

TELEDETECTION ET "ENQUETE TERRAIN" DANS LE SUD-OUEST MAURITANIEN

Optimisation des enquêtes de terrain pour la cartographie de l'occupation du sol et le diagnostic de l'état de l'environnement.

Vincent GODARD IEMVT/CIRAD

Résumé :

L'utilisation de la télédétection dans la cartographie de l'occupation du sol pour le suivi de l'environnement de deux zones tests du Sahel mauritanien a permis de fiabiliser la phase de préparation du plan d'enquête, d'améliorer les estimations de superficies et d'alléger le plan de sondage. La moyenne des coefficients de variation associés aux estimations de superficies passe de 61,5 p.100 (sans la télédétection) à 14,5 p.100 (avec la télédétection).

Abstract :

The use of remote sensing in a land use mapping programme and environment assessment in two test areas of Mauritanian Sahel allowed a more reliable preparation of survey programme, improved surface estimates and reduced of sampling frequency. The mean coefficient of variation associated to the surface estimates decreased from 61.5 % (without remote sensing) to 14.5 % (with remote sensing).

Mots clefs :

Cartographie d'occupation du sol ; Enquête de terrain ; Environnement ; Mauritanie ; Sahel ; Sondage aréolaire ; Télédétection ;

Pour réaliser des cartes d'occupation du sol, une reconnaissance de terrain est indispensable, malgré l'utilisation des satellites à haute résolution du type de SPOT1 ou Landsat TM. Cette enquête permet d'établir des relations entre les thèmes étudiés sur le terrain et les images correspondantes. Dans une seconde étape, ces relations sont généralisées à un espace plus vaste.

Le but de cette recherche est d'optimiser l'utilisation de l'imagerie satellitaire lors des phases suivantes :

- préparation de l'enquête de terrain ;
- collecte des données ;
- traitements des résultats et généralisation.

Le satellite fournit une vision instantanée, globale et "objective" du terrain, utilisable comme base de sondage. Un sondage de type aléatoire lui est appliquée. Il permet d'estimer les superficies des différents postes d'occupation du sol, la précision qui leur est associée, ainsi qu'une cartographie des thèmes reconnus. Pour aboutir à ces résultats, plusieurs étapes se sont succédé dans le cadre d'une étude d'occupation du sol en zone sahélienne du Sud de la Mauritanie (programme financé par le Ministère de la coopération).

Nous aborderons successivement :

- les problèmes de la détermination du taux de sondage et de la taille des segments d'enquête ;

- la précision de la localisation du relevé de terrain ;
 - les procédures statistiques et cartographiques utilisées ;
- enfin,
- les résultats obtenus seront analysés.

1. La phase de terrain et sa préparation

Quel chercheur, au retour d'une campagne terrain, ne s'est pas interrogé sur la pertinence et la validité de sa collecte de données. Pour s'assurer d'une représentativité la plus complète possible de sa zone d'étude, il est préférable de ne pas se laisser guider par sa seule expérience mais plutôt par des procédures de sondage d'où la subjectivité est absente. Le choix d'un échantillonnage sans biais s'est donc imposé.

Nous nous plaçons ici dans le cadre des sondages aréolaires, ou spatiaux. Selon ce principe la zone d'enquête est subdivisée en segments : "On entend par segment, non une portion de droite, mais une portion de territoire formée de parcelles contiguës (on a procédé à une segmentation du territoire d'où le terme)"¹. Enfin, le segment de par sa forme est dit carré et comme l'enquête ne porte que sur les parcelles incluses dans les limites des segments, il est considéré comme fermé.

Il était possible de procéder à un tirage systématique non aligné comme celui réalisé dans le cadre du diagnostic rapide des palmeraies². Bernard Husch³ fait très justement remarquer que ce tirage ne permet d'évaluer qu'approximativement l'erreur maximale. Le sondage aléatoire fournit, lui, l'erreur moyenne qui autorise le calcul de la précision des estimations de superficie. C'est, de ce fait, une méthode proche des sondages aléatoires qui a été retenue (Cf. développement en 2^e partie).

Il s'agit en fait d'un sondage que Philippe Fournier⁴ appelle "Sytématique Aléatoire Stratifié". Il est systématique, pour assurer une meilleure répartition des segments d'enquête à l'intérieur de la zone d'étude (ou zone pilote, Cf. figure 1), et aléatoire pour apprécier la précision des estimations des superficies. La stratification est implicite, c'est la zone d'enquête.

Cependant, la répartition même systématique-aléatoire des segments d'enquête dans la zone d'étude ne garantit pas l'obtention d'une vision exhaustive du terrain d'enquête. Le nombre et la taille des segments d'enquête ont une très grande importance sur la qualité des résultats, tant statistiques que cartographiques. Il en va de même pour la précision de la localisation des segments, puis de leur report sur le document numérique.

Il convient, avant tout, de rappeler le type de données collectées sur le terrain pour réaliser cette cartographie d'occupation du sol, cartographie pour laquelle l'activité agropastorale est notre principal centre d'intérêt. Ces données se répartissent en deux catégories.

- La première concerne directement l'enquête par segment. Elle fournit les renseignements physiques comme le cortège floristique du segment, les espèces ligneuses ainsi que la proportion de sols nus, leur nature, la quantité de phytomasse herbacée, etc. D'un point de vue sociologique, elle renseigne sur l'utilisation du sol, la présence de troupeaux, le piétinement, les cultures, donc pour partie sur la nature et l'origine des dégradations subies par le milieu.

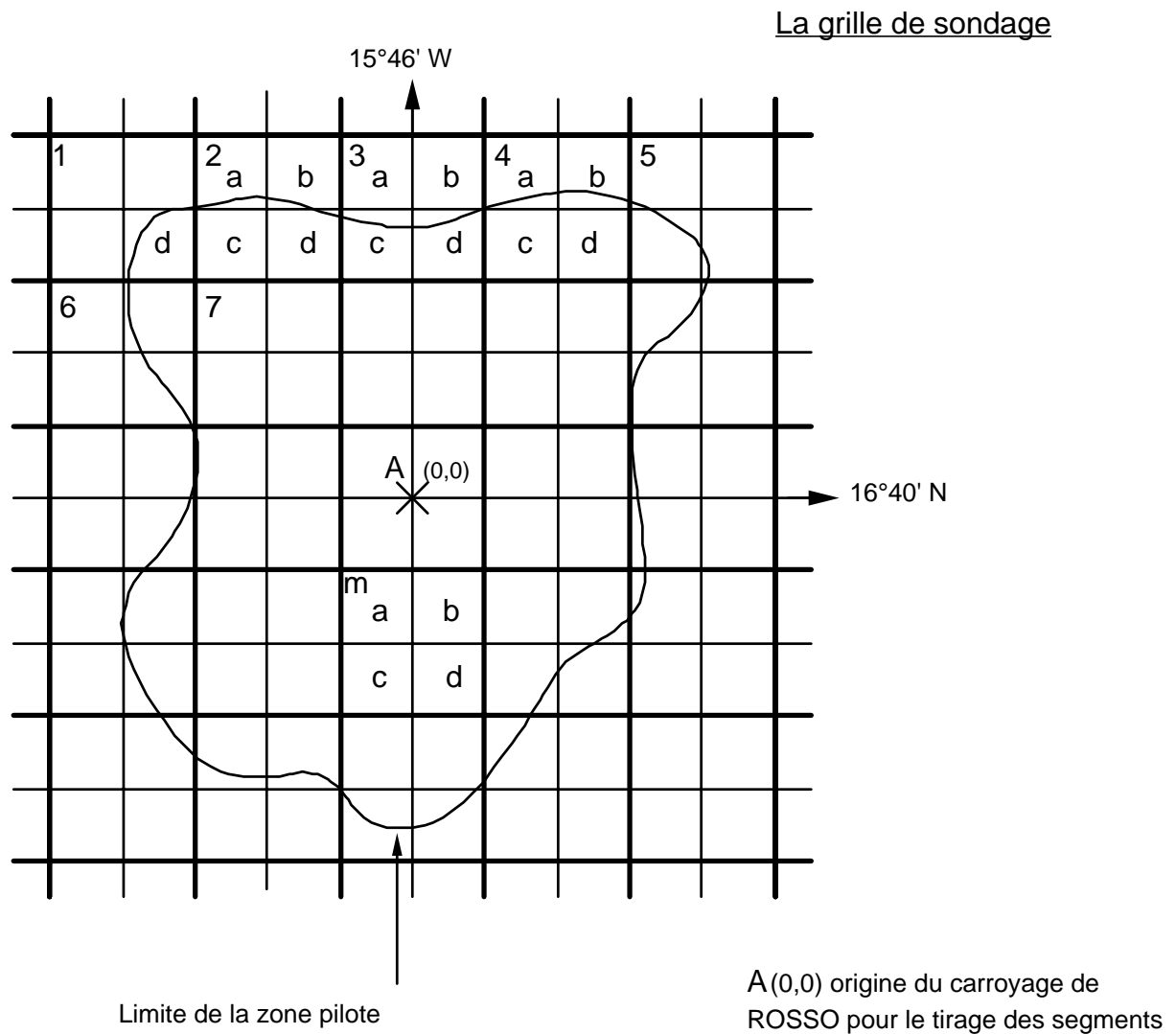
¹FOURNIER ; GILG 1985 p.8

²GODARD ; DOLLE ; VAYSSIERE 1988

³HUSCH 1971 p.47

⁴FOURNIER 1986 p.16

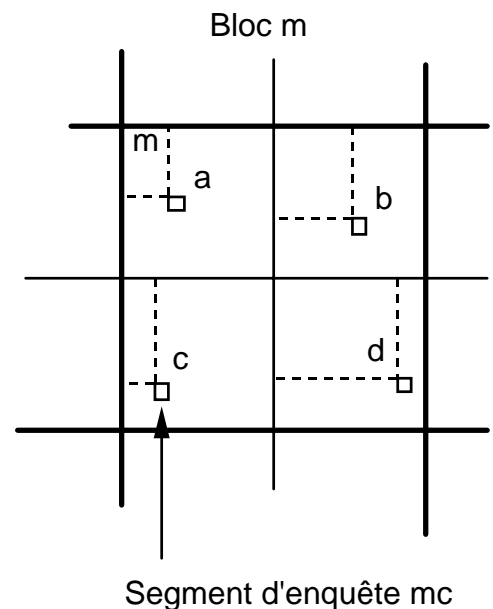
fig. 1 **Dispersion des segments d'enquête**



Le sondage "Systématique Aléatoire Stratifié"

Tirage à l'intérieur du bloc m des 4 segments d'enquête de 36 ha. Puis, répétition à l'ensemble des blocs.

Un bloc est constitué de 4 carreaux de 6 km de côté.



GODARD (V.) - 1990 - Optimisation des enquêtes de terrain pour la cartographie de l'occupation du sol et le diagnostic de l'état de l'environnement. in : *Téledétection et Tiers Monde, Actes de la table ronde internationale CEGET-CNRS : Apports de la téledétection spatiale à l'étude des paysages et des systèmes agraires tropicaux*. Talence 25-27 octobre 1989 : 87-105.

- La deuxième est d'ordre purement topographique. Elle vise à mettre à jour les documents, bien souvent anciens, qui servent d'appui à la cartographie thématique. Dans le Sahel de l'Afrique de l'Ouest, ces documents sont en général issus de l'IGN (cartes au 1/200 000) et datent, fréquemment, de la fin des années cinquante. Il faut donc réactualiser les pistes, villages, points d'eau, etc.

Pour obtenir tous ces renseignements, la préparation de la mission revêt une importance primordiale. Or, bien souvent, le temps manque pour élaborer le plan d'enquête. Il faut donc déterminer une méthode simple et rapide de préparation.

La méthode de sondage développée ici reprend celle que le Service Central des Enquêtes et Etudes Statistiques (SCEES) a utilisée en Ile-de-France¹. En fonction d'un taux de sondage prédéfini, un certain nombre de segments d'enquête sont tirés au hasard.

Il n'existait pas, à notre connaissance, d'enquête analogue en milieu naturel, nous avons donc retenu un taux de sondage s'inscrivant dans une fourchette comprise entre 1 et 2 p.100, c'est-à-dire proche des taux utilisés pour le suivi agricole. Le BDPA² relate les taux suivants : 1 p.100 pour l'enquête TERRUTI France, 1,5 p.100 pour les inventaires régionaux de la CEE et 1,7 p.100 pour "l'enquête maïs" au Kenya.

Le taux de sondage est déterminé par la taille des segments mais aussi par leur nombre. Ce type de sondages aréolaires prend en compte des surfaces. Avant de déterminer la taille et le nombre des segments, il faut caractériser la précision de la localisation et celle du report de l'information terrain vers les fichiers numériques. La précision du résultat final en dépend.

Jusqu'à présent, nous ne disposions bien souvent que du compteur kilométrique du véhicule, d'une boussole, voire d'un Topofil pour nous localiser. Cela a été notre cas pour les premières campagnes de terrain. Il est très difficile, dans ces conditions, d'apprécier l'exactitude de sa position, sauf pour du parcellaire agricole de type "culture irriguée". En forêt ou dans un champ de dunes, on est à plusieurs dizaines, voire centaines de mètres de la position voulue.

Actuellement, pour un coût très faible (moins de 30 000 FF TTC), des localisateurs par satellites positionnent à 30 mètres près. On constate, toutefois, qu'une dégradation de la précision accompagne les différentes étapes de la procédure d'enquête. Il est possible de connaître cette perte de précision, dans la localisation des résultats, en cumulant les dégradations mesurables.

Les données satellitaires sont relativement coûteuses. La plupart des utilisateurs les achètent avec un niveau de traitement minimal, de type 1B. Celui-ci permet de s'orienter sur le terrain, surtout avec des agrandissements photographiques au 1/50 000. En revanche, il ne comporte pas de mise en géométrie et la précision moyenne des amorces du carroyage (pour les films et tirages) est de 720 mètres (+/- 860 m)³. Il n'est pas question de localiser les relevés de terrain avec ce type de documents. Une mise en géométrie du fichier numérique s'impose : pour utiliser le localisateur par satellite et positionner les données collectées.

Les corrections géométriques entraînent, lors de la prise d'amers et de la rectification de la matrice, une approximation de l'ordre du pixel (la position des segments sera donc calculée à +/- 20 m). Le localisateur est précis à +/- 30 m, soit 1,5 pixel. Le report sur le document de terrain n'est pas censé induire d'erreur car les coordonnées de l'observation sont conservées et servent au repositionnement sur le document numérique, en laboratoire. La somme des imprécisions mesurables est au minimum de 50 m. En retenant 60 m, nous conservons une marge et rendons cette distance compatible avec les pixels SPOT, l'imprécision est donc de +/- 3 pixels.

¹FOURNIER 1986

²BDPA 1988 p.17

³SPOT 1986 p.5-40

Cette mesure de l'imprécision nous permet de calculer l'aire minimale d'échantillonnage. Townshend et Justice¹ déterminent cette aire (A) par la formule suivante :

$$A = L (1 + 2N)$$

où L est la taille du pixel (20 m pour SPOT XS), N le nombre de pixels d'erreur (ici 3). Ici, cette aire minimale (A) est de 140 m de côté, soit 7 pixels. On prendra donc soin sur le terrain d'effectuer les prélèvements sur des placeaux dont la taille s'inscrit à l'intérieur d'aires homogènes d'environ 2 ha (7*7 pixels), ceci pour réduire les erreurs d'affectation. Pour ce travail, les prélèvements ont été effectués sur des placeaux de 1 ha (5*5 pixels).

On peut aussi parer à une dégradation du positionnement par satellite (mauvaise configuration spatiale lors de l'enquête) en calculant une aire minimale de sécurité.

On aura compris que **la taille des segments d'enquête ne doit pas être inférieure à 2 ha en raison du niveau de précision de la localisation** définie précédemment. Cette condition préalable posée, il reste à établir la taille optimale de ces segments ; celle-ci est fonction du taux de sondage, donc de leur taille et de leur nombre.

Pour cela aucune connaissance préalable du terrain n'est nécessaire. En effet, la méthode développée ici utilise la téledétection et l'exhaustivité de sa vision du terrain pour déterminer le nombre et la taille des segments. Les radiométries que le satellite enregistre sur le domaine sahélien sont autant de variables qui nous permettent d'apprécier l'hétérogénéité du milieu et donc l'intensité du sondage à lui appliquer. De la même manière, il est possible de stratifier le paysage. Cette approche n'a pas été retenue lors des premiers essais.

Les seuls éléments nécessaires pour réaliser cette préparation du sondage sont les données numériques enregistrées une ou plusieurs années auparavant, mais à une saison comparable, pour retrouver des paysages analogues lors de la mission de terrain. De même nous avons apprécié de disposer d'agrandissements photographiques SPOT au 1/50 000 dont la finesse est d'une aide précieuse sur le terrain.

Le taux de sondage est fonction de deux paramètres : la taille des segments et leur nombre. Les techniques issues des méthodes géostatistiques sont une approche intéressante pour déterminer la taille optimale du segment d'enquête. Le fait de tirer des aires, et non des points, entraîne une certaine homogénéité à l'intérieur du segment, mais cette homogénéité n'est pas complète. Plus on s'éloigne d'un pixel de référence, plus l'hétérogénéité croît (le risque de passer d'un thème à un autre augmente avec la distance). Cet accroissement de l'hétérogénéité est mesuré par des calculs de variance qui permettent d'estimer à partir de quelle distance l'accroissement se stabilise et donc le seuil pour lequel la représentation du paysage est la plus diversifiée pour le moins de pixels à enquêter. Cette technique très intéressante a été jugée un peu lourde en première approche, tout comme celle des écarts types locaux. Nous avons donc préféré retenir une troisième solution qui nous est apparue plus légère en regard des moyens dont nous disposions.

Comme il apparaît que la dimension des segments en milieu agricole européen doit se moduler sur la taille moyenne du parcellaire et sur celle des exploitations agricoles², c'est la taille moyenne des entités thématiques qui va jouer ce rôle en milieu naturel. Nous allons calculer cette superficie moyenne en utilisant les données numériques d'un enregistrement datant d'une ou plusieurs saisons antérieures. Pour cela une classification non supervisée du type "nuées dynamiques" est réalisée. Le nombre maximal de classes est calculé par la formule de Huntberger³ où :

¹in de KEERSMAECKER 1987 p.202

²FOURNIER 1986 p.14

³CHADULE 1974 p.32

$$Nc = 5 \text{ Log } n$$

avec n le nombre de points traités. Pour une image 1024*1024, Nc est de 30, donc supérieur à celui qui sera retenu lors de la cartographie finale. Puis l'image issue de cette classification est filtrée, binarisée, ébarbulée, etc. On ne conserve plus que les taches dont la superficie est suffisante pour être enquêtée. Enfin, un dénombrement et une paramétrisation sur les surfaces sont effectués. Ils nous donnent le nombre et la surface moyenne des parcelles.

Le tableau 1 résume la paramétrisation obtenue. A titre d'exemple, pour un segment de 30 ha, la taille moyenne des parcelles qu'il contient est de 7,7 ha et leur nombre de quatre environ.

Nous n'avons pas cherché à évaluer de résultats pour des segments supérieurs à 100 ha, car il ne paraissait pas raisonnable de parcourir des zones d'enquêtes dont les côtés dépasseraient 1 000 mètres. De plus, nous pensions pouvoir réaliser 4 à 5 observations par segment sans faire d'oublis ou de redondances en raison de nos expériences antérieures. Nous avons retenu pour cette raison un segment carré de 36 ha, donc une forme géométrique simple dont chaque côté de 600 m est divisible par 20 ou 30, c'est-à-dire un multiple des pixels SPOT ou Landsat TM. Le dépouillement des données de terrain a confirmé les résultats du dénombrement effectué sur la classification automatique de préparation. Nous avons enquêté en moyenne 4,4 parcelles par segment, l'écart type associé étant de 1,3.

Tableau 1 : Nombre et surface des parcelles par segment

Taille du segment	Surface moyenne des parcelles	Nb. moyen de parcelles par segments
20	6,5	3,1
30	7,7	3,9
40	8,6	4,7
50	9,5	5,3
60	10,6	5,7
80	11,7	6,8
100	12,5	8,0

Les superficies sont exprimées en ha.

Le nombre des segments, deuxième paramètre du taux de sondage, est apprécié de manière plus pragmatique. La moyenne des temps d'enquête était en Ile-de-France d'une heure trente¹. Pour un paysage sahélien analysé comme moins complexe, le temps d'enquête a été estimé à une heure, le temps d'approche à une heure également. A raison de 7 jours de travail par zone d'étude, et 3 sites par jour (6 heures de travail plus une marge d'un tiers), il devait être possible d'étudier 21 segments. Soit un taux de sondage de 1,8 p.100 (30*30 pixels * 21 segments / 1024*1024). En réalité, seuls 20 segments ont été tirés portant le taux de sondage à 1,7 p.100. Celui ci est donc bien conforme à la fourchette retenue initialement, pour mémoire entre 1 et 2 p.100.

Après analyse, ces 20 segments se sont révélés être légèrement insuffisants (leur nombre a été porté à 30 pour la campagne de terrain 1989). Il est probable qu'un tiers de segments en plus aurait permis d'améliorer la précision des résultats en l'absence de données satellitaires contemporaines de l'enquête de terrain (Cf. 3^e partie), mais aussi d'appréhender des postes d'occupation du sol de faible superficie que la répartition des points de sondage n'a pu saisir. La pratique a d'ailleurs démontré que l'on pouvait, sans trop de risques, contrôler 4 à 5 segments par jour.

¹FOURNIER 1986 p.24

2. Aspects statistiques et cartographiques

La méthode d'enquête terrain précédemment décrite tente d'affranchir le chercheur d'une certaine subjectivité vis-à-vis des prélèvements. Elle est relayée en laboratoire par des traitements statistiques et cartographiques caractérisés par des estimateurs sans biais et indépendants. Ils fournissent les estimations des superficies des différents postes d'occupation du sol.

La procédure de sondage est issue des méthodes développées par Ph. Fournier¹. Elle a été mise au point dans une optique "statistiques agricoles" pour des paysages comme ceux d'Ile-de-France. Nous l'avons adapté à un milieu naturel sahélien peu anthropisé, aussi bien pour l'aspect terrain que dans la phase de dépouillement des données.

En laboratoire, nous avons voulu optimiser notre utilisation des données satellitaires :

- en réalisant une cartographie d'occupation du sol ;
- en affinant la précision de nos statistiques par le biais d'un estimateur par la régression.

Ces différentes étapes méthodologiques méritent que l'on s'y attarde. Nous allons en détailler les aspects les plus importants pour la cartographie et les statistiques d'occupation du sol.

Pour obtenir la superficie, segment par segment, de chacun des thèmes enquêtés, il n'est pas possible de procéder comme dans un milieu où le parcellaire est contigu. Dans le Sahel mauritanien, il est rare de rencontrer des limites tranchées, comme elles peuvent exister en Europe entre deux champs différents. Il n'y a fréquemment qu'une gradation ténue entre deux formations végétales (proche du phénomène de lisière décrit par les forestiers). C'est le chercheur qui va attribuer tel groupe de pixels à telle formation plutôt qu'à telle autre. La décision est facilitée par l'analyse des radiométries en complément du découpage analogique (limite visuelle). En effet, après positionnement du thème enquêté sur le support numérique, ses valeurs radiométriques en plus de sa forme et de sa couleur, vont le caractériser. Nous disposons dans cette étude, des trois canaux bruts plus deux néo-canaux, les indices de végétation (IV) et de brillance (IB) :

$$IV = \frac{XS3 - XS2}{XS3 + XS2} \quad IB = \sqrt{XS3^2 + XS2^2}$$

Ces informations supplémentaires vont nous permettre de discriminer plus facilement les points ambigus.

La procédure est la suivante. Un sondage d'une cinquantaine de points est réalisé sur tous les segments. Il nous fournit la proportion, par segment, de chacun des thèmes. Il est possible de déterminer quels sont les thèmes dont la fiabilité est satisfaisante et ceux pour lesquels le sondage n'a pas permis d'asseoir la représentativité, par la variance d'une proportion :

$$P \cdot Q / N$$

où P est la proportion, Q son complémentaire et N le nombre de points sondés. On effectue ici un premier tri qui consiste en une "mise à l'écart" des thèmes dont la superficie est inférieure à 7 p.cent par segment, c'est-à-dire dont le coefficient de variation est supérieur à 20 p.cent. Nous verrons plus loin que cette "mise à l'écart" consiste, entre autres, à ne pas utiliser ces groupes de pixels pour l'initialisation des traitements cartographiques.

A ce stade de la procédure, nous connaissons la proportion de chacun des thèmes par segment, deux possibilités s'offrent à nous :

¹FOURNIER 1986

- il n'y a pas eu de prise de vue effectuée au cours de notre séjour sur le terrain ;

- une prise de vue correspondant à notre enquête terrain a été enregistrée.

Dans le premier cas, nous nous contentons de l'enregistrement qui nous a servi pour aller sur le terrain. De quelques années antérieures, il doit relater des stades biologiques analogues à ceux que l'on a enquêtés. La méthode d'estimation, issue des travaux de Ph. Fournier (op. cité) est adaptée à nos besoins. Elle nous permet, à partir de nos fiches d'enquête, d'effectuer des estimations de superficie. Des erreurs, dans le calcul des proportions de chaque thème à l'intérieur des segments, vont se produire mais elles seront d'autant plus faibles que les formations végétales ont peu évolué entre les deux dates.

Le second cas, évidemment plus favorable, sera détaillé ultérieurement.

Tout le travail de terrain n'est donc pas compromis par une absence de données satellitaires contemporaines à l'enquête. Voici, après adaptation à nos préoccupations, comment calculer les estimations des différentes catégories d'occupation du sol.

- Estimation d'une superficie :

Soit Y_k la superficie du secteur d'étude dans la catégorie d'occupation du sol k.

$$Y_k = Y * \frac{1}{m} * \sum_{i=1}^m \frac{Y_{ki}}{Y_i}$$

Avec :

Y la superficie de la zone d'étude ;

m le nombre de segments tiré au hasard ;

Y_{ki} le nombre de pixels en catégorie de territoire k à l'intérieur du segment i ;

Y_i le nombre de points à l'intérieur du segment i.

- Variance de cette estimation :

Soit $V(Y_k)$ la variance d'une somme.

$$V(Y_k) = Y^2 * \frac{1}{m} * \frac{1}{m-1} * \frac{M-m}{M} * \sum_{i=1}^m (s_{ki} - s_k)^2$$

Avec :

$s_{ki} = \frac{Y_{ki}}{Y_i}$ estimation de la proportion de catégorie de territoire k dans le segment i ;

s_k estimation de la proportion moyenne en catégorie de territoire k ;

m nombre de segments échantillons tiré ;

M nombre total de segments couvrant la zone d'étude ;

$\frac{M-m}{M}$ très proche de 0,99 ou 0,98 dans la mesure où l'on a retenu un taux de sondage de l'ordre de 1 à 2 p.100.

Comme l'indique Ph. Fournier, un des initiateurs de cette méthode¹ : "Dans le cas d'un sondage aléatoire à un degré, avec des probabilités égales et sans remise, on utilise la formule de la variance d'une somme. Il n'est pas tenu compte du fait que le tirage des segments est plus ou moins stratifié, cet estimateur de la variance d'une somme est, en conséquence, calculé avec une surestimation certaine."

Grâce à cette procédure, et malgré l'absence supposée de données satellitaires récentes, il est possible de fournir, par poste d'occupation du sol, des estimations de superficie. Celles-ci seront accompagnées d'un indicateur de la précision des résultats. La variance, mesure de l'estimation de l'erreur aléatoire², permet de déterminer un coefficient de variation. Il nous indique la fourchette de précision (relative) des estimations (Cf. 3^e partie).

Nous avons vu que ces calculs ne requièrent pas directement l'utilisation de données numériques. Elles ne sont nécessaires qu'au moment de l'affectation des pixels aux différentes entités thématiques en vue de calculer la proportion de chacun des thèmes par segment. En revanche, contrairement aux statistiques d'occupation du sol qui se font sur des relevés de terrain récents, la cartographie ne peut se faire que sur des données numériques antérieures à la prospection, en l'absence d'une prise de vue qui lui soit contemporaine. La publication d'une cartographie, si elle s'avère indispensable, reprendra celle établie préalablement à la mission de terrain. La reprise de cette première carte à des fins de publication se justifie probablement plus que de réaliser une nouvelle classification supervisée à partir d'une imagerie déjà ancienne.

En revanche, dans l'hypothèse où l'on a obtenu la prise de vue contemporaine à l'enquête de terrain, la méthode de dépouillement va être différente, seul le calcul des proportions n'étant pas modifié. La classification supervisée par les données de terrain peut être effectuée. La cartographie, ici, précède obligatoirement l'estimation des superficies. En effet, l'estimation par régression consiste à établir une relation bi-univoque entre la "classification par télétection" et la "réalité de terrain" de la forme :

$$y = ax + b + e$$

où e est une variable aléatoire d'espérance nulle³.

Cette équation linéaire ne peut être déterminée directement, faute d'une connaissance exhaustive du terrain. En revanche, nos segments d'enquêtes, tirés aléatoirement, sont représentatifs de l'ensemble de notre zone d'étude. Nous allons donc comparer, thème par thème et segment par segment, les postes d'occupation du sol au moyen d'une régression linéaire. Celle-ci compare les scores obtenus sur le terrain et ceux obtenus par la cartographie. A ce stade, la cartographie n'est réalisée que sur les segments enquêtés. Les paramètres de classification sont d'abord étalonnés au niveau des segments, seules entités correctement analysées sur le terrain. Un peu moins de 10 p.100 des pixels "renseignés" sont prélevés pour initialiser les traitements par "maximum de vraisemblance". Les 90 p.100 restants sont utilisés comme population de contrôle. Cet étalonnage effectué, ces mêmes paramètres sont utilisés pour étendre la cartographie à l'ensemble de la zone pilote (Cf. carte hors texte).

Il est admis à partir de ces lignes que la cartographie est réalisée ainsi que les régressions thème par thème au niveau des segments. Nous allons détailler brièvement les calculs qui ont permis d'obtenir les estimations de superficies au niveau des zones d'études par un estimateur par la régression.

Cet estimateur est formulé de la façon suivante.

- au niveau du segment.

¹FOURNIER 1986 p.28

²Idem p.13

³MEYER-ROUX 81 p.12

$$Y_{reg\ k} = y_k + a \frac{(X_k - x_k)}{M}$$

Avec :

- y_k moyenne de la catégorie k , par segment, pour le terrain ;
- a pente de la droite issue de la régression entre les valeurs du terrain et celles de la classification pour le thème k ;
- X_k valeur du thème k dans la classification pour l'ensemble de la zone d'étude ;
- M nombre total de segments compris dans la zone d'étude, ici $1024*1024$ (taille du secteur d'enquête en pixels) divisé par $30*30$ (taille d'un segment en pixels) soit environ 1 165 segments ;
- x_k moyenne de la catégorie k , par segment, pour la classification.

- au niveau du secteur d'étude.

$$Y_{reg\ k} = M * y_{reg\ k}$$

De même, il est possible de calculer la variance associée à cette estimation.

- au niveau du segment.

$$\text{Var}(y_{reg\ k}) = \frac{(M-m)}{M} * \frac{1}{m} * \frac{1}{m-2} * \sum_{i=1}^{sm} e_{ki}^2$$

où $e_{ki} = y_{ki} - y_k - a (x_{ki} - x_k)$

Avec comme nouveaux éléments dans les formules :

- y_{ki} nombre de pixels du thème k reconnu sur le terrain dans le segment i ;
- x_{ki} nombre de pixels du thème k reconnu sur la classification dans le segment i .

Notons que "Pour calculer cette variance il faut disposer d'au moins trois échantillons"¹.

- au niveau de la zone d'enquête la variance devient :

$$\text{Var}(Y_{reg\ k}) = M^2 * \text{Var}(y_{reg\ k})$$

Tous les éléments pour effectuer le calcul des estimations de superficie ainsi que celui de leur précision sont réunis. Les bases statistiques et fondements théoriques sont issus de travaux récents dont on trouvera les références complètes en bibliographie².

L'estimateur, dit "estimateur par la régression", que l'on vient de détailler améliore la représentativité des données collectées sur le terrain, tout en permettant d'en collecter moins, pour une précision égale. Cela n'était pas possible sans la télédétection. Nous allons voir dans la troisième partie l'effet correctif que l'on a obtenu par l'apport des données satellitaires.

¹DESABIE 1966 p.218

²MEYER-ROUX 1981 ; DESABIE 1966 pp 215-219 ; GROSBAS 1987 pp 138-155 (estimation par le quotient et par la régression) ; BRION 1982 p. 29 (estimation par le quotient)

3. Quelques résultats

Bien évidemment, les deux approches n'offrent pas la même qualité de résultats. La comparaison des estimations, obtenues sur la zone pilote de Rosso (Cf. carte hors texte), permet d'apprécier les différences de précision qui caractérisent les deux procédures.

Tableau 2 : Les superficies et coefficients de variation des deux procédures de traitements

Thèmes	Sigles	Sans téledélect.			Avec téledélect.		
		Surfaces		C.V.	Surfaces		C.V.
		ha	%	%	ha	%	%
Surface en eau	Se	777	1,9	68,3	677	1,6	8,8
Culture irriguée	Ci	1143	2,7	90,1	307	0,7	16,4
Steppe arbustive hydromorphe	Sah	415	0,9	130,4	755	1,8	19,9
St. herbeuse. très claire des dépressions hydromorphes	Shy	2827	6,8	51,1	5314	12,7	17,4
Sol nu sablo-limoneux	Sns	12425	29,7	22,7	10363	24,8	4,9
Sol nu argileux	Sna	3040	7,3	49,5	1703	4,1	16,7
Sable vif	Sv	484	1,2	69,3	622	1,5	8,3
Steppe herbeuse très claire des dunes	Shd	11816	28,3	27,1	10996	26,3	4,4
St. herb. des ensablements et interdunes	Shi	7613	18,2	29,7	7952	19,1	15,6
St. arbustive des ensablements	Sae	951	2,3	60,9	1141	2,7	31,6
Steppe "boisée" gommeriaie	Sb	288	0,7	79,2	239	0,6	15,2
Total partiel					40069	95,9	
Non classés					1710	4,1	
Total		41779	100		41779	100	

Le total partiel et les superficies non classées sont dus aux arrondis dans les calculs.

Sans l'apport de la téledétection et avec ce taux de sondage, on retiendra que les résultats de cette approche ne sont pas très précis, voire mauvais, pour certains postes d'occupation du sol. Il faut que les thèmes atteignent des surfaces comprises entre 20 et 30 p.100 pour que le coefficient de variation associé à cette estimation se situe en dessous de 30 p.100 (cf. tableau 2). Pour les thèmes dont la superficie est inférieure à 10 p.100, le coefficient de variation est tellement élevé (Cf. la Steppe arbustive hydromorphe : 130,4 p.100) que seul l'ordre de grandeur garde une signification (la superficie est inférieure à 3 p.100).

L'absence de données satellitaires complémentaires ne permet pas une étude très fine du paysage. Seuls les thèmes de grandes extensions ont des résultats sur lesquels on peut se baser pour des opérations de diagnostic et d'aménagements ; il se trouve que ce sont aussi les postes d'occupation du sol les plus importants pour l'élevage. Les autres thèmes seront utilisés avec plus de réserve. Notons toutefois que le plan de sondage répondait à certaines contraintes dont la description des grandes unités de paysage à dominante pastorale et un temps d'enquête réduit. En cela le plan de sondage a donné satisfaction. **Cette démonstration doit aussi nous amener à nous interroger sur l'utilisation de données issues d'études où, ni la méthode d'obtention, ni la précision des résultats de superficies ne sont fournies.**

Comme on pouvait s'y attendre, l'apport de la télé-détection ne remet pas en cause les résultats précédemment obtenus. L'ordre de grandeur des estimations est comparable d'une méthode à l'autre. Les variations existantes affectent essentiellement des thèmes de superficies restreintes, comme les cultures irriguées qui régressent de 2,7 à 0,7 p.100 d'occupation du sol avec l'apport de la télé-détection, ou encore les steppes herbeuses très claires des dépressions hydromorphes qui passent de 6,8 à 12,7 p.100.

En revanche, **la caractéristique essentielle de l'apport de la télé-détection est l'amélioration de la précision des estimations.** La moyenne des coefficients de variation sans l'apport de la télé-détection est de 61,5 p.100. Avec adjonction des données satellitaires, celle-ci passe à 14,5 p.100. Il faut toutefois noter que l'amélioration n'est pas uniforme, elle affecte les thèmes les plus caractéristiques radiométriquement (en dehors de toute représentativité numérique dans le plan de sondage). Cela permet donc d'améliorer les thèmes "mal enquêtés" mais très spécifiques du point de vue de la radiométrie. Les surfaces en eau, cultures irriguées et steppes arbustives hydromorphes en sont de très bons exemples. Ces trois thèmes passent respectivement de 67,9 à 8,7 p.100, 90,1 à 16,3 p.100 et de 130,4 à 19,9 p.100. En revanche, la steppe herbeuse des ensablements et interdunes enregistre un gain de précision très minime : 14 points. L'apport de la télé-détection est ici moins spectaculaire.

L'adjonction de données satellitaires est nettement mise en évidence. Il ne reste plus qu'un seul thème dont le Coefficient de Variation (CV) dépasse les 20 p.100 (la steppe arbustive des ensablements : 31,6 p.100) ; tous les autres ont un écart type dont la valeur est inférieure à 20 p.100 de la superficie estimée. La précision de ces résultats permet ainsi un diagnostic plus rigoureux et peut donc servir de base à des propositions d'aménagement. Des constatations analogues ont été effectuées sur d'autres secteurs de suivi du Sahel sud-mauritanien¹

L'apport de la télé-détection se manifeste donc par une amélioration globale de la précision des résultats. Il permet de redresser un échantillonnage mal adapté à certains postes d'occupation du sol, comme les thèmes de faible étendue mais radiométriquement spécifiques. En revanche, **la télé-détection ne révélera pas ce que l'enquête de terrain n'a pu saisir par inadéquation du taux de sondage à la taille et à la répartition du thème considéré.** Nous en avons un très bon exemple avec le thème "surface en eau", les mares pour cette zone, qui est compact et de faible étendue. Le plan de sondage peut très bien les "éviter" avec leur superficie de 1,6 p.100. Leur signature spectrale très particulière fait que l'on constate rapidement leur absence, alors que d'autres thèmes radiométriquement moins contrastés ne seront pas enquêtés en tant que tels mais probablement intégrés à l'un de leurs voisins sans que l'on s'en aperçoive.

Pour les thèmes inventoriés, il est possible d'évaluer numériquement l'efficacité de la régression qui est effectuée entre les relevés de terrain par segment (y) et leurs homologues cartes (x) (Cf. tableau 3). Le coefficient de corrélation (r) qui caractérise cette régression permet d'évaluer l'efficacité de l'estimateur par la régression, donc de l'apport des données satellitaires, par la formule suivante² :

$$E = \frac{1}{(1-r^2)}$$

Une efficacité E égale à 2,6 correspond à un coefficient de corrélation de 0,78 et signifie que sans la télé-détection, il faudrait 2,6 fois plus de segments d'enquête pour faire l'inventaire du thème "steppe herbeuse très claire des dépressions hydromorphes" et obtenir un coefficient de variation identique (17,4 p.100 Cf. tableau 2). Nous disposons ici d'un argument supplémentaire en faveur de l'utilisation de données satellitaires lorsqu'une enquête de terrain est effectuée avec une méthodologie qui utilise un sondage aléatoire.

¹DE WISPELAERE ; GODARD ; CARRIERE (à paraître)

²IAT 1989 p.7

Tableau 3 : Efficacité de la télédétection

Thèmes	Sigles	C.de Correl.	Efficacité
		r	E
Surface en eau	Se	0,99	50,3
Culture irriguée	Ci	0,99	50,3
Steppe arbustive hydromorphe	Sah	0,58	1,5
St. herbeuse. très claire des			
dépressions hydromorphes	Shy	0,78	2,6
Sol nu sablo-limoneux	Sns	0,98	25,3
Sol nu argileux	Sna	0,98	25,3
Sable vif	Sv	0,99	50,3
Steppe herbeuse très claire des dunes	Shd	0,99	50,3
St. herb. des ensablements et interdunes	Shi	0,85	3,6
St. arbustive des ensablements	Sae	0,80	2,8
Steppe "boisée" gommerai	Sb	0,99	50,3

D'un point de vue thématique, l'étude du tableau 2, que ce soit avec ou sans l'apport de la télédétection, permet de définir la zone de Rosso comme étant essentiellement déforestée et orientée vers l'élevage en raison de la quasi absence de culture et de la prédominance des steppes.

En regroupant les quatre thèmes les plus importants en superficie (partie utilisant la télédétection) on arrive à 83 p.100 de l'occupation du sol. Ces 83 p.100 se ventilent en environ 25 p.100 de sol nu, 40 p.100 de steppe herbeuse très claire (dunes plus bas-fonds) et 20 p.100 d'une steppe herbeuse un peu plus fournie . La moyenne de leurs coefficients de variation est d'environ 10 p.100. On notera que les thèmes où la présence de l'arbre est significative représente ici moins de 3 p.100 des superficies enquêtées. De même, la dégradation touche plus de 60 p.100 de la zone étude.

Comme nous l'avons vu tout au long de cet exposé, cette méthode d'enquête se veut souple, légère, peu coûteuse en temps et en homme et enfin fiable et précise. Bien qu'encore à l'essai sur d'autres secteurs sahéliens, il est d'ores et déjà possible d'affirmer qu'elle remplit parfaitement le rôle qui lui a été assigné.

Le gain de temps, lors de la phase de préparation, est appréciable lorsqu'il s'agit de déterminer les paramètres régissant la procédure de sondage. Il en va de même pour la fiabilité des résultats qui, grâce à l'apport de l'imagerie satellitaire, s'accommode relativement bien d'un plan de sondage allégé.

Le sondage par segment carré est maintenant entré dans sa phase opérationnelle en Europe avec, à la suite du SCEES, l'équipe du CCR d'Ispra qui l'utilise en statistiques agricoles. Il n'est pas question actuellement que de telles utilisations deviennent routinières sur l'ensemble de l'Afrique sahélienne. En revanche, avec quelques aménagements, ces méthodologies, utilisant les ressources satellitaires, peuvent être d'une aide précieuse sur des opérations de développement intégré, principalement dans les phases de diagnostic et de suivi.

GODARD (V.) - 1990 - Optimisation des enquêtes de terrain pour la cartographie de l'occupation du sol et le diagnostic de l'état de l'environnement. in : Téledétection et Tiers Monde, Actes de la table ronde internationale CEGET-CNRS : Apports de la téledétection spatiale à l'étude des paysages et des systèmes agraires tropicaux. Talence 25-27 octobre 1989 : 87-105.

Ouvrages cités :

BDPA
1988

Prise de vue aérienne basse altitude échantillonnée sur l'arrondissement de Mayahi. Principes et exploitation. Paris, BDPA, 22 p.

BRION (Ph.), BERTIN (M.), LABAT (Ph.), 1982, Sondages, Paris, SCEES, Série S, 40 p., (Méthodes et Applications Scientifiques n°5)

CHADULE (Groupe), 1974, Initiations aux statistiques en Géographie, Paris, Masson, 191 p.

DESABIE (J.), 1966, Théorie et pratique des sondages, Paris, Dunod, 480 p.

DE WISPELAERE (G.), GODARD (V.), CARRIERE (M.), (à paraître), Expérimentation méthodologique pour le suivi de l'environnement par téledétection dans le Trarza, le Gorgol et l'Assaba (Mauritanie), Maisons-Alfort, IEMVT.

FOURNIER (Ph.), 1986, Enquête sur l'utilisation du territoire effectuée en 1985 par la méthode des segments, Paris, SCEES, Série S, 63 p., (Méthodes et Applications Scientifiques n°13).

FOURNIER (Ph.) ; GILG (J.P.), 1985, Téledétection et observations-terrain, Toulouse, cours du GDTA, 22P.

GODARD (V.), DOLLE (V.), VAYSSIERE (J.F.), (à paraître), Un outil de diagnostic rapide pour l'agriculture oasienne. Mise au point méthodologique pour l'utilisation de données satellitaires SPOT dans la région de l'Assaba mauritanien, Séminaire sur les systèmes agricoles oasiens, Tozeur (Tunisie) 19-20 novembre 1988.

GROSBRAS (J.M.), 1987, Méthodes statistiques des sondages, Paris, Economica, 331 p.

HUSCH (B.) , 1971, Préparation d'un inventaire forestier, Rome, FAO, 135 p. (Etudes sur les forêts et les produits forestiers n°17).

IAT, 1989, Inventaires régionaux, Flash Télé-Agri-News, Inst. des Applic. de la Téléd., Service Spécialisé Agriculture (CECD), Ispra (Italie), 2/89 : 2-15.

KEERSMAECKER (M.L. de), 1987, Stratégie d'échantillonnage des données de terrain intégrées dans l'analyse d'image satellitaire, L'espace géographique, XIV-3 : 195-205.

MEYER-ROUX (J.), SIGMAN (R.), CRAIG (M.), OZGA (M.), 1981, Cartographie et statistique en téledétection, Cahier de la statistique agricole, 6/6 nov.-déc. 1981 : 9-14.

SPOT, 1986, Guide des utilisateurs de données SPOT, Toulouse, CNES/SPOT IMAGE, Vol. 2 Guide pratique : np.

GODARD (V.) - 1990 - Optimisation des enquêtes de terrain pour la cartographie de l'occupation du sol et le diagnostic de l'état de l'environnement. in : *Téledétection et Tiers Monde, Actes de la table ronde internationale CEGET-CNRS : Apports de la téledétection spatiale à l'étude des paysages et des systèmes agraires tropicaux*. Talence 25-27 octobre 1989 : 87-105.

Annexe

Extrait de la légende de la carte d'occupation du sol du secteur de Rosso (Mauritanie) (à paraître)

Thèmes	Sigles
Surface en eau	Se
Culture irriguée	Ci
Steppe arbustive hydromorphe	Sah
St. herbeuse. très claire des dépressions hydromorphes	Shy
Sol nu sablo-limoneux	Sns
Sol nu argileux	Sna
Sable vif	Sv
Steppe herbeuse très claire des dunes	Shd
St. herb. des ensablements et interdunes	Shi
St. arbustive des ensablements	Sae
Steppe "boisée" gommeriaie	Sb