

revue trimestrielle

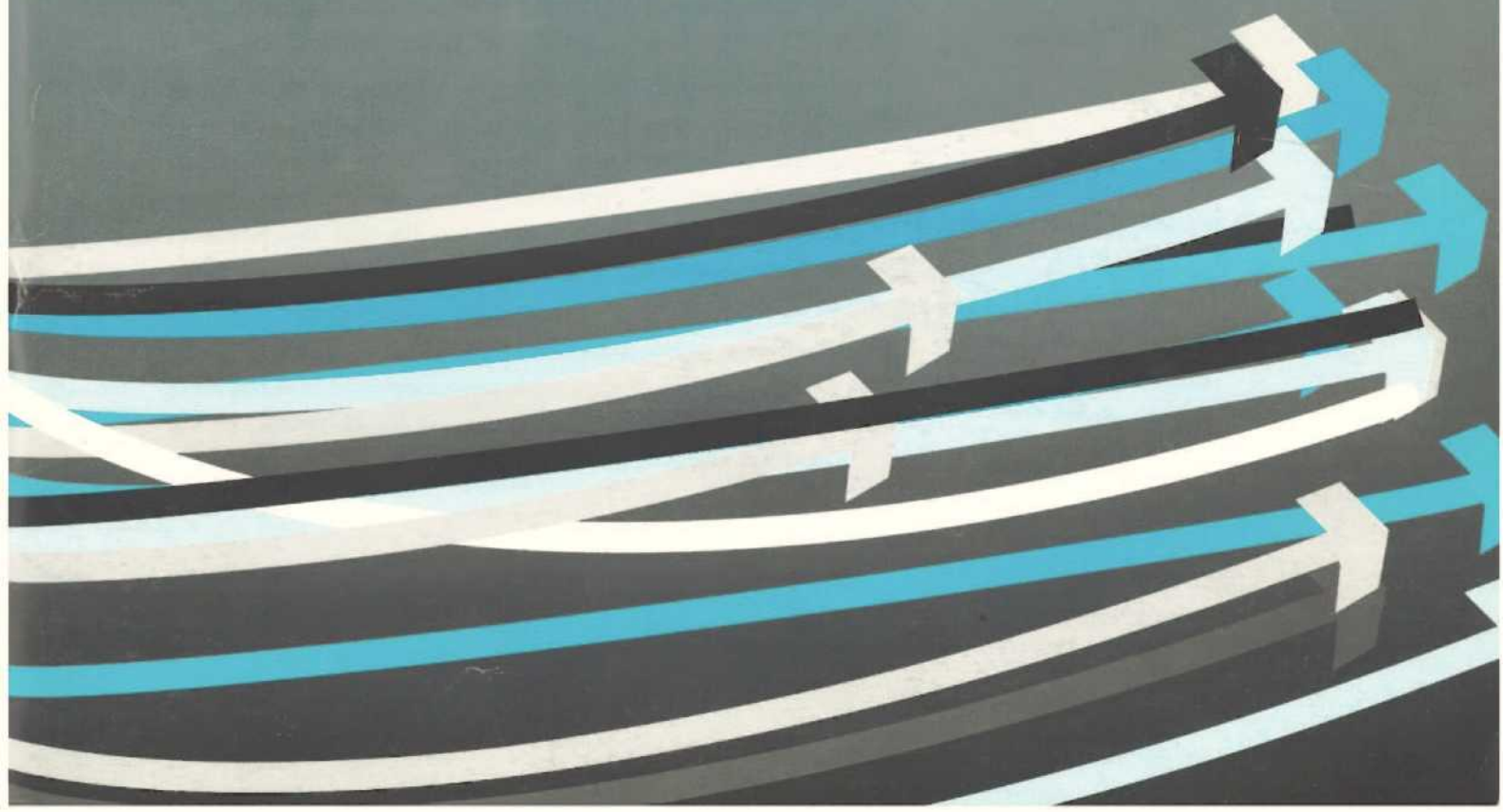
N° 61-62

Juin

1992

Courrier des statistiques

informations sur le système statistique public



Courrier des statistiques

Réalisé à l'INSEE par la direction de la coordination statistique et des relations internationales, le Courrier des statistiques est une publication **interministérielle** donnant des informations sur l'ensemble des activités du système statistique public - INSEE, services statistiques de ministères, autres organismes de production ou d'étude.

"**Grâce à cette revue**, précisait E. Malinvaud dans l'éditorial du premier numéro, **chaque statisticien devrait, même s'il a une fonction très spécialisée, être averti des travaux de ses collègues et des problèmes qu'ils rencontrent**". Information générale donc sur les différentes activités du système statistique, mais aussi information plus pointue sur l'évolution des méthodes et des moyens de l'activité statistique.

Le Courrier des statistiques souhaite rendre compte chaque trimestre des **travaux** réalisés par les statisticiens, et plus précisément de la manière dont ces travaux ont été menés, depuis les concertations qui en ont fixé les objectifs jusqu'à la publication des résultats, en passant par la mise au point des méthodes de recueil puis de traitement de l'information.

La revue présente également les **services** statistiques eux-mêmes, grandes unités de l'INSEE, services statistiques des ministères, autres organismes publics : structure, situation dans le ministère, missions, dispositif de publications...

Opérations et services sont décrits par les statisticiens qui en ont la charge, mais des **points de vue extérieurs** sont également souhaités, avec l'expérience des statisticiens étrangers, les débats du Conseil national de l'information statistique (CNIS), les comptes rendus des colloques et séminaires, l'avis des utilisateurs de l'information statistique.

Des **tribunes** permettent plus explicitement encore d'ouvrir un dialogue sur les diverses formes prises par l'investigation statistique dans le domaine économique et social, et sur l'utilisation qui peut en être faite : études, recherches, critiques...

ÉDITORIAL

de Jean-Pierre BEHMOIRAS

PAGES D'HISTOIRE

Alain DESROSIÈRES	5	La création, en 1885, du Conseil supérieur de la statistique
Gilles EVRARD	7	Au temps de la "République des Jules"
Journal Officiel	9	du 22 février 1885

DOSSIER : LE SYSTÈME STATISTIQUE PUBLIC FRANÇAIS

15

OUTILS ET MÉTHODES

Pascal ARDILLY	35	Le système d'échantillonnage des enquêtes-ménages de l'INSEE en 1992
Robert PASTORELLI	47	Superficies agricoles à partir d'images satellite

ORGANISATION ET ÉVOLUTION DES SERVICES

Jean-Marie NIVLET	53	La MEPSS, mission d'étude des problèmes du secteur des services
-------------------	----	--

Suite au verso...

Superficies agricoles à partir d'images satellite

L'investissement précoce du SCEES (Service Central des Enquêtes et Etudes Statistiques du Ministère de l'agriculture) en télédétection spatiale, appliquée aux statistiques agricoles, lui a permis de disposer dès 1988 de capacités de traitement opérationnelles.

La méthode statistique utilisée, l'estimation par la régression, améliore les estimations de superficies obtenues par enquête de terrain. Les données satellite fournissent une information complémentaire qui augmente la précision des estimations faites à partir d'un échantillon de relevés au sol.

Actuellement, l'utilisation de l'imagerie satellitaire a un coût comparable à celui qui serait dû à une augmentation de la taille de l'échantillon de l'enquête, pour un même gain de précision.

C'est en 1860 que le photographe Nadar eut l'idée d'exercer son art à partir d'une montgolfière. Depuis cet acte fondateur, les techniques d'observation à distance de la Terre, regroupées sous le vocable de télédétection, ont bien évolué.

Une nouvelle vision du monde

Ce type d'observation s'est développé avec l'aviation. Aujourd'hui une nouvelle étape est franchie avec les satellites artificiels de la Terre. Ceux-ci, avec leur "recul" de plusieurs centaines, voire plusieurs milliers de kilomètres et leur passage répétitif, offrent une vision globale et dynamique de notre planète. La télédétection spatiale est née.

Quelle est la nature de l'information ainsi recueillie ?

L'analogie avec la vision humaine n'est pas fondée. L'oeil de ces satellites, comme SPOT 1 et 2 lancés par la France en 1986 et 1990, est un radiomètre. Cet appareil "voit" le sol comme une juxtaposition de carrés élémentaires - les pixels - de 20 mètres de côté pour SPOT. Pour chacun de ces "points", il mesure le rayonnement lumineux dans certaines plages du spectre électromagnétique : jaune, rouge et proche

infrarouge pour SPOT. L'"image" satellite est en fait une énorme collection de chiffres référencés spatialement. Seul le recours aux ordinateurs et aux écrans graphiques permet de restituer un objet visuel. De même, l'information contenue dans ces images ne peut être pleinement exploitée qu'avec les méthodes de l'analyse numérique.

Quelles applications pour cette innovation ?

La télédétection spatiale fait partie de ces avancées technologiques qui précèdent leurs utilisations opérationnelles. La précision locale et la sensibilité radiométrique de ces observations laissent espérer, à tout le moins, une discrimination aisée des grands postes d'occupation du sol, en particulier dans le domaine agricole. C'est sur cette application que le SCEES travaille depuis 1980.

La Cellule télédétection du SCEES

Le SCEES n'est pas un organisme de recherche, si ce n'est dans les applications de la statistique à ses objectifs particuliers. Son Bureau "méthodes et applications scientifiques" assure cette mission ainsi que celle de l'analyse et de la réponse aux besoins informatiques du ser-

vice. C'est dans ce bureau que le Chef du Service a créé, en 1980, une cellule télédétection. Celle-ci s'est vu confier la mission d'étudier les conditions d'utilisation de l'information satellitaire dans le dispositif de la statistique agricole. Cette cellule est composée actuellement d'un ingénieur d'agronomie et d'un attaché de l'INSEE.

La cellule télédétection, après une large investigation des possibilités de l'information satellitaire, a centré ses efforts sur une seule application : déterminer les superficies des grands types de couverts du territoire à l'aide d'une enquête de terrain et d'images satellite.

Les méthodes statistiques sous-jacentes et les traitements à effectuer étaient connus. Ils avaient été évalués lors d'études préalables. La faisabilité était attestée par l'expérience américaine, menée par le Ministère de l'agriculture des USA - USDA - qui a utilisé cette technique dès la fin des années 70, pour effectuer ses propres inventaires culturaux, avec les données des satellites LANDSAT. Pour ce faire il a développé un logiciel, PEDITOR, dédié à cette application. Un accord de coopération en télédétection existe entre le SCEES et l'USDA depuis 1981. En mai 1985, un programme de développement de cette application a été décidé au SCEES.

L'objectif visé était l'intégration de cette application de la télédétection dans les opérations courantes du Service. Cet objectif passait par la réalisation préalable d'une chaîne de traitement informatique opérationnelle.

L'apport attendu était une amélioration de la précision des estimations de superficies obtenues par les enquêtes classiques.

La méthode retenue a été celle de l'estimation de superficies par une technique statistiquement fondée, "l'estimateur par la régression". Cette méthode utilise l'information contenue dans les images satellite pour accroître la précision des estimations de superficies effectuées à partir du seul échantillon de l'enquête. La méthode employée suit une démarche statistique rigoureuse, qui permet de calculer la précision des résultats obtenus.

Les enquêtes de terrain

Les conditions de mise en oeuvre de la méthode d'utilisation des données satellite nécessitent une enquête de terrain portant sur des parcelles agricoles. Or ce n'est pas le cas de l'enquête nationale du SCEES sur l'utilisation du territoire, l'enquête "TER-UTI", qui observe des points. Cette enquête est réalisée par les Services départementaux, qui relèvent le type d'occupation du sol rencontré en des points précis (environ 5000 par département), définis sur des photographies aériennes.

Cette difficulté technique à utiliser l'enquête d'occupation des sols existante au SCEES va peser sur le schéma d'intégration de l'application télédétection. Alors que l'objectif recherché était celui d'une insertion dans le dispositif classique du SCEES, une nouvelle enquête, l'enquête Segment, portant sur des parcelles et seule capable d'initialiser les traitements des images satellite a été mise en place en 1985. Le segment est une portion carrée du territoire, de 50 hectares, définie sur une photographie aérienne.

L'estimateur par la régression

C'est une méthode connue des statistiques par sondage. Elle consiste à améliorer la précision de l'estimation d'une variable, obtenue par sondage, grâce à l'information fournie par une donnée auxiliaire liée à la variable étudiée. Dans le cas présenté ici la variable étudiée est la superficie des cultures et la donnée auxiliaire est l'image satellite.

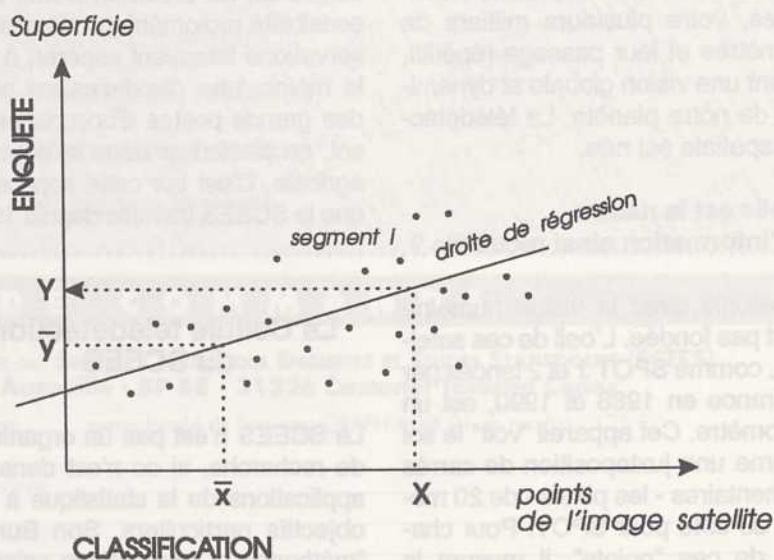
Les superficies sont observées sur l'échantillon de portions carrées de 50 hectares de territoire que sont les "segments" de l'enquête Segment (une centaine par département). Cette "réalité de terrain", issue d'un échantillon de surfaces, permet une estimation, entachée d'une erreur aléatoire, de la superficie totale existante dans la zone d'étude : le département.

D'autre part l'image satellite délivre une information exhaustive sur la zone, mais avec une erreur d'observation qui provient du processus d'affectation des points des images aux catégories d'occupation du sol : la classification des images.

A l'issue de ces deux opérations : enquête terrain et classification des données satellite correspondant aux segments, on dispose, pour une culture donnée, de deux valeurs dans chaque segment : la superficie observée (y) et le nombre de points classés dans cette même culture (x). Ces deux valeurs (x,y) déterminent un point pour chaque segment et la droite de régression qui passe au plus près de l'ensemble de ces points peut être calculée. Cette droite représente une simplification de la relation qui lie la réalité de terrain et la classification des images satellite.

Ensuite la classification de l'ensemble des points des images satellite recouvrant toute la zone permet d'obtenir le nombre total de points dans une culture ; ce dernier, divisé par le nombre théorique de segments dans la zone donnera X, le nombre moyen de points d'une culture par segment. X est en général différent de la moyenne des x mesurés sur les seuls segments. L'application des coefficients de la droite de régression à X lui fait correspondre Y, estimateur par la régression de la superficie moyenne de cette culture par segment. Il reste à multiplier Y par le nombre de segments théoriques sur la zone pour avoir l'estimation totale.

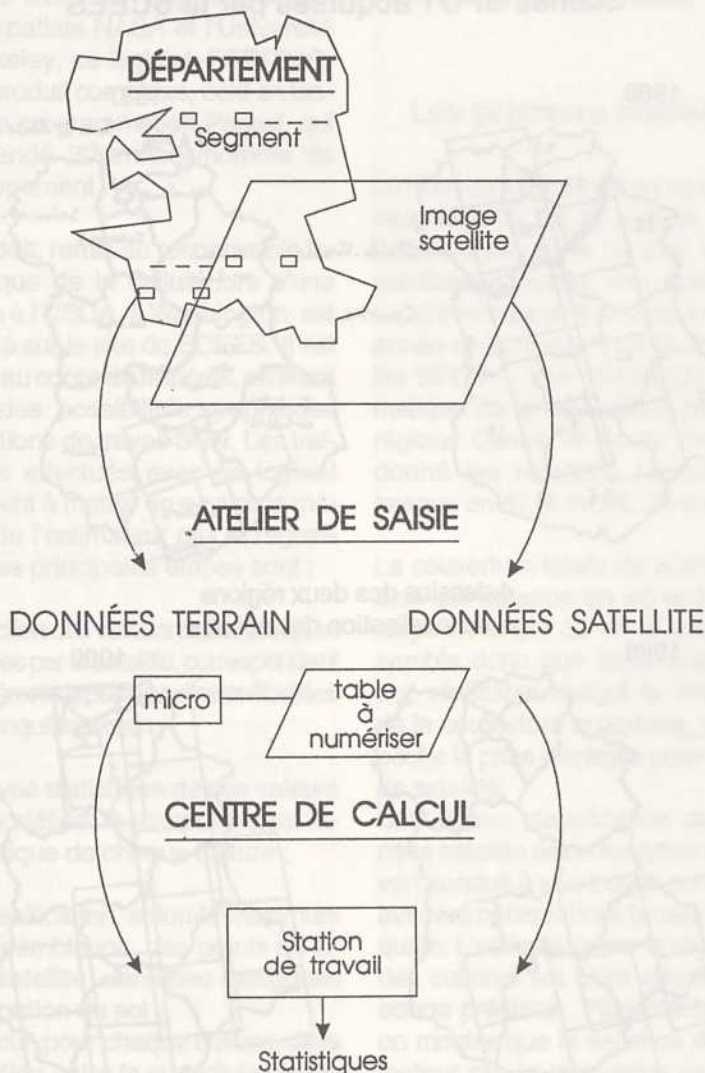
Cette estimation diffère de celle obtenue avec le seul échantillon de l'enquête. Son intérêt principal provient de sa plus grande précision au sens statistique.





L'enquêteur dessine une parcelle, sur un transparent posé sur une photographie aérienne - Photo : SCMA

De l'enquête Segment et des photos satellite aux estimations de superficie



L'échantillon comporte une centaine de segments par département. Pour chaque segment, l'enquêteur dessine le schéma du parcellaire observé et note les types d'occupation du sol. L'orientation ainsi prise, enquête Segment et télédétection, s'est développée en parallèle de l'action courante du Service pour les inventaires de cultures, l'enquête TER-UTI.

La chaîne des traitements

L'objectif préalable de mise en place d'un outil informatique de traitement était une nécessité impérieuse pour le passage de cette application en mode courant dans le Service. L'expérience américaine a servi de référence et l'effort central que supposait ce programme a porté sur l'adaptation au contexte français des méthodes et moyens sophistiqués utilisés aux USA. Ce transfert a été envisagé "tel quel", malgré la différence entre les tailles du parcellaire des deux pays, et malgré l'utilisation des données du satellite SPOT à la place de celles de LANDSAT. La cellule télédétection a réalisé, en 3 ans, une chaîne informatique, comparable à celle de l'USDA et comprenant deux grandes fonctions :

- la saisie-numérisation des questionnaires de l'enquête terrain,
- l'analyse des images satellite et le calcul des résultats statistiques.

La saisie-numérisation des questionnaires a été développée par le SCEES pour tenir compte des spécificités de son enquête. Elle permet de calculer les surfaces des parcelles, et de localiser les parcelles dans l'image satellite.

Ces objectifs sont atteints en deux étapes, à l'aide du logiciel DIGITER :

- La saisie-numérisation proprement dite. Elle consiste, à l'aide d'un ordinateur et d'un périphérique de saisie, à transformer l'information analogique représentée par un dessin de contour en une suite de coordon-

nées numériques. Le calcul des superficies est alors une simple opération mathématique.

- La localisation : elle est obtenue par une fonction mathématique de transfert déterminée à partir de quelques points remarquables repérés à la fois sur les photos aériennes, les cartes géographiques et les images satellite.

La configuration matérielle comprend une table à numériser, format A0, pilotée par un micro-ordinateur. La programmation du logiciel, en langage C et système graphique "GKS", est confiée à une société de services informatiques.

Les données saisies de l'enquête et les images satellite achetées auprès de la société SPOT IMAGE sont ensuite traitées pour produire les statistiques finales d'occupation du sol. La configuration matérielle de cette phase des traitements est fonction des caractéristiques des données et des traitements :

- volumes de données : une image SPOT correspond, au minimum, à un fichier de 27 millions d'octets !

- puissance de calcul : les traitements effectués sur une image demandent une trentaine d'heures à un mini-ordinateur !

La cellule s'est équipée du matériel suivant : 3 stations SUN-Microsystems, sous système UNIX, interconnectées en réseau local Ethernet, un dérouleur de bandes magnétiques, un disque de 380 millions d'octets et une imprimante laser.

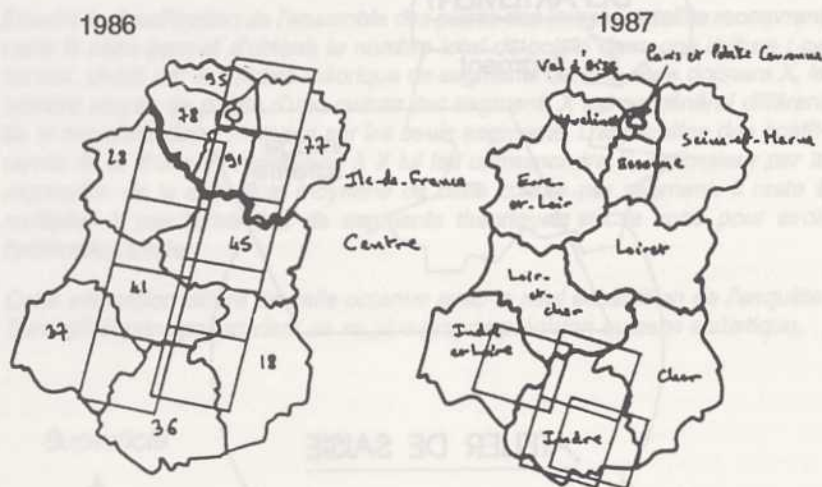
Ces stations de travail possèdent une interface utilisateur graphique, à base de "fenêtres", d'"icônes" et de "souris" sur écran de 19 pouces. Le réseau local permet de partager le disque entre les trois stations, ce qui offre un espace de travail suffisant à chaque utilisateur et permet d'avoir une seule copie du logiciel. Pour pouvoir récupérer les données d'enquête, un micro-ordinateur type PC est relié aux stations.

Le logiciel utilisé est celui du Service

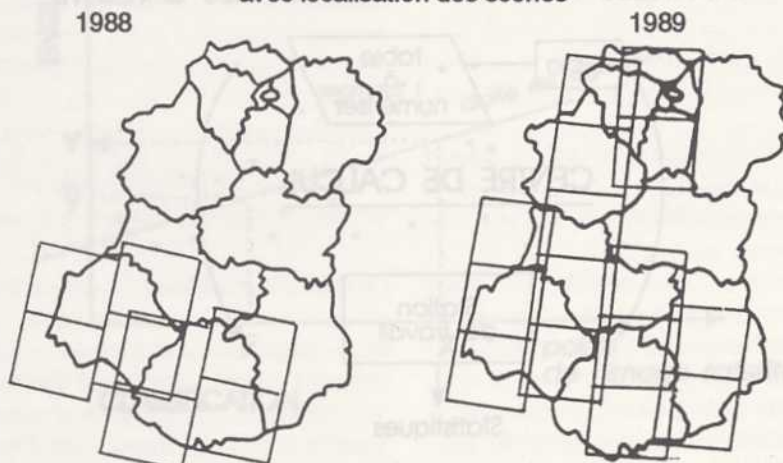


L'atelier de saisie : table à numériser, carte IGN, photo aérienne et agrandissement, transparent, curseur et écran graphique - Photo : SCEES

Scènes SPOT acquises par le SCEES



4 dessins des deux régions avec localisation des scènes



Couverture totale de la zone en 1990 et 1991

Efficacités relatives de la télédétection

Occupation du sol	Ile de France	Centre
Blé	4,5	2,6
Orge	1,4	1,1
Blé + Orge	5,1	2,6
Maïs	1,8	1,6
Toutes Céréales	4,6	2,5
Colza	5,8	3,0
Tournesol	3,0	1,8
Oléagineux	4,0	2,4
Betteraves	2,8	1,6
Légumes secs	3,5	6,0

statistique du Ministère de l'agriculture des USA. Développé avec l'agence spatiale NASA et l'Université de Berkeley, ce logiciel -PEDITOR- est un produit complexe, écrit en langage de programmation Pascal, qui a demandé 20 années/homme de développement.

Ce produit, remis au responsable informatique de la cellule lors d'une mission à l'USDA, à Washington, est implanté sur le site du SCEES. Il est adapté au contexte français, en tirant partie des possibilités graphiques des stations de travail SUN. Les traitements effectués avec ce logiciel consistent à mettre en oeuvre la méthode de l'estimateur par la régression. Les principales étapes sont :

- l'extraction des valeurs radiométriques recueillies par le satellite, correspondant aux différents types de cultures étudiées dans l'enquête terrain ;
- l'analyse statistique de ces valeurs pour modéliser le comportement radiométrique de chaque culture ;
- la classification, selon le maximum de vraisemblance, des points de l'image satellite, dans des catégories d'occupation du sol ;
- le calcul, pour chaque culture, de la corrélation entre la superficie observée par enquête et celle obtenue par la classification de l'image ;

- la détermination de l'estimation de superficie et de sa précision.

Les premiers résultats

Le bilan de 5 années d'achat de données auprès de la société SPOT IMAGE s'est avéré de plus en plus satisfaisant. Après une acquisition expérimentale de 9 images en 1986, année de la mise en orbite du satellite SPOT 1, une commande systématique de la couverture des deux régions Centre et Ile de France a donné les résultats suivants : 4 images en 87, 8 en 88, 15 en 89.

La couverture totale de la zone d'étude est obtenue en 90 et 91 avec respectivement 23 et 17 images. Il semble donc que cette acquisition soit réalisable malgré la fréquence de la couverture nuageuse, qui empêche la prise d'images pour ce type de satellite.

Une bonne classification des données satellite selon les types de couvert conduit à une bonne corrélation avec les observations terrain de l'enquête. L'estimateur par la régression des cultures est alors affecté d'une bonne précision. Plus précisément, on montre que la variance de l'estimateur par la régression est liée à celle de l'estimation des seules données d'enquête par la formule :

$$\text{Variance-régression} = (1 - R^2) \times \text{Variance-enquête}$$

Cette formule donne une valeur de la variance par régression d'autant plus faible que le coefficient de corrélation linéaire R (mesuré entre les données d'enquête et les données satellite classées) est proche de 1, c'est à dire que la classification des images satellite correspond à la réalité.

Le rapport des variances est un indicateur de l'efficacité relative de l'introduction des images satellite :

$$\text{Efficacité Relative : E.R.} = \frac{1}{1 - R^2}$$

Si E.R. = 1, l'introduction des images satellite ne sert à rien. Si E.R. vaut 3, par exemple, cela signifie que le résultat obtenu avec l'apport de l'information satellitaire est 3 fois plus précis que celui de la seule enquête terrain. Autrement dit, il faudrait multiplier par 3 le nombre de segments enquêtés pour avoir la même précision avec l'enquête seule.

Le SCEES obtient ses premiers résultats en 1988 sur le département de l'Indre, en utilisant deux images SPOT. En 1990, pour la première fois à cette échelle, des résultats sont obtenus sur deux régions françaises, soit 51 600 km², en utilisant 23 images SPOT prises en mai-juin. C'est la première application opérationnelle de cette importance ; elle permet de tirer des conclusions sur des bases solides. Les résultats ne sont pas présentés ici en termes de superficies par culture, mais en termes d'efficacité relative, de façon à mettre en évidence l'apport de l'imagerie satellitaire.

Les résultats sont meilleurs en Ile de France que dans le Centre. Cette différence peut être attribuée à des différences dans le parcellaire qui est plus petit dans le sud de la région Centre.

Sur la base du dispositif retenu (enquête Segment et télédétection) cette application peut-elle passer en phase de production routinière ?

Les conditions de fonctionnement et les délais d'obtention des résultats demandent encore à être améliorés pour une utilisation courante dans un service administratif.

L'enquête terrain par segments, nouvelle, est plus délicate à réaliser par les enquêteurs que l'enquête TER-UTI. Son dépouillement demande plus de temps et ce avec des moyens plus sophistiqués.

Le logiciel PEDITOR requiert une formation de plusieurs semaines pour un utilisateur du niveau ingénieur et demande une participation active à l'analyse. Les traitements nécessaires à la production des résultats d'un département demandent 3 semaines à une personne compétente et motivée.

Les gains de précision obtenus avec la télédétection, au stade actuel de développement des méthodes, sont trop faibles. Ils ne justifient pas le surcoût des images satellite qui est de l'ordre du double du prix de l'enquête.

Cette même dépense permettrait de tripler la taille de l'échantillon de segments.

On obtiendrait ainsi, de par le simple effet du nombre d'éléments enquêtés, un gain de précision d'un facteur 3, facteur encore trop rarement atteint avec la télédétection. Ce constat provient pour partie de la capacité du satellite SPOT à discriminer les cultures. Des progrès sont attendus avec SPOT 4 et avec les satellites utilisant les ondes radar.

L'opération conjointe SCEES-CCR

Le Centre Commun de Recherches

des Communautés Européennes - CCR -, situé à ISPRA en Italie, conduit depuis 1988, à la demande de la Direction générale de l'agriculture de la Commission Européenne, un vaste programme de recherche sur les applications de la télédétection en agriculture. Il est prévu de confier un rôle actif aux Services nationaux dans le passage à l'opérationnel des méthodes développées par le CCR. L'action 1 de ce programme est strictement identique à l'application du SCEES présentée ici.

La collaboration avec le CCR a débuté en 1988 et, en 1990, ce programme de travail est devenu une opération conjointe sur les régions Centre et Ile de France. Dans le cadre de cette collaboration, des efforts ont porté sur l'optimisation des outils de traitements au niveau matériel, logiciel et statistique. Le SCEES a participé à l'installation du logiciel PEDITOR au CCR. Ce dernier a développé, en 1990, un nouveau système de saisie des segments. Ce système utilise une caméra vidéo comme périphérique de saisie, ce qui lui confère une grande simplicité d'emploi et une rapidité 4 fois supérieure à celle de DIGITER. Aussi le SCEES a-t-il décidé, à partir de 1991, de sous-traiter la saisie de ses segments avec ce système.

Enfin des études sont en cours pour améliorer des points de méthode sous-jacents à l'application.

En particulier, avec le CCR de nouveaux essais d'utilisation de l'enquête TER-UTI sont actuellement menés. De telles études avaient été faites dès 1981 au SCEES, mais la résolution spatiale des satellites de cette époque n'avait pas permis d'obtenir des résultats satisfaisants. Aujourd'hui, la meilleure résolution des satellites et les progrès réalisés

dans les techniques de traitement de données géographiques numériques justifient ces nouvelles tentatives. La modification du type d'enquête terrain créerait un dispositif basé sur l'enquête courante du SCEES, permettant l'utilisation ou non de la télédétection. C'est la seule solution susceptible de conduire à une réelle intégration de la télédétection pour la mesure des superficies agricoles et sous le schéma méthodologique présenté ici.

Robert PASTORELLI
Cellule télédétection SCEES
Ministère de l'agriculture

Bibliographie

GDTA-SCEES (1984) - Simulations SPOT Lauragais, résultats 1981, vol 1 et 2.

FOURNIER P. (1986) - Enquête sur l'utilisation du territoire effectuée par la méthode des segments, SCEES-MAS, Série S 13, 1986.

PASTORELLI R. (1989) - PEDITOR, un logiciel de télédétection au SCEES, in "Cahiers de statistique agricole", 3/6 mai-juin 1989.

PASTORELLI R., PORCHIER J-C. (1989) - Premiers résultats sur l'utilisation des terres à l'échelle départementale à partir de données du satellite SPOT, in "Cahiers de statistique agricole", 3/6 mai-juin 1989.

CCR-SCEES-SYSAME (1991) - Projet agriculture - Action 1 : inventaires régionaux - 1990, régions Centre et Ile de France, rapport final campagne 1990.