Plan

[1] Introduction page 2

[2] Partie Theorique page 4

[3] Partie Programmation page 12

[4] conclusion et remarques. page 33

1 Introduction.

"faire une démarche analytique puis synthétique" est la composante principale de la démarche intellectuelle entreprise par Marcel Bouché et son équipe pour son travail sur les vers de terre

l'occasion de la mise en oeuvre de cette étude a été donné en territoire belge, ainsi l'équipe de chercheurs, ~~a~~ fait des prelevements sur 199 points répartis sur le territoire belge, les descripteurs (les variables) pris en compte porte sur les

1) les végétaux
2) les données topographique
3) les données pedologiques
4) les lombriciens

chaque variables a deux modalités : presence absence", le tableau de données est donc constitué de 0 et 1.

le problème qui sera traité, est de savoir comment on pourrait ~~savoir~~, reduire pour une demarche synthétique le nombre de descripteurs

Un algorithme sera proposé et

un programme en Fortran sera présenté et décrit.

Il sera aussi question des limites de cette méthode, telle qu'elle est présentée ici ...

Bibliographie :

1) Cours du D.E.A ; Analyse des données. U.S.T.L

2) " Application de l'analyse de données à l'élaboration de mini-sondages d'opinion
Massotte et Dambroise

3) " Choix de Variables en Analyse en Composante principales "
Bonifas - Escoufier - Gonzalez-Sabatte

2 Partie théorique

La méthode présenté dans cette partie consiste à faire un choix de variables, et ceci dans le but d'avoir un tableau de nombre de colonne réduit, mais possédant une bonne approximation de l'information existant dans le tableau initial. Il s'agira aussi de faire un bilan de la méthode et de montrer ses limites...

2.1 Methodologie

On commence par transformer le tableau de données en une juxtaposition de tableaux de contingence, Ainsi chaque variable présente deux modalités : (absence, présence), et par suite chaque variable donne naissance à un tableau de contingence.

On pose N : nombre d'individus
p : nombre de variables
" T_i : le tableau de contingence obtenu à partir de chaque variable

On utilisera aussi la notation $\sum_{i=1}^{p} T_i$ pour dire que T est une juxtaposition des tableaux T_i pour $i = 1, \dots, p$.

Choisir les r meilleures variables, revient à choisir une juxtaposition de r tableaux de contingence.

On rappelle que une A.F.C sur un tableau de contingence P est equivalente à une A.C.P sur l'étude statistique $(D_i^{-1} P D_j^{-1}, D_j, D_i)$

où :

D_i = matrice diagonale des poids des lignes de

D_j = " " " colonnes

2.2 Outil Fondamental :

A ayant une population de n individus, toute étude statistique est caractérisée d'une manière exclusive par l'opérateur produit scalaire.

Une comparaison avec une autre étude ne peut se faire que si celle-ci est effectuée sur les même individus, et c'est seulement là qu'elle a un sens..

soit

(X, Q, D) l'opérateur produit scalaire est XQ^tXD

(Y, B, D) " " YB^tYD

le fait d'avoir la même métrique D est justifié par la raison que c'est la métrique de poids des individus et qu'on a la même population pour les deux études.

Les deux opérateurs produit scalaires sont D-symétrique et le produit scalaire défini sur opérateurs D-symétrique : $(A, B) \longmapsto \langle A, B\rangle = Tr(A$ permettra une comparaison des deux études

Si on pose $W = XQ^tXD$

$W' = YB^tYD$

l'approximation sera quantifié par le coefficient Rv :

$$Rv(W, W') = \frac{Tr(WW')}{\sqrt{Tr(WW)}\sqrt{Tr(W'W')}}$$

de l'espace des operateurs D-symétriques, et on peut s'apercevoir aisement que le coefficient R_{IV} exprime le cosinus de l'angle entre les deux vecteurs. D'une manière matique, le problème se pose comme une recherche d'un Tableau T' extrait de Tableau initiale pour lesquels l'angle entre les deux vecteurs ω et ω' soit minimum.

2.4

Reduction du nombre de variables :

L'operateur produit scalaire associé à l'étude statistique $(D_i^{-1} T D_j^{-1}, D_j, D_i)$ est :

$$\omega = D_i^{-1} T D_j^{-1} {}^t T$$

on note $T_{(r)} = \bigcup_{i \in I} T_i$ où $I \subset \{1, \dots, p\}$ avec $\mathrm{card}\, I = r$

l'étude statistique correspondante est :

$$(D_{i(r)}^{-1} T_{(r)} D_{j(r)}^{-1}, D_{j(r)}, D_{i(r)}) \quad \text{où } D_{i(r)} \text{ et } D_{j(r)} \text{ sont définis comme precedement}$$

l'operateur produit scalaire associé est :

$$\omega_{(r)} = D_{i(r)}^{-1} T_{(r)} D_{j(r)}^{-1} {}^t T_{(r)}$$

Remarque :

On est amené ~~est~~ à faire une comparaison entre ω et $\omega_{(r)}$. mais ces deux

opérateurs ne sont pas définis par rapport à la même métrique D, car, d'une part, on a la métrique D_i (par rapport à (1)) et de l'autre part la métrique $D_{i(r)}$ et avec

$$D_i \neq D_{i(r)}$$

On montre que, dans ce cas, on peut utiliser le coefficient R_v, en effet :

si on note E l'étude statistique $(D_i^{-1} T D_j^{-1}, D_j, D)$

et $E_{(r)}$ " $(D_{i(r)}^{-1} T D_{j(r)}^{-1}, D_{j(r)}, D_{(r)})$

l'étude E est équivalente à l'étude :

$$(D^{-1} T^{(1)} Q_{(1)}^{-1}, Q_{(1)}, D)$$

où $T^{(1)}$ est défini par :

$$T^{(1)}(i,j) = \frac{T(i,j)}{\rho}$$

et ceci grace au fait que les opérateurs de produit scalaire sont homothétiques.

De la même manière $E_{(r)}$ est équivalente à l'étude : $(D^{-1} T^{(2)} Q_{(2)}^{-1}, Q_{(2)}, D)$

où $T^{(2)}(i,j) = \dfrac{T_{(r)}(i,j)}{x}$.

définir le R_v entre les opérateurs correspondant à E et $E_{(r)}$ c'est écrire le R_v correspondant aux opérateurs relatifs à ces dernières études.

Un autre raisonnement, utilisant, par exemple, l'utilisation du coefficient Rv pour comparer les deux études E et E(g) : Dans les deux études on travaille sur la même population et les individus de cette population gardent le même poids dans les deux études.

<u>Remarque 1</u>

1) Pour chaque Tableau T_i, la métrique poids des lignes est I_n (la matrice identité)

et relativement au Tableau T, cette matrice est pI_n; D_i^{-1} sera égale à $\frac{1}{p} I_n$.

l'opérateur W sera $\frac{1}{p} T D_j^{-1} {}^tT$

" $W(g)$ " $\frac{1}{r} T_{(g)} D_{(g)j}^{-1} {}^tT_{(g)}$

2) Si on pose $D = I_n$, alors les deux operateurs sont D-symétrique, en effet :

$$W = \frac{1}{p} D^{-1} T D_j^{-1} {}^tT$$

$$DW = D \frac{1}{p} D^{-1} T D_j^{-1} {}^tT = {}^t(DW)$$

même raisonnement pour $W(g)$.

On peut aussi utiliser le coefficient Rv pour quantifier l'approximation, et le problème à resoudre sera le suivant :

$$\left\{ \max_{W(g) \in \mathcal{M}(g)} Rv(W, W(g)) \right.$$ où $\mathcal{M}(g)$ est l'ensemble des operateurs de produit scalaire associés aux Tableaux $T(g)$

Remarque : 2

On peut trouver une écriture de W en fonction des operateurs W_i associés à chaque A.F.C du tableau T_i

$$W = \frac{1}{p} T D_j^{-1} {}^tT$$

$$W = \frac{1}{p} (T_1 | T_2 | \cdots | T_p) D_j^{-1} \begin{pmatrix} {}^tT_1 \\ {}^tT_2 \\ \\ {}^tT_p \end{pmatrix}$$

or on peut constater que :

$$D_j = \begin{pmatrix} D_{j1} & & 0 \\ & D_{j2} & \\ 0 & & \ddots \\ & & & D_{jp} \end{pmatrix}$$

et par suite : $W = \frac{1}{p} \sum_{i=1}^{p} T_i D_{ji}^{-1} {}^tT_i$

c'est à dire $W = \frac{1}{p} \sum_{i=1}^{p} W_i$

de même pour $W(r)$:

$$W(r) = \frac{1}{r} \sum_{i \in I} W_i \quad \text{avec} \quad I \subset \{1, \dots, p\} \text{ et } \operatorname{card} I = r$$

d'où alors : $Tr(W W(r)) = Tr\left(\frac{1}{p} \sum_{i=1}^{p} W_i \frac{1}{r} \sum_{j=1}^{r} W_{ij}\right)$

et $\|W\|^2 = \frac{1}{p^2} \sum_{i=1}^{p} \sum_{j=1}^{p} Tr(W_i W_j)$

$$\|W_{(r)}\|^2 = 1/r^2 \sum_{k=1}^{r} \sum_{j=1}^{r} Tr(W_{ik} W_{ij})$$

On déduit Rv^2 :

$$Rv^2(W\,W_{(r)}) = \frac{\sum_{k=1}^{p}\sum_{j=1}^{r} Tr(W_k W_{ij})}{\left(\sum_{i=1}^{p}\sum_{j=1} Tr(W_i W_j)\right)\left(\sum_{k=1}^{r}\sum_{j=1}^{r} Tr(W_{ik} W_{ij})\right)}$$

lorsque p et r sont grands, on ne peut plus déterminer l'optimum de manière analytique ; pour éviter l'explosion combinatoire un algorithme progressif est proposé.

2.5 : Algorithme progressif :

1) lorsque p et r sont grands, l'algorithme suivant est proposé :

On définit une suite de problème $P_i : i=1,..r$

où $P_i = \begin{cases} \text{Max } Rv(W, W_{(i)}) \\ W_{(i)} \in \mathcal{M}_i \end{cases}$

où $\mathcal{M}_i$ est l'ensemble des opérateurs de produits scalaire associés à :

$$\left(\prod_{l=1}^{i-1} T_{k_l}\right) J\, T_{k_i} \qquad \text{où } \prod_{l=1}^{i-1} T_{k_l}$$

est fixé à la $(i-1)^{\text{ème}}$ étape.

2) On utilisera aussi la décomposition de W et $W_{(r)}$ en somme d'opérateurs W_i :

En effet, on suppose qu'on ait calculé $Tr(W\,W_{(i)})$ et $\|W_{(i)}\|^2$.

à l'étape $(k+1)$:

$$Tr(W W_{(k+1)}) = \sum_{l=1}^{p} \sum_{j=1}^{k+1} Tr(W_l W_{ij})$$

$$= Tr(W W_k) + Tr(W W_{i_{k+1}})$$

de même

$$\| W_{(k+1)} \|^2 = \| W_{(k)} \|^2 + \sum_{l=1}^{k} Tr(W_{i_{k+1}} W_{il}) + \sum_{l=1}^{k} Tr(W_{il} W_{i_{k+1}})$$

$$+ Tr(W_{i_{k+1}} W_{i_{k+1}})$$

or : $Tr(AB) = Tr(BA)$ pour tout A, B

on déduit alors :

$$\| W_{(k+1)} \|^2 = \| W_{(k)} \|^2 + 2 \sum_{l=1}^{k} Tr(W_{i_{k+1}} W_{il}) + Tr(W_{i_{(k+1)}} W_{i_{(k+1)}})$$

3 Description des Programmes.

Le travail de programmation a été décomposé en trois parties.

1ère Partie : Elle consiste à faire le calcul de la norme du produit $W \times W$, où W est l'opérateur produit scalaire du tableau initiale

2e Partie : Elle consiste à donner les éléments initiaux pour déclencher le processus itératif ; c'est à dire parmi toutes les variables on cherchera celle qui donnera le "Rv^2" maximum et on calculera la norme de $W_{j_0} \times W_{j_0}$ et la norme de $W \times W_{j_0}$ où :

W est l'opérateur produit scalaire relatif au Tableau T

et

W_{j_0} est l'opérateur produit scalaire relatif au Tableau T_{j_0}, relatif à la variable j_0, avec ses deux modalités absence présence

3e partie : avec les résultats de la 2ème partie, on peut aborder cette partie ; on place la variable j_0 en 1ère colonne ; puis on cherchera dans l'ensemble des variables qui restent celle qui va donner, avec la variable j_0, le plus grand "Rv^2" ; puis après on la permute avec la 2ème variable du Tableau, et ainsi de suite ; et à chaque étape on utilise les

Résultats de l'étape précédente :

$$Tr(WW_{k+1}) = T(WW_k) + Tr(WW_{i_{k+1}})$$

De même :

$$\|W_{k+1}\|^2 = \|W_{(k)}\|^2 + 2\sum_{l=1}^{k} Tr(W_{i_{k+1}} W_{i_l}) + Tr(W_{i_{k+1}} W_{i_{k+1}})$$

1ère Partie :

Entrée : le tableau de données T

Sortie : l'operateur W et la valeur TR (W/W)

Arguments utilisés :

P : Nombre de colonne de Tableau T

PI : Nombre de variables, ou encore de tableaux de contingences. comme chaque variable presente deux modalités, on a $P = 2(PI)$

N : Nombre de prélèvements, ou d'individus

T : le tableau de données de dimension N lignes et P colonnes.

TR : le tableau transposé du Tableau T

D : la matrice diagonale ; elle sera prise comme un vecteur

W : l'operateur produit scalaire

W1 : le produit W×W

NORM : c'est la norme au carré de l'operateur W1 . NORM = TR (W1).

Sous-Programme :

comme on a affaire à plusieurs produit de matrices ; on a constitué un sous programme de produit de matrice : PRODIM

Les paramètres principaux du programme :

A, B : les deux matrices dont on veut avoir le produit

IL : Nombre de lignes de la matrice A

IC : Nombre de colonnes de la matrice A et c'est aussi le nombre de lignes de la matrice B

IN : Nombre de colonnes de la matrice B

C : Matrice résultat du produit de A par B, elle sera de IL lignes et IN colonnes.

Quantités d'operations :

une ~~evaluat~~ évaluation du nombre d'operation executé dans ce programme, donne, en fonction du nombre d'individus N et de variables P :

Nombre de soustractions : 2

" divisions : 4

" addition : $2N + N^2(N+P-2)$

" multiplication : $NP + N^2(P+N) + 1$

A.N : si $N = 199$ $P = 321$

Nombre de "×" : 33.432.200

" "+" : 33.225637

2e partie

Entrée : / le Tableau T
/ le Tableau W
/ la valeur NOR2 = TR (W.W)

Sortie : / le numéro de la meilleure variable
/ la valeur TR (Wj.Wj) où Wj est l'opérateur produit scalaire correspondant à cette variable
/ la valeur TR (W Wj)

Arguments utilisés :

N : nombre d'individus
P : nombre de colonnes
PI : nombre de variables
T : Tableau de données
PM1 : la valeur P-1
W : opérateur produit scalaire
TP : Tableau dans lequel les variables vont défiler
TPR : son transposé
WJ : opérateur produit scalaire associé à TP
WWJ : résultat du produit W.WJ
WJ1 : résultat du produit WJ.WJ
NW2 : vecteur qui rassemblera les valeurs de TR (WJ.WJ)
TRC2 : vecteur qui rassemblera les valeurs de TR (W.WJ)
D : vecteur mélique poids des colonnes

Relatif au tableau 1

RV2 : Vecteur qui rassemblera les valeurs du carré du coefficient Rv

<u>Sous programme utilisé :</u>

c'est le même que dans la 1ère partie

<u>Quantité d'opérations :</u>

en fonction du nombre d'individus N et du nombre de variable P, une évaluation du nombre d'operation donne :

'−' P fois

'+' $P\left[(2N^2+1)(N-1)+N^2\right]$ fois

'÷' 4P fois

'×' $P(N+1)(2N^2+2)$ fois

<u>A.N</u>

pour N = 199 et P = 321

'−' 321 fois

'+' 4.792.454.565 fois

'÷' 1284 fois

'×' 5.084.896.800 fois

3ème Partie

Entrée : -/ Tableau de données T
-/ operateur W
-/ valeur de NORE
-/ Numéro de la meilleure variable
-/ valeur de MWR2(1)
-/ valeur de TRC2(1)

Sortie -/ Tableau avec variables classées
-/ numéro de variable par ordre decroissant : de celle qui donne le maximum d'information à celle qui donne le moins d'information.

Arguments utilisés :

N : Nombre d'individus
P : Nombre de colonnes
PI : Nombre de variables
NOR2 : c'est tr (W.W)
T : Tableau des données
TI : Copie du Tableau T
W : operateur produit scalaire relatif à T
TJ : Tableau dans lequel les variables sont deplacer
TJR : le transposé de TJ
WJ : operateur de produit scalaire relatif au tableau TJ
WWJ : c'est le produit W.WJ

WJ2 : c'est le produit WJ.WJ

WR : operateur de produit scalaire pour les variables anterieures à la variable de la j^{ieme} étape

WR1 : c'est le resultat du produit WJ.WR

DMS : c'est la trace de WR1

DMK : c'est ce qui represente $\sum_{l=1}^{k} TR(W_{i_{k+1}} W_{i_l})$ (si on est à l'étape (k+1))

NWR : le vecteur regroupant les valeurs de TR(WJ.WJ) où WJ est l'operateur produit scalaire relatif à la j^{ieme} variable

TRC2 : le vecteur regroupant les valeurs de TR(W1.WJ)

D : Vecteur representant la métrique des colonnes.

<u>Sous-Programme</u> :

même chose que dans la partie 1 et la partie 2, on utilise une procedure de produit de matrice.

<u>Quantité d'operations</u> :

- Nombre d'additions : $\sum_{j=1,2P-1,2} \left(P - \frac{j-1}{2}\right)\left(2N^3 - N^2 + 4N + \frac{j-1}{2}(N^2+2N+2)\right)$

pour $N = 129$ $P = 321$

à l'etape $j = 1$; on aura le nombre :

5.047.070.000

- Nombre de [illegible]

c'est $\sum_{j=1,2P-1,2} \left(P - \frac{j-1}{2}\right)\left[2N(N^2+N+1) + \frac{j-1}{2}(N^3 + 2N(N+1) + 3)\right]$

pour $N = 199$, $P = 321$

à l'étape $J = 1$, le nombre est :

5.059.340000

- Pour le nombre de divisions l'évaluation est :

$$\sum_{j=1,2P-1,2} \left(P - \frac{j-1}{2}\right)\left(2 + 3\frac{j-1}{2}\right)$$

si $P = 321$ ce nombre est 642 pour $j=1$

- Pour la soustraction, l'évaluation est :

$$\sum_{j=1,2P-1,2} \left(P - \frac{j-1}{2}\right)\left(4 + \frac{j-1}{2}\right)$$

pour $P = 321$, à la 1ère étape, le nombre est : 1284

<u>Remarque</u>. Il ne faut pas espérer que le nombre d'opérations diminue quand le nombre d'itérations augmente (l'indice j augmente) car le nombre d'opérations concernant les variables avant j va augmenter.

Remarque : Pour des problèmes de temps de calcul ; il s'est avéré nécessaire de faire le travail dans un premier lieu sur un tableau réduit : les variables vers le Terre (20)

TABLEAU INITIAL
****

| | | |

Programme de la 1ère Partie :

Programme Principal

```
C*********************************************************************
C*********************************************************************
C     PROGRAMME PRINCIPAL
C BUT :CALCUL DE LA NORME DE L'OPERATEUR PRODUIT SCALAIRE !!W!!
C*********************************************************************
      INTEGER P,PI
      REAL T,D
      DIMENSION T(50,50)
      LEC=11
      IMP=6
      N=25
      P=40
C*********************************************************************
C     LECTURE DES DONNEES
C     **********************
      DO 101 J=1,N
      READ(LEC,1,END=1000) (T(J,I),I=1,P)
    1 FORMAT (8X,60F2.0)
      II=J
  101 CONTINUE
 1000 N=II
C*********************************************************************
      CALL NORME(IMP,T,N,P)
      STOP
      END
```

Sous-programme "NORME"

```
      SUBROUTINE NORME(IMP,T,N,P)
C*********************************************************************
C SOUS-PROGRAMME POUR LE CALCUL DE :!!W!!2
C*********************************************************************
      INTEGER P ,PI,PM1
      REAL TR,W,D,T,NOR2
      DIMENSION TR(50,50),W(50,50)
      DIMENSION T(50,50),D(50) ,W1(50,50)
      IDC=12
CC    DO 131 I=1,N
C     WRITE(IMP,132) (T(I,J),J=1,P)
C132  FORMAT(10X,4F5.3)
C131  CONTINUE
      PI=P/2
      PM1=P-1
C***********************
C     POIDS DES COLONNES
C***********************
      DO 31 J=1,PM1,2
            DC     =0.
            DO 32 I=1,N
                  DC     =DC     +T(I,J)
 32          CONTINUE
             D(J)=1./DC
             D(J+1)=1./(N-DC)
 31   CONTINUE
```

voir suite de la subroutine sur la page suivante

```
C  CONSTITUTION DE LA MATRICE TRANSPOSEE
C*****************************************
      DO 33 I=1,P
      DO 33 J=1,N
      TR(I,J)=T(J,I)
 33   CONTINUE
C*****************************************************
C     PRODUIT DU TABLEAU T PAR LA MATRICE DES POIDS D
C*****************************************************
      DO 3 I=1,N
      DO 3 J=1,P
          T(I,J)=T(I,J)*D(J)
 3    CONTINUE
C*****************************************************************
C   CALCUL DU LA MATRICE PRODUIT SCALAIRE W
C*****************************************************************
      CALL PRODM(T,TR,N,P,N,W)
C*****************************************************************
C   CALCUL DE  ||W||2=NOR2
C*****************************************************************
      CALL PRODM(W,W,N,N,N,W1)
C     DO 521 I=1,N
C     WRITE(IMP,520) (W1(I,J),J=1,N)
C520  FORMAT(12X,20F5.1)
C521  CONTINUE
      FC=0.
      DO 35 I=1,N
      FC=FC+W1(I,I)
 35   CONTINUE
      NOR2=FC/(PI**2)
C*****************************************************************
C  EDITION DES RESULTATS
C*****************************************************************
      WRITE(IDC,299) NOR2
 299  FORMAT(F10.7)
      DO 122 I=1,N
      WRITE(IDC,121) (W(I,J),J=1,N)
 121  FORMAT(12F10.7,/(12F10.7))
 122  CONTINUE
      RETURN
      END
```

Sous-programme 'PRODM'

```
      SUBROUTINE PRODM(A,B,IL,IC,IN,C)
C***CALCUL DU PRODUIT DE DEUX MATRICE
      DIMENSION A(50,50),B(50,50),C(50,50)
      DO 140 II=1,IL
      DO 140 JJ=1,IN
      C(II,JJ)=0.
      DO 140 KK=1,IC
      C(II,JJ)=C(II,JJ)+A(II,KK)*B(KK,JJ)
 140  CONTINUE
      RETURN
      END
```

Programme de la 2ème Partie

```
C*********************************************************************
C*********************************************************************
C       SOUS-PROGRAMME DE CHOIX DE LA MEILLEUR VARIABLE
C       OUTIL FONDAMENTAL :LE COEFICIENT RV
C*********************************************************************
C*********************************************************************
        INTEGER P,PM1,PI,DC1
        REAL T,D,NOR2,W,TRC2,NW2,TP,PND,WJ,WWJ,TRC,WJ2
        DIMENSION T(50,50),W(50,50),RV2(50),TP(50,50)
        DIMENSION TPR(50,50),WJ(50,50),WWJ(50,50),WJ1(50,50)
        DIMENSION D(2),NW2(50),TRC2(50),W2(50,50)
        LEC=11
        LEC2=12
        IMP=6
        N=25
        P=40
        PM1=P-1
        PI=P/2
C***********************
C***LECTURE DES DONNEES
C**********************
        DO 101 J=1,N
        READ(LEC,1,END=1000)(T(J,I),I=1,P)
 1      FORMAT (8X,60F2.0)
        II=J
 101    CONTINUE
 1000     N=II
        READ(LEC2,102) NOR2
 102    FORMAT(F10.7)
        DO 201 I=1,N
        READ(LEC2,202) (W(I,J),J=1,N)
 202    FORMAT(12F10.7,/(12F10.7))
 201    CONTINUE
        WRITE(IMP,1111)
 1111   FORMAT(9X,'ETAPE',18X,'RV2'/9X,5(1H*),15X,7(1H*))
C8*********************************************************************
C****ON FAIT VARIER LES VARIABLES
C*********************************************************************
        DO 50 J=1,PM1,2
C*********************************************************************
C***CONSTITUTION DU TABLEAU DANS LEQUEL LES VARIABLES VONT DEFILER :TP
C***ET DE SON TRANSPOSE :TPR
C*********************************************************************
          DO 51 I=1,N
             TP(I,1)=T(I,J)
             TP(I,2)=T(I,J+1)
             TPR(1,I)=TP(I,1)
             TPR(2,I)=TP(I,2)
   51     CONTINUE
C*************************************
C   MATRICE DES POIDS COLONNES
C*************************************
          DC1=0.
          DO 52 I=1,N
          DC1    =DC1    +TP(I,1)
```

voir la suite sur la page suivante

```
  52      CONTINUE
          D(1)=1./DC1
          D(2)=1./(N-DC1)
C*******************************
C*********CALCUL DE TRC2(J)
C*******************************
          DO 151 I=1,N
          TP(I,1)=TP(I,1)*D(1)
          TP(I,2)=TP(I,2)*D(2)
 151      CONTINUE
          CALL PRODM(TP,TPR,N,2,N,WJ)
C         DO 121 I=1,N
C         WRITE(IMP,120) (WJ(I,L),L=1,N)
C 120     FORMAT(8X,3F5.3)
C 121     CONTINUE
          CALL PRODM(W,WJ,N,N,N,WWJ)
C         DO 221 I=1,N
C         WRITE(IMP,220) (WWJ(I,L),L=1,N)
C 220     FORMAT(12X,3F5.3)
C 221     CONTINUE
          TRC=0.
          DO 53 I=1,N
                TRC=TRC+WWJ(I,I)
   53     CONTINUE
          TRCC=TRC/PI
          TRC2(J)=(TRCC)**2
C         WRITE(IMP,219)  J, TRC2(J)
C 219     FORMAT(5X,'VALEUR DE TRC2(',I4,') = ',F10.7)
C*************************************
C******** CALCUL DE NORME DE NW2(J)
C*************************************
          CALL PRODM(WJ,WJ,N,N,N,WJ1)
C         DO 321 I=1,N
C         WRITE(IMP,320) (WJ1(I,L),L=1,N)
C 320     FORMAT(10X,3F5.3)
C 321     CONTINUE
          NW2(J)=0.
          DO 55 I=1,N
                NW2(J)=NW2(J)+WJ1(I,I)
   55     CONTINUE
C         WRITE(IMP,255) J, NW2(J)
C255      FORMAT(5X,'VALEUR DE NW2(',I4,') = ',F10.7)
C*************************************************C
C   CALCUL DE RV2(J)
C*****************************************************
          RV2(J)=TRC2(J)/(NOR2*NW2(J))
          JV=(J+1)/2
          WRITE(IMP,155)  JV, RV2(J)
 155      FORMAT (9X,I4,15X,F10.7)
   50     CONTINUE
C*************************************************************
C*********TEST SUR LE RV POUR LE CHOIX DE LA MEILLEUR VARIABLE
C*************************************************************
          K=1
          RVV=RV2(1)
          DO 56 J=3,PM1,2
          IF(RV2(J).LT.RVV) GOTO 56
          RVV=RV2(J)
          K=J
```

voir suite sur la page suivante

```
   56     CONTINUE
C*****************************
C********RESULTATS A GARDER :
C*****************************
          NW2(1)=NW2(K)
          TRC2(1)=SQRT(TRC2(K))
C*****************************
          JK=(K+1)/2
C*****************************************
C  EDITION DES RESULTATS
C*****************************************
          WRITE(IMP,57)  JK,NW2 (1),TRC2(1)
   57     FORMAT(1X,'NUMERO DE LA MEILLEURE VARIABLE EST : ',I4,
     +'       NW2(1) = ',F15.8,'          TRC2(1) = ',F15.8)
          STOP
          END
```

Resultats de la 2ème Partie :

ETAPE	RV2
*****	*******
1	0.5057492
2	0.6013321
3	0.5602581
4	0.5246069
5	0.5287452
6	0.5917605
7	0.5730760
8	0.5097882
9	0.5702990
10	0.5430595
11	0.5302833
12	0.6058567
13	0.5543182
14	0.5543482
15	0.5631451
16	0.5201897
17	0.5285817
18	0.5953793
19	0.5706361
20	0.5374452

NUMERO DE LA MEILLEURE VARIABLE EST : 12 NW2(1) = 1.99999523

TRC2(1) = 1.15769482

Programme de la 3ème Partie :

```
C*********************************************************************
C*********************************************************************
C*****PROGRAMME DE CHOIX DES MEILLEURS VARIABLES
C*********************************************************************
C*********************************************************************
      INTEGER P,PM1,PI
      REAL NOR2,NW2
      DIMENSION TJ(50,50),WJ(50,50),TRC2(50),WR(50,50)
      DIMENSION WWJ(50,50),W(50,50),RV2(50),NW2(50)
      DIMENSION WJ2(50,50),T(50,50),D(2),WR1(50,50)
      DIMENSION TJR(50,50),TI(50,50)
      LEC=11
      LEC2=12
      IMP=6
      N=25
      P=40
      PI=P/2
      PM1=P-1
C*********************************************************************
C***LECTURE DES DONNEES ********************************************
C*********************************************************************
      DO 101 J=1,N
      READ (LEC,1,END=1000)  (T(J,I),I=1,P)
  1   FORMAT (8X,60F2.0)
      II=J
 101  CONTINUE
 1000 N=II
      READ(LEC2,602) NOR2
 602  FORMAT(F10.7)
      DO 701 I=1,N
      READ(LEC2,702) (W(I,J),J=1,N)
 702  FORMAT(12F10.7,/(12F10.7))
 701  CONTINUE
C*****************************************************
C  COPIE DU TABLEAU T  DANS UN TABLEAU  TI
C*****************************************************
      DO 742 I=1,N
      DO 742 J=1,P
      TI(I,J)=T(I,J)
 742  CONTINUE
C*****************************************************
C  EDITION DU TABLEAU INITIAL
C*****************************************************
      WRITE(IMP,1111)
 1111 FORMAT (20X,'TABLEAU INITIAL ',/,15X,26(1H*),/)
       DO 1002 I=1,N
      WRITE(IMP,1003) (TI(I,J),J=1,P)
 1003 FORMAT(8X,60F2.0)
 1002 CONTINUE
C*****************************************************
C  PRMUTATION DE LA VARIABLE "1"AVEC LA VARIABLE "30"
C*****************************************************=
      DO 743 I=1,N
      TEMP=T(I,1)
      T(I,1)=T(I,23)
```

voir suite page suivante

```
      TEMP=T(I,2)
      T(I,2)=T(I,24)
      T(I,24)=TEMP
 743  CONTINUE
      WRITE(IMP,903)
 903  FORMAT(//25X,'EVOLUTION DU "RV" ',/,23X,21(1H*),/,10X,'ETAPE',
     +29X,'RV2',/,10X,6(1H*),25X,8(1H*)/)
C*****************************************************************
C****ON FAIT VARIER LES VARIABLES UNE PAR UNE
C*****************************************************************
C     INITIALISATIONS
C     ***************
      TRC2(1)=1.15769482
      NW2(1)=1.99999523
C*****************************************************************
       DO   60   J=3,PM1,2
C*****************************************************************
          TRC2(1)=1.15769482
          NW2(1)=1.99999523
          DO 61 JR =J,PM1,2
C         ******************
                DO 64 I=1,N
                       TJ(I,1)=T(I,JR)
                       TJ(I,2)=T(I,JR+1)
 64             CONTINUE
C               ******
          DDC=0.
          DO 62 I=1,N
                 DDC      =DDC     + TJ(I,1)
  62      CONTINUE
          D(1)=1./DDC
          D(2)=1./(N-DDC)
          DO 63 I=1,N
          DO 63 K=1,2
                 TJR(K,I)=TJ(I,K)
   63     CONTINUE
          DO 163 I=1,N
                  TJ(I,1)=TJ(I,1)*D(1)
                  TJ(I,2)=TJ(I,2)*D(2)
  163      CONTINUE
C  ****CALCUL DE TR(WWR)*************
          CALL PRODM(TJ,TJR,N,2,N,WJ)
          CALL PRODM(W,WJ,N,N,N,WWJ)
              TRC=0.
              DO 94 I=1,N
                     TRC=TRC+WWJ(I,I)
   94         CONTINUE
C             WRITE(IMP,194)   J,TRC
C194          FORMAT(1X,I4,10X,F10.6)
               TRC2(JR)=TRC2(J-2)+(TRC/PI)
C  ****CALCUL DE RV2(R)
          CALL PRODM(WJ,WJ,N,N,N,WJ2)
          TR1=0.
          DO 73 I=1,N
                 TR1=TR1+WJ2(I,I)
  73      CONTINUE
          JM2=J-2
          DMK=0.
```

voir suite page suivante

```
      DO 12 I=1,JM2,2
      DO 2 IK=1,N
      TJ(IK,1)=T(IK,I)
      TJ(IK,2)=T(IK,I+1)
      TJR(1,IK)=TJ(IK,1)
      TJR(2,IK)=TJ(IK,2)
2     CONTINUE
      SD=0.
      DO 3 IK=1,N
      SD=SD+TJ(IK,1)
3     CONTINUE
      D(1)=1./SD
      D(2)=1./(N-SD)
      DO 4 IK=1,N
          TJ(IK,1)=TJ(IK,1)*D(1)
          TJ(IK,2)=TJ(IK,2)*D(2)
4     CONTINUE
      CALL PRODM(TJ,TJR,N,2,N,WR)
      CALL PRODM(WJ,WR,N,N,N,WR1)
      DMS=0.
      DO 5 IK=1,N
      DMS =DMS +WR1(IK,IK)
5     CONTINUE
      DMK=DMK+DMS
12    CONTINUE
C     WRITE(IMP,176) J,DMK
C176  FORMAT(3X,I4,25X,F10.4)
      NW2(JR)=NW2(J-2)+TR1+(2*DMK)
      RV2(JR)=((TRC2(JR))**2)/(NOR2*NW2(JR))
61   CONTINUE
C  ***************************************************************
C  ***TEST DE CHOIX DE LA MEIELLEURE VARIABLE  ****************
C  ***************************************************************
            K=J
            RVV=RV2(J)
            JP2=J+2
      DO 74 JR =JP2,PM1,2
      IF((RV2(JR)).LT.RVV) GOTO 74
            RVV=RV2(JR)
            K=JR
74    CONTINUE
C  ****************************************************************
C    PERMUTATION AVEC LA PREMIERE VARIABLE
C  ****************************************************************
      JV=(J+1)/2
      RV2(J)=RV2(K)
      NW2(J)=NW2(K)
      TRC2(J)=TRC2(K)
      WRITE (IMP,175)  JV,RV2(J)
175   FORMAT (7X,I4,22X,F10.7)
      DO 75 I=1,N
            TEMP=T(I,J)
            T(I,J)=T(I,K)
            T(I,K)=TEMP
            TEMP=T(I,J+1)
            T(I,J+1)=T(I,K+1)
            T(I,K+1)=TEMP
75     CONTINUE
60     CONTINUE
```

```
C*********************************************************************
C    EDITION DU TABLEAU FINAL
C*********************************************************************
          WRITE(IMP,1112)
 1112     FORMAT(///5X,'TABLEAU AVEC VARIABLES CLASSEES '/5X,32(1H*)/)
          DO 1004 I=1,N
          WRITE(IMP,1005) (T(I,J),J=1,P)
 1005     FORMAT(8X,60F2.0)
 1004     CONTINUE
C*********************************************************************
          EP=0.1
          WRITE(IMP,1113)
 1113     FORMAT(///5X,'NUMERO DES VARIABLES DANS LE TABLEAU FINAL',/,
     +5X,42(1H*),/,2X,'ETAPE',25X,'VARIABLE',/,2X,6(1H*),25X,8(1H*)/)
          DO 160 L=1,PM1,2
          LK=(L+1)/2
          DO 169 J=1,PM1,2
             CP=0.
             DO 7 I=1,N
             CP=CP+ABS(T(I,L)-TI(I,J))
  7          CONTINUE
          K=(J+1)/2
          IF (CP.LT.(EP)) WRITE(IMP,888)  LK,K
 888      FORMAT(1X,I4,25X,I4)
 169      CONTINUE
 160      CONTINUE
          STOP
          END
```

Résultats de la 3ème Partie.

EVOLUTION DU "RV"
****

ETAPE	RV2
2	0.7677463
3	0.8466663
4	0.8912321
5	0.9216561
6	0.9388496
7	0.9504602
8	0.9605723
9	0.9678366
10	0.9739749
11	0.9783671
12	0.9819758
13	0.9846935
14	0.9875411
15	0.9899437
16	0.9920952
17	0.9939063
18	0.9958079
19	0.9978501
20	1.0000000

TABLEAU AVEC VARIABLES CLASSEES
****

```
0.1.0.1.0.1.0.1.1.0.0.1.0.1.1.0.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.
0.1.1.0.0.1.0.1.0.1.0.1.0.1.1.0.0.1.0.1.0.1.1.0.0.1.0.1.0.1.0.1.0.1.0.1.0.1.
1.0.0.1.1.0.1.0.0.1.0.1.0.1.0.1.0.1.1.0.0.1.0.1.1.0.0.1.0.1.0.1.0.1.0.1.0.1.
1.0.0.1.1.0.1.0.0.1.0.1.0.1.1.0.0.1.1.0.0.1.0.1.1.0.0.1.0.1.0.1.0.1.1.0.0.1.
1.0.0.1.0.1.0.1.0.1.0.1.1.0.0.1.0.1.1.0.0.1.0.1.0.1.0.1.0.1.0.1.1.0.0.1.0.1.
1.0.0.1.0.1.0.1.0.1.0.1.1.0.0.1.0.1.1.0.0.1.0.1.0.1.0.1.0.1.0.1.1.0.0.1.0.1.
1.0.1.0.1.0.0.1.0.1.0.1.1.0.0.1.0.1.1.0.0.1.0.1.1.0.0.1.0.1.0.1.0.1.0.1.1.0.
0.1.0.1.0.1.0.1.1.0.0.1.0.1.0.1.0.1.1.0.0.1.0.1.0.1.0.1.0.1.1.0.0.1.0.1.0.1.
0.1.0.1.1.0.0.1.0.1.0.1.0.1.0.1.0.1.1.0.0.1.0.1.1.0.0.1.0.1.0.1.0.1.0.1.0.1.
1.0.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.1.0.0.1.1.0.0.1.0.1.0.1.0.1.1.0.0.1.0.1.
0.1.0.1.0.1.1.0.0.1.0.1.0.1.0.1.0.1.1.0.0.1.0.1.0.1.1.0.0.1.0.1.0.1.0.1.0.1.
1.0.0.1.1.0.0.1.0.1.1.0.0.1.0.1.0.1.1.0.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.
0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.1.0.
0.1.0.1.1.0.0.1.0.1.0.1.0.1.0.1.0.1.1.0.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.
0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.
0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.1.0.
0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.1.0.
0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.1.0.
0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.1.0.
0.1.0.1.0.1.0.1.0.1.1.0.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.
0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.1.0.1.0.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.
0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.1.0.
0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.1.0.
0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.0.1.1.0.
0.1.0.1.1.0.0.1.0.1.1.0.0.1.0.1.0.1.1.0.0.1.0.1.0.1.0.1.1.0.0.1.0.1.0.1.0.1.
```

NUMERO DES VARIABLES DANS LE TABLEAU FINAL

ETAPE	VARIABLE
1	12
2	14
3	6
4	9
5	20
6	11
7	7
8	10
9	1
10	2
11	8
12	17
13	18
14	16
15	4
16	5
17	3
18	13
19	15
20	19

I Conclusion et remarques

1) Valeurs du "Rv"

La 1ère valeur donne un Rv égale à 0,60 on peut constater que cette valeur est assez elevé; mais on a remarqué sur d'autre exemples où le nombre d'individus était important que le Rv apporté par la 1ère variable était petit (de l'ordre de 0,3); ce phenomène peut s'expliquer par le fait que plus le nombre d'individus est grand, plus il est le système est complexe, si complexe pour que l'information apporté par la 1ère variable soit importante

2) Evolution de la valeur du coefficient Rv.

La 1ère valeur qui sort est 0,605; puis ça augmente rapidement (à la 7ème etape) à 0,950; ensuite ça monte doucement et progressivement, pour arriver a 1,00000 à la dernière étape.

Le fait de n'arriver à 1 qu'au bout de la dernière étape était possible, car le nombre de variables $P = 20$ est plus petit que le nombre d'individus $N = 25$

3/ Classement des variables

Si on prend les sept premières variables qui sortent, le coefficient R_v qu'elles apportent est 0,95 ; ceci veut dire qu'elle apportent 95 % de l'information.

D'autre part si on examine de près ces variables on s'aperçoit qu'elles sont trois types de paquets de variables qui étaient classés par ordre de leur degré de contribution sur les 1er, 2eme et 3eme axes de l'A.F.C qu'on a effectué en parallèle.

On peut conclure que les 7 premières variables représentent bien le système constitués des 25 premiers prélèvements.

4/ Ordre de classement d'une variable redondante

une variable intéressante "PAS" sort à la dernière étape ; en effet cette variable signifie pas de vert de terre ; c'est alors une variable dont l'information qu'elle apporte peut-être déduite des autres. c'est donc "naturel" qu'elle sort à la dernière étape.

5) Problème du temps de calcul :

Il y a un grand problème qui se pose au niveau de la 3ème partie de la programmation ; ce problème est celui du nombre d'opérations.

Pour réduire le nombre, donc le coût de calcul, il faut, si toutefois on ne peut pas améliorer le programme (ce qui est assez improbable !), jouer sur deux facteurs :

le premier : réduire le nombre d'individus, (de prélèvements), en faisant des paquets écologiquement homogènes

le 2ème : réduire le nombre d'itération et pour cela, il faut réduire le nombre de tableau de contingence, ce qui revient à réduire le nombre de descripteurs

On peut envisager de faire, ici aussi, des paquets de variables homogènes comme l'exemple présenté ici, qui est un paquet de variable vert de terre.

LOMBRICIENS : LA DÉCOUVERTE DU SIÈCLE ...