

A PROPOS DES CERRADOS DU BRÉSIL : DE L'INFORMATIQUE DE RECHERCHE AUX SYSTEMES D'INFORMATION.

(Philippe WANIEZ et Gérard DANDOY)

RESUME - Parmi les systèmes d'information sur l'environnement, les Systèmes d'Information Géographique (SIG) occupent une place particulière. Outils très lourds, à multiples usages, ils répondent généralement plus directement aux besoins des gestionnaires du territoire qu'à ceux des chercheurs.

Une de leurs caractéristiques majeure consiste à permettre l'intégration et le traitement de données spatialisées d'origine, de forme et de nature différentes. La disponibilité d'un tel outil d'information, pour répondre à une problématique de recherche, en géographie par exemple, est donc très souhaitable.

Or, partant d'un cas concret, celui de l'étude de la dynamique spatiale des productions agricoles des cerrados du Brésil, il a fallu, faute d'avoir à disposition un véritable SIG, mettre au point une méthode facilitant la mise en relation de l'information issue de cartes thématiques sur l'environnement physique, avec des statistiques municipales de production agricole.

Cette démarche, qui n'est qu'un palliatif à l'absence de SIG, conduit à l'obtention rapide de réponses à certains aspects d'un problématique ne se limitant pas aux seules relations entretenues par le milieu naturel et la production agricole.

Suivre une telle voie, où l'optique de recherche est privilégiée (et non pas la gestion territoriale), demeure compatible avec la future mise en place d'un SIG, tout en apportant une solution partielle qui pourrait éventuellement trouver d'autres applications.

Autrement dit, la rareté des SIG fonctionnels et accessibles ne conduit-elle pas le chercheur à se forger des outils moins lourds, tout en adoptant des techniques standardisées ? Deux questions se posent alors : comment assurer la compatibilité entre bases de données thématiques en vue de leur intégration dans un système plus vaste ? Plus généralement, est-il du rôle de la recherche de prendre l'initiative de la constitution de SIG, compte tenu de l'importance des moyens qu'implique ce genre d'opération ?

1. GESTION DES DONNEES DE RECHERCHE ET SYSTEMES D'INFORMATIONS

Les organisateurs du Séminaire informatique de l'ORSTOM, SEMINFOR, ont choisi pour l'édition 1989, la troisième, d'aborder la question des systèmes d'information sur l'environnement. Pour contribuer à la discussion collective les auteurs du présent texte proposent quelques réflexions à propos de la place des systèmes d'information dans leur pratique de recherche en géographie régionale. Au delà des slogans et des projets pharaoniques ou gadgets informatiques, SEMINFOR 3 semble une excellente occasion de préciser ce qu'on doit entendre par systèmes d'information sur l'environnement, ce qu'on peut en attendre dans le cadre des programmes de recherche visant la production de connaissances nouvelles, quels sont les avantages de l'approche par les systèmes d'information sur des méthodes plus légères.

Afin d'éviter tout malentendu, disons d'emblée que la notion de système d'information nous apparaît assez différente de la gestion des données qu'un chercheur doit opérer, tant sur le plan des objectifs à atteindre que sur celui de la méthodologie à suivre.

Dans le cadre de travaux de recherche empirique, on est toujours amené à accumuler et donc à stocker une masse d'information plus ou moins importante. Cette pratique, évidemment inhérente au travail de recherche mais ne se confondant pas avec lui, était autrefois basée sur la rédaction de fiches cartonnées et sur la constitution de volumineux dossiers rassemblant un matériau hétérogène. On tend à leur substituer aujourd'hui des enregistrements informatiques qui imposent l'adoption d'autres manières de concevoir et de réaliser la collecte de l'information.

L'informatisation plus ou moins poussée du corpus constitué au cours d'une recherche, au-delà des gains habituellement procurés par l'ordinateur (volume réduit, rapidité d'accès à un document, etc.), a, selon nous, une utilité bien plus importante car elle permet de mettre en oeuvre une très grande gamme de techniques d'analyse propres à autoriser une meilleure validation des hypothèses de recherche. Cela reste, bien entendu, le principal souci d'un chercheur, quels que soient sa discipline et le cadre institutionnel dans lequel il exerce.

Notre pratique de l'informatique nous fait penser qu'il est souhaitable que l'essentiel des opérations d'organisation et d'analyse de l'information soient maîtrisées directement par le chercheur. On peut s'étonner qu'il faille encore énoncer ce genre de principe qui va de soi dans de nombreux domaines, mais qui, dans les sciences humaines reste assez peu fréquent. Mais la situation évolue - trop lentement à notre goût - grâce, en particulier, à la diffusion de progiciels conviviaux, assez faciles à utiliser et qui réalisent la majeure partie des fonctions souhaitées.

Ainsi, l'utilisation de l'informatique par un chercheur qui souhaite être en mesure de pratiquer lui-même l'organisation et l'analyse de ses données, est bien autre chose que la réalisation de systèmes d'information. Ces derniers nous paraissent différents sur plusieurs plans.

En général, les systèmes d'information ne visent pas un objectif unique et répondent plus à des préoccupations d'aménagement ou de gestion qu'à des objectifs de recherche. Ils s'adressent à des groupes plus ou moins diversifiés d'utilisateurs, dont les besoins et les qualifications peuvent être très différents. Se pose donc ici la question des besoins des usagers. En effet, la notion de système d'information, telle qu'on peut l'envisager dans le cadre du secteur tertiaire comme dans les banques ou les assurances, est sans doute difficile à transposer tel que dans le monde de la recherche. Autrement dit, la demande d'un chercheur n'est pas de même nature que celle des services techniques d'une grande métropole, ou bien encore d'une société pétrolière, même si ces organismes détiennent parfois des données utiles au chercheur. En effet, il est rare que les données détenues par un seul organisme soient suffisantes pour les besoins d'une problématique de recherche qui nécessite en général la combinaison d'informations en provenance de sources multiples. De plus, celles-ci s'appuient sur des nomenclatures hétérogènes, et donc partiellement incompatibles (par exemple les nomenclatures de métiers ou les csp) et se rapportent à des individus statistiques aux contours variables (établissements, entreprises, sociétés, etc.). En outre, ces données se présentent sous des formes variées : relevés d'observation ponctuels, cartes géographiques, documents écrits, archives, photographies, etc.. De cette complexité, qui semble atteindre son paroxysme dans les sciences qui, comme la géographie, recourent à des informations d'origine et de nature très variées, découlent deux conséquences importantes.

D'une part, notre courte expérience en la matière nous a appris que la mise sur pied de tels systèmes imposait une ségmentation plus ou moins nette des diverses tâches de collecte, de codage, d'enregistrement et de mise à disposition des données. La réalisation d'un système d'information répondant à des objectifs multiples et régulièrement mis à jour ne peut être que le fait d'une équipe disposant d'importants moyens, tant en matériel qu'en personnel. Derrière cette remarque transparaît la notion de "service de données". On peut penser que seule une modeste proportion des programmes de recherche de l'ORSTOM est en mesure de répondre à de telles exigences. Pour notre part, nous avons pu observer au Brésil la grande difficulté qu'il y a pour mettre en place et maintenir une structure, somme toute assez lourde, lorsque les cadres institutionnels sont déficients.

D'autre part, et contrairement à ce que nous affirmions plus haut, les informaticiens, en tant que détenteurs des méthodes et techniques de *systématisation* ont un rôle central à jouer dans l'élaboration de véritables systèmes d'information, entendus comme services de données d'intérêt général. Reste que cette position stratégique peut conduire à des difficultés de communication avec les différentes composantes ayant à voir avec un système donné. Lorsque l'information obtenue après de nombreuses années de terrain devient accessible à une collectivité à l'aide de moyens aussi puissants que les systèmes d'interrogation modernes, il est normal que des réticences se fassent jour pour la cession des données. Une solution peut être entrevue dans la constitution de co-

mités scientifiques composés des chercheurs concernés. Ceci signifie, qu'en aucun cas, la réalisation d'un système d'information ne devrait être conduite par un informaticien. "L'informaticien apporte un outil, aux possibilités certes immenses, mais qui reste un outil et non une machine à faire des miracles" (Dangermond (J.), 1982).

Enfin, une ambiguïté subsiste sous le vocable système d'information. Pour certains, le système se limite au logiciel qui assure les fonctions nécessaires à la mise à disposition des données. Pour d'autres, le système est constitué par l'ensemble de données lui-même. Il nous semble qu'un système digne de ce nom comprend au moins ces deux composantes.

Nous allons illustrer ce propos par l'exposé d'une expérience de simulation d'utilisation d'un système d'information géographique ("*en projet*") qu'on aurait souhaité pouvoir utiliser dans le cadre du développement d'une problématique de recherche. On tente de répondre à la question : *Quel est le degré de correspondance entre la dynamique de la production agricole des cerrados (région de savanes arborées du centre du Brésil), et les principaux contrastes géographiques du milieu naturel ?* Nous suivons ici de près l'article publié par la revue Mappemonde cité en référence (1988). Pour une présentation générale des cerrados, le lecteur pourra se reporter à l'article de P. Waniez dans les Cahiers des Sciences Humaines également cité en référence (1988).

2. MILIEU NATUREL ET PRODUCTION AGRICOLE

En mesurant, au sens plein du terme, le "*poids du milieu*", on cherche à identifier, localiser et quantifier le rôle du contexte physique sur l'émergence des structures spatiales de la production agricole. On prétend dépasser ce que L. Cambrézy nomme la "*simple mais frustrante analyse descriptive*" (du milieu naturel) en prenant en compte le temps, mais aussi en mesurant les paramètres écologiques pour les mettre en relation avec les données économiques et sociales, telles qu'elles apparaissent dans les statistiques censitaires et les enquêtes annuelles. Certes, en conférant au nombre cette place centrale, on risque de confondre l'image de la réalité construite par l'analyse statistique avec cette réalité elle-même. Cependant, comment aborder autrement cette immensité de deux millions de kilomètres carrés où vivent, depuis peu seulement, vingt millions d'habitants ? Pour contraignant qu'il soit, ce passage obligé permet de dépasser le stade des évaluations plus ou moins intuitives sur les inter-relations complexes entre les composantes écologiques et anthropiques du système géographique.

Engager une recherche sur de telles bases impose la conception d'une méthodologie particulière : il faut mettre en relation les données statistiques de production agricole avec l'ensemble d'autres informations, comme les cartes, où sont enregistrés la majeure partie des traits de l'environnement physique. Il s'agit d'une préoccupation commune à tous ceux qui s'intéressent aux Systèmes d'Information Géographique (SIG). Bien que différente des techniques utilisées dans les SIG, la méthode adoptée ici s'inscrit dans une perspective proche et

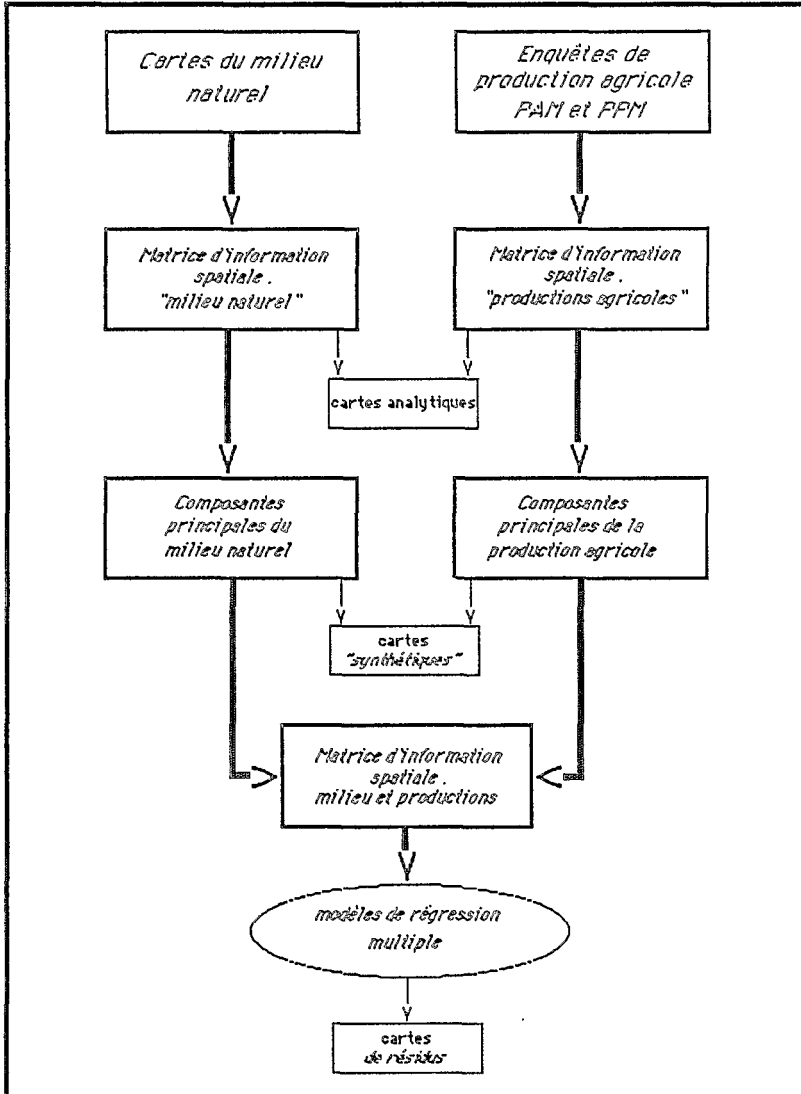
permet de déboucher rapidement sur des résultats tangibles. En effet, ne disposant ni du temps, ni des moyens nécessaires à la réalisation d'un SIG sur les cerrados, nous avons conçu une méthode, fondée sur un carroyage constituant le point de rencontre du découpage administratif de la statistique avec les multiples aires d'extension des catégories climatiques, phyto-écologiques ou pédologiques. Ceci permet d'apprécier les relations entre les composantes du milieu naturel et celles de la production agricole.

La méthode adoptée privilégie donc la mesure (figure 1). Les stocks d'information se rapportent d'une part au milieu naturel, sous forme de cartes thématiques, et de l'autre à la production agricole, sous forme d'enquêtes annuelles de 1977 à 1984. A partir de ces gisements d'information, on cherche à élaborer des matrices d'information spatiale afin d'obtenir une description plus ou moins synthétique du milieu naturel et des productions agricoles. A chaque étape, diverses cartes sont réalisées afin d'apprécier la répartition géographique des types de milieu et des types de production obtenus par analyse en composantes principales. A ce stade, on peut enfin procéder à la mise en relation des grands contrastes spatiaux exprimés par les composantes principales, c'est-à-dire à chercher à expliquer la variation observée sur chaque composante de la production par celle des différentes composantes du milieu. Il s'agit alors d'un problème classique de modélisation par régression multiple. L'étape ultérieure consiste à examiner les cartes de résidus propres à chaque régression, afin de détecter des sous-espaces ne se conformant pas au modèle. Autrement dit, l'étude des résidus permet de localiser les discordances entre les types de production et les types de milieux.

Les principaux thèmes relatifs au milieu naturel sont extraits de cartes thématiques au 1:5 000 000, publiées par L'UNESCO, la FAO et l'EMBRAPA. On procède à la mesure de chaque thème sur un carroyage à double niveau : un niveau de comptage sur plus de 9 500 carreaux et un niveau statistique composé de 97 grands carreaux. Les cartes analytiques des thèmes permettent d'apprécier, par exemple, la géographie des formations végétales (planche 1). Si le type "cerrados" existe dans la quasi totalité de la région du même nom, son aire d'extension principale se limite au Nord du Goiás et au Sud du Maranhão.

La synthèse issue de l'analyse en composantes principales de l'ensemble des variables se rapportant au climat, à la végétation, aux sols, et à la topographie (18 variables en tout) montre l'existence de trois grands domaines biogéographiques et de deux nuances, l'une climatique, l'autre topographique. Au total, ces quatre composantes représentent 60% de l'inertie totale (planche 2). Le contraste le plus marqué apparaît entre le Sud, où la végétation naturelle se fait rare (domaine des "campagnes méridionales") et le Nord, véritable domaine des "plateaux de cerrados" (carte 2-A). Sur la partie Est, le domaine "semi-aride" se rattache au monde nordestin (carte 2-B). Le climat hypertropical, ainsi que les fortes pentes (cartes 2-C et 2-D), viennent influencer localement les caractéristiques des deux premiers domaines.

Figure n°1
Méthode d'étude des relations
Milieu naturel / Production agricole



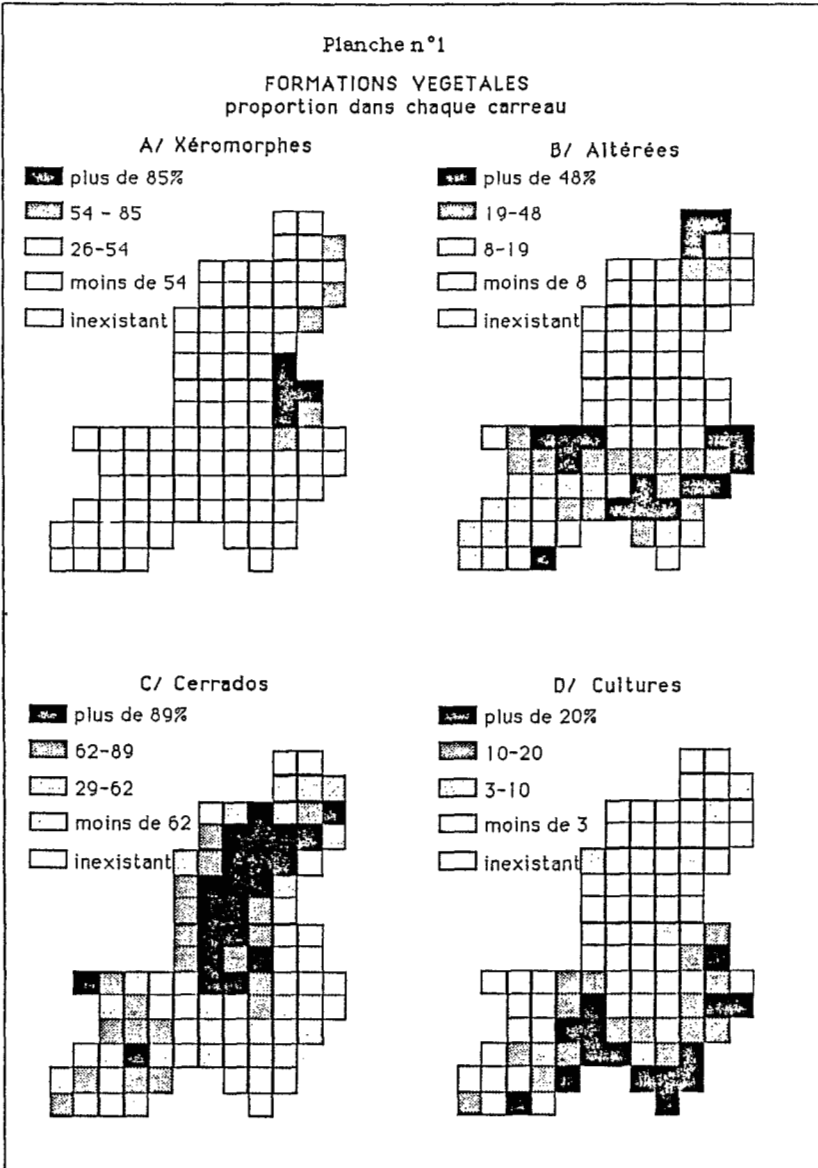







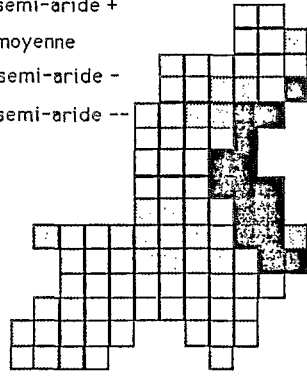
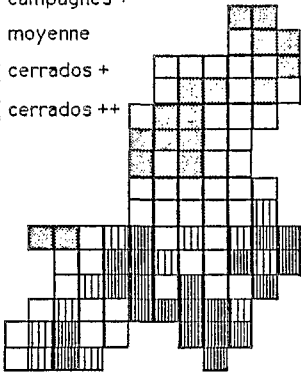
Planche n°2
 DOMAINES BIOGEOGRAPHIQUES
 homogénéité relative

A/ Plateaux de cerrados /
 Campagnes méridionales

B/ Semi-aride






-  campagnes ++
-  campagnes +
-  moyenne
-  cerrados +
-  cerrados ++






-  semi-aride ++
-  semi-aride +
-  moyenne
-  semi-aride -
-  semi-aride --

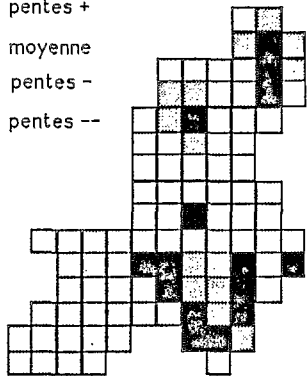
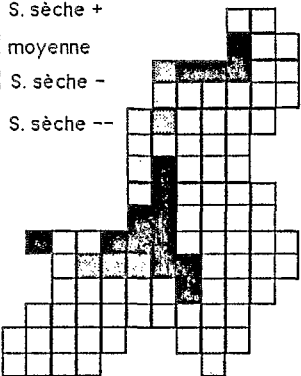


C/ Influence du climat
 hypertropical

D/ Influence des
 fortes pentes

-  Saison sèche ++
-  S. sèche +
-  moyenne
-  S. sèche -
-  S. sèche --

-  pentes ++
-  pentes +
-  moyenne
-  pentes -
-  pentes --



Parallèlement, la production agricole a fait l'objet d'une analyse détaillée, sur les 843 *municipios* de la région, pour lesquels on dispose des résultats des enquêtes de production. Par exemple, la première composante de production (carte 3-A) traduit l'opposition entre d'une part "*cultures commerciales des grands espaces*" (dominées par l'association soja et bovins) et d'autre part, les productions vivrières (principalement haricot et maïs). Pour mettre en relation les composantes du milieu naturel avec celles de la production agricole il est nécessaire d'accorder ces dernières au carroyage qui devient de ce fait le "*dénominateur géographique commun*". La planche 3 présente une application de cette technique à la première composante de la production. Après avoir placé le carroyage (carte 3-B) sur la carte originale (carte 3-A), chaque carreau se voit affecter la valeur médiane des *municipios* lui appartenant (carte 3-C). Cette division par dix du nombre d'observations (de 843 *municipios* à 97 grands carreaux), a sans doute pour prix une perte d'information, mais sans qu'on observe toutefois une modification des contrastes spatiaux. Cela prouve, a posteriori, la force et la grande stabilité des configurations géographiques dans les cerrados, condition de l'application de cette méthode. Est-elle transposable à des ensembles géographiques plus réduits, moins fortement structurés et abordés à d'autres échelles ? Nous n'en sommes pas certains...






On réitère la même opération pour les deux autres composantes de la production agricole (productions temporaires/productions permanentes et productions du Nordeste).

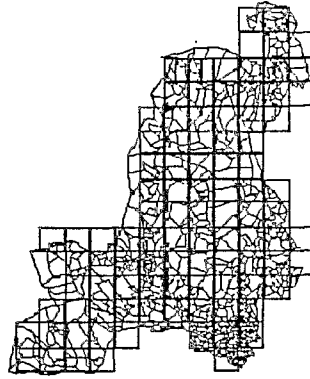
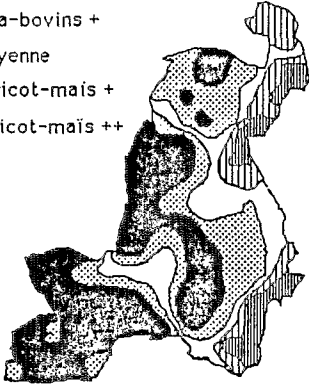
Ces deux ensembles de données (les quatre composantes biogéographiques et les trois composantes de la production agricole) se rapportant à la même maille, le carroyage, sont mises en relation par des modèles de régression multiple. Par exemple, deux composantes du milieu naturel sont corrélées de manière significative avec l'opposition "*productions commerciales des grands espaces/productions vivrières*". La figure 2, montre que ni le domaine semi-aride, ni les fortes pentes ne favorisent l'implantation des productions commerciales des grands espaces. A l'inverse, les productions vivrières se localisent de manière préférentielle dans ces types de milieux, contraignants sur le plan agricole. Cette observation met sur la voie d'une interprétation en terme de dualité des systèmes de production, mettant en valeur des espaces différents. Dans ce cas, le poids du milieu physique est incontestable, mais n'apparaît jamais comme déterminant ; d'ailleurs, le taux de variance expliquée n'atteint dans le cas présent que 35% de la variance totale. Il est clair que ce résultat, même complété par la cartographie et l'étude des résidus, ne constitue en aucun cas une fin en soi. Il ne s'agit que d'un des éléments d'interprétation d'un ensemble d'information beaucoup plus vaste. Mais cette démarche permet de mieux évaluer le degré de correspondance entre la dynamique de la production agricole des cerrados et les principaux contrastes géographiques du milieu naturel.

Planche n°3
 CULTURES ET ELEVAGE EN 1983-1984
 Facteur n°1






A/ Lissage sur 843 municipios
 (tracé approximatif)

B/ Carroyage sur
 fond municipal

-  soja-bovins ++
-  soja-bovins +
-  moyenne
-  Haricot-maïs +
-  haricot-maïs ++



C/ Carte sur carroyage

-  soja-bovins ++
-  soja-bovins +
-  moyenne
-  Haricot-maïs +
-  haricot-maïs ++

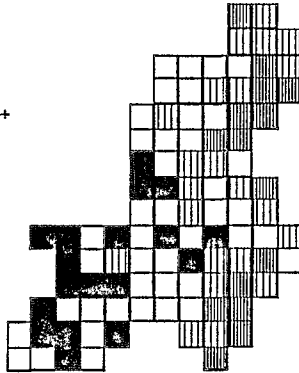
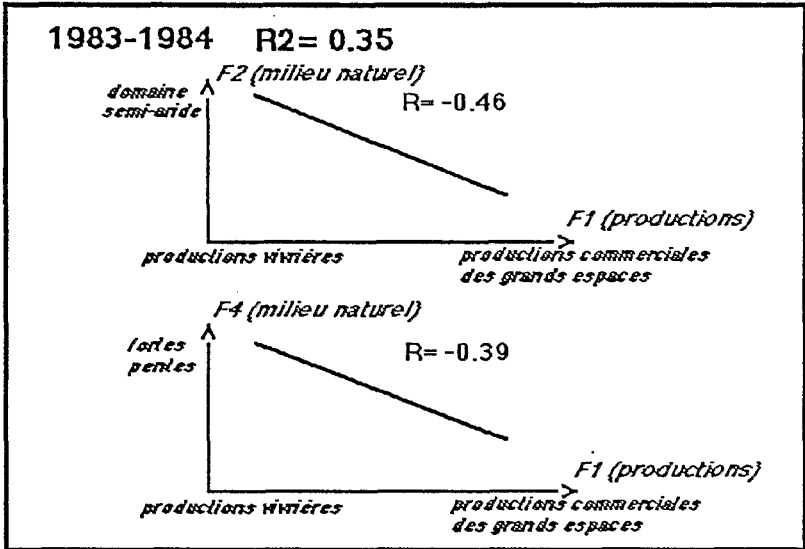


Figure n°2

**COMPOSANTES DU MILIEU NATUREL
ET PRODUCTIONS AGRICOLES**

productions commerciales des grands espaces/productions vivrières



3. CE QU'AUURAIT PU APPORTER UN SYSTEME D'INFORMATION SUR L'ENVIRONNEMENT

Bien qu'ayant choisi de ne pas adopter la voie des systèmes d'information, nous regrettons de n'avoir pu disposer de tels outils. En effet, ceux-ci présentent de nombreux avantages qu'on peut répartir en deux groupes.

D'une part, les systèmes d'information simplifient l'accès aux sources. Ils rendent disponible une grande quantité d'information de diverses origines, sur un thème large comme, par exemple, l'agriculture du Brésil. Un tel outil n'est réalisable qu'avec la participation directe des producteurs de données rassemblés dans un cadre institutionnel proportionné à la dimension du sujet traité.

D'autre part, si l'on avait disposé d'un SIG, on aurait pu, à partir des cartes conservées en ordinateur, dériver plusieurs des tableaux des données, et non pas un seul comme le traitement manuel nous a permis de la faire de manière artisanale. Ainsi, il aurait été aisé de tester la validité de plusieurs nomenclatures différentes et de choisir la mieux adaptée à la problématique. Par exemple, les caissons de la carte des sols ont été regroupés afin de rendre compte de leur valeur agronomique face à une exploitation moderne et mécanisée. Ce parti pris est parfois discutable pour certaines parties des cerrados. On aurait préféré pouvoir formuler et tester des hypothèses alternatives.

Enfin, nous avons choisi d'adopter un carroyage dont le degré de finesse dépend directement du maillage municipal. Un découpage plus fin serait souhaitable pour améliorer l'appréciation de la variété des situations et, ainsi, accroître le niveau de validité des résultats.

Ainsi, l'existence d'un SIG sur les cerrados, voire sur l'ensemble du Brésil, serait sans aucun doute de nature à améliorer la validation des résultats obtenus. En introduisant une plus grande souplesse et rapidité d'exploitation des documents originaux, on pourrait procéder à divers regroupements de variables ou d'unités spatiales, afin d'atteindre une meilleure connaissance du milieu agricole.

4. A CHAQUE RECHERCHE SON SYSTEME ?

Nous regrettons la non disponibilité d'un système d'information sur l'environnement de la production agricole du Brésil, et plus précisément encore, des Cerrados. Il constituerait, en effet, un facteur de développement important permettant de mieux apprécier et maîtriser une activité d'importance stratégique pour ce pays et cette région. Bien que l'EMBRAPA ait, depuis de nombreuses années, cette préoccupation en promouvant le système SISGEO, force est de constater que la recherche, très (trop) rapidement présentée ici et réalisée avec des moyens bien inférieurs à ceux dont dispose SISGEO, reste impossible à mener avec SISGEO.

Face à cette absence, les chercheurs géographes que nous sommes doivent-ils réaliser leur propre système d'information ? Autrement dit, faut-il subordonner la réponse à d'importantes questions que posent les rapides mutations de

l'activité agricole des cerrados, à la mise sur pied d'un système d'information, par vocation à portée plus générale, ce qui, en toute hypothèse, implique encore plusieurs années de mise au point et de travail ? Dans de telles conditions, nous avons choisi une autre voie conduisant, dans des délais plus limités, à des résultats dont le niveau de validité pourrait être amélioré, mais qui n'en constituent pas moins un apport au débat scientifique sur la question.

Néanmoins, à notre niveau, nous avons pris soin d'adopter des outils logiciels qui ne nous sont pas spécifiques et qui garantissent, de ce fait, une réelle compatibilité avec un futur système d'information. En adoptant le logiciel SAS, au demeurant le principal outil d'analyse statistique des chercheurs de l'EMBRAPA, nous avons volontairement refusé l'isolement, prévoyant notre participation à un futur SIG qui pourra être réalisé un jour, si les conditions nécessaires sont réunies. C'est dans cet esprit qu'a été implanté le module SISECSO (Système d'Informations Economiques et Sociales) sur le matériel du centre informatique de l'EMBRAPA, à Brasília.

Sans nier le grand intérêt des SIG, il ne nous semble pas que toute recherche géographique, même lorsqu'elle fait appel largement à la quantification, et donc à l'informatique, doivent nécessairement se traduire par le montage d'un système d'informations géographiques. L'importance de l'infrastructure et des moyens nécessaires à la mise en place de tels systèmes doivent nous inciter à la plus grande circonspection, et cela d'autant plus qu'on peut souvent imaginer des palliatifs, certes moins "performants", mais correspondants à la dimension des moyens dont le géographe peut disposer en général.

REFERENCES BIBLIOGRAPHIQUES

- DANDOY (G.) - 1986 *Quelques remarques sur les applications de la cartographie assistée par ordinateur en France*. ORSTOM, Colloques et séminaires "Traitement des données localisées", pp. 89-92.
- DANGERMOND (J.) - 1982 *Classification des éléments des logiciels utilisés habituellement dans les systèmes d'informations géographiques*. CFC, Bulletin n° 96, pp. 7-22.
- WANIEZ (P.) - 1988 *Base de données pour la recherche géographique : la dynamique spatiale agricole des cerrados du Brésil*. ORSTOM, Cahiers des Sciences Humaines, 24 (2), pp. 251-270.
- WANIEZ (P.), DANDOY (G.) - 1988 *Milieu naturel et productions agricoles : présentation d'une méthode pour les cerrados du Brésil*. RECLUS, Mappemonde, 1988-4, pp. 14-17.