entre les espèces, on pourra mettre un nouvel individu (anonyme) en élément supplémentaire, dont on peut calculer les facteurs directement à partir des facteurs des variables (par la formule de transition).

5 Mise au point des données

La classification hiérarchique nous a donné des limites peu précises des groupes, puisque des *poensis*, des *jouvenetae* et des *theresae* du groupe de base s'agglomèrent les uns avec les autres ; nous avons procédé alors à des analyses partielles sur les groupes obtenus pris deux à deux, où le chaos était prédominant, mais nous avons remarqué, par exemple, qu'un individu noté (DMIV) (où le V indique toujours que c'est un individu anonyme) ne se mettait avec aucun groupe dans ces analyses ; tout ceci nous a donné la preuve qu'il existait encore des erreurs dans les mesures du tableau.

On a remarqué que les profils sont à la base des différenciations entre les individus et les espèces du groupe "*crocidura*", et en général des genres et espèces animales.

Compte-tenu de ce que les individus ont des profils très voisins les uns des autres (les valeurs propres sont très faibles) on a approximativement la relation :

$$k(i,j) \# k(i) \cdot k(j) / k$$

On a donc créé le tableau des C_{ij} :

$$C_{ij} = k(i,j) \cdot k / (k(i) \cdot k(j)) \# 1$$

L'examen direct et l'analyse factorielle du tableau des C_{ij} permet de mettre en évidence les erreurs qui se manifestent par des C_{ij} anormalement écartés de 1.

Il faut dire que l'on mise sur les erreurs dans le chiffre des dizaines d'une mesure, puisque comme M. Bellier nous l'a expliqué ensuite l'on peut se tromper dans la lecture des millimètres (10 unités de mesure dans notre cas) en mesurant les crânes. Or de telles erreurs corrompent les résultats de l'analyse factorielle et de la classification.

L'analyse factorielle fait surgir quelques anomalies, mais pour les détecter toutes il faut examiner tous les individus sur presque tous les axes, et parfois ils se dissimulent ; en outre il ne faut pas toucher aux chiffres des unités, mais il faut toujours revenir au tableau initial pour vérifier les mesures et pour corriger éventuellement les erreurs.

Il faut faire des comparaisons entre les mesures semblables de l'individu et les mesures d'autres individus ayant la même taille, puisque globalement les mesures d'individus qui ont la même taille, ont un

intervalle de variabilité ne dépassant pas une dizaine d'unités. Il y a des relations entre les mesures d'un même individu, par exemple : (LRDS, LCM1, LCPC) les longueurs de la capsule, de la rangée dentaire supérieure et celle entre la capsule et le maxillaire, la somme de ces 3 mesures équivaut à la longueur totale du crâne (LTCT) avec une marge allant jusqu'à (5 à 7) unités de moins ; la différence entre deux mesures longitudinales prises dans le même sens comme (LPAL, LRDS) ne dépasse pas un certain nombre d'unités, de même (LRDS, LRMS), (ELBM, LMI, LRDI), la variabilité de la hauteur du crâne, de la hauteur du palais, et de leurs différences.

Etant donné que le problème concerne une dizaine en plus ou en moins et non des unités, on n'a en général qu'une seule solution.

Résultats : On a trouvé dans l'échantillon global une quinzaine d'individus ayant une erreur d'une seule dizaine en plus ou en moins, et deux ayant une erreur de deux sur le chiffre des dizaines.

Dans le groupe de base on a : (T2M3) et (OVM1, OEM1) les seuls *occidentalis* qui se plaçaient avec les autres espèces sur le premier axe de l'analyse du groupe principal, (ODM1), (JW2) dont les mensurations ne sont pas homogènes ensemble, et (J1M1).

6 L'étude de l'échantillon entier formé du groupe de base et des individus anonymes

L'analyse factorielle de l'échantillon entier K(I,J) où I = 588 individus.

Les valeurs propres :

$$\lambda_1 = 70.8 \times 10^{-5} \quad ; \quad \tau_1 = 45.9\%$$

$$\lambda_2 = 22.2 \times 10^{-5} \quad ; \quad \tau_2 = 14.4\%.$$

La forme globale du nuage du groupe initial est conservée, mais on voit dans le plan (1,2) une péninsule supplémentaire formée d'individus anonymes, du côté positif du second axe, qui laisse prévoir l'existence d'un nouveau groupe, proche de l'espèce *C. zimmeri*, puisque (W8F1) est à ses côtés.

Les groupes extrêmes se placent : d'un côté du premier axe les *occidentalis* L. G., et de l'autre côté au bout : les *Sylvisorex* et les *bottegi* ; alors que toutes les autres espèces se confondent entre ces deux groupes extrêmes. Il est facile et nécessaire de mettre à part celles-ci avec les individus anonymes qui s'agglomèrent avec elles, pour se consacrer à l'étude du groupement enchevêtré, et de chacun de ces groupes à part aussi.

Remarque : En soumettant à l'analyse des espèces très proches (i.e. ayant des profils presque égaux de leurs éléments comme les *theresae*, les *jouvenetae*, les *poensis*,...), il est évident qu'il faut mettre à part ou en élément supplémentaire, tout individu assez différent d'eux (i.e. relevant d'autres groupes ou espèces : *occidentalis*, ou *bottegi* par ex.), puisqu'il va se comporter comme élément aberrant qui bouleversera l'analyse et sortira, peut-être du nuage de ces groupes, vu la finesse de l'analyse factorielle.

1°) Les *occidentalis*, *lamottei* et *giffardi* (O.L.G.).

La classification automatique sur les (588) individus a groupé avec les *occidentalis* L.G. les mêmes individus anonymes que l'analyse factorielle précédente (une centaine) à 6 éléments près, elle a fourni aux groupes des autres espèces des limites assez fiables, et les individus anonymes ont été triés et mis avec les espèces aux nuages desquels

ils s'incorporent ; on a donc divisé l'échantillon global en parties assez homogènes pour faciliter le déroulement des analyses partielles ultérieures (cf la remarque plus haut).

- Les (O.L.G.) se séparent bien du groupe des *theresae* formé par la classification.

- On a eu de même, une séparation très nette avec le groupe des *poensis* ; ces deux groupes étant bien loin du groupe (O.L.G.) dans l'analyse factorielle globale.

- La séparation avec le groupe des *jouvenetae*, n'est devenue aussi nette que les précédentes qu'après avoir pondéré les *jouvenetae* identifiés par (1/17.) et les *occidentalis* par (1/126.), tous les autres individus anonymes étant mis en élém. sup., même les (3) *lamottei* et les deux *giffardi* (tous en mesures brutes).

- Nous avons parlé de l'analyse des individus pondérés des 4 espèces : *occidentalis*, *theresae*, *poensis* et *jouvenetae* ; maintenant on fait cette analyse et on met tous les autres individus en élém. sup., et l'on obtient une quasi-séparation entre les espèces avec leurs groupes d'individus anonymes dans le plan (1,2), meilleure que dans l'analyse globale précédente ; on va utiliser les résultats de cette analyse pour identifier un nouvel individu, puisqu'on pourra connaître sa place, les deux premiers facteurs étant calculés rapidement, dans le plan (1,2) ; et le classer par une méthode d'élimination.

- On a remarqué l'existence d'un individu ayant une grande taille (346) noté (WWV), on a fait alors l'analyse des 2 *giffardi* et de (3) *occidentalis*, les *occidentalis* identifiés ou non étant mis en élém. sup. : on trouve que (WWV) est un *occidentalis* et les (2) *giffardi* se retranchent du reste du nuage d'un côté du 1° axe.

- La classification nous a fourni le groupe des *lamottei* formé des (3) identifiés et de (3) nouveaux ; et l'analyse factorielle des (3) *lamottei* avec (4) *occidentalis*, tous les autres individus étant mis en élém. sup., fait retrancher les (6) *lamottei* du reste du nuage d'un côté du 1° axe. Cela donne de bons résultats comme on le voit, mais on ne peut pas s'en contenter puisque les effectifs sont faibles : les chercheurs qui travailleront sur ces données auront la possibilité de soumettre à l'analyse factorielle ces deux espèces quand on disposera d'un échantillon plus nombreux de *lamottei* identifiés.

2°) Les espèces *bottegi*, *Sylvisorex* et *lusitania* (B.S.U.).

Un noeud de la classification englobe d'une part les (4) *bottegi* avec (12) nouveaux individus, et d'autre part les (5) *Sylvisorex* avec (11) individus anonymes ; et (UJMI) s'agglomère avec (WZIV et LMTV).

Dans l'analyse factorielle globale les *theresae* sont les plus proches des (B.S.U.), on les a soumis ensemble à l'analyse : une bonne séparation se fait alors sur le premier axe, entre les (B.S.U.) et les *theresae* ; avec (UJMI) s'agglomèrent quatre individus dont deux nouveaux, on peut dire que ce sont des *theresae* ayant une mesure douteuse les rapprochant du représentant des *lusitania*.

On voit qu'il existe plusieurs groupes (espèces?) chez les *bottegi* (au moins deux), et deux individus (NMI et 5MC) se placent à part, mais avec les *bottegi*, sur le 3° axe.

- L'analyse des deux espèces (B.S.) ensemble, a mis les *Sylvisorex* d'un côté du premier axe, les *bottegi* de l'autre côté séparés en deux parties sur le second axe, dont l'une ayant (NMI et 5MC) séparés de leur groupe, soit trois groupes en tout.

- L'analyse des *Sylvisorex* ne révèle pas d'existence de sous-groupes, mais il apparaît que l'élément noté (2b9) est un peu différent des autres *Sylvisorex*.

- L'analyse des *bottegi* les met en trois groupes : le premier formé de (9) éléments d'un côté du premier axe, le second de (5) éléments se met de l'autre côté et (NM1 et 5MC) se placent à l'extrémité de ce groupe.

La classification automatique et l'analyse factorielle donnent les mêmes résultats, les représentations données suffisent et nous n'avons rien à ajouter.

7 *Les poensis, wimmeri, theresae et juvenetae* (P.W.T.J.).

On a vu qu'il existe des groupes différents de ceux que l'on connaît, ayant mis à part les groupes extrêmes : (O.L.G.) et (B.S.U.) on va essayer d'explorer le nuage restant.

L'analyse du tableau des mesures brutes de ce nuage fournit (4) groupes dans le plan (1,2), les (T.P.J.) et un groupe supplémentaire ; aussi l'analyse des individus (T.P.J.) identifiés pondérés par l'inverse de l'effectif de leur espèce, les individus restants étant mis en élém. sup., fournit quatre groupes au moins dans le plan (1,2).

Revenons à la classification globale sur le tableau des données après avoir mis à part les individus ayant des fautes dans leurs mesures ou ayant des mesures trop douteuses, soit une vingtaine ; on a avec les (14) *poensis* éparpillés deux groupes formés d'individus anonymes uniquement, l'un de (9) éléments et l'autre de (11), tous ensemble forment l'aîné d'un noeud (111) éléments dont le benjamin regroupe les *wimmeri* (17 éléments avec W8F1).

Nous avons décelé que le quatrième groupe de l'analyse précédente est formé en gros de ces (3) groupes d'individus anonymes formés par la classification : les *wimmeri* et les deux autres groupes.

L'analyse du noeud englobant les *poensis* et les *wimmeri* révèle l'existence de deux groupes supplémentaires (au moins) autres que les *poensis*, puisqu'on n'a pas de séparation entre deux groupes en présence, mais une fleur à trois pétales dans le plan (1,2).

Il n'y avait pas moyen d'omettre les individus non *poensis*, mais au contraire on pourrait soustraire les *poensis* du rassemblement. On a trié ensemble les individus des noeuds ayant des *poensis* dans leur formation, et on a fait les analyses de cet ensemble respectivement avec chacun des trois groupes cités, et l'on a eu une discrimination chaque fois entre le groupe des *poensis* d'une part et chacun des autres groupes d'autre part ; et on a rectifié l'appartenance de quelques individus, soit que l'individu appartienne au nuage des *poensis*, soit qu'il se place sur le premier axe avec l'un des groupes.

On a eu enfin un groupe autonome dont on peut dire que ses individus sont des *poensis*, puisque leur nuage a des projections convexes sur les différents plans factoriels, et que les (14) *poensis* identifiés y sont éparpillés.

Après la formation du groupe des *poensis* il nous reste à voir quelle structure formeront ces individus anonymes qui ne s'agglomèrent avec aucune des espèces dont on dispose.

Leur analyse à part nous suggère l'existence de plus de deux groupes ; leur classification en fournit trois dont les *wimmeri* (avec W8F1) ; de même leur classification avec les (14) *poensis* fournit les *poensis* à part et les mêmes trois groupes ; les arbres de classification suffisent à dire qu'il s'agit de l'espèce *wimmeri* d'une part ; les deux autres groupes peuvent en être des sous-espèces ou sinon être de nouvelles espèces ; étant donné que les discriminations avec les *poensis*, et bien sûr avec toutes les autres espèces, sont nettes et plausibles, et que ces groupes se discriminent entre eux, deux à deux, par les analyses

partielles ; la réponse ne pourra être donnée qu'après vérifications par des spécialistes de ces animaux.

Les espèces theserae, juvenetae et poensis (T.P.J.).

Les individus des nouveaux groupes brouillaient les analyses. Maintenant que nous les avons identifiés, on va voir ce qu'il en résulte pour les discriminations entre les espèces (T.P.J.), deux à deux d'une part, et entre eux et les nouveaux groupes d'autre part.

On a vérifié la discrimination entre les *juvenetae* et les nouveaux groupes rassemblés, et entre ceux-ci et les *theserae* : dans cette dernière analyse on a pondéré les individus identifiés de l'espèce *C. theserae* par l'inverse de leur effectif, et les individus des nouveaux groupes par l'inverse de leur effectif aussi, les individus anonymes classés *theserae* étant mis en éléments supplémentaires : on voit bien qu'il s'agit de trois groupes dans le plan (1,2), on peut dire que ces groupes sont autonomes.

Après les deux étapes précédentes, les discriminations par l'analyse factorielle ont été nettes, avec rectification de quelques individus, entre les groupes (T.P.J.), deux à deux ; les individus identifiés de chaque espèce étant pondérés par $(N_e(i))$ l'inverse de l'effectif de leur espèce, les autres individus anonymes classés avec l'un des deux groupes sont mis en élém. sur. .

On arrive à identifier l'espèce des individus anonymes, à l'exception d'une dizaine qui sont des très petits d'une espèce ou des très grands d'une autre ; la classification et les analyses faites ont le mérite de les indiquer et de les mettre à la disposition des chercheurs de l'O.R.S.T.O.M. pour identifier leur espèce.

On ne pourra calculer le pourcentage des bien classés pour les individus anonymes, qu'après avoir vérifié les identifications des espèces de ces individus ; mais on est sûr qu'entre *occidentalis* L.G. et les autres espèces la discrimination est parfaite, de même pour les *Sylviosorex* et les *bottegi*, et que le pourcentage des erreurs de classement des groupes (T.P.J.) les uns par rapport aux autres ne dépasse pas les (5%) même moins, soit (95%) de bien classés.

8 Une clé pour pouvoir identifier l'espèce d'un individu anonyme

Il est nécessaire que le vecteur des mensurations de l'individu (i_0) n'ait pas une composante (mesure) défectueuse, il suffit de bien prendre ses mesures, sinon et dans le cas d'un lot assez nombreux, on procédera de la même manière qu'au § 5 pour la détection des mesures erronées et de celles douteuses.

On commence par calculer les profils $(f_j^{i_0}, j = 1, 20)$ des mesures par rapport à leur somme (ou poids de l'individu i_0) ; on calcule alors les premiers facteurs (les trois premiers suffisent) de l'individu, comme élément supplémentaire, par la formule de transition :

$$F_{\alpha}(i_0) = \lambda_{\alpha}^{-1/2} \cdot \Sigma \{f_j^{i_0} \cdot G_{\alpha}(j) \mid j \in J\}$$

En se référant aux facteurs des variables $\{j \mid j \in J\}$ $\{\{G_{\alpha}(j) \mid j \in J\} \text{ et } \alpha = 1, 2, 3\}$, respectivement dans les analyses suivantes :

Premièrement on calcule ces facteurs dans l'analyse des individus des quatre espèces ayant les effectifs les plus nombreux (*occidentalis*, *theserae*, *juvenetae* et *poensis* identifiés), pondérés chacun par l'inverse de l'effectif de son espèce, (tous les autres individus étant mis

en élém. sup.) ; puisque cette analyse a le mieux séparé les espèces les unes par rapport aux autres, les *occidentalis* L.G. d'une part, et les *bottegi* - *Sylvisorex* de l'autre, de même que les groupes des *theresae*, *poensis*, *jouvenetae* et *wimmeri* prennent leur place dans le plan (1,2) que l'on reproduit.

Et grâce au calcul des facteurs l'individu à déterminer i_0 va se placer quelque part dans ce plan ; par rapport aux trois blocs :

- *occidentalis*, *lamottei* et *giffardi*,
- *jouvenetae*, *wimmeri*, *poensis*, *theresae* et *lusitania*
- Et *bottegi* et *Sylvisorex*.

On distinguera les cas suivants :

1°) Le premier facteur de l'individu i_0 ($F_1(i_0)$ du point f_j^i) est très positif, il se place avec les *occidentalis* L.G. ; on peut alors calculer les valeurs des deux premiers facteurs de i_0 dans l'analyse des *occidentalis* - *lamottei*, ou *occidentalis* *giffardi*, en élém. sup., pour pouvoir affirmer l'identification de son espèce.

Cependant, pour toutes les espèces dont l'échantillon a un effectif peu nombreux, bien que les résultats obtenus soient assez bons, ils ne sont pas suffisants : il faut constituer des échantillons d'une vingtaine au moins, d'individus identifiés.

1') f_j^i se place entre les deux premiers blocs : O.L.G. et T.P.J.W. : c'est un cas rare, on aura à calculer les deux premiers facteurs de i_0 dans l'analyse à part des individus pondérés des *occidentalis* et *jouvenetae*, pour s'assurer de l'appartenance de l'individu, puisque les *jouvenetae* sont les plus proches des *occidentalis* L.G. dans l'analyse précédente.

2°) $F_1(i_0)$ est très négatif ; i_0 serait de l'une des espèces *bottegi* ou *Sylvisorex*. Les résultats obtenus étant catégoriques en ce qui concerne ces deux espèces et leurs sous-espèces, on prendra les groupes déjà obtenus comme des entités autonomes bien précises : les individus anonymes qui se sont classés, au paragraphe précédent, avec les *bottegi* - *Sylvisorex* seront considérés comme tels.

On aura donc à faire descendre l'individu i_0 dans l'analyse à part de ces deux espèces, et calculer ses deux premiers facteurs par la formule précédente, où les $G(j)$ sont les valeurs des facteurs des profils des variables (j, f_j^i) obtenues dans cette analyse.

i_0 est "*Sylvisorex*" s'il se place du côté négatif du premier axe, il est "*bottegi* I" s'il se met dans le 2° quadrant avec cette sous-espèce, il est "*bottegi* II" s'il se met dans le premier quadrant du côté positif du second axe, et il est "*bottegi* III" s'il se place dans le premier quadrant aussi mais du côté positif du premier axe.

2') $F_1(i_0)$ est très négatif mais il y a confusion sur son appartenance aux *theresae* ou aux *bottegi* - *Sylvisorex* : on calculera alors son premier facteur par rapport aux facteurs des variables de l'analyse des *theresae* identifiés pondérés par (1/56.) avec les *bottegi* - *Sylvisorex* pondérés par (1/31.), l'élément serait de l'une ou des autres espèces selon que son premier facteur est négatif ou positif.

Remarque : Soit un individu qui n'est pas représenté sur un certain axe, le premier axe par exemple, et qui est en plus bien représenté sur l'un des autres axes, jusqu'au quatrième, et qui se place à l'une des extrémités de cet axe (le second par ex.) : cet individu n'est pas corrélié aux résultats, à la signification du premier axe, à la discrimination entre deux espèces dans notre cas ; en d'autres termes il n'est pas

de ces deux espèces, et inversement. Il faudrait donc calculer les trois premiers facteurs du nouvel individu.

3°) Les deux premiers facteurs $F_1(i_0)$ et $F_2(i_0)$ placent i_0 parmi les individus des espèces *theresae*, *poensis*, *jouvenetae*, *wimmeri* et les nouveaux groupes ; (ceux-ci se sont avérés assez nombreux ; ils se placent en élém. sup. du côté positif du second axe (plan (1,2)), dans l'analyse précédente avec pondération). Les *poensis* sont au-dessous et les *theresae* les séparent des *jouvenetae* qui sont proches des *occidentalis* et remplissent l'espace entre ceux-ci et les *theresae*.

On connaîtra donc directement à quelles espèces on doit comparer le nouvel individu à partir de l'emplacement dans le plan (1,2) du point $(f_J^{i_0})$ le représentant.

On a passé en revue les deux possibilités 1') et 2') et l'on peut être sûr de l'espèce de l'individu dans les trois cas suivants :

- Si le point $f_J^{i_0}$ se met à l'extrémité positive du second axe, l'individu est du groupe des *wimmeri*.
- Si $f_J^{i_0}$ est dans le troisième quadrant du côté du second axe, c'est un *jouvenetae*.
- Et il est *theresae* si ses deux premiers facteurs sont assez négatifs.

Dans tous les autres cas il faut calculer les trois premiers facteurs dans chacune des trois analyses suivantes (avec pondération) :

theresae, *jouvenetae*
theresae, *poensis* et
poensis, *jouvenetae*.

Et grâce à l'importante remarque précédente, on peut connaître si l'individu est de l'une de ces espèces ou non ; en effet le comportement de l'individu i_0 dans les trois analyses indique son appartenance ; s'il ne se met pas toujours avec une espèce, il n'est pas de cette espèce. De même si $f_J^{i_0}$ n'est pas représenté sur le premier axe d'une analyse qui oppose deux espèces et qu'il est bien représenté sur le deuxième ou troisième axe, il serait évident que (i_0) n'est pas de ces deux espèces ; et l'inverse est exact.

Le cas des wimmeri et groupes adjacents :

L'espèce *C. wimmeri* et les nouveaux groupes obtenus par l'analyse générale du § précédent, doivent avoir reçu la confirmation des spécialistes pour participer à cette clé. Cependant passons en revue les analyses qui les concernent, (en mettant entre accolades l'échantillon analysé) :

{Les individus identifiés des *jouvenetae* pondérés, avec les individus des *wimmeri* - groupes pondérés par l'inverse de l'effectif de leur ensemble, les autres *jouvenetae* non identifiés étant mis en élém. sup.} : on a une séparation nette dans le plan (1,3) suivant la première bissectrice.

{Ainsi pondérés, avec les *theresae* identifiés pondérés, les autres *theresae* étant mis en élém. sup.} : la séparation est nette sur le premier axe ; et les trois groupes se manifestent, et l'on voit bien qu'il s'agit de trois groupes.

On se souvient que l'on a eu du mal à les séparer des *poensis* et à identifier chaque groupe par rapport aux autres ; ainsi leur analyse d'ensemble avec les *poensis* n'aboutit qu'à indiquer qu'il s'agit de plusieurs groupes sans nette séparation. Tandis que chacun de ces groupes

une fois identifié, analysé avec les *poensis*, se sépare d'eux. On calculera donc les deux ou (3) premiers facteurs dans les trois analyses des *poensis* avec chacun de ces groupes au cas où l'on veut connaître à quel groupe cet individu i. appartiendra, étant donné qu'il n'appartient à aucune des espèces précédentes.

On se guide toujours par les remarques précédentes.

Remarque : Comme nous l'a signalé L. Bellier, cette clé peut être simplifiée si l'on tient compte non seulement de la forme des crânes (profil des mesures à partir desquelles sont calculées les valeurs des facteurs) mais de la taille ; en particulier les *giffardi* se distinguent des *occidentalis* par leur plus grande taille et par ailleurs les *C. bottegi* se distinguent des *Sylvisorex* par leur plus petite taille.