

J. P. BENZÉCRI

G. FLAMENBAUM

Y. L'HOSPITALIER

**Un algorithme d'agrégation autour de centres
variables en boules de rayons bornés**

Les cahiers de l'analyse des données, tome 4, n° 3 (1979),
p. 365-375

http://www.numdam.org/item?id=CAD_1979__4_3_365_0

© Les cahiers de l'analyse des données, Dunod, 1979, 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/>

UN ALGORITHME D'AGRÉGATION AUTOUR DE CENTRES VARIABLES EN BOULES DE RAYONS BORNÉS [ALG. AGR. RAY.]

par J. P. Benzécri ⁽¹⁾, G. Flamenbaum ⁽²⁾
et Y. L'Hospitalier ⁽³⁾

On présente ici sous forme de problème un algorithme réalisant la méthode de G. Flamenbaum : [BOULES OPTIMISEES], ce *Cahier* pp 357-364. Cet énoncé a été proposé par J.P. Benzécri pour l'examen de Logique et Linguistique (U.C.O. Juin 1979). La solution ci-jointe est due à Y.L'Hospitalier.

1. Enoncé du problème

L'objet du problème est de découvrir l'effet d'un programme ALGØL dont le listage est donnée à la fin de l'énoncé. Le programme comporte 3 procédures: dans les questions 1 et 2 on étudie respectivement les procédures DIS et GRAV. Les questions 4°, 5°, 6° concernent la procédure AFEC étudiée sur certaines données particulières. Enfin dans la question 7° on considère l'effet de la procédure AFEC sur les données les plus générales, dans le cadre du programme principal.

1.1 Quel est, en termes géométriques, l'effet de la procédure DIS ?

1.2 On suppose que le nombre entier KN et les contenus des tableaux NAFI et EFN satisfont aux conditions suivantes :

a) $\forall I \in]KI]$, $NAFI[I] \in]KN]$

b) $\forall N \in]KN]$, $EFN[N] \neq 0$

c) le tableau EFN[1:KN] peut être calculé à partir du tableau NAFI par le programme :

pour N:= 1 pas 1 jusqu'à KN faire EFN[N]:= 0 ;

pour I:= 1 pas 1 jusqu'à KI faire EFN[NAFI[I]]:=EFN[NAFI[I]] + 1 ;

Quel est, sous les conditions a,b,c l'effet de la procédure GRAV : on s'attachera comme dans 1.1, à donner une interprétation géométrique simple.

1.3 Pour étudier l'effet de la procédure AFEC on considère d'abord le cas où KN = 0 ; et on prend un exemple numérique avec KJ = 1, KI = 9 le contenu du tableau POSIJ étant :

I =	1	2	3	4	5	6	7	8	9
POSIJ[I, 1]=	5	6	7	7,5	9	3	2	1	0

(1) Professeur de statistique à l'Université P. & M. Curie ; Paris

(2) Commissariat à l'Energie Atomique, CEN/Cadarache, St Paul-les-Durance .

(3) Directeur de l'I. de M. Appliquées ; U. Catholique de l'Ouest ; Angers

Le contenu du tableau NAFI vérifie en début de procédure : $\forall i \in]KI]$, $NAFI[i] = 0$
On demande l'effet de la procédure AFEC(POSIJ, CENJ, NAFI, EFN, 9,1,0,9,KCHAN,2)
(i.e. on pose MAXN=9 ; RAY=2) on donnera les valeurs de KN, RAY et KCHAN après
exécution de la procédure, ainsi que les contenus des tableaux CENJ[1:KN,1:1],
NAFI[1:KI] et EFN[1:KN].

1.4 On reprend les mêmes données qu'en 3°, i.e. le même tableau POSIJ, et on demande
l'effet de la procédure AFEC avec de nouvelles valeurs de MAXN et de RAY :

AFEC(POSIJ, CENJ, NAFI, EFN, 9,1,0,4,KCHAN,1)

on donnera encore les valeurs de KN, RAY, KCHAN, et le contenu des tableaux CENJ, NAFI,
EFN.

1.5 On suppose réalisée la condition suivante :

$$\forall N \in \{1, \dots, KN\} : EFN[N] \neq 0 ;$$

quel est alors l'effet du bloc d'instructions par lequel se termine la procédure AFEC
NOVN:= 0 ;

.....

KN:= NOVN fin

Quel est l'effet de ce même bloc dans le cas où $EFN[1] = 0$ et les autres valeurs
de $EFN[N]$ pour $N=2, \dots, KN$ sont différentes de zéro ?

1.6 On applique encore une fois la procédure AFEC à des données quelconques, sur la
seule hypothèse que (au départ) $KN = 0$.

L'effet de la procédure est-il changé si on supprime le bloc final (qui va de
NOVN:= 0 à KN:= NOVN) ?

Quelle valeur faut-il donner à MAXN pour être assuré que ne jouera jamais l'ins-
truction :

si NOVA = 0 $\wedge$ KN = MAXN alors debut RAY:= 2 * RAY ; KCHAN:= KI ; aller à ETIQ fin ;

Décrire en termes géométriques l'effet de la procédure AFEC (dans le cas particulier
considéré ici où la valeur initiale de KN est zéro).

1.7 On considère maintenant l'effet du programme complet sur les données suivantes :

KI = 4 ; KJ = 1 ; MAXN = 4. ; RAY = 2 ; MAXIT = 3

I	1	2	3	4
POSIJ[I,1]	2	1	0	4

- Quel est l'effet du premier appel de la procédure AFEC : on donnera la valeur
de KN et KCHAN ; et le contenu des tableaux CENJ[1:KN, 1:KJ], NAFI[1:KI]
et EFN[1:KN] après exécution de AFEC.
- Quel est l'effet du premier appel de GRAV ? On donnera le contenu de
CENJ[1:KN, 1:KJ] après cet appel.
- Quel est l'effet du deuxième appel de AFEC ?
- Quel est l'effet de l'ensemble du programme ? On donnera les valeurs de
KN, RAY, KCHAN, IT ainsi que le contenu des tableaux CENJ, NAFI et EFN.

1.8 Expliquer l'effet de l'ensemble du programme.

LISTAGE 1

```

début   entier KI,KJ,KN,MAXN,MAXIT,IT ; reel RAY ;
début reel tableau POSIJ[1:KI,1:KJ],CENJ[1:MAXN,1:KJ];
        entier tableau NAFI[1:KI],EFN[1:MAXN] ; entier I ;
        LIRE (KI,KJ,MAXN,MAXIT,RAY,POSIJ) ;
commentaire : KI,KJ,MAXN,MAXIT,POSIJ sont des données inaltérables,
        RAY est une donnée que le programme peut modifier, le tableau
        POSIJ décrit un ensemble donné de KI points dans l'espace de
        dimension KJ, le tableau CENJ décrit un ensemble de points
        inconnus dans ce même espace ;
réel procédure DIS(TAB1,TAB2,M1,M2,KJ) ; corps de procédure ;
procédure GRAV(POSIJ, CENJ,NAFI,EFN,FI,KJ,N) ; corps de procédure ;
procédure AFEC(POSIJ,CENJ,NAFI,EFN,KI,KJ,KN,MAXN,KCHAN,RAY) ;
corps de procédure ;
    pour I:= 1 pas 1 jusqua KI faire NAFI[I]:= 0 ;
    IT:= 0 ; KN:= 0 ;
    ETIQAF: AFEC(POSIJ,CENJ,NAFI,EFN,KI,KJ,KN,MAXN,KCHAN,RAY) ;
    si KCHAN = 0 v IT=MAXIT alors allera FIN ;
    GRAV(POSIJ,CENJ,NAFI,EFN,KI,KJ,KN) ;
    IT:= IT+1 ;
    allera ETIQAF ;
    FIN ;
fin
fin

```

LISTAGE 2

```

reel procedure DIS(TAB1,TAB2,M1,M2,KJ) ;
    reel tableau TAB1,TAB2 ; entier M1,M2,KJ ;
    debut reel DIS2 ; entier J ;
        DIS2:= 0 ;
        pour J:= 1 pas 1 jusqua KJ faire
            DIS2:= DIS2+(TAB1[M1,J]-TAB2[M2,J])+2 ;
            DIS:= RAC2(DIS2)
        fin ;
procedure GRAV(POSIJ,CENJ,NAFI,EFN,KI,KJ,KN) ;
    reel tableau POSIJ,CENJ ; entier tableau NAFI,EFN ; entier KI,KJ,KN ;
    début entier I,J,N,NAUX ; reel RAUX ;
        pour N:= 1 pas 1 jusqua KN faire
            pour J:= 1 pas jusqua KJ faire
                CENJ[N,J] := 0 ;
                pour I:= 1 pas 1 jusqua KI faire
                    debut NAUX:= NAFI[I] ;
                        pour J:= 1 pas 1 jusqua KJ faire
                            CENJ[NAUX,J] := CENJ[NAUX,J]+POSIJ[I,J]
                        fin ;
                    pour N:= 1 pas 1 jusqua KN faire
                        debut RAUX:= EFN[N] ;
                            pour J:= 1 pas 1 jusqua KJ faire
                                CENJ[N,J] := CENJ[N,J]/RAUX
                            fin
                        fin ;
                    fin ;
    fin ;

procedure AFEC(POSIJ,CENJ,NAFI,EFN,KI,KJ,KN,MAXN,KCHAN,RAY) ;
    reel tableau POSIJ,CENJ ; entier tableau NAFI,EFN ; entier KI,KJ,KN,MAXN,KCHAN ,
    reel RAY ;
    debut reel DISA,DISB ; entier I,N,NOVA,NOVN ; entier tableau NN[1:MAXN] ;
        KCHAN:= 0 ;
        ETIQ: pour N:= 1 pas 1 jusqua KN faire EFN[N] := 0 ;
            pour I:= 1 pas 1 jusqua KI faire
                debut DISA:= RAY ; NOVA:= 0 ;
                    pour N:= 1 pas 1 jusqua KN faire
                        debut DISB:= DIS(POSIJ,CENJ,I,N,KJ) ;
                            si DISB < DISA alors debut DISA:= DISB ;
                                NOVA:= N
                            fin
                        fin ;
                    fin ;
                fin ;
            fin ;
    fin ;

```

```

si NOVA = 0  $\wedge$  KN = MAXN alors debut RAY:= 2 * RAY ;
                                KCHAN:= KI ;
                                allera ETIQ
                                fin ;
si NOVA = 0  $\wedge$  KN < MAXN alors debut KN:= KN + 1 ;
                                KCHAN:= KCHAN + 1 ;
                                NAFI[I] := KN ;
                                EFN[KN] := 1 ;
                                pour J:= 1 pas 1 jusqua KJ faire
                                CENJ[KN,J] := POSTJ[I,J]
                                fin ;
si NOVA  $\neq$  0 alors EFN[NOVA] := EFN[NOVA] + 1 ;
si NOVA  $\neq$  NAFI[I]  $\wedge$  NOVA  $\neq$  0 alors debut KCHAN:= KCHAN + 1 ;
                                NAFI[I] := NOVA
                                fin
fin ;

NOVN:= 0 ;
pour N:= 1 pas 1 jusqua KN faire NN[N] := 0 ;
pour N:= 1 pas 1 jusqua KN faire
debut si EFN[N]  $\neq$  0 alors debut NOVN:= NOVN+1 ;
                                NN[N] := NOVN ;
                                EFN[NOVN] := EFN[N] ;
                                pour J:= 1 pas 1 jusqua KJ faire
                                CENJ[NOVN,J] := CENJ[N,J]
                                fin
fin ;
pour I:= 1 pas 1 jusqua KI faire NAFI[I] := NN[NAFI[I]] ;
KN:= NOVN
fin ;

```

2 Solution du problème

Avant d'aborder explicitement la solution, nous donnerons un certain nombre d'indications générales (en particulier un glossaire) facilitant la lecture du programme.

POSIJ : tableau donnant la POSiTion des KI individus dans un espace de dimension KJ.

CENJ : tableau donnant les CENtres des MAXN classes

NAFI[I] : Numéro d'Affectation de l'individu I (i.e. l'individu I est affecté à la classe numérotée NAFI[I])

EFN[N] : Effectif de la N-ième classe

EFN[NAFI[I]] : Effectif de la classe dont le Numéro est le Numéro d'Affectation de I

KCHAN : nombre de CHANGements d'affectation des individus

RAY : RAYon des classes

MAXIT : nombre MAXimum d'ITérations de la procédure AFEC

Le programme que nous allons analyser a pour objectif de réunir en un certain nombre de classes les KI individus répertoriés dans le tableau POSIJ.

2.1 Effet de la procédure DIS

La procédure DIS opère sur deux tableaux-lignes notés :

TAB1[M1,1],..., TAB1[M1,J],..., TAB1[M1,KJ]

TAB2[M2,1],..., TAB2[M2,J],..., TAB2[M2,KJ]

Après exécution de la procédure, la case DIS2 a comme contenu :

$$\sum \{ (TAB1[M1,J] - TAB2[M2,J])^2 ; J \in]KJ] \}$$

Ainsi DIS (racine carrée de DIS2) est la distance euclidienne entre les deux points de $\mathbb{R}^{KJ}$ donnés respectivement par TAB1 et TAB2.

2.2 Interprétons d'abord les conditions a) b) et c)

a) $\forall I \in]KI], NAFI[I] \in]KN]$

Tout individu doit être affecté à l'une des KN classes.

b) $\forall N \in]KN], EFN[N] \neq 0$

L'effectif de chacune des KN classes est non-nul.

c) Calcul du tableau EFN à partir du tableau NAFI.

NAFI donne les classes d'affectation des individus ; EFN comptabilise les individus affectés à chaque classe.

- Effet de la procédure GRAV

La procédure GRAV comprend trois blocs successifs :

1 - pour N:=1...

⋮

CENJ[N,J]:=0 ;

Mise à zéro du tableau des coordonnées des centres des classes.

2 - pour I:=1...

⋮

CENJ[NAUX,J]:=CENJ[NAUX,J]+POSIJ[I,J]

fin ;

La ligne de numéro N dans le tableau CENJ contient la somme des coordonnées des points affectés à la N-ième classe.

3 - pour N:=1...

⋮

Chacune des lignes du tableau CENJ est divisée par le nombre d'individus affectés à cette ligne.

Globalement : La procédure GRAV(POSIJ,CENJ,NAFI,EFN,KJ,KN) donne dans le tableau CENJ le centre de GRAVité de chacune des KN classes.

2.3 Effet de la procédure AFEC sur le tableau de données

I	1	2	3	4	5	6	7	8	9
POSIJ[I,1]	5	6	7	7,5	9	3	2	1	0

avec MAXN = 9 et RAY = 2

Pour avoir une idée précise de la procédure, nous allons donner l'état des principaux tableaux aux différents pas de la boucle :

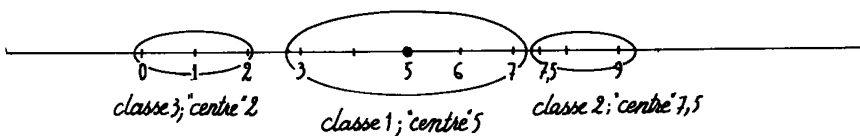
"pour I:=1 pas 1 jusqu'à KI..."

KN	0	NAFI	1	2	3	4	5	6	7	8	9
I = 1	1		1	0	0	0	0	0	0	0	0
I = 2	1		1	1	0	0	0	0	0	0	0
I = 3	1		1	1	1	0	0	0	0	0	0
I = 4	2		1	1	1	2	0	0	0	0	0
I = 5	2		1	1	1	2	2	0	0	0	0
I = 6	2		1	1	1	2	2	1	0	0	0
I = 7	3		1	1	1	2	2	1	3	0	0
I = 8	3		1	1	1	2	2	1	3	3	0
I = 9	3		1	1	1	2	2	1	3	3	3

KCHAN	0	EFN	1	2	3	4	5	6	7	8	9
1	1	EFN	1	CENJ	5						
2	1	EFN	2	CENJ	5						
3	1	EFN	3	CENJ	5						
4	1 2	EFN	3 1	CENJ	5	7,5	2				
5	1 2	EFN	3 2	CENJ	5	7,5	2				
6	1 2	EFN	4 2	CENJ	5	7,5	2				
7	1 2 3	EFN	4 2 1	CENJ	5	7,5	2				
8	1 2 3	EFN	4 2 2	CENJ	5	7,5	2				
9	1 2 3	EFN	4 2 3	CENJ	5	7,5	2				

KCHAN contient la valeur 9 ; on a 3 classes ; le rayon n'a pas changé de valeur.

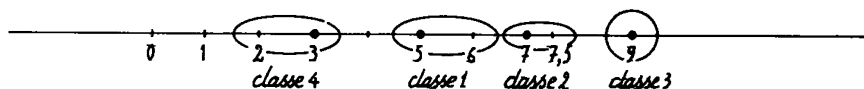
Géométriquement, on obtient :



2.4 AFEC (POS1J, CENJ, NAFI, EFN, 9, 1, 0, 4, KCHAN 1)

Effet de la procédure AFEC pour $KN = 0$, le nombre maximum de classes étant 4 et le rayon valant au départ 1.

2.4.1 Au pas 1 = 8, on a déjà fait la classification suivante :



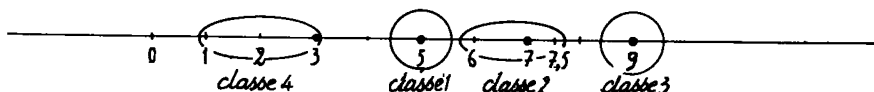
les centres des classes sont notés.

NOVA = 0 signifie que le point n° 8 de coordonnée 1 ne peut être affecté à l'une des classes existantes (sa distance à chacun des centres est supérieure à la valeur de RAY : 1).

$KN = MAXN$ signifie qu'on a atteint le nombre maximum de classes autorisé.

2.4.2 On double le rayon $RAY = 2$; KCHAN prend la valeur 9 et on retourne à ETIQ pour réaffecter les points aux centres déjà trouvés.

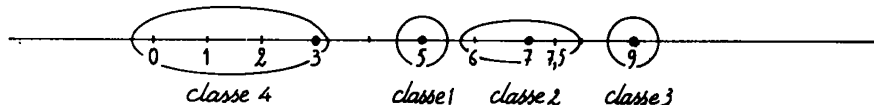
On aboutit à la classification non complète.



Comme pour $RAY = 1$, on n'a pas affecté "0", RAY devient 4.

2.4.3 $RAY = 4$; $KCHAN = 9$; Retour à ETIQ

On obtient la classification :



2.5 Effet du bloc terminant la procédure AFEC

1. dans le cas où $\forall N \in]KN]$, $EFN[N] \neq 0$

On aura toujours $NN[N] = N$ (i.e. le nouveau numéro de la classe N est le même que l'ancien).

Ainsi, ce bloc n'a aucun effet.

2. dans le cas : $EFN[1] = 0$ et $\forall N \in]KN] - \{1\}$, $EFN[N] \neq 0$.

On renumérote les centres de CENJ en supprimant la première classe à laquelle aucun point n'est affecté.

Ainsi les tableaux suivants :

N	1	2	3	4
EFN[N]	0	2	2	3

deviennent

N	1	2	3
EFN[N]	2	2	2

I	1	2	3	4	5	6	7
NAFI[I]	2	2	3	4	4	3	4

I	1	2	3	4	5	6	7
NAFI[I]	1	1	2	3	3	2	3

3. dans le cas général, le bloc terminant la procédure AFEC élimine les classes dont l'effectif est nul. Il renumérote les classes restantes de 1 à la nouvelle valeur de KN.

2.6 Effet de la procédure AFEC dans le cas général

2.6.1 KN = 0 au départ.

Aucune des classes ne sera vide... si on supprime le bloc final, l'effet de la procédure sera le même.

2.6.2 Valeur à donner à MAXN pour être assuré que l'instruction "si NOVA = 0 $\wedge$ KN = MAXN alors..." ne joue jamais.

Comme nous l'avons constaté en 4.1, le rayon étant fixé il peut se produire que le nombre de classes ne soit pas suffisant.

Cette instruction augmentant la valeur de RAY ne jouera pas si l'on accepte suffisamment de classes.

On pose MAXN = KI.

2.6.3 Description en termes géométriques de l'effet de la procédure

AFEC dans le cas particulier où KN = 0 (i.e. le tableau des centres des classes est vide).

La procédure AFEC va créer un nombre de classes ne dépassant pas le nombre MAXN autorisé et affecter les individus à ces classes, selon le principe :

- Un individu distant d'une classe créée d'une valeur inférieure à RAY est affecté à cette classe dans la mesure où il est plus éloigné du centre de toute autre classe. (En cas d'égalité, on retient le centre d'indice le plus fort).

- Un individu distant d'au moins RAY de toute classe déjà créée devient le centre d'une nouvelle classe (si MAXN n'est pas dépassé).

- Si MAXN est dépassé, on double la valeur de RAY et on recommence l'affectation en début de procédure.

2.7 Données KI = 4 ; KJ = 1 ; MAXN = 4 ; RAY = 2 ; MAXIT = 3

I	1	2	3	4
POSIJ[I,1]	2	1	0	4

a - Premier appel de la procédure AFEC

On obtient KN = 1 KCHAN = 4

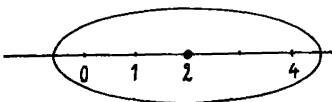
N	1
CENJ[N,1]	2

I	1	2	3	4
NAFI[I]	1	1	1	1

N	1
EFN[N]	4

c'est-à-dire :

une seule classe de centre 2 où sont affectés les 4 individus.



b - Premier appel de la procédure GRAV

Le nouveau centre est 7/4 i.e.

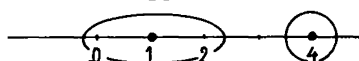
$$\frac{N}{\text{CENJ}[N,1]} \left| \frac{1}{7/4} \right. \quad \text{KCHAN} = 1$$

c - Deuxième appel de la procédure AFEC

On obtient une classification en deux classes, de centres respectifs 7/4 et 4. (La classe de centre 4 a été créée du fait que la distance de 4 à 7/4 dépasse la valeur du rayon et que le nombre maximum de classes autorisé n'est pas dépassé).

d - Effet de l'ensemble du programme

Le deuxième appel de GRAV nous donne



i.e. 2 classes de centres respectifs 1 et 4.

Au troisième appel de la procédure AFEC, on obtient $\text{KCHAN} = 0$. Ainsi le programme s'arrête ; la classification retenue est celle ci-dessus.

L'état des différentes variables et des tableaux est le suivant :

KN = 2	RAY = 2	KCHAN = 0	IT = 2																							
<table> <tr> <td>N</td> <td>1</td> <td>2</td> </tr> <tr> <td>CENJ[N,1]</td> <td>1</td> <td>4</td> </tr> </table>	N	1	2	CENJ[N,1]	1	4		<table> <tr> <td>I</td> <td>1</td> <td>2</td> <td>3</td> <td>4</td> </tr> <tr> <td>NAFI[I]</td> <td>1</td> <td>1</td> <td>1</td> <td>2</td> </tr> </table>	I	1	2	3	4	NAFI[I]	1	1	1	2		<table> <tr> <td>N</td> <td>1</td> <td>2</td> </tr> <tr> <td>EFN[N]</td> <td>3</td> <td>1</td> </tr> </table>	N	1	2	EFN[N]	3	1
N	1	2																								
CENJ[N,1]	1	4																								
I	1	2	3	4																						
NAFI[I]	1	1	1	2																						
N	1	2																								
EFN[N]	3	1																								
(2 centres de coordonnées respectives 1 et 4)		(les 3 premiers points sont affectés à la première classe le point 4 à la seconde)																								

2.8 Effet de l'ensemble du programme

But du programme : classer KI individus d'un espace de dimension KJ en au plus MAXN classes, le nombre KN de classes n'étant pas fixé *a priori* mais borné supérieurement.

Etape 1 : Initialisation de la classification : obtenue par un premier appel de la procédure AFEC.

Chaque individu est affecté à une classe suivant le principe :

(1) Un individu distant d'au moins RAY de toute classe déjà créée devient le centre d'une nouvelle classe, dans la mesure où l'on ne dépasse pas le nombre maximum MAXN de classes prévu.

(2) Si MAXN est dépassé, on double la valeur de RAY et on recommence l'affectation au départ de la procédure AFEC.

Procédure GRAV

Remplace les centres des classes de l'étape précédente par les centres de GRAVité des points affectés à chaque classe.

Utilisation de AFEC

Après le procédure GRAV, on réutilise AFEC avec les nouveaux centres définis dans GRAV. Le principe d'affectation est le même que celui de l'étape 1.

La procédure AFEC se termine par un bloc supprimant les classes ayant un effectif nul et renumérotant les classes non vides.

Le programme utilise à tout de rôle les procédures AFEC et GRAV.

Il s'arrête :

soit lorsque deux utilisations successives de AFEC ne changent pas l'affectation des individus (KCHAN = 0)

soit lorsque l'on a atteint le nombre maximum (MAXIT) d'utilisations de la procédure AFEC.