

# COPEPODES PELAGIQUES DU PLATEAU IVOIRIEN

## II-UTILISATION DE L'ARBRE DE LONGUEUR MINIMUM DANS LE CLASSEMENT DES RECOLTES

par M. Roux, M. Gaborit-Rezzouk et D. Binet

### RESUME

L'algorithme de l'arbre de longueur minimum est utilisé pour le classement de deux séries de récoltes de copépodes planctoniques.

L'algorithme consiste à relier entre elles les récoltes les plus proches les unes des autres, sans créer de cycle et de façon à ce que la somme des longueurs des segments soit minimale.

Nous avons déterminé la distance entre récoltes de deux manières : en utilisant un coefficient d'association basé sur les présences ou absences d'espèces (l'indice de Jaccard) et avec la métrique du  $\chi^2$ . L'indice de Jaccard est jugé inintéressant ; la distance du  $\chi^2$  permet par contre de comparer les résultats avec ceux de l'analyse des correspondances.

Les résultats sont discutés d'un point de vue écologique.

### ABSTRACT

The minimum spanning tree algorithm is used to classify two sets of planktonic copepods samples. This algorithm links the samples the distance of which is minimum, without doing a loop and so that the sum of the segment lengths be minimum.

We estimated the distance between samples by two different ways : by a coefficient of association - the Jaccard's index - and by the  $\chi^2$  distance. Jaccard's index is not retained but the use of the  $\chi^2$  distance allows to compare with the "analyse factorielle des correspondances".

The results are discussed from an ecological point of view.

Des différentes méthodes d'analyse numérique dont dispose l'écologiste, aucune n'est parfaitement satisfaisante, parce qu'elle ne peut rendre compte à la fois simplement et totalement de tous les aspects de la réalité. Aussi en utilise-t-on plusieurs et en cherche-t-on toujours de nouvelles.

Nous exposerons ici le principe de l'"arbre de longueur minimum" et les résultats obtenus avec une série de données déjà soumises à l'analyse des correspondances (BINET, GABORIT et ROUX, 1972).

Il s'agit des effectifs des 119 espèces de copépodes récoltées au cours d'un cycle annuel (mars 1969-avril 1970) en plusieurs points du plateau continental ivoirien.

Ces récoltes se divisent en deux séries:

- La radiale Grand Bassam (Nord-Sud) comprend 3 stations, numérotées de 1 à 3, situées respectivement au-dessus des fonds de 30, 50 et 200m. Les pêches sont verticales, au filet ICITA, du fond jusqu'en surface, elles ont eu lieu toutes les 4 semaines.

- La station côtière, à deux milles à l'ouest de Vridi, au-dessus des fonds de 35m, était visitée approximativement une fois par semaine. Les échantillons proviennent également de pêches verticales fond-surface, au filet ICITA.

A chaque station nous avons effectué 2 traits identiques. Les effectifs de chaque espèce étaient dénombrés à la loupe binoculaire sur une fraction provenant du mélange des 2 récoltes.

#### 1.- ALGORITHME DE L'ARBRE DE LONGUEUR MINIMUM

La construction de l'arbre de longueur minimum (en anglais, minimum spanning tree) entre  $n$  points dont on connaît toutes les distances deux à deux correspond à l'origine au problème pratique suivant: étant données  $n$  villes séparées par ces distances, quel doit

être le tracé de la ligne téléphonique la plus économique, telle que de chacune de ces villes on puisse entrer en communication avec n'importe laquelle des  $n-1$  autres, les points d'articulation du réseau (noeuds de l'arbre) ne pouvant se trouver que dans les villes.

L'arbre devra être connexe, c'est-à-dire qu'il ne pourra être scindé en deux blocs sans liaisons entre eux, pour que de tout point on puisse joindre n'importe quel autre; il sera en outre sans boucle (cycle) afin d'être le plus court possible (économique). Mathématiquement ce problème consiste à construire un graphe connexe, sans cycles, tel que la somme des longueurs de ses arcs (liaisons) soit minimale.

La solution est la suivante: on range d'abord tous les arcs (segments joignant deux points) par ordre de longueurs croissantes, sur une liste A. On construit, ensuite, pas à pas, la liste B des arcs de l'arbre cherché. A chaque pas on examine un nouveau segment de la liste A. Si celui-ci permet de faire une boucle en l'adjoignant aux segments déjà présents en B, au pas considéré, alors on passe au segment suivant de A; sinon on le considère, désormais, comme faisant partie de la liste B. On arrête ce processus lorsque B contient  $n-1$  segments, car l'adjonction d'un segment supplémentaire provoquerait nécessairement l'apparition d'une boucle.

Remarque: Un tel arbre n'est pas hiérarchique contrairement à ceux que l'on construit généralement en classification automatique. Les noeuds de l'arbre ne représentent pas des niveaux de liaisons entre points ou groupes de points mais bien les points eux-mêmes. Cet arbre est donc une schématisation de l'ensemble des relations existant entre les objets étudiés. Il n'est autre que le "dendrite" de la taxonomie de Wroclaw.

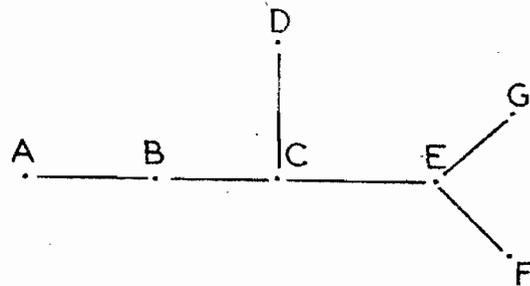
Exemple: Soit le tableau des distances mutuelles entre 7 points.

	A	B	C	D	E	F
B	17					
C	23	16				
D	19	26	18			
E	42	37	22	41		
F	43	25	30	47	14	
G	65	39	36	54	13	15

La liste A sera constituée ainsi:  
EG, EF, FG, BC, AB, CD, AD, CE, AC,  
etc...

tandis que la liste B sera:  
EG, EF, BC, AB, CD, CE.

On représentera l'arbre en étalant  
ses branches au maximum



## 2.- METRIQUES UTILISEES

On peut utiliser la technique de l'"arbre de longueur minimum" pour schématiser les relations entre différents objets, à partir de leurs distances mutuelles. En écologie on l'utilise pour classer les relevés suivant la ressemblance de leur composition en espèces, ou réciproquement pour classer les espèces suivant les biotopes dans lesquels elles se trouvent. Encore faut-il définir une distance entre récoltes, laquelle doit traduire la similitude de composition spécifique des récoltes.

On peut utiliser un indice d'association basé sur les seuls critères de présence ou d'absence; ou bien, tenir compte également de l'abondance de chaque espèce.

### 2.1.- Indice de présence - absence

ROUX et ROUX (1967) concluent à la supériorité des coefficients de présence-absence dans les relevés de végétaux terrestres, pour l'étude des associations. Ils considèrent la présence d'une espèce comme

un caractère "génétique" d'une population, alors que son abondance serait un caractère "morphologique". DANDONNEAU (1971) obtient de bons résultats avec ce type de coefficients, appliqués à l'étude du cycle annuel du phytoplancton devant Abidjan.

La possibilité de remplacer les comptages spécifiques par un examen de présence-absence ferait réaliser une économie de temps considérable; nous avons transformé les données d'abondance en données de présence-absence, pour tester l'intérêt de cette méthode.

Nous avons choisi l'indice de Jaccard (In SOKAL & SNEATH, 1963) en raison de sa simplicité et de son emploi assez répandu.

La distance entre les récoltes j et j' est:

$$s_{jj'} = \frac{c_{jj'}}{c_{jj'} + p_j + p_{j'}}$$

avec  $c_{jj'}$  = nombre d'espèces communes aux récoltes j et j',

$p_j$  = nombre d'espèces présentes dans j mais pas dans j'

$p_{j'}$  = nombre d'espèces présentes dans j' mais pas dans j.

Les résultats se sont avérés confus, les groupes de récoltes correspondant mal aux saisons hydrologiques; nous n'avons pu les interpréter. Nous avons jugé cette distance inintéressante et en avons cherché une tenant compte des variations d'effectifs.

En effet, la majorité des espèces de copépodes présentes dans ces récoltes peuvent supporter d'importantes variations du milieu sans disparaître totalement. Par contre l'abondance maximale d'une espèce au cours de l'année, a lieu lorsque les conditions écologiques optimales sont réalisées et que l'espèce a pu se multiplier activement.

## 2.2.- Distance basée sur les abondances

Nous avons travaillé sur les effectifs par  $m^3$  en utilisant comme pour l'analyse des correspondances la distance du  $\chi^2$  :

STATION COTIERE

N°	Date	T°	S‰	N°	Date	T°	S‰
1	28.2	23.4	35.46	24	16.9	18.9	35.82
2	18.3	28.6	35.07	25	23.9	18.7	35.83
3	25.3	28.3	35.05	26	30.9	20.0	35.72
4	1.4	29.2	34.87	27	7.10	21.8	35.23
5	8.4	29.1	34.99	28	14.10	21.6	35.
6	16.4	29.1	34.95	29	21.10	22.4	35.45
7	22.4	29.2	34.95	30	28.10	24.4	34.86
8	30.4	29.7	34.87	31	5.11	25.1	35.06
9	6.5	29.7	34.83	32	12.11	26.4	34.57
10	12.5	29.5	34.59	33	2.12	26.5	34.50
11	19.5	27.6	34.65	34	9.12	25.9	34.96
12	27.5	28.2	34.25	35	19.12	26.0	34.78
13	10.6	27.5	33.89	36	22.12	25.7	34.81
14	18.6	26.3	31.01	37	30.12	26.4	34.68
15	24.6	24.3	33.75	38	7.1	25.8	34.85
16	1.7	23.6	33.57	39	13.1	23.7	35.02
17	9.7	23.6	34.06	40	20.1	26.7	34.60
18	15.7	21.0	34.64	41	28.1	26.7	34.55
19	30.7	19.7	35.79	42	6.2	26.9	34.48
20	5.8	18.7	35.78	43	13.2	26.9	34.71
21	12.8	19.2	35.78	44	20.2	23.7	35.43
22	18.8	17.5	35.72	45	28.2	27.0	35.20(?)
23	2.9	18.6	35.78	46	10.3	22.1	35.57
				47	17.3	24.9	35.42
				48	24.3	27.6	35.14

NUMEROTATION ET DATE DES RECOLTES

CONDITIONS HYDROLOGIQUES A 10 m



$$d^2_{(jj')} = \sum_{i=1}^n \frac{1}{P(i)} \left[ \frac{P(i,j)}{P(j)} - \frac{P(i,j')}{P(j')} \right]^2$$

$P(i, j)$  est l'effectif de l'espèce  $i$  dans la récolte  $j$ .

$P(i)$  est la somme des effectifs de l'espèce  $i$  dans toutes les récoltes.

$P(j)$  est la somme des effectifs de toutes les espèces dans la récolte  $j$ .

### 3.- RESULTATS

#### 3.1.- Station côtière (fig. 1 et 2)

On retrouve les groupements obtenus par l'analyse des correspondances lorsque ceux-ci étaient nets\* :

a) - Les 3 parties de la saison froide:

A1: 15 juin au 10 juillet, c'est la mise en place de la saison froide, les eaux se refroidissent, mais la salinité diminue notablement, c'est la grande saison des pluies.

A2: 15 juillet au 20 août, c'est la première partie de la saison froide, le refroidissement s'accroît et la salinité augmente.

A3: premier septembre au 15 octobre, c'est la deuxième partie de la saison froide, les eaux se réchauffent et la salinité diminue légèrement.

b) - Les récoltes de saison chaude, en eau tropicale (groupe D) sont scindées en deux parties.

c) - Les récoltes de petites saisons chaude et froide (B1 et B2) sont réparties en deux groupes.

---

\* Nous reprenons les mêmes dénominations que pour l'analyse de correspondances (BINET, GABORIT & ROUX, 1972).

d) - Les diverses récoltes effectuées en saison chaude mais lors des conditions peu caractéristiques (C1 et C2) se répartissent en milieu d'arbre entre les groupes de petites saisons et ceux correspondant aux eaux tropicales.

On remarquera que les principales oppositions ne se font pas entre saison chaude et saison froide mais entre début et fin d'une même saison.

Notons par ailleurs la position centrale des groupes de "petites saisons" aux caractères intermédiaires. La récolte 43 (13.2.70) est le point de départ des principales ramifications, elle est très proche de l'origine des axes de l'analyse des correspondances, mais fait partie d'un ensemble de récoltes de petites saisons ou d'intersaison.

La récolte 43 semble jouer un rôle très particulier, nous l'avons supprimé dans une seconde analyse. L'arbre s'en trouve plus étalé; les groupes de grande saison froide sont rapprochés les uns des autres, sans être modifiés, ce qui prouve leur homogénéité. L'opposition se fait entre les groupes A de grande saison froide et C1, C2 et D de grande saison chaude. Les groupes intermédiaires se trouvent au milieu (fig.2).

### 3.2.- Radiale Grand Bassam (fig.3 et 4).

Les groupements de récolte correspondant à des saisons sont moins nets ici. On retrouve cependant:

- Le groupe de grande saison froide (A) à l'exception des récoltes du mois de septembre. Il est situé en extrémité d'arbre.

- Les groupes de petites saisons aux stations néritiques (1 et 2): petite saison froide (C) et petite saison chaude (D). L'un et l'autre sont situés vers le centre du dendrite.

RADIALE BASSAM

Date	Station 1			Station 2			Station 3		
	N°	T°	S‰	N°	T°	S‰	N°	T°	S‰
13. 3. 69.	1	28. 90.	35. 05	15	28. 90	35. 05	29	29. 19	35. 02
8. 4.	2	29. 31	34. 85	16	29. 29	34. 96	30	29. 15	34. 95
8. 5.	3	29. 57	34. 82	17	29. 70	34. 60	31	29. 46	34. 47
5. 6.	4	28. 18	33. 24	18	28. 14	33. 70	32	28. 40	34. 14
2. 7.	5	23. 98	34. 19	19	25. 56	34. 56	33	24. 80	34. 60
30. 7.	6	19. 78	35. 83	20	21. 48	35. 26	34	18. 94	35. 61
28. 8.	7	20. 66	35. 42	21	21. 18	35. 47	35	22. 75	35. 44
26. 9.	8	18. 71	35. 85	22	18. 45	35. 83	36	18. 73	35. 83
23. 10.	9	24. 11	34. 01	23	23. 18	35. 40	37	24. 07	34. 85
20. 11.	10	28. 27	33. 60	24	28. 30	33. 83	38	27. 76	33. 90
18. 12.	11	28. 23	34. 02	25	27. 61	34. 10	39	27. 65	34. 15
19. 1. 70.	12	25. 53	34. 96	26	27. 68	34. 26	40	27. 07	34. 43
14. 2.	13	27. 41	34. 68	27	27. 52	34. 64	41	27. 63	34. 56
12. 3.	14	25. 43	35. 39	28	25. 68	35. 36	42	26. 73	35. 31

NUMEROTATION, DATE ET CONDITIONS HYDROLOGIQUES DES RECOLTES.  
TEMPERATURE ET SALINITE A 10 METRES

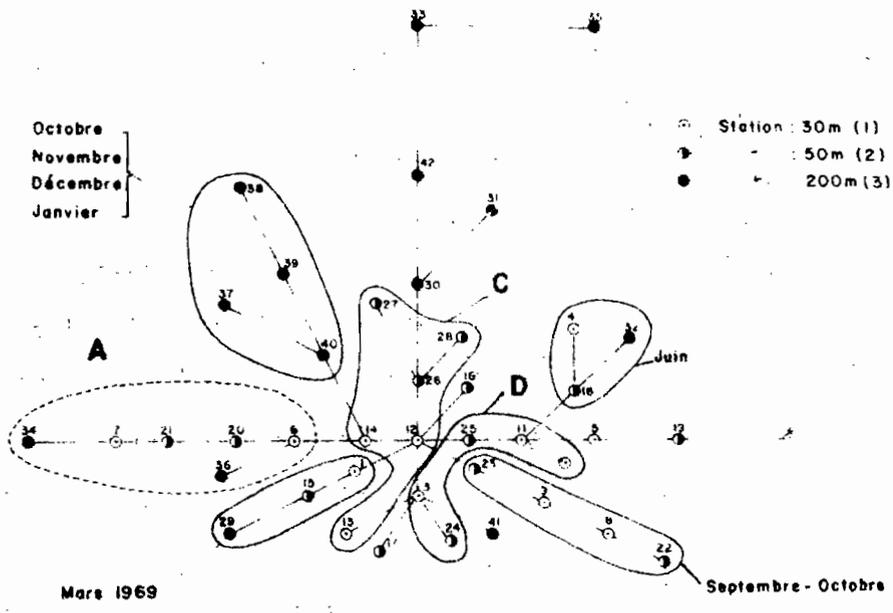


Fig. 3: - Arbre de longueur minimum des récoltes de la radiale Bassam.

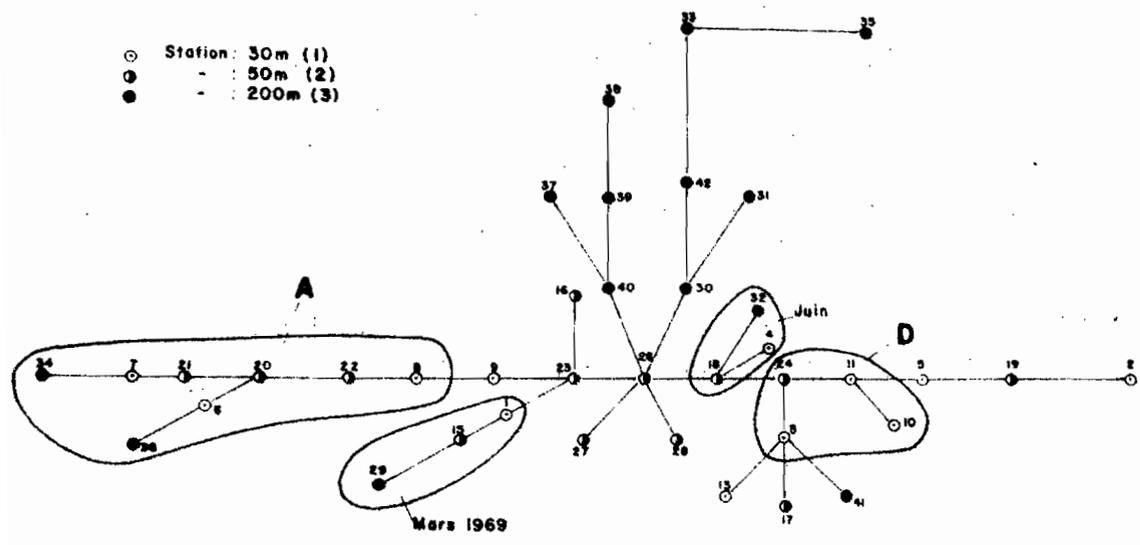


Fig. 4: - Arbre de longueur minimum des récoltes de la radiale Bassam, récoltes 12, 14 et 25 exclues.

- Les récoltes d'une même radiale peuvent se retrouver groupées: mars 1969, ou juin (équivalent du groupe A1 cité précédemment). Il apparaît entre stations d'une même radiale une plus grande similitude qu'avec l'analyse des correspondances.

- On peut dans un cas ~~retrouver~~ les points récoltes de la station 3 par saison: 37, 38, 39, 40 ~~correspondent~~ aux mois d'octobre à janvier. Pour les autres récoltes de la station 3 ce n'est pas possible.

Les noeuds des principales ramifications de l'arbre sont occupés par les stations 1, les stations 3 sont toujours aux extrémités de l'arbre.

Dans ce dernier cycle de récoltes, le prélèvement 12 joue également un rôle très important; nous avons procédé à une nouvelle analyse après suppression des récoltes 12, 14 et 25. On retrouve à peu près la même disposition des récoltes, mais les points de ramification sont occupés par des récoltes de la station 50m (fig.4).

#### 4.- DISCUSSION ET CONCLUSION

Il n'est pas surprenant de trouver avec l'algorithme de l'arbre de longueur minimum des résultats voisins de ceux obtenus par l'analyse des correspondances, puisqu'on utilise la même métrique dans les deux cas.

Nous voudrions plutôt dégager les différences et tenter de les expliquer.

4.1.- L'absence de boucle dans un tel classement, l'étalement maximum du dendrite contraint à couper l'année en certains endroits,

au contraire, la projection des récoltes sur les premiers axes de l'analyse des correspondances ne masque pas l'aspect cyclique de la succession des saisons\*.

Dans le cas de la station côtière, au lieu d'obtenir les grandes saisons chaude d'un côté, froide de l'autre avec les petites saisons entre les deux, on observe, dispersées en croix autour des petites saisons, le début et la fin de la grande saison froide, le début et la fin de la grande saison chaude. L'évolution des populations se fait de manière continue, les transitions entre saisons sont progressives pour le zooplancton. Le changement de faune en cours de "saison" peut être relativement plus important que lors du changement de saison. Mais la suppression de la seule récolte 43 rétablit l'opposition entre saisons chaudes et saisons froides, très nettement.

4.2.- Parmi les récoltes de la radiale Bassam, on n'observe une cohésion assez forte (distances faibles) entre les pêches des 3 stations qu'aux mois de juillet et d'août. Tout le plateau est en effet recouvert par des eaux d'upwellings, les différences entre stations sont moins marquées qu'aux autres périodes.

Dans deux autres cas (mars et juin 1969) les 3 stations d'une même radiale sont aussi rassemblées, ce qui prouve qu'il y a plus d'affinité entre des récoltes prélevées de la côte au rebord du talus, à quelques heures les unes des autres, qu'entre des récoltes issues d'une même station, mais à 4 semaines d'intervalle.

Autrement dit le facteur saisonnier est plus important que le gradient côte-océan dans la détermination des populations du plateau.

---

\* La méthode de l'arbre de longueur minimum pourrait présenter plus d'intérêt dans l'analyse de récoltes provenant d'une même aire géographique, avec peu de décalage chronologique entre les récoltes.

4.3.- On devrait s'attendre à ce que les points-récoltes de la station de 50m soient souvent aux noeuds de ramification, puisque cette station se trouve entre les stations de 30 et 200m.

Or les points représentatifs de la station de 30m sont en général proches les uns des autres et pour joindre entre eux les points récoltes de la station 50m, il faut fréquemment passer par ceux de la station de 30m. Mais cette anomalie disparaît avec la suppression de quelques récoltes des fonds de 30m. Il en est de même pour un certain nombre de points récoltes de la station de 200m, quoique la majorité d'entre eux soient reliés les uns aux autres, mais par des distances importantes. Les distances entre points-récoltes sont faibles pour la station de 30m et augmentent au fur et à mesure qu'on s'éloigne vers la périphérie de l'arbre, qui correspond souvent nettement aux points-récoltes de la station de 200m. On pourrait interpréter ceci de la façon suivante:

- Il existe près de la côte (station 30m) des populations qui évoluent avec les variations saisonnières du milieu (influences terrigènes, upwellings...); les distances entre récoltes sont faibles. La plupart de ces récoltes sont au centre du dendrite.

- Au milieu du plateau (station 50m) les influences saisonnières sont encore très nettes. Dans la mesure où les points-récoltes de cette station ne peuvent être reliés entre eux que par l'intermédiaire de ceux de la station de 30m, on peut dire que l'évolution des populations au-dessus des fonds de 50m dépend davantage de celle des populations des fonds de 30m, que de la sienne propre.

- Au-dessus du talus (station 200m) les populations ne se rattachent à celles du reste du plateau que lorsqu'il y a un refroidissement et un apport d'eau profonde près de la côte.

Durant les petite et grande saisons chaudes les populations du talus s'éloignent de celles du plateau et sont également très distantes les unes des autres; comme si leur évolution était entièrement distincte de celles des populations côtières. Les différents états de ces populations sont tellement éloignés les uns des autres que l'on peut se demander s'il s'agit de l'évolution d'une même population ou bien d'un changement de faune.

Nous retrouvons ce qui semble un caractère général des variations zooplanctoniques au-dessus d'un plateau continental en milieu tropical, lorsqu'il est soumis à un régime d'upwellings saisonniers (BINET & DESSIER, 1971 - BINET, GABORIT & ROUX, 1972):

- Avancée des populations hauturières vers la côte, relative uniformisation des peuplements durant les périodes froides.
- Séparation entre les populations du plateau et celles du large en période de réchauffement.

#### Remerciements:

Nous remercions les responsables du laboratoire de Taxinomie Végétale Expérimentale et Numérique associé au C.N.R.S., dont l'amabilité nous a permis de réaliser ces calculs, sur l'ordinateur UNIVAC de la faculté des Sciences d'Orsay.

#### BIBLIOGRAPHIE

- BINET (D.) et DESSIER (A.), 1971.- Premières données sur les copépodes pélagiques de la région congolaise. I.- Liste des espèces et notes écologiques. Cah.ORSTOM, sér.Océanogr., 9, 4: 411-457

- BINET (D.), GABORIT (M.) et ROUX (M.), 1972.- Copépodes pélagiques du plateau ivoirien. Utilisation de l'analyse des correspondances dans l'étude des variations saisonnières. Doc.Scient. Centre Rech. Océanogr. Abidjan, 3, 1: 47-79
- DANDONNEAU (Y.), 1971.- Etude du phytoplancton sur le plateau continental de Côte d'Ivoire. I - Groupes d'espèces associées. Cah. ORSTOM, sér. Océanogr., 9, 2: 247-266
- GOWER (J.C.), and ROSS (G.J.S.), 1969.- Minimum spanning tree and single linkage cluster analysis. Applied Statistics, 18: 54-64
- FLOREK (K.), and coll., 1951.- Tacksonomia Wroclawia, Przegl. Antrop 17. Poznan: 193-207
- ROUX (G.) et ROUX (M.), 1967.- A propos de quelques méthodes de classification en **phytosociologie**. Rev. Statistique appliquée, 15, 2: 59-72
- ROUX (G.), 1971.- Le problème des pelouses à Nardus stricta: une approche par les méthodes d'analyse numérique. Colloque interdisciplinaire sur les milieux naturels supra-forestiers des montagnes du bassin occidental de la Méditerranée. pp. 299-314. Centre Universitaire de Perpignan. (Cet article contient un exemple d'application et des comparaisons de méthodes).
- SOKAL (R.R.), and SNEATH (P.H.A.), 1963.- Principles of Numerical Taxonomy. Freeman & Cie U.S.A., 359 p.

\*  
\* \*