

UNIVERSITÉ LOUIS PASTEUR

Institut
de Recherches
sur l'Enseignement
des Mathématiques

10, rue du Général Zimmer
67084 STRASBOURG CEDEX

Informatique, quand tu nous tiens

Ou comment utiliser l'informatique dans
l'enseignement secondaire sans en faire
une matière supplémentaire

1974/75

Tome 2



UNIVERSITÉ LOUIS PASTEUR

Institut
de Recherches
sur l'Enseignement
des Mathématiques
10, rue du Général Zimmer
67084 STRASBOURG CEDEX

Informatique , quand tu nous tiens

Ou comment utiliser l'informatique dans
l'enseignement secondaire sans en faire
une matière supplémentaire

1974/75

Tome 2

INFORMATIQUE, QUAND TU NOUS TIENS

Ou comment utiliser l'informatique dans l'enseignement
secondaire sans en faire une matière supplémentaire

1974 - 1975

CHAPITRE 1 : Simulations

CHAPITRE 2 : Pour une géographie quantitative en second cycle de lycée

Ont participé aux travaux de ce groupe :

Christiane BERNICOT

Walter BIEBER

Jean Paul BOMANS

Gérard CLOSS

Claire DUPUIS

Thierry HATT

Jean Claude KEYLING

Patrice KOULLEN

Micheline LELLIG

Jean Marie PONCELET

Frédéric REISS

Gilbert REMY

Bernard RIEHL

Jacques VINCENT

Jean WENCKER

Geneviève WIRRMANN

Pour revenir succinctement sur le fascicule 1973-74, disons que le chapitre sur l'art du calcul a été apprécié et utilisé avec différentes variantes sans qu'il soit pour autant utile d'y revenir cette année.

Les organigrammes de recherche de limites n'ont, soyons francs, rencontré aucun succès. Nous les enterrons donc pour l'instant. Par contre, l'utilisation de l'informatique en géographie s'est considérablement développée comme en témoigne le chapitre 2.

SIMULATIONS

Nous avons cherché à réaliser des simulations utilisables par les élèves et ayant un caractère aléatoire. Nous en avons retenu deux et présentons des exécutions de ces deux simulations sur Mitra 15.

Il faut cependant remarquer que l'utilisation de la loi normale dans la simulation de l'évolution d'une valeur boursière peut rendre sa compréhension plus difficile pour les élèves quoiqu'elle ne constitue pas un obstacle majeur, la "normalité" étant une notion presque intuitive.

Par contre, la simulation de l'évolution de deux populations ne présente aucun problème et peut être comprise par des élèves assez jeunes. On peut, avant d'utiliser l'ordinateur, faire simuler les rencontres aléatoires au moyen d'un échiquier, de pions de différentes couleurs et d'une urne pour tirages. Les similitudes entre ce problème et celui de la grippe sont évidents (Bulletin de l'A.P.M. N° 296, décembre 1974 - merci Josèphe).

Ce programme peut être utilisé par des collègues de sciences naturelles ou de géographie. Ce ne sont pas les variantes sur ce thème qui manquent.

SIMULATION DE L'EVOLUTION D'UNE
"VALEUR BOURSIERE "

L'évolution d'une valeur dans une bourse, que ce soit la bourse des actions, des matières premières ou des fleurs est un phénomène complexe que nous ne prétendons pas avoir traduit dans son intégralité. Nous avons simplement simulé la situation suivante : la valeur VAL (t) d'un produit à l'instant t dépend de ses valeurs antérieures mais de façon non déterministe. Si rien ne perturbe l'évolution de cette valeur, elle tend à se stabiliser à une moyenne. Cette situation simpliste risque de manquer d'intérêt. C'est pourquoi nous supposons qu'il existe des phénomènes extérieurs appelés CRISES qui tendent à modifier les valeurs.

Quand une crise se déclenche, il peut y avoir auto-entretien de la crise ou amortissement par la tendance à la stabilisation. On peut expliquer l'auto-entretien par le fait qu'une crise traduit une tendance qui se prolonge éventuellement pendant plusieurs "instants".

Modèle et présentation du programme.

- La valeur VAL (t) à l'instant t est une variable aléatoire qui suit une loi normale (de Gauss) dont la moyenne MU et l'écart type SIGMA sont calculés à partir d'un certain nombre de valeurs antérieures, pondérées.

On peut par exemple choisir :

$$MU = \frac{1}{12} V(t - 4) + \frac{1}{6} V(t - 3) + \frac{1}{4} V(t - 2) + \frac{1}{2} V(t - 1)$$

$$SIGMA = \sqrt{\frac{1}{12} V^2(t - 4) + \frac{1}{6} V^2(t - 3) + \frac{1}{4} V^2(t - 2) + \frac{1}{2} V^2(t - 1) - MU^2}$$

MU et SIGMA sont ici la moyenne et l'écart type de la distribution

valeurs	V(t - 4)	V(t - 3)	V(t - 2)	V(t - 1)
pondérations	1/12	1/6	1/4	1/2

Dans le programme fourni nous n'avons pas imposé la pondération : ainsi on appelle NOMBR le nombre de valeurs antérieures dont on tient compte, VAL (I) et POND (I), pour I variant de 1 à NOMBR, les valeurs et pondérations de la distribution servant au calcul de MU et de SIGMA. Ainsi

$$\text{MU} = \sum_{I=1}^{\text{NOMBR}} \text{VAL (I)} \cdot \text{POND (I)}$$

$$\text{SIGMA} = \sqrt{\sum_{I=1}^{\text{NOMBR}} \text{POND (I)} \cdot \text{VAL}^2(\text{I}) - \text{MU}^2}$$

La réalisation-simulation de la valeur VAL (t) à l'instant t sera obtenue par un tirage au hasard suivant une loi normale $\mathcal{N}(\text{MU}, \text{SIGMA})$ de moyenne MU et d'écart-type SIGMA. On peut évidemment aussi tirer un nombre N(t) suivant une loi normale centrée réduite $\mathcal{N}(0, 1)$. On calculera alors VAL (t) = MU + (SIGMA * N(t)).

A chaque instant t on recalcule évidemment MU et SIGMA.

- Une CRISE à un instant t est caractérisée par le fait que la moyenne MU de la loi normale de VAL (t) est multipliée par un coefficient KCRI positif. Les crises sont indépendantes des valeurs obtenues. La probabilité de déclenchement d'une crise à un instant t (il n'y a pas eu de crise à l'instant t-1) est constante. (PCRI).

S'il y a crise elle peut être faible, moyenne ou forte et dans chaque cas à la hausse ou à la baisse. L'intensité de la crise est déterminée par un tirage (LCRI) au hasard, les probabilités respectives des crises étant LAMC, MAMC et 1 - LAMC - MAMC.

La hausse ou la baisse sont supposées aussi probables l'une que l'autre (YCRI). Ces deux éléments permettent de calculer KCRI. La probabilité de prolongement peut être différente suivant le type de crise (PROP 1, PROP 2, PROP 3).

Réalisation pratique des tirages au hasard.

En simulation il est très fréquent d'utiliser des suites de nombres pseudo-aléatoires (nul n'est parfait) qui présentent l'intérêt d'être reproductibles.

La méthode classique utilisée pour engendrer une suite pseudo-aléatoire est due à Lehmer.

La relation de récurrence $x_i \equiv a x_{i-1} + c \pmod{m}$ fournit une suite x_i . m est un entier de grande taille, en général 2^α où α est fonction du plus grand nombre que votre ordinateur peut contenir. La période de la suite x_i est alors m à condition que c soit un pair et que a soit égal à un multiple de 4 plus 1.

La suite $\frac{x_i}{m}$ est utilisée comme suite pseudo-aléatoire de nombres compris entre 0 et 1 $-\frac{1}{m}$; on peut considérer que la suite $\frac{x_i}{m}$ simule un tirage au hasard suivant la loi uniforme sur $[0, 1[$ (précision : $1/m$).

Pour obtenir une simulation d'un tirage suivant la loi normale centrée réduite $\mathcal{N}(0, 1)$ on utilise le fait que si X_1, X_2 sont deux variables aléatoires indépendantes, uniformes sur $[0, 1[$, les variables :

$$Y_1 = \sqrt{-2 \text{Log } X_1} \cdot \cos(2 \pi X_2)$$

$$Y_2 = \sqrt{-2 \text{Log } X_1} \cdot \sin(2 \pi X_2)$$

sont indépendantes et de loi $\mathcal{N}(0, 1)$

En pseudo-aléatoire, on remplace X_1 et X_2 par deux suites indépendantes du type décrit précédemment.

Sur ce sujet il existe une nombreuse littérature, par exemple J.M. Hammersley et D.C. Handscombe : Les méthodes de Monte-Carlo (Dunod).

J. Agard Les méthodes de simulation (Dunod)

D.N. Chorafas La simulation mathématique et ses applications (Dunod)

L. Lebart et J.P. Fénelon Statistique et informatique appliquée (Dunod).

Ces méthodes sont en général programmées par les constructeurs. Dans le programme présenté, la procédure `NORM(R, S, U, V)` calcule, à partir des suites "aléatoires" fournies par la fonction `ALE`, le résultat d'un tirage au hasard suivant une loi $\mathcal{N}(R, S)$ normale de moyenne R et d'écart type S .

NOMBRE ←
PONDERATIONS ←
VALEURS ←
PROBABILITES ←
TEMPS DE L'ETUDE ←
ECART TYPE MAXIMAL ←
PERIODE D'AFFICHAGE ←
LAM ← 0

FAIRE POUR T ← 1 JUSQU'À T MAX

CALCUL DE MU ET SIGMA

OUI NON
A L'INSTANT T-1, Y-AVAILL-IL UNE CRISE?

OUI NON
cette crise se prolonge-t-elle?

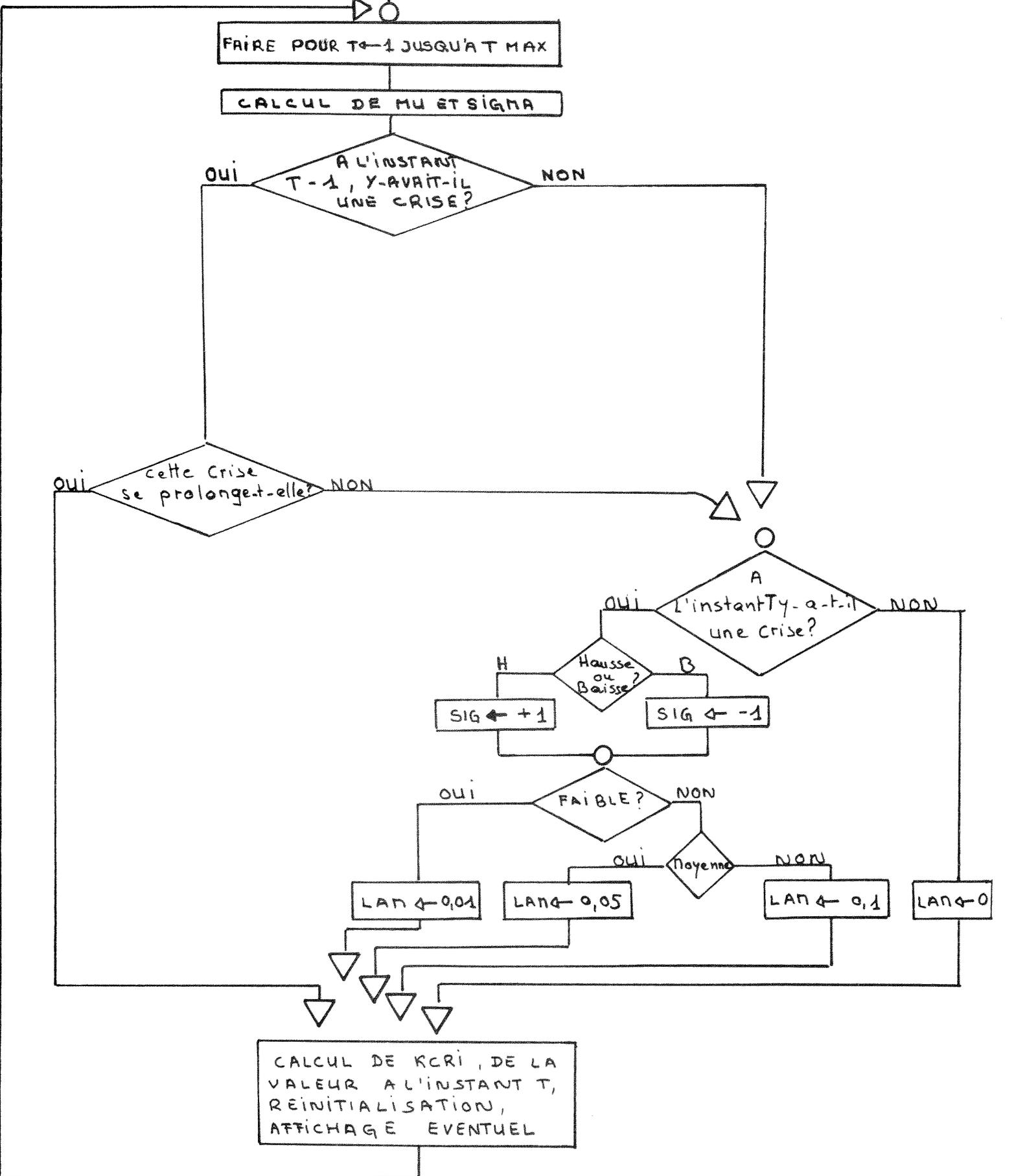
OUI NON
A L'INSTANT T, Y-A-T-IL UNE CRISE?

H B
Hausse ou Baisse?
SIG ← +1 SIG ← -1

OUI NON
FAIBLE?

OUI NON
Moyenne
LAM ← 0,01 LAM ← 0,05 LAM ← 0,1 LAM ← 0

CALCUL DE KCRi, DE LA VALEUR A L'INSTANT T, REINITIALISATION, AFFICHAGE EVENTUEL



PROGRAMME EN LANGAGE SYMBOLIQUE D'ENSEIGNEMENT

Simulation d'évolution d'une valeur boursière.

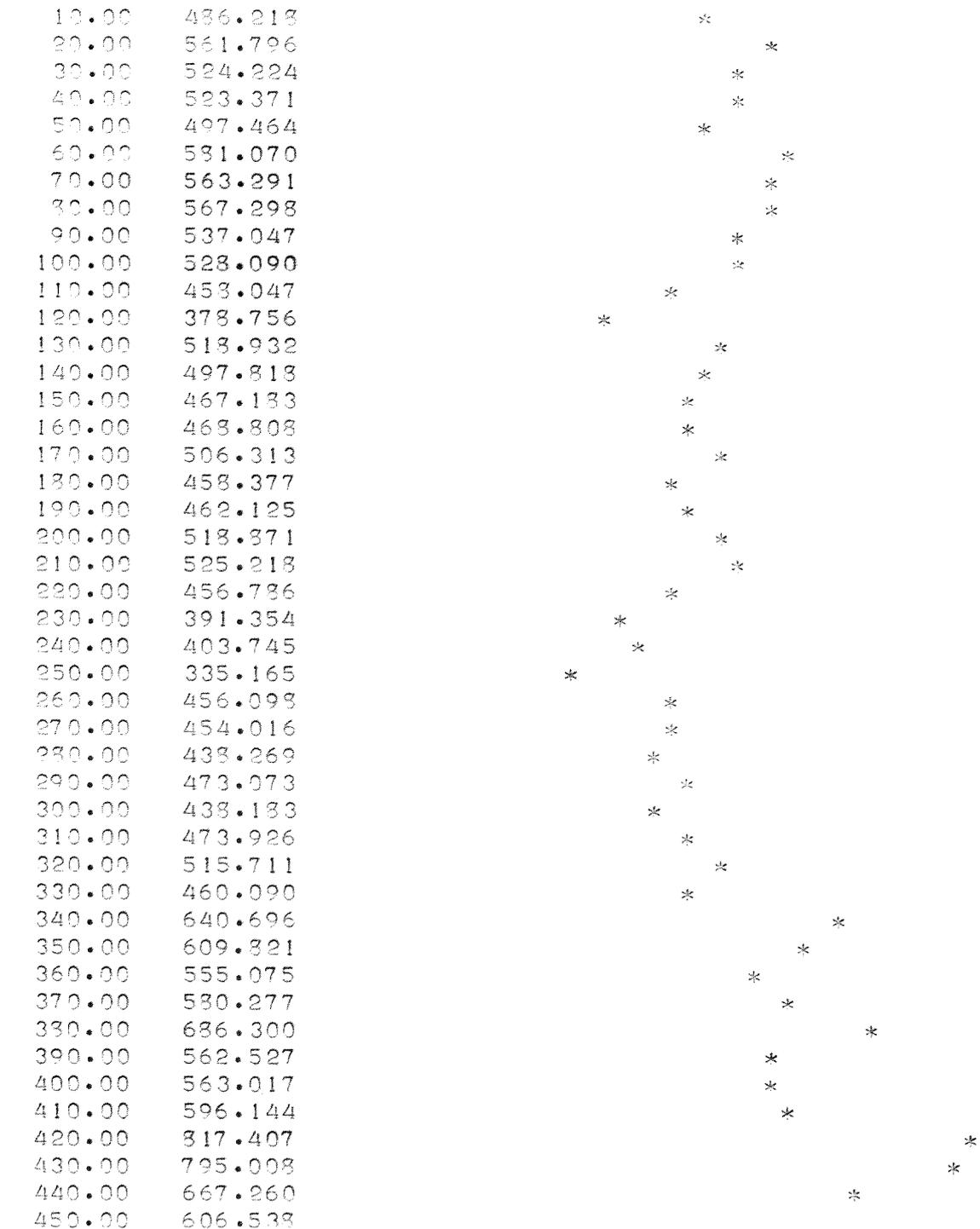
```
APPELER CRIS2
PERFORMER A PARTIR DE 1
1*CRIS2
20*
22 AFFICHER(/,'DONNER LE NOMBRE DE VALEURS DE REFERENCE 'I;LIRE NOMBR
24 TABLEAU POND[NOMBR],VAL[NOMBR+1]
26 AFFICHER(/,'DONNER LES PONDERATIONS : 'I]
28 FAIRE 28 POUR I+1 JUSQUA NOMBR;LIRE POND[I]
30 AFFICHER(/,'DONNER LES VALEURS : 'I]
32 FAIRE 32 POUR I+1 JUSQUA NOMBR;LIRE VAL[I]
34 VAL[I]+0
36 XCRI+1;LAM+0
37 SIG+0
38 AFFICHER(/,'INITIALISEZ LES 7 VARIABLES ALBATOIRES : 'I]
40 LIRE LCRI,POCRI,XCRI,YCRI,ZCRI,UU1,VV1
42 AFFICHER 'PROBA DE PROLONG. DE CRISE FAIBL,MOYENNE,FORTE '
44 LIRE PROP1,PROP2,PROP3
46 AFFICHER(/,'PROBABILITE DE NOUVELLE CRISE 'I]
48 LIRE PCRI
50 AFFICHER(/,'ECART TYPE MAXI RETENU 'I]
52 LIRE SISUP
54 AFFICHER(/,'TEMPS MAXIMUM DE L'ETUDE : 'I]
56 LIRE TMAX
58 AFFICHER(/,'PERIODE D'AFFICHAGE : 'I;LIRE PEP
60 AFFICHER 'PROBA DE CRISE FAIBLE 'I;LIRE LAMC
62 AFFICHER 'PROBA DE CRISE MOYENNE 'I;LIRE MAMC
64 AFFICHER 'DONC PROBA DE CRISE FORTE= 'I,1-LAMC-MAMC
67 FAIRE 113 POUR T+1 JUSQUA TMAX
68 MU+0;XX+0
69 FAIRE 70 POUR J+1 JUSQUA NOMBR;MU+MU+VAL[J]*POND[J]
70 XX+XX+VAL[J]+2*POND[J]
71 SIGMA+RAC(ABS(XX-MU+2))
72 SI LAM=0 ALORS ALLER EN 84
76 POCRI-ALE(POCRI)
78 SI LAM=0.01 ALORS SI POCRI<PROP1 ALORS ALLER EN 100
80 SI LAM<0.05 ALORS SI POCRI<PROP2 ALORS ALLER EN 100
82 SI LAM=0.1 ALORS SI POCRI<PROP3 ALORS ALLER EN 100
84 XCRI-ALE(XCRI)
86 SI XCRI<1-PCRI ALORS DEBUT LAM+0;ALLER EN 98 FIN
87
88 YCRI-ALE(YCRI);SI YCRI<0.5 ALORS SIG+1 SINON SIG+-1
90 LCRI-ALE(LCRI)
92 SI LCRI<LAMC ALORS DEBUT LAM+0.01;ALLER EN 98 FIN
94 SI LCRI<LAMC+MAMC ALORS DEBUT LAM+0.05;ALLER EN 98 FIN
96 LAM+0.1
98 ZCRI-ALE(ZCRI)*LAM;KCRI+1+ZCRI*SIG
100
102 MU+MU*KCRI;SI SISUP<SIGMA ALORS SIGMA+SISUP
104 WA+&NORM(MU,SIGMA,UU1,VV1)
106 VAL[NOMBR+1]-WA
108 FAIRE 108 POUR K+1 JUSQUA NOMBR;VAL[K]+VAL[K+1]
110 VAL[K]-0;SI T/PER#ENT(T/PER) ALORS ALLER EN 113
112 AFFICHER(/,X,F4.2,X,F5.3,*X,'*']T,WA,ENT(WA/20)
113
120 TERMINER
122 PROCEDURE &NORM(P,S,U,V)
124 U-ALE(U);V-ALE(V)
126 RESULTAT S*RAC(-2*LGN(U))*COS(2*3.141592*V)+R
```

DONNER LE NOMBRE DE VALEURS DE REFERENCE 4
DONNER LES PONDERATIONS : 0.1 0.2 0.3 0.4
DONNER LES VALEURS : 123 34 F45\\234 567
INITIALISEZ LES 7 VARIABLES ALEATOIRES : 0.1 0.2 0.3 0.4 0.5 0.6 0.7
PROBA DE PROLONG. DE CRISE FAIBLE, MOYENNE, FORTE 0.9 0.9 0.9
PROBABILITE DE NOUVELLE CRISE 0.2
ECART TYPE MAXI RETENU 10
TEMPS MAXIMUM DE L'ETUDE : 1000
PERIODE D'AFFICHAGE : 10
PROBA DE CRISE FAIBLE 0.2
PROBA DE CRISE MOYENNE 0.2
DONC PROBA DE CRISE FORTE= 0.6

10.00	306.663				*
20.00	289.244				*
30.00	293.560				*
40.00	302.821				*
50.00	263.573			*	
60.00	233.368		*		
70.00	251.426		*		
80.00	313.968			*	
90.00	281.149			*	
100.00	357.038				*
110.00	344.714				*
120.00	324.328			*	
130.00	246.423		*		
140.00	213.309		*		
150.00	183.408		*		
160.00	170.908	*			
170.00	176.672	*			
180.00	167.675	*			
190.00	168.023	*			
200.00	180.756	*			
210.00	193.611	*			
220.00	212.129	*			
230.00	120.277	*	*		
240.00	99.181	*	*		
250.00	52.684	*	*		
Est					
740.00	304.445				*
750.00	304.908				*
760.00	304.635				*
770.00	387.524			*	
780.00	411.164			*	
790.00	482.880				*
800.00	490.531				*
810.00	482.166				*
820.00	471.431			*	
830.00	336.517		*		
840.00	297.841		*		
850.00	378.092			*	
860.00	397.803			*	
870.00	488.334				*
880.00	568.717				*
890.00	589.781				*
900.00	487.757			*	
910.00	369.544		*		
920.00	290.099		*		
930.00	267.706		*		
940.00	261.734		*		
950.00	250.799		*		
960.00	238.409		*		
970.00	219.000		*		
980.00	182.483	*			
990.00	183.883	*			
1000.00	134.026	*			

TERMINE

DONNER LE NOMBRE DE VALEURS DE REFERENCE 4
DONNER LES PONDERATIONS : 0.1 0.2 0.3 0.4
DONNER LES VALEURS : 123 234 345 739
INITIALISEZ LES 7 VARIABLES ALEATOIRES : 0.1 0.2 0.3 0.4 0.5 0.6 0.7
PROBA DE PROLONG. DE CRISE FAIBL, MOYENNE, FORTE 0.5 0.5 0.5
PROBABILITE DE NOUVELLE CRISE 0.3
ECART TYPE MAXI RETENU 1000
TEMPS MAXIMUM DE L'ETUDE : 1000
PERIODE D'AFFICHAGE : 10
PROBA DE CRISE FAIBLE 0.2
PROBA DE CRISE MOYENNE 0.4
DONC PROBA DE CRISE FORTE= 0.40000006



SIMULATION DE L'EVOLUTION DE

DEUX POPULATIONS

Dans une région se trouvent en présence des RENARDS et des COQS de BRUYERE. Les coqs ou victimes sont atteints d'une maladie qui n'est pas mortelle mais qui les rend vulnérables aux renards. Plus précisément, si un renard rencontre un coq de bruyère malade il le mange ; si le coq est sain, le renard n'arrive pas à l'attraper. On suppose en outre que si un coq malade rencontre un coq sain, il le contamine. Les coqs ne manquent pas de nourriture dans la région mais s'ils sont peu nombreux les renards peuvent manquer de nourriture et un certain nombre d'entre eux disparaîtra. Chaque population, isolément, a un taux de reproduction constant.

Modèle.

Nous simulons la région où évoluent ces deux populations par un échiquier*. Les renards et les coqs se déplacent au hasard sur cet échiquier c'est-à-dire qu'à chaque instant on tire, au hasard avec remise, le numéro de la case de chaque animal. Deux animaux se rencontrent s'ils sont dans la même case à un instant : un renard mange un coq malade et un coq malade contamine un coq sain. A chaque instant, une proportion (à déterminer) des renards n'ayant rien mangé disparaît.

L'intérêt du programme, réalisé en conversationnel, est de permettre aux élèves en faisant varier les paramètres (nombre d'individus, proportion de malades, taux de reproduction, proportion de renards qui disparaît lorsque la nourriture manque) de constater des phénomènes très différents allant de l'extinction de la population de renards à un équilibre relatif des renards et des coqs. Ainsi, en modifiant la surface de la région, on peut augmenter ou diminuer la probabilité de rencontre.

Le programme prévoit dans ce but une période d'affichage (PER) qui peut varier en cours d'exécution et permettre aux élèves ayant bien vu l'évolution pas à pas de saisir plus rapidement les répercussions d'un paramètre (équilibre ou extinction) en lui donnant de nombreuses valeurs distinctes.

* ou plutôt une surface découpée en cases numérotées, 8 ha correspondant à 80 cases.

Le tirage se fait à partir de la fonction ALE qui donne un nombre au hasard suivant la loi uniforme sur $[0,1[$, le numéro de la case étant la partie entière de $80 \cdot (ALE(0)) + 1$ pour une surface de 8 ha.

LISTER A PARTIR DE 1

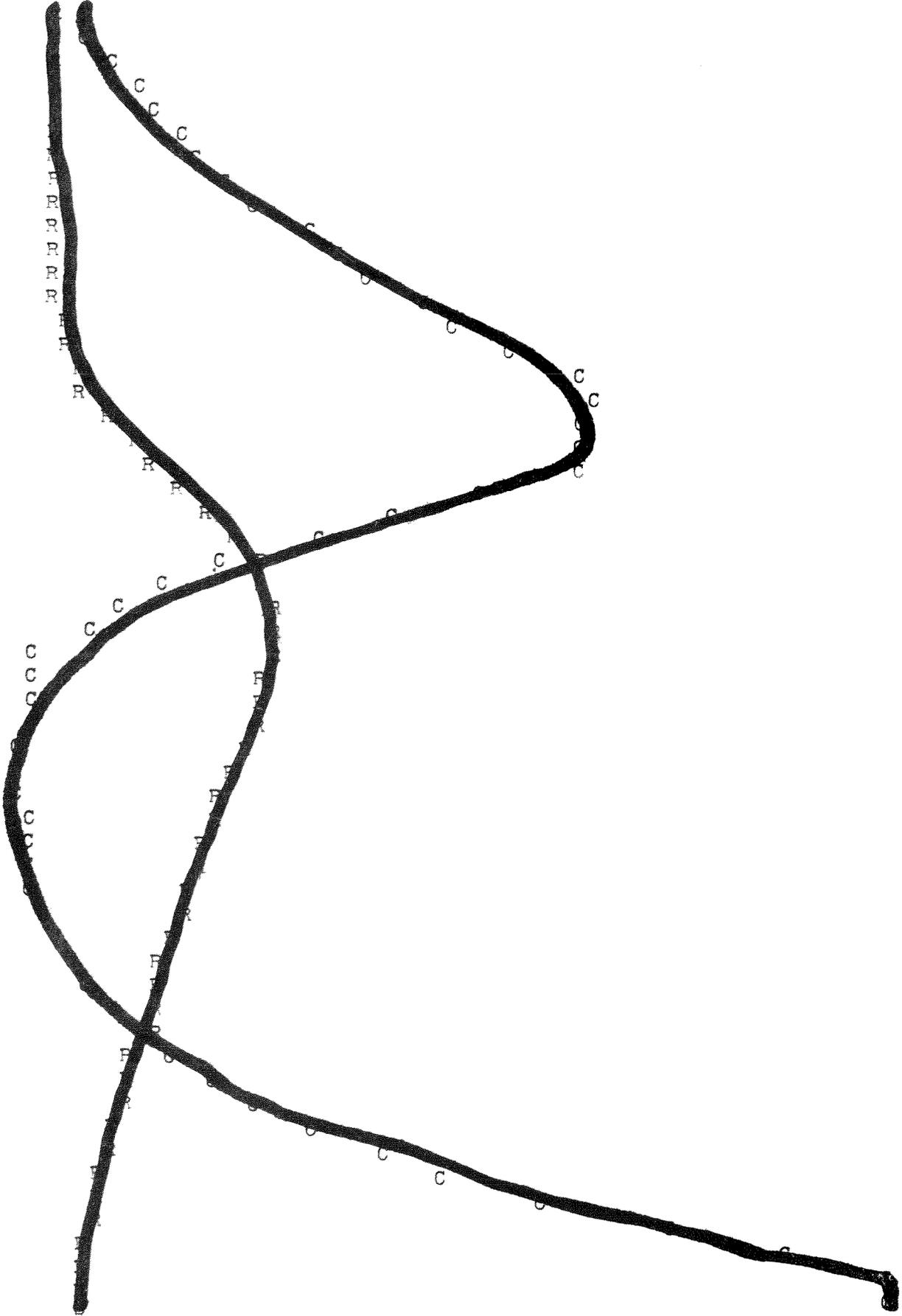
```

1
7*****ENTREE DES DONNEES*****
8 AFFICHERI/, 'DONNER L 'OPTION 1:NOMBRES 2: COURBES 'J;LIRE PER
10 AFFICHERI/, 'DONNER LE NOMBRE DE RENARDS 'J;LIRE NA
12 AFFICHERI/, 'DONNER LE NOMBRE DE COQS 'J;LIRE NCOQ
14 AFFICHERI/, 'DONNER LA PROPORTION DE MALADES 'J;LIRE PROP
15*****CALCUL DU NOMBRE DE COQS MALADES ET EN BONNE SANTE*****
16 NB1=NCOQ*(1-PROP)
17 NB2=NCOQ*PROP
20 AFFICHERI/, 'DONNER LE TAUX DE CROISSANCE DES RENARDS ET DES COQS 'J
22 LIRE TA,TB
24 AFFICHERI/, 'DONNER LA PROPORTION DES MORTS DE FAIM 'J;LIRE FA
26 AFFICHER 'DONNER TMAX PER 'J;LIRE TMAX,PER
28 AFFICHERI/, 'DONNER LA SURFACE EN APES 'J;LIRE NS
29*****CALCUL DU NOMBRE DE CASES DE 10M2*****
30 NS=NS*10
31*POUR DES RAISONS DE PROGRAMME
32 TABLEAU TABA[1], TAB1[1], TAB2[1]
34 TABA[1]--1; TAB1[1]--2; TAB2[1]--3
35
39*****SIMULATION*****
40 FAIRE 98 POUR T-1 JUSQUA TMAX
41*****PLACEMENT DES RENARDS*****
42 &PLAC(NA, NS, TABA)
43*****PLACEMENT DES COQS SAINS*****
44 &PLAC(NB1, NS, TAB1)
45*PLACEMENT DES COQS MALADES *****
46 &PLAC(NB2, NS, TAB2)
47*****RENCONTRE DES RENARDS ET DES COQS*****
48 BOF=0
50 FAIRE 56 POUR I-1 JUSQUA NA
52 FAIRE 55 POUR J-1 JUSQUA NB2
54 SI TABA[I]#TAB2[J] ALORS ALLER EN 55; TAB2[J]--1; BOF=BOF+1
55
56
59*****RENCONTRE DES COS MALADES ET DES COQS SAINS*****
60 CONT=0; FAIRE 70 POUR I-1 JUSQUA NB2
62 FAIRE 69 POUR J-1 JUSQUA NB1
64 SI TAB2[I]#TAB1[J] ALORS ALLER EN 69
66 TAB1[J]--1; CONT=CONT+1
69
70
71****COMPTAGE DES EFFECTIFS REDUCTIONS ET AUGMENTATIONS *****
72 NA=(NA-(NA-BOF)*FA)*(1+TA)
73 NB1=(NB1-CONT)*(1+TB)
74 NB2=(NB2-BOF+CONT)*(1+TB)
75 NCOQ=NB1+NB2; SI NCOQ=0 ALORS ALLER EN 79
77 PROP=NB2/NCOQ
79 SI T/PER#ENT(T/PER) ALORS ALLER EN 98
81 SI REP#1 ALORS ALLER EN 84
82 AFFICHERI/, ' T= ', F4.0, ' NA= ', F4.0, ' NC= ', F4.0]T, NA, NCOQ
83 AFFICHER(X, F2.2, ' % 'ENT(100*PROP); ALLER EN 98
84 COR=0.5
86 AFFICHERI/, F2.2, ' % 'ENT(100*PROP); SI NA<NCOQ ALORS ALLER EN 90
87 AFFICHER[*X, 'C', *X, 'R']ENT(NCOQ*COR), ENT((NA-NCOQ)*COR)
89 ALLER EN 98
90 AFFICHER[*X, 'R', *X, 'C']ENT(NA*COR), ENT((NCOQ-NA)*COR)
98
99 TERMINER
100 PROCEDURE &PLAC(U, V, B) LOCAL V, U
102 LIBERER B
103 SI ENT(U)=0 ALORS DEBUT TABLEAU B[1]; B[1]--19; RETOUR FIN
104 TABLEAU B[ENT(U)]
106 Z=0; FAIRE 108 POUR I-1 JUSQUA ENT(U)
108 Z=ALE(Z); B[I]=ENT(V*Z+1)
109 RETOUR

```

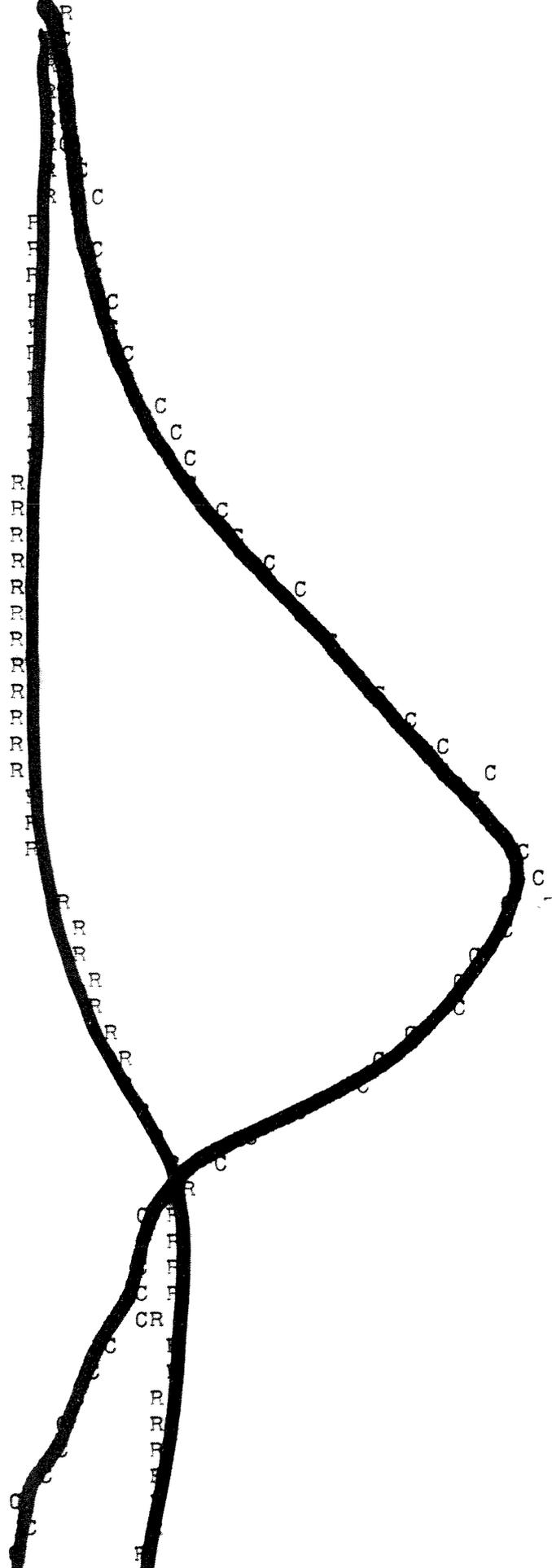
DONNER L'OPTION 1:NOMBRES 2:COURBES 2
DONNER LE NOMBRE DE RENARDS 10
DONNER LE NOMBRE DE COQS 10
DONNER LA PROPORTION DE MALADES 0.4
DONNER LE TAUX DE CROISSANCE DES RENARDS ET DES COQS 0.2 0.2
DONNER LA PROPORTION DES MORTS DE FAIM 0.2
DONNER TMAX,PER1000 1
DONNER LA SURFACE EN ARES 8

40.00 %
34.00 %
34.00 %
34.00 %
36.00 %
36.00 %
41.00 %
46.00 %
56.00 %
67.00 %
70.00 %
76.00 %
89.00 %
91.00 %
97.00 %
97.00 %
98.00 %
98.00 %
98.00 %
97.00 %
98.00 %
97.00 %
96.00 %
97.00 %
96.00 %
94.00 %
91.00 %
81.00 %
74.00 %
64.00 %
58.00 %
38.00 %
23.00 %
23.00 %
40.00 %
31.00 %
20.00 %
8.00 %
8.00 %
8.00 %
8.00 %
1.00 %
1.00 %
1.00 %
1.00 %
1.00 %
1.00 %
1.00 %
1.00 %
1.00 %
1.00 %
1.00 %
1.00 %
1.00 %
1.00 %
1.00 %
3.00 %
11.00 %



DONNER L'OPTION 1: NOMBRES 2: COURBES 2
 DONNER LE NOMBRE DE RENARDS 10
 DONNER LE NOMBRE DE COQS 10
 DONNER LA PROPORTION DE MALADES 0.4
 DONNER LE TAUX DE CROISSANCE DES RENARDS ET DES COQS 0.2 0.1
 DONNER LA PROPORTION DES MORTS DE FAIM 0.2
 DONNER TMAX, PER1000 1
 DONNER LA SURFACE EN ARES 8

37.00 %
 37.00 %
 30.00 %
 30.00 %
 22.00 %
 14.00 %
 14.00 %
 14.00 %
 22.00 %
 29.00 %
 19.00 %
 19.00 %
 14.00 %
 14.00 %
 9.00 %
 9.00 %
 9.00 %
 9.00 %
 12.00 %
 16.00 %
 16.00 %
 22.00 %
 23.00 %
 24.00 %
 29.00 %
 38.00 %
 48.00 %
 68.00 %
 80.00 %
 90.00 %
 94.00 %
 96.00 %
 99.00 %
 98.00 %
 98.00 %
 98.00 %
 98.00 %
 98.00 %
 97.00 %
 99.00 %
 99.00 %
 99.00 %
 98.00 %
 98.00 %
 98.00 %
 97.00 %
 97.00 %
 96.00 %
 96.00 %
 96.00 %
 94.00 %
 93.00 %
 92.00 %
 99.00 %
 99.00 %
 99.00 %
 99.00 %
 99.00 %
 99.00 %



(suite de la p. 14)

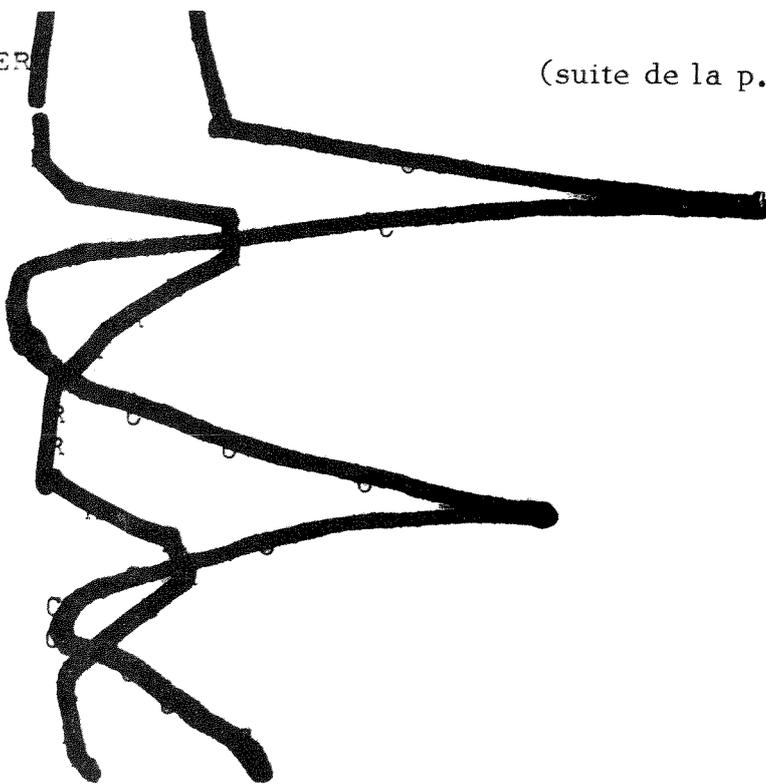
99.00	%		
98.00	%		
98.00	%		
98.00	%		
98.00	%		
98.00	%		
98.00	%		
97.00	%		
97.00	%		
97.00	%		
96.00	%		
96.00	%		
96.00	%		
96.00	%		
96.00	%		
96.00	%		
96.00	%		
96.00	%		
92.00	%		
92.00	%		
89.00	%		
89.00	%		
84.00	%		
84.00	%		
75.00	%		
75.00	%		
75.00	%		
75.00	%		
75.00	%		
64.00	%		
64.00	%		
64.00	%		
64.00	%		
64.00	%		
64.00	%		
64.00	%		
64.00	%		
64.00	%		
84.00	%		
81.00	%		
81.00	%		
81.00	%		
81.00	%		
77.00	%		
73.00	%		
73.00	%		
73.00	%		
73.00	%		
87.00	%		
86.00	%		
86.00	%		
86.00	%		
84.00	%		
82.00	%		
80.00	%		
78.00	%		
75.00	%		
75.00	%		
75.00	%		
73.00	%		
70.00	%		
67.00	%		
67.00	%		
67.00	%		
81.00	%		
81.00	%		
81.00	%		

Temps accéléré x 10

PER-10
CONTINUER
C

(suite de la p.15)

81.00 %
99.00 %
99.00 %
99.00 %
99.00 %
91.00 %
68.00 %
68.00 %
57.00 %
57.00 %
79.00 %
92.00 %
99.00 %
96.00 %
93.00 %
66.00 %
42.00 %
59.00 %
93.00 %
97.00 %
99.00 %



POUR UNE GEOGRAPHIE QUANTITATIVE EN SECOND
CYCLE DE LYCEE.

PLAN.

INTRODUCTION.

- A- Informatique et second degré
- B- Le système SESAM
- C- Buts poursuivis avec les élèves
- D- Méthodes de travail avec les élèves
 - 1° Analyse régionale qualitative
 - 2° Analyse régionale quantitative

I LA CLASSIFICATION REGIONALE ET SES DONNEES.

- A- Les matrices géographiques
- B- Une représentation multi-dimensionnelle
- C- Etapes préalables au traitement multi-dimensionnel

II CLASSIFICATION REGIONALE ET ANALYSE EN COMPOSANTES PRINCIPALES

- A- Les principes de la méthode
- B- Les étapes
 - 1° La matrice factorielle des "saturations"
 - 2° La matrice factorielle des "poids locaux"
 - 3° Exemples traités avec des élèves.
- C- Inconvénients de la méthode.

III CLASSIFICATION REGIONALE ET ANALYSE HIERARCHIQUE.

- A- La mesure de la "ressemblance" entre régions
 - 1° La distance euclidienne
 - 2° Métrique et ultramétrique
- B- Exemples de travaux avec les élèves

IV UNE METHODE MANUELLE : LA MATRICE ORDONNABLE DE BERTIN.

- A- Passage de la matrice quantitative à la matrice graphique ordonnable
- B- Simplification de l'image graphique et typologie

V AUTRES APPLICATIONS : SIMULATIONS, BANQUES DE DONNEES, TRAITEMENTS D'ENQUETES.

- A- Les simulations de phénomènes économiques et géographiques
- B- Les banques de données
- C- Les enquêtes
- D- Pour mémoire

CONCLUSION.

Qu'en pensent les élèves ?

INTRODUCTION.

A- Informatique et second degré.

L'introduction par l'I.N.R.D.P. de l'informatique dans le second degré long, la formation de 400 collègues de toutes disciplines et l'équipement de 53 établissements en mini-ordinateurs ont donné l'idée et les moyens d'appliquer au niveau du second degré des méthodes courantes dans l'enseignement supérieur.

Les simulations, les indexations lexicographiques de textes longs, les traitements statistiques élaborés de chroniques ou de données économiques sont devenus possibles.

B- Le système SESAM.

L'auteur, dont l'établissement a été équipé en novembre 1974, a mis au point pour l'expérience informatique I.N.R.D.P. un logiciel qui permet des traitements statistiques perfectionnés de matrices de données géographiques (voir graphe 3).

Ce système a été dénommé SESAM pour : Software d'Etudes Spatiales et d'Analyse Multidimensionnelle. Ce système écrit en LSE a été conçu pour être entièrement utilisé par les élèves de seconde et première.

C- Buts poursuivis avec les élèves.

Nous souhaitons avec ce système :

1° Permettre aux élèves une véritable recherche scientifique sur des données réelles et à jour sur des thèmes précis, par exemple, la recherche dans une masse importante de données des types principaux de régions économiques françaises ou les principales classifications possibles des végétations ou des climats terrestres.

Les exemples traités ici sont les suivants : classification des régions françaises (22 régions de programme en classe de première) dans le domaine de la démographie, de l'industrie. Typologie des principaux climats terrestres (en classe de seconde).

- 2° Donner aux élèves l'habitude et les moyens de dominer une documentation massive ce qui sera une des difficultés majeures de leur vie future de citoyen.
- 3° Donner aux élèves une attitude critique face aux documents statistiques : les traitements statistiques raffinés ne rendent pas sensées des données aberrantes.
- 4° Donner aux élèves un comportement méthodologique plutôt que des connaissances périmées aussitôt qu'apprises.
- 5° Accéder à une géographie plus quantitative et plus formalisée. Les méthodes proposées sont une amorce de solution pour abandonner une géographie de guide bleu purement descriptive et qualitative.
- 6° Amener les élèves à se rendre compte de ce qu'est un ordinateur, ce qu'on peut lui demander, ce qu'on ne peut pas lui demander, et démystifier la machine auprès de ses futurs utilisateurs.

D- Méthodes de travail avec les élèves.

Le travail des élèves est organisé pendant 1 à 2 mois (soit 30 heures) sur le même thème de géographie ou d'histoire. Les travaux sont faits par équipe sur des sujets différents dont chaque équipe rend compte par polycopie distribuée à tous et par exposé oral. Le professeur fixe les cadres de travail, introduit le thème par un cours de quelques heures. Chaque équipe rédige le compte-rendu de son travail et distribue le résultat à toute la classe.

Il y a dans chaque dossier deux parties principales :

1° Analyse qualitative.

Dans un premier temps les élèves rédigent des compte-rendus à partir d'articles de revues ou de chapitres d'ouvrages distribués par le professeur. Cette moisson de compte-rendus (entre 10 et 20) forme la première partie du dossier qui est mis en commun avec un bref exposé de 5 minutes de chaque équipe.

2° Analyse quantitative.

Un élément essentiel du travail est la collaboration du professeur de mathématiques qui traite toutes les notions statistiques qui seront nécessaires pour la suite (Moyenne, écart-type, variance, standardisation des séries etc ...).

Quelques séances sont consacrées à l'initiation au maniement des consoles de visualisation.

La bases des données pour l'étude des régions françaises est fournie par le volume annuel de "STATISTIQUES ET INDICATEURS DES REGIONS FRANCAISES" [1].

Le cadre des 22 régions de programme est très pratique pour le second cycle, (mise à jour annuelle, faible coût de la source, volume de données raisonnable).

La recherche des séries statistiques susceptibles de représenter le phénomène que l'on veut étudier se fait à l'aide d'ouvrages (pour l'exemple démographique D. NOIN [21]). Les élèves définissent par équipe les notions utilisées, en font la critique. [7] [8].

Puis l'ensemble de la démarche expliquée dans les paragraphes suivants est poursuivie grâce au système SE SAM sur console conversationnelle. Les équipes d'élèves sont responsables de diverses tâches de relevés numériques, cartographiques etc ... Les résultats sont photocopiés et figurent dans le dossier de chacun. (Les cartes et graphiques sont photocopiés en couleurs et ne sont pas publiables, quelques exemplaires ont été refaits en noir et blanc pour cette publication).

=====

I LA CLASSIFICATION REGIONALE ET SES DONNEES.

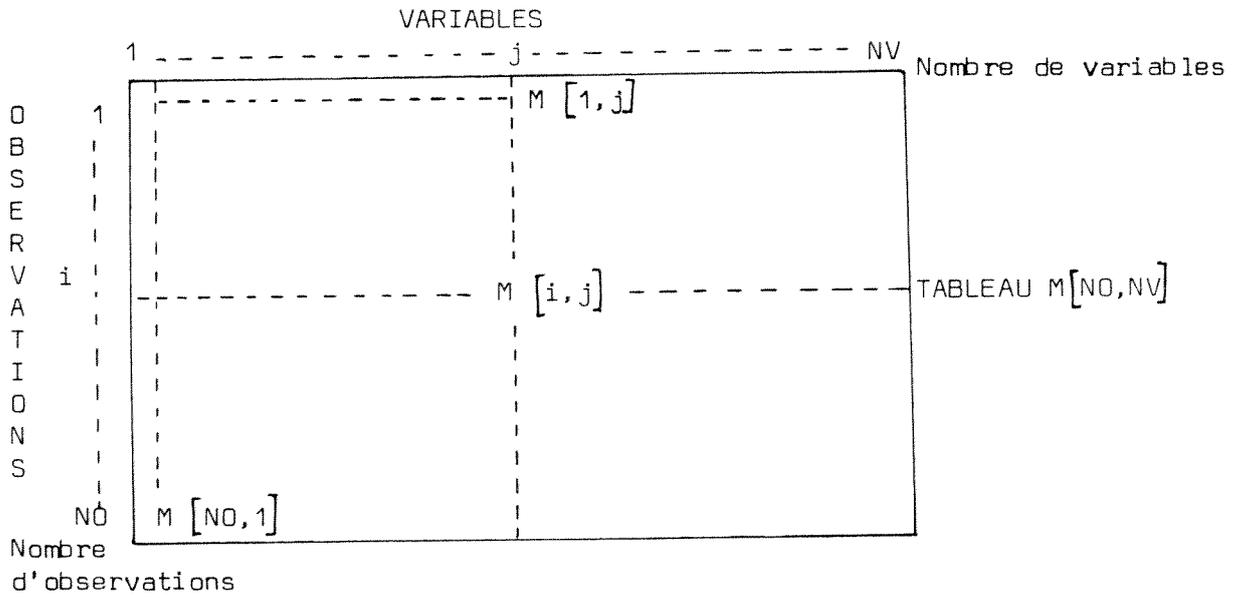
A- Les matrices géographiques.

Les données de base se présentent sous la forme de tableaux à double entrée (voir graphe 1). En colonne on trouve les variables (ou indicateurs ou caractères) affectées aux régions qui se trouvent en ligne.

Pour ce qui concerne l'exemple de démographie, la liste des variables se trouve graphe 5 ; l'exemple des régions industrielles françaises est ici moins détaillé : seul le résultat est présenté.

LA REPRESENTATION MATRICIELLE

G1



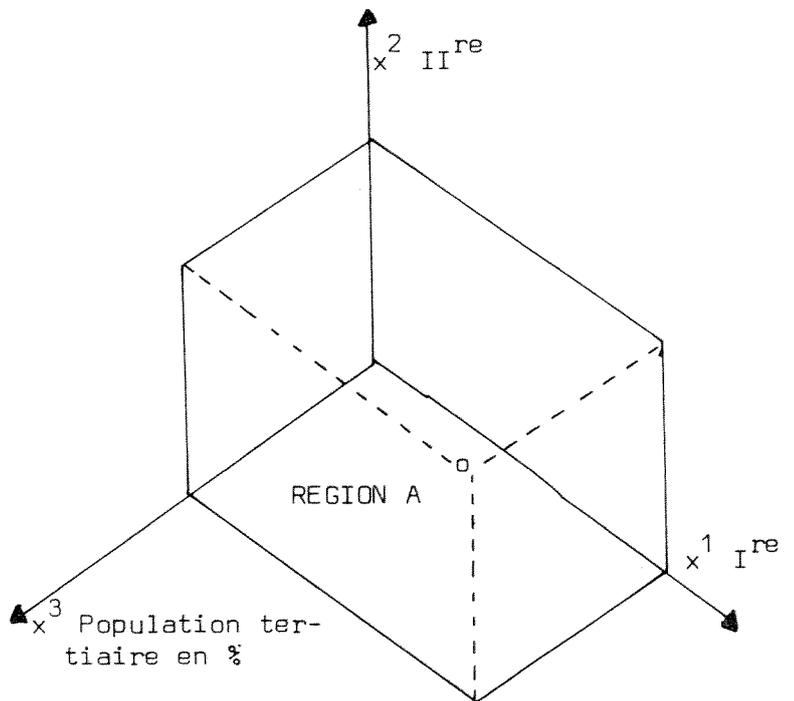
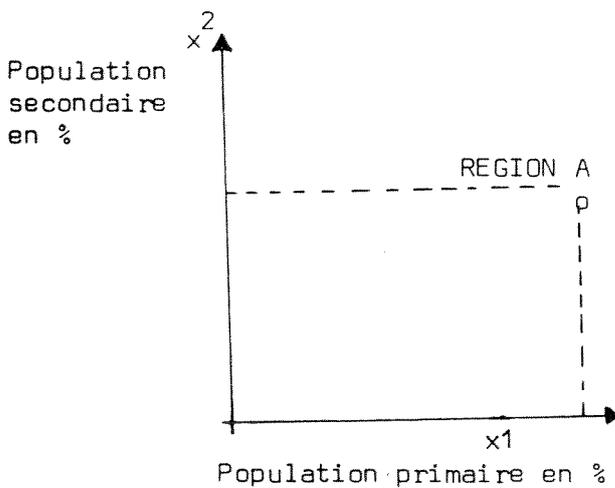
LA REPRESENTATION MULTIDIMENSIONNELLE

G2

2. Représentation d'un point

DANS UN ESPACE A DEUX DIMENSIONS,

A TROIS DIMENSIONS



etc, essayez de rajouter 18 dimensions !

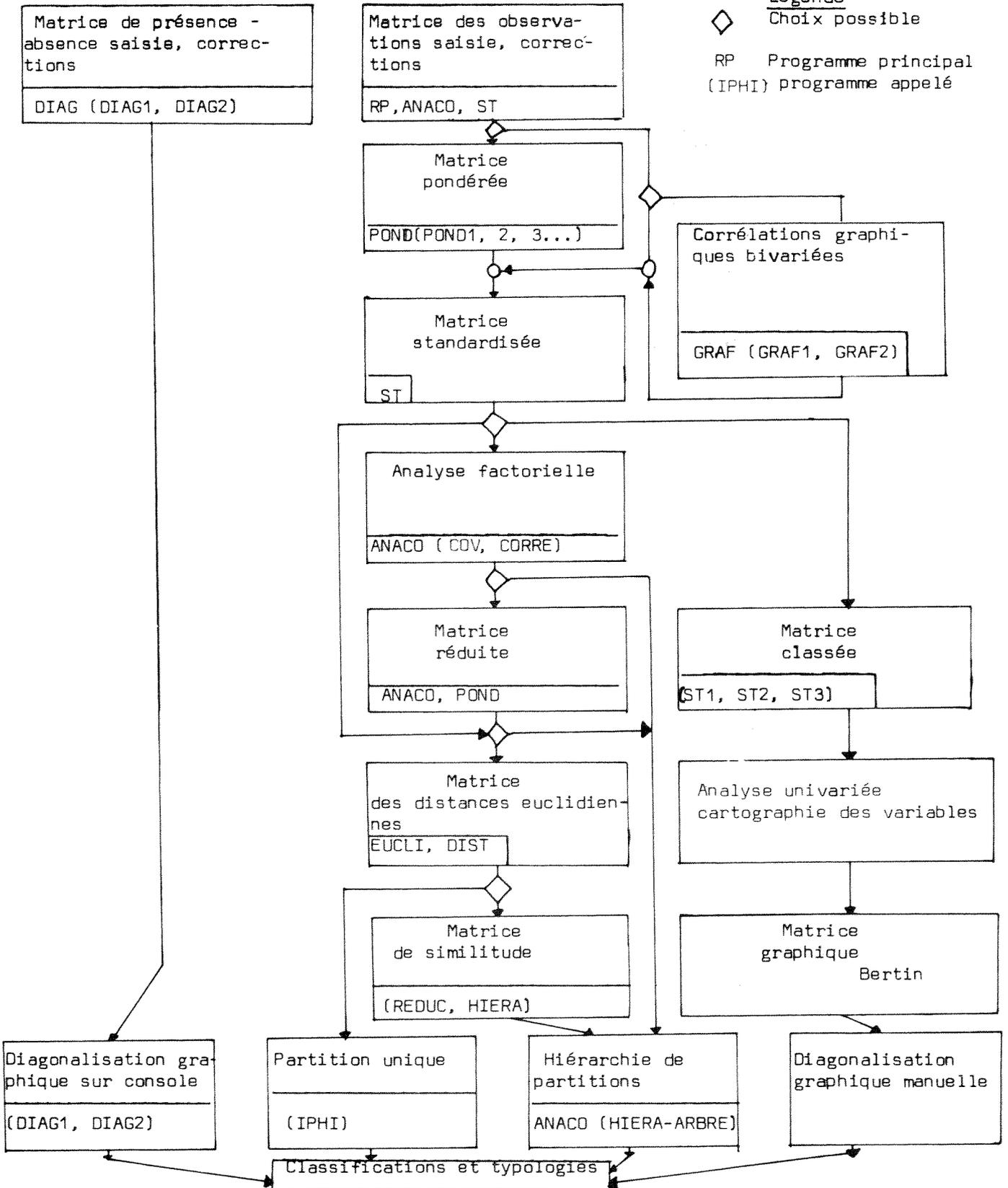
SOFTWARE D'ETUDES SPATIALES ET D'ANALYSE MULTIDIMENSIONNELLE

MATRICE BOOLEENNE

MATRICE QUANTITATIVE OU BOOLEENNE

Légende
◇ Choix possible

RP Programme principal
(IPHI) programme appelé



Le système proposé comporte 58 modules et 16000 mots MITRA 15.

LES OPPOSITIONS DEMOGRAPHIQUES
DES REGIONS FRANCAISES

65

N° de colonne dans la
matrice du fichier DEMO

N° de pages dans
SIRF 73

VARIABLES DEMOGRAPHIQUES

	<u>I - Les classes d'âge</u>	
1	Hommes 0-25 ans 1968 (% de la population totale)	17
2	Femmes 0-25 ans 1968 (% de la population totale)	
3	Hommes 25-44 ans 1968 (% de la population totale)	
4	Femmes 25-44 ans 1968 (% de la population totale)	
5	Hommes 65 et plus 1968 (% de la population totale)	
6	Femmes 65 et plus 1968 (% de la population totale)	
7	Indice de vieillesse = $\frac{100 \cdot V [65 \text{ et } +]}{J [0-25]}$	
	<u>II - Fécondité - mortalité</u>	
8	Taux de fécondité générale (67-69)	34
9	Durée moyenne de vie Hommes (67-69)	
10	Durée moyenne de vie Femmes (67-69)	
	<u>III - Comportement global</u>	
11	Taux net de reproduction	
12	Taux d'accroissement naturel	32

CODE DES 22 REGIONS DE PROGRAMME

PAR	REGION PARISIENNE	BRE	BRETAGNE
CHA	CHAMPAGNE ARDENNE	PCH	POITOU CHARENTES
PIC	PICARDIE	AQU	AQUITAINE
HNO	HAUTE NORMANDIE	MID	MIDI PYRENEES
CTR	CENTRE	LIM	LIMOUSIN
BNO	BASSE NORMANDIE	RHA	RHONE ALPES
BOU	BOURGOGNE	AUV	AUVERGNE
NOR	NORD	ROU	LANGUEDOC ROUSSILLON
LOR	LORRAINE	PCA	PROVENCE CÔTE D'AZUR
ALS	ALSACE	COR	CORSE
FCT	FRANCHE COMTE	FRA	FRANCE
LOI	PAYS DE LA LOIRE		

Quant à l'exemple des climats, il concerne 16 stations qui sont mentionnées (graphe 22).

Dans le premier cas on a une matrice de 22 lignes (régions) \times 12 colonnes (variables).

Dans le deuxième cas 22 régions et 13 variables.

Dans le troisième cas 16 stations et 13 variables. Le terme $M(i, j)$ de la matrice M correspond à la valeur de la variable j pour la région ou station i . Nous appellerons NO le nombre de lignes et NV le nombre de colonnes du tableau.

Les variables fournies par l'INSEE ne sont pas toujours de bonne qualité et si l'on veut caractériser les régions il faut disposer du plus grand nombre possible de variables. Des matrices de 70 variables ne sont pas rares en géographie et supposent donc le maniement de $22 \times 70 = 1540$ nombres pour élaborer une typologie ! Si l'échelle du découpage est plus petite on peut avoir 95 départements \times 70 variables, soit 6650 nombres.

B- En quoi cette représentation matricielle est-elle multidimensionnelle ?

Le graphe n° 2 explique ce dont il s'agit : on considère que chaque région est représentée par un point à NV (autant que de variables) coordonnées dans un espace à NV dimensions. De même chaque variable est représentée par un point à NO coordonnées dans un espace à NO dimensions. Au-delà de trois dimensions, ce type d'espace n'est plus utilisable autrement que par des méthodes mathématiques.

C- Etapes préalables au traitement multidimensionnel. (Voir graphe 3)

1° Saisie des données par les élèves.

(Graphe 3 programmes RP). Les équipes de travail "rentrent" en machine les données tirées des documents INSEE. Ce travail se fait au clavier des consoles, des procédures de vérification et de corrections sont prévues. Chaque équipe de travail est responsable d'une ou deux variables.

2° Eliminer l'effet de taille sur la matrice des données brutes.

Les valeurs brutes de la matrice origine doivent être ramenées à des pourcentages. Les programmes conversationnels sont prévus pour que cette étape soit faite par le calculateur sous la direction des élèves. (Graphe 3 programme POND).

3° "Standardisation" de la matrice.

(Graphe 3 programme ST). L'hétérogénéité des variables du point de vue des unités utilisées et des nomenclatures rend nécessaire cette étape de "standardisation".

Cette opération est simple : les élèves mettent en oeuvre le programme qui opère cette transformation. Que fait le programme ?

- a) Calcul de la moyenne de la série des 22 observations pour la variable considérée, puis de l'écart-type.
- b) Centrage de la série : on soustrait à chaque valeur de la série sa moyenne.
- c) Réduction : on divise le résultat par l'écart-type.

Chaque valeur est donc remplacée par son rapport entre son écart à la moyenne et l'écart-type. Cette étape est d'autant plus intéressante pour le géographe qu'elle va permettre une cartographie univariée très facile.

4° Cartographie univariée.

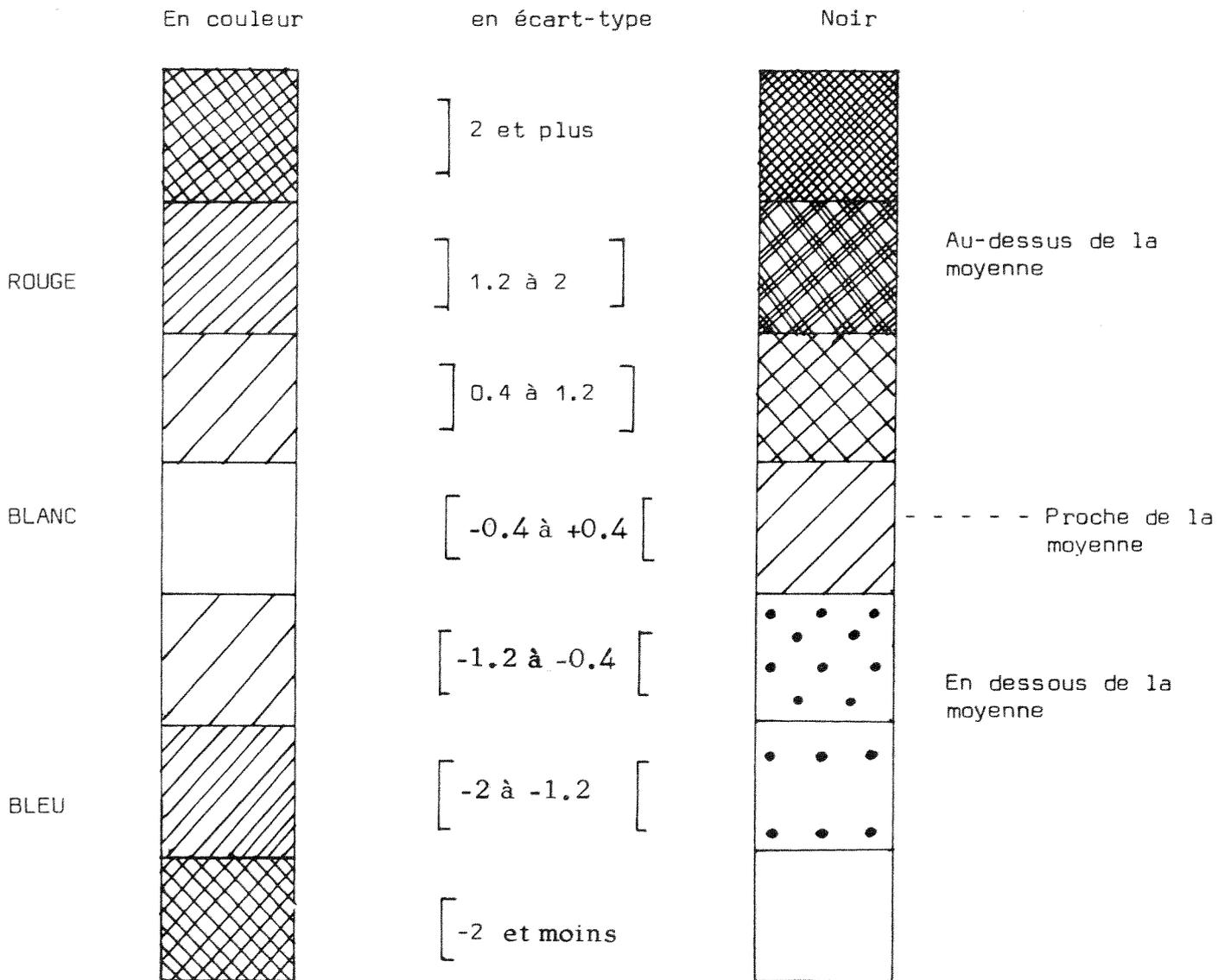
Cette cartographie n'a évidemment rien de nouveau : c'est une étape essentielle fondamentale du travail géographique. La standardisation facilite seulement la cartographie et la comparaison des séries entre elles. En effet, si l'on étudie les "oppositions démographiques de la France", il est très intéressant de raisonner en terme d'écarts à la moyenne pour se donner une base de comparaison.

Les classes cartographiables qui vont permettre de traduire les valeurs centrées réduites en paliers visuels, figurent sur le graphe N° 4. La légende des cartes en couleurs est faite pour mettre en valeur les écarts à la moyenne (rouge pour les écarts positifs, bleu pour les écarts négatifs, blanc pour les valeurs proches de la moyenne).

Les élèves élaborent NV cartes univariées, c'est-à-dire autant que de variables. (Exemples graphe N° 7). Ces cartes sont au nombre de 12 pour le dossier démographique, 13 pour le dossier industriel etc...

CLASSES GRAPHIQUES CARTOGRAPHIABLES
D'UNE MATRICE QUANTITATIVE

G4



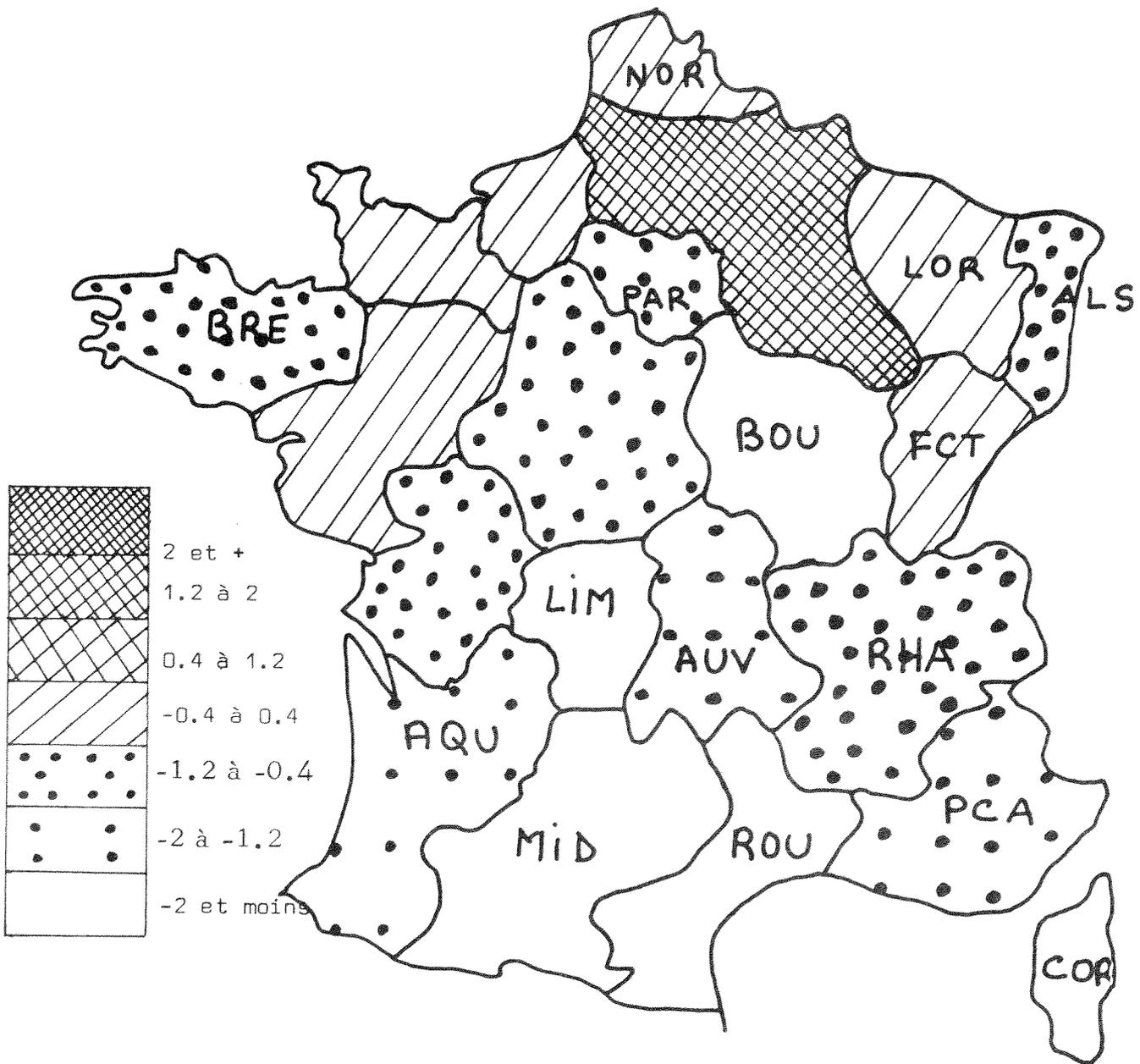
Solution choisie : découpage arbitraire en 0.8 écart-type.

OPPOSITIONS DEMOGRAPHIQUES DE LA FRANCE : MATRICE 12 VARIABLES

TAUX DE FECONDITE GENERALE 1969-1971

G7

(Variable centrée et réduite découpée en 7 classes de 0.8 écart-type centrées sur la moyenne).



Ces cartes polycopiées figurent également au dossier de chaque élève. Les équipes recherchent dans les ouvrages les explications de la répartition des différentes variables dans l'espace français et rendent compte de ce travail devant la classe.

5° Corrélations graphiques bivariées.

Les élèves peuvent rechercher sur console l'allure graphique des corrélations des variables par paire (Graphe 3 programme GRAF). Une fois la matrice entièrement standardisée, on peut aborder l'étape suivante.

II CLASSIFICATION REGIONALE ET ANALYSE EN COMPOSANTES PRINCIPALES.

A- Les principes de la méthode.

([10], [11], [12], [13], [17], [18].)

La démarche du géographe consiste après avoir élaboré les cartographies univariées à essayer de mettre en corrélation visuellement les différentes cartes qu'il étale devant lui. Cette opération facile avec 4-5 cartes devient très rapidement impossible. Les élèves se rendent très vite compte que l'oeil se sature et que l'on oublie des éléments importants de comparaison. Il faut pourtant, si l'on veut déboucher sur une typologie tenir compte des dimensions multiples en présence. Le modèle mathématique de l'analyse en composantes principales permet cette opération et peut rendre compte de manière synthétique d'un grand nombre de variables. Les "corrélations" entre variables ne sont plus visuelles mais calculées.

- Le programme rend possible la recherche des variables synthétiques ou "facteurs" qui résument au mieux le nuage multidimensionnel.
- Le programme opère la projection des points "observations" ou des points "variables" (qui se trouvent dans un "hyper-espace" inintelligible) dans un ou plusieurs espaces de dimension 2.

B- Les étapes.

1° La matrice des saturations.

Le programme calcule les coordonnées des points variables sur les "facteurs" retenus comme significatifs et ordonnés par importance décroissante : ce sont les "saturations". La matrice des saturations est tronquée aux facteurs retenant plus de 5 % de la variance (de l'information) ce qui amène dans les cas les plus fréquents à une matrice de (NV lignes) \times (3 colonnes-facteurs).

Seuls sont retenus les plans factoriels qui sauvegardent le maximum d'information.

L'analyse en composantes principales permet de substituer les facteurs aux variables de la matrice d'origine. La représentation graphique habituelle consiste à représenter les coordonnées des variables dans l'espace des facteurs sur les trois premiers plans déterminés par les axes 1 et 2, 1 et 3, 2 et 3.

(exemple graphe 8).

Le grand intérêt de cette projection est que les axes sont interprétables.

En effet, si géométriquement les axes factoriels sont les axes principaux d'inertie du nuage de points on peut également dire (GUIGOU) Bib. n° 12 que ce sont les plus grands communs dénominateurs des variables rangés en ordre décroissant.

Certaines machines trop petites ne peuvent "tenir" en mémoire qu'un nombre limité de variables (17 au maximum dans notre cas) la méthode permet de réduire une matrice de 68 variables, traitées en 4 étapes de 17 variables en une matrice de 4×3 facteurs = 12 colonnes, sauvegardant 80 à 90 % de l'information.

2° La matrice des "poids locaux".

La matrice des "facteurs scores" ou des "poids locaux" est la matrice fondamentale pour le géographe. Elle fournit les coordonnées des projections des points observations dans les plans factoriels. Les régions sont ordonnées sur chaque axe ce qui permet d'en tirer une typologie. (Ex. graphe 10).

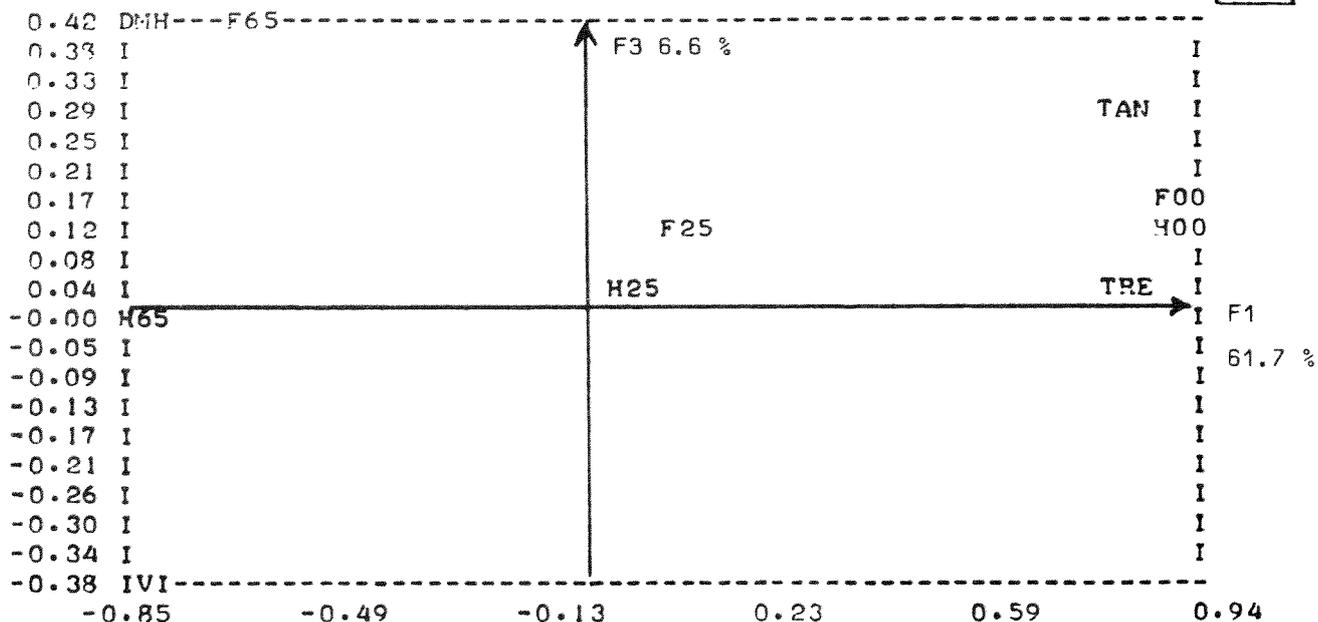
Une fois l'analyse faite, les élèves cartographient les 2 premières colonnes de la matrice des facteurs scores (exemple graphe n° 11) Pour des raisons de commodité cartographique les scores sont standardisés.

Par ailleurs, en délimitant les proximités sur les plans factoriels on peut déterminer une typologie, fonction de la signification des axes.

OPPOSITIONS DEMOGRAPHIQUES DES REGIONS FRANCAISES : FICHER 12 VARIABLES

PROJECTION DES POINTS VARIABLES DANS L'ESPACE DES FACTEURS AXES 1 ET 3

G8



POINTS DOUBLES:
DMH DMF F00 TFG

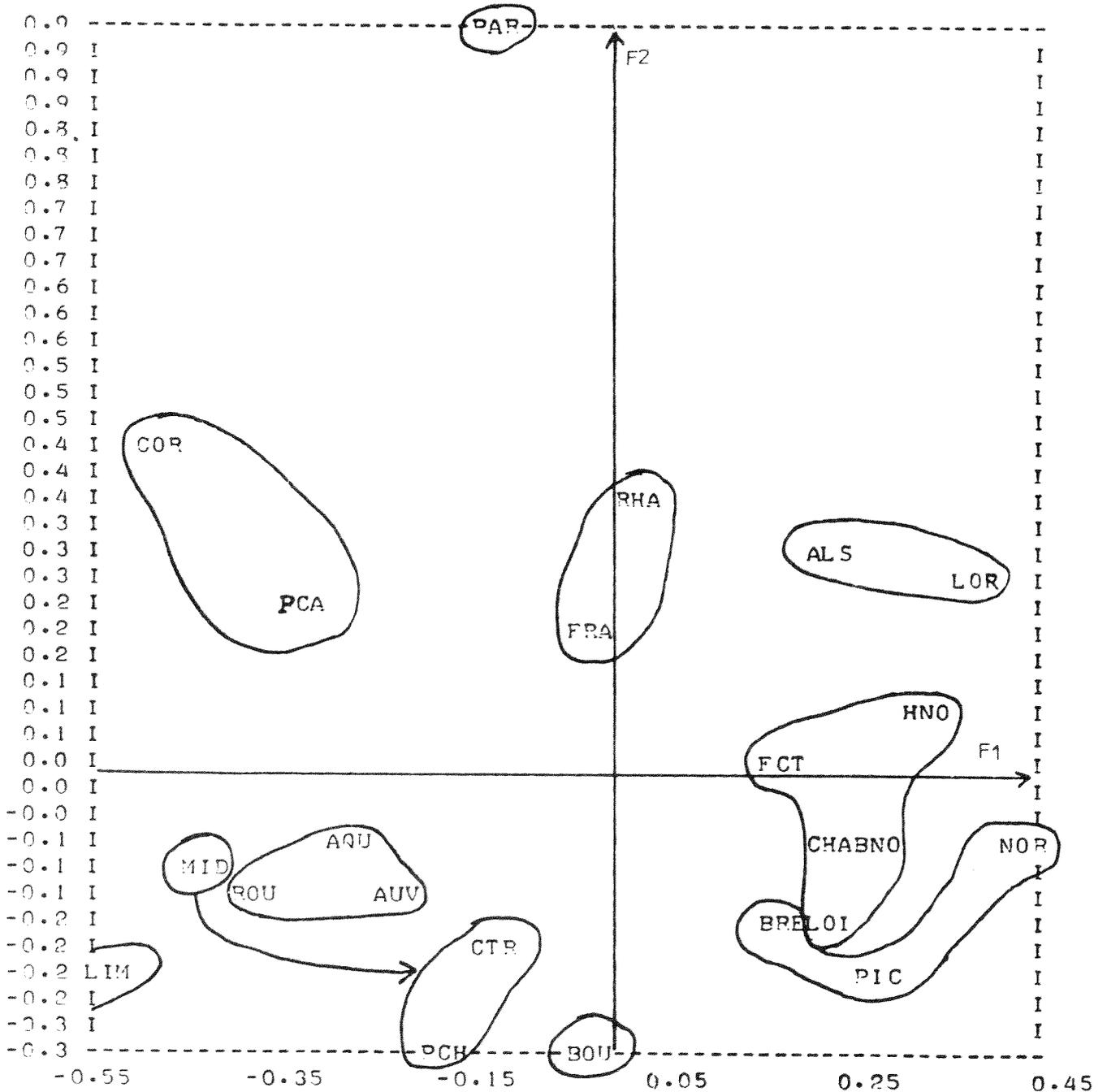
INTERPRETATION DE L'AXE I : Opposition des classes d'âges extrêmes.

CODES DES VARIABLES

H00	HOMMES 0 25 ANS 1968	
F00	FEMMES 0 25 ANS 1968	
H25	HOMMES 25 44 ANS 1968	
F25	FEMMES 25 44 ANS 1968	
H65	HOMMES 65 ANS ET + 1968	
F65	FEMMES 65 ANS ET + 1968	
IVI	INDICE DE VIEILLESSE VIEUX DE 65 ET +/JEUNES DE 0 25 ANS	1968
TFG	TAUX DE FECONDITE GENERALE	1967 1969
DMH	DUREE DE VIE HOMMES	1967 1969
DMF	DUREE DE VIE FEMMES	1967 1969
TRE	TAUX NET DE REPRODUCTION	1967 1969
TAN	TAUX D'ACCROISSEMENT NATUREL	1967 1969

OPPOSITIONS DEMOGRAPHIQUES DES REGIONS FRANCAISES : FICHIERS 12 VARIABLES
ANALYSE FACTORIELLE : PROJECTIONS DES POINTS OBSERVATIONS DANS L'ESPACE DES
FACTEURS AXES 1 ET 2

N.B. Les regroupements indiqués sont issus du traitement par analyse hiérarchique du fichier des trois premiers facteurs scores. (arbre, coupé au niveau 0.850 de similitude).



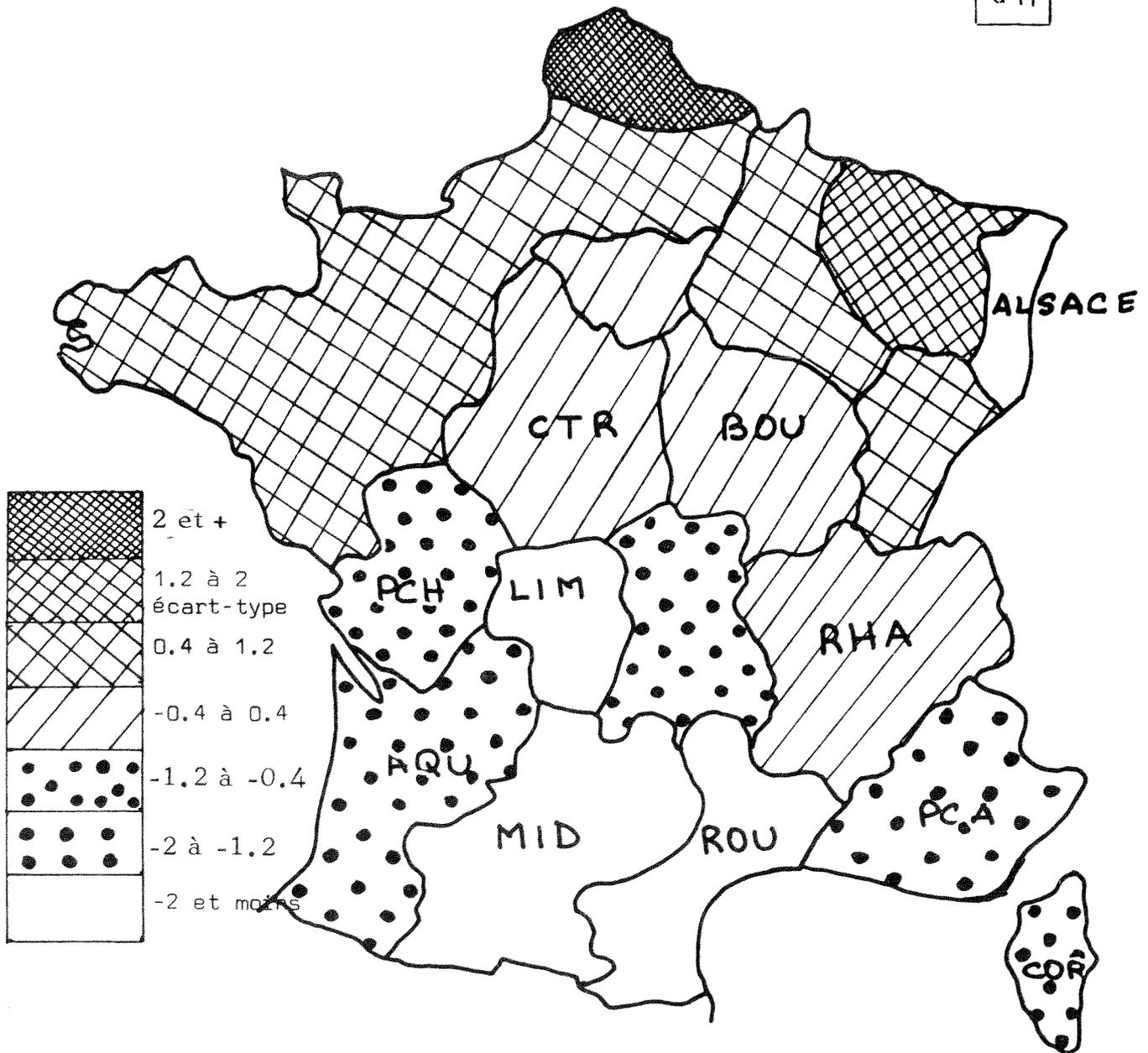
Le code des 22 régions de programme se trouve graphe 5.

CLASSIFICATION DES REGIONS DEMOGRAPHIQUES FRANCAISES : MATRICE 12 VARIABLES

PREMIER FACTEUR DE L'ANALYSE FACTORIELLE : OPPOSITION DES CLASSES D'AGES EXTREMES.

(Facteurs scores standardisés et découpés en 7 classes de 0.8 écart-type centrées sur la moyenne).

G 11



3° Exemples traités avec les élèves.

a) Opposition démographique des régions françaises (graphes N° 8-10)

- Espace réduit des variables (graphes n°8-9). L'analyse résume les 12 variables de la matrice en trois facteurs dont les deux premiers sont facilement interprétables :

Le premier facteur est corrélé négativement avec les classes d'âge 65 ans et + (H et F) ainsi qu'avec l'indice de vieillesse ; il est corrélé positivement avec TRE : taux net de reproduction 1967/69 et TAN : taux d'accroissement naturel 1967-69 et les classes d'âge 0-25 ans (H et F). On peut l'interpréter comme "l'opposition des extrêmes" ou comme un facteur de dynamisme démographique.

Le deuxième facteur est corrélé négativement avec HOO, FOO ainsi que H 65, F 65 ; positivement avec F 25 et H 25. On peut l'interpréter comme l'opposition entre actifs et inactifs. L'interprétation du 3e pose quelques problèmes. Les trois premiers facteurs de cette analyse sauvegardent 87,8 % de l'information.

- Espace réduit des observations.

Le premier "facteur score" est cartographié sur la carte n° 11 et confirme le bien fondé de l'interprétation l'opposition démographique Nord Sud. On peut également utiliser le premier plan factoriel (graphe 10) pour élaborer une typologie. Les groupes de régions déterminés "à la main" sur les plans factoriels restent assez arbitraires et assez grossiers. Nous avons traité la matrice des poids locaux avec une méthode automatique, l'analyse hiérarchique pour tenir compte de toutes les coordonnées sur les 3 premiers facteurs. Le résultat de cette analyse figure sur les graphes 13-14.

b) Opposition régionale de l'industrie française (graphe n° 17)

Le résultat de cette étude que nous ne détaillerons pas est intéressant en soi. Les facteurs 1 et 2 portent respectivement 65 et 15 % de la variance de l'information soit pour le 1er plan de projection 80 % de l'information. Ceci permet de tracer une typologie sans les projections des plans un et trois, deux et trois.

Les subdivisions à l'intérieur des grands groupes ont été obtenues par analyse hiérarchique de la matrice des saturations en trois 1ers facteurs. Les oppositions essentielles sont très claires : sur le graphique factoriel apparaissent séparées, les régions Nord, Lorraine, Rhône-Alpes, Région Parisienne, Corse d'une part, opposées à toutes les autres régions d'autre part.

DEMOGRAPHIE : ANALYSE HIERARCHIQUE DES FACTEURS SCORES

ARBRE DE CLASSIFICATION

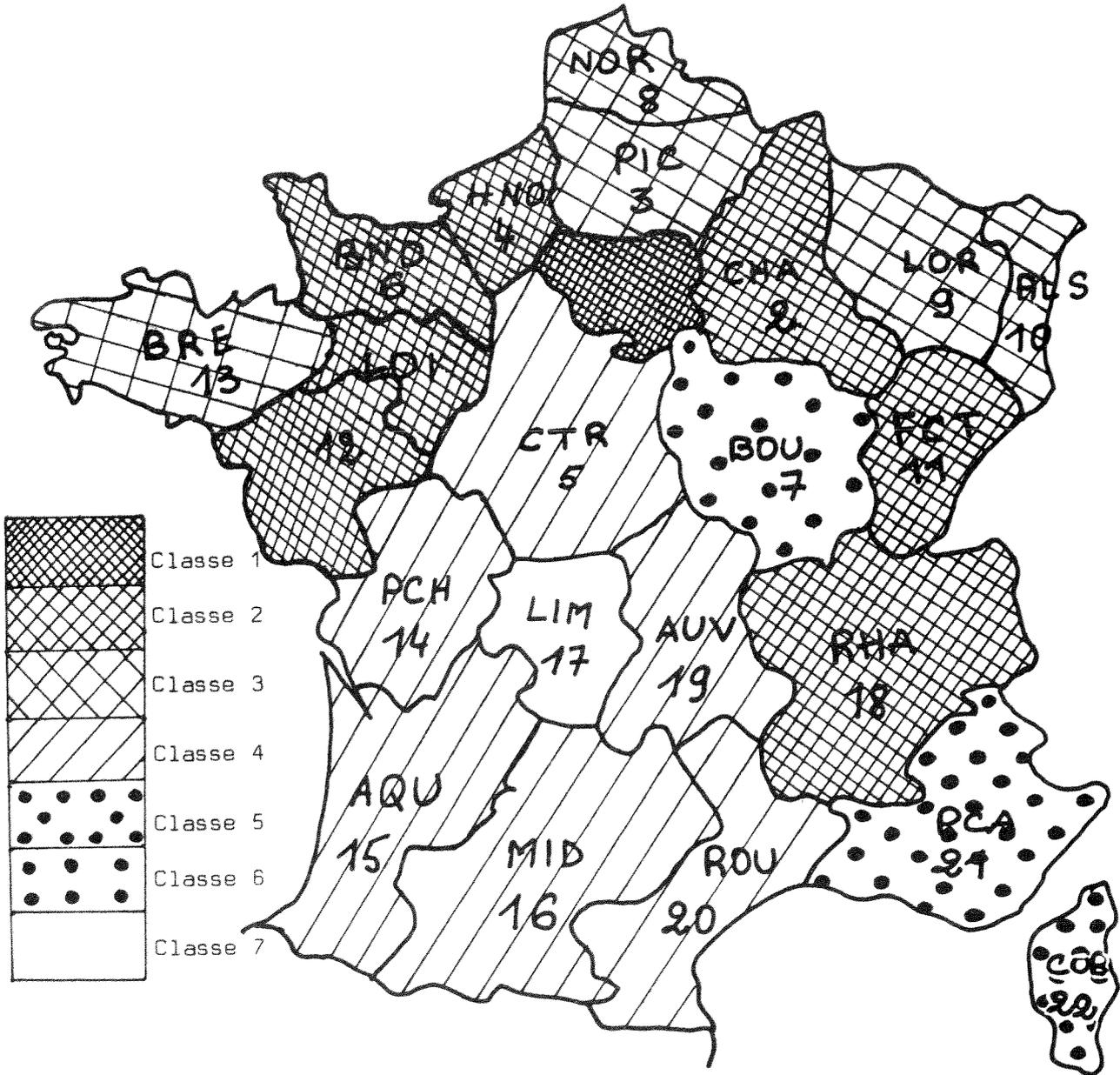
NIVEAUX DE SIMILITUDE DE 1.0 A 0.0

G13

Classe	1.0	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1	0
	+	+	+	+	+	+	+	+	+	+	+
1 PAR-----*											*
CHA-----*								I ⁴⁵			*
I ²⁴	I	I	I	I	I	I	I	I			
ENO--* I ²⁵	I	I	I	I	I	I	I	I			
I	I	I	I	I	I	I	I	I			
HNO-----* I ³¹	I	I	I	I	I	I	I	I			
I ²⁶	I	I	I	I	I	I	I	I			
2 FCT-----* I ³⁸	I	I	I	I	I	I	I	I			
I	I	I	I	I ⁴⁰	I	I	I	I			
LOI-----* I	I	I	I	I	I	I	I	I			
I	I	I	I	I	I	I	I	I			
RHA-----* I	I	I	I	I	I	I	I	I			
I ³⁴	I	I	I	I	I	I	I	I			
FRA-----* I	I	I	I	I	I	I	I	I			
I	I	I	I	I	I	I	I	I			
PIC-----* I	I	I	I	I	I ⁴²	I	I	I			
I ²⁹	I	I	I	I	I	I	I	I			
ERE-----* I ³⁶	I	I	I	I	I	I	I	I			
I	I	I	I	I ³⁹	I	I	I	I			
3 NOR-----* I	I	I	I	I	I	I	I	I			
I	I	I	I	I	I	I ⁴³	I	I			
LOR-----* I	I	I	I	I	I	I	I	I ⁴⁴			
I ³³	I	I	I	I	I	I	I	I			
ALS-----* I	I	I	I	I	I	I	I	I			
I	I	I	I	I	I	I	I	I			
CTR-----* I	I	I	I	I	I	I	I	I			
I ³⁰	I	I	I	I	I	I	I	I			
PCH-----* I ³⁵	I	I	I	I	I	I	I	I			
I	I	I	I ³⁷	I	I	I	I	I			
4 MID-----* I	I	I	I	I	I	I	I	I			
I	I	I	I	I ⁴¹	I	I	I	I			
AQU-----* I	I	I	I	I	I	I	I	I			
I ²⁸	I	I	I	I	I	I	I	I			
AUV-----* I	I	I	I	I	I	I	I	I			
I ²⁷	I	I	I	I	I	I	I	I			
ROU-----* I	I	I	I	I	I	I	I	I			
I	I	I	I	I	I	I	I	I			
PCA-----* I	I	I	I	I	I	I	I	I			
I ³²	I	I	I	I	I	I	I	I			
5 COR-----* I	I	I	I	I	I	I	I	I			
I	I	I	I	I	I	I	I	I			
6 EOU-----* I	I	I	I	I	I	I	I	I			
I	I	I	I	I	I	I	I	I			
7 LIM-----* I	I	I	I	I	I	I	I	I			

(L'arbre a été coupé au niveau de similitude 0.650). (3 facteurs)

G14



Il est clair que pour faire une classification des régions françaises il faut éliminer au moins Paris de la matrice d'origine.

Nous avons fait ailleurs cette 2e analyse en éliminant Paris et la Corse de la matrice et le graphique factoriel nous a donné une typologie beaucoup plus étalée sur l'axe 1.

c) Typologie des climats terrestres.

La matrice analysée comprenait 12 moyennes mensuelles de températures et une moyenne annuelle qui sont les variables observées par 16 stations.

Espace réduit des variables : (graphes 19)

Axe 1 : complémentarité de variables fortement corrélées.

Axe 2 : oppositions des saisons chaudes et froides (rôle du soleil)

Axe 3 : 1,36 % de l'information ne porte plus assez de variance pour être interprétable.

Espace réduit des observations (graphes 20)

Les plans 1 et 2 portant presque toute l'information on obtient une très jolie typologie sans avoir à traiter les autres coordonnées.

C- Inconvénients de la méthode.

L'analyse en composantes principales opère un ajustement linéaire sur le nuage de points. Or en géographie les relations entre variables ne sont pas toujours des fonctions linéaires. Un faible coefficient de corrélation linéaire peut cacher une forte liaison non linéaire. Il faut dans ce cas essayer de se ramener à la linéarité en traitant par exemple les logarithmes des valeurs. On peut également recourir au coefficient de corrélation des rangs de Spearman, ou au coefficient de Pearson.

III CLASSIFICATION REGIONALE ET ANALYSE ASCENDANTE HIERARCHIQUE.

(Graphe 3 Prog. DIST, EUCLD)

A- Mesure de la ressemblance entre les régions.

[10], [11], [12], [13], [17], [18]

1° La distance euclidienne.

Comment mesurer la ressemblance et la différence entre les régions de manière à pouvoir les séparer, les réunir, les classer ?

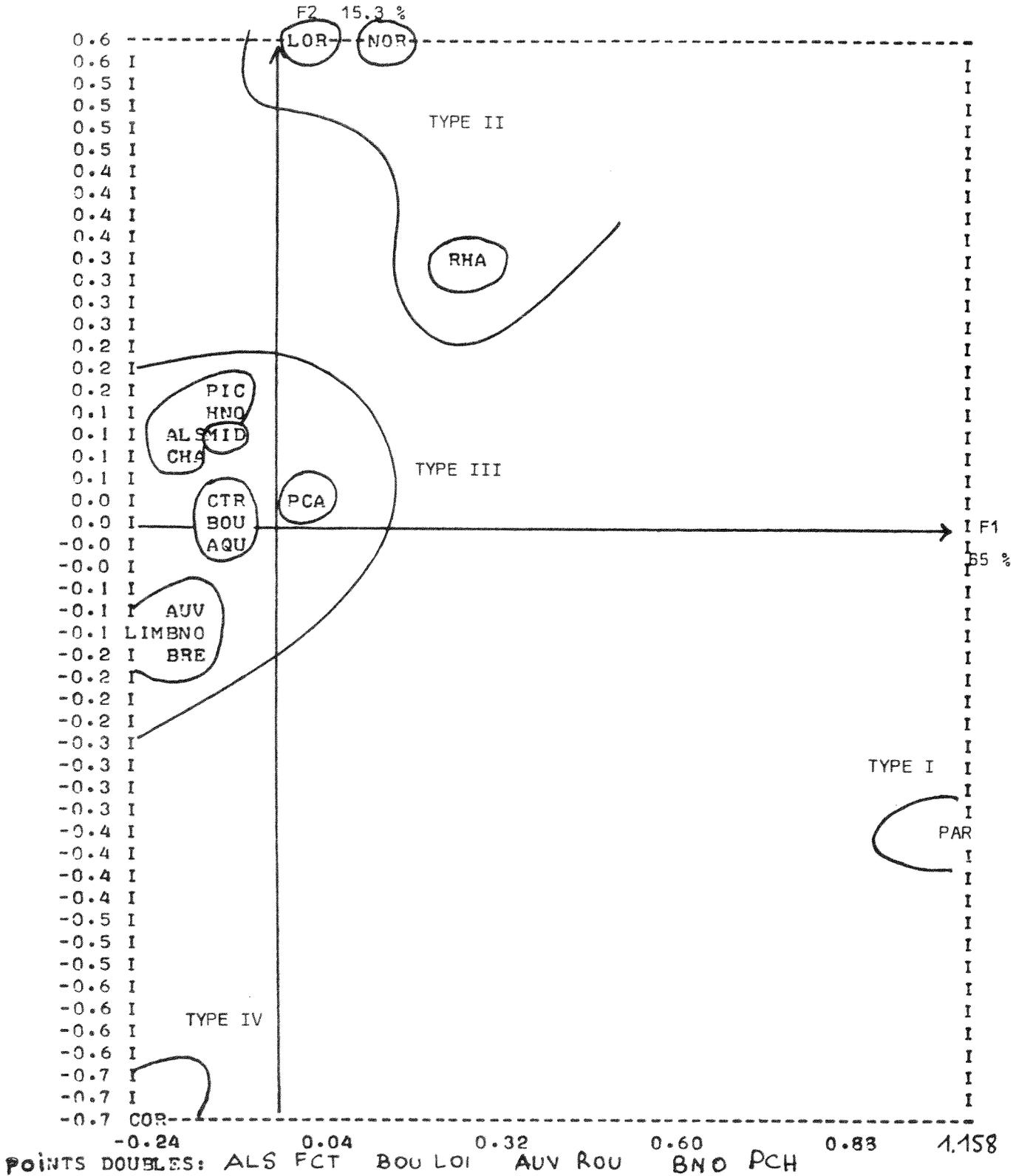
LES OPPOSITIONS REGIONALES DE L'INDUSTRIE FRANCAISE : FICHER 13 VARIABLES

(Variables non pondérées ; standardisées).

ANALYSE FACTORIELLE : PROJECTION DES POINTS OBSERVATIONS DANS L'ESPACE DES FACTEURS.

G 17

NB : Les classes issues de l'analyse hiérarchique sont entourées (similitude de 0.896).



Les codes des 22 régions se trouvent graphe 5.

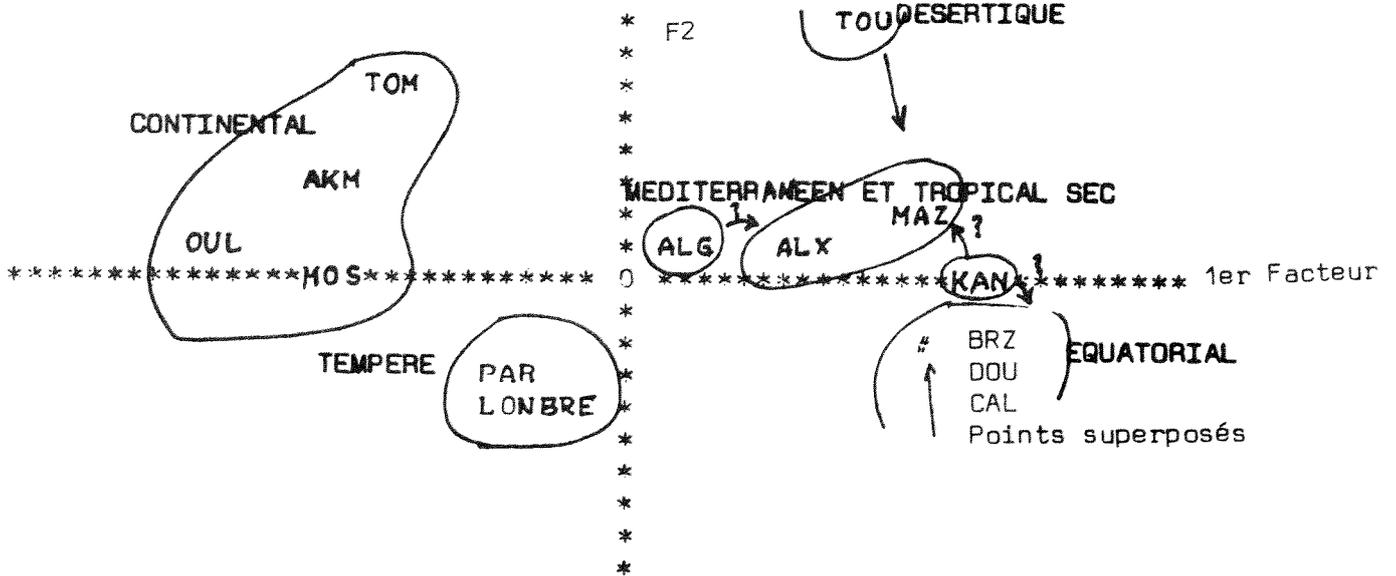
TYPOLOGIE DES CLIMATS TERRESTRES

ANALYSE FACTORIELLE : AXES 1/2 ET 1/3

G20

NU. DES 2 AXES DU PLAN A REPRESENTER ? 12

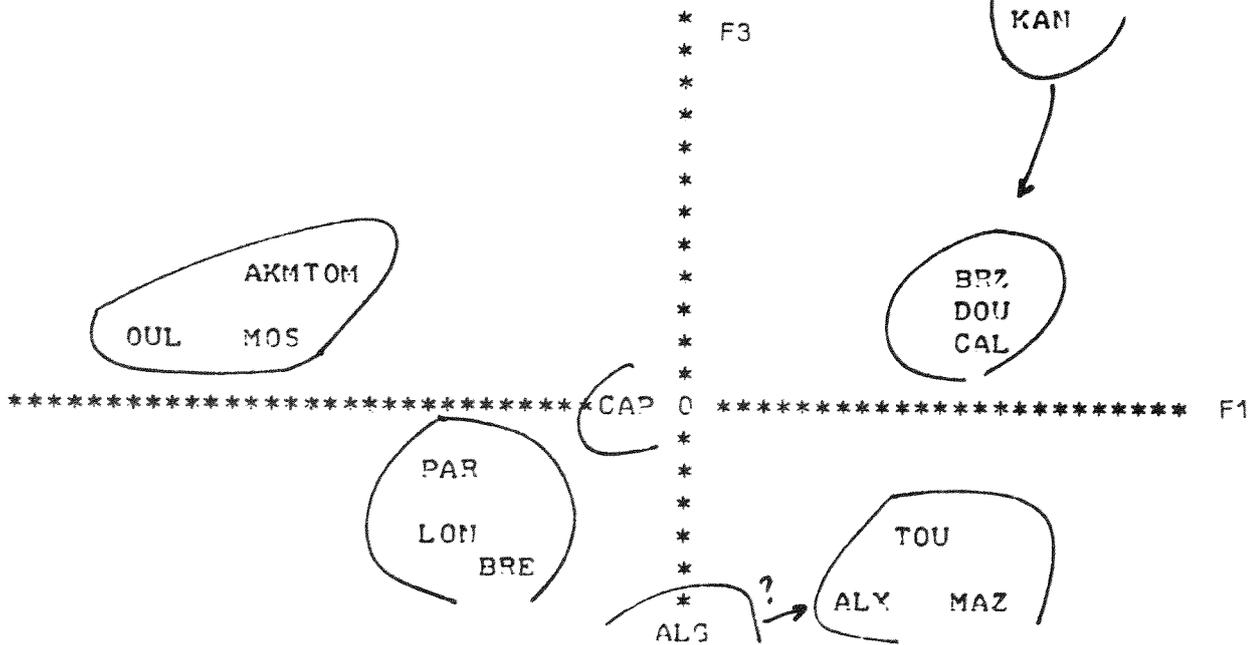
MAX MIN SUR X 0.363 -0.483 MAX MIN SUR Y 0.555 -0.706



MEDITERRANEEN (HEMISPHERE S)

NU. DES 2 AXES DU PLAN A REPRESENTER ? 13

MAX MIN SUR X 0.363 -0.483 MAX MIN SUR Y 0.671 -0.387



En classe de 1ère on peut difficilement expliquer la métrique KH I 2 ; il est par contre très facile d'exposer la distance euclidienne (notée d) et de l'utiliser dans un espace à plusieurs dimensions qui sera ici l'ensemble des régions (ou des stations.)

Une distance sur un espace E possède trois propriétés :

- 1- de séparation $\forall x_i, x_k \in E \quad d(x_i, x_h) = 0$ si et seulement si $x_i = x_h$
- 2- de symétrie $\forall x_i, x_k \in E \quad d(x_i, x_h) = d(x_h, x_i)$
- 3- d'inégalité triangulaire $\forall x_i, x_h, x_e \in E \quad d(x_i, x_h) \leq d(x_i, x_e) + d(x_e, x_h)$

La distance euclidienne de deux points x_i, x_h de E se définit comme la racine carrée de la somme des carrés des différences des coordonnées respectives des deux points et s'écrit donc

$$d(x_i, x_h) = \left[\sum_{j=1}^{NV} (x_{ij} - x_{hj})^2 \right]^{1/2}$$

où $(x_{i1}, \dots, x_{ij}, \dots, x_{iNV})$ sont les coordonnées de x_i , $(x_{h1}, \dots, x_{hj}, \dots, x_{hNV})$ celles de x_h .

La standardisation des variables sera faite systématiquement car il est évident que cette distance est sensible aux unités choisies.

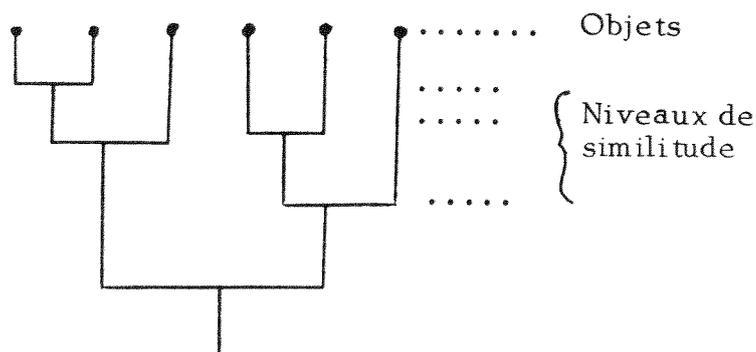
2° Métrique et ultramétrie.

L'analyse ascendante hiérarchique veut définir en partant des régions une hiérarchie de partitions qui se traduit par un arbre.

Pour définir une partition sur les régions on doit passer par une ultramétrie où la condition de l'inégalité triangulaire est remplacé par une condition plus forte l'inégalité ultramétrique. :

$$\forall x_i, x_h, x_e \in E \quad d(x_i, x_k) \leq \text{SUP} [(d x_i, x_e), (d x_e, x_h)]$$

Le regroupement des régions se présente comme un arbre dont les points de départ sont les individus :



(Ex : graphes 15 - 16 - 21).

La distance euclidienne ne conduit pas à une ultramétrie on introduit donc un "indice de similitude" S ("réduit" pour demeurer entre zéro et 1, 0 = différence maximum, 1 = égalité)

$$S(x_i, x_h) = \frac{D - d(x_i, x_h)}{D} \quad \text{où } D = \text{SUP}_{i,h} [d(x_i, x_h)] \text{ et } 1 - S(x_i, x_h) = \frac{d_{ij}}{D}$$

On peut en effet à toute distance ultramétrique faire correspondre une hiérarchie indicée unique dont $S = 1 - d$ est l'indice associé. Sur cette hiérarchie indicée nous pouvons lire la partition qui correspond à un niveau donné de similitude" (J.C. Chevailler biblio. n° 13). Ce passage par un indice de similitude est nécessaire car une distance permet d'établir une ordonnance et non pas une hiérarchie de partitions.

Comment procède le programme ? ;

- 1 - Recherche des 2 régions les plus proches i et j
- 2 - Fusion des deux régions i et j remplacées par leur point moyen qui sera considéré comme un individu nouveau lors de l'étape suivante.
- 3 - Pour l'individu k suivant, la distance de k à i et de k à j est remplacée par la distance au point moyen : d' . Cette substitution remplace l'indice de similitude par un indice qui correspond lui à une ultramétrie puisque les inégalités ultramétriques sont vérifiées sur le trio i, j, k pour d' .
- 4 - A la fin du processus, on a organisé la population sous la forme d'un arbre. L'intérêt de la méthode est de permettre un choix très souple dans l'arbre soit du nombre de classes que l'on se fixe, soit du niveau de similitude où l'on souhaite couper l'arbre.

B- Exemples de travaux avec les élèves.

Les élèves ont en main les photocopies des arbres hiérarchiques et cartographient les résultats selon des classes cartographiables du type de celles du graphe N° 4.

1° Les oppositions démographiques.

Un premier exemple traite les coordonnées de la matrice des poids locaux pour optimiser l'utilisation des plans factoriels (graphes n° 11 - 13 - 14 - 15 - 16). Le fichier complet traité par analyse hiérarchique figure sur les graphes. La bonne concordance des cartes 14 et 16 prouve la qualité de l'outil factoriel.

ANALYSE HIERARCHIQUE DE LA MATRICE DES 12 VARIABLES
DEMOGRAPHIE.

G 15

ARBRE DE CLASSIFICATION

NIVEAUX DE SIMILITUDE DE 1.0 A 0.0

	1.0	0.9	0.8	0.7	0.65 ⁰	0.6	0.5	0.4	0.3	0.2	0.1	0
	+	+	+	+		+	+	+	+	+	+	+
PAR-----												*
CHA-----									I 45			*
FCT-----		I 24	I I I	I I	I	I	I	I	I			
LOI-----		I 25	I I I	I I	I	I	I	I	I			
HNO-----		I 29	I I	I I	I	I	I	I	I			
ENO-----		I 30	I	I I	I	I	I	I	I			
PI C-----		I 26	I	I I	I	I	I	I	I			
NOR-----		*	I	I I	I	I	I 43	I 44				
LOR-----			I 34	I	I	I	I	I				
ALS-----			*	I	I	I	I	I				
ERE-----				I 37	I	I	I	I				
RHA-----				*	I	I	I	I				
FRA-----			I 31		I	I	I	I				
BOU-----			*		I	I	I	I				
CTR-----							*					
PCH-----		I 25		I	I	I I						
AQU-----		*		I 38	I	I I						
ROU-----		I 27	I I I			I I						
AUV-----		*	I 32	I I		I I	41 I I 42					
MID-----			I I I			I I						
PCA-----			* 33	I I 35		I I						
LIM-----				I		I I						
COR-----				*		I						*

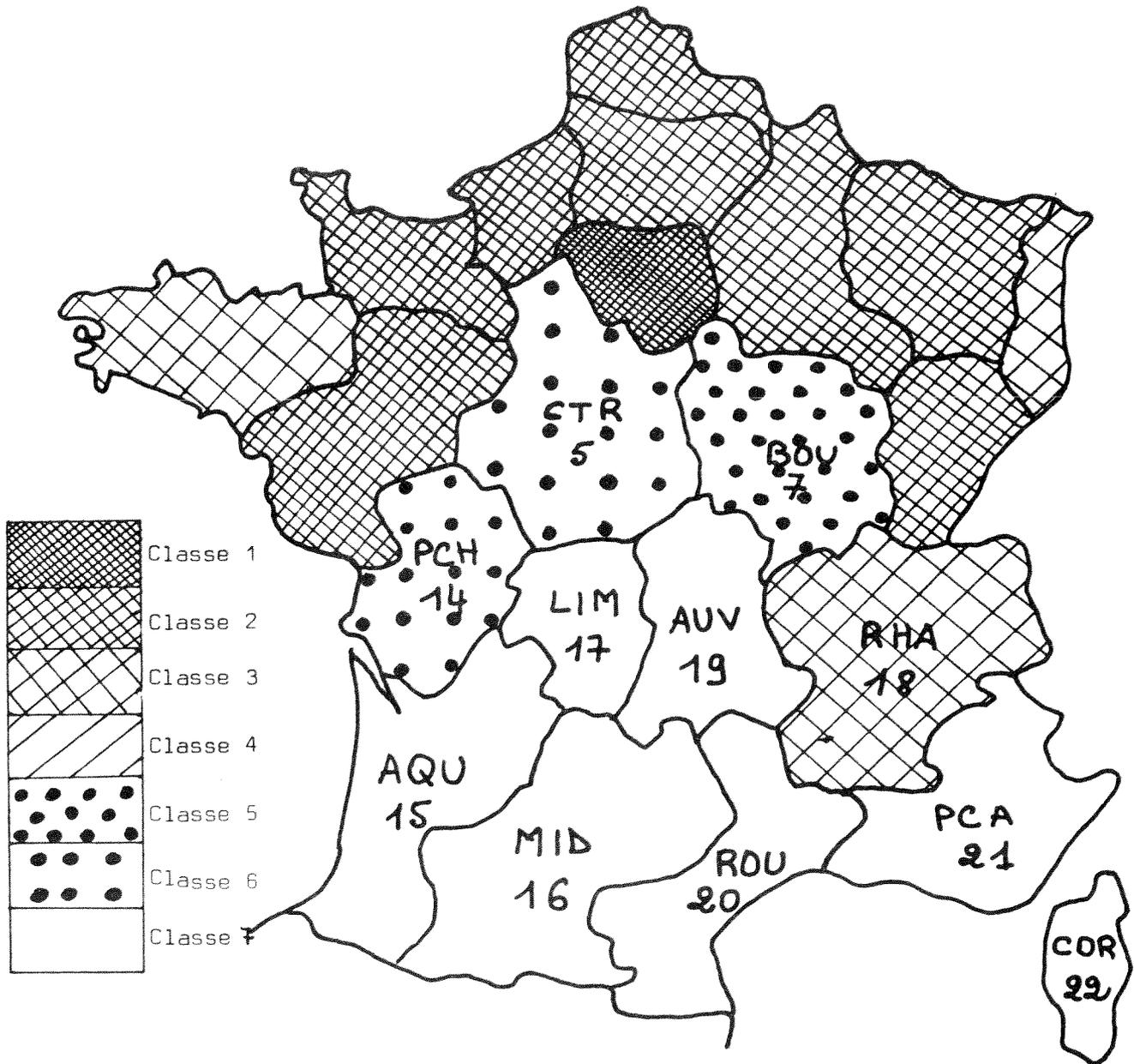
TERMINE

CLASSIFICATION DES REGIONS DEMOGRAPHIQUES FRANCAISES : MATRICE 12 VARIABLES

ANALYSE HIERARCHIQUE ASCENDANTE DE LA MATRICE DES VARIABLES STANDARDISEES

(L'arbre a été coupé au niveau de similitude 0.650).

G16



2° La matrice des températures.
(Graphe 21)

IV UNE METHODE MANUELLE : LA MATRICE GRAPHIQUE DE BERTIN.
[14], [15], [16]

A- Passage de la matrice des données à la matrice graphique.

1° On passe du tableau des nombres à la matrice graphique en faisant correspondre les valeurs numériques du tableau à des valeurs visuelles (établies avec le graphe 4 par ex.). Ce travail de correspondance a en fait déjà été élaboré par les élèves au niveau de la cartographie univariée ; il suffit donc qu'une équipe reprenne l'ensemble des cartographies univariées dans une synthèse graphique.

La matrice Bertin a autant de colonnes que de variables : au lieu de retrouver en $M(i, j)$ la valeur numérique, on retrouve un palier graphique.

2° Cette image complexe est ensuite simplifiée par la procédure suivante :

a) Il faut rapprocher entre elles les lignes qui se ressemblent visuellement, éloigner les lignes qui se différencient.

b) Faire la même opération sur les colonnes.

De quoi s'agit-il en fait ? Il s'agit simplement de rapprocher les variables qui ont même profil visuel et les régions qui ont même profil. Il en découle une typologie.

Pour la permutation des lignes et des colonnes ciseau, colle, papier mince suffisent. L'image simplifiée est organisée autour de la diagonale du tableau c'est pourquoi Bertin parle de "diagonalisation de matrice ordonnable". Elle est ordonnable dans la mesure où cela a un sens de permuter colonnes entre elles et lignes entre elles. Cela n'aurait pas de sens sur une matrice où l'on aurait en ligne une chronologie (par exemple.)

3° Exemples traités avec les élèves.

Les matrices graphiques élaborées par les élèves sur le dossier industriel étant en couleurs ne sont pas publiables, l'exemple exposé ici (graphes 22-23) a été établi sur la matrice des climats avec les élèves de seconde.

la matrice graphique a été établie de la manière suivante :

a) Recherche par les élèves des valeurs de coupure qui permettent de caractériser les différents climats (ce travail est mené sur des graphiques calculés automatiquement avec le programme GRAF).

TYPOLOGIE DES CLIMATS

G21

ARBRE DE CLASSIFICATION

NIVEAUX DE SIMILITUDE DE 1.0 A 0.0

TYPE	1.0 +	0.9 +	0.8 +	0.7 +	0.6 +	0.5 +	0.4 +	0.3 +	0.2 +	0.1 +	0 +
1 AXI											*
12 MOS		I ²²	I I		I		I				
13 OUL			I ²⁴ I	I ²⁶	I		I ³⁰				
15 TOM			I		I		I				
3 ALG					*		I ³¹				
5 BRE				I ²⁷	I		I				
10 LON		I ¹⁹		I			I				
14 PAR		I ¹⁸		I ²⁹			I				
9 CAP				I			I				
2 ALX											*
11 MAZ		I ³¹	I	I							
4 BRZ			I ²⁵	I							
6 CAL		I ²⁰	I	I ²⁸							
7 DOU		I ¹⁷	I ²³	I							
8 KAN			I	I							
16 TOU			*	I							

- b) Elaboration d'une matrice (remplie de zéro-un) à partir des valeurs de coupure : et est-ce que la i ème station a une moyenne du mois de décembre inférieure à -15° ? Si oui on met un dans la case sinon zéro. Ce travail est fait pour les 9 variables.
- c) La matrice peut être "diagonalisée" graphiquement sur console conversationnelle (Graphe 3 programmes DIAG).

Typologie sur la matrice "diagonalisée".

L'image simplifiée présente un maximum de un proche de la diagonale et de zéros dans deux coins opposés. Il faut ensuite choisir les discontinuités en ligne qui permettront de classer les éléments (voir graphe 23).

C- Inconvénients de la méthode (et avantages).

Elle est relativement pesante à réaliser avec les élèves. Mais c'est une méthode qui donne de bons résultats et c'est la seule méthode qui soit applicable sans l'équipement d'une machine. Elle est beaucoup plus lente qu'une analyse factorielle, 3 à 4 h, au lieu de une minute et demie mais les résultats correspondent à peu près à ce que donne le premier plan factoriel. Bertin a, lui, mis au point un système de dominos (cubes traversés par deux tunnels perpendiculaires) avec lequel les manipulations sont beaucoup plus rapides que si l'on utilise le matériel simplifié (papier, colle, ciseaux).

Les dominos, rassemblés en une grande matrice, sont manipulés par lignes ou colonnes entières à l'aide de tringles métalliques passées dans les tunnels.

Elle ne permet pas de traiter plus d'une dizaine de variables et l'information perdue n'est pas calculée.

V AUTRES APPLICATIONS : LES SIMULATIONS, LES BANQUES DE DONNEES, LES TRAITEMENTS D'ENQUETE.

A- Les simulations.

1° La mise en oeuvre de modèles par simulation est une démarche très formatrice : passage d'une perception de la réalité à une représentation de cette réalité par des symboles.

MATRICE LOGIQUE BERTIN AVANT DIAGONALISATION GRAPHIQUE

G 22

		Moyenne des températures annuelles > 20	Moyenne des températures mensuelles de Déc. > 7	MTM (12) < 0	MTM (12) < -15	MTM (7) ou (12) > 20 (Hém. 5)	Moyenne annuelle de précipitations > 1000	800 < MAP < 1000	250 < MAP < 800	MAP < 250
		1	2	3	4	5	6	7	8	9
AKMOLINSK	1	1			4	5				9
ALEXANDRIE	2		2			5				9
ALGER	3					5		7		
BRAZZAVILLE	4					5	6			
BREST	5								8	
CALABAR	6					5	6			
DOUALA	7					5	6			
KANO	8					5			8	
LE CAP	9					5			8	
LONDRES	10								8	
MAZATLAN	11					5			8	
MOSCOU	12				3				8	
OULAN BATOR	13				4					9
PARIS	14								8	
TOMSK	15					5			8	
TOUGGOURT	16					5				9

MATRICE LOGIQUE BERTIN APRES DIAGONALISATION GRAPHIQUE

VARIABLES

G23

T Y P O L O G I E

		1	2	3	4	5	6	7	8	9		
ALEXANDRIE	2	■	■	■	■						←	
BRAZZAVILLE	4	■	■	■		■					Lignes à rapprocher	
CALABAR	6	■	■	■		■						
DOUALA	7	■	■	■		■						
KANO	8	■	■	■			■					
MAZATLAN	11	■	■	■			■					
TOUGGOURT	16	■	■	■	■							←
LE CAP	9		■	■			■					
ALGER	3		■	■						■		
AKMOLINSK	1			■	■				■			
TOMSK	15			■			■	■				
MOSCOU	12						■	■				
OULAN BATOR	13				■				■			
LONDRES	10						■					
PARIS	14						■					
BREST	5						■					

- 2° Un modèle même très simple nécessite une somme considérable de calculs. La rapidité de calcul de l'ordinateur permet de faire jouer des interactions complexes. Cette démarche introduit une méthode éducative tout-à-fait neuve : on remplace en effet les définitions statiques et isolées par la dynamique de systèmes à interactions complexes bien plus proches de la réalité.
- 3° La démarche de modélisation est très formatrice dans la mesure où elle contraint à formaliser à expliciter à fond les hypothèses de travail ce qui n'était que très peu fait auparavant. Les élèves peuvent critiquer ce modèle sans avoir l'impression de critiquer le professeur.
- 4° Exemples de programmes déjà disponibles et testés avec les élèves.
- a) Simulation simple de croissance démographique (2 paramètres continus, affichage de la pyramide des âges par bond de 5 ans).
 - b) Simulation sur le même thème mais beaucoup plus complexe (75 paramètres tous continus). Grande variété des sorties graphiques et numériques.
 - c) La croissance d'une économie et ses interactions à travers les agrégats de la comptabilité nationale. L'élève peut jouer sur 5 paramètres et le programme lui indique les conséquences de ses choix.
 - d) Conséquences de taux de dévaluation choisis par l'élève en fonction de diverses balances commerciales et de diverses élasticités à l'importation et à l'exportation.
 - e) Gestion d'une exploitation agricole avec tirage aléatoire d'événements tels que des hausses et baisses de prix ou des mauvaises récoltes.
 - f) Dynamique d'un système écologique : la pollution d'une masse d'eau (11 paramètres dont 9 discrets et 2 continus)
 - g) Satellisation autour d'une planète (3 paramètres).
 - h) En préparation :
 - simulation des trajets des ondes sismiques à travers le globe
 - érosion continentale avec ou sans compensation isostatique

B- Elaboration et exploitation de banques de données.

- 1° La saisie des matrices de données concernant les 22 régions de programme se faisant sur plusieurs années, au bout de quelque temps, les élèves auront à leur disposition une véritable banque de données régionales directement accessibles.
- 2° En séismologie : nous avons obtenu les listings de traitement sur ordinateur des principaux séismes mondiaux élaborés par le Centre International de Séismologie de Strasbourg. Les élèves de 2des ont "rentré" l'année dernière 800 séismes de l'année 1972 avec dates, lieux, coordonnées géographiques, magnitudes, profondeurs, dégâts ou morts signalés. Des programmes conversationnels d'interrogation de ces fichiers permettent de faire un tri parmi ces séismes suivant un certain nombre de critères. Ceci peut servir de base à une cartographie.
Nous prévoyons d'utiliser sur ces fichiers des programmes de tris croisés écrits par un collègue du Creusot, Mr Massard.
- 3° En climatologie et météorologie les élèves auront progressivement un nombre croissant de stations climatiques à leur disposition. Les programmes d'interrogation des fichiers sont encore à faire.
En météorologie nous prévoyons de stocker les observations quotidiennes relevées toutes les trois heures pour 14 paramètres, pour quelques stations françaises.

C- Enquêtes.

Il serait également possible de réaliser avec les élèves des enquêtes de type sociologique. Ceci n'a pas encore été fait. Un travail de ce genre permettrait de poser clairement les problèmes des sondages dont les élèves voient maintenant chaque jour les résultats dans les journaux.

Définition de l'échantillon ? Définition des termes du questionnaires ?

Mise en page, codage pour exploitation, prise de conscience de l'ambiguïté des questions et des réponses et des difficultés d'interprétation.

D- Pour mémoire je rappellerais simplement que les enseignants de second degré sont également chargés de l'histoire. Nous avons adopté la même démarche : les programmes de saisie de données concernant la révolution industrielle 1800 1913 sont prêts, ont déjà "tourné" ainsi que les programmes de traitement (calculs d'indices sur bases de temps au choix, ajustement de la série à un polynôme, calcul direct de salaires réels, calculs de taux de croissance par durée de périodes au choix etc...)

QU'EN PENSENT LES ELEVES ?

Nous avons travaillé trois ans selon ce type de méthodes. D'abord deux ans sans machine dans l'établissement en passant sur le compte du Lycée Fustel et de l'I.R.E.M., sur l'UNIVAC 1110 du Centre de Calcul Universitaire de Strasbourg en Fortran et Algol. Le Lycée est équipé d'ordinateur avec consoles depuis novembre 1974.

Nous n'avons pas cherché à évaluer systématiquement les acquis ou l'intérêt des élèves et les remarques qui suivent sont celles des enseignants et non des élèves directement.

- Un inconvénient d'abord qui n'est pas lié à l'utilisation des statistiques mais plutôt à la méthode du dossier. La cohérence du travail des équipes est assurée par le professeur. La tâche des élèves est parcellisée et cette parcellisation empêche certains élèves de comprendre comment ils s'insèrent dans le travail d'ensemble. Il est pourtant difficile de leur demander de construire toute la progression pour les rendre actifs face à l'organisation du travail.
- Un avantage ; l'intérêt soutenu de nos élèves dont certains travaillent depuis plus d'un an avec cette méthode.
On peut se hasarder à quelques proposition d'explications :
 - Plaisir de la manipulation technique, de la démystification de l'instrument, de la puissance de l'homme sur la machine.
 - Absence d'élément moralisateur dans l'appréciation des résultats.
La console ne juge pas l'élève.

- Calme de la demie classe de 16 à 19 élèves, dans un lycée situé au centre ville et plutôt exigu.
- Indépendance de chacun dans sa "progression" aux consoles.
- Elaboration d'un travail de véritable recherche et "fabrication de toutes pièces d'un dossier.
- Variété des jeux, simulations, travaux demandés.
- Applications concrètes de notions apprises en mathématiques.
- Prise de conscience et compréhension de méthode ou d'information réputées difficilement accessibles.

Les programmes L.S.E dont certains ont été élaborés en collaboration avec P. KOULLEN de l'I.R.E.M. de Strasbourg sont à la disposition des collègues intéressés : Thierry HATT, Lycée Fustel de Coulanges, Place du Château 67000 STRASBOURG.

BIBLIOGRAPHIE

I. SOURCES.

A) Régions : base fondamentale :

- [1] Statistiques et Indicateurs des Régions Françaises, coll. INSES R N° 14 (annuel depuis 1968, parution le 1er trimestre de l'année scolaire) 1974 - 30 F
- [2] R n° 13 "Les comptes régionaux de l'agriculture" 1973, 18 F.
- [3] R n° 16 "L'emploi dans les grands établissements, 1962, 1967" 1970, 10 F.
- [4] "Les établissements de 1000 salariés et plus en France en 1968" INSEE 1969.

B) Climats :

- [5] M. GARNIER : Direction de la Météorologie Nationale "Memorial de la Météorologie Nationale" Climatologie de la France , Paris 1967.
- [6] Ch. P. PEGUY : "Précis de Climatologie" 2° éd. revue remaniée 1970.

II. CRITIQUE DES SOURCES.

A) Economiques :

- [7] O. MORGENSTERN : "Précision et incertitudes des données économiques" Dunod 1972.
- [8] M. LEVY : "L'Information statistique" Seuil 1975.

B) Climats :

- [9] R. ARLERY : "Climatologie, méthodes et pratiques" G. Villars 1973.

III. METHODES ET PROGRAMMES D'ANALYSE DES DONNEES.

A) Analyse numérique :

- [10] J.P. BENZECRI "L'analyse des données" T.1. "Taxinomie", T.2. "Analyse des correspondances" Dunod 1973, très difficile à lire pour non mathématicien.

- [11] L.LEBART et J.P. FENELON : "Statistique et Informatique appliquées"
2e Ed. Dunod 1973. Excellent, facile à lire, pédagogique.
- [12] J.L. GUIGOU : "Analyse des données et choix à critères multiples "
Dunod 1974. Facile à lire mais redondant, sans programme, algorithmes
pas assez explicites, vue d'ensemble assez large.
- [13] J.C. CHEVAILLER : "Classification en analyse économique spatiale"
1974 CUJAS n° 7 TEM Espace ; lecture facile, excellente vue d'ensemble.

B) Les matrices graphiques :

- [14] J. BERTIN : "Sémiologie graphique" Mouton 1967 ; excellent.
- [15] J. BERTIN : "Le traitement graphique de l'information" dans l'Informatique
janv. 1971 p.53 et sq.
- [16] J. BERTIN : "Généralisation du traitement graphique de l'information"
A.E.S.C. 1969 n° 1 p. 70.

IV. METHODES D'ANALYSE DES DONNEES EN GEOGRAPHIE.

- [17] GROUPE CHADULE : "Initiation aux méthodes statistiques en géographie"
Masson 1974 ; incomplet pour les analyses multicritères.
- [18] J.B. RACINE et H. REYMOND : "L'analyse quantitative en géographie"
PUF 1973 ; fondamental (les pièges de l'algorithme factoriel p 184).
- [19] P. HAGGETT : "L'analyse spatiale en géographie humaine" Colin , Coll. U
1973 ; une revue des parutions de géographie quantitative anglo-saxonnes.
- [20] L.J. KING : "Statistical analysis in géographie" Prentice Hall 1969 ;
classique.

V. QUELQUES EXEMPLES D'APPLICATIONS EN GEOGRAPHIE ET ECONOMIE.

A) Matrices graphiques :

- [21] D.NOIN : "Géographie démographique de la France" PUF 1973.
- [22] J.D. GRONOFF : "La zone herbagère des Ardennes ; analyse typologique
par matrice ordonnable." Etudes Rurales p. 169-214.
- [22b] Documentation photographique N° 6014 - 1974 : L'inventaire des moyens
de connaissances.

B) Analyse factorielle :

- [23] E. DALMASSO : "Essai d'application de quelques méthodes statistiques à la région milanaise" Bull. Association des Géographes français av. 1972.
- [24] Ph. AYDALOT : "La mobilité des activités économiques : l'exemple français" Coll. TEM Espace G. Villars 1971 n° 2. "Mobilité et croissance spatiale" idem n° 1.
- [25] Actes des colloques de géographie quantitative de Besançon depuis 1972.
- [26] Revue "L'espace géographique".