

ANDRIANASOLO Haja
O.R.S.T.O.M.
L.I.A.-TELEDETECTION
UR 3F Maitrise de la Sécurité Alimentaire

*Réduire l'incertitude sur une région rurale:
une approche relevant de l'estimation de surface,
à travers une modélisation reposant principalement
sur les données satellitaires, sous contrainte d'opérationnalité.*

INTRODUCTION

Un handicap majeur à la conduite d'une gestion et politique de développement conséquentes dans les pays pauvres, est l'absence d'une connaissance fiable et concrète des réalités devant les sous-tendre. Sachant que l'économie de tels pays repose principalement sur le rural et l'agricole, un enjeu essentiel est donc celui de la capacité à réduire au maximum l'incertitude sur l'organisation et le fonctionnement de l'espace rural d'une part, et de quantifier précisément les productions agricoles d'autre part.

Pour cela divers démarches et systèmes d'information, généralement issus des schémas des pays riches, ont été appliqués ou plutôt plaqués aux réalités de ces pays, avec comme conséquence une inflation de coûts (dans la mesure où ils sont chiffrables) et de résultats plus contradictoires les uns que les autres <1>. En fait on peut dire que ces schémas ont globalement échoué, du fait que les concepts et définitions ayant présidé à leur élaboration ne rencontrent pas les spécificités induites par les réalités sociales, économiques, culturelles, géographiques, ... du tiers-monde: les modèles de représentation étaient inexacts. Il est alors nécessaire de s'orienter vers des méthodes reposant au maximum sur des représentations rendant de la façon la plus directe possible les réalités mesurables, physiques et concrètes constituant l'espace. Une représentation qui soit accessible et sur lequel il soit possible de travailler qualitativement et quantitativement.

L'objectif central se situe dans la problématique de la maîtrise de la sécurité alimentaire, et tend à l'élaboration d'un système d'information obéissant à ces préceptes, et capables de fournir des éléments pour la réduction de l'incertitude, sur la totalité des phénomènes présents et influents sur l'espace, et d'aboutir à une estimation quantifiée de l'extension spatiale de postes précis d'une nomenclature. Une représentation adéquate doit rendre compte de façon synthétique de l'ensemble de ces phénomènes.

De par leurs caractéristiques (elles rendent compte de façon exhaustive et synoptique des états et types de phénomènes, de leur forme, géométrie, extension, localisation et organisation) les données satellitaires constituent une telle représentation. Elles seront le point de départ, le système d'incertitude maximal, le modèle du prototype de la réalité (ramenée à la manifestation et l'inscription dynamique dans les états de surface de l'action de l'ensemble des phénomènes inter-agissant à son niveau), dont il faut réaliser le potentiel. Ce qui nécessite la définition de tout un système d'information qui les intègre.

O.R.S.T.O.M. Fonds Documentaire

N° :

B 30050

Cote :

ex 1

21.6.90 PL III

Pour répondre réellement à la demande et aux attentes des pays du tiers-monde toute solution se doit de répondre à des contraintes précises, relevant des coûts d'exploitation dans le cadre d'une opérationnalité réelle et routinière. Ce qui dans le cas présent sera ramené à une contrainte centrale d'efficacité, définie comme étant le rapport de la fiabilité et de la précision des résultats à la somme des coûts. Au niveau de la définition du système, sont principalement inclus dans les coûts, ceux qui ont une influence directe sur l'exploitation: les enquêtes sur le terrain, l'achat des images, le temps de réponse. Sont exclus ceux qui se rapportent globalement aux investissements initiaux ou de grande périodicité, tels les équipements (informatique...), l'acquisition du savoir scientifique et technique, la mise en place du système d'information... Au niveau de cette contrainte, à défaut de disposer d'éléments précis (qui ne peuvent que ressortir de directives planificatrices précises) pouvant servir de base à une étude d'optimisation (recherche opérationnelle...), le modèle tel que présenté ici tendra à minimiser systématiquement chaque poste de coûts.

L'hypothèse de base étant que l'incertitude initiale est maximale -niveau de connaissance zéro (il va sans dire que la mise en oeuvre réelle intègre l'utilisation de données exogènes)- le modèle comporte deux grandes parties: organisation de l'espace et estimation de surface, intégrées dans une démarche statistique globale.

Sont successivement présentés:

- le modèle statistique générale et son articulation
- la méthode de structuration globale de l'espace
- la méthode d'estimation de surface (stratification, reconnaissance d'un thème, validation, extrapolation)

au travers desquels sont posés les problèmes d'échantillonnage d'images "optimales", et de leur traitement en adéquation avec les objectifs.

La région d'application se situe dans le nord-ouest malgache (province de Mahajanga) et porte sur une sous-préfecture (Marovoay). Elle porte sur la capacité à fournir une estimation de l'extension du riz de seconde saison (irriguée).

MODELE STATISTIQUE GENERAL ET CONTRAINTES D'OPTIMISATION

DESCRIPTION DU MODELE

Soit:

- une aire S
- une culture quelconque i

l'objectif est d'estimer l'extension de i dans l'aire S .

Définition:

- on appelle segment une portion de territoire, aux limites bien définies (E. Houseman, 1979) <2>.

Dans les méthodes d'échantillonnage d'aires, la surface totale S est subdivisée en segments. Un segment constitue l'unité élémentaire d'échantillonnage.

Soit:

- N le nombre total de segments de l'univers
($N=S/s$, tel que s: aire d'un segment)
- E un échantillon de m segments ($m < N$)
- j un segment quelconque

Par enquête on obtient:

$x(i,j)$ la surface de i dans j faisant partie de E

Qui permet l'obtention de la superficie moyenne de i par segment:

$$\text{Moy}(x(i)) = 1/m \cdot \sum_{j=1}^m x(i,j)$$

Un estimateur de l'extension totale de i dans S est par expansion directe:

$$\text{Est}(x(i)) = N \cdot \text{Moy}(x(i))$$

Dont la variance est:

$$\text{Var}(\text{Moy}(x(i))) = (N - m)/(N - 1) \cdot 1/(m(m - 1)) \cdot \sum_{j=1}^m (x(i,j) - \text{Moy}(x(i)))^2$$

D'autre part étant démontré que tout découpage de l'univers en strates est source de gain, dans le cas de ce modèle il suffit de considérer chaque strate séparément et de sommer les résultats sur toutes les strates.

Ainsi si S est découpé en H strates, on aurait si h représente une strate:

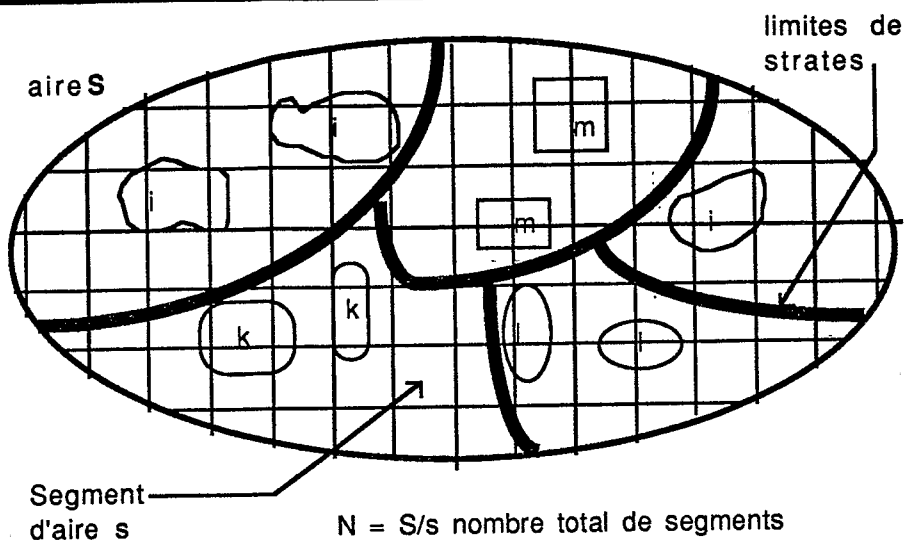
$$\text{Moy}(x(i,h)) = \sum_j x(i,h,j)/m(h)$$

$m(h)$ est le nombre de segments de h
 $x(i,j,h)$ est l'extension de i dans j de h
 $\text{moy}(x(i,h))$ est l'extension moyenne de i par segment de h

L'estimateur de l'extension de i est alors:

$$\text{Est}(x(i)) = \sum_{h=1}^H N(h) \cdot \text{Moy}(x(i,h))$$

IDENTIFICATION du MODELE d'ESTIMATION STATISTIQUE (Sondage aréolaire)



E un échantillon de M segments

$x(i,j)$: surface de i dans j, j appartenant à E

$$\text{Moy}(x(i)) = \frac{1}{M} \sum_{j=1, M} x(i,j)$$

$\text{Est}(x(i)) = N \cdot \text{Moy}(x(i))$: estimation de l'aire de i dans S

Stratification en H strates

h étant une strate:

$$\text{Moy}(x(i,h)) = \sum_{j=1, M} x(i,h,j)/m(h) \quad \text{tel que : } \begin{array}{l} \bar{m}(h) \text{ nombre de segments de h} \\ x(i,j,h) \text{ aire de i dans j de h} \\ \text{moy}(x(i,h)) \text{ aire moyenne de i} \\ \text{par segment de h} \end{array}$$

Estimateur de l'aire de i:

$$\text{Est}(x(i)) = \sum_{h=1, H} N(h) \cdot \text{Moy}(x(i,h))$$

CONTRAINTES D'OPTIMISATION ET OPÉRATIONNALITE

Malgré son apparente simplicité, la mise en oeuvre pratique et efficace de ce modèle d'estimation exige que certaines conditions soient remplies. En particulier que la variation entre segments soit la plus petite possible (que chaque segment possède tous à peu près les mêmes proportions des phénomènes existant), et qu'ils soient tous à peu près de la même taille. La contrainte de la minimisation des variations inter-segments, exige formellement une partition en strates efficaces de l'aire étudiée. En effet plus les strates sont "pures" du point de vue des phénomènes ciblés, plus les variations entre les segments dans chaque type de strate diminuent. A la limite dans le cas où une strate correspond à un phénomène (culture), il suffirait de ne faire les enquêtes que dans cette strate, pour estimer ce phénomène dans l'univers.

Un point crucial est donc l'élaboration d'un modèle apte à générer des strates tel que leur variance interne soit minimisée, sur le critère mesuré sur les segments.

Un autre point crucial est celui lié aux coûts:

. en effet le modèle ainsi défini requiert de façon routinière une couverture exhaustive de la totalité du territoire étudié: ce qui est prohibitif. Nécessité est donc de le penser tel qu'il ait une capacité à fournir des résultats sur la base d'un échantillon minimal d'images (pas de couverture exhaustive sur l'exploitation annuelle).

. L'un des postes, sinon le poste certainement le plus lourd dans une élaboration routinière d'information, telle qu'envisagée ici, est celui représenté par les opérations de terrain. Sachant que le territoire couvert est potentiellement celui national (ou même transnational dans le cadre de politiques régionales), il est d'une nécessité absolue de le contrôler très rigoureusement à travers une optimisation des déplacements des enquêteurs, ce qui implique une détermination précise et objective des lieux à enquêter, du niveau de précision désiré sur les résultats pour l'estimation du taux d'échantillonnage, et donc du nombre de ces lieux. L'optimisation aura alors comme paramètres principaux la localisation des points d'enquêtes (dispersion géographique et représentativité), le nombre de ceux-ci, l'aire de chacun d'eux, les infrastructures et logistiques existantes, le nombre d'enquêteurs disponibles (ou à disposer), leur capacité d'enquête en unités de temps, et le capital alloué.

Le modèle, sur la base de l'intégration des données satellitaires, de leurs traitements et des autres sources d'information disponibles (données exogènes), doit être à même de fournir un maximum d'éléments à la concrétisation de ces paramètres.

En résumé, la capacité à réduire l'incertitude sur un espace rural à travers l'estimation de surface de termes d'une nomenclature, se ramène à l'optimisation du rapport précision sur coût, par l'élaboration d'un modèle d'approche de l'espace comportant plusieurs niveaux de perceptions imbriqués, dont les paramètres s'estiment sous contraintes opérationnelles. Contraintes qui se traduisent par diverses étapes et problèmes d'optimisation:

.pour un capital donné, comment allouer les différents types de données satellitaires aux objectifs des différents niveaux du modèle? Les variables utilisées sont celles recouvertes par les attributs espace total couvert exhaustivement, résolution spatiale, résolution spectrale et types de bandes du spectre électromagnétique utilisés, résolution temporelle des prises de vue, et la résolution temporelle de disponibilité effective (le laps de temps entre les fourchettes de dates de disponibilité utiles au modèle, et la disponibilité effective dans les circuits commerciaux).

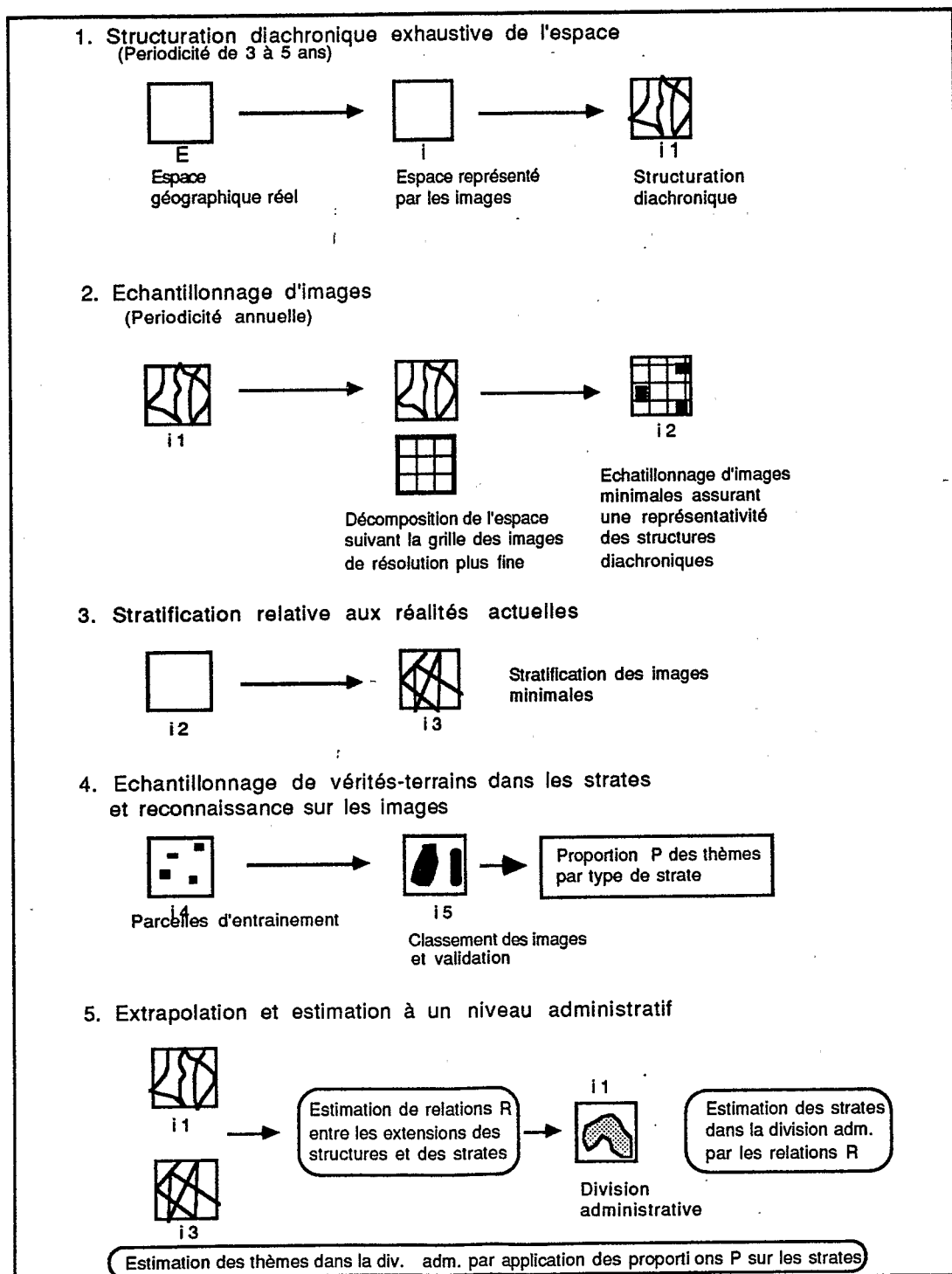
.Pour un capital temps global donné, comment le décomposer et l'allouer aux différentes phases dynamiques du modèle, pour une capacité à un temps de réponse en adéquation avec les objectifs? Ceci sachant que chacune de ces phases constitue en soi un modèle à optimiser.

.Pour un capital "ensemble des traitements possibles des images" et "totalité des données exogènes", comment assurer la maximisation de la précision des résultats à travers un enchaînement optimal et systématique de ceux-ci? de telle façon que chaque élément soit exploité au maximum de ses possibilités, sans gachis.

Pour un capital enquêteurs, comment obtenir un maximum de résultats terrains optimaux (sans élément de subjectivité), sous les contraintes de leur capacité d'enquête?

Le présent article expose les grandes lignes d'un modèle, qui répond de façon générale à l'ensemble de ces points, en tendant à chaque étape à maximiser la précision des résultats tout en minimisant les coûts d'obtention. Dans le cadre d'une exploitation routinière réelle, les différents paramètres du modèle seraient précisément évalués, et sa mise en oeuvre s'inscrirait dans des stratégies répondant à des objectifs économiques précis: utilisation d'un certain volume de main d'oeuvre...

SCHEMA GENERAL



STRUCTURATION EXHAUSTIVE ET DIACHRONIQUE DE L'ESPACE

Ainsi que le montre le schéma général, l'aptitude du modèle à fournir des résultats statistiques de façon opérationnelle, en quelque localisation que ce soit, repose sur une structuration préalable de l'espace. Les données satellitaires constituant un modèle de représentation, leurs traitements adéquats peuvent fournir une telle structuration <3>. Structuration qui une fois obtenue conservera une validité temporelle dont une périodicité de révision serait de l'ordre de trois à cinq ans, puisqu'à moins de grandes catastrophes naturelles, l'évolution de l'espace par rapport à la problématique n'est perceptible qu'à ces échelles de temps.

Par ailleurs mis à part son rôle purement statistique de base d'extrapolation, ce type de résultat qui est en fait une approche des modes d'organisation, de fonctionnement et d'évolution de l'espace, constitue une représentation intégrante et synthétisante de la résultante de tous les facteurs influençant celui-ci. Son étude d'un point de vue multidisciplinaires et systémique pourrait s'avérer riche d'enseignement.

Une structuration de l'espace à partir d'une approche modélisatrice de donnée satellitaire prise à une date et de leurs traitements étant acquise <3>, il s'agit plus précisément dans le cas présent d'obtenir un tel résultat en s'affranchissant de la relativité de celui-ci aux moments des prises de vue. Il doit rendre compte de l'espace sur le continuum du temps. Les images étant prises à des instants discrets, la démarche sera diachronique tel que les données sont traitées indépendamment les unes des autres, l'organisation "intemporelle" de l'espace étant obtenue par superposition spatiale de tous les résultats ainsi générés <8>, <9>.

Une telle démarche induit une contrainte sur l'attribut date de prise des images: il faut qu'elles traduisent le maximum de variation de la vie de l'espace sur l'ensemble d'un cycle annuel. Le climat jouant dans ce cadre un rôle primordial (il influe directement sur les activités humaines, animales, végétales... et donc sur les variations d'aspect de l'espace), il faudrait à priori que les images se répartissent sur les saisons les plus contrastées.

S'agissant d'organiser exhaustivement l'espace, une contrainte se pose aussi au niveau des attributs résolutions spatiale et spectrale des images. Il faut d'une part qu'elles couvrent la plus grande superficie possible, qu'elles permettent des études jusqu'à des échelles géographiques de l'ordre du $1/250000^e$, et que d'autre part elles soient capables d'assurer une bonne discrimination des différents phénomènes végétatifs et minéraux. On parle ici de données à résolution moyenne.

Dans le cadre de la présente application deux images issues des capteurs Landsat MSS (environ 80 mètres de résolution spatiale, et 4 canaux de prises de vues se répartissant du bleu au proche infra-rouge) sont utilisées, prises en fin de saison des pluies et en milieu de saison sèche.

TRAITEMENT D'UNE IMAGE

STRUCTURES POTENTIELLES

L'espace en tant que tel et à travers ses éléments de surface constitue une énorme masse d'information. Une représentation d'un aspect particulier de cette information, à travers les images satellitaires se fait par individualisation de zones, suivant des critères en cohérence avec les objectifs. Aspect particulier, car suivant le niveau (la résolution) de l'information recherchée, les formes des zones, l'organisation même des structures tendront vers une apparence de plus en plus morcelée et complexe. Ce qui est la normalité des choses tel qu'on le retrouve à travers la théorie des fractals <16>, ou d'une autre manière de la notion d'échelle en géographie.

Dans les perspectives des objectifs méthodologiques ici développés, deux questions doivent être à priori résolues:

a). Etant donné qu'on dispose d'un critère <4> apte à découvrir les structures (partitions) les plus fortes de l'espace en un nombre quelconque k fixé de classes, un tri doit être réalisé parmi toutes ces partitions. En effet même si elles sont toutes justifiées du point de vue du critère, de celui des réalités structurales de l'espace proprement dit, certaines d'entre elles ne le sont pas. L'organisation qu'elles font ressortir tiennent moins des réalités que du critère de partition.

Rappelons que le critère de partition est tel que pour un nombre k de classes, la décomposition de la variance dans la relation d'analyse de celle-ci ($T=W+B$, T variance totale, W variance à l'intérieur des classes, B variance entre les classes) tend vers une minimisation de W , corrélativement une maximisation de B (rendre les classes les plus homogènes possibles, et les plus hétérogènes possibles entre elles).

Une multitude de partitions peut donc être générées, dont le nombre P est théoriquement borné par le nombre d'individus traités. Ce qui dans la pratique et sous la contrainte "temps de réponse" de l'efficacité, est irréalisable. Aussi toutes les structures potentielles seront elles générées jusqu'à concurrence de P' obtenu suivant la formule de décomposition d'histogrammes:

$$P' = 5 \log(n), \text{ tel que } n \text{ est l'effectif d'un échantillon représentatif, dans le cas présent à un niveau de confiance de 95\%}$$

Comme une image de l'espace par les données satellitaires est en fait de nature multidimensionnelle (une image Landsat comporte 4 canaux), une valeur P' est calculée pour chaque dimension, et est utilisée celle qui est le supérieur ($\text{Sup}(P')$).

Sont donc successivement calculées toutes les partitions, de 2 à P' classes, intégrant différents niveaux d'information.

STRUCTURES REELLES ET STRUCTURES DEGENEREEES.

A l'issue du processus précédent, sont donc obtenues:

$$K = \sum_{k=2}^{P'} (k) \text{ somme des entiers de 2 à } P'$$

classes individualisées, tel qu'au niveau de partition qu'elles représentent, elles identifient à la fois les structures les plus homogènes (W minimum) et les plus différentes entre elles (B maximum).

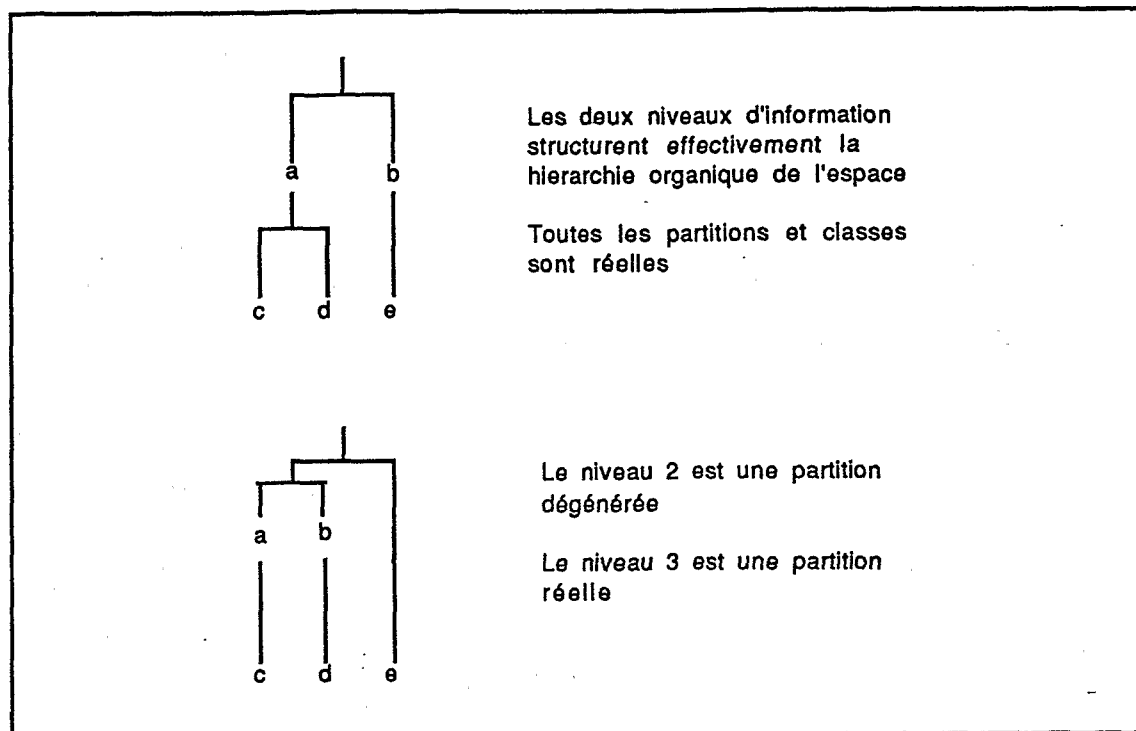
Sachant d'une part que les classes d'une partition donnée constituent au sens de l'homogénéité, la structuration la plus forte en ce nombre de classes, et d'autre part que chacune de ces structurations n'a qu'une *portée locale* relative à son niveau, une capacité de lecture structurelle et organique de l'espace, requiert de chacune des structurations une *portée globale*. Ce qui constitue une *contrainte supplémentaire en vue de l'élaboration d'une grammaire* <14> <15>.

En fait sachant que l'existence des k classes de toutes les partitions résulte du critère de minimisation de la variance à un *niveau local*, cette contrainte exige que cette minimisation soit aussi effective au *niveau global*.

Dans ce cadre, la compréhension intrinsèque de l'organisation de l'espace à un niveau quelconque de détail, est assurée par une approche structurelle et organique de celui-ci. A savoir que les composantes de base expliquant les classes d'un niveau, sont constituées par les classes de niveaux inférieurs. Une telle compréhension passe alors par une représentation hiérarchique des réalités tel que traduites par le modèle, chaque niveau de la hiérarchie étant organisé par les classes correspondant à son niveau de partition. *Dans l'élaboration d'une telle hiérarchie structurelle, si une partition s'avère inapte à organiser son niveau, sa puissance "locale"*

d'organisation ne résiste pas aux contraintes d'une organisation "globale", on dira qu'elle est dégénérée. Dans le cas contraire, on dira qu'elle est réelle, car maintenant sa structuration au niveau de toute la hiérarchie organique. La représentation hiérarchique ainsi obtenue sera appelée "graphe structurel".

Par exemple supposons que l'on ait généré 2 partitions, en 2 et 3 classes. Les deux cas de figures qui peuvent apparaître sont alors les suivants:



CRITERE DE HIÉRARCHISATION

Sachant qu'une partition quelconque équivaut à une décomposition de la variance totale T , et que le critère de classification utilisé est tel que dans cette décomposition, la variance W tend à être minimisée, ce même critère doit aussi gouverner la hiérarchisation.

Dans cette perspective, partant initialement de la totalité des classes, à chaque niveau de la hiérarchie doivent être agglomérées les deux classes dont la réunion minimise la variance totale à l'intérieur des classes formées. Alors les k classes d'une partition en k , si une telle partition n'est pas dégénérée, devraient se retrouver organisant la hiérarchie au niveau k . A l'issue de ce processus, partant des structures potentielles, sont obtenues les structures réelles, au sens de leurs capacités organisatrices sur le critère d'homogénéité interne (d'hétérogénéité inter).

b). La deuxième question qui se pose est qu'étant donné que toutes ces structurations sont réelles et ne se différencient que par leur niveau de détails, laquelle utiliser?

En fait une telle question ne devrait pas être posée, car relevant d'un point de vue ramenant à une description statique de la réalité, alors même que le résultat obtenu -le graphe structurel- constitue une vision dynamique et fonctionnelle de l'espace (vision fonctionnelle dont l'exploitation doit intégrer la nature du modèle spectral de la réalité par les données satellitaires: deux objets de nature différente peuvent avoir une même signature spectrale, et un même type d'objet peut se traduire par plusieurs signatures. L'exploitation du graphe structurel dépasse le cadre de cet article et fera l'objet de prochaines publications).

Deux types de réponse sont cependant possibles à cette question. Soit l'utilisateur décide à priori du niveau de détail auquel il mène son étude, soit il se donne un critère "objectif" pour

désigner une "meilleure" partition. Dans le premier cas le problème est résolu. Dans le second cas nous partirons du fait que *chacun des niveaux du graphe n'est qu'une représentation d'un état particulier de l'existence de la réalité (tel que perçu par le modèle), et qu'en tant que tel peut lui être associée une probabilité*. Probabilité qui doit refléter la nature réelle ou dégénérée de la structure, ainsi que sa puissance d'organisation, traduisant la puissance de la réalité à exister dans cet état.

CRITERE "OBJECTIF" DE SELECTION D'UNE PARTITION.

La méthode de partition utilisée <7> est celle dite des "centres mobiles avec initialisation par tirage au hasard des individus centraux". Ce qui entraîne que pour un même nombre de classes sur un même corpus de données, et quoique la convergence de cette démarche soit démontrée, différentes exécutions conduisent théoriquement à des résultats différents. En fait ces différences se situent aux marges des classes. En conséquence, plus une structure en un nombre fixé de classes est faible, plus des fluctuations seront observées au niveau des classes, induisant des variations au niveau des variances W et B (décomposition de T). Une étude de W (ou de B) sur plusieurs exécutions de l'algorithme en un nombre fixé de classes, indique alors la capacité structurante de ce nombre de classes. En effet plus W (ou B) fluctue, moins est forte cette capacité structurante.

Nous définissons la variable aléatoire "Niveau de partition" associée aux événements "2 à P' nombre de classes", dont la distribution de probabilité est tel que:

quelque soit un événement k,

$$Prob(x=k) = 1/n \cdot \sum_{i=1}^n (1/(EC_i(W_e)+1)) \cdot s$$

tel que:

k est le nombre de classes du niveau de partition

s = 1 si la partition est réelle

s = 0 si la partition est dégénérée

EC_i est l'écart-type de la variable W (variance dans les classes) sur e réalisations de celle-ci, sachant que la probabilité de l'événement (x = k) est calculée sur n réalisations de cet écart-type EC_i.

Cette variable aléatoire est tel que plus une structure a tendance à exister, plus sa probabilité est forte. Il est alors possible d'établir une relation d'ordre entre les structures, en fonction de leur puissance d'organisation.

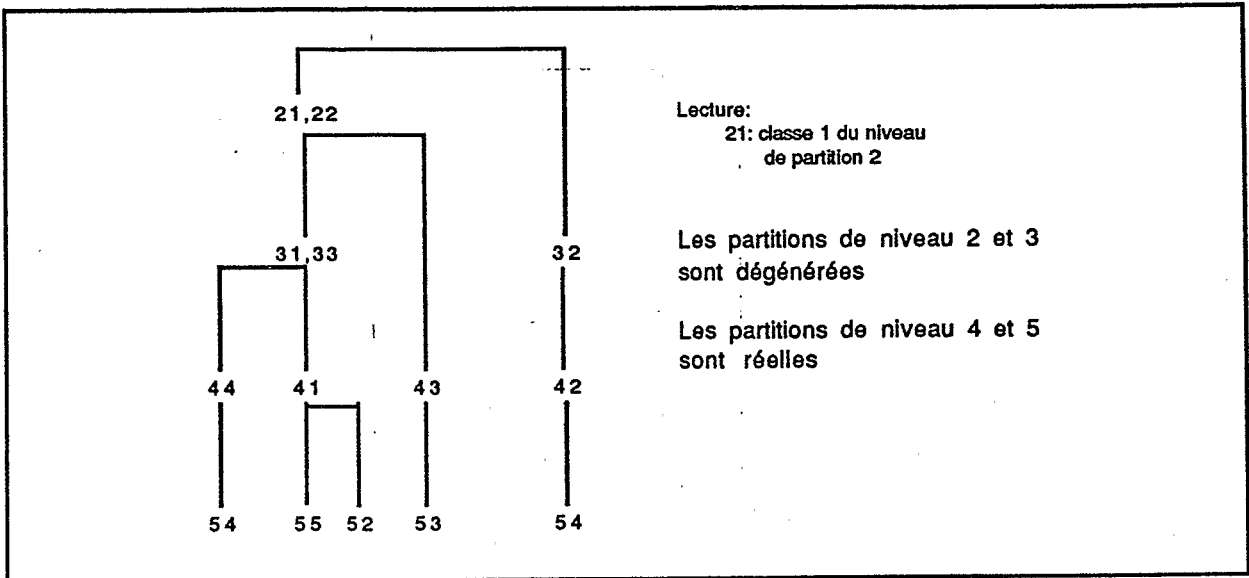
Par ailleurs une étude de la distribution de ces probabilités permet de les comparer effectivement, et de leur attribuer des niveaux de probabilités.

APPLICATION : STRUCTURATION DE L'ESPACE.

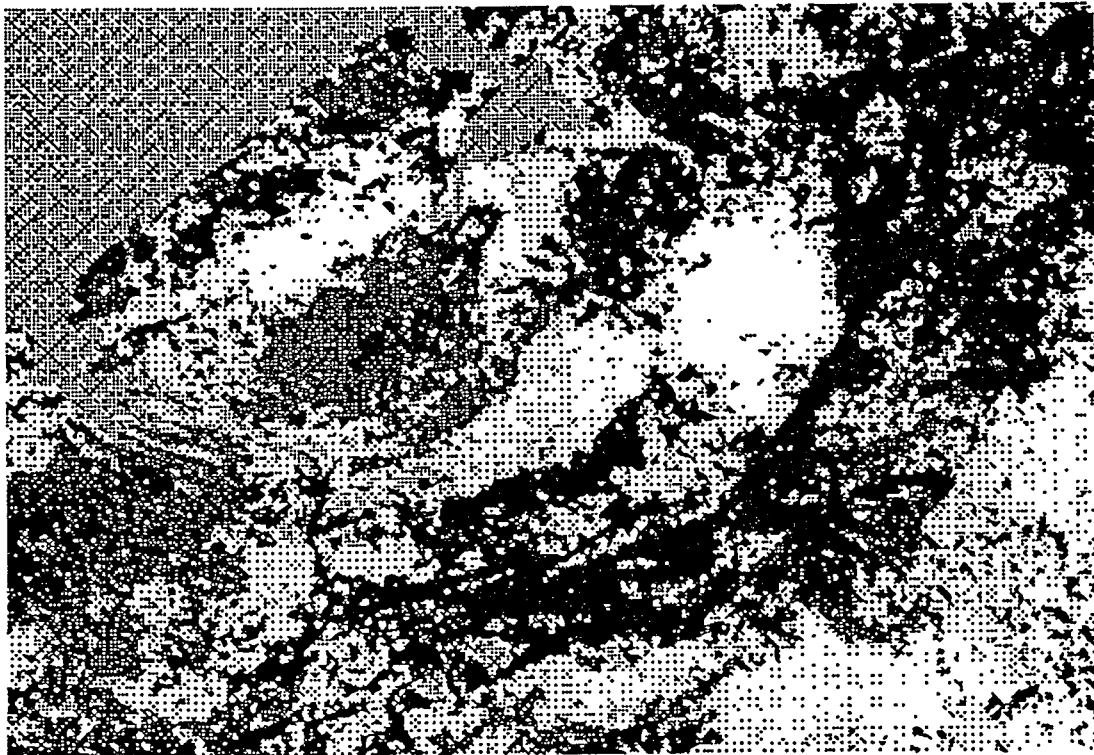
Image 1: fin de saison des pluies, 24 avril 1981.

L'effectif d'un échantillon représentatif pour un niveau de confiance de 95% (moyenne de la population se trouvant dans un intervalle de 1 autour de la moyenne estimée) est de 3537 pixels. La règle de décomposition de l'histogramme donne $P=17$ classes maximales.

L'arbre de hiérarchisation est représenté jusqu'au niveau 5 de partition.



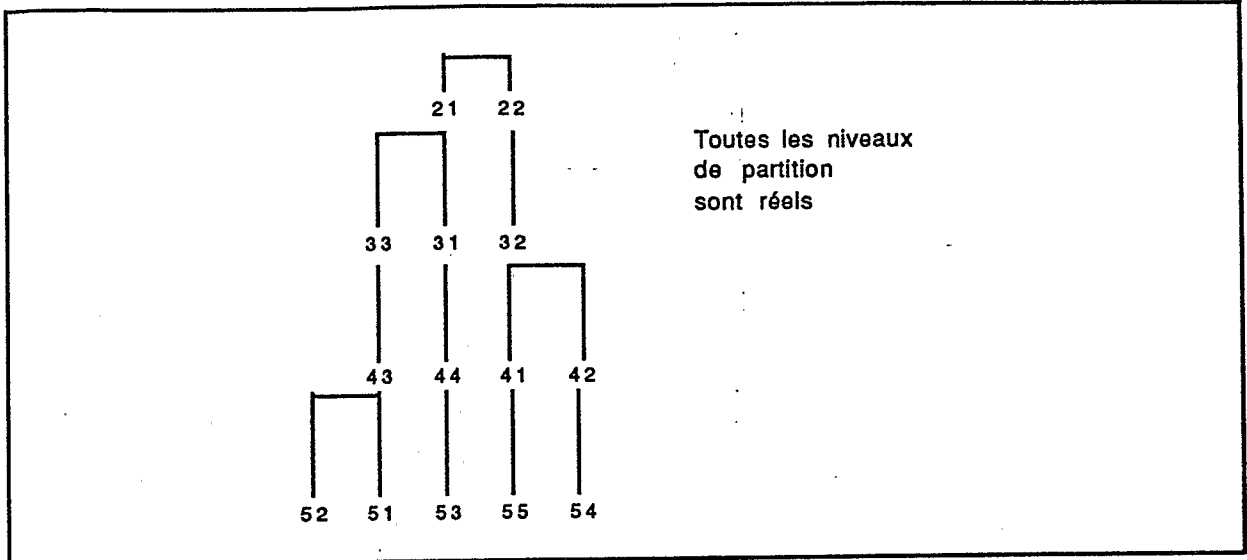
La structure en quatre possède la plus forte probabilité d'existence.



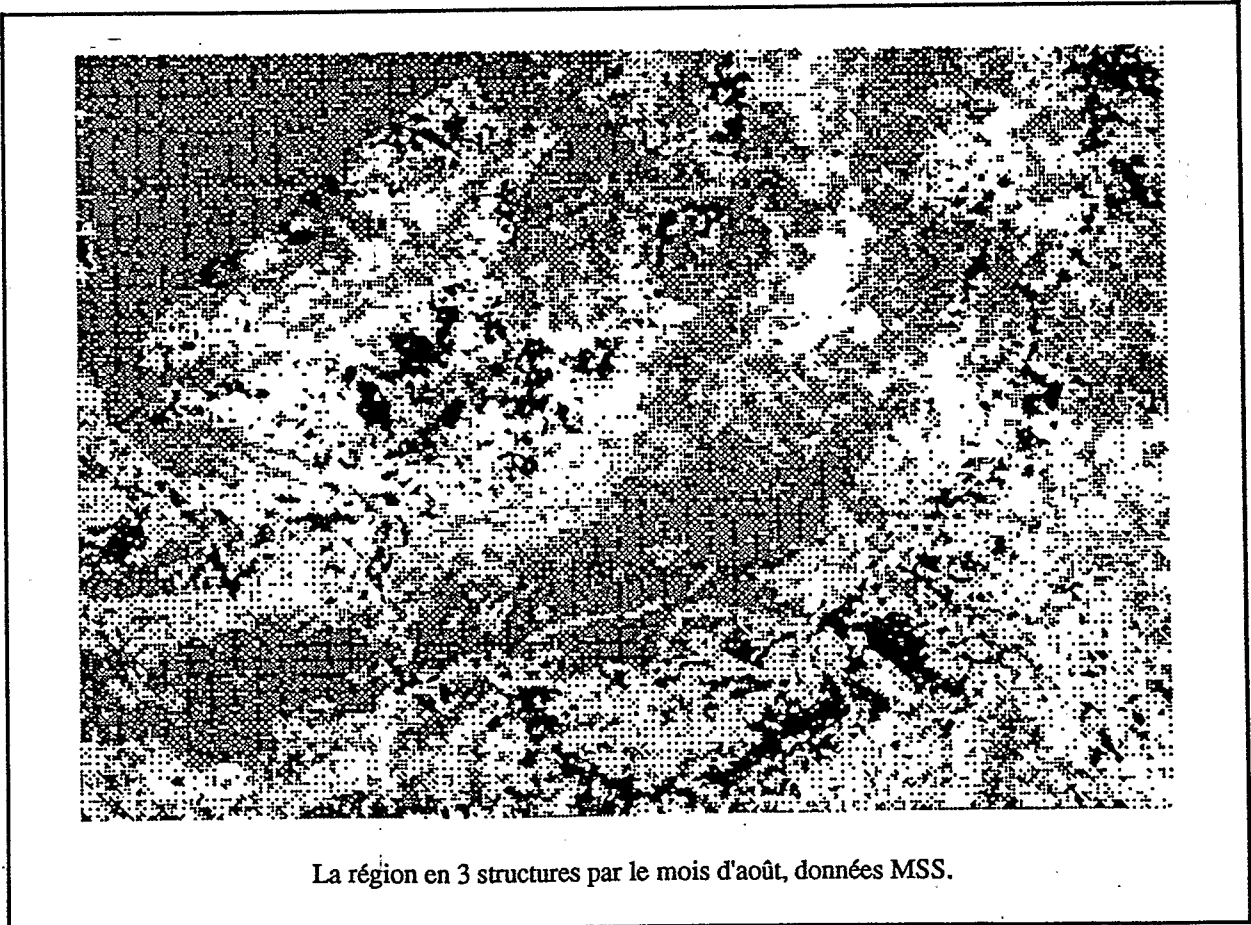
La région en 4 classes par le mois d'avril, données MSS

Image 2: date de milieu de saison sèche, 28 août 1981.
 Taille de l'échantillon: 3500 pixels.
 Nombre maximal de classes: 17

Arbre de hiérarchisation jusqu'au niveau 5 de partition.

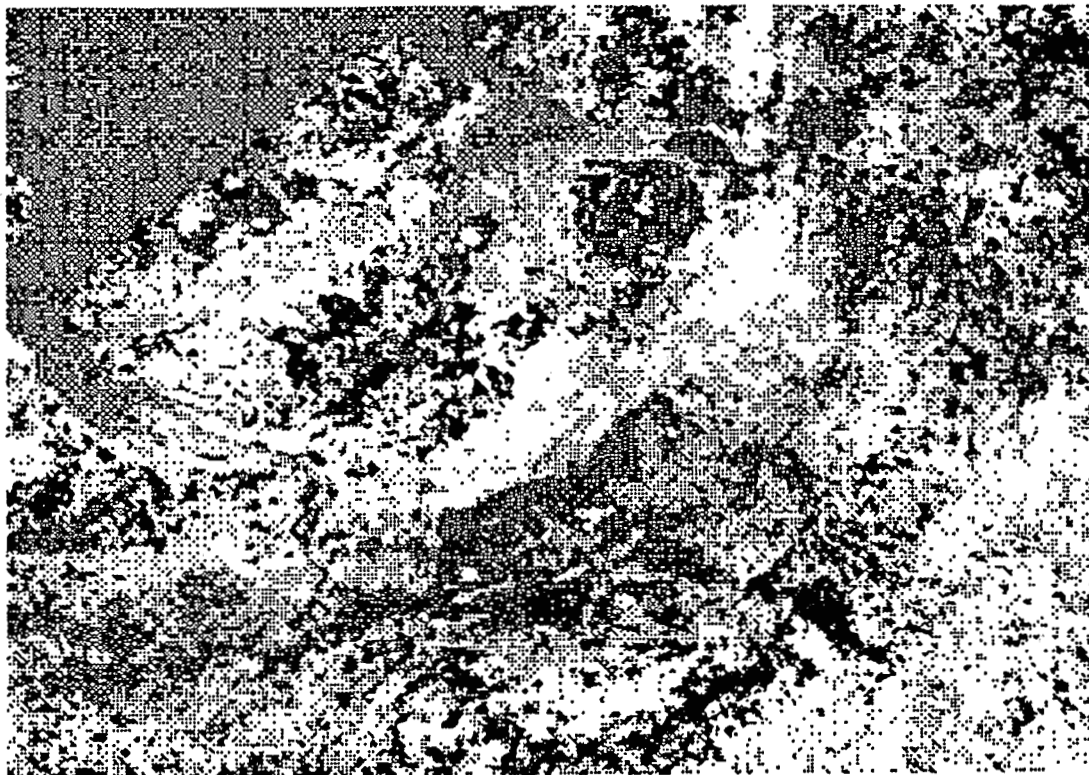


La structure en trois possède la plus forte probabilité d'existence.



STRUCTURATION DIACHRONIQUE.

Elle est obtenue par superposition des structurations obtenues aux deux dates. Le nombre de classes résultat est égal au produit du nombre de classes de chacune des dates.



Structuration diachronique en 12 types

ÉCHANTILLONNAGE D'IMAGES A HAUTE RESOLUTION

L'exploitation annuelle requiert l'achat d'images de résolutions fines qui puissent permettre et de stratifier la région étudiée, et de reconnaître les thèmes suivis. Obeissant à la règle de l'efficacité, il ne peut être envisagé l'achat de l'ensemble des images couvrant exhaustivement ladite région. Un échantillon minimal d'images représentatives doit alors être constitué.

On dispose des deux ensembles suivants:

- I=(scènes-images)
- U=(structures)

tel que:

- cardinal de I=n
- cardinal de U=k
- chaque élément de U est caractérisé par:
 - .le nombre d'éléments de I qu'il possède
 - .l'extension de chacun de ces éléments

La représentativité que l'échantillon sous-ensemble de I doit assurer, est relative dans le cas le

EXTENSION DES STRUCTURES DIACHRONIQUES DANS LES SCENES SPOT

A priori n'importe laquelle des 9 scènes Spot couvrant l'aire régionale étudiée conviendrait: il suffirait d'en tirer une aléatoirement. Cependant la thématique de l'application (riz de seconde saison) exige l'utilisation d'une scène, telle que l'on puisse disposer sur le terrain d'aires d'entraînement et de validation, qui soient accessibles. Incluant cette contrainte, l'image Spot de trace et rang 166/382 est sélectionnée car comportant une partie des grandes plaines rizicoles de Marovoay.

STRATIFICATION.

La stratification a pour rôle d'accroître la précision des résultats. Pour cela elle doit être efficace, et donc avoir un fort rapport de corrélation avec la variable étudiée. Une précédente publication <4> nous a déjà permis d'établir une modélisation de stratifications efficaces par les images satellitaires, et de démontrer que pour cela il fallait que les valeurs des images sur les attributs localisation spatiale, dates de prises de vue, critères de partitions, et variables utilisées (canaux et néo-canaux) devaient respecter des contraintes induites par le modèle et ses objectifs. Les problèmes de la localisation spatiales et des traitements ont déjà été résolus au niveau des paragraphes précédents.

DATE DE PRISE DE VUE.

Une contrainte sur la date de prise de vue, inhérente à l'objectif de stratification par les images satellitaires, est que les phénomènes existant dans la région doivent être dans des états les plus différents possibles. Ce qui nécessite une étude précise de leur calendrier.

Celui-ci étant constitués d'instant élémentaires, son étude pour la détermination de dates efficaces doit aboutir à la définition de classes de ces instants élémentaires, caractérisées par la présence et l'état des phénomènes de la région. Pour qu'une telle étude soit le plus riche d'enseignements, il faut pouvoir aussi disposer d'une organisation de ces classes d'instant élémentaires entre eux. Pour cela, on soumet à une classification ascendante hiérarchique, les individus constitués par les instants élémentaires, et décrits par les phénomènes et leurs états <10>.

Au niveau de l'attribut résolution spatiale des images à utiliser, la contrainte se ramène à une capacité à distinguer au mieux entre eux les phénomènes. Comme de plus on ne cherche pas ici une couverture exhaustive (échantillonnage minimal) l'extension spatiale des images n'est pas une contrainte. Aussi des images Spot (20 mètres de résolution spatiale, et 60 kilomètres carré de couverture exhaustive) ou Thematic Mapper (30 mètres de résolution spatiale et 90 kilomètres carré de couverture exhaustive) conviennent. Dans la présente étude des images Spot ont été utilisées.

APPLICATION: LOCALISATION TEMPORELLE.

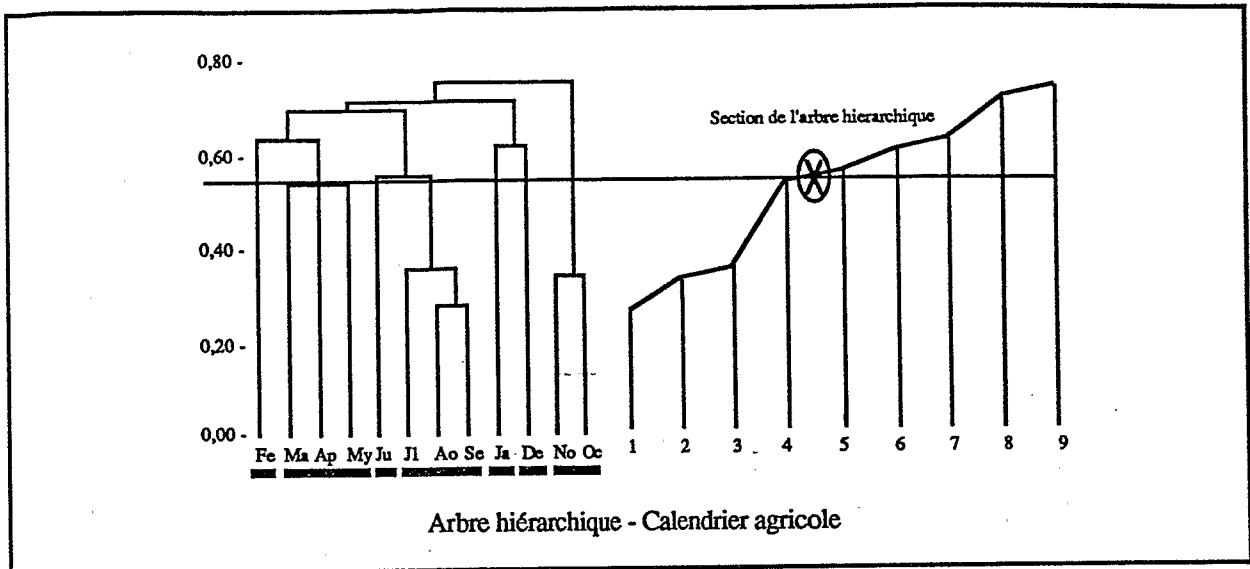
Travaillant au niveau de cultures, le calendrier soumis à analyse est celui du calendrier agricole:

- . Instant élémentaire: le mois.
- . Etats des cultures: préparation du sol/semis, plantation/repiquage, entretien, récolte.

Les données se présentent sous la forme présence/absence (1/0): quand une culture donnée est dans un état donné, il est codé à 1 (0 sinon).

TRAVAUX	JAN	FEV	MAR	AVR	MAI	JUIN	JUILL	AOU	SEP	OCT	NOV	DEC
SEMIS												
Arachide saison1	1	0	0	0	0	0	0	0	0	1	1	1
Arachide saison2	0	0	1	1	1	0	0	0	0	0	0	0
Coton	0	0	0	1	1	0	0	0	0	0	0	0
Maïs saison 1	0	0	0	0	0	0	0	0	0	0	1	1
Maïs saison 2	0	0	0	1	1	1	0	0	0	0	0	0
Manioc	0	0	0	0	0	0	0	0	0	1	1	0
Riz saison 1	0	0	0	0	0	0	0	0	0	0	0	1
Riz intermédiaire	1	0	0	0	0	0	0	0	0	0	0	0
Riz saison 2	0	0	1	1	0	0	0	0	0	0	0	0
Tabac	1	1	1	1	0	0	0	0	0	0	0	0
REPIQUAGE												
PLANTATION												
Arachide saison 1	0	0	0	0	0	0	0	0	0	0	0	0
Arachide saison 2	0	0	0	0	0	0	0	0	0	0	0	0
Coton	0	0	0	0	0	0	0	0	0	0	0	0
Maïs saison 1	0	0	0	0	0	0	0	0	0	0	0	0
Maïs saison 2	0	0	0	0	0	0	0	0	0	0	0	0
Manioc	0	0	0	0	0	0	0	0	0	0	1	0
Riz saison 1	1	1	0	0	0	0	0	0	0	0	0	1
Riz intermédiaire	0	1	0	0	0	0	0	0	0	0	0	0
Riz saison 2	0	0	0	0	1	0	0	0	0	0	0	0
Tabac	0	1	1	0	1	1	0	0	0	0	0	0
ENTRETIEN												
Arachide saison 1	0	1	1	0	0	0	0	0	0	0	0	0
Arachide saison 2	0	0	0	0	0	1	1	0	0	0	0	0
Coton	0	0	0	0	0	1	1	1	1	0	0	0
Maïs saison 1	1	1	0	0	0	0	0	0	0	0	0	0
Maïs saison 2	0	0	0	0	0	0	0	0	0	0	0	0
Manioc	0	1	1	1	0	0	0	0	0	0	0	0
Riz saison 1	0	1	1	0	0	0	0	0	0	0	0	1
Riz intermédiaire	1	0	1	0	0	0	0	0	0	0	0	0
Riz saison 2	0	0	0	0	0	1	1	1	0	0	0	0
Tabac	0	0	0	0	0	0	0	0	0	0	0	0
RECOLTE												
Arachide saison 1	0	0	0	1	1	1	0	0	0	0	0	0
Arachide saison 2	0	0	0	0	0	0	0	1	1	0	0	0
Coton	0	0	0	0	0	0	0	0	0	1	1	0
Maïs saison 1	0	0	1	1	1	0	0	0	0	0	0	0
Maïs saison 2	0	0	0	0	0	0	1	1	1	0	0	0
Manioc	0	0	0	0	0	1	1	1	1	0	0	0
Riz saison 1	0	0	0	0	1	1	0	0	0	0	0	0
Riz intermédiaire	0	0	0	0	0	1	1	0	0	0	0	0
Riz saison 2	0	0	0	0	0	0	0	0	1	1	1	0
Tabac	0	0	0	0	0	0	1	1	1	1	0	0

Calendrier agricole



ANALYSE GLOBALE

Une première étude globale de l'arbre hiérarchique permet de vérifier à la lumière de la réalité, la validité ou non de la classification obtenue. En cheminant du haut vers le bas, suivant une démarche désagréative, différents regroupements apparaissent :

- à un premier niveau, les mois d'octobre et novembre se désolidarisent des autres, ce qui traduit les faits que :
 - c'est l'instant charnière entre saison sèche et saison des pluies, puisque c'est en octobre que s'achève la sécheresse.
 - comme les cultures et leurs calendriers ne sont pas indépendants des saisons, ces deux mois correspondent au fait qu'en eux se situe la majeure partie des changements d'états de l'occupation du sol. En effet des récoltes s'y achèvent, des terrains sont mis en jachère et d'autres sont semés, plantés ou préparés pour des cultures à venir, arrivant avec les pluies.
- A un deuxième niveau, un autre groupe composé des mois de décembre et janvier est constitué. C'est le premier moment des fortes pluies de décembre pendant et après lesquelles sont mises en place les cultures pluviales de saison humide.
- A un troisième niveau, deux grands groupes se forment comprenant d'un côté les mois de février, mars, avril, et mai, et d'un autre ceux de juin, juillet, août et septembre. Globalement :
 - le premier groupe correspond à la deuxième moitié de la saison des pluies, moment où sont semées, plantées ou repiquées les cultures de saison sèche ou irriguée, et où arrivent à maturation les cultures pluviales de saison humide.
 - le second groupe correspond à la pleine saison sèche, moment où arrivent à maturation et sont récoltées les cultures irriguées de saison sèche.

Cette première analyse globale révèle la validité de la classification. En effet sont révélées ici des structures recoupant les saisons globales et agricoles connues d'une année.

CONSTITUTION DE CLASSES DE MOIS ET DATES OPTIMALES

La section de l'arbre hiérarchique au quatrième palier, suivant le critère de l'inflexion maximale <12> génère les 7 classes suivantes:

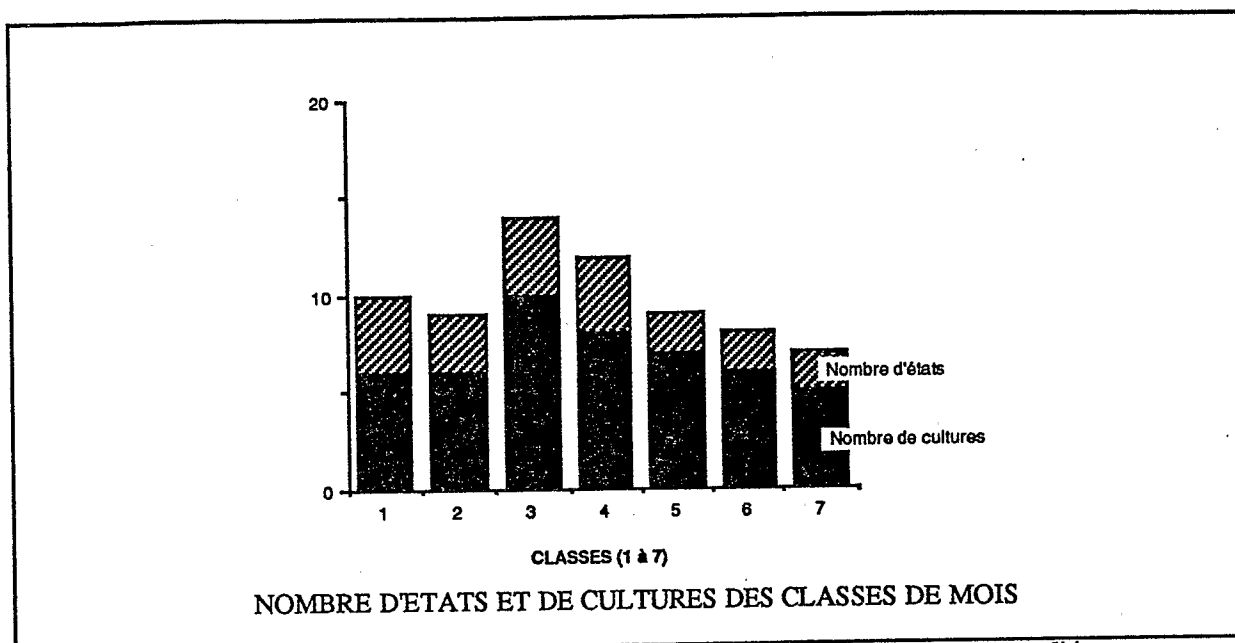
- classe 1 = janvier / classe 2 = février / classe 3 = mars, avril, mai / classe 4 = juin
- classe 5 = juillet, août, septembre / classe 6 = octobre, novembre / classe 7 = décembre

Pour disposer de résultats précis quant aux capacités des différentes classes de mois, celles-ci doivent être qualifiées par la contribution de chaque culture et état à leur formation. Ce qui permettra alors de les ordonner suivant le nombre de cultures et d'états différents.

CULTURES	CLASSES						
	1	2	3	4	5	6	7
SEMIS							
Arachide saison 1	0	0	0	0	16,6	20	20
Arachide saison 2	0	12,5	0	0	0	0	0
Coton	0	8,3	0	0	0	0	0
Maïs saison 1	0	8,3	0	0	0	20	10
Maïs saison 2	0	0	11,1	0	0	0	0
Manioc	0	0	0	0	0	0	20
Riz saison 1	0	0	0	0	0	20	0
Riz intermédiaire	0	0	0	0	16,6	0	0
Riz saison 2	0	8,3	0	0	0	0	0
Tabac	14,2	8,3	0	0	16,6	20	0
REPIQUAGE-PLANTATION							
Arachide saison 1	0	0	0	0	0	0	0
Arachide saison 2	0	0	0	0	0	0	0
Coton	0	0	0	0	0	0	0
Maïs saison 1	0	0	0	0	0	0	0
Maïs saison 2	0	0	0	0	0	0	0
Manioc	0	0	0	0	16,6	20	0
Riz saison 1	14,2	0	0	0	16,6	0	0
Riz intermédiaire	14,2	0	0	0	0	0	0
Riz saison 2	0	4,1	0	0	0	0	0
Tabac	0	4,1	11,1	0	0	0	0
ENTRETIEN							
Arachide saison 1	14,2	4,1	0	0	0	0	0
Arachide saison 2	0	0	11,1	5,2	0	0	0
Coton	0	0	11,1	15,7	0	0	0
Maïs saison 1	14,2	0	0	0	16,6	0	0
Maïs saison 2	0	0	0	0	0	0	0
Manioc	14,2	8,3	0	0	0	0	0
Riz saison 1	14,2	4,1	0	0	0	0	0
Riz intermédiaire	0	4,1	0	0	0	0	0
Riz saison 2	0	0	11,1	10,5	0	0	0
Tabac	0	0	0	0	0	0	0
RECOLTE							
Arachide saison 1	0	8,3	11,1	0	0	0	0
Arachide saison 2	0	0	0	10,5	0	0	0
Coton	0	0	0	0	0	0	20
Maïs saison 1	0	12,5	0	15,7	0	0	0
Maïs saison 2	0	0	0	15,7	0	0	0
Manioc	0	0	11,1	0	0	0	0
Riz saison 1	0	4,1	11,1	0	0	0	0
Riz intermédiaire	0	0	11,1	5,2	0	0	0
Riz saison 2	0	0	0	5,2	0	0	20
Tabac	0	0	0	15,7	0	0	10

CONTRIBUTION DES CULTURES ET ETATS AUX CLASSES

Sous forme graphique, la mise en évidence de l'intérêt de chaque classe apparait immédiate:



La relation d'ordre entre les classes, suivant la contrainte nombre de cultures et d'états différents est alors:

classe 3 > classe 4 > classe 5 = classe 1 > classe 2 > classe 6 = classe 7

Une image à une date donnée, pour une stratification optimale de cette région, doit à priori être sélectionnée dans la classe 3, soit parmi les mois de mars, avril et mai. Dans cette application une image Spot du 21 mai 1986 est utilisée.

STRATIFICATION

Après traitement selon les critères et démarches précédemment définis, une partition en 10 classes de l'aire recouverte par l'image est obtenue.



Stratification en 10 classes par l'image SPOT

ÉCHANTILLONNAGE VÉRITES-TERRAIN - CLASSEMENT DE L'IMAGE

Une étude préalable des classes permet de les diviser en deux groupes de strates, d'un côté celles susceptibles de par leurs caractéristiques de porter la culture riz de seconde saison, et de l'autre le reste. En effet sachant qu'à cette date les rizières de seconde saison sont inondées, leur comportement spectral ne peut que les ranger dans les classe-strates dont les caractéristiques spectrales sont celles des thèmes humides (rappelons qu'en télédétection les signatures spectrales de grandes thèmes globaux: sols nus, végétation active, eau... sont connus et répertoriés). Ainsi il apparaît que seules quatre classes parmi les dix sont à priori des strates riz, car ayant un comportement caractéristique des phénomènes humides.

L'image satellitaire constituant la représentation sur laquelle on travaille, et quoique sa résolution spatiale induit un changement d'échelle, elle sera encore à la base de l'échantillonnage de vérités-terrain. En particulier pour le nombre de segments à visiter, et leur localisation.

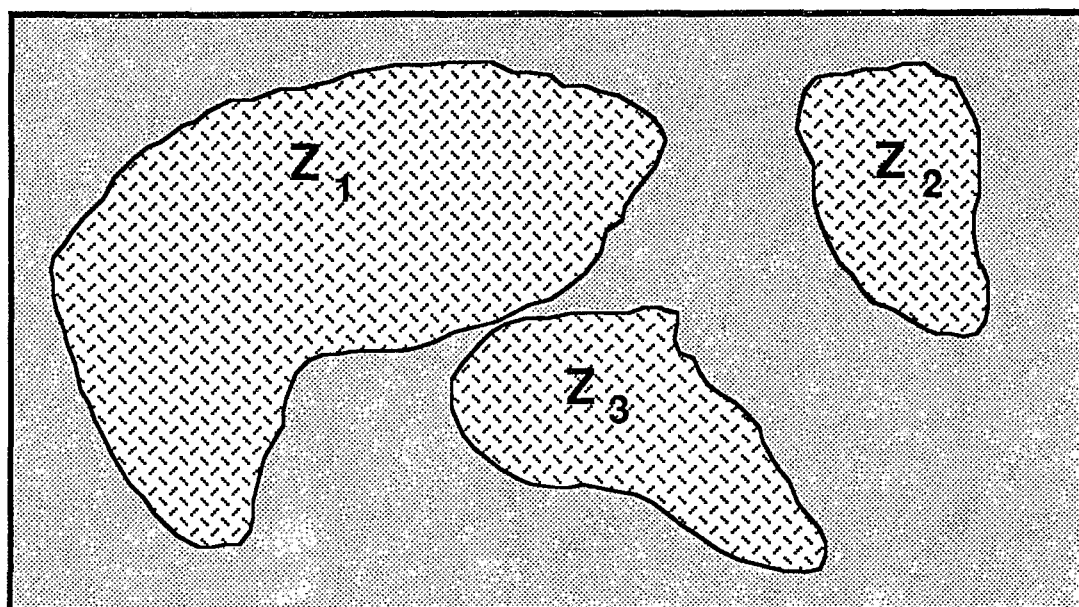
La taille de l'échantillon par strates, en nombre de pixels est calculable selon, soit la règle de sondage stratifié représentatif, soit celle du sondage stratifié optimal (Neyman). Dans le modèle, comme la stratification est obtenue par une démarche d'analyse typologique, la première règle peut s'appliquer <6>. Soit:

$$n_s = N_s \cdot (n / N) \text{ tel que } n_s : \text{taille de l'échantillon de la strate } s$$

N_s : taille de la strate s
 n : effectif total de l'échantillon à constituer
 N : effectif de la population.

Connaître le nombre de pixels à enquêter est insuffisant, car en effet il est impossible d'identifier un pixel sur le terrain. Ceux sont des segments qui sont enquêtés, donc des groupes de pixels. Si la capacité moyenne d'un enquêteur oscille entre 25 et 75 hectares par jour, nous avons adopté 50 hectares comme taille de segment. La taille d'un pixel Spot étant de 0,04 hectares, il en faut 1250 pour arriver à 50 hectares. Le segment sera alors un carré de 35 pixel de côté, soit de 700 mètres de côté sur le terrain. Connaissant le nombre de pixels (n_s) à tirer par strates, la division de celui-ci par 1250 donne le nombre de segments à enquêter.

Sachant que généralement une strate est dispersée sur plusieurs zones, la détermination des segments à enquêter pour la constitution des vérités-terrain, est faite de la façon suivante:



Soit une strate dispersée sur p zones ($p=3$) : z_1, \dots, z_p

- s_x : nombre calculé de segments de z_x
- s_1 : nombre calculé de segments de z_1
- s_2 : nombre calculé de segments de z_2
- s_3 : nombre calculé de segments de z_3
- Nombre total de segments sur la strate : $s = s_1 + s_2 + s_3$

L'échantillon devant être constitué de k segments, tirage aléatoire de k nombres appartenant à l'intervalle $1, s$.

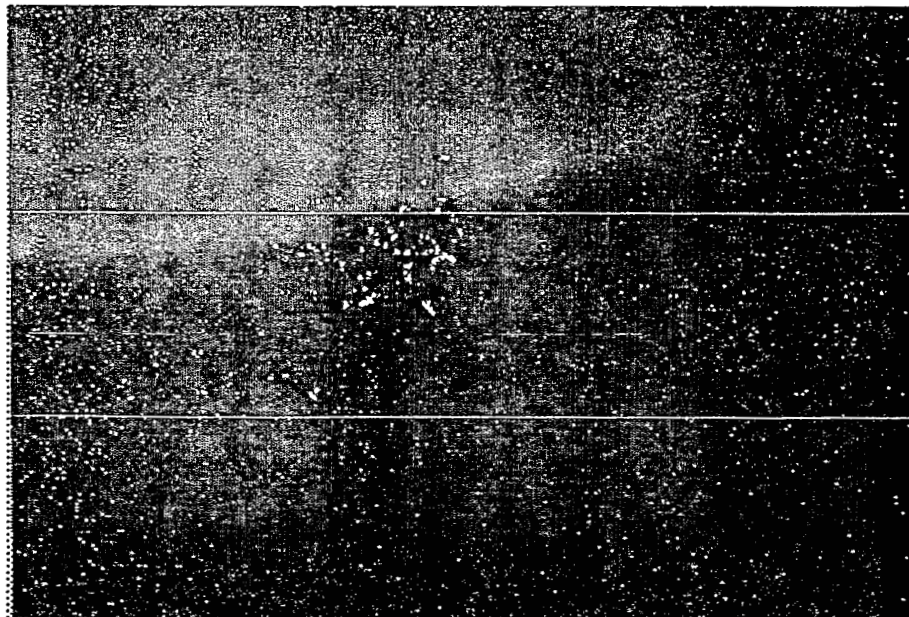
Chaque nombre tiré détermine la zone dans laquelle l'échantillonnage effectif du segment est effectué :

- si $0 < k < s_2$: échantillonnage dans z_1
- si $s_1 < k < s_1 + s_2 + 1$: échantillonnage dans z_2
- si $s_1 + s_2 < k < S + 1$: échantillonnage dans z_3

Le tirage aléatoire du segment lui-même, s'effectuant de la manière suivante:

- tirage d'un point aléatoire dans la zone échantillonnée
- construire une aire rectangulaire de cinquante hectares autour de ce point
- enquêtes de ces segments, dans le cas de cette application, détermination de vérités-terrains.

Dans la pratique cette démarche est à amender suivant les réalités du terrain. Cependant le fait de se tenir au maximum à ces indications assure un fondement statistique aux résultats, et leur confère une fiabilité qualifiable en niveaux de probabilité.



Parcelles d'entraînement du riz

RECONNAISSANCE DE THEMES (RIZ) - CLASSEMENT - VALIDATION

L'ensemble des vérités-terrain obtenu a été divisé en deux groupes, dont les rôles respectifs sont d'une part de réaliser le classement de l'image, et d'autre part de faire une validation de ce classement.

Deux contraintes ont guidé la détermination de la méthode de classement: qu'elle soit non paramétrique, et permette l'existence d'une classe de rejet. Nous avons donc utilisé une démarche de discrimination affectant une probabilité à chaque classe à priori sur la base de la proximité des k-plus proches voisins du pixel à classer.

Principe:

On se munit de la distance euclidienne

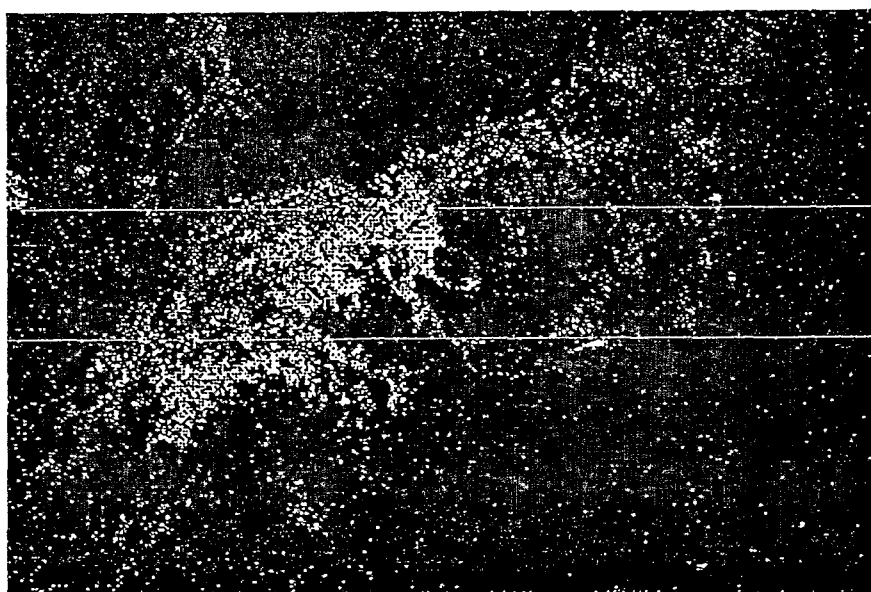
On se donne un seuil de probabilité s

Soit: un pixel quelconque p à classer
un ensemble de n classes d'entraînement

- 1- calcul de la distance euclidienne de p aux individus issus des classes d'entraînements
- 2- détermination des k plus proches voisins de p
- 3- affectation d'une probabilité à chaque classe d'entraînement sur la base de leur fréquence en tant que faisant partie des k -plus proches voisins
- 4- p est affecté à la classe de plus forte probabilité, si cette probabilité est supérieure à s sinon p est rejeté.

Dans la présente application, le seuil est de 0,75 et k est à 30.

Au niveau de l'application la seule date du mois de mai s'est avérée insuffisante pour la reconnaissance du riz. En effet nous sommes alors en saison des pluies, et autant les rizières que les marais, fleuves... sont dans des états quasi-identiques. Aussi a-t-il fallu recourir aux résultats de l'étude du calendrier agricole <5> <10> pour déterminer une fourchette de date pouvant idéalement compléter celle-ci. La classe 5, juillet, août et septembre, a été alors sélectionnée car déterminé en partie par le riz à l'état chlorophyllien. Ce qui assure une optimalité de sa discrimination d'avec les phénomènes qui lui ressemblaient au mois de mai, puisque ceux-ci étant tous théoriquement du domaine de l'eau, demeure globalement de l'eau. Rappelons que le modèle de classement ainsi posé, on peut tendre vers une précision accrue des résultats en introduisant d'autres variables, tel par exemple un modèle numérique de terrain, des paramètres de textures, de structures...



Le riz de seconde saison après classement par les images SPOT

VALIDATION

Sur le groupe de parcelles dévolu à la validation, après classement il est possible par confrontation à la vérité-terrain de déterminer si oui ou non un pixel est bien classé. En d'autres termes chaque fois qu'un pixel est considéré, il n'y a que l'alternative 0/1 sur son état. L'ensemble des pixels de validation définit donc une distribution de probabilité, tel que pour chacun d'entre eux pris au hasard, on ait une variable aléatoire de probabilité p d'être bien classé, et q d'être mal classé, tel que $p+q=1$. On se retrouve alors dans un cas d'application de la loi binomiale, avec p étant la proportion de bien classé. Comme de plus le volume de pixels utilisé en télédétection est grand, on peut approximer cette loi binomiale par une distribution de Laplace-Gauss. *Ce qui permet alors d'affecter un niveau de confiance à la démarche de classification et au résultat obtenu <5>*.

APPLICATION

L'échantillon de validation comporte 810 pixels, et 684 d'entre eux sont bien classés. La loi binomiale est $B(810,0.84)$.

Ses paramètres sont alors:

$$\begin{aligned} \text{moyenne} &= np = 680.4 \\ \text{écart-type} &= \text{sqr}(npq) = 10.43 \end{aligned}$$

Elle est approximée par la normale $N(680.4,10.43)$.

Comme dans une distribution normale 99.9% de la population se trouvent dans un intervalle de plus ou moins 3 écarts-types autour de la moyenne, on peut alors dire avec un niveau de confiance de 99.9%, qu'au minimum ($m-3s$) individus sont bien classés par la démarche classificatoire. En rapportant cette limite inférieure de bien classé à l'effectif total servant à la validation on obtient le pourcentage minimum de bon classement auquel on peut s'attendre avec un niveau de probabilité de 99.9% en utilisant cette démarche de classement. Suivant la répartition de la loi normale il est possible de faire varier ce niveau de probabilité. Ainsi :

- probabilité de 68% pour une limite inférieure de bon classement de ($m-1s$)
- probabilité de 95% pour une limite inférieure de bon classement de ($m-1.96s$)
- probabilité de 99% pour une limite inférieure de bon classement de ($m-2.58s$).

Dans l'actuelle application on arrive à au moins 649 individus bien classés, soit que la démarche classificatoire a une limite inférieure de bonne classification de 80%.

Il est possible de faire varier à son gré cette limite inférieure en agissant sur le seuil de probabilité de classement / rejet de la méthode discriminatoire des k -plus proches voisins. Ainsi dans ce cas on se satisfait d'une erreur maximale de classification de 20%. Il est possible de le diminuer en diminuant le seuil de rejet (une diminution excessive ou le retrait du seuil de rejet aboutit à une surestimation, auquel cas il est nécessaire de tester la limite supérieure de bon classement).

ESTIMATION A UN NIVEAU ADMINISTRATIF QUELCONQUE.

Ainsi que vu précédemment, du fait de la contrainte d'efficacité, seul un minimum d'images à haute résolution spatiale est utilisé pour la reconnaissance effective des thèmes visés. L'estimation proprement dite est assurée par la partie statistique du modèle.

Il a été vu lors de la présentation de l'estimation statistique aréolaire que l'extrapolation au niveau de la totalité des strates, est obtenue par simple expansion directe. A savoir qu'il suffit de multiplier l'estimation de l'extension moyenne du thème par segment, par le nombre total de segments de la strate pour avoir l'estimation sur la totalité de la strate. Il est donc nécessaire de disposer de l'extension des strates pour avoir leur nombre de segments. Ce dont on ne dispose précisément pas du fait de l'utilisation d'images minimales.

On dispose par contre de façon exhaustive, de l'extension des structures diachroniques. Pour disposer de l'extension des strates en quelque localisation que ce soit, il suffit d'estimer sur la base des surfaces communes aux structures diachroniques et aux strates, des relations fournissant les strates à partir des structures. Deux démarches sont possibles: estimation statistique, ou estimation sur la base de recalage des images.

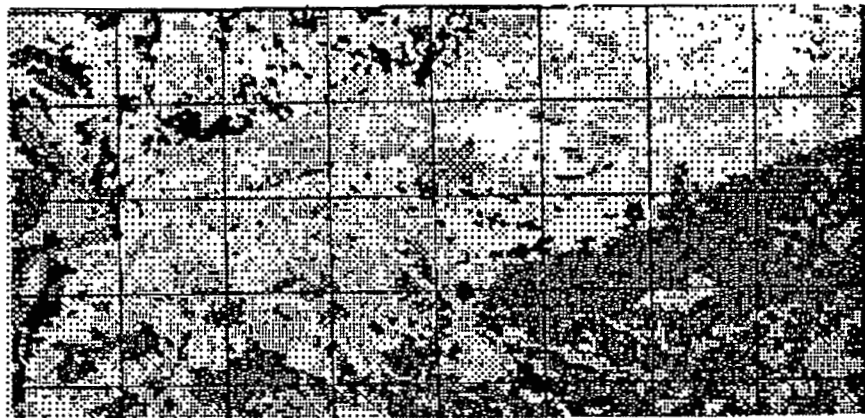
a). Estimation statistique: du fait que globalement chacune des deux représentations est influencée par les mêmes facteurs, les relations recherchées seraient du type linéaires. Ainsi pour une strate quelconque s , et pour k structures diachroniques d_j , on aurait:

$$s = \sum_{j=1}^k (a_j \cdot d_j) + e$$

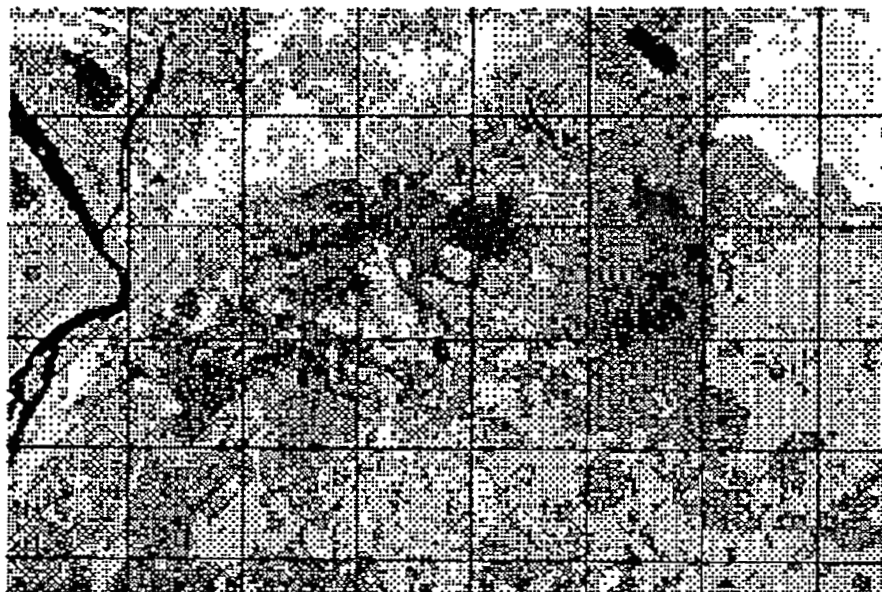
tel que : a_j facteur estimé, e constante, et d_j l'extension de la structure j

L'estimation de telles relations ne pouvant se faire que sur la base d'un ensemble d'individus, on définit une population de quadrats par décomposition de l'aire commune aux deux types d'information (structure et strate) <5>.

Ainsi dans l'actuelle application obtient-on les deux décompositions suivantes en quadrats:



Aire couverte par les images LANDSAT en quadrats



Aire couverte par les images SPOT en quadrats

b). Estimation par recalage des images: après recalage des images, il est possible d'obtenir simplement pour chaque type de structure la proportion de chaque strate. Comme la représentativité des structures a été assurée au niveau de l'échantillon minimal d'images à hautes résolutions, ces proportions sont directement utilisables.

Ainsi on aurait pour une strate et pour une structure:

$$E_{s,d} = P_{s,d} \cdot E_d$$

tel que:

- $E_{s,d}$ extension de la strate s dans l'aire administrative totale pour la structure d
- $P_{s,d}$ proportion de la strate s dans la structure d à partir de l'aire commune aux deux types d'image Landsat et Spot
- E_d extension de la structure d dans toute l'aire administrative.

Ayant la capacité d'obtenir l'extension des strates en quelque localisation que ce soit, il suffit de disposer de la proportion du thème visé par type de strate, pour pouvoir l'y estimer. Ainsi dans le cas présent, au niveau de la sous-préfecture telle que représentée ci-dessous, à partir de l'extension des structures diachroniques peut-on obtenir celle des strates.

Dans le cas où ni les relations linéaires ne sont significatives et ni le recalage possible, l'obtention de résultats demeure possible du fait de la base statistique même du modèle. En effet d'une part les images à hautes résolutions sont échantillonnées (en particulier détermination de leur nombre et localisation spatiale) de telles façons qu'elles soient représentatives des structures diachroniques et de leurs extensions. Et d'autre part à un second niveau, le prélèvement de vérités-terrain est réalisé suivant un processus probabiliste stratifié, assurant un optimum de représentativité à la fois en termes de localisation (dispersion) spatiale, et en termes d'état (différentiations phénologique, variétale, culturelles...).

Dans ce cas de figure, on ne passe plus par l'estimation aréolaire stratifié, mais on extrapole directement à l'aire entière. Pour cela:

.on calcule la proportion du thème sur le nombre total de pixels des images à hautes résolutions après classement

.on applique cette proportion au nombre total, calculé, de pixels hautes résolutions de la subdivision administrative entière.

C'est ce troisième cas de figure -le plus défavorable- qui est illustré dans cette présentation (rappelons que le premier cas de figure a été illustré dans la référence bibliographique <5>).



Sous - préfecture de Marovoay organisée en 12 structures par les données MSS

APPLICATION

L'aire totale d'entraînement (image à haute résolution) se compose de 2567205 pixels.

Le nombre de pixels classés en riz de seconde saison est de 50133.

L'extension de la sous-préfecture étant de 1755688 pixels de l'image à moyenne résolution (MSS), sachant qu'un pixel MSS fait environ 12 pixels SPOT, on calcule qu'elle fait 21068256 pixels SPOT.

L'expansion directe donne alors $411425 - ((50133 * 21068256) / 2567205)$ - pixels en riz. Soit comme le pixel Spot fait 0.04 hectares, 16457 hectares.

Ce résultat est tel qu'à chaque étape où un échantillonnage a été réalisée, un niveau de probabilité de 95% a été retenu de manière à ce qu'un intervalle de un autour des moyennes estimées contienne la moyenne de la population. De plus au niveau des classifications (reconnaissance du thème) on sait avec un niveau de confiance de 99.9% que la limite inférieure du pourcentage de bien classé est de 80%.

Partant de l'hypothèse que tout a été contrôlé par le modèle, y compris la représentativité statistique des vérités-terrain et donc la validité des classifications relativement aux réalités du terrain, ce résultat ne sera pas expressément comparé aux chiffres publiés par ailleurs. Ceci car ainsi que vu <1> les chiffres disponibles sont souvent contradictoires, sans niveau de précision, et élaborés de façon inconnue. Rappelons cependant que par exemple en 1981 un chiffre obtenu officiellement a été de 13600 hectares.

BIBLIOGRAPHIE

- <1> BONNEFOND P., Rapport sur les sources statistiques de 12 pays en voie de développement - C.E.E., 1988.
- <2> HOUSEMAN E., Area frame sampling in agriculture, Statistical Reporting Service U.S. Department of Agriculture- Washington D.C., 1979.
- <3> ANDRIANASOLO H., Première approche d'un modèle de décomposition de l'espace (pour une approche intégrée). Journées Télédétection de l'ORSTOM, 1988.
- <4> ANDRIANASOLO H., Modélisation de la stratification d'un espace régional par la télédétection, pour l'estimation de surface. Séminaire Informatique de l'ORSTOM, 1988.
- <5> ANDRIANASOLO H., Analyse statistique des données de télédétection, statistiques agricoles, application sur Madagascar. Doctorat 1987.
- <6> DEROO-DUSSAIX, Pratique et analyse des enquêtes par sondage. PUF, 1980.
- <7> ANDERBERG, Cluster analysis for application. Academic Press, 1973
- <8> ANDRIANASOLO H.-MALEK C., Définition et détermination des unités dynamiques d'un paysage sahélien (OURS) pour un diagnostic des facteurs de dégradation et de régénération des ressources naturelles et renouvelables. Congress of ISPRS, KYOTO, 1988.
- <9> ANDRIANASOLO H.-MALEK C., Proposition d'une méthode de structuration du paysage à partir d'images Spot, application en zone sahélienne. Congrès des Sociétés Savantes, Strasbourg 1988.
- <10> ANDRIANASOLO H., Calendrier agricole et acquisition des images satellitaires; EHESS, 1987.
- <11> ANDRIANASOLO H., Agricultural statistics and space images in Madagascar - Estimation of second rice season acreage in a sub-préfecture. Congrès of ERIM, Nairobi 1987.
- <12> DIDAY, LEMAIRE, POUGET, TESTU, Eléments d'analyse de données. DUNOD, 1982
- <13> AMERICAN SOCIETY OF PHOTOGRAMMETRY, Manual of remote sensing. Second Edition, two volumes, 1983.
- <14> DUDA, HART, Pattern recognition and scene analysis. Wiley New-York, 1973.
- <15> FU K. S., Syntactic methods in pattern recognition. Academic Press, New-York, 1974.
- <16> MANDELBROT B., Les objets fractals. Flammarion, 1984.