

# Échantillonnage à deux degrés dans des bases aréolaires sur des segments carrés pour des sondages agricoles

F.J. GALLEGO, J. DELINCÉ et E. CARFAGNA<sup>1</sup>

## RÉSUMÉ

Dans le projet MARS (Monitoring Agriculture with Remote Sensing ou surveillance de l'agriculture par télédétection) de la CE (Communauté européenne), on utilise des bases aréolaires définies à l'aide d'une grille carrée pour estimer les superficies au moyen d'enquêtes sur le terrain et d'images-satellites à haute résolution. Quoique coûteuses, ces images-satellites sont utiles pour l'estimation des superficies; leur utilisation pour l'estimation du rendement n'est pas encore opérationnelle. Pour combler cette lacune, on emploie aussi les éléments de l'échantillon (segments) de l'enquête aréolaire pour échantillonner les exploitations agricoles à l'aide d'un gabarit de points superposé au segment. Le plus souvent, on utilise un nombre fixe de points par segment. On demande aux exploitants agricoles de fournir des données globales pour leur exploitation et on calcule les estimations à l'aide d'une approche de type Horvitz-Thompson. Les principaux problèmes sont la difficulté de repérer les exploitants et la vérification des cas où les instructions ont été mal comprises. On obtient de bons résultats pour la superficie et pour la production des cultures principales. Les bases aréolaires doivent être complétées par des listes (bases de sondage multiples) pour donner des estimations fiables dans le cas des animaux d'élevage.

MOTS CLÉS: Base aréolaire; sondage par points; sondage par segments; sondage d'exploitations agricoles.

## 1. INTRODUCTION

L'objet principal de cet article est de présenter la méthode utilisée dans le cadre du projet MARS (Monitoring Agriculture with Remote Sensing ou surveillance de l'agriculture par télédétection) de la Communauté européenne (CE) pour prélever un échantillon d'exploitations agricoles dans une base aréolaire. Dans ce projet, le prélèvement d'un échantillon d'exploitations n'est pas une activité essentielle, mais plutôt une façon d'éviter le problème de la capacité limitée des images-satellites, en particulier pour l'estimation du rendement. Nous allons d'abord présenter brièvement le projet MARS lui-même, étant donné que très peu d'articles en parlent dans les périodiques statistiques (Ambrosio 1993; Gallego 1992). On peut en trouver d'autres présentations dans des communications présentées lors de conférences (Meyer Roux 1990; Delincé 1990; Sharman et coll. 1992; Carfagna et coll. 1994) ou dans des périodiques traitant de télédétection (González et coll. 1991; Gallego et coll. 1993).

## 2. LE PROJET MARS DE LA COMMUNAUTÉ EUROPÉENNE

Le projet MARS a été lancé en 1988 pour évaluer et développer des applications opérationnelles de la télédétection dans le domaine des statistiques agricoles. Ce projet est réalisé par l'Institut des applications de la télédétection (IAT) du Centre commun de recherche (CCR) de la CE. La plupart des activités réalisées de 1988 à 1993 ont été divisées en quatre parties principales nommées "actions":

- 1) Inventaires régionaux des cultures.
- 2) Surveillance de la végétation.
- 3) Modèles agrométéorologiques.
- 4) Estimations rapides au niveau de la CE.

Des travaux sont aussi effectués dans d'autres domaines connexes, tels que l'échantillonnage à l'aide de bases aréolaires. Nous nous intéresserons ici à une méthode d'échantillonnage utilisée dans le cadre de l'action 1 "inventaires régionaux", mais nous allons d'abord parler brièvement des autres actions.

### 2.1 Surveillance de la végétation

Dans cette action, on s'intéresse aux images-satellites à faible résolution obtenues à l'aide du AVHRR (Advanced Very High Resolution Radiometer ou radiomètre perfectionné à très haute résolution) des satellites de la NOAA. Sur ces images, chaque pixel couvre une superficie d'environ 1 km<sup>2</sup> en visée verticale. Les principaux objectifs sont l'élaboration d'un logiciel convivial pour le prétraitement de ces images et la constitution d'une banque de données de séries chronologiques d'indices de végétation et d'autres indicateurs pour environ 3,000 unités de surveillance dans la CE. Ces unités de surveillance n'ont pas encore été définies de façon définitive. Elles devraient être des régions géographiques couvrant en gros de 500 à 1,000 km<sup>2</sup> avec un indice de végétation ou de coloration verte plus ou moins homogène (Houston 1984; Goward 1991).

### 2.2 Modèles agrométéorologiques

On élabore actuellement des modèles généraux et propres aux cultures à partir des données provenant d'un réseau

<sup>1</sup> F.J. Gallego, Centre commun de recherche des Communautés européennes, tp. 440, 21020 Ispra (Varèse) Italie; J. Delincé, Commission européenne DG VI, Loi 120, 4-23/1049 Bruxelles, Belgique; E. Carfagna, Département de Statistiques, Université de Bologne, V. Belle Arti 41, 40126 Bologne, Italie.

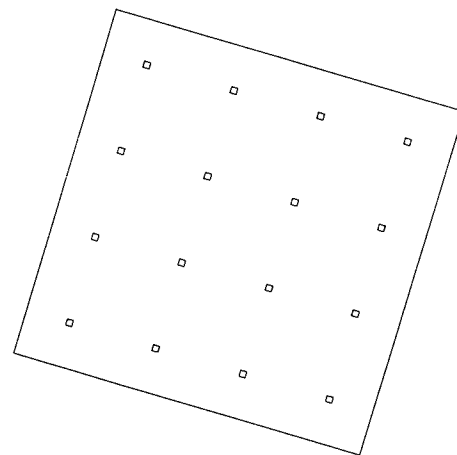
d'environ 650 observatoires météorologiques d'Europe et des régions voisines. Ce modèle, le CGMS (Crop Growth Monitoring System ou système de surveillance de la croissance des cultures), développé en collaboration avec le WOFOST (World Food Studies Centre – ou centre mondial d'études sur les aliments – situé à Wageningen aux Pays-Bas), fait aussi appel à d'autres données, par exemple sur le sol et l'altitude, ainsi qu'à des informations sur la physiologie des plantes (van Diepen 1989; van Lanen 1992). La télédétection (images à faible résolution) sera utilisée plus tard pour l'interpolation géographique de données météorologiques observées au sol. Actuellement, on calcule les paramètres du modèle pour chaque case d'une grille de 50 km sur 50 km.

### 2.3 Estimations rapides au niveau de la CE

Le but principal est d'obtenir des estimations rapides de la variation de la superficie et du rendement des cultures annuelles, comparativement aux données correspondantes de l'année précédente, au moyen d'un plan de sondage à deux degrés: on utilise 53 sites (figure 1) de 40 km sur 40 km avec un échantillon de 16 segments carrés de 700 m sur 700 m (figure 2) dans chacun des sites. Les données sont acquises par interprétation photographique d'images au moyen des capteurs TM des satellites Landsat et XS des satellites SPOT. On analyse en moyenne trois images pour chaque site avec un minimum d'information de terrain, c'est-à-dire une connaissance générale des cultures dominantes dans chaque région. Un relevé de terrain est effectué en vue d'une validation a posteriori de la photo-interprétation. Un rapport mensuel est produit (de mars à novembre) avec une mise à jour des estimations. Chaque rapport doit utiliser toutes les images acquises plus de 15 jours avant la production du rapport.



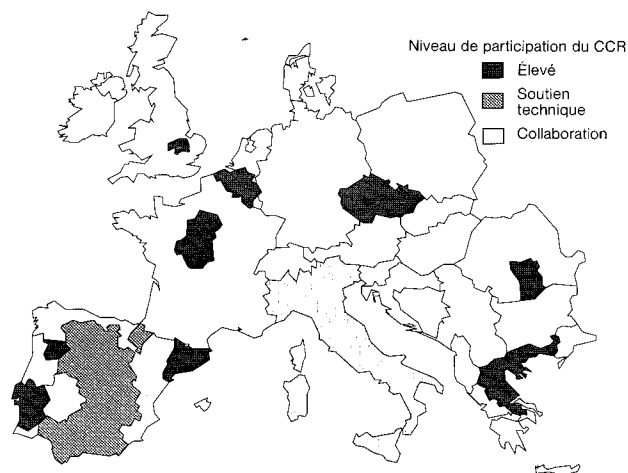
**Figure 1.** Échantillon de 53 sites pour l'estimation rapide des cultures dans la CE.



**Figure 2.** Segments dans un site (estimations rapides dans la CE).

### 2.4 Inventaires régionaux des récoltes au moyen d'un sondage par segments et de la télédétection

L'objectif de cette action était de mettre en oeuvre, d'adapter et d'évaluer des méthodes d'estimation de la superficie cultivée et de la production agricole fondées sur un échantillonnage effectué à partir d'une base aréolaire et sur des images-satellites. Quand l'IAT a commencé en 1988 à mettre cette action en oeuvre sur cinq zones pilotes d'environ 20,000 km<sup>2</sup> chacune, la priorité a été accordée aux cultures annuelles: blé tendre, blé dur, orge, colza, légumineuses à grain, tournesol, maïs, coton, tabac, betterave à sucre, pommes de terre, riz et soja, ainsi qu'aux jachères. On accorde de plus en plus d'attention aux cultures permanentes, aux pâturages et aux utilisations non agricoles des terres.



**Figure 3.** Régions européennes dans lesquelles des enquêtes par segments ont été réalisées en 1992.

Depuis 1990, l'IAT a progressivement transféré l'initiative aux administrations régionales ou nationales qui souhaitent utiliser des sondages effectués à l'aide de bases aréolaires fondées sur des segments. De façon générale, les activités ont été transférées aux pays du sud de la CE et aux anciens pays communistes d'Europe centrale, qui ont manifesté beaucoup d'intérêt pour cette méthode (figure 3). Dans certains cas, comme en Italie, il n'y a qu'un échange de points de vue entre le projet national et l'IAT.

**2.4.1 Prélèvement d'un échantillon de segments sur une grille carrée**

Il existe deux façons principales de constituer une base aréolaire fondée sur des segments: on peut dessiner les segments sur des cartes topographiques ou cadastrales en suivant les routes, les rivières ou les limites des champs (parfois appelés segments cadastraux). L'échantillon est habituellement tiré au moyen d'une procédure à deux degrés avec unités primaires d'échantillonnage intermédiaires pour réduire le fardeau que représente l'établissement de la base de sondage (Cotter 1987), travail qui demeure, dans tous les cas, une opération lourde.

Nous utilisons habituellement des bases aréolaires fondées sur une grille carrée (Gallego et Delincé 1994), que l'on peut définir beaucoup plus rapidement. Nous employons en général (mais pas nécessairement) des images-satellites pour effectuer une stratification avant l'échantillonnage.

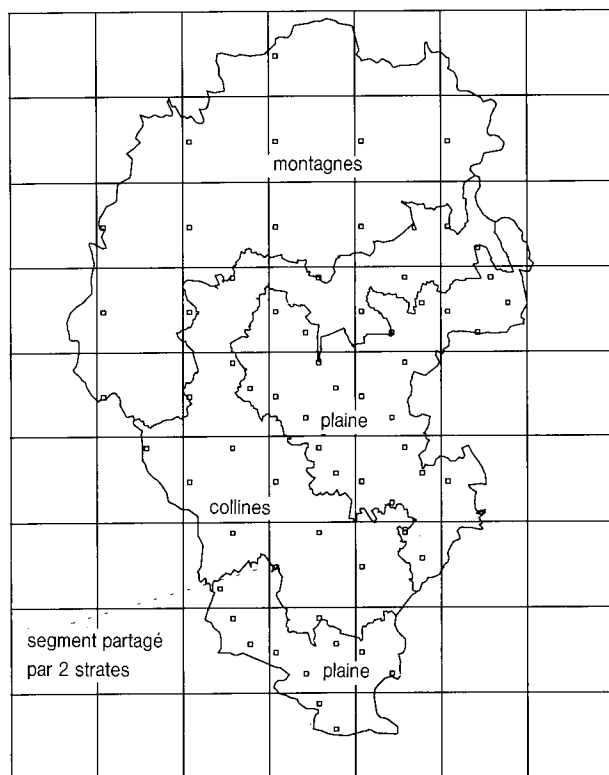


Figure 4. Exemple d'échantillon aréolaire avec segments carrés et blocs carrés.

La figure 4 donne un petit exemple de ce genre d'échantillon avec une stratification très simple et des segments de 25 ha (hectares). L'échantillonnage est systématique, avec répétition d'une configuration dans des blocs carrés. Ici, la dimension des blocs est de 10 km sur 10 km et la configuration a quatre répétitions dans la strate la plus agricole (la plaine), deux répétitions dans les collines et une dans les montagnes.

Le principal inconvénient de cette méthode est le problème lié à la gestion des segments qui chevauchent la limite entre deux strates (figure 5). Trois solutions sont mises à l'essai pour résoudre ce problème: 1) adapter la stratification à la grille d'échantillonnage, 2) séparer les segments qui chevauchent les limites en pièces appartenant à différentes strates et 3) conserver seulement la plus grosse de ces pièces.

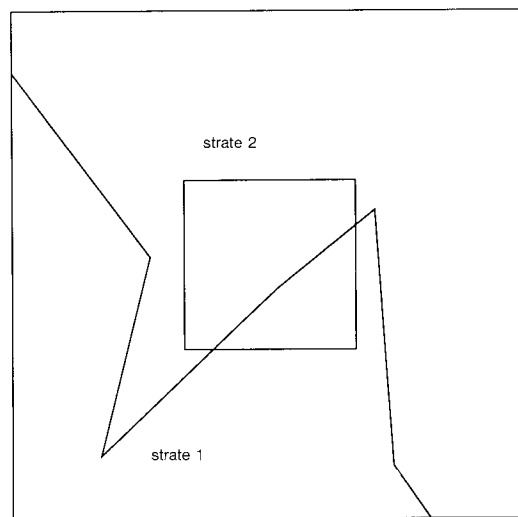
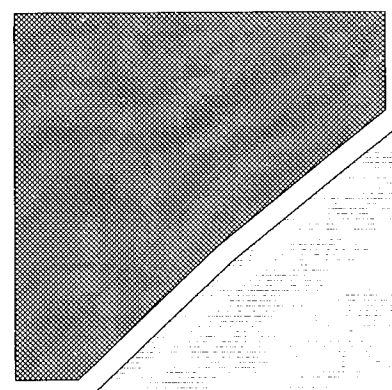


Figure 5. Un segment peut être séparé par une limite de strate.

Il n'existe pas de forte corrélation entre les erreurs non dues à l'échantillonnage les plus fréquentes – c'est-à-dire les déplacements de l'emplacement et l'inexactitude dans la forme ou la dimension du segment – et l'utilisation du terrain. On n'a pas trouvé d'influence importante sur les estimations de la superficie ou sur leur précision.

La configuration d'échantillonnage à répéter dans chaque bloc est tirée au hasard, avec une restriction quant

à la distance entre les segments pour éviter de choisir des segments trop rapprochés. Ici, on peut utiliser des estimateurs par grappe plutôt que les formules normales pour l'échantillonnage aléatoire (Fuentes 1994; Ambrosio 1993). Avec l'échantillonnage systématique, on risque d'introduire un biais s'il existe un effet cyclique dans le paysage avec une période qui coïncide avec la taille du bloc (10 km dans l'exemple), mais cela est très peu probable. Le seuil de distance entre les segments peut amener une surestimation des erreurs types si la corrélation géographique a une valeur positive élevée pour des distances inférieures au seuil.

La taille des segments varie d'une région à l'autre selon le paysage agricole, en particulier selon la superficie des champs. Dans la République tchèque, la superficie d'un segment était de 400 ha. Pour l'enquête aréolaire, des enquêteurs repèrent les segments, tracent les champs sur une feuille transparente placée sur une photographie aérienne et notent l'utilisation du terrain. Des surveillants visitent de nouveau 5 à 10% des segments pour s'assurer qu'il n'y a pas eu d'erreur dans le travail sur le terrain. On n'utilise d'images-satellites ni pour l'enquête proprement dite ni pour le sondage agricole, mais ces images peuvent être employées pour améliorer la précision des estimations des superficies, comme nous allons le voir dans la prochaine section.

#### 2.4.2 Amélioration des estimations de superficie au moyen d'images-satellites

On a évalué l'emploi des images-satellites à haute résolution obtenues à l'aide des capteurs TM des satellites Landsat et XS des satellites SPOT, et on les utilise encore, dans une certaine mesure, pour améliorer les estimations obtenues à partir de l'enquête sur le terrain portant sur un échantillon de segments. La méthode la plus habituelle utilise un estimateur par régression sur des images classées. On a aussi fait l'essai d'un autre estimateur basé sur des grilles de correction, et les résultats sont très proches de ceux obtenus pour l'estimateur par régression (Hay 1988; Gallego 1994).

Les conclusions de cette évaluation sont semblables à celles du Département de l'agriculture des États-Unis (Allen 1990): l'utilisation d'images-satellites pour l'estimation des superficies est opérationnelle, mais encore trop coûteuse pour le gain d'efficacité obtenu. On peut atteindre le seuil de rentabilité en améliorant l'automatisation du traitement des images, car, sur le marché européen, le coût du traitement des images à cette fin est très supérieur au coût des images elles-mêmes. Ce seuil a presque été atteint en Grèce avec les images du capteur TM des satellites Landsat. Giovacchini (1992) présente des conclusions différentes sur l'analyse des coûts.

### 3. ÉCHANTILLONNAGE D'EXPLOITATIONS AGRICOLES AU MOYEN DE POINTS

Pour les enquêtes agricoles dans la Communauté européenne, les exploitations agricoles sont traditionnellement choisies au moyen d'une liste (Eurostat 1991). La liste est

un recensement des exploitations qui dépassent une certaine taille. Dans de nombreux pays, un recensement agricole est effectué tous les dix ans et il est rarement sinon jamais mis à jour. Il peut donc y avoir une différence importante entre la base de sondage et la population réelle au moment de l'enquête. La situation est pire dans les pays d'Europe centrale faisant partie de l'ancien bloc oriental (la région comprise entre la Pologne et la Roumanie-Bulgarie), où le changement dans la structure de la propriété foncière est si rapide que le recensement peut ne plus exister pour les fermes privées et être trop ancien pour les coopératives.

On peut facilement définir des bases aréolaires sur des segments carrés quand on connaît les limites géographiques de la région. Dans plusieurs pays, on utilise aussi un sous-échantillon de ces segments pour sélectionner des exploitations à l'aide d'un gabarit de points superposé au segment. Cette méthode a fait l'objet d'expériences en Allemagne, au Portugal, en Italie (Carfagna 1991) et en Espagne, et on l'utilise maintenant couramment en Grèce, en Roumanie et en République tchèque.

Le gabarit est le même pour tous les segments d'une strate et il est habituellement symétrique de façon à réduire le risque de biais attribuable à un lieu géographique particulier. Les données sont obtenues seulement pour les exploitations correspondant aux points qui tombent sur une superficie agricole utilisée (SAU).

La définition de la SAU employée pour le travail sur le terrain est adaptée à chaque système national. Les bâtiments agricoles et les pacages sont inclus dans certains pays, exclus dans d'autres. Ce qui importe, c'est que la définition utilisée soit compatible avec la définition de la colonne SAU employée pour le calcul (tableau 1).

**Tableau 1**  
Observations produites par les points sélectionnés dans le segment de la figure 6

Segment	Point	SAU	Cultures permanentes	Blé		Orge	
				Superficie	Production	Superficie	Production
1	1	19	4	12	64	0	0
1	2	0	0	0	0	0	0
1	3	0	0	0	0	0	0
1	4	35	0	24	131	3	12
1	5	35	0	24	131	3	12
2	...	...	...	...	...	...	...

Dans l'exemple de la figure 6, le point 3 est tombé sur un terrain boisé et le point 2, sur une zone bâtie. Ces deux points produiront deux enregistrements de valeur zéro dans le fichier des exploitations agricoles. L'enquêteur devra trouver les exploitants dont les terres correspondent aux trois autres points. L'exploitation correspondant au point 1 a, dans le segment, d'autres champs qui seront inclus implicitement dans l'enquête, mais l'enquêteur n'aura pas à trouver si ces champs existent. Les points 4 et 5 appartiennent à la même exploitation, qui figurera donc deux fois dans le fichier des exploitations agricoles (tableau 1).



Figure 6. Segments avec un gabarit de 5 points pour la sélection d'exploitations agricoles.

On trouve les exploitants et on leur demande de fournir des données globales sur l'exploitation agricole, y compris sur la superficie totale consacrée à chaque culture visée par l'enquête et sur la production correspondante. On ne pose pas de question sur la production de chaque champ ou de l'ensemble des champs qui se trouvent dans le segment. Cela n'est pas nécessaire parce que, dans les formules finales utilisées pour calculer les estimations (formules 2 et 3 de la section 4.1), on n'a pas besoin de la superficie cultivée ou de la production agricole de la parcelle de terre.

Les instructions relatives à l'enquête sur le terrain sont habituellement transmises du CCR aux administrations nationales. Ces dernières expliquent les instructions aux coordonnateurs régionaux, qui donnent les renseignements aux enquêteurs. Les instructions peuvent être modifiées à certaines de ces étapes. Il est parfois difficile de s'assurer que les instructions n'ont pas été mal comprises, en partie du fait que les différences linguistiques sont un obstacle sérieux à la communication directe avec les enquêteurs. Dans certains pays (par exemple en Espagne), les exploitants agricoles vivent principalement dans des centres urbanisés assez grands et sont difficiles à trouver; ce qui fait qu'il peut manquer beaucoup de données.

#### 4. ESTIMATIONS FONDÉES SUR DES EXPLOITATIONS AGRICOLES ÉCHANTILLONNÉES PAR POINTS

Nous supposons que la population  $\Omega$  des segments est divisée en strates  $\Omega_h$ ,  $h = 1, \dots, H$ , que la taille totale de la population est de  $N$  segments ( $N_h$  pour la strate  $\Omega_h$ ) et que la taille de l'échantillon est de  $n$  segments ( $n_h$ ). La taille de notre échantillon de points dans chaque segment sera  $K_i$ , déjà fixée; en général, nous avons  $K_i = K$ , qui

est une valeur constante pour toutes les strates; de ce nombre  $F_i$  correspondent aux exploitations agricoles sur lesquelles tombent ces points. Chaque segment  $i$  a une SAU totale  $U_i$ .

Nous utilisons un plan de sondage à deux degrés. Au premier degré, le segment  $i$  est choisi avec probabilité  $p_i = 1/N_h$  dans chacun des  $n_h$  essais. Au deuxième degré, l'unité n'est pas l'exploitation agricole mais la parcelle (SAU dans un segment appartenant à la même exploitation). La parcelle  $k$  du segment  $i$  a une superficie  $T_{ik}$ . La SAU totale de l'exploitation est  $A_{ik}$  pour tous les segments.  $U_i$  est la somme de la superficie  $T_{ik}$  des parcelles dans le segment  $i$ .

La méthode présentée ci-après ressemble beaucoup à la méthode dite de l'"estimateur pondéré pour un segment" utilisée aux États-Unis et au Canada (Nealon 1984).

#### 4.1 Estimations fondées sur des points correspondant à une exploitation agricole et sur des points ne correspondant pas à une exploitation agricole

Il y aura  $K - F_i$  observations (exploitations agricoles fictives) avec valeur 0 correspondant aux points qui se trouvent à l'extérieur de la SAU.

L'échantillonnage par points signifie que les parcelles sont choisies avec remise et avec une probabilité  $p_{ik}$  proportionnelle à la superficie  $T_{ik}/D_i$  (il n'est pas nécessaire de connaître la valeur de  $T_{ik}$ ), où  $D_i$  est la taille du segment déterminée par la conception de la base de sondage. Nous supposons implicitement que la région étudiée est plane. Il peut s'introduire un léger biais du fait que les cultures annuelles sont habituellement exploitées sur des terres qui ne sont pas parfaitement planes et que les pâturages et les régions ne correspondant pas à une SAU se trouvent souvent sur des terrains avec une pente plus abrupte.

L'échantillonnage est effectué avec remise: une exploitation peut être choisie plus d'une fois, ce qui donne des formules plus simples pour estimer la variance. À proprement parler, la probabilité de sélection conjointe des exploitations  $k$  et  $k'$  dans l'échantillon est  $p_{ikk'} \neq p_{ikk} \times p_{ikk}$  comme cela serait le cas si l'on tirait de façon indépendante les différents points du gabarit, puisqu'il y a habituellement une distance relativement grande entre ces points. Nous ne tiendrons pas compte de ce fait ici.

$W_{ik}$  sera une quantité additive pour une exploitation agricole, le plus souvent la production ou la superficie d'une culture particulière. Il est évident que le rendement  $n$ 'est pas une variable additive.

Comme nous n'avons pas de renseignements sur la distribution de  $W_{ik}$  dans l'exploitation agricole, nous créons une variable fictive  $X$  qui est uniformément distribuée et qui a, par définition, le même total que  $W$  pour chaque exploitation:

$$X_{ik} = \frac{T_{ik}}{A_{ik}} W_{ik}. \quad (1)$$

L'estimation du total de  $X$  et du total de  $W$  sont des problèmes équivalents.

La version à deux degrés de l'estimateur de Horvitz-Thompson pour le total de  $X$  dans la strate  $\Omega_h$  donne:

$$\hat{X}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} \frac{\hat{X}_i}{p_i} = \frac{N_h}{n_h} \sum_{i=1}^{n_h} \frac{1}{K_i} \sum_{k=1}^{K_i} \frac{X_{ik}}{p_{ik}} = \frac{N_h}{n_h} \sum_{i=1}^{n_h} \frac{D_i}{K_i} \sum_{k=1}^{K_i} \frac{W_{ik}}{A_{ik}}. \quad (2)$$

Cela signifie que, même si l'unité secondaire d'échantillonnage est la parcelle, nous n'avons pas besoin de connaître sa superficie ni  $X_{ik}$ ; nous avons seulement besoin des renseignements globaux sur l'exploitation agricole.

L'estimateur est une fonction linéaire des estimations des segments sélectionnés. Sa variance dans la strate  $\Omega_h$  peut être estimée par la méthode suivante (Cochran 1977, section 11.6):

$$\hat{V}(\hat{X}_h) = \frac{N_h^2}{n_h} \left(1 - \frac{n_h}{N_h}\right) \sum_{i=1}^{n_h} \frac{(\hat{X}_i - \bar{X}_h)^2}{n_h - 1} + \frac{N_h}{n_h} \sum_{i=1}^{n_h} \frac{1}{K_i(K_i - 1)} \sum_{k=1}^{K_i} \left(\frac{W_{ik}D_i}{A_{ik}} - \hat{X}_i\right)^2. \quad (3)$$

Les estimations du total sont:

$$\hat{X} = \sum_{h=1}^H \hat{X}_h \quad \hat{V}(\hat{X}) = \sum_{h=1}^H \hat{V}(\hat{X}_h). \quad (4)$$

Les superficies cultivées sont actuellement estimées à partir de l'enquête par segments avec des données de terrain plus objectives (une observation directe faite par l'enquêteur sur le terrain), bien qu'un certain biais puisse apparaître à cause de la localisation imparfaite des segments sur le terrain. Les sondages agricoles fournissent des estimations à la fois de la superficie et de la production, mais ces estimations peuvent être entachées d'un biais plus important à cause de la non-réponse et d'une tendance subjective de l'exploitant agricole qui peut dépendre du fait qu'au moment de l'enquête ce dernier est préoccupé plus par les impôts que par les subventions ou vice-versa. La comparaison des deux estimations de la superficie, celle obtenue à partir de l'enquête par segments et celle obtenue à partir du sondage agricole, peut être utile pour vérifier si un biais n'a pas été introduit dans l'estimation de la production fondée sur le sondage agricole.

Il est aussi possible d'obtenir des estimations pour les bovins, mais les résultats seront présumément mauvais s'il y a un nombre élevé d'exploitations qui n'ont pas du tout de SAU et qui ne seront donc pas échantillonnées: la couverture de la base aréolaire ne sera pas complète dans ce cas. Par contre, il se peut qu'il n'y ait pas de corrélation entre le nombre d'animaux d'élevage et la SAU et, par conséquent, la probabilité de sélection. Les estimations seraient alors inefficaces.

On a été écrit pour les ordinateurs personnels (PC) (Dicorato 1993) un programme en langage C qui permet de calculer des estimations à l'aide de cette méthode. La principale partie du programme a d'abord été écrite pour calculer des estimations dans le cadre d'une enquête par segments.

#### 4.2 Estimation fondée seulement sur des exploitations agricoles correspondant à des points

Nous allons voir une autre option, où l'on utilise seulement des points qui tombent sur la SAU. Dans ce cas, nous commençons par déterminer  $F_i$ , c'est-à-dire le nombre de points qui tombent sur la SAU. (Il arrive souvent que  $F_i = F_h$ , constante dans chaque strate). Dans le segment  $i$ , nous observons autant de points qu'il le faut pour obtenir  $F_i$  points dans la SAU. Si le segment  $i$  ne comprend pas de SAU, nous ajoutons une observation (une exploitation agricole fictive) avec valeurs 0. Cela constitue en fait une stratification implicite au second degré ou une stratification des unités primaires (segments) en deux strates: SAU et superficies qui ne sont pas des SAU. La strate des superficies qui ne sont pas des SAU n'est pas échantillonnée. Dans ce cas, les équations (2) et (3) doivent être adaptées en remplaçant  $K_i$  par  $F_i$  et  $D_i$  par  $U_i$ . Il se peut que certaines incohérences surgissent dans les régions vallonnées parce que  $A_{ik}$  provient de la déclaration de l'exploitant agricole et  $U_i$ , de segments tracés sur des photographies aériennes.

$$\hat{X}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} \frac{\hat{X}_i}{p_i} = \frac{N_h}{n_h} \sum_{h=1}^{n_h} \frac{1}{F_i} \sum_{k=1}^{K_i} U_i \frac{W_{ik}}{A_{ik}}, \quad (5)$$

$$\hat{V}(\hat{X}_h) = \frac{N_h^2}{n_h} \left(1 - \frac{n_h}{N_h}\right) \sum_{i=1}^{n_h} \frac{(\hat{X}_i - \bar{X}_h)^2}{n_h - 1} + \frac{N_h}{n_h} \sum_{i=1}^{n_h} \frac{1}{F_i(F_i - 1)} \sum_{k=1}^{K_i} \left(\frac{W_{ik}U_i}{A_{ik}} - \hat{X}_i\right)^2. \quad (6)$$

Le deuxième membre de l'équation (6) est nul pour des segments qui n'ont pas de SAU. Ce membre ne peut être calculé si  $F_i = 1$  à cause de la non-réponse. On peut attribuer une valeur 0, bien que cela amènera une sous-estimation de la variance à l'intérieur des segments, qui est relativement faible selon les calculs faits à partir des données disponibles (Carfagna 1992).

Cette méthode n'a été utilisée qu'une fois pour résoudre un problème lié à la mauvaise compréhension d'instructions relatives au travail sur le terrain, qui aurait dû être effectué d'après la méthode décrite à la section 4.1. Toutefois, les avantages et les inconvénients des deux méthodes ne sont pas évidents et, jusqu'ici, aucune comparaison systématique n'a été faite pour la même région et la même année. Le fait de n'utiliser que des points correspondant à des exploitations agricoles peut accroître le coût de l'enquête si le nombre de points par segment doit être maintenu

constant, mais les points relatifs à des superficies autres que des SAU et qui sont supprimés correspondent à des valeurs nulles de  $W_{ik}$ , et leur suppression peut entraîner une réduction de la variance.

### 4.3 Exploitations agricoles avec des champs dans différentes strates

À première vue, l'estimateur (2) semble supposer qu'une exploitation  $k$  choisie par un point dans la strate  $\Omega_h$  est entièrement incluse dans cette strate. Il est évident qu'une exploitation peut avoir des champs dans différentes strates, et il faut se demander si ce fait peut avoir une incidence négative sur la fiabilité des résultats.

Rappelons que la variable utilisée n'est pas réellement  $W_{ik}$ , mais  $X_{ik}$ , définie pour chaque parcelle. Le total des  $W$  ne coïncide pas avec le total des  $X$  dans chaque strate, mais c'est le cas pour toute la région pourvu que

$$\sum_i T_{ik} = A_k. \quad (7)$$

Il faut remarquer que  $A_k$  est identique à ce que nous avons déjà désigné par  $A_{ik}$ , où l'indice  $i$  n'est employé que pour indiquer que l'exploitation  $k$  a été choisie dans l'échantillon par l'intermédiaire du segment  $i$ .

Cette identité est vraie pour toute la population, quelle que soit la procédure d'échantillonnage, si les exploitations se trouvent entièrement dans la région et si la géométrie du document (la photographie aérienne) utilisé pour le relevé de terrain est exacte.

On suppose faible la perturbation attribuable aux exploitations ayant des champs dans différentes régions parce que la proportion de ces derniers est faible (en général inférieure à 1 ou 2%) et parce qu'il y a une certaine compensation du biais attribuable au fait qu'il y a des champs situés à l'intérieur de la région qui appartiennent à des exploitations dont le siège se trouve à l'extérieur de la région par le biais attribuable au fait qu'il y a des champs situés à l'extérieur de la région qui appartiennent à des exploitations dont le siège est dans la région. Nous supposons que le total de  $W$  est calculé pour les exploitations dont le siège se trouve dans la région visée par l'enquête.

### 4.4 Non-réponse

Nous parlons ici des estimateurs fondés sur des points correspondant à une exploitation agricole et sur des points ne correspondant pas à une exploitation agricole (section 4.1). Si un exploitant refuse de collaborer ou si l'on ne peut le retracer, on remplace la ou les rangées correspondant à cette personne dans le tableau utilisé en entrée (tableau 1) par les valeurs moyennes pour les exploitations du segment qui ont répondu, s'il y en a; autrement, on les remplace par la moyenne des exploitations qui ont répondu pour tous les segments de la strate traitée.

Si, au deuxième degré (sélection d'exploitations dans le segment), nous considérons les points correspondant à des exploitations et ceux qui ne correspondent pas à des

exploitations et que nous attribuons la valeur 0 aux points qui tombent sur des terres non agricoles, il est évident que l'exclusion des non-répondants introduirait un biais important, parce que les valeurs nulles correspondant à des superficies autres que des SAU ne manquent jamais. Ces points ne sont pas employés pour calculer les valeurs de l'"exploitation agricole moyenne" servant à remplacer les valeurs manquantes. Il y a encore un risque d'introduire un biais si les exploitants non retracés ou refusant de collaborer ont un comportement particulier, par exemple si leurs exploitations sont de façon générale plus petites ou moins efficaces.

Nous aurions pu étudier une autre façon de surmonter ce problème: l'élimination à la fois des valeurs manquantes et d'un nombre proportionnel de valeurs 0 correspondant aux points qui ne se rapportent pas à des SAU. Les deux méthodes donnent la même estimation du total, mais la seconde solution est moins commode parce que, au deuxième degré, la taille de l'échantillon n'est plus un nombre entier.

L'introduction de valeurs pour l'"exploitation agricole moyenne" mènera à un biais négatif pour la variance. Pour compenser, l'exploitation n'est pas incluse dans la taille de l'échantillon  $K_i$  pour le calcul des variances.

## 5. RÉSULTATS: DEUX EXEMPLES

Nous traitons ci-après de certains résultats provenant de deux régions: l'Émilie-Romagne (Italie) et la République tchèque. En République tchèque, ce sont les deux méthodes présentées respectivement dans la sous-section 2.4, 3 et dans la section 4.1 qui ont été employées; aucune donnée ne manquait. En Émilie-Romagne, le plan d'enquête général ne suivait pas exactement la procédure décrite plus haut. Les données manquantes ont été traitées selon la méthode décrite à la section 4.4.

### 5.1 Émilie-Romagne, 1990

En Émilie-Romagne, une superficie de 19,500 km<sup>2</sup> a été divisée en 4 strates, qui excluaient les régions montagneuses. Un échantillon de 313 segments "cadastraux" (avec des limites physiques) a été tiré au moyen d'une procédure à deux degrés avec unités primaires d'échantillonnage (upé) d'environ 10 km<sup>2</sup>. La taille des segments était d'environ 50 ou 100 hectares, selon la strate. On a tiré au hasard cinq points par segment à partir d'une grille ayant un pas de 50 mètres.

Des 1,565 points échantillonnés, il y en avait 326 qui ne correspondaient pas à des SAU, 206 qui correspondaient à une SAU mais pour lesquels on n'a pu trouver l'adresse de l'exploitant, 38 pour lesquels l'exploitant n'a pu être repéré et 32 où l'exploitant a refusé de collaborer. On avait des données valables pour 963 points se rapportant à des SAU et provenant de 285 segments; ces points correspondaient à 617 exploitations, dont certaines figuraient plus d'une fois dans l'échantillon.

Quand nous ne pensons qu'à l'estimation de la superficie, nous pouvons considérer l'enquête par segments comme plus objective et plus complète, puisqu'il n'y a pas de données manquantes et que les observations ne sont pas fondées sur les réponses des exploitants. Si nous acceptons ce principe, nous pouvons avoir une idée d'un biais possible dans le sondage agricole en faisant une comparaison avec les estimations de la superficie obtenues dans l'enquête par segments. Nous pouvons comparer, au tableau 2, les estimations pour les cultures principales de la région. Il y a une bonne correspondance entre les chiffres pour les céréales, à l'exception du blé dur, et pour les cultures permanentes, mais nous constatons pour la betterave à sucre et le soja certains problèmes qui pourraient être liés à une mauvaise compréhension de la façon de déclarer les deuxièmes récoltes au cours de la même année et le même champ ou à un biais attribuable à des valeurs manquantes. Les statistiques officielles sont produites en tenant compte de divers renseignements. Le blé dur est déclaré séparément à cause de la signification spéciale que donne à cette culture l'importante subvention accordée par la CE pour chaque hectare cultivé.

Tableau 2

Résultats de l'enquête par segments et du sondage agricole pour les cultures principales en Émilie-Romagne (1990)

Émilie-Romagne	Enquête par segments		Sondage agricole				ISTAT Superficie
	Superficie × (1,000 ha)		Superficie × (1,000 ha)		Production × (1,000 tm)		
	Esti-mation	CV %	Esti-mation	CV %	Esti-mation	CV %	
Blé tendre	212	5.7	208	6.9	1,177	8	212
Blé dur	46	14.9	48	15.2	260	14	72
Orge	43	11.2	50	17.7	184	17	38
Riz	-	-	4	59.0	23	61	6
Betterave à sucre	111	7.1*	96	9.6	5,474	28	119
Soja	76	6.0*	55	11.6	321	39	47
Vignes	78	13.3*	76	18.7			75
Vergers	91	13.1*	96	19.7			85

\* Estimation corrigée au moyen d'une régression portant sur une image-satellite classée.

ISTAT: Statistiques officielles. Aucun renseignement n'a été fourni sur la précision.

Dans le sondage agricole, les coefficients de variation ont un comportement logique pour les céréales, mais ils deviennent plus difficiles à comprendre dans le cas de la betterave à sucre et du soja. Le CV (coefficient de variation) élevé pour la production peut être attribuable à des rendements plus élevés dans des exploitations agricoles plus grosses et plus spécialisées.

L'estimation de la production peut être corrigée à l'aide de la différence entre les estimations de superficie obtenues dans le cadre de l'enquête par segments et du sondage agricole. Une méthode qui fait appel à un estimateur par régression pourrait être une bonne solution.

Le nombre d'animaux d'élevage est très sous-estimé (tableau 3) puisque beaucoup de propriétaires d'animaux d'élevage ne possèdent pas de terres agricoles. Nous avons utilisé une méthode mixte pour les bestiaux et les porcs: recensement à l'aide d'une liste, pour les cinquante plus grosses exploitations, échantillonnage par points pour les autres exploitations. La procédure est utilisable pour les porcs, mais les CV ne sont pas encore satisfaisants.

Tableau 3

Résultats du sondage agricole sur la base aréolaire et sur la base combinée pour les animaux d'élevage en Émilie-Romagne (1990)

En milliers d'unités	Recen-sement	Base aréolaire		Base combinée	
		Estimation	CV %	Estimation	CV %
Bestiaux	869	829	14	894	13
Porcs	1,876	1,312	37	1,818	27
Moutons	90	38	74		

## 5.2 République tchèque 1992

Les bases aréolaires semblent particulièrement utiles dans les anciens pays communistes d'Europe à cause du changement rapide dans la structure de la propriété foncière. Les statistiques agricoles sont le plus souvent produites sans erreur d'échantillonnage en additionnant les données déclarées par chaque exploitation ou coopérative d'État. Cette procédure ne pourra plus s'appliquer au cours des prochaines années. Il sera extrêmement difficile d'avoir une idée du nombre d'exploitations existantes, et un recensement de l'agriculture sera dépassé avant que les données puissent être produites. Il se peut que les bases aréolaires soient la meilleure solution.

Le territoire de la République tchèque (environ 80,000 km<sup>2</sup>) a été distribué en six strates par interprétation photographique d'images obtenues par le capteur TM des satellites Landsat. Il a fallu quinze jours à une personne pour effectuer cette stratification. En 1992, on a effectué une enquête avec un échantillon de 417 segments carrés de 400 ha en répétant une structure fixe de blocs de 40 km sur 40 km. On a visité les segments et obtenu l'estimation de la superficie selon la procédure expliquée à la sous-section 2.4.1.

Les exploitations ont été échantillonnées à l'aide d'une grille fixe de cinq points dans chaque segment. La grille de cinq points avait la forme d'un 'x', comme dans la figure 6. Cette procédure a permis d'obtenir 2,085 points: 858 correspondaient à des régions non agricoles tandis que les 1,227 autres provenaient de 458 exploitations. Aucune donnée manquante n'a été enregistrée: toutes les exploitations ont été repérées et aucun exploitant n'a refusé de collaborer. Cela s'est produit surtout du fait que l'ancienne structure des grosses exploitations était encore presque intacte.



Dans le tableau 4, on compare les résultats de l'enquête par segments (observations directes sur le terrain), ceux du sondage agricole (exploitations échantillonnées par points) et les statistiques officielles pour les principales cultures du pays. On obtient les statistiques officielles en additionnant les chiffres déclarés par toutes les exploitations ou coopératives d'État. Il y a une assez faible différence dans les estimations de superficie pour le blé, le maïs et les pommes de terre. Il ne faut pas exclure des réponses fournies par les exploitants agricoles un biais qui s'explique par l'autoconsommation de produits agricoles.

**Tableau 4**

Résultats de l'enquête par segments et du sondage agricole en République tchèque (1992)

En milliers d'hectares	Enquête par segments		Sondage agricole				BST	
	Superficie	CV (%)	Superficie	CV (%)	Prod.	CV (%)	Superficie	Prod.
Blé	824	5.4	757	3.7	3,412	4.9	780	3,413
Orge	655	5.1	630	3.8	2,521	4.3	640	2,512
Colza	140	11.6	137	6.8	310	7.5	136	296
Betterave à sucre	119	11.5	127	8.1	4,172	11.0	125	3,874
Maïs	361	7.5	326	4.8	8,884	4.3	361	8,904
Pommes de terre	109	13.6	92	7.9	1,706	8.7	111	1,969

BST: Bureau statistique tchèque.

Les coefficients de variation (CV) des estimations de la superficie sont plus faibles pour le sondage agricole que pour l'enquête par segments. Cela n'est pas surprenant puisque le sondage agricole donne des renseignements à propos des champs qui se trouvent à l'extérieur des segments. Les 458 exploitations sélectionnées représentent plus de 15% de la SAU du pays. Les CV sont un peu plus élevés pour les estimations de la production que pour les estimations de la superficie (plus faibles dans le cas du maïs). Cela semble montrer que la variabilité des rendements contribue moins que la variabilité des superficies à la variabilité de la production.

## 6. CONCLUSIONS ET RECOMMANDATIONS

Les bases aréolaires fondées sur des grilles carrées sont une solution de remplacement pragmatique des bases aréolaires fondées sur des éléments au sol délimités par des caractéristiques physiques. Elles sont beaucoup moins chères à constituer et ne semblent pas comporter d'inconvénients importants quant aux résultats définitifs. Toutefois, un certain travail théorique reste à faire pour qu'on puisse déterminer dans quelles conditions les erreurs d'emplacement attribuables aux limites autres que physiques ont un effet négligeable sur les estimations.

L'échantillonnage de points dans des segments aréolaires est une méthode réaliste qui permet de constituer des bases de sondage pour l'échantillonnage d'exploitations

agricoles. Cette méthode est extrêmement utile dans les cas où la mise à jour des listes (recensement) est mal effectuée ou si de telles listes n'existent pas. L'échantillonnage de quelques points par segment peut être beaucoup moins coûteux qu'une enquête dans toutes les exploitations possédant des champs dans le segment. Il semble que cinq points par segment soit un choix raisonnable.

L'utilisation de bases aréolaires seules donne de mauvais résultats pour les animaux d'élevage lorsqu'il n'existe pas de forte corrélation entre le nombre d'unités et la superficie agricole utilisée dans l'exploitation.

## REMERCIEMENTS

Nous remercions les diverses administrations nationales et régionales qui nous ont apporté leur collaboration pour ce travail. Nous remercions également A. Burrill et O. O'Hanlon, qui ont bien voulu réviser cet article. Les nombreuses observations faites par les arbitres ont beaucoup contribué à rendre cet article plus utile pour les lecteurs.

## BIBLIOGRAPHIE

- ALLEN, J.D. (1990). A look at the remote sensing applications program of the national agricultural statistics service. *Journal of Official Statistics*, 6, 4, 393-409.
- AMBROSIO, L., ALONSO, R., et VILLA, A. (1993). Estimación de superficies cultivadas por muestreo de áreas y teledetección. Precisión relativa. *Estadística Española*, 35, 91-103.
- CARFAGNA, E., RAGNI, P., ROSSI, L., et TERPESSI, C. (1991). Area frame: un Nuovo Istrumento per la Realizzazione delle Statistiche Agricole in Italia. *Contributi alla Statistica Spaziale*. University of Parma.
- CARFAGNA, E., et DELINCÉ, J. (1992). Farm survey based on area frame sampling. The case of Emilia Romagna in 1990. Conference on the Application of Remote Sensing to Agricultural Statistics, Belgirate. Office for Publication of the E.C. Luxembourg.
- CARFAGNA, E., et GALLEGGO, F.J. (1994). Extrapolating intra-cluster correlation to optimize the size of segments in an area frame. Conference on Applied Statistics to Agriculture, Kansas State University, Manhattan, KS.
- COCHRAN, W. (1977). *Sampling Techniques*. New York: John Wiley and Sons.
- COTTER, J., et NEALON, J. (1987). Area frame design for agricultural surveys. U.S. Department of Agriculture. National Agricultural Statistics Service.
- DELINCÉ, J. (1990). Un premier bilan de l'action I Inventaires Régionaux du Projet Agriculture après deux années d'activité. Conference on the Application of Remote Sensing to Agricultural Statistics, Varese. Office for Publication of the E.C. Luxembourg.
- DICORATO, F. (1993). AIS estimation programs. User documentation. JRC Ispra.

- VAN DIEPEN, C.A., WOLF, J., VAN KEULEN, H., et RAPPOLDT, C. (1989). WOFOST: A simulation model for crop production. *Soil Use and Management*, 5, 16-24.
- EUROSTAT 1991. Working party, Crop Products Statistics. Methodological reports. Document AGRI/PE/333, Luxembourg.
- FUENTES, M., et GALLEGO, F.J. (1994). Stratification and cluster estimator on an area frame by squared segments with an aligned sample. Conference on Applied Statistics to Agriculture, Kansas State University of Manhattan, KS.
- GALLEGO, F.J. (1992). Flächenschätzungen für einjährige Feldfrüchte mit Hilfe Fernerkundung. *Neue Wege raumbbezogener Statistik. Forum der Bundesstatistik*, 20, 109-120. Wiesbaden: Statistisches Bundesamt.
- GALLEGO, F.J. (1994). Using a confusion matrix for area estimation with remote sensing. *Atti Convegno AIT*, Roma, 99-102.
- GALLEGO, F.J., et DELINCÉ, J. (1994). Area estimation by segment sampling. Dans *Euro-Courses Remote sensing applied to Agricultural Statistics*.
- GALLEGO, F.J., DELINCÉ, J., et RUEDA, C. (1993). Crop area estimates through remote sensing: Stability of the regression correction. *International Journal of Remote Sensing*. 14, 18, 3433-3445.
- GIOVACCHINI, A. (1992). Agricultural statistics by remote sensing in Italy: an ultimate cost analysis. Conference on the Application of Remote Sensing to Agricultural Statistics, Belgirate. Office for Publication of the E.C. Luxembourg.
- GONZÁLEZ, F., LOPEZ, S., et CUEVAS, J.M. (1991). Comparing two methodologies for crop area estimation in Spain using landsat TM images and ground gathered data. *Remote Sensing of the Environment*, 32, 29-36.
- GOWARD, S.N., MARKHAM, B., DYE, D.G., DULANEY, W., et YANG, J. (1991). Normalized difference vegetation index measurements from the advanced very high resolution radiometer. *Remote Sensing of the Environment*, 35, 257-277.
- HAY, A.M. (1988). The derivation of global estimates from a confusion matrix. *International Journal of Remote Sensing*, 9, 1395-1398.
- HOUSTON, A.G., et HALL, F.G. (1984). Use of satellite data in agricultural surveys. *Communications in Statistics Theory and Methods*, 13, 23, 2857-2880.
- VAN LANEN, H.A.J., VAN DIEPEN, C.A., REINDS, G.J., DE KONING, G.H.J., BULENS, J.D., et BREGT, A.K. (1992). Physical land evaluation methods and GIS to explore the crop growth potential and its effects within the European Communities. *Agricultural Systems*, 39, 307-328.
- MEYER-ROUX, J. (1990). Présentation du projet pilote de télédétection appliquée aux statistiques agricoles. Conference on the Application of Remote Sensing to Agricultural Statistics. Office for Publications of the E.C. Luxembourg.
- NEALON, J.P. (1984). Review of the multiple and area frame estimators. U.S. Department of Agriculture, Statistical Reporting Service, Report 80, Washington, D.C.
- SHARMAN, M., et de BOISSEZON, H. (1992). Action IV: de l'image aux statistiques, bilan opérationnel après deux années d'estimations rapides des superficies et des rendements potentiels au niveau Européen. Conference on the Application of Remote Sensing to Agricultural Statistics, Belgirate. Office for Publications of the E.C. Luxembourg.