

U S T A

ECOLE NATIONALE POLYTECHNIQUE

DEPARTEMENT D'ELECTRONIQUE

PROJET DE FIN D'ETUDE

Thèse d'ingénieur en électronique

Analyse en Composantes Principales

Logiciel & Applications

المدرسة الوطنية للعلوم الهندسية
المكننة
ECOLE NATIONALE POLYTECHNIQUE
BIBLIOTHÈQUE

Proposé par:

M^r A. ABDELLAOUI

D^r de spécialité

Etudié par:

CHOUGUI Ahmed

MOKRANI Karim

BIBLIOTH

U S T A

ÉCOLE NATIONALE POLYTECHNIQUE

DÉPARTEMENT D'ÉLECTRONIQUE

PROJET DE FIN D'ÉTUDE

Thèse d'ingénieur en électronique

ANALYSE EN COMPOSANTES PRINCIPALES

Logiciel & Applications

Proposé par :

M^r A. ABDELLAOUI

D^r de spécialité

Étudié par :

CHOUGUI Ahmed

MOKRANI Karim

Janvier 82



 REMERCIEMENTS

Nous tenons à remercier notre promoteur Mr. A.ABDELLAOUI Directeur du Groupe de Recherche en Télédétection de nous avoir accueillis dans son laboratoire pour l'élaboration de notre projet.

Qu'il veuille bien trouver ici l'expression de notre sincère gratitude.

Nos remerciements vont à Mr. CHIFFLET dont les conseils nous ont été très utiles.

Que tous ceux qui, ont contribué de près ou de loin à l'élaboration de ce projet trouvent ici l'expression de notre profonde reconnaissance.

SOMMAIRE

INTRODUCTION

I - PRESENTATION DE LA TELEDETECTION

- 1- Généralités .
- 2- Acquisition de données
- 2-1- Acquisition par systèmes aéronautiques
- 2-2- Acquisition par satellite

II - TRANSFORMEE DE KARHUNEN-LOEVE OU ANALYSE EN COMPOSANTES PRINCIPALES

- 1- Généralités sur les transformations linéaires orthogonales.
- 2- Analyse en composantes principales
 - 2-1- Position du problème
 - 2-2- Démonstration
 - 2-3- Formalisation
 - 2-4- Conclusion
- 3- Applications de l'analyse en composantes principales
 - 3-1- Application à la télédétection
 - 3-2- Application aux sondages

III- PROGRAMMATION DE L'ANALYSE EN COMPOSANTES PRINCIPALES:

IV - RESULTATS DES APPLICATIONS

- 1- Resultats relatif a la télédétection
- 2- Résultats relatifs aux sondages

CONCLUSION

ANNEXES .

TELEDETECTION : C'est un ensemble de techniques mises en oeuvre à partir d'avions, de ballons, de satellites et qui ont pour but d'étudier, soit la surface de la terre (ou d'autres planètes), soit l'atmosphère en utilisant les propriétés des ondes électromagnétiques émises, réfléchies ou diffractées par les différents corps observés. Plus généralement, ce terme peut qualifier toute technique de mesure basée sur les propriétés énoncées ci-dessus, qui n'implique pas de contact direct entre le capteur et l'objet de la mesure

Une chaîne de télédétection comprend :

- 1 -- une zone à étudier
- 2 -- un milieu de propagation
- 3 - une plate-forme d'observation
- 4 - des capteurs
- 5 - un système d'enregistrement ou de transmission de données
- 6 -- un centre de recueil et de prétraitement de données
- 7 - un centre de traitement (ou d'interprétation)

Notre étude concerne le sixième point de cette chaîne.

Depuis l'apparition des expériences de ressources terrestres, on cherche encore la meilleure façon d'orienter le traitement des données issues de ces campagnes aéroportées. Le plus gros problème étant la quantité d'information numériques à manipuler.

Après un bref exposé, au chapitre I, de la présentation de la télédétection et au chapitre II, de la méthode d'analyse en composantes principales, nous décrivons au chapitre III, la mise en oeuvre de cette méthode. Au chapitre IV sont expliqués les résultats obtenus pour deux types de données auxquelles nous avons appliqué notre programme : données de télédétection et données de sondages.

I PRESENTATION DE LA TELEDETECTION .

=====

I- 1 Generalités

D'une manière generale, on peut dire qu'il y a télédetection l'orsqu'on acquiert des informations sur un objet quelconque sans qu'il y ait contact physique entre cet objet et l'opérateur.

En somme, l'homme fait de la télédetection depuis toujours en utilisant trois de ses cinq sens pour acquérir des connaissances du monde extérieur , sans contact physique: La vue, l'ouïe et l'odorat .

La technique de télédetection apparait comme l'ensemble des moyens nécessaires pour étendre les possibilités humaines au dela de ces étroites limites.

En fait, on applique d'office des restrictions importantes à cette définition générale, nées du developpement historique de la technique de télédetection et de ses applications.

Actuellement, la définition recouvre seulement les applications de detection utilisant le rayonnement electro-magnétique comme porteur d'informations, a l'exclusion des autres techniques permises par la physique (onde sonore, champ électrique, magnétique ...)

Les ressources de rayonnement electromagnétique sont des sources naturelles (soleil et la terre) ou artificielles (laser ou radar) .

Le soleil et la terre émettent de l'énergie selon des lois physiques bien connues qui représentent, en particulier, la repartition spectrale de l'emission en fonction de la température (rayonnement du corps noir): le rayonnement du soleil est celui d'un corps noir (substance radiative idéale qui absorbe tous les rayonnements incidents et dont le rayonnement dépend uniquement de sa

température) après de 6000°K , c'est à dire que sont maximum se place à une longueur d'onde voisine de 0,5 micron, dans le jaune.

Le rayonnement de la terre est sensiblement celui d'un corps noir a la temperature de 3000°K . Le maximum de rayonnement se trouve alors dans l'infrarouge thermique, vers 10 microns.

Tout corps émet un rayonnement électromagnétique résultant de l'agitation thermique de ses atomes et molécules .

Les ondes électromagnétiques qui véhiculent l'information doivent traverser un milieu, l'atmosphère, dont il est par conséquent important de connaître les propriétés .

L'atmosphère est constituée de gaz (oxygène, azote, ozone, ...) et particules (gouttelettes d'eau des nuages, fumées, ...). Ces composants ont pour effet de diffuser et d'absorber le rayonnement selon des lois qui dépendent des dimensions des particules ou molécules diffusantes par rapport a la longueur d'onde. Il est, par exemple, bien connu que les gaz atmosphériques donnent ^{au ciel} sa couleur bleue, alors que les gouttelettes d'eau des nuages l'air donnent un aspect blanc et que les mêmes nuages qui sont opaques dans les longueurs d'onde du visible et du proche infra-rouge, deviennent transparents dans les longueurs d'ondes utilisées par les radars ..

Theoriquement, tout le spectre électromagnétique pourrait être utilisé pour l'exploration à distance, de l'ultraviolet aux ondes radio. Mais a l'état actuel de la technique, il n'existe de détecteurs que pour certaines régions du spectre .

D'autre part l'atmosphère joue un rôle perturbateur dans la propagation de ces rayonnements et seules quelques "fenêtres" peu affectées par l'absorption restent utilisables .

Il faut noter, en particulier, que l'atmosphère est transparente dans les bandes d'émission du soleil et de la terre, phénomène qui rend possible la télédétection .

En schématisant, on peut dire que l'atmosphère est transparente dans trois domaines principaux :

- * fenêtre du visible (0,4 à 0,75 μ m) et du proche infrarouge (0,75 à 11 μ m)

Utilisés pour l'enregistrement des rayonnement solaires réfléchis par les objets terrestres . C'est le domaine de la photographie aérienne classique et plus récemment des appareils à balayage multispectraux . Le domaine du proche infra rouge joue un rôle important dans la mesure où les longueurs d'ondes correspondantes sont celles des ondes fortement réfléchies par la chlorophylle, il formera donc la composante essentielle de la signature spectrale des différentes plantes .

- * fenêtre de l'infrarouge thermique (3 à 5 μ m et 8 à 14 μ m)

Dans ces deux bandes, on reçoit l'énergie électromagnétique émise par les corps terrestres, qui est fonction de leur température et de leur émissivité, c'est à dire le rapport de l'énergie émise réellement par le corps à celle qu'émettrait un corps parfait, appelé corps noir .

- * fenêtre des micro-ondes (de 1mm à quelques mètres)

Utilisable soit avec des capteurs actifs possédant leur propre source d'illumination, les radars, soit avec des radiomètres passifs recueillant l'énergie, très faible, émise par les corps terrestres dans ces longueurs d'onde .

Si on en vient aux techniques proprement dite de la télédétection, il faut étudier les capteurs, les véhicules et les moyens techniques de traitement des données .

Un capteur est un appareil capable de recevoir le rayonnement qui vient du sol, dans une certaine gamme de longueurs d'ondes, et de le transformer en un signal permettant la mémorisation de l'information. Il existe des capteurs très variés adaptés à diverses études, aussi nous ne décrirons que les principaux.

* Appareils photographiques :

Le premier et le plus connu de tous les capteurs est l'appareil photographique qui fournit une image sur film de la scène observée, image que l'on examine a loisir ensuite .Avec cet appareil photographique on ne peut obtenir des clichés que dans une bande de longueur d'onde qui peut aller de 0,4 à 0,9 micron, en utilisant diverses émulsions :

- Panchromatique: c'est le film noir et blanc **classique** entre 0,4 et 0,7 micron, c'est à dire dans le domaine du visible .

Couleurs naturelles : c'est là aussi le film couleur bien connu qui, dans le même domaine du visible, restitue une information de couleur grâce au trois couches sensibles du film .

- Infrarouge noir et blanc : Cette émulsion ressemble au panchromatique, la seule différence (essentielle) étant que la gamme de sensibilité s'étend dans le proche infrarouge de 0,7 a 0,9 micron .

- Infrarouge couleur : film sensible entre 0,5 et 0,9 micron, l'information est restituée sous forme d'une image en couleur ou, par un décalage spectral apparaissent en rouge les informations infrarouges, en jaune les informations rouges et en bleu les informations jaunes-vertes

.../...

Ce décalage fournit donc des couleurs non réelles, ce qui explique le terme de " fausses couleurs " souvent utilisé pour designer cette emulsion .

* Radiomètre multispectral à balayage :

Cet instrument se compose d'un dispositif optique a miroir tournant permettant d'analyser le sol point par point, ligne par ligne, à la manière du spot d'un écran de télévision. Puis cette image ponctuelle est dispersée à l'aide d'un prisme ou d'un réseau afin de fournir un spectre qui est capté par un certain nombre de détecteurs (photomultiplicateurs ou photodiodes) qui transforment le rayonnement en un signal électrique que l'on enregistre sur bande magnétique .

Cet appareil a de nombreux avantages :

- il permet de quantifier le rayonnement reçu par calibration de la chaîne complète à l'aide de sources étalons
- par ailleurs, il permet de travailler dans une gamme de longueurs d'ondes plus large que la caméra photographique, pouvant aller jusqu'à 15 microns dans l'infrarouge.
- enfin , il fournit une quantité d'information beaucoup plus grande que la photographie car le nombre de canaux n'est plus limité à 3 comme en photo couleurs, mais peut atteindre dix, voire davantage .

* Autres systemes :

Dans les longueurs d'ondes plus grandes que celles de l'infrarouge, il faut s'adresser à une nouvelle classe de capteurs, les récepteurs hyperfréquence, qui comprennent essentiellement le radiomètre millimétrique passif qui restitue une image du sol dans les longueurs d'onde millimétriques, et le radar a antenne latérale,

.../...

instrument actif qui eclaire le paysage avec des ondes hyperfréquences, reçoit l'énergie réfléchie et restitue une image du sol.

Les instruments hyperfréquence sont encore assez mal connus des utilisateurs potentiels, mais il est certain que la faculté de travailler " à travers les nubes " leur autorise un avenir certain.

Les véhicules utilisables sont tous ceux qui permettent de placer les capteurs à une distance suffisante de la surface examinée pour les études à faire .

On peut donc distinguer :

- le camion équipé d'un bras télescopique qui soulève une nacelle à quelques mètres du sol.
- les systèmes aéronautiques (avions, hélicoptères, ballons captifs ou dirigeables troposphériques) qui travaillent entre quelques centaines et quelques milliers de mètres d'altitude
 - les avions de haute altitude entre 10 et 20.000 mètres.
 - les ballons stratosphériques entre 30 et 40 Km.
 - en fin les fusées sonde et les satellites pour des altitudes supérieures à 20 Km.

Les performances de ces véhicules sont évidemment très différentes.

D'une manière générale, on considérera que l'on obtient une échelle d'autant plus petite qu'on est plus haut, la résolution au sol étant corrélativement diminuée et la surface observée élargie

Les données fournies par les capteurs doivent subir un traitement afin d'être ~~4~~ exploitables .

Dans l'exemple le mieux connu, celui de la photographie, on sait que ce traitement consiste en développement et tirage, opérations qui restituent une image que l'oeil humain peut examiner .

Dans d'autres cas, ce traitement peut devenir relativement complexe, mettant en oeuvre des machines capables de reconstituer des images a partir des informations portées par les bandes magnetiques. Il est aussi possible de traiter ces bandes par ordinateur pour analyser les données reçues et qui dans certains cas, peut aller jusqu'a une exploitation entièrement automatique en fournissant a l'utilisateur des cartes elaborées .

I. 2 ACQUISITION DES DONNEES .

I-2.1 Acquisition par système aéronautique :

A ses débuts, la photographie aerienne etait prise a partir de ballons et d'avions dont le gros problème etait la stabilité de ces appareils.

Actuellement, on utilise des avions spécialement aménagés. Les principaux modèles actuels de capteurs sont des chambres a film permettant, en fonction de la hauteur de vol, un large éventail d'échelles ($1/500$ à $1/110.000$).

On utilise une batterie de quatre appareils photographiques montés sur une platine de fixation s'adaptant sur les trappes des avions utilisés. Ces appareils sont équipés d'un objectif. Le déclenchement des quatre appareils est simultané , il est commandé par un intervalomètre . Celui-ci autorise un déclenchement réglable des prises de vues, ce qui permet d'obtenir un recouvrement réglable des images. L'ouverture des diaphragmes étant réglée manuellement avant chaque séries de prises de vues .

Suivant les objectifs assignés, on utilise différents types d'émulsions associées a des filtres, mais on ne peut couvrir que la fenêtre du visible et du proche IR (jusqu'a 0,9 micron avec des emulsions sensibles au proche IR)

- film panchromatique : d'usage le plus courant, employé d'abord pour la photogrammétrie puis pour diverses thématiques, géologie, forestière,...

- L'infrarouge noir et blanc : qui se caractérise en particulier par une absorption totale du rayonnement IR par l'eau libre.

D'autre part, cette émulsion permet d'étudier les végétaux dont les réponses dans l'IR varient fortement selon leurs états.

- Couleurs naturelles : très intéressantes pour les études de pollution et d'hydrologie côtière. En effet la pénétration des rayonnements de courte longueur d'onde, très bonne en eau claire, varie avec la teneur en particules solides en suspension dans l'eau

- l'infrarouge couleur dit encore "fausse couleur" comprenant par superposition sur un même support, deux couches sensibles aux radiations visibles et une troisième sensible à l'infrarouge, ce qui est très intéressant pour l'étude de la végétation.

Une mission photographique aérienne se déroule selon un plan de vol précis, l'appareil effectuant des passages successifs au-dessus de la zone à étudier afin d'obtenir des bandes parallèles de photographies.

Les photographies sont prises avec recouvrement pour pouvoir restituer la zone étudiée.

La restitution permet de passer des photographies aériennes à la représentation du terrain à l'aide d'appareil de restitution photogrammétrique pour les besoins de la photo-interprétation (la photogrammétrie est une technique qui permet à partir de photographies de reconstituer avec précision la forme et les dimensions d'un objet)

La photo-interpretation est une technique utilisée depuis fort longtemps pour l'observation du sol et s'est appliquée essentiellement à l'analyse des photographies aériennes .

Des spécialistes entraînés, au vu des photos, sont capables de fournir des renseignements sur la région observée. Le degré d'analyse dépend essentiellement des expériences acquises sur les images passées .

La méthode est extrêmement puissante, l'oeil et le cerveau humain, biens entraînés, peuvent tirer parti des aspects multiples d'une image .

- Valeur(et éventuellement couleur)
- formes globales, formes locales
- éléments structuraux
- texture ou grain.

Par contre, depuis l'avènement de nouveaux systèmes d'observation, les limites de la méthode sont apparues nettement :

- le nombre limité de niveaux de gris distinguables par l'oeil est trop faible pour apprécier la finesse des observations.
- L'analyse est plutôt qualitative que quantitative .
- le coût d'une photographie est élevé

I-2-2 Acquisition des données par satellite .

Les enregistrements de la "lumière " réfléchiée ou diffractée par les corps terrestres et les phénomènes naturels sont généralement effectués sur les émulsions photographiques dont l'objet est d'être des capteurs chimiques mais leur usage est limité.

Ces difficultés ont trouvé leur solution par l'utilisation de capteurs à "balayage"(les scanners), embarqués sur des satellites, l'énergie réfléchiée par les objets est transformée en courant

d'intensité variable, puis digitalisée pour être en fin retransmise par ondes radio à des stations de réception terrestres.

Par processus inverse, l'information est visualisée pour fournir des "images" ou plus exactement des "pseudo-images". L'enregistrement à balayage permet en outre d'obtenir les réponses spectrales des objets dans différentes bandes du spectre électromagnétique.

L'analyse des données multispectrales permet de définir les caractères spécifiques d'un objet appelé "signature".

Parmi les véhicules utilisables comme plate-formes d'observation, le satellite permet, grâce à sa grande altitude de vol, d'avoir une vue synthétique de larges régions; de plus, la configuration de son orbite permet une certaine optimisation des caractéristiques de survol de certaines parties du globe, ainsi que la possibilité de répétitivité de passages. C'est donc un moyen qui peut répondre à certaines applications.

Dès le moment du lancement du satellite Landsat-1 le 23 Juillet 1972, l'attention de la communauté scientifique internationale était attirée sur le rôle que joueraient dans le futur ces familles de satellites destinés à la télédétection de ressources terrestres.

Les critères généralement retenus pour convaincre les expérimentateurs et les encourager à proposer des expériences étaient les suivantes :

a) Le caractère de répétitivité de l'acquisition des données relatives à une zone donnée : théoriquement, on pouvait compter sur une prise de données tous les 18 jours avec un satellite en orbite, ou tous les 9 jours avec 2 satellites (possibilité de suivre les phénomènes évolutifs);

b) Le caractère synoptique des prises de données, on a une observation et un enregistrement rapide de données sur de vast

.../...

vastes superficies .

c) L'intérêt des systèmes de détection donnant lieu à des mesures de luminance énergétique, répétées simultanément dans plusieurs bandes spectrales, et de nature à fournir des réponses discriminantes en ce qui concerne les propriétés physiques et, en conséquence, la nature des cibles de la surface terrestre.

La première couverture systématique de la surface de la terre fut entreprise par le satellite américain LANDSAT-1, lancé le 23 juillet 1972, a cessé son activité en janvier 1978. Le deuxième satellite LANDSAT-2 poursuivant les mêmes objectifs, fut lancé le 22 janvier 1975 .

Le troisième fut lancé en Mars 1978 a des difficultés de fonctionnement (au point de vue capteurs). LANDSAT-IV est prévu pour la fin 1981.

Le satellite Landsat est placé sur une orbite circulaire quasi - polaire, à une altitude de vol de 920Km environ, ce qui permet l'enregistrement à une échelle à peu près constante. Sa période de révolution est de 103 mn, soit 14 fois le tour du globe en 24 heures. Il assure l'enregistrement complet de la surface du globe, à l'exception des zones polaires, tous les 18 jours. ainsi tout point de la surface terrestres (sauf les régions polaires) est survolé 20 fois par an à la même heure locale .

Le système de détection utilisé est le MSS (Multispectral Scanner : subsystem).

Le MSS est un instrument electro-optico-mécanique, constitué d'un radiomètre multispectral à balayage opérant dans 4 canaux correspondants au visible et le proche I.R. (4,5,6, et 7).

Canal	longueur d'onde(nm)	Bande spectrale
4	500 à 600	vert
5	600 à 700	jaune-orangé
6	700 à 800	Rouge-proche IR
7	800 à 1100	proche IR-IR

Chaque canal comporte 6 détecteurs qui ne sont pas strictement identiques, on peut remarquer que :

Le violet et le bleu ne sont pas enregistrés .

Les Canaux 6 et 7 du radiometre à balayage sont dans le proche I.R., en dehors du visible: ils nous donnent des images de la vue dans une lumière que notre oeil ne voit pas .

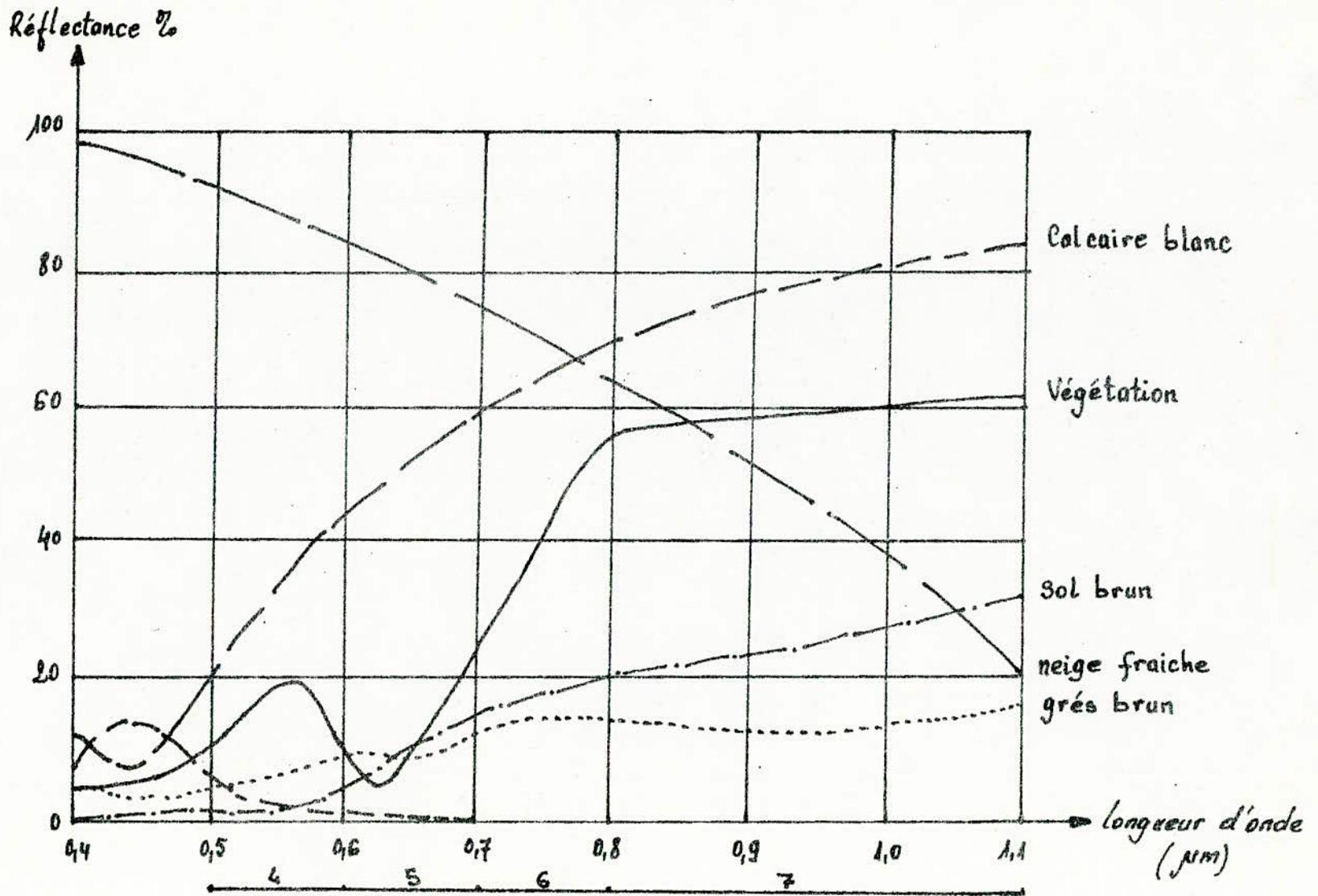
Canal MSS 4 : Donne un meilleur contraste des zones urbaines et des constructions par rapport à l'environnement , fait apparaître les fonds faiblement immergés et délimite les masses d'eau.

Canal MSS 5 : Révèle le couvert végétal, donne les traits caractéristiques du substratum geologique, du réseau hydrologique, des formes du relief. Les parcelles des terres agricoles et les voies de communication dans cette bande spectrale y sont le mieux représentées.

Canal MSS 6 : Est intermédiaire entre les canaux MSS5 et MSS7 notamment en ce qui concerne le réseau hydrographique.

Canal MSS7 : Permet de délimiter toutes les masses d'eau qui apparaissent en noir en raison de la forte absorption du rayonnement I.R. Les sols saturés en eau y sont bien discernables.

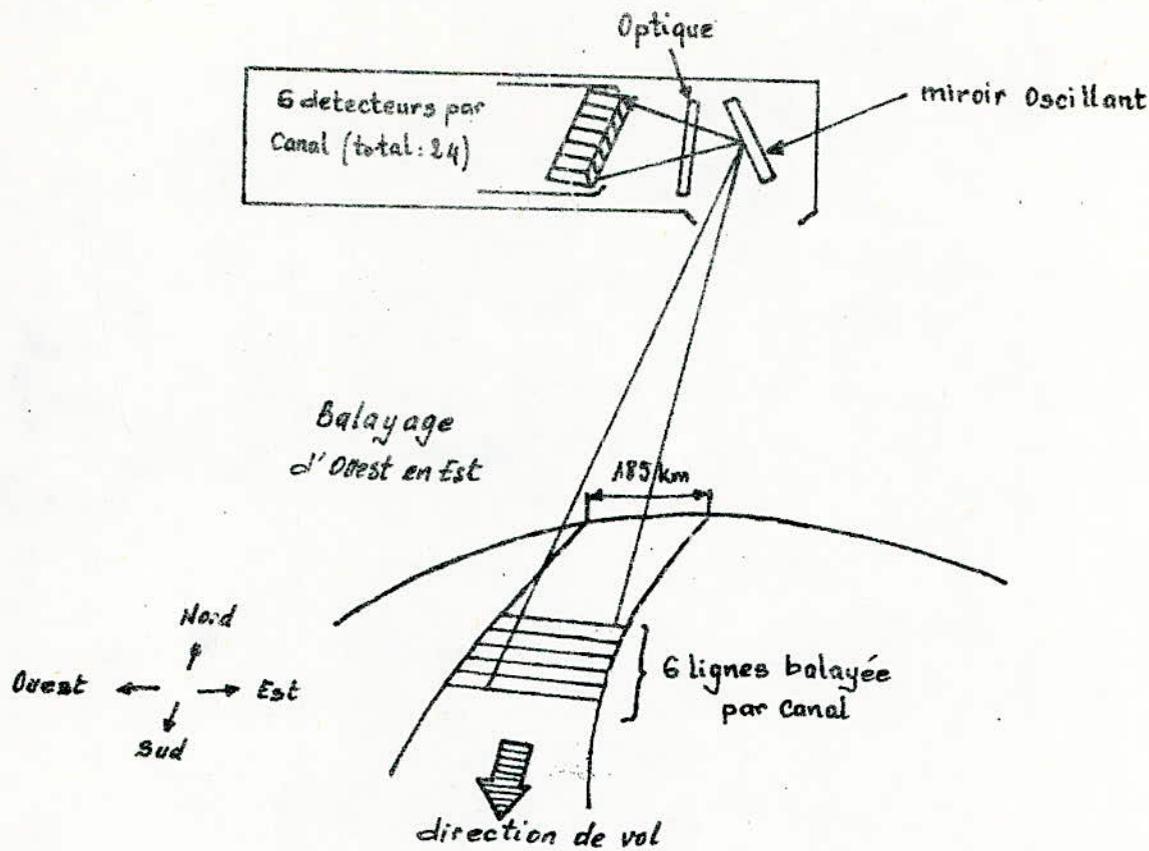
Ce canal donne la meilleure représentation du système hydrographique et du relief disséqué des zones montagneuses .



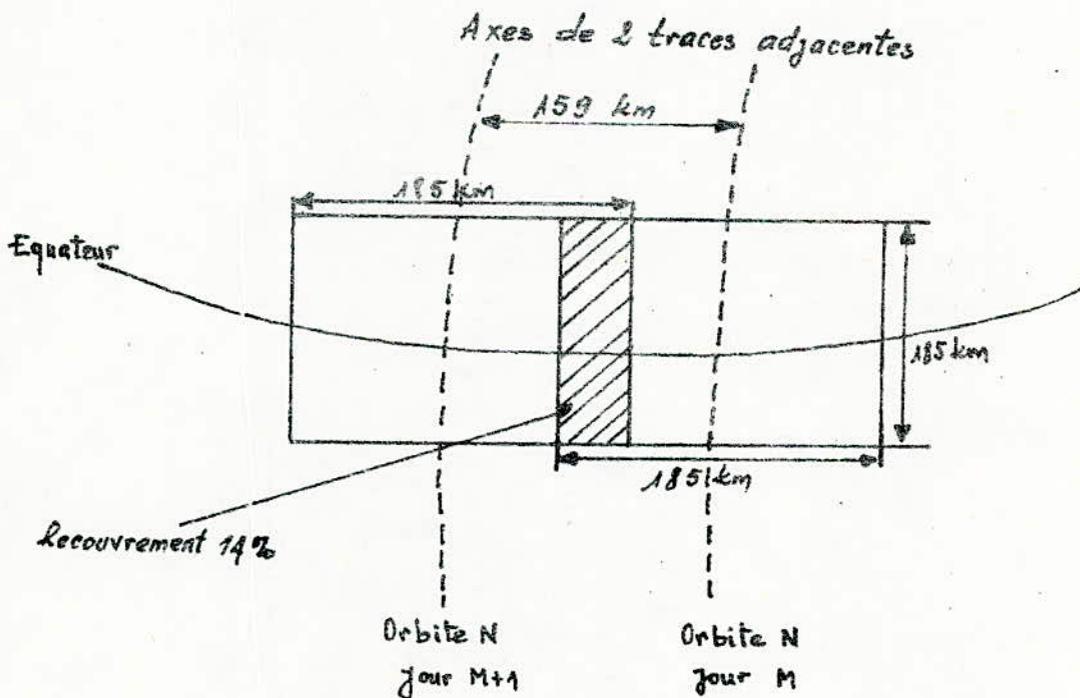
Courbes schématiques de réflectance pour différents types de Couverture du Sol.

A titre indicatif , nous donnons le tableau suivant :

λ (μm)	Spectre		DAEDALUS	λ (μm)	LANDSAT-1	λ (μm)
0,3-	U.V					- 0,3
0,4		0,38				
	VIOLET	0,42	1			
		0,45	2			0,4
0,5	BLEU	0,5	3	0,5		0,5
	VERT	0,55	4		1	
0,57		0,6				
0,6	JAUNE	0,6	5	0,6		0,6
0,65	ORANGE	0,65	6		2	
0,7	ROUGE		7			
0,75		0,7		0,7		0,7
			8		3	
0,8-	I.R.	0,8		0,8		0,8
		0,89	9			0,9
0,9-		0,92				
1-			10		4	1
1,1-		1,1				1,1



Principe de radiomètre multispectral à balayage (MSS) du Satellite LANDSAT



Recouvrement latéral des traces d'orbites à l'équateur

Ce radiomètre fournit la presque totalité des observations de la terre transmises par les satellites LANDSAT .

Le principe de ce radiomètre spectral est le balayage de la surface terrestre le long de la trace de la trajectoire du satellite suivant une bande continue de 185 Km de long.

Le balayage s'effectue par par oscillation d'un miroir qui envoie sur un système optique un faisceau lumineux. Un prisme, compris dans le système optique, décompose le faisceau lumineux en 4 faisceau de longueurs d'ondes différentes (canaux 4, 5, 6 et 7). Chaque faisceau est reçu par un ensemble de 6 détecteurs qui sont des cellules photosensibles distinctes pour chaque type de radiations en développant une tension proportionnelle à l'énergie reçue. La mesure de cette tension donne la grandeur de l'énergie des composantes spectrales .

Chaque ligne balayée est analysée point par point d'une manière discrète.

Chaque signal correspond à l'information spectrale d'une surface élémentaire de 79m de côté, ces surfaces se chevauchant sur la ligne balayée avec un pas de 57m de côté correspondant à la résolution spatiale des images du satellite LANDSAT . Ces images couvrent une superficie de $185 \times 185 \text{ Km}^2$ (34225 Km^2) et est formée de 2340 lignes, dont chacune est composée de 3240 unités élémentaires d'information (ou pixels) au dimension de $79 \times 57 \text{ m}^2$ soit 4500 m^2 environ. La Scène entière contient donc 7581 600 pixels .

L'énergie émise par chaque détecteur est enregistrée sur un support magnétique.

Pour rendre cette information utilisable par l'ordinateur, il faut la numériser. Cette numérisation se fait en 2 opérations.

- L'échantillonnage qui consiste à représenter un carré élémentaire par un seul nombre .
- La quantification qui arrondit à une valeur autorisée dans une échelle discrète de valeurs .

Théoriquement nous avons 256 valeurs possibles ou "niveaux de gris " qu'on peut coder sur 8 bits .

Après amplifications des signaux analogiques et transformations en signaux numériques par un codeur, ces derniers sont :

- Soit transmis directement au sol, si le satellite est dans la zone de visibilité d'une station de réception, où ils sont enregistrés sur bandes magnétiques .
- Soit conservés en mémoire pour être transmis au sol ultérieurement .

Les enregistrements ainsi obtenus sont du type multispectral.

C'est -à-dire fournis simultanément dans 4 bandes spectrales différentes; ils permettent donc une meilleure différenciation des objets de signatures spectrales voisines .

Le flot des données correspondant à une partie d'orbite est découpé en une succession de fichiers magnétiques correspondant à 2340 lignes du radiomètre multispectral. Toutes les images ne sont qu'une reproduction visuelle d'une partie des valeurs numériques contenues sur support magnétique. Il est possible d'acheter une copie sur bande magnétique du fichier qui a servi à établir les premières images d'une vue. C'est sous cette forme que l'on obtient l'information la plus complète et la plus précise .

Les valeurs numériques pour chaque point varient entre 0 et 63 pour le canal 7 et entre 0 et 127 pour les autres canaux.

Ces données sont présentées à l'utilisateur sous forme de bande magnétiques . Les utilisations des données des satellites LANDSAT sont très variées .

La répétitivité des prises de vues et la possibilité d'avoir des images de plusieurs dizaines de milliers de Km², sont deux des caractéristiques originales des données LANDSAT .

Nous rappelons ici brièvement les caractéristiques essentielles des données .

- Une vue se compose de 7581 600 pixels : 2340 lignes et 3240 colonnes
- . Chaque point a une dimension au sol de 57 x 79 m² , soit 4500 m²
- . la prise de vue n'est pas instantannée; elle dure 27,6 sec .
- Pour chaque vue, on dispose de quatre images prises sur des canaux de longueurs d'onde différentes (deux dans le visible, deux dans l'infrarouge), on a donc près de 30 millions d'éléments d'information .

La répétitivité théorique des vues est de 9 jours avec les deux satellites; en pratique elle est au moins d'un mois.

- Pour chaque zone, la vue est prise à la même heure locale, le soleil est donc toujours au même azimut .

La masse considérable de données ainsi obtenue est cependant redondante et pose dès lors le problème du traitement de l'information numérique .

Devant cette situation , il est nécessaire d'automatiser tout ou une partie de l'analyse des images.

Les problèmes à résoudre sont :

- Compression selective des données de télédétection. Il faut compresser pour extraire la partie utile à l'application envisagée, mais il faut aussi compresser pour alléger le volume des données .
- Décorrélation des canaux afin de supprimer la redondance de l'information dans les différents canaux .

La compression porte ici sur le nombre de bandes spectrales prises en compte .

Une partie de ce choix est souvent déjà faite lors de la prise de vue. Des études très importantes doivent être alors réalisées pour optimiser ce choix en fonction des missions à assurer.

Dans certaines conditions, il peut être intéressant de créer artificiellement de nouveaux canaux (combinaisons linéaires, quotient, ou d'autres combinaisons encore ,.....) mais leur définition à priori basée sur la connaissance des phénomènes est difficile.

Dans ce cas, on construira les n canaux de façon optimisée vis-à-vis d'un critère donné (maximum d'information conservée).

C'est en particulier de cette façon que l'on établit les composantes principales d'une image .

II - ANALYSE EN COMPOSANTES PRINCIPALES

II- 1- Généralités sur les transformations linéaires orthogonales :

Soient E un espace vectoriel de dimension n sur un corps
- K .

Soient B et B' deux bases distinctes de E

$$B : (e_1, \dots e_n)$$

$$B' : (e'_1, \dots e'_n)$$

Un vecteur X_i de E aura pour composantes $(X_{i1}, \dots X_{in})$ dans la base B et $(X'_{i1}, \dots X'_{in})$ dans la base B'

On peut définir une matrice M de passage de la base B à la base B' .

$$\text{On aura : } X'_i = M X_i \quad M \text{ étant une matrice } n \times n$$

Si l'on considère que l'information portée par les X_i dans B est donnée par leurs inerties, alors pour retrouver la même quantité d'information dans la nouvelle base B' il faut que cette transformation linéaire conserve angles et distances entre les vecteurs .

Si la transformation est orthogonale (c'est -à-dire $M^T = M^{-1}$), elle aura des propriétés caractéristiques :

$$\text{Soient } X'_1 = M X_1$$

$$X'_2 = M X_2$$

$$\vec{X}'_1 \cdot \vec{X}'_2 = X'^T_1 \cdot X'_2 = (M X_1)^T (M X_2) = X^T_1 M^T M X_2 = X^T_1 X_2 = \vec{X}_1 \cdot \vec{X}_2$$

$$\|\vec{X}'_1\|^2 = X'^T_1 \cdot X'_1 = (M X_1)^T \cdot (M X_1) = X^T_1 \cdot M^T \cdot M \cdot X_1 = X^T_1 \cdot X_1 = \|\vec{X}_1\|^2$$

$$\text{Si } \theta = \text{angle entre } \vec{X}_1 \text{ et } \vec{X}_2$$

$$\theta' = \text{angle entre } \vec{X}'_1 \text{ et } \vec{X}'_2$$

$$\text{alors} \quad \cos \theta' = \frac{\vec{x}_1' \cdot \vec{x}_2'}{\|\vec{x}_1'\| \cdot \|\vec{x}_2'\|} = \frac{\vec{x}_1 \cdot \vec{x}_2}{\|\vec{x}_1\| \cdot \|\vec{x}_2\|} = \cos \theta$$

Les transformations linéaires orthogonales conservent donc angles et distances entre les vecteurs . On retrouve donc la même information .

Il existe, bien sur, autant de transformations orthogonales que de bases orthogonales dans R^n , mais l'intérêt est de trouver une transformation qui nous permette une compression de données .

II-2- Transformation de Karhunen-Loeve ou Analyse en composantes principales :

II-2-1. Position du problème :

On dispose d'une population E constituée de N individus et on a P observations de chacun individu .

On aura donc un tableau de valeurs numériques ayant N lignes et P colonnes, chaque ligne représente un individu de E sur lequel les P mesures ont été effectuées . La valeur dans le tableau se trouvant à l'intersection de la $i^{\text{ème}}$ ligne et de la $j^{\text{ème}}$ colonne représentera la valeur prise par la $j^{\text{ème}}$ observation sur le $i^{\text{ème}}$ individu .

Ce tableau véhicule de l'information qui peut être considérée comme l'inertie de ses éléments .

Le nombre d'individus étant très important, on aura un tableau de grande dimension. On cherchera, pour des facilités de traitement (humain ou par ordinateur) à réduire le nombre de variables .

Il faudra donc trouver un nouveau tableau de plus faible dimension telle que l'information véhiculée par ce dernier soit la plus

.../...

proche possible de l'information du tableau initial.

Si on munit le tableau initial de la distance euclidienne alors on pourra l'identifier à un espace de R^P , et ses éléments à des points de cet espace. Si on fait ensuite une représentation dans cet espace, on obtiendra un nuage de points ayant la forme d'un ballon de rugby.

On cherche à visualiser seulement les directions d'allongement dominantes du nuage en vue d'une interprétation plus aisée, mais dans ce cas on perd de l'information. On doit pouvoir évaluer cette perte.

On cherchera donc une représentation de ce nuage de points dans un sous-espace R^k ($k < p$) de telle sorte que l'information véhiculée par cette dernière soit la plus proche possible de l'information véhiculée par le tableau initial. *initial d'information véhiculée par les lignes du tableau* sous-espace de R^p peut être considérée comme la somme des carrés des distances entre les lignes (la distance dans R^p étant la distance euclidienne classique).

Cette information représente en quelque sorte l'inertie du nuage.

II₂ - Démonstration :

Soit $I =$ ensemble des individus, un individu est représenté par une ligne du tableau R :

$$R = \begin{bmatrix} x_1^1 & \dots & x_i^1 & \dots & x_p^1 \\ \vdots & & \vdots & & \vdots \\ x_1 & \dots & x_i & \dots & x_p \\ \vdots & & \vdots & & \vdots \\ x_1^n & \dots & x_i^n & \dots & x_p^n \end{bmatrix} \rightarrow \begin{matrix} j^{\circ} = \text{individu } X^j \\ \\ \\ \\ i^{\circ} \text{ observation } X_j \end{matrix}$$

L'inertie de I :
$$J = \sum_{i=1}^n m_i \cdot d^2(X^i, G)$$

X^i : représente le i° individu

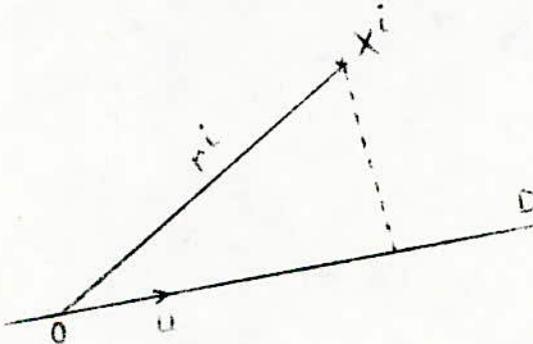
G : centre de gravité

m_i : poids du i° individu

Si on centre et on réduit les données, l'expression de l'inertie sera plus simple, la distance considérée étant la distance euclidienne classique. Les individus auront un même poids que l'on prend égal à un.

donc
$$J = \sum_{i=1}^n \|X^i\|^2$$

En projetant sur un axe D



U étant un vecteur unitaire sur D

$$U^T \cdot U = 1.$$

$$U = \begin{bmatrix} U_1 \\ \vdots \\ U_p \end{bmatrix}; \quad r^i = \begin{bmatrix} x_1^i \\ \vdots \\ x_p^i \end{bmatrix}$$

On remarque que : $U^T \cdot r^i = r^i \cdot U$

$$J(D) = \sum_{i=1}^N (U^T \cdot r_i)^2$$

Pour minimiser les pertes d'information, on doit maximiser $J(D)$.

$$J(D) = \sum_{i=1}^N (U^T \cdot r_i)^2 = \sum_{i=1}^N (U^T \cdot r_i)(r_i^T \cdot U) = U^T \cdot \left(\sum_{i=1}^N r_i \cdot r_i^T \right) \cdot U$$

$$J(D) = U^T \cdot R^T \cdot R \cdot U$$

Donc le problème revient à maximiser $U^T R^T R U$ avec la contrainte $U^T \cdot U = 1$

D'autre part, on définit la matrice de covariance des observations Q ayant pour terme général:

$$\text{COV} (X_i, X_j) = \frac{1}{N} \sum_{k=1}^N (x_i^k - \bar{x}_i) (x_j^k - \bar{x}_j)$$

On a par hypothèses des variables centrées, la moyenne est donc nulle

$$\text{d'où : } \text{COV} (X_i, X_j) = \frac{1}{N} \sum_{k=1}^n x_i^k \cdot x_j^k$$

Le terme général de la matrice $R^T \cdot R$ (i° ligne, j° colonne) est:

$$x_i^1 \cdot x_j^1 + x_i^2 \cdot x_j^2 + \dots + x_i^n \cdot x_j^n = \sum_{k=1}^n x_i^k \cdot x_j^k$$

On remarque bien que $R^T \cdot R$ est proportionnelle à Q , donc elles admettent les mêmes vecteurs propres et les mêmes valeurs propres. Q étant la matrice de covariance donc c'est une matrice symétrique, définie positive ; elle admet des vecteurs propres orthogonaux et unitaires et des valeurs propres associées réelles et positives.

Si on note par V : la matrice contenant les vecteurs propres de $R \cdot R^T$ on a bien $V \cdot V^T = I$ I étant la matrice identité.

Soit Λ la matrice diagonale contenant les valeurs propres de $R^T \cdot R$

$$\Lambda = \begin{array}{|c|} \hline \lambda_1 & & & \\ & \ddots & & \\ & & \lambda_p & \\ & & & 0 \\ \hline \end{array}$$

aux $\lambda_1 \dots \lambda_p$ valeurs propres correspondent respectivement les $v_1 \dots v_p$ vecteurs propres tel que :

$$v_i = \begin{bmatrix} v_{i1} \\ \vdots \\ v_{ip} \end{bmatrix} ; V = \begin{bmatrix} v_{11} & \dots & v_{p1} \\ \vdots & \ddots & \vdots \\ v_{1p} & \dots & v_{pp} \end{bmatrix} \quad \text{avec } V \cdot V^T = I$$

Par definition on a :

$$R^T \cdot R v_1 = \lambda_1 v_1$$

$$R^T \cdot R v_p = \lambda_p v_p$$

en ecriture matricielle on aura .

$$R^T \cdot R = \Lambda \quad V = V \Lambda$$

En multipliant à droite par V^T l'expression $R^T R V = V \Lambda$

$$\text{On obtient : } R^T \cdot R \underbrace{V \cdot V^T}_I = V \Lambda V^T$$

$$R^T \cdot R = V \Lambda V^T$$

$$\Rightarrow U^T \cdot R^T \cdot R \cdot U = U^T \cdot V \Lambda V^T \cdot U = (V^T \cdot U)^T \Lambda (V^T \cdot U)$$

$$\text{posons } y = V^T \cdot U \quad \Rightarrow \quad y^T = U^T \cdot V$$

avec ce changement on peut écrire :

$$U^T \cdot R^T \cdot R \cdot U = y^T \Lambda y$$

La contrainte $U^T \cdot U = 1$ devient :

$$U^T \cdot \underbrace{V \cdot V^T}_I \cdot U = (V^T \cdot U)^T \cdot (V^T \cdot U) = y^T \cdot y = 1$$

Le problème se simplifie et revient ainsi à maximiser $y^T \Lambda y$ avec la contrainte $y^T \cdot y = 1$.

$$y^T \Lambda y = \sum_{i=1}^p \lambda_i y_i^2 \quad \text{avec } y^T \cdot y = \sum_{i=1}^p y_i^2 = 1$$

Si λ_{i_1} correspond à la plus grande valeur propre de $R^T \cdot R$
 maximiser $y^T y$ revient à prendre $y_{i_1} = 1$ et $y_j = 0 \quad \forall j \neq i_1$

$$\text{Or } y = V^T U \quad \Rightarrow \quad u = V \cdot y.$$

Donc : le vecteur u cherché qui rende $U^T \cdot R^T \cdot R \cdot U$ maximum
 est le vecteur propre correspondant à la plus grande
 valeur propre de $R^T \cdot R$. C'est le vecteur unitaire porté
 par le premier axe d'inertie maximum (axe principal).

Pour trouver le deuxième axe d'inertie maximum, on doit chercher
 un vecteur unitaire t orthogonal à U tel que $t^T \cdot R^T \cdot R \cdot t$ soit
 maximum

De la même manière

$$\text{On pose } s = V^T \cdot t, \quad s^T = t^T \cdot V$$

On aura à maximiser $s^T \Lambda s$ avec la contrainte $s^T \cdot s = 1$

$$s^T \Lambda s = \sum_{\substack{i=1 \\ i \neq i_1}}^P \lambda_i s_i^2, \quad s^T s = \sum_{i=1}^P s_i^2 = 1$$

De même le vecteur unitaire t sera le vecteur propre corres-
 pondant à la deuxième plus grande valeur propre de $R^T \cdot R$ et sera
 porté par le deuxième axe principal.

Les axes d'inertie maximum suivants porteront les vecteurs
 propres correspondant aux valeurs propres suivantes et seront rangés
 par ordre décroissant.

Quant on a des variables centrées, on a vu que la matrice
 $R^T \cdot R$ est proportionnelle à la matrice de covariance. Généralement
 on a des variables quelconques, dans ce cas on fait un changement
 de variables, on remplace les termes x^k par $x_1^k - \bar{X}_1$ pour centrer
 le nuage de points.

ou $\bar{X} \mathbf{1} = \frac{1}{n} \sum_{k=1}^n x^k$ est la moyenne pour la l^o observation

alors $R = \begin{bmatrix} x'_1 - \bar{X}_1 & \dots & x'_p - \bar{X}_p \\ \vdots & & \vdots \\ x^n_1 - \bar{X}_1 & & x^n_p - \bar{X}_p \end{bmatrix}$

Le probleme est identique avec cette nouvelle matrice, le terme general de $R^T R$ etant

$$(x^1_i - \bar{X}_i)(x^1_j - \bar{X}_j) + \dots + (x^n_i - \bar{X}_i)(x^n_j - \bar{X}_j) = \sum_{k=1}^n (x^k_i - \bar{X}_i)(x^k_j - \bar{X}_j)$$

On retrouve bien l'expression de la covariance au terme n pres

$$\text{COV}(X_i, X_j) = \frac{1}{n} \sum_{k=1}^n (x^k_i - \bar{X}_i)(x^k_j - \bar{X}_j)$$

donc $R^T \cdot R$ et Q ont les m\^emes valeurs et vecteurs propres on aura aussi

$$\text{cor}(U_i, U_j) = \cos(U_i, U_j) = \begin{cases} 0 & \text{si } i \neq j \\ 1 & \text{si } i = j \end{cases}$$

En resum\^e :

Pour trouver les axes principaux, il suffira de chercher les vecteurs propres de la matrice de covariance et de les ranger par valeur propres associes decroissante .

Si l'on consid\^ere que la variance est representative de la quantite d'information presente, le premier axe principal contiendra le plus d'information et les autres de moins en moins .

Ainsi on a trouv\^e de nouvelles "pseudo-observations" decorrel\^ees avec une concentration de l'information au niveau des premi\^eres

" pseudo.observations " .

II - 2 - 3 Formalisation de l'Analyse en composants Principales.

On prend l'exemple d'un tableau à n individus et p caractères

1) calcul de la moyenne $\mu_1 \dots \mu_p$ pour chaque caractère

$$\mu_k = \frac{1}{n} \sum_{j=1}^n X_k(j) \quad k = 1, p$$

$X_k(j)$ étant la valeur du k^{e} caractère pour le j^{e} individu .

2) centrer et réduire les caractères s'ils ne sont pas homogènes.

$$\frac{X_k - \mu_k}{\sigma_k} \text{ avec } \sigma_k = \frac{1}{n} \sum_{j=1}^n (X_k(j) - \mu_k)^2$$

3) calcul des covariance entre caractères .

$$\text{Cov}(k, l) = \frac{1}{n} \sum_{j=1}^n (X_k(j) - \mu_k) (X_l(j) - \mu_l).$$

nous obtenons la matrice de covariance Q .

nous rappelons que $\text{Cov}(X, X) = \text{Var}(X)$.

$$Q = \begin{bmatrix} \text{Var}(X_1) & \text{cov}(X_1, X_2) & \dots & \text{cov}(X_1, X_p) \\ \text{cov}(X_2, X_1) & & & \vdots \\ \vdots & & & \vdots \\ \text{Cov}(X_p, X_1) & \dots & \dots & \text{Var}(X_p) \end{bmatrix}$$

Remarque : si les caractères sont centrés réduits leurs variances égale à l'unité .

$$\text{cov}(X, Y) : \text{cor}(X, Y) \cdot \sqrt{\text{Var}(X)} \cdot \sqrt{\text{Var}(Y)} = \text{cor}(X, Y)$$

on obtient la matrice de corrélation .

Q étant symétrique définie positive, ses valeurs propres sont réelles positives et correspondent à l'information contenue dans les nouveaux caractères. Leur somme est égale à la trace de la matrice Q. La quantité d'information est retrouvée

$$\text{on a bien } \text{tr}(Q) = \sum_{i=1}^p \text{Var}(X_i) = \text{Information totale}$$

4. calcul des vecteurs et des valeurs propres de la matrice Q.

et on ordonne par valeurs propres associées décroissantes.

Pour obtenir les caractères transformés :

$$\text{on fait : } X' = M \cdot X$$

M étant la matrice $p \times p$ ayant pour lignes les vecteurs propres de Q rangés par valeur propres associées décroissantes.

Les vecteurs étant orthogonaux et unitaires on a bien $M^t \cdot M = I$

En prenant les premiers caractères transformés, on garde le maximum d'information, et ces caractères sont de plus totalement décorrelés.

II 2 - 4 Conclusion .

Etant donné un ensemble de variables mesurées sur un ensemble d'individus, on se propose de réduire l'information (en général redondante) donnée par l'ensemble des variables, de telle sorte que seules quelques combinaisons linéaires des variables initiales, indépendantes entre elles, appelées facteurs, permettent de caractériser les individus.

Ainsi la première composante principale relativement à P variables correspondra à la direction de R^P suivant laquelle les individus seront les mieux différenciés, et la méthode nous donne successivement les premières, deuxièmes, - - - $p^{\text{ième}}$ combinaisons linéaires des variables initiales dont les variances vont ainsi en décroissant.

Ces composantes principales sont précisément les variables transformées et leurs variances sont les valeurs propres de la matrice de covariance .

La première composante principale explique la plus grande partie possible de la variance totale, alors qu'inversement la dernière composante principale aura la variance la plus faible possible; ainsi en analyse en composantes principales, on résumera souvent l'ensemble des variables par les deux ou trois premières composantes principales .

On remarque d'ailleurs que cette analyse en composantes principales permet un certain regroupement de l'information en minimisant les pertes .

Du fait de sa généralité et de la faiblesses des hypothèses qu'elle nécessite, l'analyse en composantes principales présente un grand intérêt pour analyser des données quantitatives et ce dans tous les domaines notamment les sciences humaines, médicales, économiques , les sondages, - -

Avec l'avènement de la télédétection, vu la masse considérable de données fournies en permanence par les satellites, il s'est avéré intéressant d'appliquer cette méthode en vue d'un prétraitement d'images pour accroître la rapidité des traitements souvent lents.

III-3- APPLICATIONS DE L'ACP

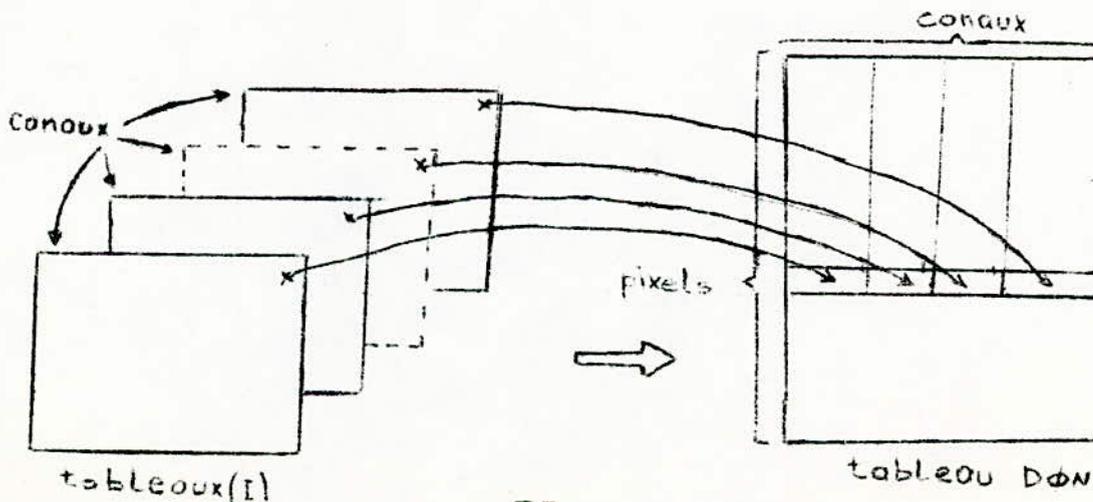
III - 3 - 1-Application à la télédétection :

Nous rappelons que les données fournies par le satellite sont sous forme d'images multispectrales .

Numériquement parlant, une image est composé de plusieurs pixels, chacun d'eux étant caractérisé par ses N niveaux de luminances mesurés dans les N canaux. On se trouve donc devant N images (une par canal), qui sont des tableaux de valeurs numériques à n lignes et n colonnes représentant l'organisation spatiale des points sur le terrain. On peut appeler ces tableaux des tableaux images (tableau (I)) .

Relativement aux données LANDSAT nous disposerons donc de 4 tableau (I).

Une première étape du traitement consiste à transformer cet ensemble de quatre tableaux(I) en un seul tableau $D \times N$ pour données. Cette "mise à plat " schématisée par la figure suivante, consiste à concaténer les éléments de chaque tableau en une colonne unique .



L'ordre de concaténation peut être quelconque: il suffit que ce soit le même pour chaque tableau (I)

Chaque colonne du tableau $D \emptyset N$ peut être identifiée à un canal

Chaque ligne du tableau $D \emptyset N$ peut être identifiée aux radiométries de chaque pixel mesurées dans les quatre canaux .

Finalement on obtient un tableau de chiffres représentant tous des radiométries, il n'est pas nécessaire de réduire ces données .

Ce tableau véhicule de l'information; on voudrait réduire les dimensions de ce tableau tout en conservant le maximum d'information

Nous avons appliqué pour ce faire notre programme .

On obtient alors un nouveau tableau véhiculant sensiblement la même information . Les colonnes ne sont plus interprétables comme des canaux; ce sont des combinaisons linéaires des colonnes du tableau initial .

Les " pseudo canaux " obtenus sont complément décorrelés .

Dans ce cas on passe d'un tableau à 4 dimensions à un tableau de 2 dimensions .

Ainsi il est possible de visualiser les pixels d'après leurs pseudo-radiométries en faisant une représentation dans le plan formé par les 2 pseudo-canaux .

(Sous-programme HISBID). Les pixels ayant des réponses spectrales voisines seront proches dans le plan de représentation . L'étape suivante consiste à regrouper les points voisins en classes (si on a une bonne différenciation entre les différentes classes, on dira qu'on a classé les pixels), puis à déterminer les bornes de de chaque classe, ensuite introduire ces valeurs dans le programme. On restitue l'image en faisant une impression d'un chiffre distinct
.../...

désignant chaque classe (sous -programme REIMA)

Ainsi on représente la région étudiée par une "image numérique

La projection des points sur un plan pourrait nous donner une fausse idée de la proximité de deux points dans R^4 . Mais en télédétection les deux pseudo-canaux sont choisis dans les directions du plus grand étalement du nuage et ils contiennent plus de 90 % de l'information initiale, donc on peut dire que l'on a une bonne différenciation dans le plan de ces deux axes .

Vu le grand nombre de pixels , alors le nombre minime de pixels mal représentés dû à leur projection ne pourrait altérer le choix des classes .

III-3-2 Application aux sondages :

Nous citons un exemple paru dans l'observateur de l'OCDE(n°50, 1971) qui publiait un ensemble de tableaux donnant un aperçu de la diversité des économies des 22 pays membres et permettant d'esquisser la silhouette économique de chacun .

L'auteur, après avoir procédé aux diverses opérations de dépouillement et de l'étude des corrélations, donnait le tableau codé des variables (caractéristiques) et des individus(pays) ainsi que le tableau reproduisant les données d'entrées que nous reprenons ici-(voir tableau I et II).

Donc nous avons appliqué notre programme à cet exemple .

Dans le cas de l'application de l'analyse en composantes principales à la télédétection , il n'était pas nécessaire de centrer et de réduire les variables car elles représentaient toutes des radiométries et les différents canaux avaient la même contribution.

Mais dans le cas des sondages , les réponses pour un individu peuvent différer d'unité ce qui entraînent que certaines caractéristiques sont plus favorisées que d'autres .Exemple Si on prend le caractère " population totale " en millions d'habitants pour d'autres certains individus (pays) et en milliers d'habitants pour d'autres , on n'aura pas les mêmes résultats.

C'est pour cela qu'il est nécessaire de centrer et de réduire les variables .

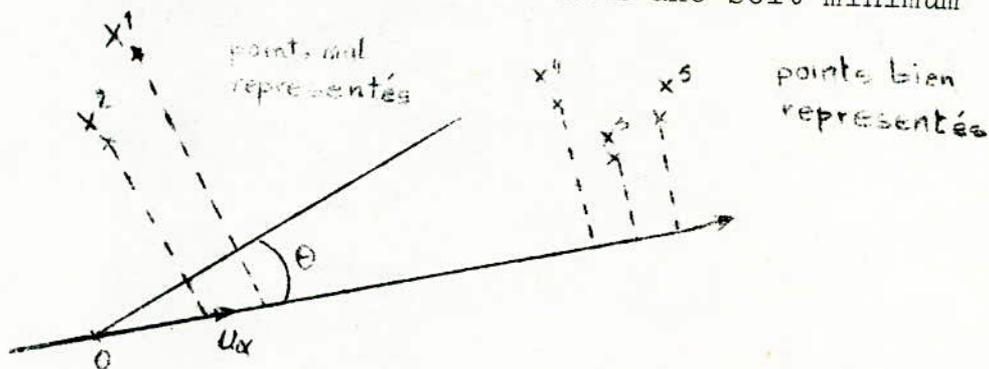
(le Sous-programme REDUCT permet de faire cette opération).

Une fois les nouvelles variables trouvées (projections des variables sur les nouveaux axes), le problème est de savoir si les individus sont bien représentés. Pour cela on fait appel aux aides à l'interprétation .

Soit U_α un axe qui correspond à une nouvelle variable .

Pour qu'un individu X^i soit bien représenté il faut que :

- sa projection sur l'axe soit maximum
- l'angle entre le vecteur X^i et l'axe soit minimum



Il faut avoir donc une idée de la qualité de la nouvelle représentation que l'on a trouvé .

Si θ = angle entre le vecteur X^i et l'axe U

$$\text{alors } \cos \theta = \frac{U_\alpha^T \cdot X^i}{\|X^i\|}$$

En général on prend le carré du cosinus que l'on note $CO2(i)$
 (pour le i^{eme} individu) avec :

$$CO2 (i) = \left(\frac{U^T \cdot X^i}{\|X^i\|} \right)^2$$

On considère que si $CO2 (i) > 0,5$ (angle $\theta < 45^\circ$) alors l'in-
 dividu i est bien représenté si sa projection sur l'axe est grande
 ($> \sqrt{\frac{1}{N}}$, N = nombre d'individus).

La somme des $CO2(i)$ sur les nouvelles variables nous renseigne
 ra sur la qualité de la représentation du i^{eme} individu . Plus cette
 somme est proche de 1 meilleure est la représentation de cet indivi-
 du .

Par ailleurs , en analysant les nouvelles variables trouvées
 (qui sont des combinaisons linéaires des anciennes variables), on
 peut les interpréter et voir leur contribution à l'information to-
 tale et les individus qu'elles opposent .

III PROGRAMMATION DE L'ANALYSE EN COMPOSANTES PRINCIPALES

Le programme que nous avons élaboré permet l'analyse aussi bien des données de télédétection que les données de sondages. Ces dernières sont fournies sur cartes, par contre les premières, vu leur quantité sont sur bande ou sur disque, leur lecture nécessite un programme propre .

Le programme de lecture sur bande que nous avons utilisé est celui que nous ^{avons} emprunté à nos camarades et que nous avons donné en annexe 1.

Le programme que nous avons élaboré se compose d'un programme principal et de plusieurs sous-programmes .

- Les dimensions affectées aux tableaux utilisés dans le programme que nous donnons ci-joint sont celles utilisées pour l'application aux sondages (matrice de données 20X20).

- L'option choisie est donnée par la valeur de $K\emptyset P$ ($K\emptyset P=1$ pour la télédétection, $K\emptyset P=0$ pour les sondages).

-Le taux d'information que l'on désire retrouver est donné par $TINF$; le nombre de caractères à retenir pour cela est donné par NCT .

Toutes ces indications ($K\emptyset P$, $TINF$) sont choisies au départ ainsi que le nombre de caractères (NC) et le nombre d'individus (NI).

Dans le cas de l'application de ce programme à la télédétection ou les données sont sur bandes et que ce sont des valeurs entières donc il faudra procéder aux changements suivants :

- . Changement des dimensions
- . Ajouter 2 cartes (INTEGER $D\emptyset N$, X): l'une au niveau du programme principal, l'autre au niveau de la SUBROUTINE CALMO, puis 2 autres cartes (INTEGER $D\emptyset N$): l'une au niveau de CALCOV, l'autre au niveau de KARLO.

Par ailleurs à la sortie du listing et après examen de l'histogramme bidimensionnel, on définit les bornes de chaque classe (choisie en fonction du regroupement des points) et quatre vecteurs de même dimension (égale au nombre de classes) tel que:

. KS1 (ou KI1) contient les valeurs des bornes supérieures (ou inférieures) dans le ~~deuxième~~^{premier} pseudo-canal.

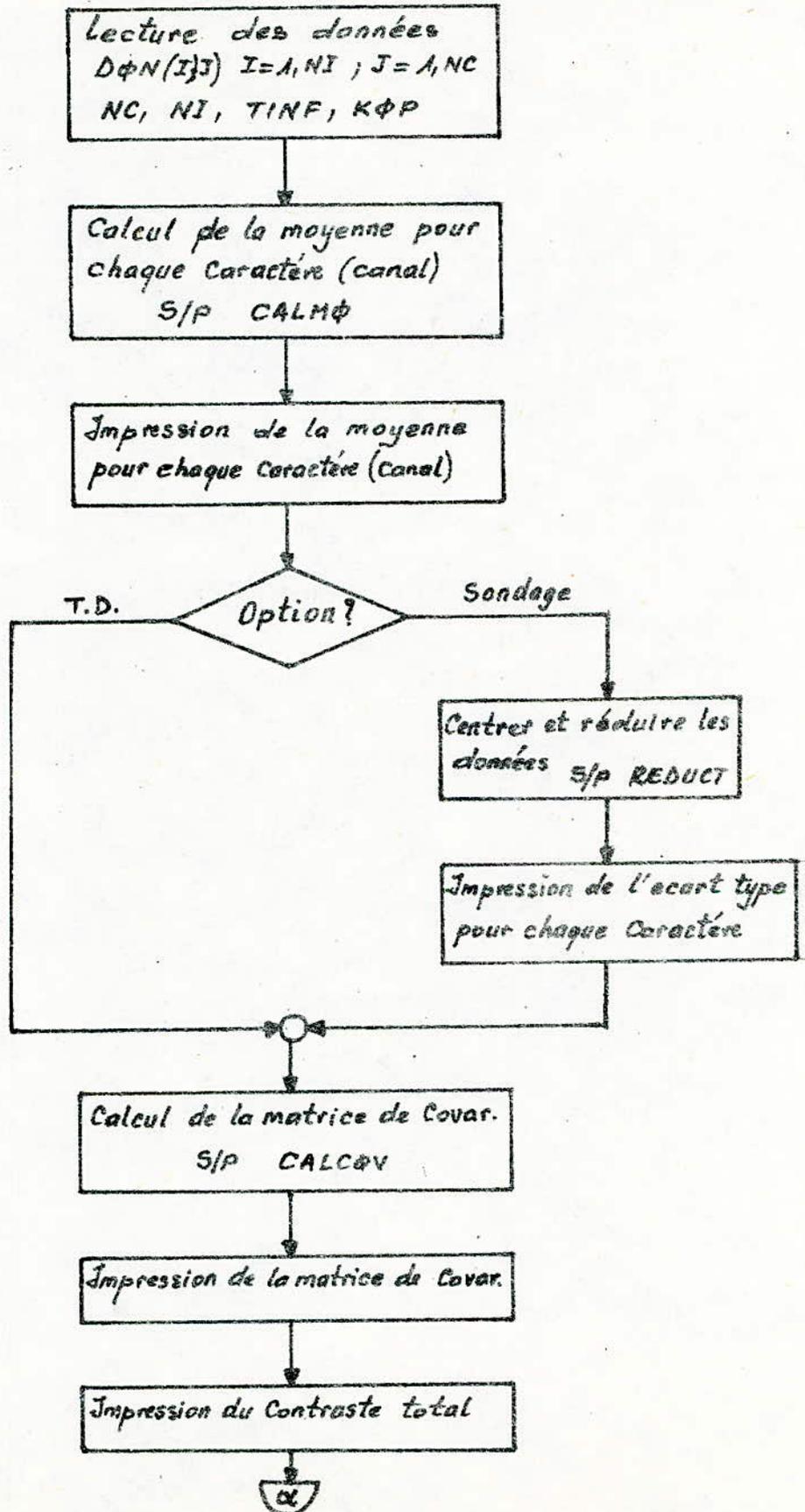
KS2 (ou KI2) contient les valeurs des bornes supérieures (ou inférieures) dans le deuxième pseudo-canal.

Ensuite, on introduit ces données par cartes et on refait un deuxième passage qui nous permettra de restituer une image numérique sur le listing .

Dans ce qui suit nous donnons l'organigramme général et les différents organigrammes propres aux différents sous-programmes.

Puis nous donnons le programme FORTRAN IV .

Organigramme Programme Principal



α

Calcul des vecteurs propres
et des valeurs propres
S/P MSPRO

Rangement des vecteurs propres
associés
par valeurs propres décroissantes
S/P RAP0D

Impression des vecteurs propres,
des valeurs propres et du
pourcentage d'information

Calcul du nombre de Caractères
(canaux) choisis (fonction du TINF)

Calcul des nouvelles Variables
S/P KARL0

Option?

T.D.

Sondage

NON

$NCT \neq 2$

OUI

tracé de
l'histogramme
bidimensionnel
S/P HISBID

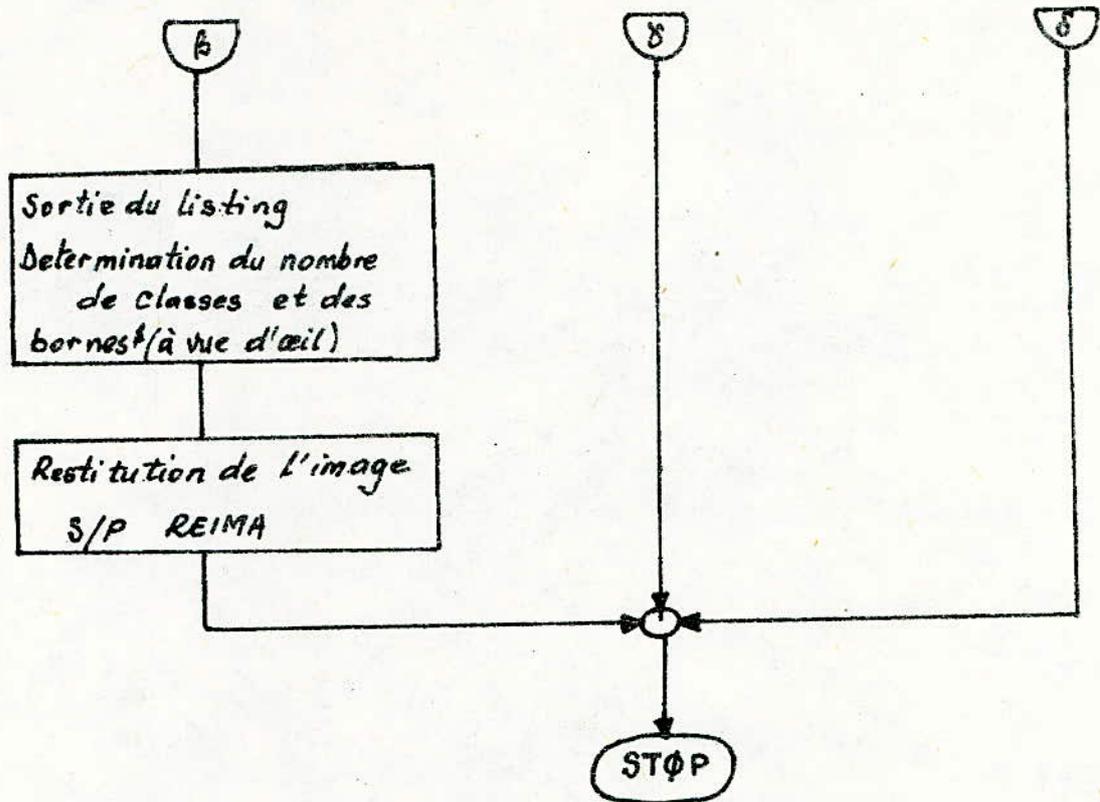
Calcul de la matrice $CO2(I, J)$, $I=1, NI$
 $J=1, NCT$
et de $\sum CO2(J)$, $J=1, NCT$
S/P S0NDE

Impression des nouvelles variables,
matrice $CO2$ et de $\sum CO2$

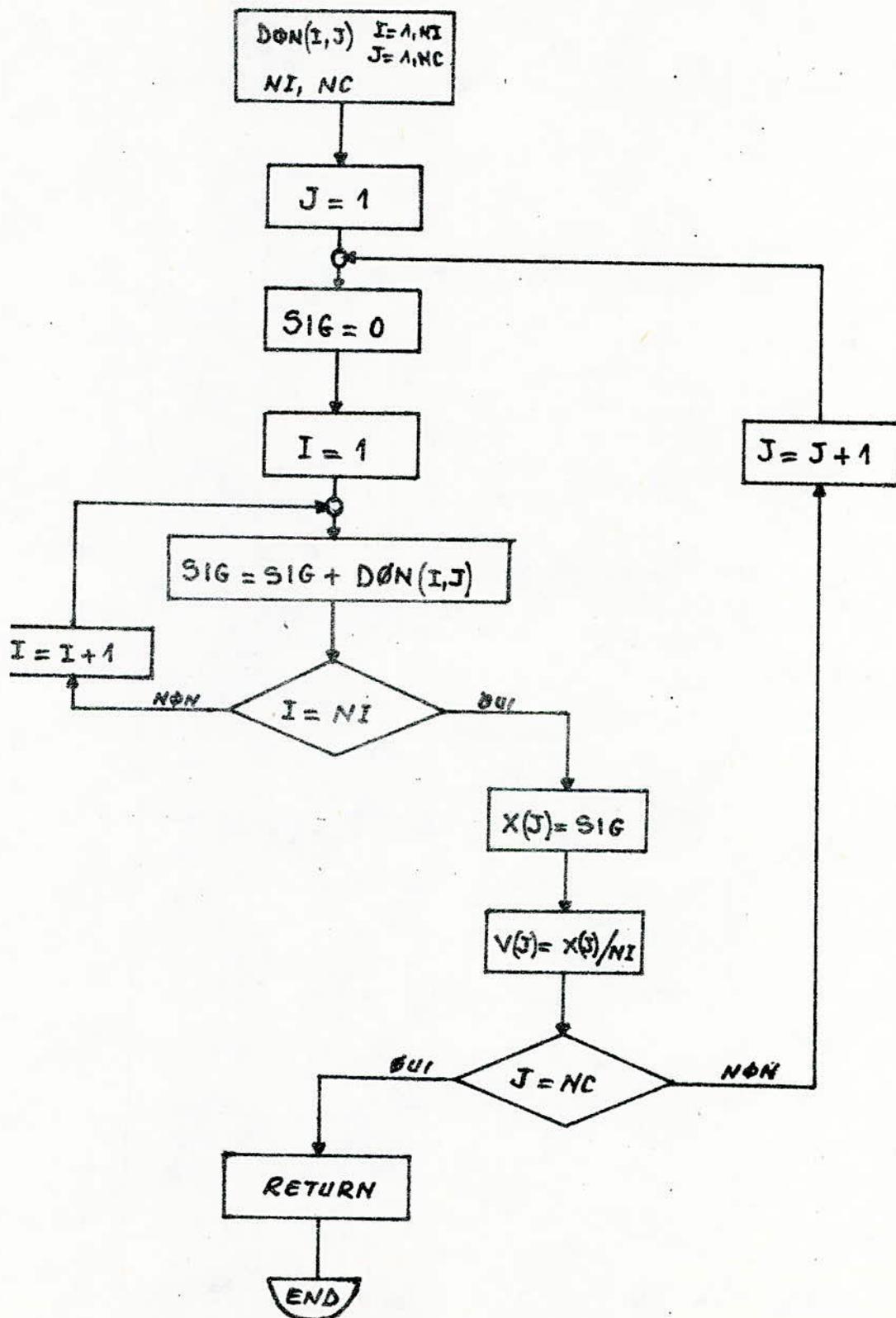
β

γ

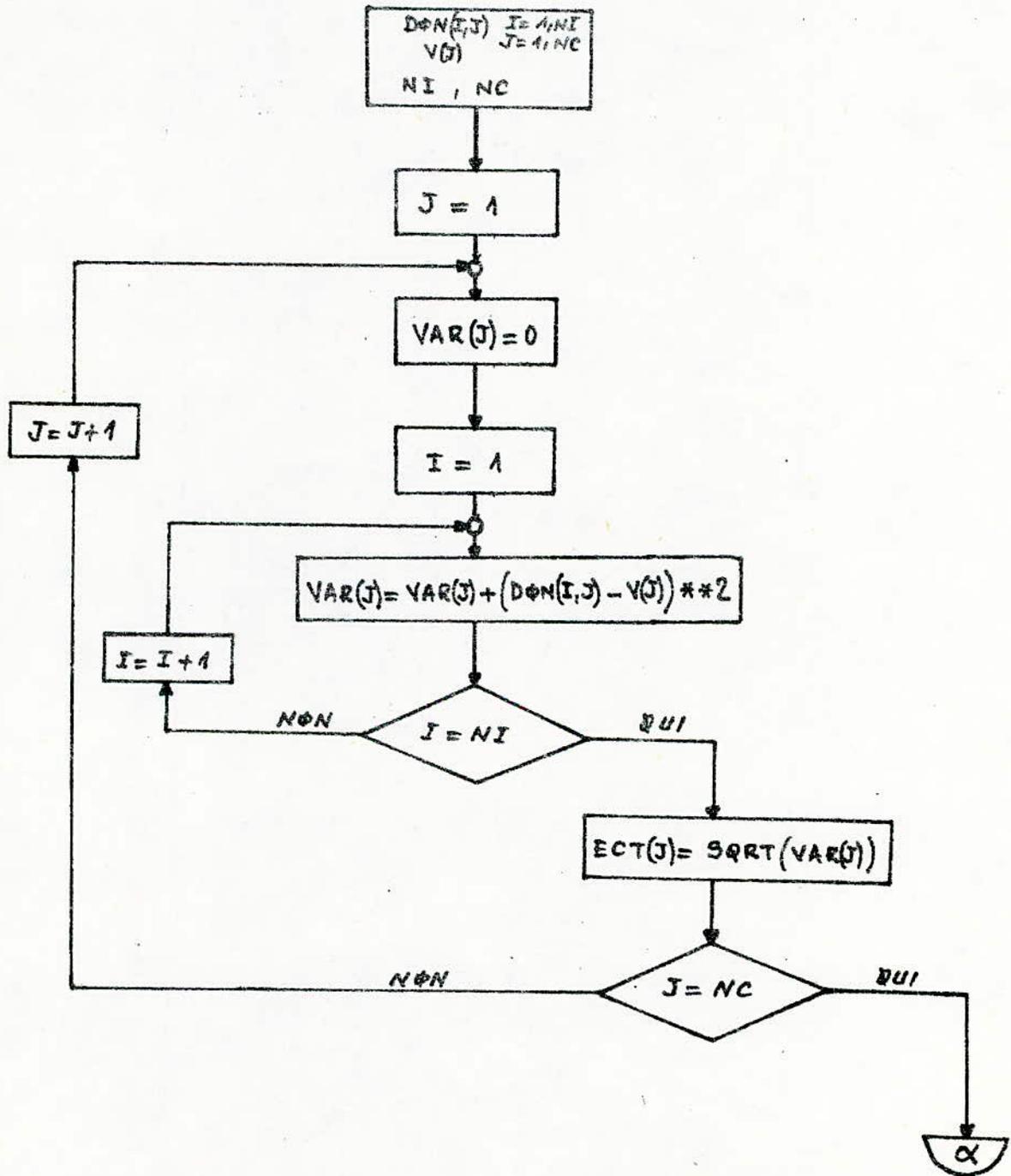
δ

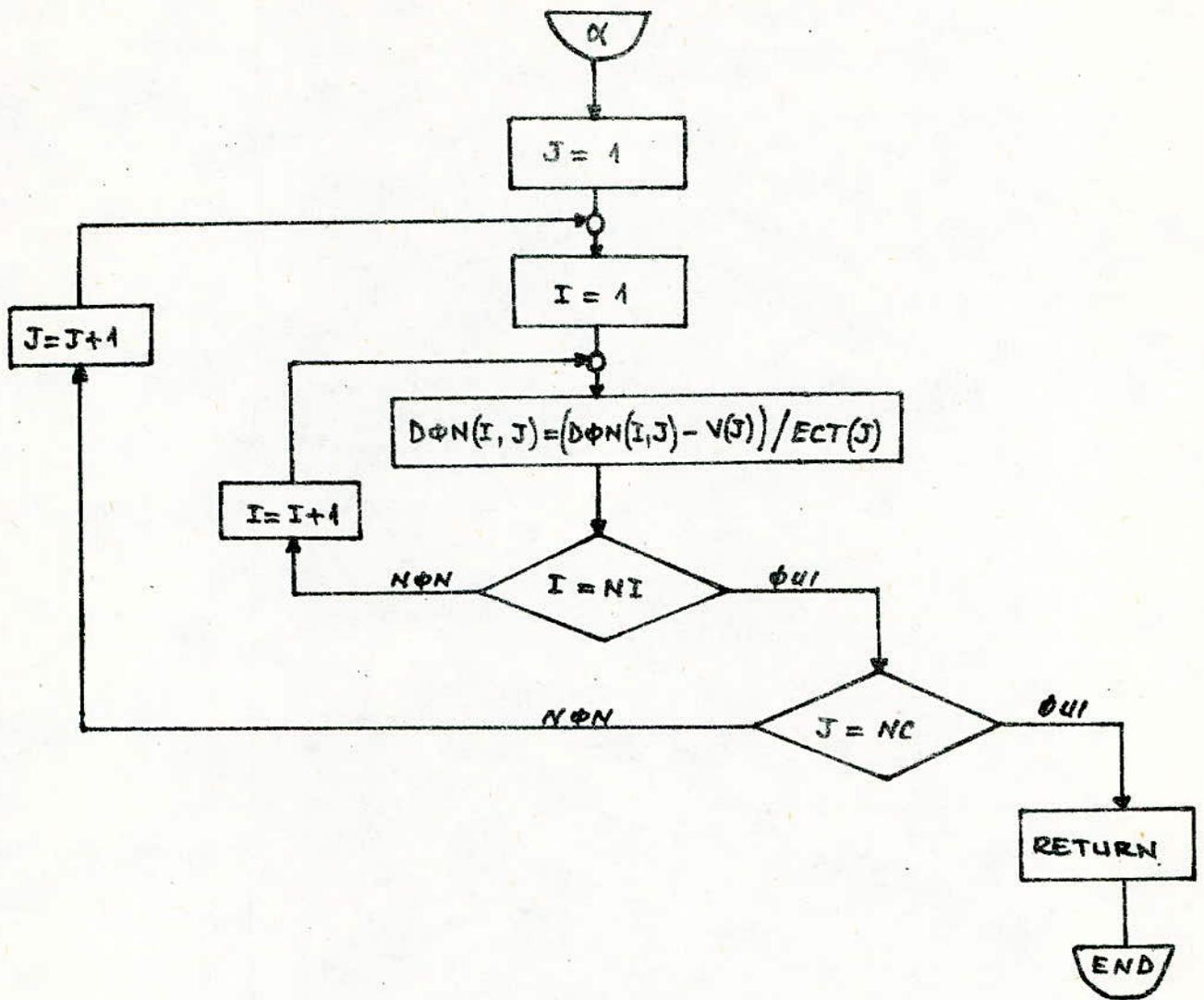


Organigramme Subroutine CALMA

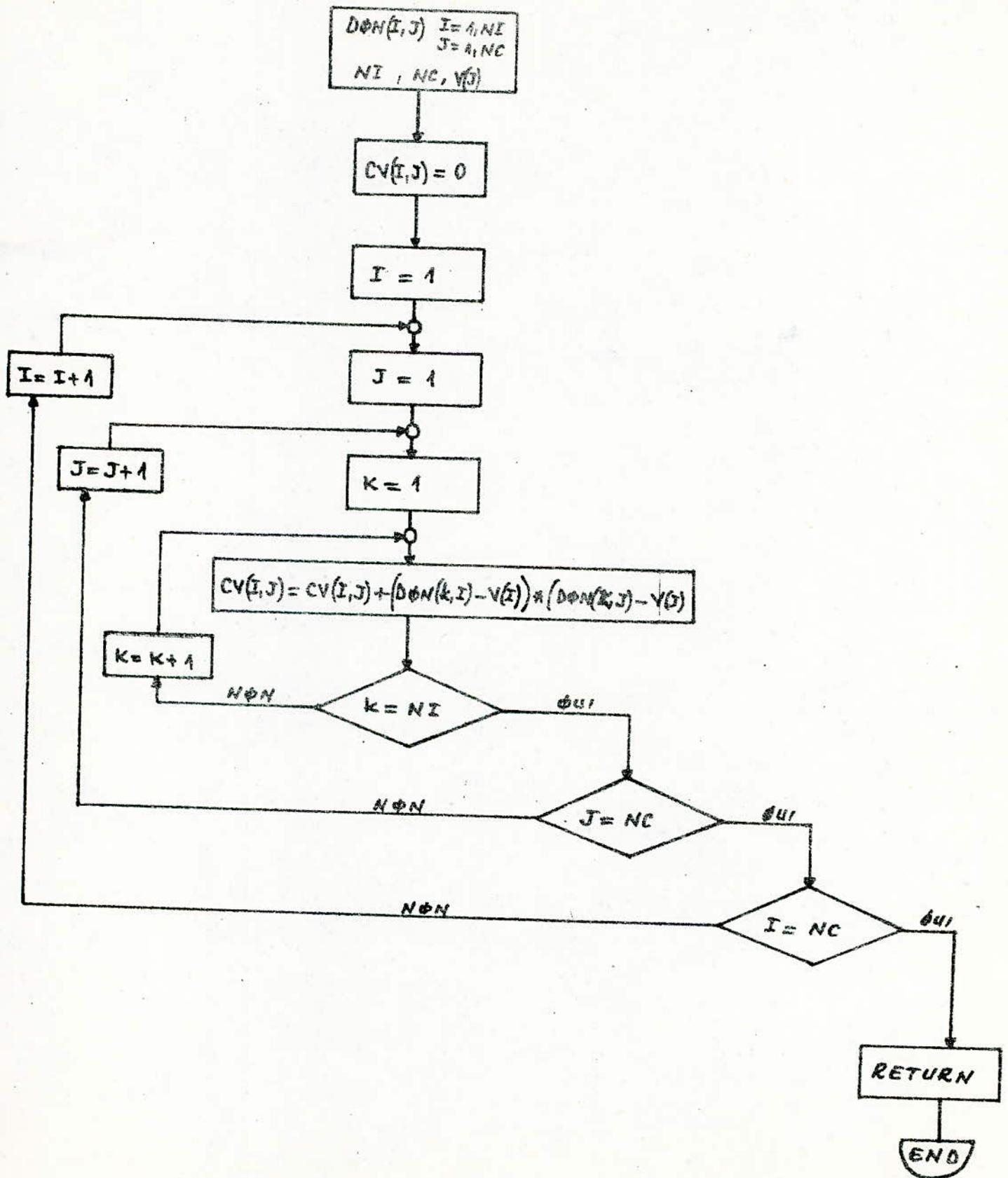


Organigramme Subroutine REDUCT

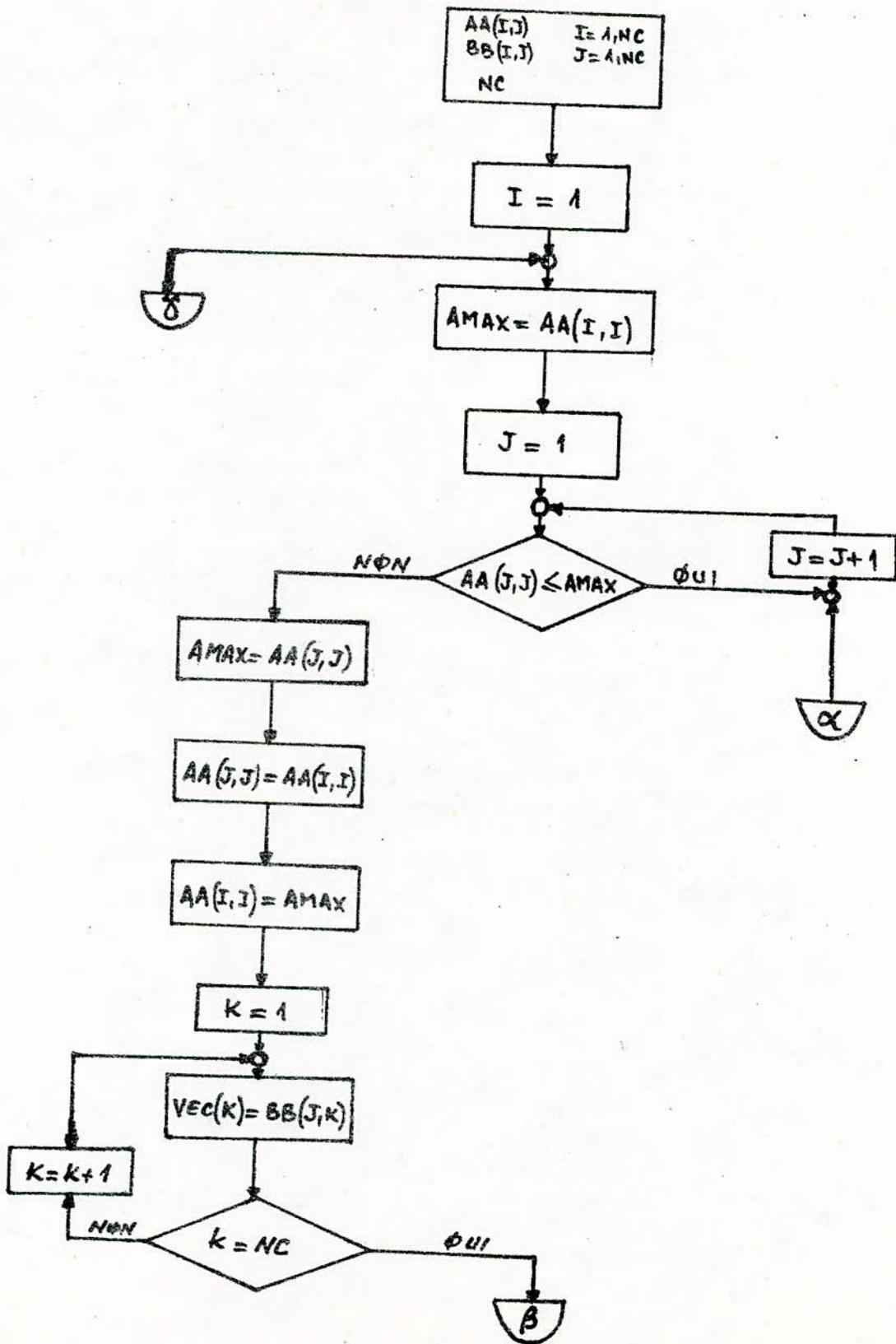


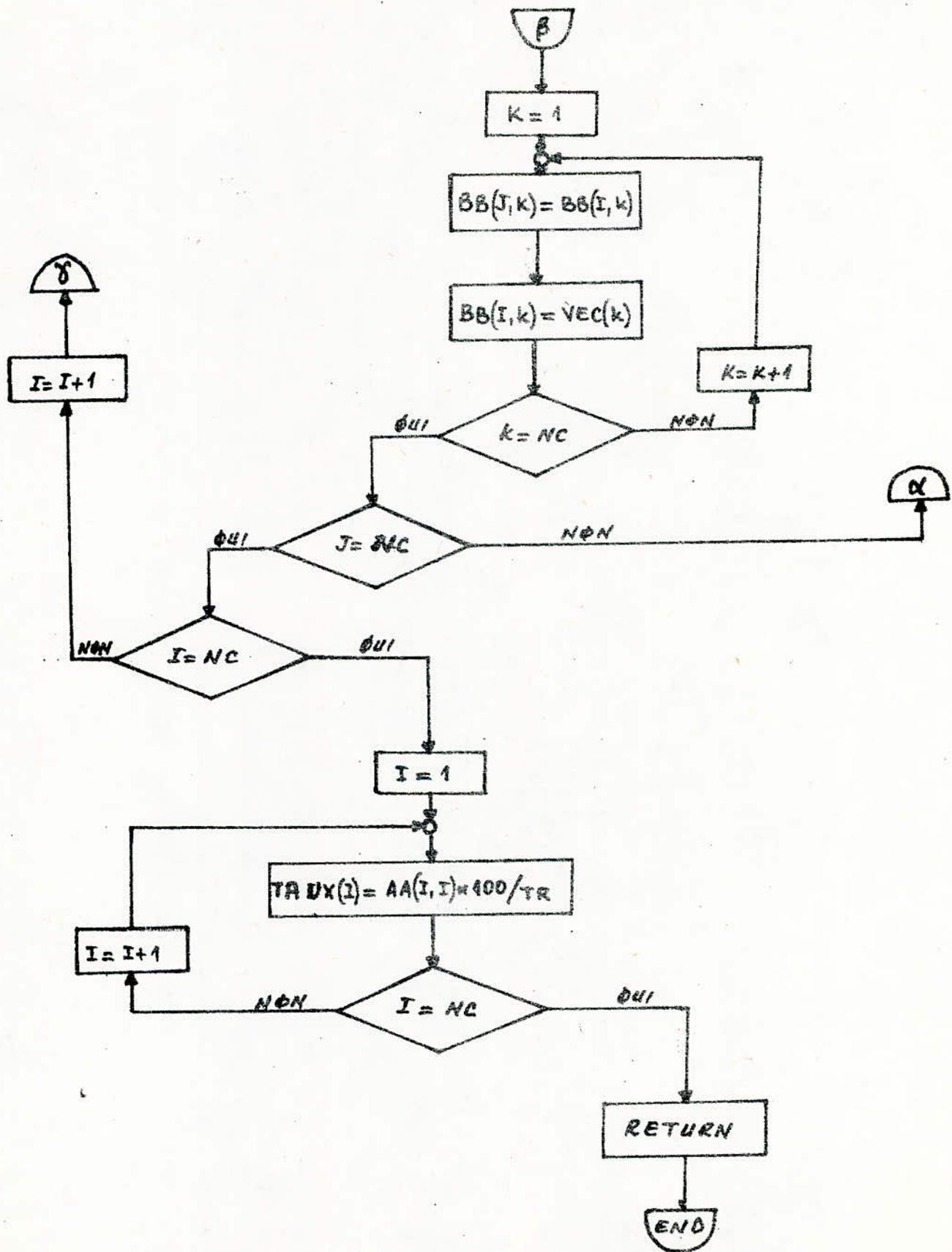


Organigramme Subroutine CALCΦV

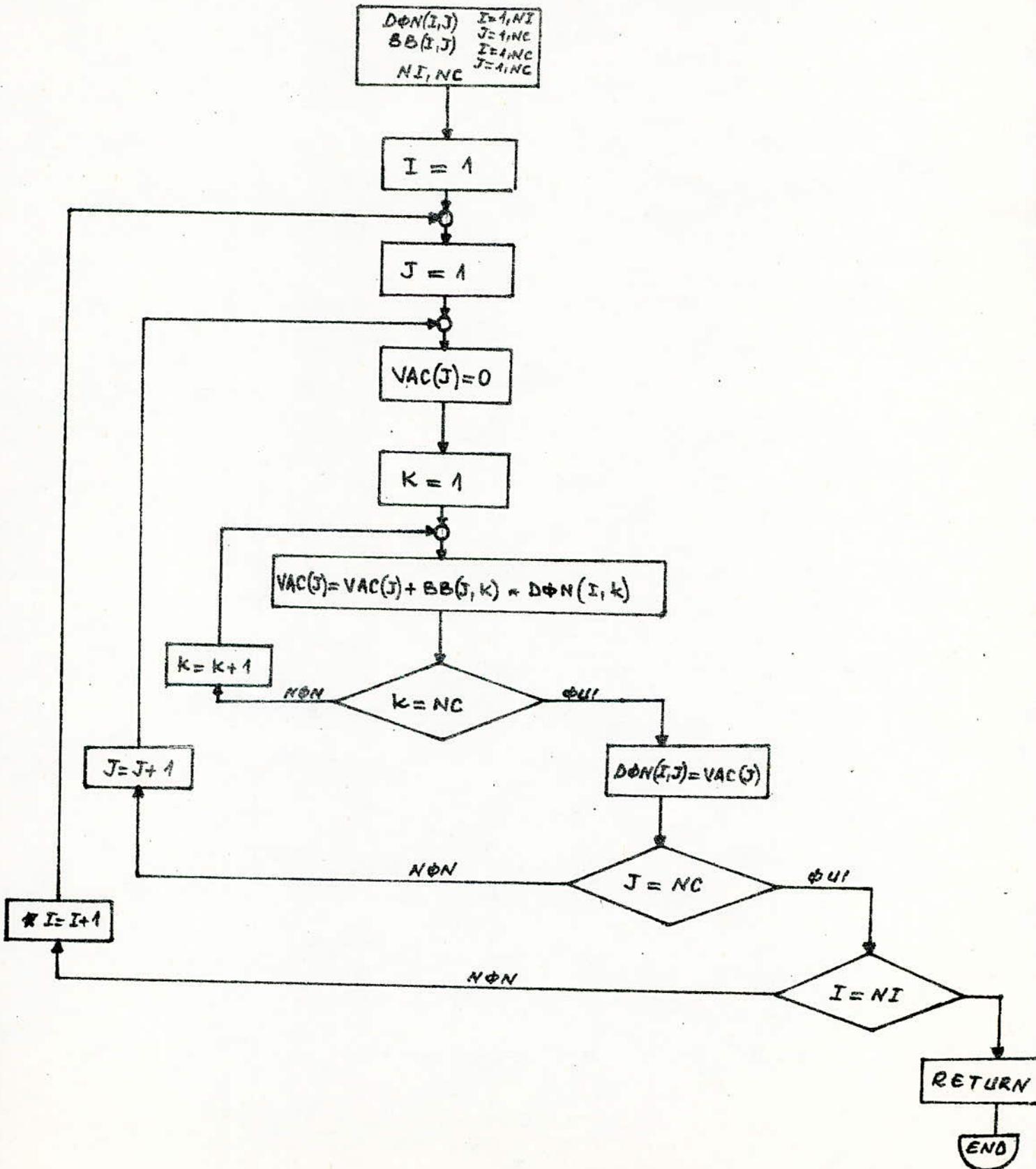


Organigramme Subroutine RAPØD

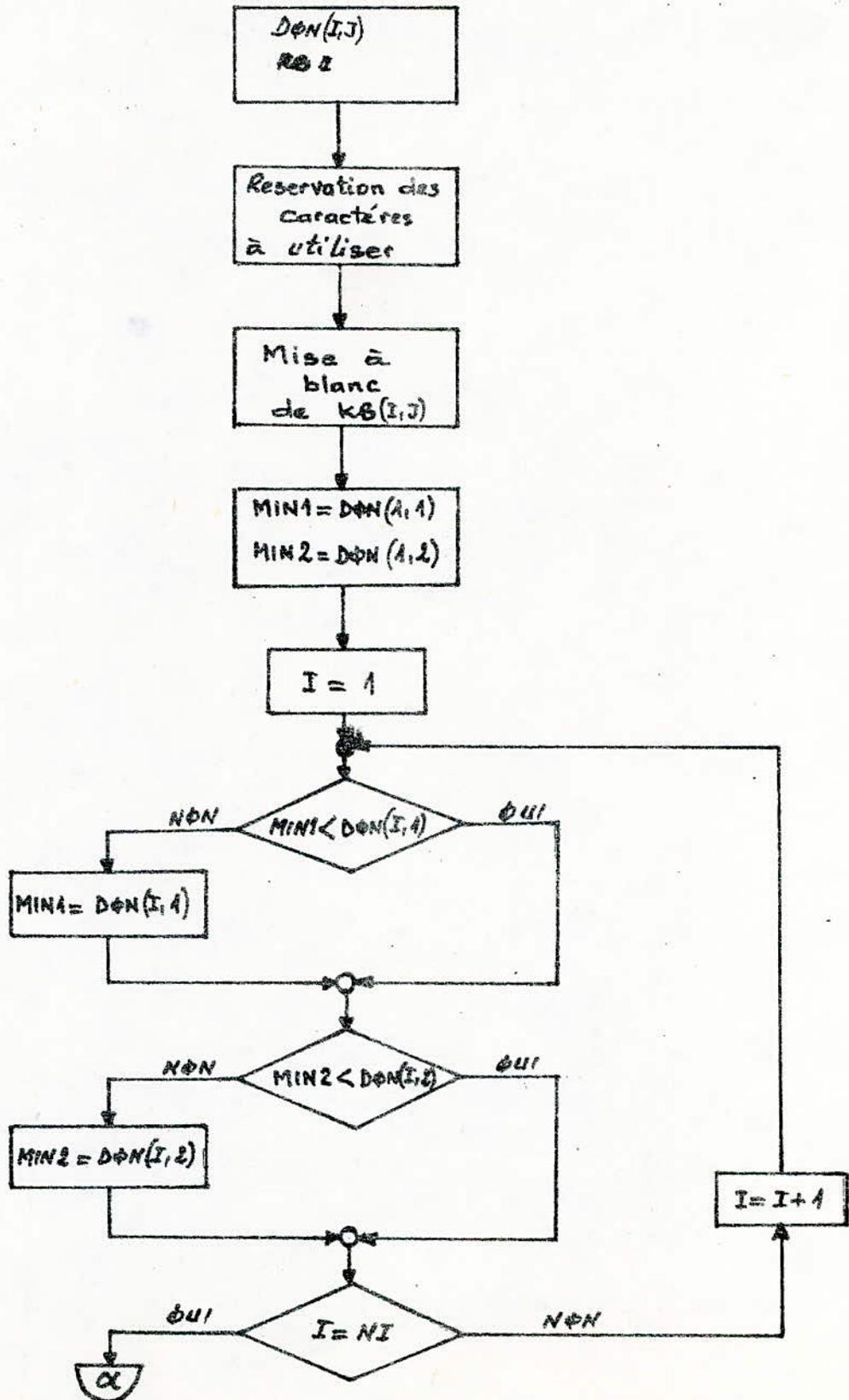


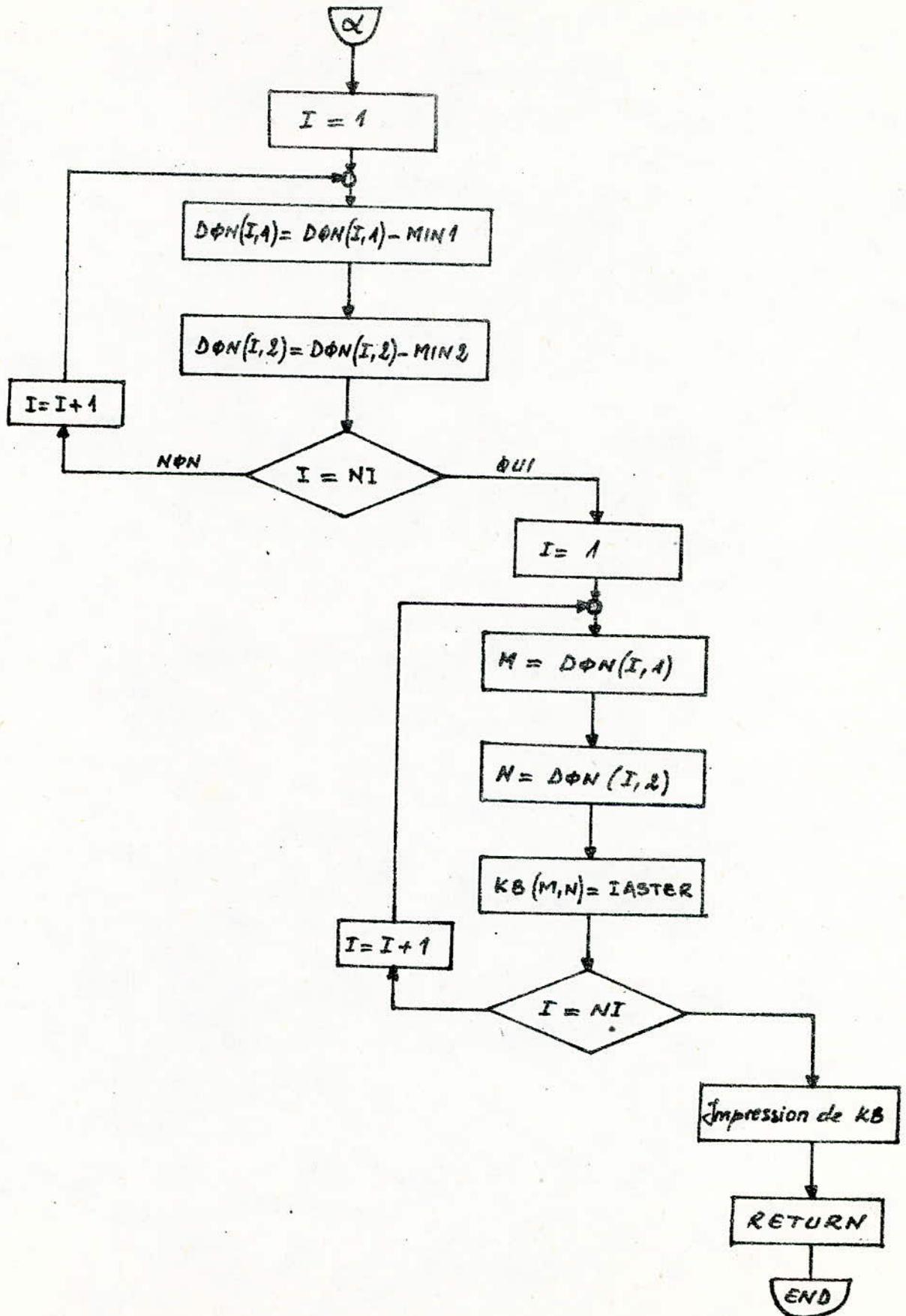


Organigramme Subroutine KARLφ

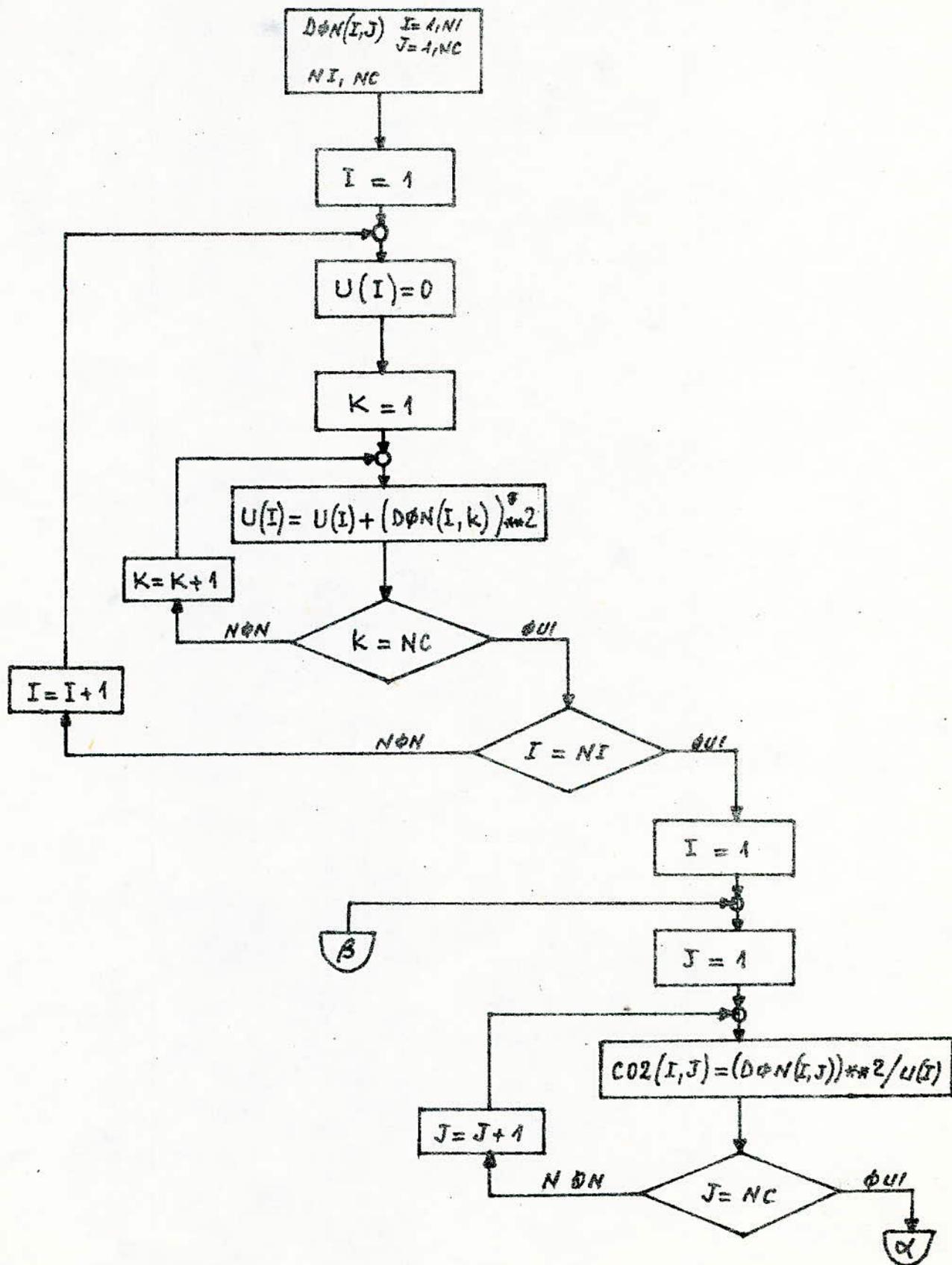


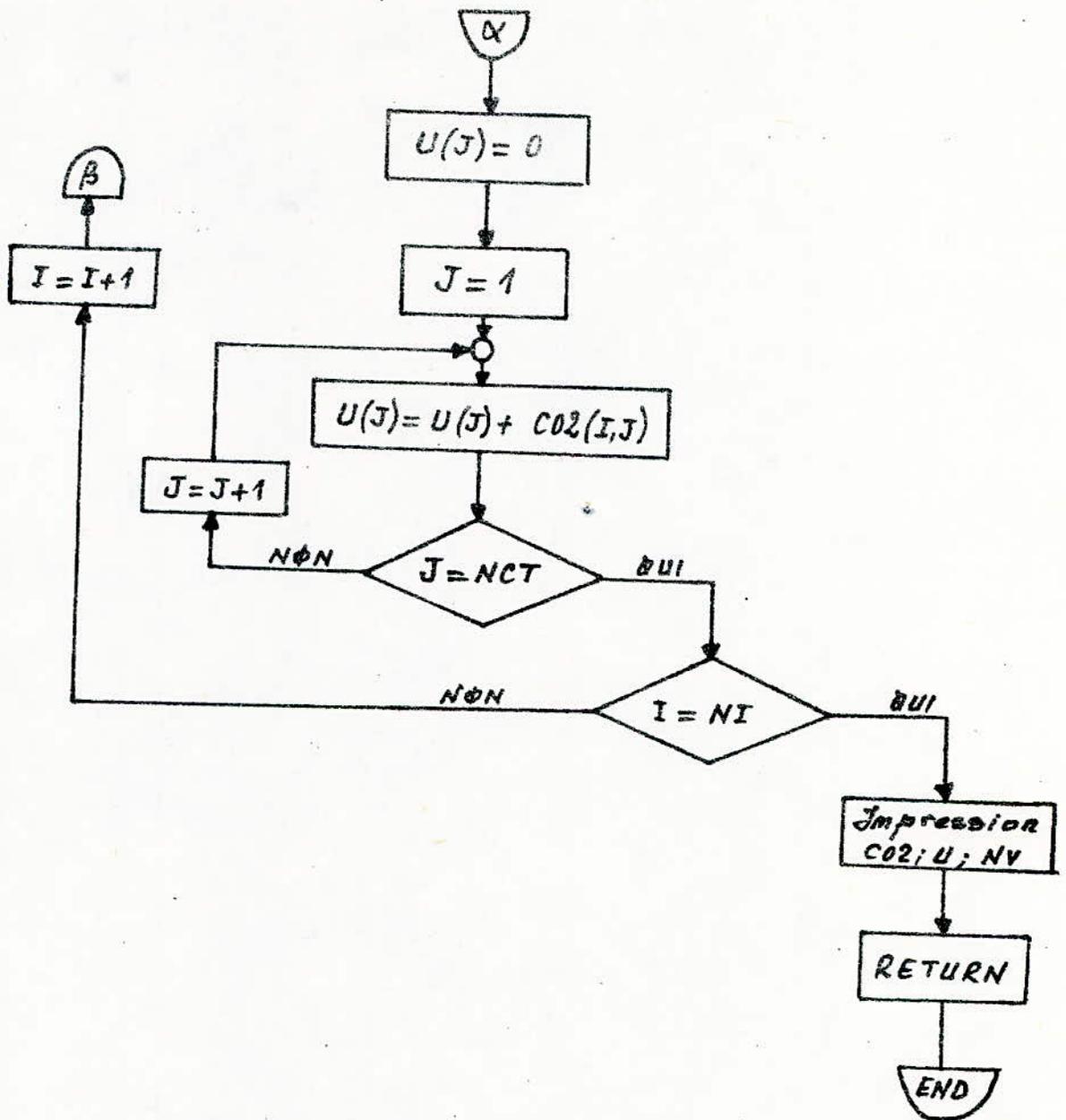
Organigramme Subroutine HISBID



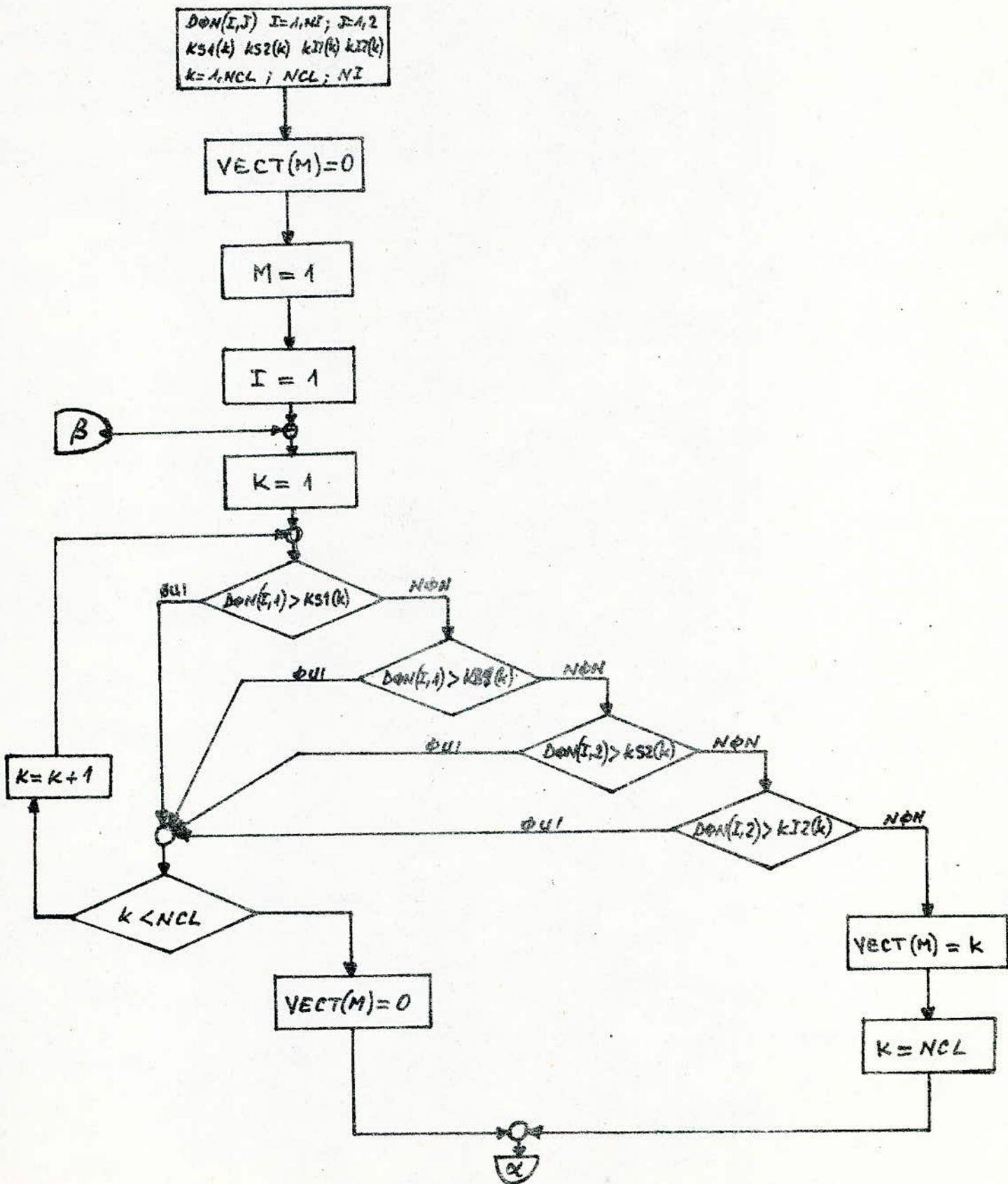


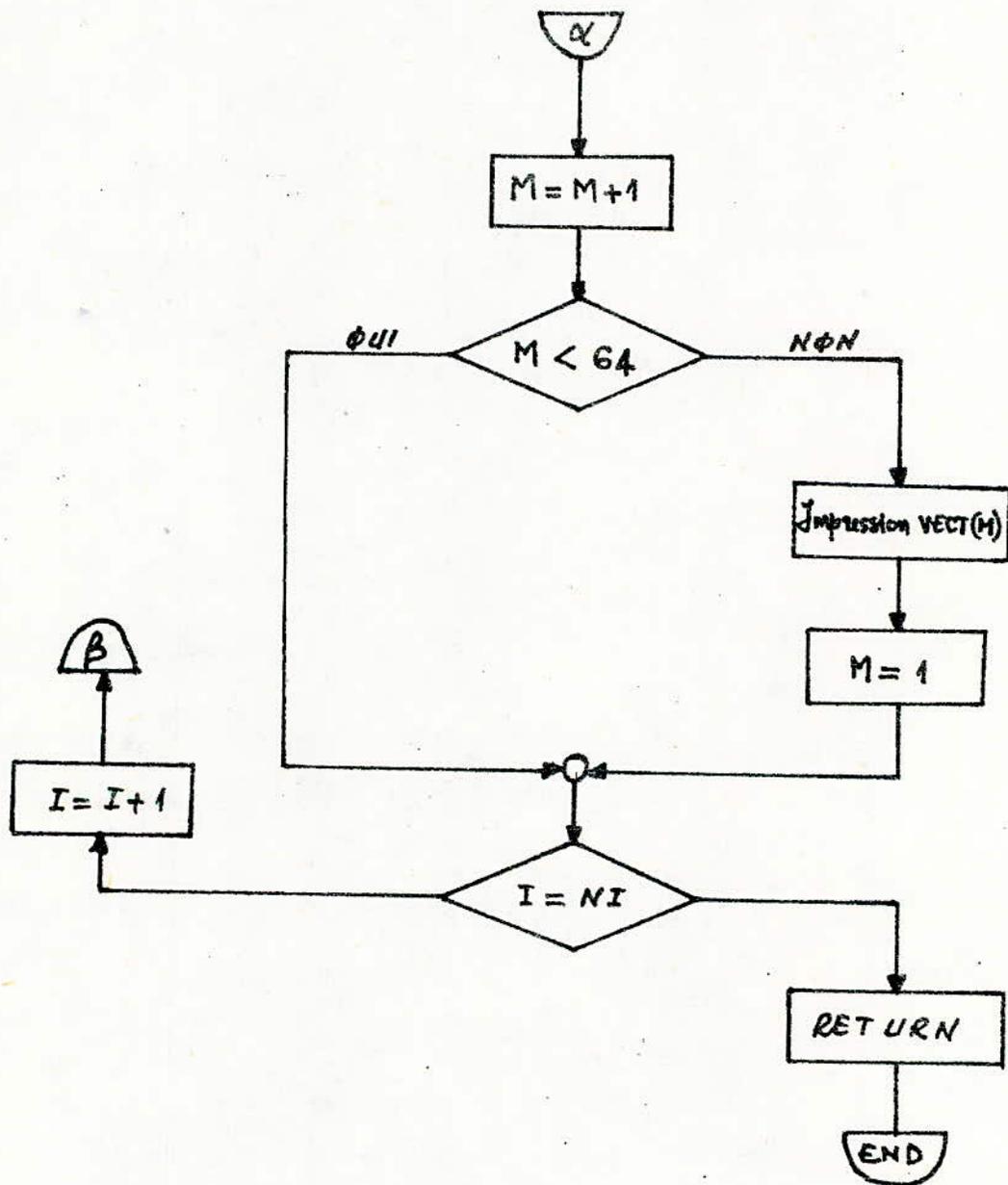
Organigramme Subroutine SONDE





Organigramme Subroutine REIMA





PROGRAMME PRINCIPAL

```

C   PRØGRAMME D'ANALYSE EN CØMPØSANTES PRINCIPAL
C   IL PERMET LA CØMPRESSIØN DES DØNNES
C   NC = NØMBRE DE CARACTERES (ØU CANAUX)
C   NI= NØMBRE D'INDIVIDUS (ØU PIXELS)
C   TINF=TAUX D'INFØRMATIØN DESIRE
C   KØP = ØPTIØN

DIMENSIØN DØN(20,20), CV(20,20), AA(20,20), BB(20,20),KB(100,80
A(210),B(400),VEC(20),TAUX(20),VAC(20),X(20),V(20),VECT(64)
KS1(5),KS2(5),KI1(5),KI2(5),ECT(20),VAR(20),CO2(20,20),U(20)
READ (105,100) NC,NI,TINF,KØP
PRINT 95 ,(NC,NI)
DØ 1 I=1,NI
1  READ(105,101)(DØN(I,J),J=1,NC)
   CALL CALMO(DØN,X,V,NI,NC,KØP)
   IF(KØP.EQ.1) GØTØ 2
   CALL REDUCT(DØN,V,NC,NI,VAR,ECT)
2  CALL CALCØV(DØN,V,NC,NI,CV,KØP)
   PRINT 102
   DØ 3 I=1,NI
   PRINT 103,(CV(I,J), J=1,NC)
3  CØNTINUE
C   CØNTRASTE TØTAL ØU TRACE DE CV :
   TR=0;
   DØ 4 I=1,NC
   TR=TR + CV(I,I)
   PRINT 104,TR

```

```

K= 0
DØ 5 J=1,NC
DØ 5 I=1,J
K = K+1
5 A(K)= CV(I,J)
N=NC
KØD=0
EPS=0.001
CALL MSPRØ(A,B,N,KØD,EPS)
K=1
DØ 6 J=1,NC
DØ 6 I =1,J
AA(I,J)=A(K)
K=K+ 1
6 CØNTINUE
L=1
DØ 7 I=1,NC
DØ 7 J=1,NC
BB(I,J) = B(L)
L= L+1
7 CØNTINUE
CALL RAPØD(AA,BB,VEC,TAUX,TR,NC)
PRINT 108
PRINT 109
DØ 8 I=1,NC
8 PRINT 110,((BB(I,J),J=1,NC),AA(I,I),TAUX(I))
NCT=0
PO=0.

```

```

DØ 9 I=1,NC
PO=PO + TAUX (I)
NCT = NCT + 1
IF (PO.GE.TINF) GØTØ 10
10 PRINT 111,NCT
CALL KARLØ(DØN,BB,VAC,NC,NI)
IF(KØP.EQ.1) GØTØ 11
CALL SØNDE (DØN,CO2 , U, NC, NI, NCT)
GØTØ 12
IF (NCT.NE.2) GØTØ 12
CALL HISBID ( DØn,KB,NC,NI)
CALL REIMA(DØN,KS1,KS2,KI1,KI2,NCL,NI,VECT)
12 STØP
95 FØRMAT (///// ,10X, NC=',I2,5X,'NI=',14 , //// )
100 FØRMAT(I2,I4,I2,I1)
101 FØRMAT(8 F9.2,F9.2)
102 FØRMAT(///// , 40X,'MATRICE DE CØVARIANCE',//)
103 FØRMAT(1X,18(F5.2,1X))
104 FØRMAT(///,10X,'CØNTRASTE TOTAL=,'F10.2)
108 FØRMAT(///// ,90X,' VALEURS)
109 FØRMAT(40X,'VECTEURS PROPRES,'34X,' PRØPRES','1X,' TAUX,'//)
110 FØRMAT(1X,18(F5.2,1X),1X,F10.2,1X,F5.2,' %',/)
111 FØRMAT(///,10X,'NØMBRE DE CARACTERES CHØISIS=,'I1,////)
END

```

```

SUBROUTINE CALMØ(DØN,X,V,NI,NC,KØP)
C   CALCUL DE LA SØMME ET DE LA MØYENNE PØUR CHAQUE
C   CARACTERE ( ØU CANAL).
C   LES SØMMES SØNT DANS LE VECTEUR X(J)
C   LES MØYENNES SØNT DANS LE VECTEUR V(J)

```

```

DIMENSION DON(20,20),X(1),V(1)
IF (KØP.EQ.0) GØTØ 52
DØ 51 J=1,NC
SIG =0
DØ 50 I=1,NI
50  X(J) = X(J) + DØN (I,J)
51  V(J) = FLØAT(X(J))/NI
GØTØ 55
52  J=1
DØ 54 J=1,NC
SIG = 0.
DØ 53 I= 1,NI
53  SIG = SIG + DØN(I,J)
X(J) = SIG
54  V(J) = X(J)/FLØAT (NI)
55  RETURN
END

```

```

SUBROUTINE REDUCT(DON.V.NC,NI,VAR.ECT)
C   PERMET DE CENTRER ET DE REDUIRE LES
C   DONNEES DANS LE CAS DES SONDAGES

DIMENSION DON (20,20),V(20),VAR(20),ECT(20)
DO 6 J=1,NC
VAR(J)=0.
DØ 5 I=1,NI
5   VAR(J)=VAR(J) * (DØN(J,J)-V(J))XX2 /FLOAT(NI)
6   ECT(J)= SQRT(VAR(J))
DØ 7 J=1,NC
DØ 7 I=1,NI
7   DØN(J,J)=(DØN(J,J)-V(J))/ECT(J)
PRINT 113
DO 8 I=1,NC
8   PRINT 114,(V(I),ECT(I))

113  FORMAT(///,11X,'MOYENNES:',21X,'ECART TYPE',//)
114  FORMAT(10X, F10.2,20X, F10.2,/)
RETURN
END

```

SUBROUTINE CALC^{ON}V (V,NC,NI,CV, ,KOP)

C CALCUL DE LA MATRICE DE COVARIANCE :

DIMENSION DON(20,20),CV(20,20),V(1)

IF(KOP.EQ.1)GOTO 52

DO 51 J=1,NC

51 V(J) =0.

52 I =1

DO 53 I=1,NC

DO 53 J=1,NC

CV(I,J)=0.

DO 54 K=1,NI

H = FLOAT (DON(K,J))

G = FLOAT (DON(K,I))

54 CV(I,J)=CV(I,J)+(G-V(I))*(H-V(J))/FLOAT(NI)

53 CONTINUE

RETURN

END

```

SUBROUTINE RAPOD(AA,BB,VEC,TAUX,TR,NC)
C  RANGEMENT DES VECTEURS PROPRES PAR VALEURS PROPRES -
  ASSOCIEES .
C  PAR ORDRE DECROISSANT ET CALCUL DU POURCENTAGE D INFORMATION.
C  DANS CHAQUE NOUVEAU CANAL.
  DIMENSION AA(20,20),BB(20,20),VEC(20),TAUX(20)
  DO 57 I=1,NC
    AMAX=AA(I,I)
    DO 56 J=I,NC
      IF(AA(J,J).LE.AMAX) GOTO56
      AMAX = AA(J,J)
    AA(I,I)=AMAX
  DO 54 K=1,NC
54  VEC(K) =BB(J,K)
    DO 55 K=1,NC
      BB(J,K)=BB(I,K)
55  BB(I,K)=VEC(K)
56  CONTINUE
57  CONTINUE
    DO 58 I=1,NC
58  TAUX (I)=AA(I,I)*100/TR
    RETURN
  END

```

```

SUBROUTINE KARL(DON, BB, VAC, NC, NI)
C   DETERMINATION DES NOUVELLES VARIABLES

DIMENSION DON(20,20), BB(20,20), VAC(1)
DØ  58 I= 1, NI
DØ  55 J= 1, NC
VAC(J) = 0.
DØ  55 K=1, NC
55  VAC(J)= VAC(J)+(BB(J,K))*FLØAT(DØN(I,K))
DØ  57 J=1, NC
57  DØN(I,J)=VAC(J)
58  CØNTINUE

RETURN

END

```

```

SUBROUTINE SONDE(DON,U, NC, NI, NCT)
C PERMET DE DETERMINER LA QUALITE DE LA REPRESENTATION
C ET L'IMPRESSION DES MATRICES DES CO2,  $\Sigma$  CO2, ET DES
C NOUVELLES VARIABLES
DIMENSION DON (20,20), CO2(20,20),U(20)
DO 74 I=1,NI
U(I)=0.
DO 74 K=1,NC
74 U (I) =  $U(I) + (DON (J,K))XX2$ 
DO 77 I = 1, NI
DO 75 J = 1, NC
75 CO2 (J,J) = (DON(J,J) XX2)/U(I)
U(I) = 0.
DO 76 J= 1,NCT
76 U(I) = U(I) + CO2 (J,J)
77 CONTINUE
PRINT 301
PRINT 302
DO 78 I=1,NI
78 PRINT 300, ((DON (I,J), J = 1, NCT), (CO2 (I,J), J = 1,NCT), U(I))
300 FORMAT (10X,3(F6.2,2X), 20X,3(F6.3,2X),10X,F6.3)
301 FORMAT (////, 6x, 'PROJECTION DES OBSERVATIONS')
302 FORMAT (8X, 'SUR LES NOUVEAUX AXES', 30X, 'MATRICE CO2', 15X, 'SOMME
DES CO2', //)
RETURN
END

```

```

SUBROUTINE HISBID(DON,KB,NC,NI)
C   DETERMINATION DES 2 NOUVEAUX CANAUX CONTENANT LE MAXIMUM
C   D'INFORMATION ET TRACE DE L'HISTOGRAMME BIDIMENSIONNEL
      INTEGER DON
      DIMENSION DON(4100,4),KB(100,80)
      DATA I BLANC, I ASTER/1H ,1H*/
      DO61 I=1,100
      DO61 J=1,80
61   KB (I,J)=IBLANC
      MIN1 =DON(1,1)
      MIN2 =DON(1,2)
      DO 62 I=1,NI
      IF(MIN1,LE?,DON(I,1)) GOTO62
      MIN1 = DON(I,1)
62   CONTINUE
      DO 63 I=1,NI
      IF(MIN2,LE,DON(I,2)) GOTO 63
      MIN2 = DON(I,2)
63   CONTINUE
      DO 64 I=1,NI
      DON(I,1)=DON(I,1)-MIN 1
      DON(I,2)=DON(I,2)-MIN 2
64   CONTINUE
      DO 65 I=1,NI
      M=DON(I,1)
      N=DON(I,2)
65   KB(M,N) = I ASTER
      PRINT 105

```

.../...

```
DO 66M=1,100
66 PRINT 106,(KE(M,N),N=1,80)
PRINT 105
105 FORMAT(10X,82(1HI))
106 FORMAT(10X,1HI,80A1,1HI)
RETURN
END
```

```

SUBROUTINE REIMA (DON,KS1,KS2,KI1,KI2,NCL,NI,VECT)
C   PERMET LA RESTITUTION DE L'IMAGE AVEC IMPRESSION
C   D'un CHIFFRE DISTINCT POUR CHAQUE CLASSE
INTEGER DON,VECT
DIMENSION DON(4100,4),KS1(5),KS2(5),KI1(5),KI2(5),VECT(64)
DO 80N=1,64
80  VECT(N)=0
PRINT 119
M=1
DO 83 I=1,NI
DO 83 K=1,NCL
IF (DON(I,1).GT.KS1(K)) GO TO 82
IF (DON(I,1).LT.KI1(K)) GO TO 82
IF (DON (I,2). GT.KS2(K)) GO TO 82
IF ( DON (I,2).LT.KI2(K)) GO TO 82
VECT(M)=K
K= NCL
81  M= M+1
IF (M.LE.64) GO TO 83
PRINT 120,(VECT(M),M=1,64)
M=1
GO TO 83
CONTINUE
82  IF (K.NE.NCL)GO TO 83
VECT (M)=0
GO TO 81
83  CONTINUE

```

.../...

```
PRINT 119
119 FORMAT (30X, 66(1HI))
120 FORMAT( 30X, 1HI, 64I1, 1HI)
RETURN
END
```

IV-RÉSULTATS DES APPLICATIONS:

I-Résultats relatifs à la télédétection:

Les données auxquelles nous avons appliqué notre logiciel se trouvent sur une bande magnétique et correspondent aux radiométries d'une image (512 x 512) des monts de Palni en INDE. Nous rappelons qu'une scène complète fournit près de 30 millions d'éléments d'information. Le traitement numérique de ces images nécessite de gros moyens, aussi s'est-on contenté de traiter une fenêtre (64x64).

La région de Palni est une région montagneuse recouverte d'une végétation pas très diversifiée, donc le fait de prendre une petite fenêtre ne nous a pas permis d'avoir des zones bien distinctes, car cette image a été prise au moment de la saison sèche c'est-à-dire une époque où une grande quantité de végétation est difficilement perceptible. C'est pour cela que nous avons dû lire plusieurs fenêtres afin d'en trouver une qui soit quelque peu différenciée. Afin d'illustrer ce choix, nous donnons ici deux exemples.

On s'est fixé un taux d'information (TINF) de 99%.

Première fenêtre:

Cette fenêtre correspond à une zone où l'on trouve des types de végétations sèches très similaires. Le canal 6 est le plus contrasté
($VAR(\text{canal } 6) = 1063,00$).

On remarque par ailleurs que le premier pseudo-canal contient plus de 99% de l'information utile; il est combinaison linéaire des quatre canaux initiaux, le canal 6 à la plus grande contribution car le plus contrasté (0.76). Les trois autres pseudo-canaux contiennent moins de 1% de l'information utile à eux trois.

Nous avons néanmoins fait une représentation bidimensionnelle; on remarque que le nuage de points est très allongé suivant le premier axe et beaucoup moins suivant le deuxième. On ne peut pas distinguer de classes, on se retrouve dans le cas monospectral, en ne gardant que ce premier pseudo-canal.

Deuxième fenêtre:

Cette fenêtre correspond à une zone où l'on a une végétation humide. Les canaux 5 et 6 sont les plus contrastés (2369,52 et 2342,52).

Le premier pseudo-canal contient 98,54% de l'information utile, il est combinaison linéaire des canaux initiaux. Les canaux 5 et 6 ont une plus grande contribution (0.67) car les plus contrastés. S'étant fixé au départ un taux d'information de 99%, on le retrouve avec les deux premiers pseudo-canaux.

La représentation bidimensionnelle nous montre un nuage diffus, néanmoins, on remarque des regroupements mal délimités, cependant nous avons choisis trois groupes (classes).

En définissant les bornes de chaque classe ainsi formée, on fait une représentation numérique de l'image, en affectant à chacun de ces points un chiffre distinct correspondant au numéro de la classe à laquelle il a été assigné. Les points non classés sont représentés par le chiffre "0". La classe 2 est la plus dominante.

La première composante principale est une combinaison linéaire des quatre canaux initiaux, tous ses facteurs sont positifs, elle se rapproche le plus d'une image pancromatique.

FENETRE 1

NC= 4

NI=4096

MATRICE DE COVARIANCE

263.99				
267.02	274.56			
524.93	530.18	1063.00		
263.41	265.89	534.49	270.67	

CONTRASTE TOTAL= 1872.21

VECTEURS PROPRES

.37	.38	.76	.38
.38	.76	-.40	-.34
.69	-.48	.16	-.52
.50	-.20	-.49	.69

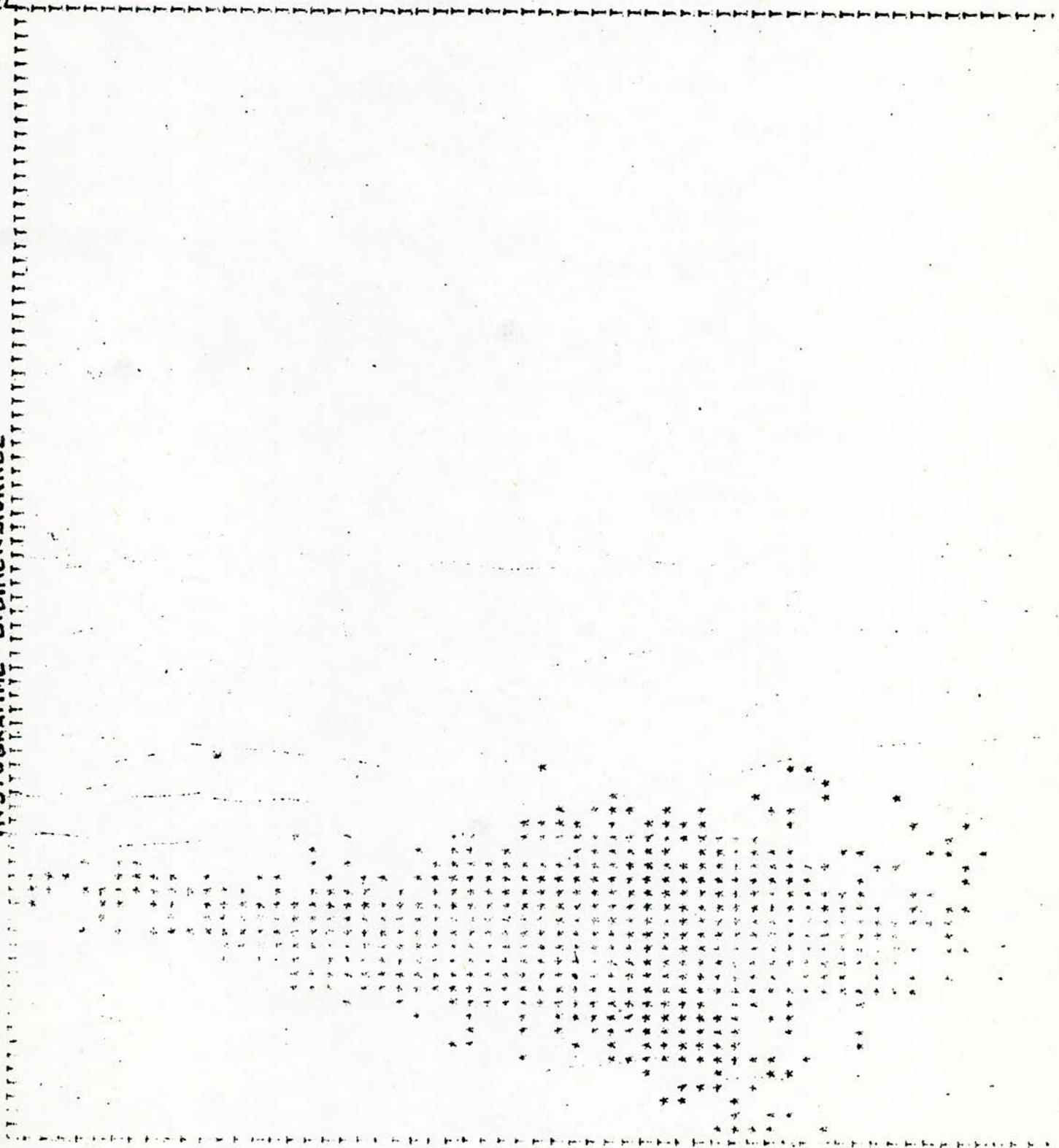
VALEURS PROPRES TAUX

1858.92	99.29%
10.31	.55%
1.70	.09%
1.28	.07%

NOMBRE DE CARACTERES CHOISIS=1

xe2

HISTOGRAMME BIDIMENSIONNEL



Axe2

FENETRE: 2

NC= 4 NI=4096

MATRICE DE COVARIANCE

259.79			
768.17	2369.52		
768.68	2297.75	2342.52	
256.41	767.59	766.94	266.30

CONTRASTE TOTAL = 5238.12

VECTEURS PROPRES				VALEURS PROPRES	TAUX
.22	.67	.67	.22	5161.79	98.54%
.06	-.72	.69	.06	58.61	1.12%
.24	-.14	-.25	.93	12.99	.25%
.94	-.07	-.14	-.29	4.71	.09%

NOMBRE DE CARACTERES CHOISIS=2

IV-2 Résultats relatifs au sondage .

Les données étant présentées sur cartes en vue de traitement, les résultats sont fournis sur listing d'ordinateur .

Nous donnons les résultats dans les tableaux : III, IV, V et VI

- Dans le tableau III, on trouve les moyennes et les écarts types pour chaque variable .

- Dans le tableau IV, on trouve les covariances entre les variables, nous n'avons imprimé que la partie inférieure, la matrice étant symétrique.

Les variables étant centrées et réduites, l'inertie totale est égale à 18(nombre de variables)

- Le tableau V donne les résultats fondamentaux de l'analyse à savoir les vecteurs propres et les valeurs propres associées en ordre décroissant ainsi que le pourcentage d'information relatif à chaque valeur propre.

Pour retrouver le taux d'information désiré (on a pris 65 %), il faudrait tenir compte de 3 caractères .

Le tableau VI-a) donne les coordonnées des observations (pays) les trois axes principaux.

Le tableau VI-b) donne la qualité de la représentation de chacun des pays par chacun des trois premiers axes. La première colonne donne le cosinus carré des observations avec le premier axe, la deuxième avec le deuxième axe et la troisième avec le troisième axe.

Quand au tableau VI-c) il nous donne la qualité de la représentation de chacun des pays dans le sous espace constitué par les trois premiers axes, il donne la somme des cosinus carrés correspondant à ces trois axes. Un pays est d'autant mieux représenté que le

.../...

chiffre correspondant est proche de 1 valeur que l'on obtiendrait si on retenait 18 axes .

Exemple :

Le premier caractère correspondant au vecteur propre associé à la plus grande valeur propre tient compte de 39,71 % de l'information totale .

Il est combinaison linéaire des anciens caractères, ceux qui ont une plus grande contribution sont :

TV(0.35), PNB(0.33), IMPT(0.32), EXPT(0.32) à opposer à PIBA(-0.34) et (ASP(-0.33)

Les pays qui sont relativement les mieux représentés sont : Grèce (0.716), Etats-Unis(0.661), Espagne(0.629) et Portugal(-0.589) mais si l'on regarde maintenant leurs contributions on remarque que l'Espagne(-3.07) et le Portugal(-4.07) accusent une contribution inférieure à $\sqrt{18} = \pm 4,24$.

On en déduit finalement que seuls les Etats-Unis (6.11) et la Grèce(-4.90) sont bien représentés sur le premier axe .

Le deuxième caractère quand à lui tient compte de 16,88 % de l'information totale .

Les anciens caractères qui ont une plus grande contribution sont :

POPU(0,37) et RESO(0,35) à opposer à RECC(-0,43) et EDUC(-0,41)

Le pays qui est relativement le mieux représenté est :

Suede(0,638) mais avec une contribution de (-3.62) seulement

Le troisième caractère tient compte de 12.88 % de l'information avec les caractères LOG(0,49) et FBCF(0,47) à opposera CAL(-0,48)

Le pays le mieux représenté est le JAPON (0,540) mais avec
Une contribution de (3.71) seulement.

En faisant la somme des CO2, on peut dire que la Suede(0,869) et la Grece(0,869) sont les pays les mieux représentés sur les trois nouvelles variables.

Ainsi en passant de \mathbb{R}^{18} à \mathbb{R}^3 on ne perd que 30% de l'information . Bien sur, on peut approfondir l'etude de ces tableaux, en voyant la proximité des differents pays... cela sort du cadre de notre projet de fin d'étude, on laisse ce soin aux economistes.

Tableau I. - Code des Caractéristiques et des pays.

Caractéristique (Variable)		N°	Pays (Individus)	
Population totale (1000 hab.)	POPU	1	ALLEMAGNE	D
Densité (au km ²)	DENS	2	AUTRICHE	A
Taux d'accroissement total de la population	TATO	3	BELGIQUE-LUXEMBOURG	B
% d'actifs dans l'agriculture, la sylviculture et la pêche	AASP	4	CANADA	CDN
% d'actifs dans l'industrie	AIND	5	DANEMARK	DK
Produit National Brut en dollars par habitant	PNB	6	ESPAGNE	E
% du P.I.B. en agriculture	PIBA	7	ETATS-UNIS	USA
Formation brute du capital fixe en % du P.N.B.	FBCF	8	FINLANDE	SF
Recettes courantes en % du P.N.B.	RECC	9	FRANCE	F
Réserves officielles (en millions de dollars)	RESO	10	GRECE	G
Taux d'escompte officiel	TESC	11	IRLANDE	SE
Importations marchandises totales (en millions de dollars)	IMPT	12	ITALIE	I
Exportations marchandises totales (en millions de dollars)	EXPT	13	JAPON	JAP
Calories par habitant et par jour	CAL	14	NORVEGE	N
Nombre de logements achetés pour 1000 hab.	LOG	15	PAYS-BAS	NL
Consommation d'électricité en kWh par habitant et par an	ELEC	16	PORTUGAL	P
Dépenses publiques d'éducation en % du P.N.B.	EDUC	17	GRANDE-BRETAGNE	RUN
Nombre de T.V. pour 1000 habitants	TV	18	SUEDE	S

TABLEAU II - DONNEES

caracté- risti- ques Pays	1	2	3	4	5	6	7	8	9
D	50848.00	245.00	1.05	9.60	49.10	2520.00	3.60	24.40	37.90
A	7373.00	88.00	.05	19.10	39.90	1690.00	7.00	23.20	37.50
B	9984.00	332.00	.06	5.40	44.80	2352.00	5.40	23.10	37.10
CDN	21089.00	2.00	1.85	8.20	32.30	3460.00	5.90	21.70	35.20
DK	4893.00	114.00	.75	11.90	38.50	2860.00	8.90	22.00	37.10
E	32949.00	65.00	.95	30.70	37.40	870.00	15.00	22.00	22.40
USA	203213.00	22.00	1.35	4.60	33.70	4660.00	2.90	16.70	31.50
SF	4706.00	14.00	.70	24.50	34.50	1940.00	14.70	23.80	35.90
F	50325.00	91.00	1.05	15.10	40.60	2770.00	6.00	25.40	38.10
GR	8856.00	67.00	.70	48.20	22.50	950.00	20.30	29.70	26.90
SE	2921.00	42.00	.25	28.40	29.70	1040.00	19.70	19.90	30.70
I	54120.00	180.00	.85	21.50	43.10	1520.00	11.30	20.50	33.30
JAP	102380.00	277.00	1.05	18.80	35.00	1630.00	8.70	35.20	21.20
N	3851.00	12.00	.80	14.70	36.80	2530.00	6.50	25.30	43.40
-NL	12873.00	352.00	1.25	7.50	41.30	2190.00	7.00	25.50	41.90
P	9883.00	105.00	.90	31.50	35.50	600.00	17.70	18.40	24.00
RUN	55643.00	228.00	.65	2.90	46.80	1970.00	3.00	17.30	39.00
S	7969.00	18.00	.70	8.80	40.40	3230.00	5.90	23.60	48.10
caracté- risti- ques Pays	10	11	12	13	14	15	16	17	18
D	10940.00	6.50	24926.00	29052.00	2990.00	8.60	3322.00	3.00	231.00
A	1563.00	5.00	2825.00	2412.00	2990.00	6.60	2647.00	4.40	134.00
B	2406.00	7.00	9934.00	10069.00	3150.00	5.00	2814.00	5.30	184.00
CDN	3846.00	6.00	13137.00	13754.00	3160.00	8.20	8199.00	5.70	279.00
DK	384.00	9.00	3800.00	2958.00	3180.00	9.00	2413.00	6.00	244.00
E	1519.00	6.50	4233.00	1900.00	2750.00	6.40	1245.00	2.10	84.00
USA	12306.00	5.75	36052.00	37988.00	3210.00	7.70	7013.00	5.10	392.00
SF	379.00	6.00	2023.00	1985.00	2900.00	7.90	3836.00	6.30	193.00
F	4617.00	7.50	17392.00	15020.00	3160.00	8.20	2407.00	4.80	185.00
GR	290.00	6.50	1594.00	554.00	2910.00	10.10	823.00	2.40	9.00
SE	694.00	7.31	1413.00	891.00	3450.00	4.00	1577.00	4.20	141.00
I	4642.00	5.50	12450.00	11729.00	2940.00	5.10	1810.00	5.80	146.00
JAP	3072.00	6.00	15024.00	15990.00	2450.00	11.90	2734.00	4.50	190.00
N	1607.00	4.50	2943.00	2203.00	2910.00	8.80	12976.00	5.80	175.00
NL	2621.00	6.00	10991.00	9965.00	3240.00	9.70	2565.00	6.70	197.00
P	1442.00	3.50	1234.00	883.00	2930.00	4.30	607.00	1.40	29.00
RUN	2469.00	7.00	19956.00	17515.00	3180.00	7.70	3690.00	4.20	263.00
S	586.00	7.00	5899.00	5698.00	2750.00	13.40	6803.00	7.40	288.00

MO=18 MI= 18

MOYENNES:	ECART TYPE
36310.33	48567.77
125.22	111.73
.89	.75
17.30	11.54
37.87	6.19
2154.56	1001.74
9.42	5.51
27.21	4.26
34.35	7.12
3016.78	3345.08
6.25	1.19
10326.28	9329.66
10028.11	10147.98
3014.40	222.42
7.92	2.42
3748.30	3080.49
4.73	1.60
185.22	91.11

TABLEAU III

VECTEURS PROPRES

VALEURS PROPRES TAUX

.23	.06	.20	<u>-.33</u>	.19	<u>.37</u>	<u>-.70</u>	-.02	.19	.27	.06	<u>.32</u>	<u>.32</u>	.09	.11	.19	.19	<u>.35</u>	7.148	39.71%
<u>.37</u>	.10	.16	.15	-.06	-.11	.09	-.06	<u>-.47</u>	<u>.35</u>	-.16	.28	.29	-.05	-.21	-.24	<u>-.41</u>	-.10	3.038	16.88%
.14	-.16	.25	.14	-.29	.09	.01	<u>.47</u>	-.17	-.02	-.14	-.00	.02	<u>-.48</u>	<u>.49</u>	.22	.05	.03	2.318	12.88%
-.06	.61	-.13	-.11	.40	-.20	-.15	.37	.00	-.03	.18	.04	.04	-.24	.19	-.31	-.01	-.09	1.982	11.01%
.11	-.07	-.08	.11	-.32	.14	.17	.11	-.12	-.01	.74	.08	.06	.27	.19	-.31	.08	.13	1.268	7.04%
-.27	.34	.72	-.08	-.20	.07	.00	.17	-.06	-.05	-.10	-.06	-.06	.40	-.09	-.05	.12	-.11	.674	3.74%
.32	.32	-.43	.00	-.31	-.02	.13	.24	-.04	-.01	-.32	.05	.07	.20	-.18	.10	.50	.02	.434	2.41%
.19	.07	.20	-.22	.06	-.00	.05	-.32	-.44	-.32	.06	-.12	-.17	-.37	-.17	-.19	.36	.28	.403	2.24%
.01	.12	-.14	-.27	-.13	.01	-.28	.18	-.38	-.27	.24	-.03	-.05	.08	-.21	.53	-.39	.03	.283	1.57%
.13	.25	-.06	-.12	-.25	-.13	-.01	-.40	.08	-.35	-.25	.08	-.00	.19	.56	-.10	-.28	.13	.199	1.11%
-.07	.20	.11	.11	.13	-.43	.51	-.20	.03	.08	.25	.15	.17	-.07	.08	.52	.12	.07	.102	.56%
-.38	.27	-.19	-.15	-.14	.40	.29	-.09	-.15	.39	-.09	-.35	.07	-.15	.06	-.01	-.16	.29	.070	.39%
-.26	-.36	-.04	-.32	-.01	-.48	.06	.31	-.07	0.00	-.17	.04	.21	.14	-.03	-.19	-.03	.48	.045	.25%
.47	-.10	.07	-.25	.30	-.10	.19	.12	-.06	.29	-.01	-.41	-.39	.29	.22	.07	-.07	-.03	.021	.12%
.04	-.11	-.01	-.32	.24	.40	.57	.19	.07	-.38	-.10	.29	.10	.01	-.05	-.05	-.14	-.18	.011	.06%
.08	.11	.05	.60	.32	.19	.04	.16	-.00	-.24	-.11	-.04	-.08	.18	-.08	.05	-.15	.56	.004	.02%
.16	.15	.09	-.13	-.31	-.09	.10	.08	.50	.10	.10	.19	-.45	-.27	-.31	-.06	-.24	.26	.001	.00%
.27	.04	.13	-.03	-.09	-.05	.02	.02	.34	-.26	.10	-.59	.57	-.10	-.18	-.04	-.12	-.02	0.000	0.00%

OMBRE DE CARACTERES CHOISIS: 3

PROJECTION DES OBSERVATIONS
SUR LES CANONICAL AXES

3.10	1.93	-.45
-1.35	-.69	-.67
.61	-.53	-2.08
2.29	-.52	1.07
.20	-2.08	-.88
-3.07	1.67	.23
6.11	3.15	.47
-1.58	-1.40	.54
1.34	.10	-.27
-4.90	.95	2.06
-3.38	-.12	-2.74
-.26	.93	-1.03
-.10	1.71	3.71
.16	-2.58	1.44
1.26	-1.12	-.40
-4.07	2.27	-.83
2.11	-.06	-2.10
1.54	-3.62	1.53

a)

MATRICE CO2

.478	.186	.010
.315	.083	.078
.038	.029	.443
.308	.016	.067
.004	.414	.074
<u>.629</u>	.186	.004
<u>.661</u>	.176	.003
.370	.290	.043
.382	.002	.011
<u>.716</u>	.027	.126
.489	.001	.235
.013	.161	.195
.000	.115	<u>.540</u>
.001	.374	.116
.139	.110	.014
<u>.589</u>	.182	.025
.362	.000	.360
.116	<u>.638</u>	.115

b)

SOMME DES CO2

.673
.476
.510
.391
.492
.819
.801
.704
.396
<u>.869</u>
.724
.360
.656
.491
.263
.796
.722
<u>.869</u>

c)

STOP

CONCLUSION

À partir, notre but était l'étude de l'analyse en composantes principales et la mise au point d'un logiciel destiné à aider les non informaticiens devant travailler en conversationnel avec la machine au moyen d'un terminal de visualisation.

Généralement, les images multispectrales de télédétection sont formées d'un nombre important de points, on est alors amené à travailler avec des tableaux très grands de nombres.

L'application de l'analyse en composantes principales à ces images nécessite de gros moyens, aussi s'est-on limité à prendre une petite fenêtre d'une image.

Néanmoins l'application de cette A.C.P.S. a montré qu'il était possible de réduire la masse de données et de décorréler les canaux avec un minimum de perte d'information.

Il serait intéressant de compléter notre travail pour pouvoir étudier une image multispectrale complète.

BIBLIOGRAPHIE

- A. ABDELLAOUI -R2SUMES de séminaires sur la télédétection
- J. DESACHY Thèse de docteur de spécialité (Nov.1980)
Contribution à l'étude de la texture en télédétection
- B- LACAZE thèse de docteur ingénieur
Le traitement des données multispectrales acquises
par télédétection (applications à l'étude de la végétation).
- S.CASTAN - Eléments de traitement d'images
- C N E S - Principes physiques et mathématiques de la télédétection.
- Contribution à l'analyse écologique des zones arides de Tunisie avec l'aide de la télédétection spatiale. (dec.1978)
- Identification des caractéristiques des images du détecteur multispectrale à balayage utiles pour la recherche hydrogéologique Tome 1 (janv.1976).
- Revue du CNES - La télédétection des ressources terrestres (1973)
- Cours de technologie - Méthodes d'analyse en télédétection des
CNES
Spatiale au ressources terrestres (1976)
- Revue:O.R.S.T.O.M : - Les satellites d'observation de la terre :
LANDSAT (1977)
- J.VIGNES et M. LA PORTE:- théorie et pratique de la programmation
FORTRAN .
- P.BERTIER et J.M.BOUROCHE - Analyse de données multidimensionnelles

ANNEXE

ANNEXE 1

Programme de lecture sur bandes :

Ce programme permet de lire sur la bande une fenêtre NBLOC X NBLOC choisie (64X64). Il comporte un programme principal et un sous-programme (SAUT). Nous avons effectué un petit changement au niveau du P.P. : On lit les valeurs et on les transfère dans notre matrice de données $DON(I,J)$.

Nous indiquons que nous avons empruntés .Ce programme fait par nos camarades dans le cadre de leur étude .

Programme Principal :

```
DEFINE FILE 10=T(R:128,RW,BN,DV)
INTEGER BUF(64), FUB(128),DON(4100,4)
READ(105,400) N1,N2,N3,NBLOC
M=N2
N=N3
DØ25 IP=1,4
CALL SAUT (N1,M,N)
JB=0
JB= JB+1
READ(10) BUF
DØ 22 I=1,NBLOC
J=(2*I)-1
K=2*I
FUB(J)= BUF(I)/256
22 FUB(K)= BUF(I) *256/256
DØ 23 I=1,3
23 READ(10) BUF
```

```

DØ 24 I=1,NBLØC
IA =I +(JB-1)* NBLOC
24 DØN(IA,IP)= FUB(I)
IF(JB.LT.NBLØC) GØTØ 21
M= (512-NBLØC)
N= N3
25 CØNTINUE
400 FØRMAT (4I4)
SUBRØUTINE SAUT(N1,M,N)
INTEGER BUF (64)
N2=M
N3=N
IF(N1.EQ.1) GØTØ 2
L= (N1-1)* 2049
DØ 1 I=1,L
1 READ(10) BUF
2 IF(N2.EQ.1) GØTØ 4
L=(N2-1)*4
DØ 3 I = 1,L
3 READ (10) BUF
4 I F(N3.EQ.1) GØTØ 6
L=N3-1
DØ 5 I=1,L
5 READ (10) BUF
6 RETURN
END

```

ANNEXE 2

Le problème du calcul des vecteurs propres et des valeurs propres est un problème classique qui est utilisé très souvent.

Au niveau du centre de calcul du CSTN il existe un sous-programme permettant de résoudre ce problème (il s'applique à des matrices symétriques). On utilise uniquement la matrice triangulaire supérieure, la lecture se faisant colonne par colonne, et on transfère ces valeurs dans un vecteur.

Nous donnons ici ce sous-programme:

```
SUBROUTINE MSPRO(A, B, N, KOD, EPS)
```

```
DIMENSION A(1), B(1)
```

```
C 1. A: TABLEAU DANS LEQUEL ON RANGE LA PARTIE DE LA MATRICE
C SYMETRIQUE SE TROUVANT AU-DESSUS DE LA DIAGONALE. SI N EST:
C LA DIMENSION DE LA MATRICE, LA DIMENSION DU TABLEAU A EST :
C  $N*(N+1)/2$ .
C 2. B: TABLEAU DANS LEQUEL SONT RANGEES EN COLONNES LES VECTEURS
C PROPRES DE LA MATRICE A. LA DIMENSION DU TABLEAU B EST  $N*N$ .
C APRES L'EXECUTION DU PROGRAMME, LES VALEURS PROPRES DE A SE
C TROUVENT RANGEES SUR LA DIAGONALE DE A DANS UN CERTAIN ORDRE.
C ON TROUVE LE MEME ORDRE DE RANGEMENT DES VECTEURS PROPRES DANS
C LA MATRICE B, C'EST -A-DIRE A LA IEME VALEUR PROPRE SE
C TROUVANT
C SUR LA DIAGONALE DE A CORRESPOND UN VECTEUR PROPRE DONT LES
C ELEMENTS SONT RANGEES DANS LA IEME COLONNE DE B .
C B, EST INUTILE SI KOD = 1, CEPENDANT IL EST NECESSAIRE QUE
C L'ARGUMENT B FIGURE DANS L'INSTRUCTION D'APPEL DU SOUS-PRO-
C gramme
C MSPRO, DANS CE CAS, ON DEFINIT B COMME UN TABLEAU D'UN SEUL
C ELEMENT.
```

C 3. N: DIMENSION DE LA MATRICE
C 4. KOD: PARAMETRE INDIQUANT L'OPTION CHOISIE .SI L'ON DESIRE
C CALCULER SEULEMENT LES VALEURS PROPRES KOD=1:SI L'ON VEUT
C AVOIR EN PLUS LES VECTEURS PROPRES CORRESPONDANTS KOD =0.
C LA VALEUR DE KOD DOIT ETRE PRECISEE DANS LE PROGRAMME APPELANT
C 5. EPS: PRECISION DESIREE.

C

IF(KOD.EQ.1)GOTO 2

I1=-N

DO 1 J=1,N

I1=I1+N

DO 1 I=1,N

IJ=I1 +1

B(IJ)=0.

IF(I.NE.J)GOTO 1

B (IJ)=1.

1 CONTINUE

C

2 RN=0.

DO 3 I=1,N

DO 3 J=1,N

IF(I.EQ.J) GOTO 3

IJ =I+(J*J-)/2

RN = RN +A(IJ)**2

3 CONTINUE

IF(RN.LE.EPS) GOTO 19

RN= SQRT(2.*RN)

RX=RN*EPS/ FLOAT(N)

```

9 IM=M1+I
  GOTO 11
10 IM=I1+M
11 IF(I.GE.L) GOTO 111
  IL=L1+I
  GOTO 12
111 IL=I1+L
12 P=A(IL)*C-A(IM)*S
C   A(IM)=A(IL)*S+A(IM)*C
  A(IL)=P
C 13 IF(KOD.EQ.1) GOTO 14
  ILB=IL 1+I
  IMB=IM1+I
  P=B(ILB)*C-B(IMB)*S
  B(IMB)=B(ILB)*S+B(IMB)*C
  B(ILB)=P
14 CONTINUE
  P=2.*A(LM)*SC
  Q=A(LL)*CC+A(MM)*SS-P
  P=A(LL)*SS+A(MM)*CC+P
  A(LM)=(A(LL)-A(MM))*SC+A(LM)*(CC-SS)
  A(LL)=Q
  A(MM)=P
C 15 IF(M.EQ.N)GOTO 16
  M=M+1
  GO TO 7

```

C

INI =0

SEUIL=RN

4 SEUIL=SEUIL/FLOAT(N)

5 L=1

6 M=L+1

C

7 M1=(M*M-M)/2

L1=(L*L-L)/2

LM=L+M1

IF(ABS(LM)).LT.SEUIL)GOTO 15

INI=1

LL=L1+L

MM=M1+M

P= . 5*(A(LL)-A(MM))

Q=-A(LM)/SQRT(A(LM)**2+P**2)

IF(P.GE.EPS) GOTO 8

Q=-Q

8 S=Q/ SQRT(2.*(1.+(SQRT(1.-Q**2))))

SS=S*S

C= SQRT(1.-SS)

CC=C*C

SC=S*C

C

IL1=N*L-N

IM1=N*M-N

DO 14 I=1,N

I1=(I*I-I)/2

IF(I.EQ.L)GOTO 13

IF(I-M) 9,13,10

```
16 IF(L.EQ.(N-1)) GOTO 17
    L=L+1
    GOTO 6
17 IF(INT.NE.1) GOTO 18
    INT =0
    GOTO 5
18 IF (SEUIL.GT.RX) GOTO 4
C 19 CONTINUE
    RETURN
    END
```