

DECOMPOSITION D'UN ESPACE REGIONAL POUR UNE APPROCHE INTEGREE

Haja ANDRIANASOLO

INTRODUCTION

L'hypothèse de base est qu'une aire régionale quelconque constitue un système complexe organisé, à propos duquel aucune information a priori n'est disponible. C'est un système d'incertitude maximale.

L'approche, sujet de nos recherches actuelles, tendrait à obtenir une réduction de cette incertitude, en réalisant une compréhension "structurelle" d'un tel système.

Démarche qui réduit l'incertitude (l'entropie) en décomposant le système en sous-systèmes "élémentaires" dont il serait plus aisé d'en réaliser l'analyse et la compréhension.

Les problèmes posés sont alors ceux recouverts par la découverte des sous-systèmes, leur qualification et l'appréhension de leurs inter-relations. Soit, que sont les sous-systèmes?, comment sont-ils organisés?, comment structurent-ils la région?.

En fait la première difficulté est de disposer d'une représentation adéquate de cette réalité qu'est l'aire régionale. Si la réduction de l'incertitude visée concerne la totalité des phénomènes présents et influents sur cet espace, la représentation recherchée doit rendre compte de façon synthétique de ces phénomènes. Les données satellitaires constituent une telle représentation. Représentation apte à fournir des éléments d'analyse tant du point de vue de l'analyse spectrale que spatiale.

La donnée satellitaire constitue alors le système d'incertitude maximale, duquel doivent être découverts les sous-systèmes servant à l'analyse.

Cet article se limitera à une approche basée sur les propriétés spectrales des images satellitaires d'une part, et d'autre part est illustré par une problématique "Sécurité Alimentaire", qui a plus particulièrement conduit à la définition des contraintes de l'efficacité et de l'adéquation du modèle. Les

données utilisées sont des images "Landsat MSS", et le thème est le riz.

I. CONSTRUCTION DE SOUS-SYSTEMES

Générer des sous-systèmes revient à réaliser des partitions sur la représentation que constitue l'image. Partitions qui sont usuellement réalisées suivant deux grands types de démarche: dirigée et non-dirigée.

De par l'hypothèse de départ: "incertitude maximale sur le système", la démarche dirigée ne peut s'appliquer. Qu'utiliserait-on en guise de connaissance à priori?

La seule approche possible est celle non-dirigée. Mais alors quels critères de partition employer pour assurer une efficacité et une adéquation optimale aux objectifs? et comment assurer une interprétation systematisable des classes?.

La recherche d'une solution dans ce sens se fera à travers une démarche modélisatrice.

II. MODELISATION

II.1. Criteres de partitions

Soit:

- une population P
- une variable mesurée x sur les individus de P

On note T la variance totale sur x.

A une partition quelconque de P, correspond une décomposition de T en:

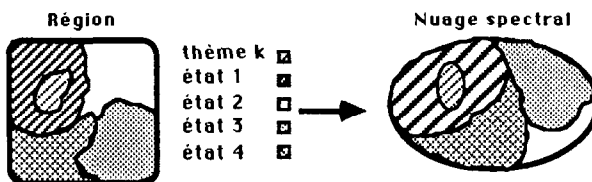
W: variance dans les classes

B: variance entre les classes

$$T=W+B$$

Les images satellitaires rendent en particulier compte et de la nature des phénomènes, et de leurs états. Ainsi par exemple la représentation des phénomènes végétatifs (nature) est en étroite relation avec leur stade phénologique (état).

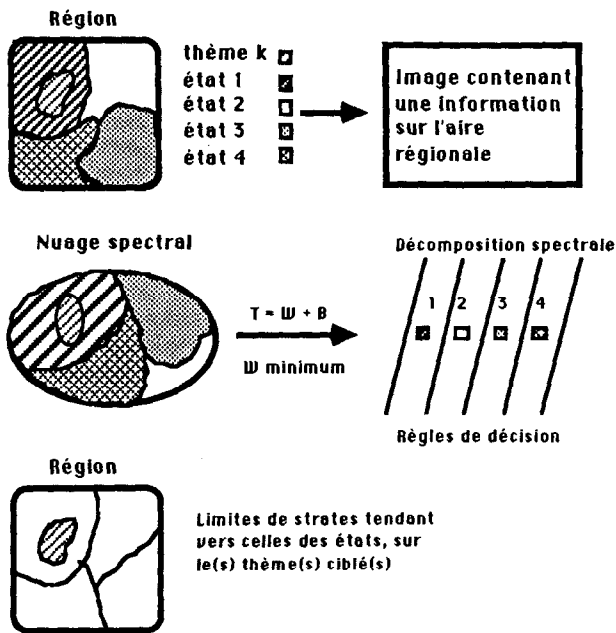
Spectralement les phénomènes se retrouvent plus ou moins regroupés suivant leur état et nature: un thème donné appartient à la partie de l'espace radiométrique de sa classe d'état (voir figure).



Une partition est optimale, si les critères de sa réalisation font effectivement ressortir les différentes composantes du nuage. Ils doivent en découvrir les limites.

La décomposition doit faire ressortir les composantes homogènes du nuage. Un critère apte à découvrir celles-ci est celui de la minimisation de la variance totale à l'intérieur des classes. Plus celle-ci est petite pour une partition en un nombre de classes donné, plus les composantes qui ressortent sont homogènes, et tendent à recouvrir des classes "intrinsèques" de thèmes et d'états.

Le modèle de décomposition de T suivant W et B, est donc tel que W soit minimisée. Modèle dont une réalisation sont les méthodes de classifications par réallocations (Forgy, Jancey...).



Schématisation:

II.2. Variables

Indépendamment du modèle de classification, la minimisation du rapport $\frac{W}{T}$ est aussi fonction des variables utilisées pour la description des individus. Le traitement des images satellitaires reposent souvent, de façon empirique, sur les canaux originaux, les indices, les axes factoriels, et les ratios.

Tel que l'approche est réalisée ici, les variables adéquates doivent exacerber les différences entre les diverses structures présentes dans les données: capacité à maximiser la variance entre les classes.

Ce qui équivaut à maximiser dans "l'absolu" le rapport $\frac{B}{T}$. Ainsi modélisées les variables recherchées sont celles d'une fonction discriminante (axes canoniques) sur les classes d'une partition.

La génération de telles variables nécessite donc l'existence d'une partition (dans le cas présent nous nous limiterons à la génération d'une variable).

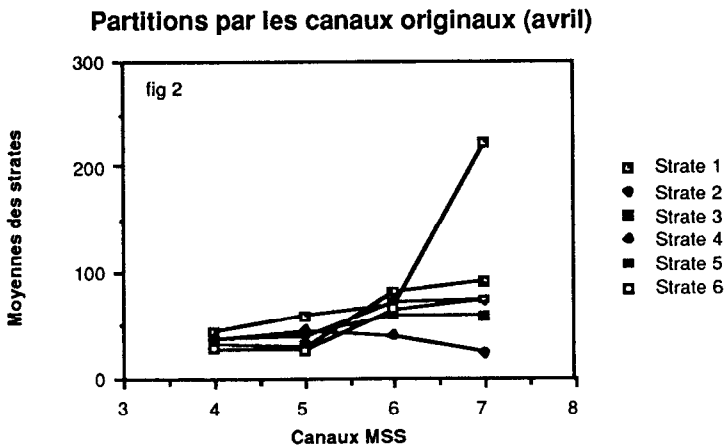
Pour la construction d'une variable ainsi définie, l'existence d'au moins deux classes est nécessaire, tel qu'elles soient les plus différentes possibles au sens du rapport $\frac{B}{T}$. Il suffit donc de construire une partition en deux par le modèle de classification précédemment défini, pour remplir cette condition.

Le modèle général ainsi défini, une autre catégorie de paramètres a une importance capitale quant aux résultats: celui lié aux images elles même. Nous illustrerons cette importance à travers un plan d'expérience mettant en évidence le rôle des dates de prise de vues et des variables utilisées.

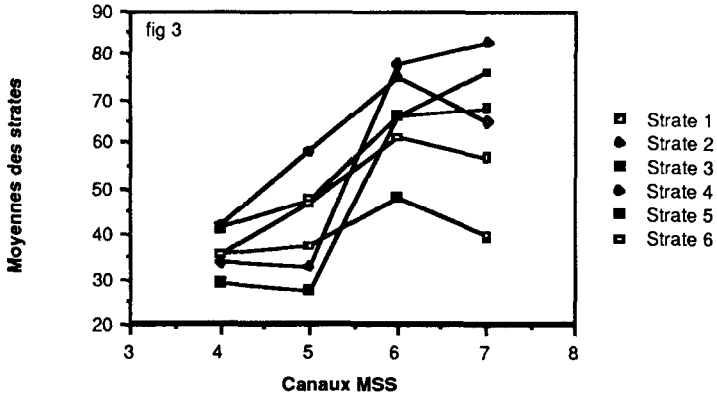
Rappelons que l'adéquation du modèle est vérifiée par les contraintes liées à la stratification pour l'estimation de surfaces rizicoles. L'estimation des W est donc réalisée à partir de la variance de l'extension du riz dans les classes. Une telle vérification sur un thème fixé et unique, n'est pas restrictive de la validité du modèle, puisqu'ainsi que déjà démontré, les partitions sont réalisées par des regroupements de thèmes suivant leur état et nature.

Une partition en six est réalisée, tel que sont comparés les résultats obtenus sur la base des différentes variables et images.

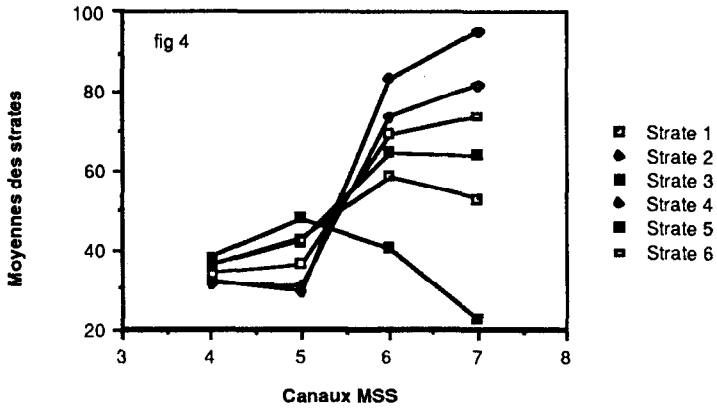
Les figures suivantes illustrent le fait que les partitions obtenues diffèrent suivant les variables utilisées.



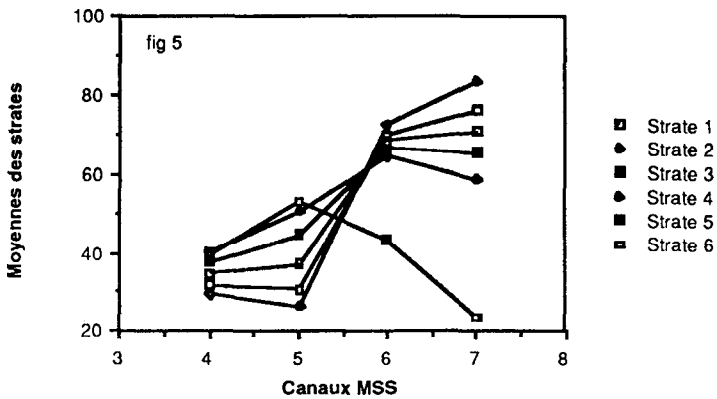
Partition par les axes factoriels (avril)



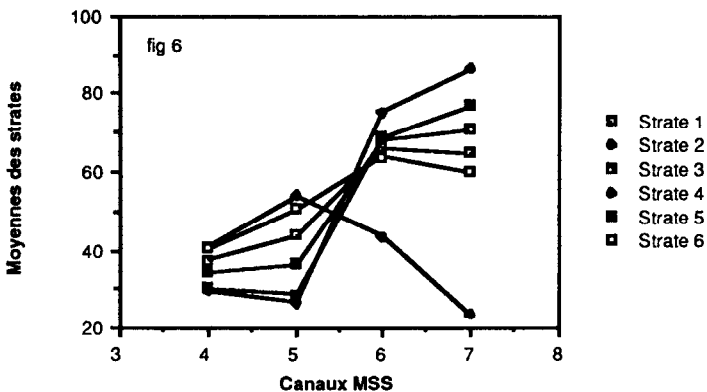
Partition par l'axe unique (avril)



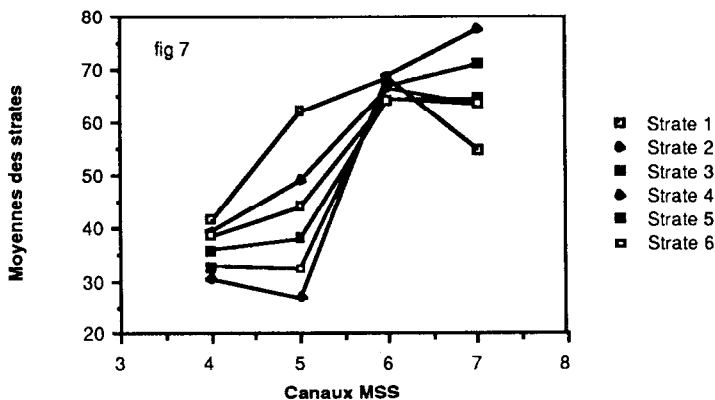
Partitions par indice de végétation verte



Partition par indice végétation jaune

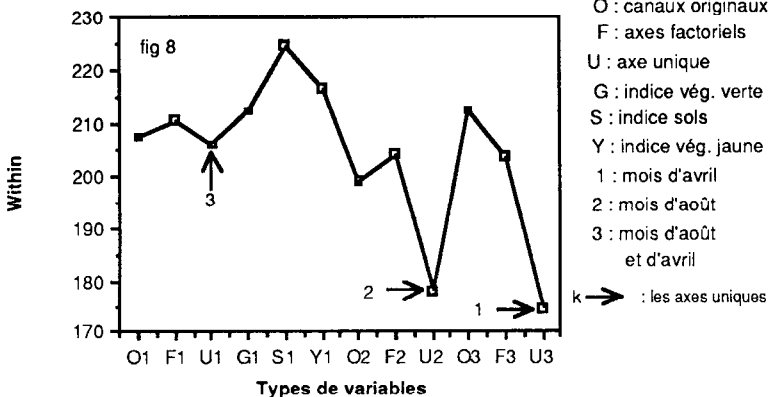


Partition par indice des sols (avril)



Une synthèse des résultats, quant aux aptitudes des différentes combinaisons des données de base est représentée sur la figure suivante.

Variations intra (W) des strates



Il apparait bien que la variable telle que modélisée minimise dans l'absolu (relativement aux autres données de comparaison) W aussi bien sur la date d'avril que d'août, et que d'avril et d'août. Dans le cas présent il s'avère que l'utilisation des deux dates (U3) simultanément amène peu par rapport à la seule date d'août (U2): saison sèche, période de maturation du riz.

Les critères de partition ainsi acquis, comment aborder la réduction de l'incertitude sur une aire régionale?

III. PRINCIPES DE L'ANALYSE

Résolu le point de la génération de sous-systèmes, restent les problèmes liés à la reconnaissance des classes de sous-systèmes dont ils sont issues, et à l'établissement des inter-relations existant entre eux.

Les hypothèses de base de l'approche sont les suivantes:

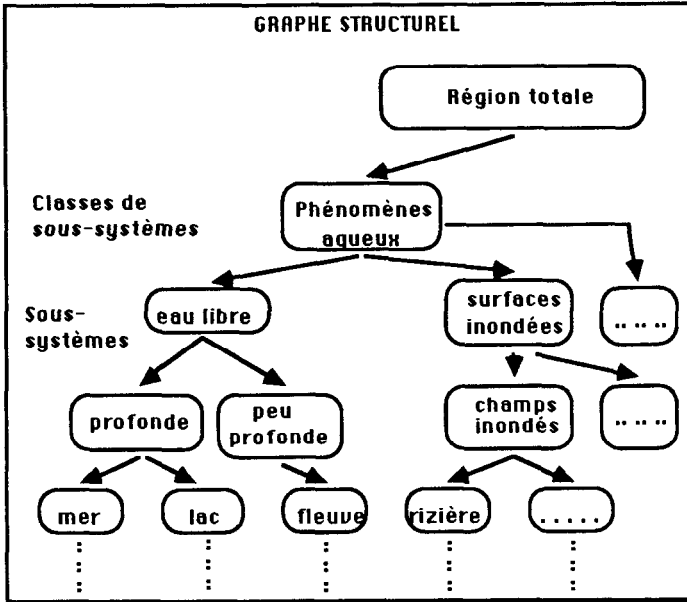
- l'univers est indivise dans son organisation et dans son fonctionnement
- le détail de l'étude que l'on peut en faire est lié à l'échelle et au niveau de nomenclature auquel on se situe
- L'aire régionale peut donc être vue et comprise à différents niveaux. Comme elle est une réalité complexe et organisée, ses différents niveaux, en sus des relations entre leurs sous-systèmes constitutifs propres, ont des relations entre eux. Puisque ceux sont les mêmes réalités que l'on regarde à des niveaux différents.

La compréhension (réduction de l'entropie) de cet univers va donc se faire à travers une démarche, une décomposition hiérarchisée de celle-ci.

III.1. Graphe structurel

(Nous rappelons que les démarches présentées ici sont en cours de recherches et ne sont donc pas figées, en particulier il en est ainsi du vocabulaire).

Le principe de partition tendant à découvrir des classes suivant l'état et la nature des phénomènes posé, la hiérarchisation des niveaux de finesse concernera une vision de plus en plus détaillée des états, jusqu'à arriver idéalement à chaque thème élémentaire (mais qu'est donc un thème élémentaire?) présent dans l'univers. Une représentation de cette hiérarchisation est ce que nous appellerons graphe structurel, dont un exemple est fourni par la figure ci-après.



Sur cette représentation, un niveau constitue un ensemble de sous-systèmes et le niveau qui lui est supérieur un ensemble de classes de sous-systèmes.

Les questions qui se posent sont ceux de la construction d'un tel graphe, et de son interprétation. D'autant qu'un des objectifs de nos recherches est la systématisation et l'automatisation de cette interprétation.

III.2. Hierarchisation

Le modèle de représentation utilisé à travers les données satellitaire est celui recouvert par le domaine spectral. Dans ce cas il ressort de tout ce qui précède, que chaque sous-système appartient de manière récursive à un sous-système de niveau supérieur, jusqu'à l'appartenance à la totalité régionale.

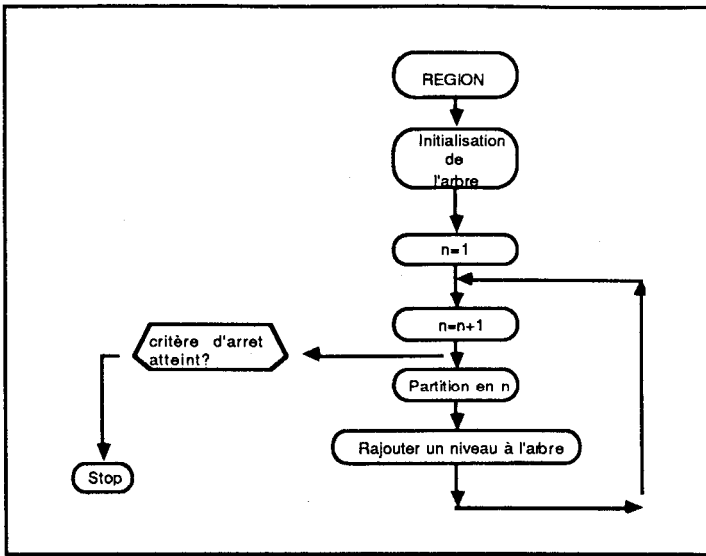
L'arbre structurel doit donc traduire cette structuration, qui pourrait-on dire est la "grammaire naturelle" de cette sorte de représentation.

Il doit montrer une résolution de détail de plus en plus fine, suivant des segmentations dont la règle d'évolution soit l'apparition progressive des structures les plus fortes en sous-systèmes (décomposition de plus en plus fine du nuage spectral).

Chaque niveau de l'arbre sera alors constitué par une partition.

Une partition à un niveau k donné, comporte k classes, et ainsi de suite jusqu'au regroupement de l'ensemble des sous-systèmes pour former la totalité de l'aire régionale.

L'algorithme d'élaboration en serait le suivant:

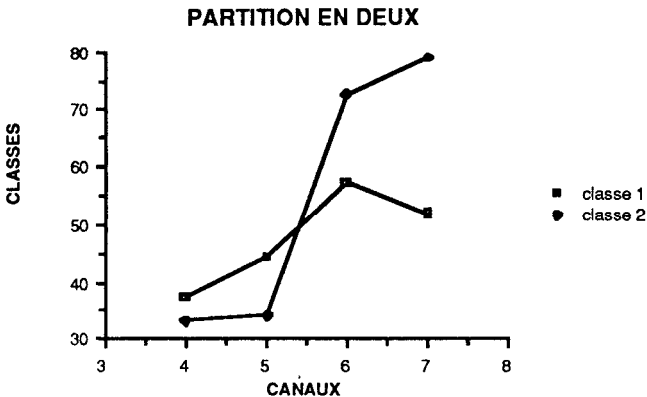


Dans l'application illustrant notre propos, des partitions successives de deux à six classes ont été générées, dont suivent les caractéristiques:

Une étude succincte permet de dégager les comportements suivants: (lecture: (1.2) signifie: la classe 1 de la partition en 2)

Partition en deux:

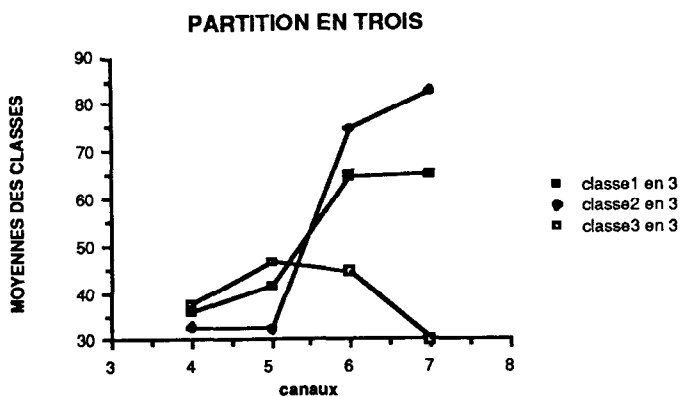
La première classe (1.2) regroupe les phénomènes globalement humides: lacs, fleuves, marais, terres agricoles inondées, savanes et steppes plus ou moins humides....La seconde classe (2.2) regroupe les phénomènes où dominent les thèmes végétatifs (chlorophylle) et secs, tels la forêt, certaines savanes ou steppes....



Partition en trois:

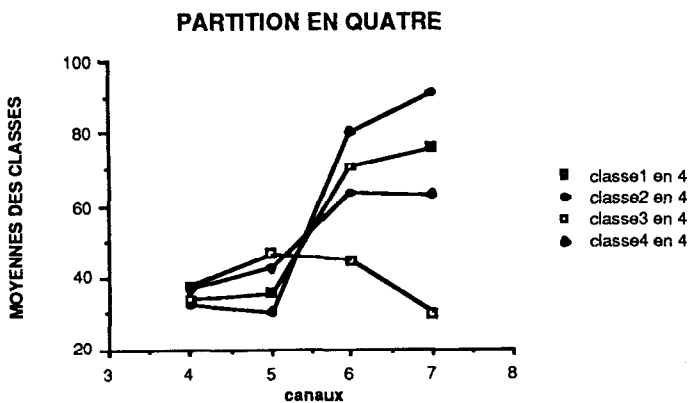
Subdivision du thème global humide (1.2) du niveau précédent en deux sous-thèmes: tout ce qui est vraiment inondé (1.3) d'une part, et le reste (2.3) d'autre part. La classes regroupant les thèmes végétatifs et secs (3.3) de-

meurrant globalement la même.



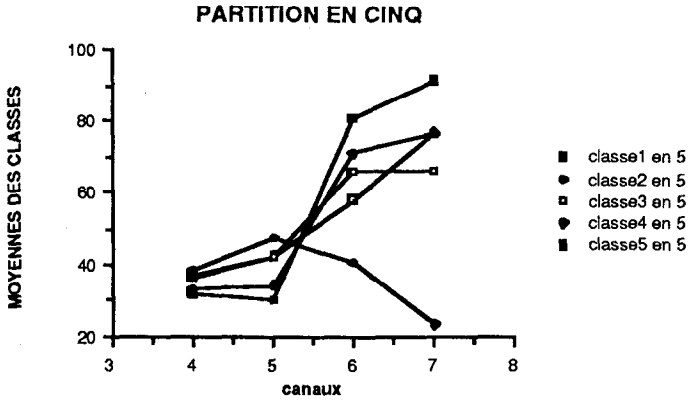
Partition en quatre:

Subdivision de la classe végétative et sèche (3.3) en deux sous-classes: d'un coté la masse forestière (1.4) et de l'autre les savanes et steppes plus ou moins sèche, et plus ou moins arborée (2.4).



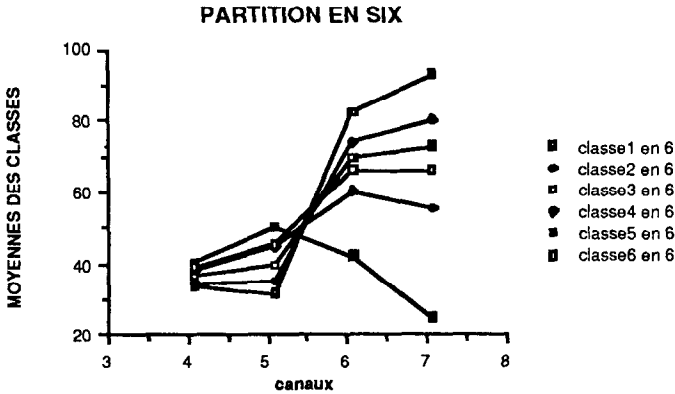
Partition en cinq:

Subdivision de la classe (2.3) en deux sous-classes: d'un coté des phénomènes contenant des traces d'humidité (1.5), et de l'autre des phénomènes où la tendance est vers une domination du végétatif sur l'humide (2.5).



Partition en six:

Subdivision de la masse forestière (1.4) en deux sous-classes (1.6) et (2.6).



Mais une fois constituées ces partitions, comment les relier entre elles pour élaborer l'arbre structural proprement dit? et permettre ainsi une réduction de plus en plus importante de l'incertitude sur la région?.

CONCLUSION

Ces premiers résultats et expérimentations nous confortent dans nos hypothèses et notre approche. En particulier les principes ici exposés ont donné lieu à de premières utilisations pour l'estimation de superficies rizicoles à Madagascar, l'étude de l'écosystème en région aride (Egypte), l'analyse et le diagnostic d'un paysage sahélien (Burkina Faso), dont les résultats par ailleurs exposés ont fait montre d'une efficacité certaine.

Cependant bien des questions sont encore à résoudre, et nous tenons à souligner que cette présentation préliminaire ne constitue qu'une introduction. Ainsi n'ont pas été abordé les problèmes de l'élaboration de l'arbre lui même, ni de celui de sa lecture et de son interprétation.

Tout au plus pouvons nous signaler que la résolution de ces problèmes, dans une optique d'opérationnalité, nous oriente actuellement vers le test de solutions intégrant sur les systèmes experts, utilisés conjointement avec des systèmes de gestion de base de données (SGBD) localisées (SIG).

ELEMENTS BIBLIOGRAPHIQUES

- ANDRIANASOLO H. Analyse statistique des données de télédétection - Statistiques agricoles - Application sur Madagascar. Doctorat en Mathématiques Appliquées aux Sciences Sociales - Paris 1987.
- ANDRIANASOLO H. Agricultural statistics and space images in Madagascar - Estimation of second rice season acreage in a sub-prefecture. Congress of ERIM, Nairobi 1987.
- ANDRIANASOLO H. and MALEK C. Définition et détermination des unités dynamiques d'un paysage sahélien (OURSIL) pour un diagnostic des facteurs de dégradation et de régénération des ressources naturelles et renouvelables. Congress of ISPRS, KYOTO 1988.
- ANDRIANASOLO H. Modélisation de la stratification d'un espace régional par la télédétection, pour l'estimation de surface. Séminaire Informatique de l'ORSTOM (SEMINFOR), Montpellier 1988.