

SISECSO OU L'INFORMATIQUE AU SERVICE D'UNE RECHERCHE EN GÉOGRAPHIE HUMAINE : LA DYNAMIQUE SPATIALE DES CERRADOS AU BRÉSIL

Philippe WANIEZ

(Géographie)

(UR 502, Cadres spatiaux de l'Indépendance Alimentaire)

avec la collaboration de Jozeneida PIMENTA de AGUIAR
technicienne à l'EMBRAPA/CPAC, Planaltina (D.F.), Brésil
et de Violette CABOS
cartographe, ingénieur au CNRS, Montpellier, France

RESUME - Un système d'information géographique : pour quoi faire ?

SISECSO est le module "*socio-économique*" d'un ensemble plus vaste nommé SISGEO, réalisé conjointement par l'ORSTOM et l'Empresa Brasileira de Pesquisa Agropecuaria (EMBRAPA) et visant la mise au point d'un système d'information sur l'environnement. Il s'agit de "*capitaliser*" toutes les informations existantes relatives aux conditions de développement de l'agriculture brésilienne, conditions physiques, mais aussi humaines.

Initialement, le projet SISECSO apparaissait plus comme une réponse à une demande de type technocratique qu'à un programme de recherche centré sur une problématique. Rien de bon ni d'intéressant n'aurait pu sortir d'une telle approche faisant l'impasse sur 1/ *la définition précise d'un problème scientifique* et l'énoncé de quelques hypothèses y afférent, 2/ *élaboration d'un corpus statistique* en parfaite adéquation avec les concepts qu'il est censé représenter, 3/ *l'immédiate mobilisation* de la totalité de l'arsenal *des techniques d'analyse*.

Ce diagnostic a eu pour conséquences le choix d'autres finalités et d'autres moyens que ceux prévus initialement.

Problématique et méthodologie : la dynamique régionale des cerrados

Les *cerrados* sont composés de la partie périphérique de plusieurs régions brésiliennes très diverses couvrant une grande variété de niveaux de développement. Le géographe est directement conduit à observer et comprendre le processus de formation de ce qui n'est pas une vraie région géographique mais qui pourrait le devenir.

La base de données réalisée dans le cadre de SISECSO constitue à la fois le moyen et un produit de la recherche. C'est le moyen car il est préférable d'asseoir l'essentiel du travail de recherche sur la petite échelle. C'est aussi un produit de la recherche car elle renferme l'essentiel des informations nécessaires à la compréhension du processus d'émergence régionale.

Cette approche doit être mise en perspective avec ce que certains ont pu appeler la "*révolution quantitative*" affectant la géographie française depuis une quinzaine d'années.

Sources statistiques, aire d'étude et unités d'observations

L'Institut Brésilien de Géographie et de Statistique (IBGE) est l'organe central de gestion et de restitution des données générées par le fonctionnement de l'appareil administratif. L'IBGE assure également la réalisation périodique de recensements.

Ainsi, l'IBGE conserve les fichiers et les diffuse auprès des entreprises qui en font la demande. Dans le cadre de SISECSO, toutes les données censitaires de 1970 à 1980 ont été acquises : agriculture, industrie, services et démographie. Elles ont été complétées par les données relatives aux productions de l'agriculture et de l'élevage.

La définition précise de l'aire d'étude s'est appuyée sur les travaux antérieurs du Centre de Recherches Agro-pastorales des cerrados (CPAC). C'est un ensemble de 885 *municípios*, unité politico-administrative de base, formant une maille extrêmement évolutive, caractéristique dont découlent quelques problèmes délicats pour l'organisation des données.

Adopter un modèle d'informatisation.

Préalablement au choix d'un modèle de mise en oeuvre de l'informatique, il a fallu procéder à une analyse des diverses possibilités. L'option "*progiciel clés en mains*" a été choisie en raison de la brièveté des délais nécessaires à sa mise en oeuvre.

Le Statistical Analysis System (SAS) utilisé depuis longtemps à l'EMBRAPA et correspondant bien à la nature des données à organiser constitue l'outil informatique de réalisation de SISECSO, implanté sur un gros système IBM (2 ordinateurs IBM 4341) et accessible en batch sous OS/VS2 et en conversationnel sous VM/CMS.

La place de SAS au cours des principales étapes de réalisation de SISECSO

En raison de la grande variété de fonctions proposées par SAS, toutes les étapes de réalisation de SISECSO y ont fait appel, à des degrés divers : 1/ *la constitution du tableau d'identification des municípios*, 2/ *la lecture des bandes de l'IBGE*, 3/ *la constitution des tableaux finals*.

Seule la numérisation de la carte des municipios des cerrados a nécessité la conception et la programmation d'un petit ensemble de programmes générant un fichier "*fond de carte*" compatible avec SAS/GRAPH.

Composition et avenir de SISECSO

Au printemps 1987, SISECSO se compose de 3 bases de données informatisées, implantées sur le centre informatique de l'EMBRAPA et accessibles via le progiciel SAS. A la fin de l'année 1987, le CPAC disposera d'une banque de données de près de 3 000 variables relevées sur 856 municipios. Mais ces données relatives à la décennie 1970-1980 devront être complétées au fur et à mesure de la sortie des nouveaux recensements.

1. UN SYSTEME D'INFORMATION GÉOGRAPHIQUE : POUR QUOI FAIRE ?

SISECSO est un programme de coopération initié il y a plusieurs années, visant "*la mise au point d'un système d'information sur l'environnement*". La finalité de ce projet apparaît parfois comme étant la "*réalisation d'un système d'informations géographiques à vocation agricole*". La prétention de SISGEO est donc de couvrir à la fois l'ensemble des questions relatives à l'environnement de l'activité agricole, ainsi que la totalité du territoire brésilien. La division du travail est conçue de la manière suivante : d'une part, six modules thématiques traitent du climat, de la végétation, des sols, de l'hydrologie, des données phyto-sanitaires et de l'information socio-économique ; d'autre part, deux modules méthodologiques sont voués l'un à la cartographie automatique, l'autre à la télédétection. Avec sept départements, services ou centres de recherche de l'EMBRAPA, 56 chercheurs, ingénieurs ou techniciens brésiliens et quatre de l'ORSTOM, SISGEO s'inscrit naturellement dans le cadre de l'Unité de Recherche 502 de l'ORSTOM (Cadres spatiaux de l'indépendance alimentaire).

Le module relatif aux données économiques et sociales, SISECSO, a été entrepris bien après ceux ayant trait aux sols et au climat ; c'est au Centre de Recherche Agro-Pastorale des Cerrados (CPAC) qu'est revenu sa conception. En effet, la forte expansion de l'agriculture des cerrados en fait un excellent objet de recherche, cristallisant l'essentiel des problèmes scientifiques que peut présenter un tel projet. Malheureusement, une fois à pied d'oeuvre, on s'est rapidement rendu compte que dans cette institution, la "*socio-économie*" n'était qu'un tiroir de plus, un nouveau facteur à optimiser pour accroître les rendements, améliorer les conditions de la commercialisation, mieux définir la demande des consommateurs

potentiels. On avait donc affaire à une demande purement technocratique pouvant se résumer de la manière suivante : accumulons la plus grande masse d'information dans tous les domaines ayant un rapport avec l'activité agricole ; interrogeons la banque de données ainsi constituée et produisons des "*caractérisations*" plus ou moins fouillées en fonction de la "*demande sociale*" (en fait la demande de l'état planificateur) ; puis, effectuons des "*croisements synthétiques*" montrant sur quels éléments il faudrait agir pour atteindre un optimum donné. Dans cette perspective, l'informatique n'intervient que pour rendre plus efficiente, plus systématique l'oeuvre de planification.

Sur un tel programme, plusieurs points de vue peuvent être adoptés. Celui de l'archiviste : plus il y a de données, meilleur c'est ; cela servira peut-être un jour et tant mieux si l'informatique permet de s'y retrouver plus facilement. Celui du technocrate : avec une telle masse de données, je vais pouvoir truffer mes rapports de tableaux de chiffres, judicieusement choisis pour venir à l'appui de mes conceptions. Celui du technicien agricole : enfin, je vais pouvoir définir avec précision où je peux planter mes carottes. Celui du responsable scientifique : l'avenir c'est l'informatique ! d'ailleurs, voyez la recherche s'informatise à toute allure. Celui du chercheur agronome ou sociologue classique : tout cela ne remplacera jamais ma solide expérience de terrain ; ce qu'il faut, c'est "*sentir les liaisons*" et les illustrer par quelques données bien choisies. Curieusement, le géographe concevant sa discipline comme l'analyse des structures de l'espace sera peut-être le plus déçu par ce genre de réalisation. En effet, pour conduire à quelques découvertes intéressantes, la méthodologie quantitative requiert au moins trois conditions préalables : 1/ la définition précise d'un problème scientifique et l'énoncé de quelques hypothèses y afférent, 2/ la constitution d'un corpus statistique en parfaite adéquation avec les concepts qu'il est censé représenter, 3/ l'immédiate mobilisation de la totalité de l'arsenal des techniques d'analyse. Dans le cas du programme SISECSO, aucun de ces trois points n'était pleinement représenté ; le premier était totalement ignoré, le second ne faisait l'objet que de discussions de détail sur la véracité de telle ou telle valeur relevée dans un annuaire statistique, et le troisième était réduit à sa plus simple expression : statistique descriptive et tabulations.

En appliquant à SISECSO les propos du sociologue Wright Mills, on peut se demander si ses promoteurs n'étaient pas victimes de "*l'inhibition méthodologique*", c'est-à-dire d'une confusion entre l'objet et la méthode. Le fait que le projet insiste plus sur l'acquisition et l'organisation informatique des données que sur l'usage qui doit en être fait dans le cas d'une recherche précise montre bien où se situe le problème. Si l'on ajoute à cela la position

dominante des informaticiens et la multiplication des modules spécialisés segmentant le poids des divers interlocuteurs, on aboutit à une situation où le chercheur est littéralement dépossédé de son objet de recherche qui devient la matière première d'une équipe de techniciens naturellement plus enclins à mettre l'accent sur la méthode que sur l'objet. L'épuisement de ce processus se traduit par l'existence d'un produit informatique peu utile à la recherche, mais que peuvent s'accaparer les technostructures pour leurs fins propres. C'est pourquoi il ne faut pas s'étonner des encouragements réitérés exprimés par les structures bureaucratiques de la recherche, de l'EMBRAPA notamment. Si l'on admet sans grande discussion le bien fondé de telles réalisations à des fins de gestion du territoire et, dans une moindre mesure, de planification, on doit fermement contester leur utilité pour la recherche en géographie humaine. Il y a là une confusion des genres : la recherche, base du développement, devrait être autre chose qu'une activité de service informatique pour le seul profit de l'Etat planificateur. Ce diagnostic a eu pour conséquence directe le choix d'autres finalités et d'autres moyens pour la réalisation de SISECSO.

2. PROBLÉMATIQUE ET MÉTHODOLOGIE : DYNAMIQUE RÉGIONALE DES CERRADOS

Etudier, décrire, même de manière très fouillée, les transformations des cerrados, ne constitue en aucun cas une problématique de recherche en géographie humaine. Après avoir pris connaissance de la maigre littérature sur le sujet, et demandé le point de vue de plusieurs chercheurs et aménageurs (français...), il a été possible de définir un ensemble de questions méritant d'être étudiées.

Les cerrados constituent un immense ensemble biogéographique s'étendant sur deux millions de kilomètres carrés, près d'un quart du territoire brésilien. Sur le plan physionomique, le cerrado typique est une savane composée d'une couverture herbacée de 50 à 70 centimètres de hauteur, et d'un étage supérieur discontinu d'éléments arborés ou arbustifs. Ce "*cerrado tipo*" donne à l'ensemble son unité, et résulte, sur le plan écologique, de la conjonction de sols acides et d'un climat à deux saisons, l'une sèche, l'autre humide ; ces conditions sont réunies sur un peu plus des deux tiers du territoire. Ailleurs, des variantes générées par des facteurs limitants comme la qualité des sols ou les conditions hydrologiques sont à l'origine des formations plus ouvertes telles que les "*campo cerrado*", "*campo sujo*", "*campo limpo*", ou plus fermées comme le "*cerradão*" sur à peu près un quart de la superficie. L'aire d'extension des cerrados est une bande de largeur variable, d'orientation NE/SW, couvrant la majeure partie des états du

Goiás, Mato Grosso, Mato Grosso do Sul, Minas Gerais, et une partie importante du *Maranhão, Piauí, Bahia*. Cette rapide description de l'homogénéité biogéographique relative n'aurait pas d'intérêt si elle n'avait pas été à l'origine de la transformation de ce grand espace en un enjeu considérable pour l'Etat brésilien : le contrôle et l'intégration nationale de "*l'Interior*".

Limités à leur partie la plus massive, les cerrados appartiennent à plusieurs grandes régions définies par l'Instituto Brasileiro de Geografia e Estatística (IBGE), (Figure 1), à la fois au Nordeste, au Sudeste et au Cento-Oeste. Cette observation conduit directement à se poser la question de l'intérêt (autre que technocratique) de réaliser une banque de données socio-économiques à propos d'un espace chevauchant les principales divisions régionales du Brésil. C'est précisément à partir de cette remarque qu'est née la problématique de recherche : les cerrados sont composés de la partie périphérique de plusieurs régions brésiliennes très diverses, couvrant une grande variété de niveaux de développement. Qu'y a-t-il de commun entre, d'une part, la frange sud du Goiás ou du Mato Grosso do Sul, jouxtant l'Etat de São Paulo, phare de l'économie brésilienne, et, d'autre part, les zones dépressives du Piauí ou de l'ouest de Bahia ? Rien, si ce n'est d'appartenir à une même région biogéographique, les cerrados. Cette caractéristique est si forte, que nombreux sont ceux qui, au niveau des grands organismes d'aménagement de l'Etat, n'hésitent pas à parler de la "*Région de Cerrados*". S'agit-il d'un simple abus de langage, ou bien, comme c'est peut-être le cas, de la volonté déclarée d'ériger les cerrados en témoin avancé d'une rupture avec le sous-développement, un peu comme l'avait déjà fait le Président Juscelino Kubitschek, en installant la nouvelle capitale, Brasília, au coeur des cerrados ? Si cette hypothèse s'avérait exacte, les cerrados seraient une sorte de creuset, ou mieux, un laboratoire vivant, où l'on pourrait observer des transformations rapides résultant des jeux d'influences convergentes et se répercutant sur tous les aspects de la vie économique et sociale.

Le géographe est directement conduit à observer et comprendre le processus de formation de ce qui n'est pas encore vraiment une région mais qui pourrait le devenir. L'inscription dans l'espace géographique des effets des forces d'intégration en présence est une problématique de recherche dont les résultats peuvent apporter des informations utiles au développement des cerrados.

Une problématique sans méthodologie, c'est poser une série de questions sans se donner la possibilité de leur apporter une réponse. La banque de données SISECSO constitue à la fois le moyen et le produit de la recherche. C'est le moyen car il apparaît préférable d'asseoir l'essentiel du travail de recherche sur la petite



échelle, c'est-à-dire d'abord sur la totalité des deux millions de kilomètres carrés. Ce n'est pas une pratique habituelle à l'ORSTOM, mais c'est la meilleure façon d'apprécier la formation de la région prise dans son ensemble. SISECSO sera aussi le produit de la recherche en renfermant la plupart des informations nécessaires à l'analyse et à la compréhension des processus.

Cette approche différente doit être mise en perspective avec ce que certains ont pu appeler la *"révolution quantitative"* affectant la géographie française depuis une quinzaine d'années. Il s'agit d'une profonde mutation, qui ne peut être réduite à un recours au nombre plus fréquent qu'auparavant. En modifiant leur approche des phénomènes, en adoptant le modèle d'analyse systémique, en mettant l'accent sur la nécessité de ne pas privilégier un niveau d'analyse ou une échelle particuliers aux dépens des autres, c'est une autre manière de penser et de pratiquer la géographie qui a fait son apparition. La lecture de la revue *"L'Espace Géographique"* donne une bonne idée des travaux produits par cette *"nouvelle géographie"* durant la décennie 1970- 1980. La différence la plus aisée à apprécier relève de la présentation de tableaux numériques ou de graphiques comme on en avait peu vu jusqu'alors, issus principalement des techniques d'analyse statistique multivariée : droites de régression, plans factoriels, arbres de classification. C'est probablement pour cette raison que la discipline scientifique *"géographie"* a été affublée de l'épithète *"quantitative"* par ses détracteurs, tenants de la géographie *"classique"* qui doit se pratiquer *"les pieds dans la boue !"* Les pionniers de la géographie quantitative française ayant démontré l'intérêt et l'efficacité de leur méthode et de leurs techniques d'analyse, toutes les branches de la discipline y font aujourd'hui appel, mais à des degrés assez divers. SISECSO est un exemple de pénétration de la géographie quantitative dans un secteur jusqu'alors épargné !

La méthodologie peut être résumée de la manière suivante : à partir d'informations (statistiques pour l'essentiel) relevées sur la totalité du territoire des cerrados, à des dates différentes permettant d'évaluer les évolutions, et enregistrées dans une banque de données informatisée, on cherche à mettre en évidence les principaux mouvements d'intégration régionale, à l'aide de techniques mathématiques et graphiques d'analyse de l'espace. Cependant, rien n'interdit la vérification de certains résultats en allant a posteriori sur le terrain, mais sûrement pas l'inverse. On pourrait objecter que les données statistiques brésiliennes ne constitueraient pas une source d'information fiable. Il semble plus constructif d'examiner s'il n'est pas possible de leur faire dire quelque chose. Après tout, n'est-ce pas une des formes de la pratique de la recherche : analyser en connaissance de cause des données partielles, entachées d'erreurs, à l'aide de techniques appropriées ?

3. SOURCES STATISTIQUES, AIRE D'ÉTUDE ET UNITÉS D'OBSERVATION

Le Brésil est un pays de tradition statistique ancienne. La philosophie positiviste d'Auguste Comte, encore si présente aujourd'hui, n'est peut-être pas étrangère à cette demande de données numériques. En 1808, un dénombrement montrait l'existence de 4 millions d'habitants, mais il ne s'agissait que d'une estimation visant à informer la métropole portugaise de l'état de ses domaines en Amérique du Sud. Ce n'est qu'en 1872 qu'eut lieu le premier vrai recensement démographique, digne de ce nom en raison du contrôle effectif de chacune des phases de sa réalisation. Suivirent, en 1890 et 1900, deux autres recensements de la population, puis en 1920 furent ajoutés ceux de l'agriculture et de l'industrie. 1940 marque un tournant de l'appareil statistique brésilien : outre le recensement démographique, furent réalisés les recensements économique, agricole, industriel, commercial, des transports et des communications. Cette immense tâche avait été rendue possible par la fondation, en 1934, de l'Institut National de Statistique, devenu, en 1938, l'Institut Brésilien de Géographie et de Statistique (IBGE). Chargé à la fois des enquêtes et études statistiques, géographiques et cartographiques, l'Institut a conservé son nom jusqu'à nos jours, en étendant ses activités au domaine de la planification. En 1950, 1960 et 1970, l'IBGE réitéra ses opérations de 1940 en les affinant et les adaptant aux nouvelles caractéristiques économiques et sociales du pays. En 1975, les recensements économiques devinrent quinquennaux. En 1980 eut lieu le dernier en date des recensements démographiques, en même temps que les recensements économiques, dont les plus récents datent de 1985 (et ne sont pas encore disponibles).

Réduire la production statistique de l'IBGE à la réalisation des recensements serait une grossière erreur. Comme tous les grands instituts du même type dans le monde, l'IBGE tend à devenir l'organe central de gestion et de restitution des données générées par le fonctionnement de l'appareil administratif. Cette évolution prend en particulier la forme de la coordination des services statistiques existant dans chaque état de la Fédération, mais également de la conservation de divers fichiers dans son grand centre informatique de Rio de Janeiro. Par exemple, l'IBGE publie, depuis 1973, deux recueils relatifs aux productions municipales agricoles et à l'élevage (PMA et PMP), réalisés à partir du croisement d'informations multiples, en provenance des agriculteurs, des conseillers agricoles et des coopératives. Ces données, paraissant avec deux ans de retard, sont considérées par les spécialistes de

l'économie agricole comme étant plus fiables que les recensements de l'agriculture.

La part la plus importante des statistiques de l'IBGE fait apparaître des tableaux au niveau des états de la Fédération ; les agrégats nationaux n'ont de sens que s'ils sont examinés à la lumière des disparités territoriales. Heureusement, pour tout ce qui concerne les données de base (démographie et études économiques notamment), l'IBGE utilise comme unité territoriale le "*município*" : c'est l'unité politico-administrative la plus petite, la moins hétérogène (bien que très disparate dans les états les moins peuplés), celle qui permet d'asseoir une recherche géographique.

En l'absence de travaux préliminaires au CPAC, c'est à ces données de base disponibles au niveau municipal qu'on s'est attaché en premier lieu.

La délimitation précise de l'aire d'extension des cerrados est difficile et a été à l'origine de nombreuses controverses. D'un côté, on trouve ceux qui réduisent les cerrados à leurs formes les plus typiques, excluant les marges et les enclaves ayant d'autres compositions biogéographiques. A l'opposé, on peut penser qu'une délimitation large favorisera ultérieurement le choix d'un périmètre plus spécifique, basé non plus sur les seules caractéristiques physiques homogènes, mais aussi sur l'appartenance à une région géographique. Cette seconde option paraissant plus opératoire pour la recherche envisagée et correspondant aux choix du CPAC, il a suffi de calquer les limites des cerrados définies par l'EMBRAPA sur une carte municipale du Brésil (à l'échelle du 1/2 500 000) obtenue à l'IBGE après bien des péripéties (la plupart des travaux faits par les Brésiliens se limitent le plus souvent aux états ou groupes d'états formant des régions ; demander une carte à une échelle unique, couvrant à la fois les grands espaces du Mato Grosso et ceux, plus densément peuplés du Minas Gerais, apparaissait un peu extravagant...mais cette carte existait au service des atlas de l'IBGE, perdu dans la périphérie nord de Rio de Janeiro).

Contrairement au maillage ancien et peu évolutif de la France (le Brésil n'ayant pas encore achevé son processus de contrôle territorial), la maille municipale des cerrados évolue assez rapidement. Sur les 885 *municípios* référencés dans le recensement de 1980, plus de 400 n'existaient pas en 1950 ; de 1970 à 1980, plus d'une quarantaine de *municípios* ont été créés. Il n'est pas possible d'ignorer ces transformations car elles faussent les éventuelles comparaisons entre les recensements de 1970 et 1980. De plus, les *municípios* concernés sont précisément ceux où l'expansion démographique a eu pour conséquence la reconnaissance d'une communauté, se traduisant par la création d'une municipalité. C'est la raison pour laquelle il a fallu recalculer les valeurs données par le recensement de 1980 dans le cadre de la maille municipale de 1970.

Lorsque les nouveaux étaient de simples partitions d'un seul municipio ancien, la solution était simple : il suffisait d'additionner les chiffres de 1980. Par contre, lorsqu'un nouveau municipio était constitué avec des parties de plusieurs autres (comme c'est souvent le cas dans le Mato Grosso), il a fallu modifier la maille de 1970 pour former un municipio théorique comparable avec la réunion de ceux ayant donné une portion au nouveau. Par exemple, dans l'état du Mato Grosso, les municipios *Nobres et Chapada do Guimaraes* ont vu leur superficie se réduire comme une peau de chagrin entre 1970 et 1980 ; quatre nouveaux municipios ont été créés à leurs dépens : *Paramatinga, Nova Brazilandia, Sinop et Colider*, Sinop chevauchant d'ailleurs le territoire des deux municipios originaux. Pour comparer les données de 1970 et de 1980, un municipio théorique rassemble les deux de 1970 et les six de 1980 (fig. 2). De 885 municipios en 1980, la banque de données n'en connaît donc que 856 strictement identiques en 1970 et 1980. Bien entendu, le même travail devra être réalisé pour rendre semblables les mailles de 1980 et 1985 ou 1970 et 1985.

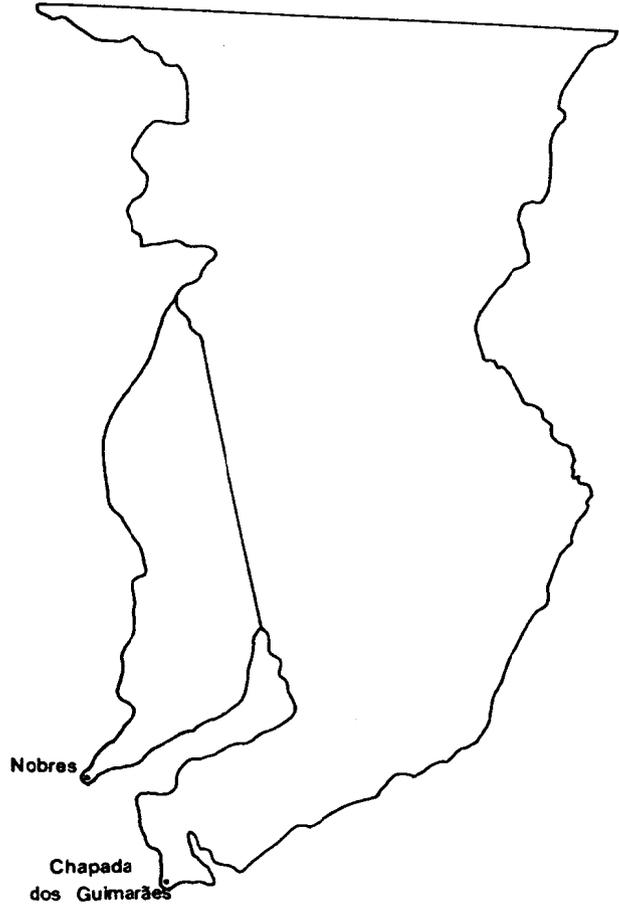
4. ADOPTER UN MODELE D'INFORMATISATION

Préalablement au choix d'un modèle de mise en oeuvre de l'informatique, il a fallu procéder à une analyse des diverses possibilités. Résultat : trois options différant sur le plan des délais nécessaires à la programmation, du niveau technique requis et du mode d'utilisation des données.

La première option, connue sous le nom de "*système d'information*", a pour vocation d'enregistrer la totalité des documents de nature diverse traitant d'un thème général, par exemple l'environnement, ou d'un espace tel qu'une ville ou une région. Un cas particulier est celui des systèmes d'information géographique, dans lesquels sont rassemblées à la fois des cartes, des enquêtes de terrain, des séries statistiques, des mesures de télédétection, etc.. De telles réalisations, devant répondre non seulement aux besoins des chercheurs mais aussi à ceux des administrateurs et même des citoyens, requièrent des moyens humains et matériels considérables, ainsi qu'une politique à long terme ; en effet, l'ensemble ne peut être vraiment opérationnel, c'est-à-dire apte à proposer un large éventail d'applications, qu'après de longues étapes de mise au point du progiciel gouvernant le stockage et la restitution de l'information, mais aussi l'enregistrement des documents de base à proprement parler.

La seconde option, diamétralement opposée à celle qui précède, met l'accent sur la capacité de l'ordinateur à réaliser un traitement plus rapide de données plus volumineuses, rassemblées

1970



1980

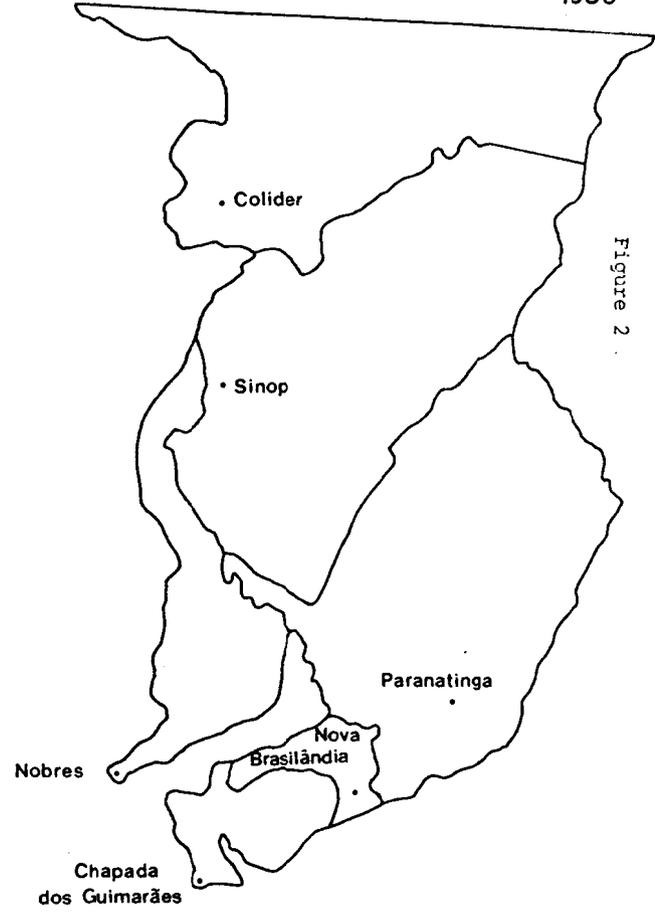
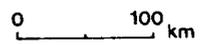


Figure 2



dans le cadre spécifique d'une problématique de recherche. La fonction d'archivage est, dans ce cas, réduite au minimum, c'est-à-dire à l'élaboration d'un fichier enregistré sur un support magnétique où les programmes de traitement, le plus souvent statistiques, viendront lire les données. Ce procédé a montré ses avantages mais aussi ses limites : les résultats sortent plus vite, mais les données ayant permis de les obtenir, souvent collectées au prix d'un grand effort, sont délaissées par la suite et deviennent inaccessibles à la communauté scientifique, faute du minimum de gestion requis : l'analyse secondaire est impossible.

Enfin, une troisième option s'impose depuis quelques années, celle des bases de données alliant la rapidité des réalisations à la rigueur de gestion. Des progiciels "*clés en main*" permettent d'enregistrer facilement une information rassemblée dans le cadre d'un problème particulier, tout en facilitant sa préparation pour l'analyse ; certains d'entre-eux intègrent en un seul ensemble les fonctions d'archivage et d'analyse. Ainsi est-il possible d'obtenir dans un délai raisonnable, de l'ordre de 18 mois à deux ans, des résultats de recherche, et cela tout en préservant l'avenir. En effet, rien ne s'oppose à la récupération des données par un système d'information plus ambitieux, avec un avantage substantiel : on connaît mieux leur signification et leur domaine de validité. Dans le cadre institutionnel de SISECSO, cette dernière option présentait de nombreux avantages sur les autres et a été finalement adoptée, provoquant une rupture avec l'équipe d'informaticiens de SISGEO, favorables à la première possibilité.

Acheter les droits d'utilisation d'un progiciel, c'est s'assurer, de manière contractuelle, un certain nombre d'avantages. Par exemple, on est sûr que les programmes ont été testés par des professionnels, c'est-à-dire que la probabilité de l'occurrence d'une erreur ou d'un incident est très réduite. En second lieu, la documentation livrée avec le système est souvent abondante, parfois didactique, et lorsqu'elle ne l'est pas, parfois relayée par des ouvrages en librairie. C'est ensuite la faculté d'échanger ses expériences avec d'autres utilisateurs et de progresser ainsi dans la maîtrise de cet outil. Enfin, et toujours dans le même ordre d'idée, il est quelquefois possible de recourir aux services d'une assistance technique qui peut à la fois dépanner les usagers et repertorier leurs critiques. Ces avantages évidents ne dispensent pas pour autant de l'élaboration d'un cahier des charges pour le choix définitif du produit ou, plus exactement de l'exposé des points clé, ceux qui conduiront à une discrimination immédiate de l'offre du marché. Bien entendu, cet examen est directement dépendant du type d'information à traiter, elle-même fonction de la problématique du chercheur, les données constituant un matériel adapté à une recherche et non l'inverse. Par exemple, si on considère le cas des

matrices d'information spatiale, tableaux rectangulaires où les lignes figurent des unités spatiales et les colonnes des attributs géographiques mesurés sur ces unités, les points clé sont les suivants : 1/ *facilité de gestion des matrices*, de sélection, d'association et de codification des attributs ou des unités, 2/ *grande variété de traitements statistiques*, univariés ou multivariés, descriptifs ou inférentiels, 3/ *représentation graphique des résultats* sous la forme de cartes en deux dimensions (cartes ponctuelles ou choroplèthes) ou trois (blocs diagrammes), 4/ *interactivité* (langage établissant le dialogue entre le système et le chercheur), *convivialité* (facilité du dialogue), *ouverture et portabilité* complétant cette liste non limitative. Il n'y a que très peu de produits sur le marché répondant à la totalité de ces spécifications.

Les avantages du *Statistical Analysis System* ont été décrits ("*SAS, un outil à la hauteur des besoins des géographes*") dans l'ouvrage de G. Dandoy intitulé "*l'Infographie à l'ORSTOM*".

Aux arguments généraux présentés dans ce texte, il faut en ajouter un plus particulier à SISECSO : le volume relativement important d'informations a fait préférer le Centre de Calcul de l'EMBRAPA (CCE) de Brasilia à une solution bâtie autour d'un micro-ordinateur. Doté de deux ordinateurs IBM 4341, l'un destiné au temps partagé sous VM/CMS, l'autre au batch sous OS/VS2, le CCE propose SAS à ses utilisateurs qui le pratiquent depuis de nombreuses années (malheureusement limités aux produits BASE, FSP et ETS). SAS apparaît ainsi comme un langage commun à un grand nombre de chercheurs de l'EMBRAPA pouvant exploiter eux-mêmes des bases de données SAS, en toute indépendance vis à vis des informaticiens.

5. PLACE DE SAS DANS LES PRINCIPALES ÉTAPES DE SISECSO

En raison de la grande variété des fonctions proposées par SAS, toutes les étapes de réalisation de SISECSO y ont fait appel, à des degrés divers.

5.1. La constitution du tableau d'identification des municipios

L'identification des unités d'observation est un préalable à toute constitution de banque de données. A l'aide de PROC FSEDIT, ce tableau est saisi en "*plein écran*", de la manière suivante :

1/ *le nom du municipio* sur 40 caractères,

2/ *son code* sur la carte de repérage, composé de l'identifiant de l'État de la Fédération sur deux lettres, suivi du numéro d'ordre sur sept chiffres,

3/ *ses codes numériques* figurant dans les fichiers de l'IBGE en 1980 et 1970,

4/ *d'autres informations complémentaires* comme les coordonnées géographiques du siège et la population en 1980.

FSEEDIT est très bien adapté à ce travail nécessitant un temps de saisie assez long et de nombreuses corrections successives, en particulier de l'orthographe des noms des municipios. Les nouveautés apportées par la version 5, notamment l'initialisation du tableau à l'appel de la procédure et surtout l'option AUTOSAVE ajoutent un confort certain à ce travail assez ingrat.

5.2. La lecture des bandes des recensements de l'IBGE

Les données fournies sur bande magnétique par l'IBGE sont enregistrées de deux manières. S'il s'agit de données récentes diffusées à l'aide du progiciel local SIDRA, on a affaire à un format "*caractères*". Pour les informations plus anciennes (de 1970 et 1975), l'enregistrement, réalisé par des programmes PL/1, se présente en format binaire. Les formats internes à SAS, ici \$8., IB4. ou RB4., simplifient considérablement l'opération de lecture des bandes, sans avoir à écrire un programme ad hoc.

5.3. La constitution des tableaux définitifs

La mise en forme et le stockage des tableaux définitifs résultent de l'exécution de deux opérations. La première a trait aux nomenclatures qu'il faut rendre compatibles d'un recensement à l'autre. Une étape DATA assure cette homogénéisation de chaque tableau : les diverses variables relatives à une nomenclature très désagrégée ou à des modalités de tableaux croisés sont regroupées en une seule nomenclature commune à tous les recensements.

La seconde opération, plus délicate, revient à agréger les nouveaux municipios, ceux qui sont apparus entre 1970 et 1980, de manière à exprimer les valeurs dans une maille administrative unique. PROC MEANS réalise cette opération selon le principe suivant :

```
PROC MEANS DATA=ENTREE SUM NOPRINT ; VAR V01-V99 ;
      OUTPUT OUT=SORTIE SUM=S01-S99 ;
      BY CODAGREG ;
```

Dans le tableau ENTREE figurent les observations d'origine, correspondant à une maille municipale différente pour chaque recensement ; le tableau SORTIE contient les observations agrégées selon une maille municipale unique où chaque observation est repérée par le code CODAGREG, variable supplémentaire du tableau des codes présenté plus haut.

5.4. Numérisation de la carte des municipios des cerrados

Une banque de données à base géographique doit être en mesure de restituer l'information sous forme de cartes. La carence quasi totale de l'EMBRAPA en informatique graphique n'a pas permis de réaliser des cartes thématiques. Cependant, le CPAC disposait d'un micro-ordinateur CANON AS-100 et d'une table à digitaliser japonaise (MUTOH) au grand format A0. Lorsqu'on a voulu numériser le fond de carte, cela n'a pas été possible immédiatement : ce matériel était inutilisable faute d'un progiciel adapté. Il a donc semblé utile de concevoir et de programmer en langage BASIC (le plus facilement utilisable sur CANON AS-100) un petit ensemble de programmes pour saisir le fond de carte.

Réalisé avec l'aide de Violette Cabos, ingénieur cartographe au CNRS, ce travail a donné lieu à l'élaboration d'un manuel pratique de numérisation publié en France (par le GIP RECLUS), dans sa version PC DOS. La figure 3 présente le résultat de cette numérisation qui prend la forme d'un fond de carte directement utilisable par SAS GRAPH sur les données de base.

6. CONCLUSION : COMPOSITION ET AVENIR DE SISECSO

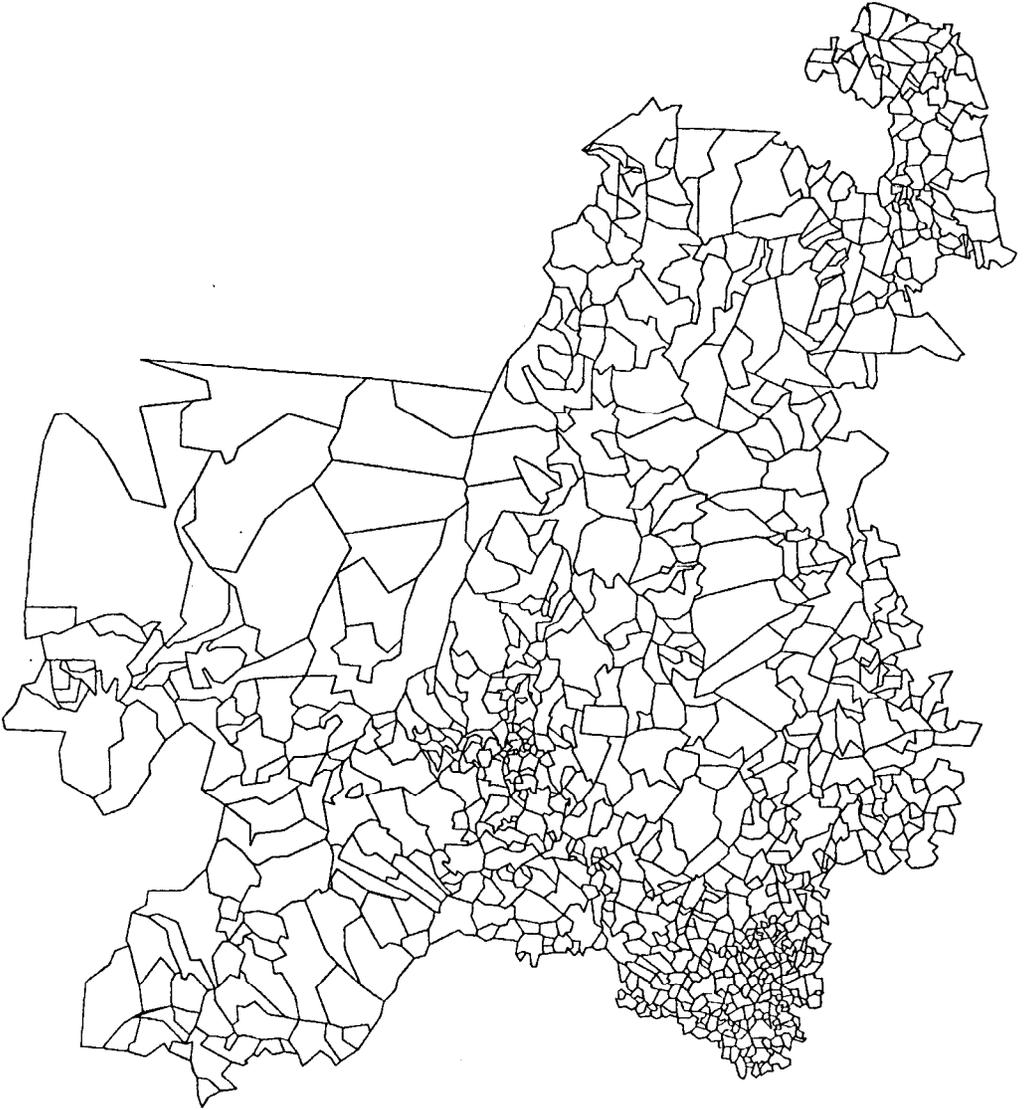
Au printemps 1987, SISECSO se compose de trois bases de données informatisées implantées sur le centre informatique de l'EMBRAPA.

La base de données agricoles comprend plus d'une centaine de variables pour chacun des recensements de 1970, 1975 et 1980. Ces variables sont strictement comparables et couvrent les domaines suivants : propriété de la terre, utilisation du sol, production agricole et élevage, taille des exploitations.

Les productions de l'agriculture et de l'élevage sont stockées dans une base de données spécifiques. Pour quelques 55 cultures, on dispose pour chaque année de 1977 à 1984, de la quantité produite, de la surface récoltée et de la valeur de la production ; pour l'élevage, on connaît la quantité et la valeur d'une quinzaine de productions.

La base de données démographiques, plus réduite, environ 80 variables, comprend les données des recensements de 1970 et 1980. Comme précédemment, ces variables sont directement comparables. Il s'agit à la fois de données strictement démographiques, comme le nombre d'habitants, les migrations ou la structure familiale, mais aussi d'informations sur le logement permettant de se faire une idée à la fois sur les habitants et sur leurs conditions de vie.

Figure 3



L'achèvement de la base de données 1970 - 1980 aura lieu fin 1987. Par la suite, les travaux pourront s'orienter selon trois directions : 1/ *l'exploitation scientifique* des données, 2/ *l'extension de la banque de données* à la totalité du Brésil, 3/ *l'alimentation en informations nouvelles*.

En l'absence de moyens informatiques graphiques suffisants, on n'est pas en mesure de faire un travail sérieux d'exploitation des données de SISECSO au Brésil. Aucune solution locale à court terme n'étant envisageable, il n'y a pas d'autre possibilité que de réaliser ce travail en France, sur l'un des systèmes du CNUSC ou du CIRCE. Il est regrettable que l'EMBRAPA ne puisse mettre à disposition les matériels et progiciels indispensables ; en leur absence, il est impossible de mener une recherche intéressante, au moment où elle devient réalisable, c'est-à-dire une fois acquises les sources statistiques.

Peut-être serait-il utile qu'un chercheur brésilien, intéressé par SISECSO, ait la possibilité de se former à l'analyse quantitative de l'espace en France ; une fois rentré au Brésil, il pourrait sans doute, plus efficacement que moi, convaincre ses interlocuteurs du bien-fondé de ces méthodes.

SISGEO est un projet couvrant la totalité du Brésil. En réalité, à part le module SISOLOS, tous les autres ne couvrent qu'une partie seulement du territoire brésilien. SISECSO n'échappe pas à la règle. En effet, chaque module est pris en charge par un centre de l'EMBRAPA, ce qui explique que SISECSO n'ait trait qu'aux seuls cerrados. Il est donc indispensable d'entreprendre un travail du même type, ailleurs au Brésil ; il n'y a que l'embarras du choix. Mais il serait vain de vouloir faire strictement la même chose et retomber ainsi dans un *"empirisme abstrait"* stérile. Il y a certainement d'autres questions qui se posent ailleurs, pouvant être abordées avec la même méthodologie quantitative.

A la fin de l'année 1987, le CPAC disposera d'une banque de données composée d'un peu plus de 3 000 variables, relevées sur 856 municipios. La forme sous laquelle ces données auront été stockées en ordinateur leur garantira la plus grande facilité d'accès. Mais ces données, relatives à la décennie 1970-1980, devront être complétées au fur et à mesure de la sortie de nouveaux recensements.

**ANNEXE : TRAVAUX DE PHILIPPE WANIEZ EN
RAPPORT AVEC SISECSO**

1/ *Les données et le territoire* (ouvrages techniques)

- Initiation au traitement informatique des données spatialisées. Paris, ORSTOM, coll. IDT, 119 p.

- Initiation à la numérisation pour la cartographie thématique. Montpellier, GIP RECLUS, 54 p.

- Initiation à l'analyse en surfaces de tendances. A paraître, 36 p.

2/ *Communications*

- Du système du technocrate à celui du géographe. VIIIème séminaire d'économie et de sociologie rurales. Montpellier, CIRAD/MESRU, 14-18 Septembre 1987

- Comment monter une base de données géographiques avec SAS. Contribution à la réunion annuelle des utilisateurs du système SAS. Chantilly, 21-22 Octobre 1987

3/ *Articles*

- Approche d'un grand espace : la maille municipale des cerrados brésiliens. Montpellier, GIP RECLUS, Mappemonde, n°3-1987, pp.38-41.

- Base de données pour la recherche géographique : la dynamique spatiale agricoles des cerrados au Brésil. Paris, ORSTOM, Cahiers de Sciences Humaines, à paraître.