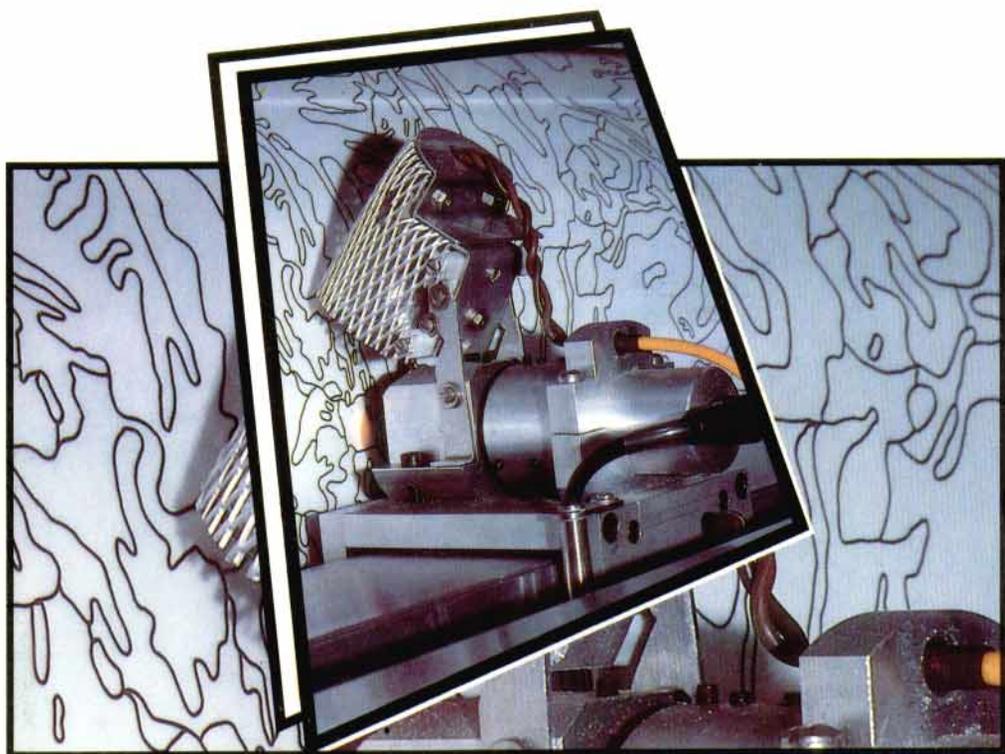


SÉMINFOR 3

Systemes d'information pour l'environnement



Troisième séminaire informatique
de l'ORSTOM
Bondy, du 26 au 28 septembre 1989

Editions de l'ORSTOM

SÉMINFOR 3

Systemes d'information pour l'environnement

Troisième séminaire informatique
de l'ORSTOM
Bondy, du 26 au 28 septembre 1989

Editions de l'ORSTOM

INSTITUT FRANÇAIS DE RECHERCHE SCIENTIFIQUE POUR LE DÉVELOPPEMENT EN COOPÉRATION

Collection **COLLOQUES et SÉMINAIRES**

PARIS 1990

Commission scientifique des sciences de l'ingénieur
et de la communication, Mission technique
Informatique & Mission technique Formation

**Séminfor 3 : Troisième
séminaire informatique de
l'ORSTOM**

(Bondy, 26-28/09/89)

*"Systèmes d'information pour
l'environnement"*

Textes recueillis et présentés par Patrick Séchet

La loi du 11 mars 1957 n'autorisant, aux termes des alinéas 2 et 3 de l'article 41, d'une part, que les «copies ou reproductions strictement réservées à l'usage privé du copiste et non destinées à une utilisation collective» et, d'autre part, que les analyses et les courtes citations dans un but d'exemple et d'illustration, «toute représentation ou reproduction intégrale, ou partielle, faite sans le consentement de l'auteur ou de ses ayants droit ayants cause, est illicite» (alinéa 1er de l'article 40).

Cette représentation ou reproduction, par quelque procédé que ce soit, constituerait donc une contrefaçon sanctionnée par les articles 425 et suivants du Code pénal.

AVANT-PROPOS

Les séminaires informatiques de l'ORSTOM, Séminfor, sont des réunions organisées conjointement par la Commission Scientifique des Sciences de l'Ingénieur et de la Communication, CS07, et la Mission Technique Informatique, MTI, avec un soutien financier de la Mission Technique Formation, MTF. Elles se proposent essentiellement de réunir les acteurs de l'informatique à l'Institut, quelles que soient leurs disciplines de rattachement, autour d'un thème méthodologique relevant des techniques de traitement automatisé de l'information.

Ainsi, Séminfor 1 s'est tenu au siège de l'ORSTOM à Paris, du 06 au 08 octobre 1987, pour répondre à la question : "*Bases de données et systèmes d'information : quelles méthodes*". Le deuxième séminaire a eu lieu au Centre ORSTOM de Montpellier, du 26 au 28 septembre 1988, pour aborder un sujet intitulé : "*La modélisation : aspects pratiques et méthodologie*".

La troisième édition de Séminfor a été organisée au Centre ORSTOM de Bondy, du 26 au 28 septembre 1989, sur le thème des "*Systèmes d'information pour l'environnement*", dans le but de faire connaître l'état actuel des réalisations de l'Institut dans ce domaine, et de favoriser une réflexion collective.

Séminfor 3 a été ouvert par Monsieur Gérard WINTER, Directeur général de l'ORSTOM : son discours introductif est reproduit ici. En parallèle aux exposés préparés par les intervenants, il a été l'occasion de présentation de "posters" et de plusieurs démonstrations de logiciels, tant sur PC-compatibles que sur stations de travail sous environnement UNIX.

Vingt-deux communications ont été présentées et organisées en trois sessions : *capitalisation des données, de la donnée à l'information (aide à l'interprétation) et stratégies de spatialisation*. Chacune des sessions a été brièvement introduite par l'animateur, tandis que le rapporteur a fait un résumé des points les plus marquants des débats : tous ces textes sont recueillis dans ce volume.

COMITE DE LECTURE

Bernard LORTIC, Yveline PONCET et Pascal RENAUD.

SECRETARIAT DU SEMINAIRE

Patrick SECHET (+)

SOMMAIRE

Ouverture.

Discours introductif. Gérard WINTER (Directeur général de l'ORSTOM) 9

Première session : capitalisation des données.

(animateur : Pascal Renaud, rapporteur : Jean-Jacques Lechauve)

Exposé introductif de l'animateur de session : Systèmes d'information sur l'environnement, nouveaux outils, nouveaux concepts. Christian MULLON & Pascal RENAUD ; 17

PLUVIOM : un logiciel de gestion des données pluviométriques. Patrick RAOUS, Jacques GUISCAFRE, Jean-François BOYER & Anne CRESPIY ; 37

La base de données hydropluviométriques du Laboratoire d'hydrologie. Gilbert JACCON & Michel TRAVAGLIO ; 49

SISOND : une conception de l'informatique mise au service de l'utilisateur. Daniel CARDON ; 61

SISPRAGAS : un abordage de système d'information pour le contrôle des insectes phytophages. Patrick SECHET, José MARTINS, Luiz BELARMINO & Alci LOECK ; 79

Acquisition, traitement et transmission de données de volcanologie. J. TABBAGH, C. PAMBRUN & M. LARDY ; 91

Intégration d'imagerie satellitaire dans un système d'information géographique. Bernard LORTIC ; 101

Compte-rendu du rapporteur de session. 111

Deuxième session : de la donnée à l'information, aide à l'interprétation.

(animateur : Jean-Yves Martin, rapporteur : Patrick Séchet)

Exposé introductif de l'animateur de session ; 117

Utilisation de l'informatique dans la gestion et le traitement de données biologiques. Daniel CARDON ; 119

Modèles numériques de terrain : le logiciel BABEL. Marc SOURIS & Dominique REMY ; 135

Cartographie synthétique sur MacIntosh d'un ensemble de variables régionalisées par un logiciel intégré d'analyse factorielle de correspondances, de classification et de cartographie automatiques. Application à la régionalisation agricole de la Sierra équatorienne. Marie-Madeleine THOMASSIN ; 141

A propos des cerrados du Brésil : de l'informatique de recherche aux systèmes d'information. Philippe WANIEZ & Gérard DANDOY ; 159

Détermination assistée par ordinateur : base de données, clé informatisée ou système expert ? L. TITO DE MORAIS ; 173

Traitement de données satellitaires pour la connaissance de l'environnement en milieu tropical : applications en Guyane. Marc LOINTIER (travail présenté sous forme de "poster") ;	187
Compte-rendu du rapporteur de session.	201

Troisième session : stratégies de spatialisation.

(animateur : Gérard Dandoy, rapporteur : Dominique Rambaud)

Exposé introductif de l'animateur de session ;	205
SISIMAGEM : une méthodologie pour l'utilisation de la télédétection à l'EMBRAPA. Deusdeth SANTOS, Eduardo ASSAD & Severino JUNIOR ;	211
Traitement de données météorologiques satellitaires et conventionnelles pour la connaissance de l'environnement. Dominique DAGORNE ;	227
NEPTUNE : un système pour la constitution et l'exploitation de bases de données sur l'environnement. Anne-Marie AUBRY, Cossi-Jean HOUNDAGBA, Jean-François RICHARD & Patrick SECHET ;	241
Un prototype de système d'information géographique pour la planification urbaine et son utilisation à Quito (Equateur). Marc SOURIS ;	253
Exemple de manipulation de données intégrées à un système d'information géographique. Dominique COURET ;	261
Extraire, visualiser et archiver une carte avec LEICA. François DELCLAUX & Claudio LAWIRA ;	277
SISGEO : une méthodologie d'administration des données de l'environnement rural. Patrick SECHET, Gérard COCHONNEAU & Paulo FRESNEDA ;	287
Système d'information géographique pour l'état de Veracruz (Mexique) : programme pilote INEGI/ORSTOM. Luc CAMBREZY.	319
Proposition d'un schéma de collecte de l'information pour le milieu physique. Hervé LE MARTRET & Alain BEAUDOU ;	331
Compte-rendu du rapporteur de session.	349

Liste des participants 353

Index par auteur 361

EXPOSE INTRODUCTIF DE SEMINFOR 3

(Gérard WINTER, Directeur général de l'ORSTOM)

INTRODUCTION

Je me réjouis de la tenue de cette rencontre, organisée conjointement par la Commission scientifique des sciences de l'ingénieur et de la communication (CS7) et la Mission technique informatique (MTI). Je réponds avec satisfaction à l'invitation de Jean-Yves Martin & Patrick Séchet, de la CS7, d'introduire les travaux de ce séminaire, comme mon prédécesseur l'avait fait pour Séminfor 1, en 1987.

C'est avec un grand plaisir que nous accueillons un représentant de nos partenaires, *Ingénieur Miguel Angel Vizconde*, directeur de géographie à l'INEGI au Mexique, institut avec lequel nous menons un de nos projets de coopération les plus réussis. Soyez le bienvenu Monsieur Vizconde. Je vous remercie de nous faire l'honneur de participer à cette rencontre : je suis sûr que vous nous aiderez à ne pas perdre de vue la finalité de toutes les actions de nos équipes, le développement.

Voici donc réunie pour la troisième fois la communauté informatique de l'ORSTOM, à l'initiative de la CS7 qui joue là pleinement son rôle d'animation des activités transversales de l'Institut, comme elle l'a fait la semaine dernière à Montpellier, avec les "Journées Laboratoire", dans un autre secteur relevant de sa compétence.

Séminfor 1, le premier séminaire de cette série s'est tenu à Paris en octobre 1987, sur le thème des méthodes et outils pour les bases de données ; pour Séminfor 2, réalisé en septembre dernier à Montpellier, on avait choisi le sujet de la modélisation. Séminfor 3 constitue une occasion de débattre d'une question très importante pour l'ORSTOM et ses partenaires : les systèmes d'information pour l'environnement.

Tous ces sujets méthodologiques concernent l'ensemble des disciplines de l'Institut, de sorte que Séminfor constitue un lieu de rencontre privilégié, permettant de favoriser ou consolider le développement d'une pluridisciplinarité effective à l'Institut, mais aussi de susciter des échanges entre les différentes équipes et d'identifier les complémentarités, malgré les contextes extrêmement variés dans lesquels elles doivent s'inscrire et travailler. Il faut en effet chercher une convergence entre nos actions et éviter les concurrences entre les équipes.

Je continuerai donc à apporter mon soutien pour la réalisation de ces séminaires annuels, en espérant que Séminfor 4 et les suivants abordent quelques uns des nombreux sujets que constituent les instruments et méthodes modernes d'acquisition, gestion et traitement de l'information, comme la télédétection, l'acquisition automatique et la télétransmission, les systèmes experts, le génie logiciel, etc., et l'assimilation de celles-ci par les équipes de l'Institut.

COMMENTAIRES SUR SEMINFOR 3

Pour revenir à Séminfor 3, permettez-moi de faire une remarque préliminaire : pourquoi ne pas avoir choisi comme thème "*Systèmes d'information pour le développement*", plutôt que "*Systèmes d'information pour l'environnement*" ?

Le terme "*environnement*" a pour moi un double sens : celui qui se rapporte à l'écosystème local et celui, global, de l'échelle planétaire.

Il devient de plus en plus évident que bon nombre des problèmes de développement des pays tropicaux trouvent leur origine dans les difficultés rencontrées pour la préservation du patrimoine que constituent les ressources naturelles d'un territoire. Le plus souvent, ces difficultés sont elles-mêmes issues de l'inadéquation entre les activités humaines et la gestion de l'espace où celle-ci se développe.

Il faut se garder toutefois de ramener tous les problèmes de développement à des problèmes de gestion d'environnement. De plus, l'ORSTOM n'a pas pour objet d'étudier l'environnement planétaire. S'il le fait parfois c'est parce que le développement doit se réaliser sous la contrainte de préserver cet environnement.

Le travail d'élaboration des systèmes d'information pour l'environnement répond ainsi directement à la mission assignée à l'Institut, qui est de contribuer au développement des pays partenaires, par la voie des recherches qu'il mène en coopération avec ceux-ci. Il y répondra d'autant mieux que seront pris en compte tous les aspects de l'environnement physique, biologique et humain, d'où l'intérêt particulier que représente les *Systèmes d'information géographique, SIG* pour le développement. Il est évident que l'ORSTOM occupe une position privilégiée pour la mise au point de ce type d'outils : nous devons en faire une de nos priorités.

J'ai constaté à l'examen de la programmation de cette rencontre que les organisateurs ont identifié trois sessions distinctes : *la capitalisation de l'information, de la donnée à l'information et stratégies de spatialisation*. C'est à partir de cette grille de lecture que je me pose un certain nombre de questions, que je soumetts à la réflexion des participants à ce séminaire et auxquelles je souhaiterais que des éléments de réponse soient apportés.

Capitalisation de l'information

En ce qui concerne cette première session, au cours de laquelle seront présentées plusieurs bases de données monothématiques, je ferai trois constatations.

La première concerne la question de la sélection de nos actions dans ce domaine, face à l'accumulation des données existantes et à la progression du volume de celles-ci, surtout depuis l'apparition de l'acquisition automatique dans certains cas. Comme j'ai déjà eu l'occasion de le dire, l'ORSTOM ne peut faire tout, partout, très bien : là comme ailleurs, il nous faut faire des choix. Il faut donc réfléchir à la pertinence des bases de données que nous réalisons, eu égard aux enjeux stratégiques du développement. Cette réflexion ne doit pas considérer uniquement la nature des données, mais aussi leur degré d'agrégation, les fonctionnalités du système, la possibilité de croisement avec d'autres données, l'extension géographique et le domaine de validité de la base.

Je pense par ailleurs qu'une analyse comparée des travaux de structuration des bases de données effectués par les différentes disciplines de l'Institut serait particulièrement intéressante, compte tenu des différentes natures des données collectées et administrées.

Les uns, je pense aux hydrologues, aux océanographes physiciens, aux géophysiciens ou encore aux climatologues, collectent essentiellement des données quantitatives, le plus souvent sous forme de séries chronologiques. Ils ont donc rapidement utilisé l'outil informatique pour développer et perfectionner progressivement des traitements appropriés à l'acquisition, la validation, la gestion, au traitement et à la synthèse de ces informations. Les avancées technologiques dues à nos équipes dans tous ces domaines sont significatives.

D'autres, comme les pédologues, les botanistes ou les océanographes biologistes, par exemple, se sont heurtés d'emblée au problème de la complexité taxonomique et de la manipulation de données descriptives, pour lesquelles des méthodes d'administration satisfaisantes ont été mises au point, mais qui restent difficiles à traiter. Enfin, plus récemment, la dimension géographique commence à être prise en compte par l'intermédiaire des SIG, pour la représentation des informations dites spatialisées.

L'ORSTOM est encore bien placé pour faire en sorte que les avancées des uns et des autres soient assimilées par tous, tant il est vrai que les données de l'environnement présentent toutes, à des degrés divers, ces trois dimensions temporelle, taxonomique et spatiale. Toutes ces données d'observation constituent en effet des mesures : même lorsqu'elles sont quantitatives, il ne faut pas oublier qu'elles sont associées à une clé qui représente le rattachement de cette mesure à un lieu, une date, etc., laquelle est qualitative.

La dernière préoccupation que j'évoquerai à propos de cette première session est inhérente au risque de séparation que je perçois entre le gestionnaire des données et les utilisateurs potentiels de la base. Les premiers systèmes d'information réalisés à l'ORSTOM étaient le fait de quelques spécialistes thématiques, non informaticiens : ils avaient donc pour le moins ceux-là comme utilisa-

teurs. De plus en plus, les nouveaux systèmes sont pris en charge par des équipes spécialisées dans ce type de travail, parce qu'ils sont plus complexes à mettre en oeuvre (même en tenant compte des facilités nouvelles d'aide à la réalisation) : il faut néanmoins veiller à ce qu'ils répondent bien aux besoins des scientifiques et favoriser l'appropriation de cet outil par ces derniers.

De la donnée à l'information, aide à l'interprétation

La deuxième session m'est apparue peu développée et je m'interroge sur les raisons de cette relative pauvreté, d'autant que le reproche a souvent été fait aux chercheurs de l'ORSTOM de se contenter de collecter des données, et de laisser à d'autres le soin de les interpréter ! Pour certains cet état de fait n'a pas fondamentalement changé.

Je crois toutefois nécessaire de pondérer cette opinion. En effet, capitaliser n'est pas simplement accumuler, de sorte que les bases de données et systèmes d'information apportent en réalité une valorisation significative des données. Les données y sont mises en forme et sont disponibles avec un certain nombre de fonctionnalités de base (d'extraction, de sélection, d'inventaire, d'édition, etc.), ce qui constitue déjà en soi un premier effort d'utilisation et une aide substantielle à l'interprétation.

Pourtant, l'acquisition directe par un chercheur confère à celui-ci un pouvoir critique sur les données, difficilement transférable, de sorte qu'il possède un privilège énorme pour l'interprétation des observations qu'il a lui-même réalisées.

La remarque qui me vient à l'esprit à l'occasion de ce deuxième groupe d'exposés concerne le problème de propriété des données. Cette question est fondamentale, dans la mesure où l'on prend progressivement conscience de la valeur de l'information, qu'elle soit ou non scientifique. Les données collectées sur l'environnement appartiennent évidemment (aussi) à nos partenaires : d'ailleurs, l'ORSTOM a d'ores et déjà assuré un certain nombre d'opérations de transfert (base de données hydrologiques, fonds documentaires, etc.).

Il faut en tout cas se préparer à négocier avec le partenaire le droit d'accès aux données, en vue de réaliser les applications souhaitées : les programmes scientifiques, qu'ils prévoient ou non la réalisation d'un système d'information, devraient désormais se préoccuper de cet aspect.

Dans cette optique, la politique menée par les hydrologues me paraît judicieuse. Elle consiste en effet à utiliser les logiciels de gestion de données déjà mis au point par l'Institut (HYDROM, par exemple) comme monnaie d'échange pour garantir l'accès aux données.

Il faut bien sûr pouvoir toujours disposer librement de notre savoir-faire (constitué par ces logiciels), ce qui suppose qu'on l'utilise avec discernement lors des négociations, dans l'intérêt de l'Institut dans son ensemble, toutes disciplines confondues.

Stratégies de spatialisation

La dernière session est consacrée aux *stratégies de spatialisation* : les SIG sont des outils tout à fait novateurs, encore qu'il me semble que ce concept revêt, même à l'intérieur de notre maison, diverses acceptions, qu'il est pour le moins nécessaire de clarifier. C'est un sujet très important et la réflexion qui va être amorcée jeudi devra se poursuivre la semaine prochaine au niveau de l'UR 3J (Analyse des organisations régionales et gestion des milieux agro-pastoraux) du département MAA, à l'occasion du groupe de travail *Analyse comparée des modèles pour l'information spatialisée, ACMIS*.

Je crois important de susciter une réflexion vis à vis de l'échelle d'intégration pluridisciplinaire. En effet, à petite échelle, c'est à dire sur de grands espaces, la mise en regard des informations n'aura pas le même degré d'efficacité que l'intégration des grands volumes de données dont on disposera avec un SIG (sur la même région géographique, mais avec un niveau de détail beaucoup plus poussé). Néanmoins, il me semble qu'il existe une combinaison disciplinaire féconde à chaque échelle.

CONCLUSION

En résumé, je voudrais insister sur l'importance que revêtent à mes yeux les instruments et méthodes de traitement de l'information à l'ORSTOM, et plus généralement les activités scientifiques transversales regroupées sous le vocable de *Sciences de l'ingénieur et de la communication*. Ces disciplines, sans objet de recherche propre dans notre maison, constituent des partenaires pour les autres domaines scientifiques représentés par les programmes de recherche en cours. Elles peuvent ainsi jouer un rôle pour favoriser la pluridisciplinarité autour des démarches de l'informatique, de l'analyse physico-chimique ou encore de la communication.

Comme je l'ai annoncé aux récentes *journées de septembre*, il nous faut réfléchir sur notre dispositif institutionnel dans ce domaine, constitué par les missions techniques. Je suis toutefois convaincu qu'il nous faut maintenir et renforcer ces compétences : ces ingénieurs et techniciens jouent un rôle indispensable de veille technologique et nous permettent l'appropriation des nouvelles méthodes.

Je déclare ouvert le troisième séminaire informatique de l'ORSTOM, Séminfor 3 : je souhaite que vos travaux soient féconds et que chacun tire les meilleurs bénéfices de cette rencontre.

Session 1 :

CAPITALISATION DES DONNEES

SYSTEMES D'INFORMATION SUR L'ENVIRONNEMENT : NOUVEAUX OUTILS NOUVEAUX CONCEPTS

(*Christian MULLON & Pascal RENAUD*)

RESUME - Les auteurs présentent dans ce document quelques développements récents de l'informatique et montrent comment ceux-ci peuvent trouver des applications dans les programmes de recherche développés à l'Institut.

Dans un premier temps, en prenant le cas d'une banque de données naturalistes, les possibilités d'implantation à partir des modèles relationnels, sur micro ou mini-ordinateur, et du modèle hypertexte sont évoquées. La seconde partie est consacrée à l'étude d'un grand programme pluridisciplinaire et aux possibilités d'application des méthodes orientées objet. Une comparaison, sur ce cas particulier, avec des méthodes plus classiques, est également avancée.

INTRODUCTION

Nous avons donné à cette communication d'introduction au troisième séminaire informatique de l'ORSTOM, un titre très général. Il est souhaitable, pour en comprendre les objectifs, de la replacer dans le cadre du développement de l'informatique scientifique à l'ORSTOM, et notamment dans celui de la politique d'ensemble mise en place par la Mission Technique Informatique. Cette politique est basée sur l'idée qu'il faut concevoir le processus d'informatisation d'un organisme de recherche de façon dynamique, comme une dialectique entre de nouvelles possibilités technologiques et de nouvelles expressions de besoins, et bien entendu en se donnant des impératifs d'adaptation et d'efficacité.

C'est une politique :

- de standards (Unix, tcp/ip, SQL, X-Windows, Postscript) ;
- de projets adaptés (RIO, réseau informatique ORSTOM) ;
- de veille technologique et d'animation scientifique (Séminfor, Séminaires de formation) ;
- de collaboration avec les instituts de recherche en Informatique (INRIA, LAFORIA de Paris VI, LRI de Paris XI).

Il est évident que le thème des systèmes d'information sur l'environnement est un des thèmes récurrents des programmes scientifiques de l'ORSTOM. De nombreux projets sont en effet confrontés à cette thématique :

- observatoires (celui du Sahel) ;
- systèmes d'informations localisées (Atlas de Quito, SIG Vera-Cruz, etc.) ;
- modèles de changement global (Projet HAPEX) ;

- systèmes de production (Programme Delta Central du Niger).

De nombreuses expériences et des projets de mise en place de systèmes d'information sur l'environnement sont conduites à l'ORSTOM, et seront présentées lors de ce séminaire.

Nous souhaitons ici étudier, d'une façon générale, quelles sont les implications des développements informatiques actuels dans deux domaines particuliers : celui des bases de données naturalistes et celui des programmes multi-disciplinaires. Cela afin de montrer qu'en prenant le point de vue de la recherche en informatique, pour aborder des thèmes de recherche spécifiques de l'ORSTOM, on peut définir une problématique originale. Notre objectif est ici de montrer comment les outils et les concepts actuellement élaborés en informatique s'appliquent aux besoins des chercheurs de l'ORSTOM, de montrer comment l'originalité des systèmes d'information sur l'environnement peut être un sujet de recherche en informatique et enfin, de façon pragmatique, de donner des indications pour des solutions techniques immédiatement opérationnelles.

1. BASES DE DONNEES NATURALISTES

1.1. Introduction

Les exemples de bases de données naturalistes sont nombreux à l'ORSTOM. On peut citer un certain nombre de programmes qui visent à constituer :

- un herbier (herbier de Cayenne) ;
- un inventaire faunistique (systématique des poissons d'eau douce en Afrique de l'Ouest) ;
- un inventaire géologique ;
- un inventaire pédologique (RCA).

Aujourd'hui, il est naturel d'en envisager l'informatisation. C'est naturel, ce n'est pas automatique, ce qui est heureux. On peut classer les solutions d'informatisation en deux grandes catégories basées sur deux modèles alternatifs de représentation de l'information : le modèle relationnel et le modèle hypertexte.

1.2. Le modèle relationnel

1.2.1. Principes

Le modèle relationnel est décrit dans le cadre d'une partie de la théorie mathématique des ensembles, l'algèbre relationnelle, qui formalise des opérations abstraites d'intersection, d'appartenance, d'inclusion, de projection, de produit, etc.. On peut en présenter les grandes lignes de façon plus suggestive en mettant en avant les concepts :

- Table (ou relation). Information mise sous la forme d'un tableau rectangulaire. Chaque colonne est appelée un attribut, par exemple :

Table Individus			
nom	age	ville	profession
Jérôme	34	Bordeaux	Viticulteur
Alain	24	Bondy	Commerçant
Hélène	56	Paris	Médecin
Laurent	18	Angoulême	Sans
Marcelle	68	Bordeaux	Sans

- Base de données. Ensemble de tables pouvant avoir des attributs communs. Par exemple :

Table Individus			
nom	age	ville	profession
Jérôme	34	Bordeaux	Viticulteur
Alain	24	Bondy	Commerçant
Hélène	56	Paris	Médecin
Laurent	18	Angoulême	Sans
Marcelle	68	Bordeaux	Sans

Table Ville		
Ville	Population	Aéroport
Paris	2.000.000	oui
Bondy	45.000	non
Angoulême	87.000	non
Lyon	800.000	oui

- Requête. Manipulation élémentaire de l'information contenue dans la base (interrogation, mise-à-jour. Exemples : "Sélectionner tous les individus de la base ayant une profession et les classer par age croissant". Ou encore "Sélectionner tous les individus de la base habitant dans une ville ayant un aéroport".

- Jointure. Opération qui consiste à croiser des informations se trouvant dans deux tables distinctes d'une même base de données.

Ce qui est important, et qui explique en partie les succès du modèle relationnel, est l'existence d'un langage de requête standard : SQL. Il existe d'autres langages de requête (QUEL, QBE, QBF), mais en général ils sont liés à l'offre d'un développeur et ne sont pas normalisés.

Un autre trait important du modèle relationnel réside dans le principe de normalisation des bases de données relationnelles. Il existe un ensemble de règles contraignantes sur la structuration des tables. Ainsi la première règle de normalisation consiste à imposer une forme rectangulaire pour toutes les tables. Les autres règles de normalisation visent à :

- éviter toute redondance de l'information ;
- assurer son intégrité, c'est à dire à éviter des contradictions dans l'information stockée ;
- la protéger dans le cas d'interrogation et de mise à jour simultanées d'une même base, par un grand nombre d'utilisateurs.

On doit noter également que les systèmes mis en place dans le cadre du modèle relationnel permettent d'en assurer la répartition, c'est à dire qu'il y a possibilité pour les diverses tables de se trouver physiquement sur des ordinateurs différents et connectés en réseau (figure 1).

Nous présentons maintenant, avec leurs implications pratiques, deux possibilités d'implantation de bases de données relationnelles.

1.2.2. Solution Micro-ordinateur avec dBASE

On doit remarquer en premier lieu que dBASE n'est pas à proprement parler un système relationnel et qu'en particulier, jusqu'à de récentes versions, il n'intégrait pas le langage SQL.

Marche à suivre :

- 0 première estimation des volumes d'information et des principaux traitements, débouchant sur une base de taille moyenne et sur une exploitation mono-utilisateur ;
- 1 achat du logiciel et du matériel. Micro-ordinateur de type IBM/PC (20 000 à 50 000 francs). Logiciel dBASE III (environ 8 000 francs) ;
- 2 structuration de la base (définition des tables et des fonctions principales) ;
- 3 programmation des principales fonctions. Programmes de saisie, de mise à jour et d'interrogation, programmes de calcul suivant les cas (de 3 à 12 mois de programmeur, 30 000 francs par mois en prestation de service) ;
- 4 mise en forme et saisie du rétrospectif (coût de saisie en soustraction d'environ 2 centimes le caractère) ;
- 5 mise en exploitation.

Avantages et limitation :

- volume d'information. En théorie illimité, de fait difficulté d'exploiter des bases de l'ordre de 10 Mégaoctets ;

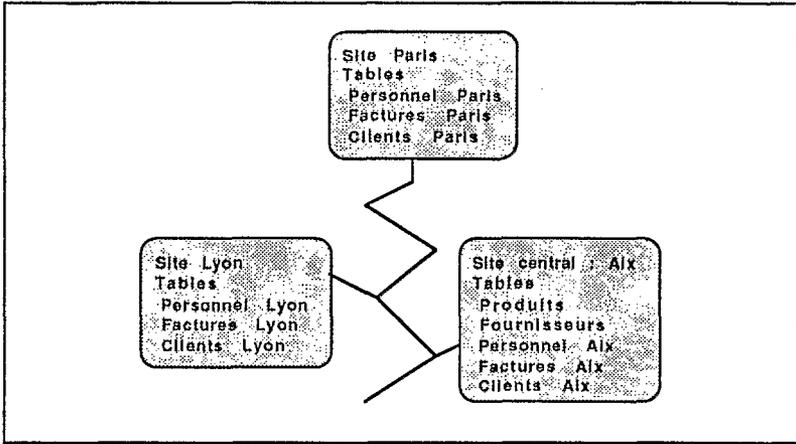


Figure 1 Exemple de base de données répartie

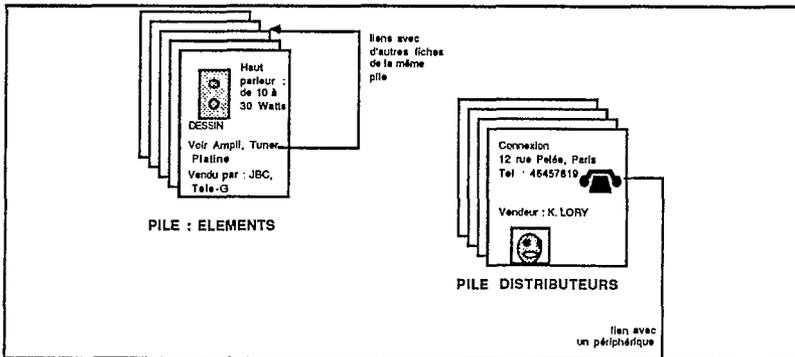


Figure 2 Exemple de base de données hypertexte

- structure relationnelle. En apparence seulement (pas de contrôle d'intégrité, pas de langage de requête déclaratif) ;
- langage de programmation (procédural) propre à dBASE.

1.2.3. Mini-ordinateur avec Ingres

Marche à suivre :

- 0 première estimation des volumes d'information et des principaux traitements, débouchant sur une base de grande taille, sur des gros débits et sur une exploitation multi-utilisateurs ;
- 1 achat du logiciel et du matériel. Mini-ordinateur sous le système UNIX (à partir de 60 000 F). Logiciel INGRES (environ 100 000 francs, mais gratuit sur SUN pour les applications "recherche") ;
- 2 structuration de la base (définition des tables et des fonctions principales) ;
- 3 programmation des principales fonctions. Programmes de saisie, de mise à jour et d'interrogation, programmes de calcul suivant les cas (de 12 à 36 mois de programmeur, 30.000 francs par mois en prestation de service) ;
- 4 mise en forme et saisie du rétrospectif (coût de saisie en soustraction d'environ 2 centimes le caractère) ;
- 5 mise en exploitation.

Avantages et limitation :

- volume d'information en fait illimité ;
- interrogation à distance possible ;
- répartition effective ;
- structure relationnelle effective et complète, avec contraintes d'intégrité ;
- gros débit transactionnel ;
- langage de requêtes standard (SQL).

1.3. Le modèle Hypertexte

1.3.1. Principes

Le modèle Hypertexte est un des premiers modes proposés pour structurer l'information contenue dans les bases de données multi-média, c'est à dire de grands ensembles d'informations de structures différentes (textes, tableaux numériques, photographies, schémas, sons, films).

Le modèle Hypertexte, appelé parfois suggestivement *modèle navigationnel*, est ainsi une façon de structurer une base de données multi-média de façon à pouvoir en extraire des informations pertinentes (figure 2). Les concepts les plus importants du modèle Hypertexte sont :

- carte. Support élémentaire d'information ;
- pile. Ensemble de cartes pouvant être consultées à la suite les unes des autres ;
- base de données hypertexte. Ensemble de piles contenant des cartes qui peuvent être liées les unes aux autres de façon très souple.
- fond. Structure de cartes.

Il est très important de noter ce qui différencie ce modèle des précédents modèles de représentation de l'information : il peut y avoir dans une même pile des cartes de fond différent.

Le modèle Hypertexte est particulièrement adapté pour un logiciel se substituant à ce qui est un outil fondamental du chercheur naturaliste : le *cahier de terrain*. Celui-ci peut comporter, sous une forme décidée par le seul chercheur, selon ses manies, des informations de toute sorte, comme :

- des relevés d'observations ;
- des adresses ;
- des collages photographiques ;
- des notes de lecture ;
- des remarques sur le sujet ;
- des remarques hors sujet ;

qu'il veut pouvoir consulter très librement, sans procédures pré-établies.

De la même façon que pour le modèle relationnel, on peut indiquer les implications pratiques du choix du modèle Hypertexte au sein d'une opération de recherche.

1.3.2. Modalités d'implantation : Hypercard

Marche à suivre:

- 0 évaluation des volumes et des principales fonctions ;
- 1 acquisition d'un Macintosh (de 20 000 à 50 000 francs, le logiciel Hypercard est fourni en standard avec le Macintosh) ;
- 2 éventuellement acquisition d'un Scanner (12 000 francs) ;
- 3 réalisation de la saisie rétrospective ;
- 4 mise en exploitation.

1.3.3 Modalités d'implantation : constitution d'un CD-ROM

Un CD-ROM est un disque laser vendu dans le commerce. On peut aujourd'hui de façon très simple, le connecter à un micro-ordinateur, obtenant ainsi un périphérique :

- de très grande capacité (500 MO) ;
- d'accès assez rapide (un petit peu plus lent qu'un disque dur traditionnel) ;
- d'accès en lecture seulement ;
- de gros débit.

Marche à suivre :

- 0 analyse préalable des volumes et des besoins des utilisateurs ;
- 1 recueil de l'information ;
- 2 programmation de l'interface utilisateur, de un (solution Hypercard) à six mois (solution adaptée) de programmeur. Coût par un façonnier : 100 000 francs ;
- 3 saisie de l'information ;

- 4 presse du CD-ROM. Coût : 50 000 francs ;
- 5 diffusion du CD-ROM. Coût du lecteur : 10 000 francs et coût du disque : 400 francs. On notera que le marché français est actuellement loin d'être équipé.

1.3.4. Comparaison des deux modèles

	Relationnel	Hypertexte
données	numériques texte	numériques texte images films, son
structuration	rigide	souple
interrogation	langage de requête normalisé : SQL	langage de requête graphique
calcul	interfaçage naturel des tables produites par une requête avec les langages de calcul les plus courants	au coup par coup

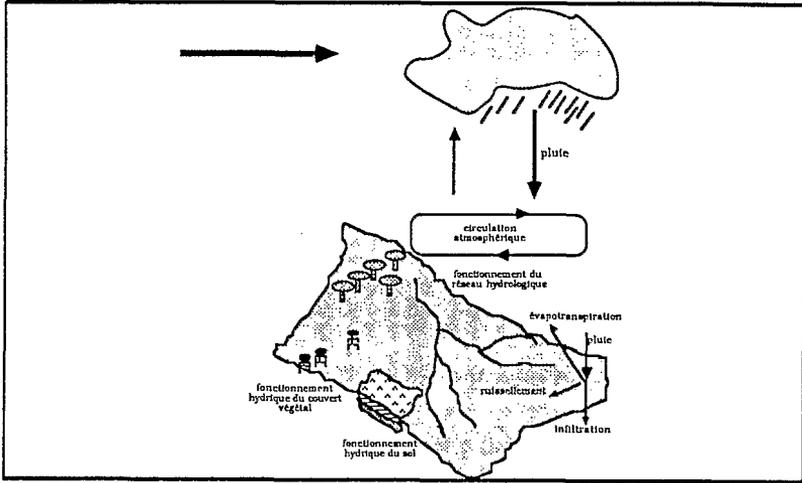
2. PROGRAMMES MULTIDISCIPLINAIRES

2.1. Un exemple de programme multidisciplinaire : le projet HAPEX

HAPEX signifie Hydrological Atmospheric Pilot Experimental. C'est un grand projet scientifique international placé sous la responsabilité de Michel HOEPFFNER, responsable de l'UR "Continent, Atmosphère, Séries Climatiques" du département TOA. Les organismes participants sont fort divers : ORSTOM, CNRM, LAMP, Université de Dijon, Université de Toulouse III. Il s'agit d'un projet de modélisation du bilan hydrologique et du bilan énergétique en climat tropical sec. L'ensemble des phénomènes pris en compte dans un tel projet est représenté sur la figure 3 ci-après (la zone d'étude est de un degré carré : 110km x 110 km), tandis que la figure 4 présente le dispositif expérimental mis en oeuvre.

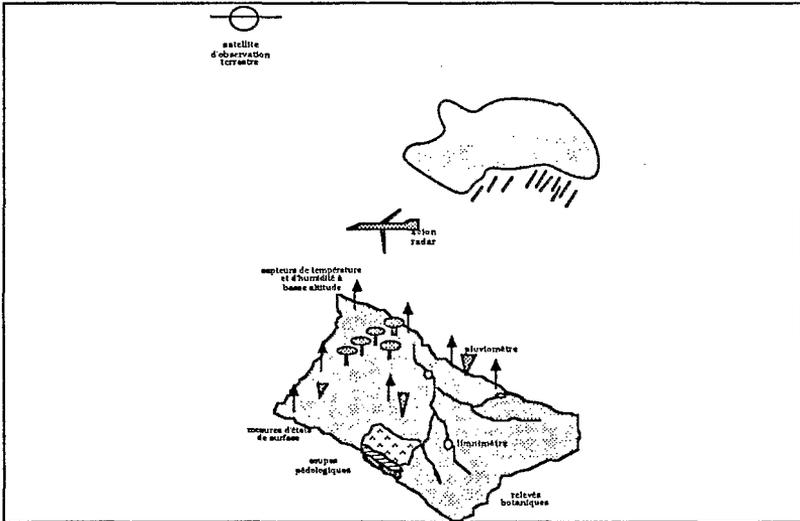
2.2. Problématique de l'informatisation d'un projet multidisciplinaire

On conçoit les difficultés de l'informatisation d'un programme comme HAPEX. Elles proviennent de multiples sources.



LE PROJET HAPEX : VUE D'ENSEMBLE

Figure 3



LE PROJET HAPEX : DISPOSITIF EXPERIMENTAL

Figure 4

2.21. Volume des données

En première estimation, l'ensemble des données recueillies serait d'environ 1000 bandes magnétiques de densité 6250 bpi, ce qui représente un total de 30 milliards d'octets.

2.22. Origine des données

La multidisciplinarité d'un projet comme HAPEX se traduit dans la diversité de la forme des données. Les hydrologues, les climatologues, les pédologues, les géologues, les botanistes ont tous une pratique du traitement des données qui leur est propre, et ont défini des modes de structuration de l'information pertinents pour leur problématique propre.

2.23. Forme des données

Il faut noter également que la diversité des échelles employées pour recueillir des données se traduit également au niveau de leur structuration. Cette diversité est particulièrement impressionnante dans HAPEX (figure 5). On note des mesures de phénomènes aux échelles suivantes :

- échelle du globe et sur une longue durée (modèles climatologiques, images des satellites météorologiques) ;
- échelle du pays (cartes pédologiques, cartes du couvert végétal) ;
- échelle du bassin versant (modèles hydrologiques) ;
- échelle ponctuelle (coupes pédologiques) ;
- échelle moléculaire (modèles d'équilibre chimique).

2.24. Utilisation des données

Une des difficultés de l'informatisation d'HAPEX provient de ce qu'à la différences des projets de gestion, dans un projet de recherche :

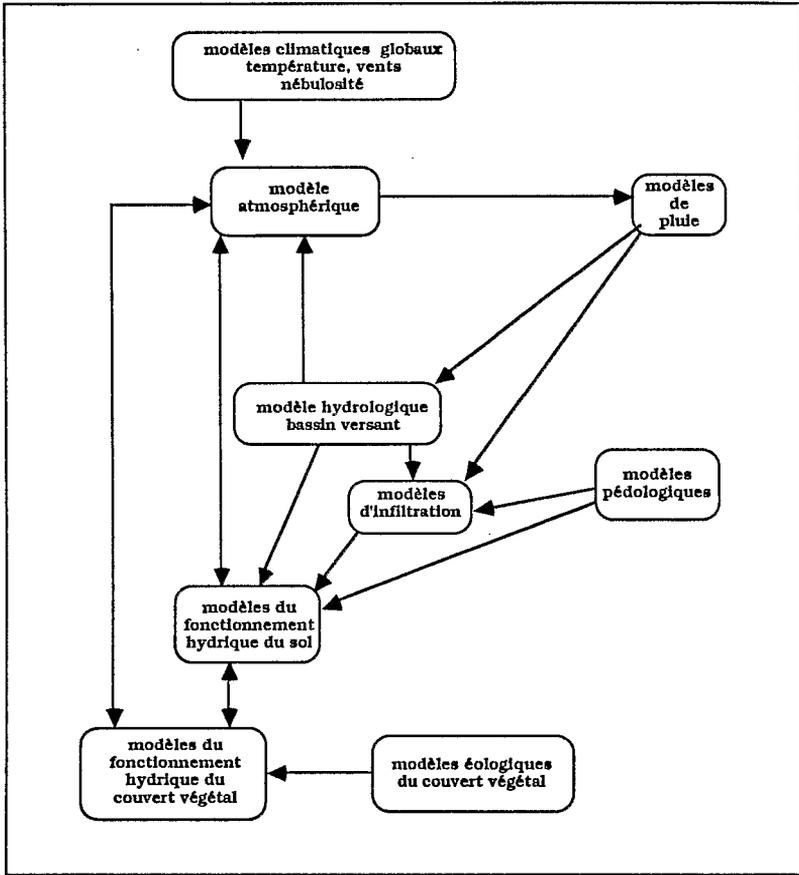
- le but est de trouver un modèle acceptable, non de faire tourner un modèle existant ;
- la mise en place du modèle sera progressive ;
- l'objectif n'est pas la répétabilité, mais une étude globale.

Or la plupart des méthodes d'informatisation et des produits logiciels ont été définis pour la mise en place de projets de gestion et se révèlent peu adaptés.

2.3. Solutions pour l'informatisation d'un projet multidisciplinaire

Le problème de l'informatisation d'un projet multidisciplinaire réside dans l'intégration, dans un ensemble unique, d'un acquis informatique au sein de chaque discipline et de nouvelles questions provenant de contacts multidisciplinaires.

Nous proposons ici d'étudier comment se présentent trois solutions standards que nous définissons de façon schématique comme : pragmatisme, méthode et objet.



LE PROJET HAPEX : SCHEMA LOGIQUE

Figure 5

2.3.1. La solution pragmatique

Cette solution fait appel au bon sens et recommande de :

- laisser chaque équipe mono-disciplinaire utiliser les outils qu'elle maîtrise déjà ;
- définir des règles pour la constitution et l'utilisation des bases de données constituées ;
- définir des formats de données standards qui seront utilisés lors des rencontres multidisciplinaires.

2.3.2. La solution méthodique

Cette solution, qui a été très en vogue il y a quelques années et qui après expérimentation semble voir son champ d'application se restreindre progressivement, consiste à baser toute la stratégie informatique sur l'utilisation d'une méthode d'informatisation (MERISE, AXIAL) qui :

- repose sur l'idée qu'une formalisation du système d'information dans toutes ses composantes (organisationnelle, opérationnelle, temporelle) est possible, et qu'il existe des outils de représentation adaptés (schémas logiques et conceptuels des données et des traitements) ;
- préconise un certain nombre d'étapes, allant du général au particulier (schéma directeur, analyse préalable, analyse détaillée, planification du développement, réalisation du développement et mise en exploitation ;
- aboutit à la mise en place d'un atelier de génie logiciel et d'une base de données (la plupart du temps avec un modèle relationnel).

2.3.3. La solution objet

Le principe de base de la programmation objet repose sur la constatation qu'au fur et à mesure des développements de système d'information, la problématique générale de l'informatisation s'est déplacée et que l'on met de moins en moins l'accent sur la *variabilité des données* et de plus en plus sur la *variabilité des traitements*. Le but d'une informatisation n'est plus d'appliquer des traitements constants à des jeux de données différents, mais d'appliquer des traitements en évolution permanente à une base d'information, remise à jour certes, mais dont la structure est à peu près constante.

Les concepts de base de la programmation-objet sont les suivants :

- objet. Entité reconnue par le système et identifiée par une référence unique. Par exemple : point, rectangle, individu ;
- classe. Collection d'objets ayant les mêmes propriétés statiques (attributs) et dynamiques (méthodes). Par exemple : l'ensemble des points (coordonnées, déplacer, afficher), l'ensemble des rectangles (points extrêmes, agrandir, afficher), l'ensemble de tous les individus (nom, prénom, âge, adresse, changer_d_adresse) ;
- encapsulation. Une méthode n'est applicable à un objet que lorsqu'elle fait partie des méthodes de la classe de cet objet ;
- message. Donnée déclenchant l'exécution d'une méthode liée à une classe et s'appliquant à un objet de cette classe (le destinataire). Par exemple : "agrandir le rectangle X", "donner l'âge de l'individu Marcel" ;

- hiérarchie des classes. Une classe A est une sous-classe de B, si les objets de la classe A appartiennent à la classe B. Ainsi la classe des rectangles de surface 1 est une sous-classe de la classe des rectangles ;

- héritage. Transmission des propriétés d'une classe vers une sous-classe. Par exemple : toute méthode applicable à la classe des rectangles sera applicable aux rectangles de surface 1. Mais attention, ne pas confondre sous-classe et composants : une sous-classe est une spécialisation (les triangles rectangles sont une sous-classe des triangles), alors que les composants sont des parties (les sommets sont les composants d'un triangle).

Il est important de noter que dans un système orienté-objet les requêtes sont dynamiques. Une requête (un message de l'utilisateur) se propage dans le système, déclenchant à son tour, en chaîne, toute une série de nouveaux messages (figure 6). Dans un système traditionnel, les requêtes sont plus statiques : le système répond immédiatement à toute requête qui lui est adressée.

On notera également que les bases de données hypertexte sont conçues à partir des principes de la "représentation objet".

2.4. Représentation orientée objet de l'information d'un programme multidisciplinaire

Pour donner une idée de la souplesse de la représentation orientée-objet, essayons de donner un premier schéma de la construction d'un sous-ensemble des données d'un projet comme HAPEX. Ce schéma est manifestement incomplet et inexact. Mais en essayant de le corriger, on verra comment les notions de classes et de méthodes associées se prêtent à la ré-utilisation du code.

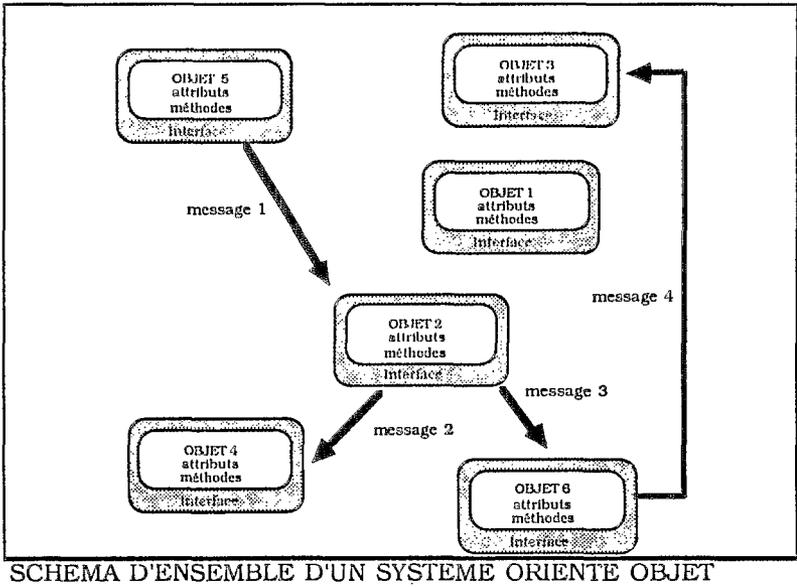


Figure 6

2.4.1. Messages

message	émetteur	destinataire
rasterisation	carte ponctuelle	carte raster
cartographie des états de surface	carte raster (satellite)	carte zone
simulation pluie	utilisateur relevés pluvio relevés pluvio relevés pluvio débits infiltration infiltration pedo eco évapo atmosphère	relevés pluvio débits évapo infiltration évapo pedo eco évapo évapo atmosphère relevés pluvio
classification supervisée	relevés ponctuels	image raster

2.4.2. Classes

TERRITOIRE

description géographique

édition de la description géographique

TERRITOIRE.HYDRO

ensemble de cours d'eau

ensemble de points cotés (MNT)

*extraction de bassins versants**représentation 3-D*

TERRITOIRE.PEDO

ensemble de zones pédologiques

représentation 2-D

TERRITOIRE.ECO

ensemble de zones écologiques

représentation 2-D des flux

ZONES

suite de points (contour)

calcul de surface

ZONES_PEDO

caractéristiques pédo

calculs de masses

ZONES_ECO

caractéristiques éco

modèles écologiques

OBS. PONCTUELLES

OBS. PONCTUELLES PEDO

ensemble de coupes

OBS. PONCTUELLES ECO

ensemble d'éco-systèmes localisés

OBS. PONCTUELLES PLUVIO

ensemble de séries pluvio

OBS. PONCTUELLES LIMNI

ensemble de séries limni

OBS. PONCTUELLES ATMO

ensemble de séries atmo

SÉRIES DE MESURES

suite de points T,F(T)
 représentation graphique
 calculs statistiques

SÉRIES PLUVIO

lieu
 description du pluviomètre
 calculs de cumuls

SÉRIES LIMNI

lieu
 description du limnimètre
 courbe d'étalonnage
 calcul de débits

SÉRIES ATMOSPHERIQUES

description du capteur
 calculs de flux

CARTE PONCTUELLE

ensemble de points de cartes

CARTE RASTER

ensemble de pixels

CARTE ZONE

ensemble de zones

2.4.3. Représentation objet et multidisciplinarité

Au sein d'une discipline on manipule naturellement des objets, des classes et des méthodes et, en quelque sorte, on pratique la représentation objet, comme Monsieur Jourdain faisait de la prose.

Dans la pratique multidisciplinaire, l'emploi maîtrisé de la représentation objet offre un intérêt supplémentaire. Les difficultés de dialogue proviennent de ce qu'au sein d'une discipline on a tendance à rester "propriétaire" de son information (dans sa définition et son contenu). Avant que l'échange ne se fasse avec une autre discipline, l'information disciplinaire doit être synthétisée en un objet structuré, propre à l'échange; et c'est dans cette opération de synthèse préalable que surgissent les hiatus de dialogue. Avec la représentation objet, notamment avec le concept d'"encapsulation", les tenants d'une discipline, en associant des méthodes à des objets, contrôlent naturellement les modes d'accès à leurs données et peuvent donc les mettre simplement à la disposition de leurs partenaires dans les autres disciplines.

2.4.4. La solution objet, implications pratiques

On trouve actuellement sur le marché des langages orientés objet : Smalltalk, C++, Object Pascal, Microsoft Pascal V.O, etc.. On trouve également, mais souvent sous forme de prototype, des SGBD orientés objets : ORION, VBASE, GBASE, IRIS, GEMSTONE, O2, et des "extensions objet" des SGBD relationnels : SABRINA, POSTGRES, EMPRESS, etc..

Ainsi, alors que la solution orientée-objet semble très séduisante pour l'informatisation des projets multidisciplinaires, il est nécessaire aujourd'hui de la considérer sous un angle expérimental. Les utilisateurs actuels doivent envisager d'essayer les plâtres tant qu'une solution bien standardisée ne se sera pas imposée.

2.5. Comparaison des trois solutions

	pragmatique	méthodique	objet
mise en place	rapide	différée	expé- ri- mentale
aide à la conceptualisation	nulle	importante très contraignante	importante souple
efficacité du système de gestion	faible	forte	?
efficacité du système de développement	moyenne	moyenne	forte
aide à la mise en place de modèles	nulle	faible	importante
forme des requêtes	statique	statique	dynamique
possibilité de prototypage	faible	faible	naturelle

CONCLUSION

Nous avons essayé dans cette communication de montrer qu'un certain nombre d'outils, actuellement développés en informatique, semblent particulièrement convenir à la mise en place des systèmes d'information sur l'environnement, notamment dans le domaine des bases de données naturalistes et celui des grands programmes multidisciplinaires. Parmi ces outils, certains sont immédiatement opérationnels, (le modèle relationnel et le modèle Hypertexte) et nous recommandons très instamment aux chercheurs de l'Institut de tenir compte de leurs possibilités dans le montage de leurs opérations de recherche. D'autres, principalement issus de la formalisation-objet, en sont à une étape expérimentale et leur utilisation dans les programmes doit être envisagée en collaboration avec les laboratoires de recherche dans le domaine. Il faut savoir que ceux-ci sont demandeurs d'expériences en grandeur nature: "*AI-DB integration requires theoretical and experimental work. Experimental AI-DB systems should be developed now to help identify difficult issues and to build a technology base for AI-DB integration. Do not wait for the theoretical problems to be solved before constructing AI-DB systems*" M.L. Brodie, "Future Intelligent Information Systems: AI and DataBase Technologies Working Together", Readings in Artificial Intelligence and Databases, 1988.

La problématique des systèmes d'information sur l'environnement peut ainsi donner lieu à la mise en place de collaborations de recherche entre l'ORSTOM, avec sa vocation multidisciplinaire, sa connaissance du terrain et la richesse de ses données, et les organismes de recherche en informatique dans le domaine des bases de données-objet, mais aussi dans ceux des bases de données déductives et des modèles connexionnistes d'apprentissage.

REFERENCES BIBLIOGRAPHIQUES

- ADIBA, M. et alii (1987). *Le PRC Bases de données -BD3*, Techniques et sciences informatiques, Vol 6, n° 5, p.498-505.
- ABITEBOUL, S. & GRUMBACH, S. (1987). *Bases de données et objets structurés*, Techniques et sciences informatiques, Vol 6, n° 5, p.382-404.
- BAILLY, C., CHALINE, J.F., FERRI, H.C., GLOES, P.Y. & MARCHESIN, B. (1987). *Les langages orientés-objet*, Cepadues-Editions.
- BANCILHON, F. & ROMAKRISHNAN, R. (1989). *An Amateur's Introduction to Recursive Query Processing Strategies*, in Readings in Artificial Intelligence and Databases, J. Mylopoulos et M.L. Bordie (editors), Morgan et Kaufman, pp.376-431.
- MASINI, G., NAPOLI, A., COLNET, D., LEONARD, D. & TOMBRE, K. (1990). *Les langages à objets*. Inter-éditions.
- GARDARIN, G. & VALDURIEZ, P. (1990). *SGBD avancés, bases de données objet, déductives et réparties*, Eyrolles.
- GARDARIN, G. & VALDURIEZ, P. (1987). *Bases de données : les systèmes et leurs langages*, Eyrolles.

- GARDARIN, G. & SIMON, E. (1987). *Les systèmes de gestion de bases de données déductives*, Techniques et sciences informatiques, Vol 6, n° 5, pp.347-382.
- CONKLIN, E.J. (1987). *Hypertext: an introduction and survey*. IEEE Computer 2,9, september 1987.
- PINON J.M. (1989). *Multimedia Communication System*. Tutorial presented at the 13th Urban Data Management Symposium (UDMS), Lisbon, Portugal, June 1989.
- VALDURIEZ, P. (1987). *Objets complexes dans les systèmes de bases de données relationnels*, Techniques et sciences informatiques, Vol 6, n° 5, pp.405-418.
- ZDONIK, S.B. & MAIER, D. (1990). *Fundamentals of object-oriented database systems*, in Readings in object-oriented database systems, Zdonik et Maier editors, Morgan et Kaufmann, pp.1-32.

PLUVIOM : UN LOGICIEL DE GESTION DES DONNEES PLUVIOMETRIQUES.

(Patrick RAOUS, Jacques GUISCAFRE, Jean-François BOYER & Anne
CRESPY)

RESUME - Développé au Laboratoire d'Hydrologie de l'ORSTOM, PLUVIOM est un logiciel qui permet la saisie et la gestion de données provenant d'un réseau d'observation pluviométrique. L'information de base se présente sous diverses formes :

- bordereaux remplis par des observateurs plus ou moins fréquemment (une ou plusieurs fois par jour, ou quelques fois par mois) ;
- enregistrements graphiques provenant de pluviographes de diverses natures ;
- enregistrement sur mémoire EPROM, provenant également de pluviographes ;
- de rapports de tournées d'inspection ou d'installation de postes de mesure.

Le logiciel dispose de toutes les fonctions permettant l'introduction de ces données de base dans les divers fichiers correspondants (identification, historique des stations, pluviométrie journalière, etc.), soit à partir de bordereaux (saisie au clavier par un opérateur), soit à partir de périphériques spécialisés (numériseur pour les diagrammes pluviographiques, lecteur de cartouches EPROM). Il permet également d'obtenir des éditions, sous des formes diverses destinées au contrôle ou à la publication de tout ou partie des données disponibles. Enfin, il est équipé de fonctions de communication assurant, aussi bien la mise à jour par des données provenant d'autres systèmes ou logiciel, que la mise à disposition de ces données pour des applications non assurées par PLUVIOM.

Deux versions compatibles sont utilisables :

- une version micro-ordinateur (sous système d'exploitation MS-DOS ou compatible), munie de facilités d'utilisation (fonctions sélectionnées dans les menus, masques d'écran rendant plus aisée la saisie, aide affichée à l'écran, documentation détaillée, etc.) ;
- une version plus sommaire, implantée sur gros système (ordinateur IBM 3090 du Centre National Universitaire Sud de Calcul), assurant le stockage, certaines éditions (identiques à celles de la version micro-ordinateur) et les communications.

1. INTRODUCTION

La grande dispersion géographique des unités de recherche de l'ORSTOM n'a pas favorisé la construction d'outils communs de traitement de données de base comme la pluviométrie. Plusieurs logiciels ont été créés conduisant à la constitution de fichiers divers, et sans possibilité de communication ou d'échange des données entre eux.

A l'occasion d'un gros travail de mise en forme des données de pluies journalières et fort de l'expérience acquise lors de l'élaboration et de l'utilisation du logiciel de gestion de données hydrométriques HYDROM, le Laboratoire d'Hydrologie a décidé de se doter d'un outil similaire de gestion des données pluviométriques et d'en assurer la diffusion.

L'intérêt d'une telle démarche est multiple. La réalisation d'un outil commun à diverses unités libère ces dernières des tâches de maintenance et de documentation, et assure un développement correspondant à des besoins généraux formulés par les utilisateurs. De plus, l'uniformisation du mode de stockage, garanti par l'utilisation d'un logiciel unique, permet une gestion plus efficace des données, qu'il s'agisse de la constitution ou de la mise à jour d'une banque à partir de données acquises de façon décentralisée ou de la cession de données à partir de la banque centrale. (schéma n°1)

PLUVIOM est conçu de façon à permettre une gestion de toutes les données concernant les stations pluviométriques : il s'agit aussi bien des informations signalétiques (localisation, équipement, vie de la station) que des valeurs de hauteur de pluie qui y sont mesurées.

Les fonctions qu'il assure sont accessibles simplement, par l'intermédiaire de menus ; l'utilisateur est aidé soit par des messages émis par le logiciel au cours des manipulations, soit par une documentation regroupant les informations ou recommandations à suivre pour mener à bien un traitement.

2. DESCRIPTION DES DONNEES GERES PAR PLUVIOM

2.1. Les types de données

PLUVIOM gère deux types distincts de données.

2.1.1. Données relatives aux stations pluviométriques

- Identification des postes de mesure. Il s'agit de l'ensemble des informations connues lors de l'installation du poste : localisation (nom du site, coordonnées, altitude), date d'installation, type de mesure, capteurs climatologiques associés.

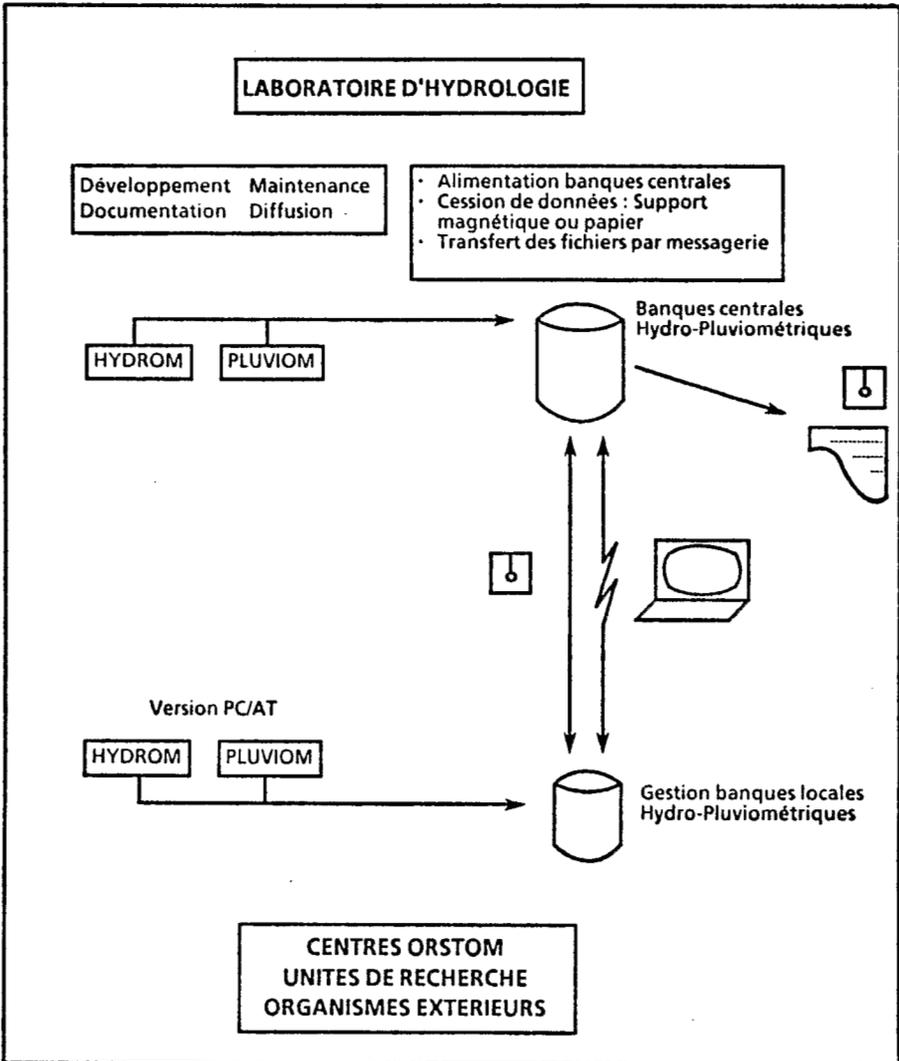


Schéma n°1: Les banques de données hydropluviométriques.

- Description des appareils. Ces informations regroupent toutes les caractéristiques des enregistreurs, qu'il s'agisse du système de mesure (augets, siphon, etc .) ou du type d'enregistreur (tambour, table déroulante, codeur). Elles sont nécessaires pour permettre un traitement correct des supports de données recueillies (diagrammes, bandes, cartouches à mémoire) à l'aide de périphériques appropriés (numériseur, lecteur de cartouches).

- Historique. Toutes les informations non quantitatives et datées, susceptibles de caractériser les postes pluviométriques, figurent dans cette rubrique. Il s'agit essentiellement de changement d'observateur, de déplacement du poste, de modification de l'environnement, de date de visite ou d'entretien, de changement d'appareil, etc.. Elles permettent à l'utilisateur de s'informer sur toute modification survenue au cours de la période d'observation.

2.1.2. Données de pluie

- Pluies journalières. Sont regroupées dans cet ensemble les hauteurs de pluie mesurées de façon quotidienne au moins. Des codifications permettent de distinguer qualitativement les valeurs mesurées (rosées, traces, neige, relevés groupés) et de leur attribuer un critère de fiabilité.

- Pluie à pas de temps variable. Ce sont les données issues du traitement des enregistrements de diverse nature (bandes, cartouches, etc.), obtenus sur des pluviographes. Les données pluviographiques se présentent sous la forme de couples instant-hauteur de pluie précipitée depuis l'instant précédent. Elles peuvent inclure d'autres valeurs permettant à l'utilisateur de contrôler le bon fonctionnement du système enregistreur (dérive d'horloge) ou la stabilité du système de mesure (hauteur recueillie au seau totalisateur).

2.2. Structure des données

Les données décrites précédemment sont réparties dans cinq fichiers principaux :

- identification des stations (en association avec une table d'identification des pays) ;
- description de l'appareillage ;
- dossier des stations ;
- pluviométrie journalière ;
- données pluviographiques.

Ces fichiers sont organisés en enregistrements identifiés par une clé et contenant toutes les données qui y sont associés.(voir tableau ci-après). Pour limiter au maximum la place qu'elles occupent, les données sont stockées dans un format binaire. Pour tous ces fichiers, l'organisation dite "séquentielle - indexée" a été retenue. Elle permet une localisation extrêmement rapide de chaque enregistrement dans des fichiers qui peuvent atteindre des tailles considérables et dans lesquels une recherche séquentielle conduirait à des temps de réponse très longs.

Fichier	Clef	Contenu de l'enregistrement
Identification des stations	Numéro de code (10 chiffres)	nom de la station, localisation, type d'appareil, mesure associée.
Pluies journalières	Numéro de code + date (année, mois)	31 valeurs de hauteurs journalières + codes associés, hauteur mensuelle et commentaires.
Données pluviographiques	Numéro de code + date (année, mois, jour, heure, min)	successions de couples. instant-hauteur

2.3. Organisation des données

Le logiciel est conçu de façon que l'utilisateur puisse séparer les deux types de données gérées en deux répertoires distincts :

- un répertoire "*identification*", qui regroupe les fichiers d'intérêt commun à tous les utilisateurs de données relatives aux stations ;
- un répertoire "*données*" contenant les fichiers de données propres à une station, une étude, une région, une période, etc..

L'utilisateur peut être amené à constituer plusieurs répertoires d'un même type pour des raisons de commodité ou de sécurité. Ainsi, une organisation possible pour un service gérant des données de pluie peut se concevoir de la manière suivante :

- un répertoire R0 contient les modules du logiciel PLUVIOM ;
- un répertoire R1 contient les identifications des stations ;
- un répertoire R2 capitalise les données pluviométriques du service depuis l'ouverture des stations ;
- un répertoire R3 regroupe les données collectées au cours d'une année.

Les accès à R2 sont exceptionnels et réservés aux opérations de mise à jour de la banque locale (par exemple par les données du répertoire R3, avec une périodicité de mise à jour au choix du gestionnaire) et d'extraction de données (en direction de services extérieurs ou pour des besoins internes de fonctionnement).

3. DESCRIPTION DU LOGICIEL

3.1. Les fonctions disponibles (schéma n°2)

PLUVIOM dispose de trois fonctions.

3.1.1. Mise à jour

La mise à jour des données gérées par PLUVIOM peut être réalisée de différentes manières, selon le type de données et le support de stockage de l'information.

- Traitement de documents écrits (bordereaux d'observateur, rapport de tournée, etc.). La saisie (ou correction) s'effectue directement dans le fichier de la donnée correspondante à partir du clavier.

- Transformation de données stockées sur support magnétique. Sous réserve qu'elles soient dans un format reconnu, des données externes peuvent être transformées aux fins de mise à jour de la banque gérée par PLUVIOM. L'utilisateur précise alors les séquences de station-année devant faire l'objet de la transformation.

- Dépouillement automatique d'enregistrements graphiques. Par l'intermédiaire d'un numériseur, l'utilisateur peut traiter les enregistrements obtenus sur les appareils pluviographiques. De nombreuses options sont paramétrées afin de permettre le dépouillement d'enregistrements très divers, que ce soit par le type d'enregistrement (diagrammes, bandes continues), par la périodicité des relevés (journalier, hebdomadaire, etc.), par le système de mesure (augets, siphon, pesée, etc.), par le mode de graduation des supports (équitemps verticaux, curvilignes) ou par le déroulement (vers la droite ou la gauche).

En raison de la profusion des modèles disponibles et pour ne pas imposer aux utilisateurs tel(s) modèle(s) compatible(s), PLUVIOM est adaptable à de nombreux (tous ?) numériseurs. Pour cela, l'utilisateur doit préciser au logiciel les paramètres de communication à prendre en compte et la structure du message émis par le numériseur en fonction des caractéristiques de son matériel.

- Traitement de cartouche à mémoire morte EPROM. Certains pluviographes sont équipés de centrales (fabriqués par la société ELSYDE) assurant le codage des données pluviographiques et leur stockage sur des cartouches à mémoire morte de grande capacité (64000 octets) sous forme de date de basculement des augets. PLUVIOM peut traiter les cartouches provenant des diverses générations de centrales actuellement disponibles, par l'intermédiaire d'un lecteur spécialisé, et permet une mise à jour des données après divers contrôles.

3.1.2. Edition

Ce terme recouvre l'impression d'inventaires de données disponibles dans les fichiers, d'annuaires ou de tableaux de données pour des séquences de station-année définies par l'utilisateur. De même, la liste des stations pluviométriques répondant à des critères particuliers (localisation géographique, date d'installation, etc.), peut être éditée.

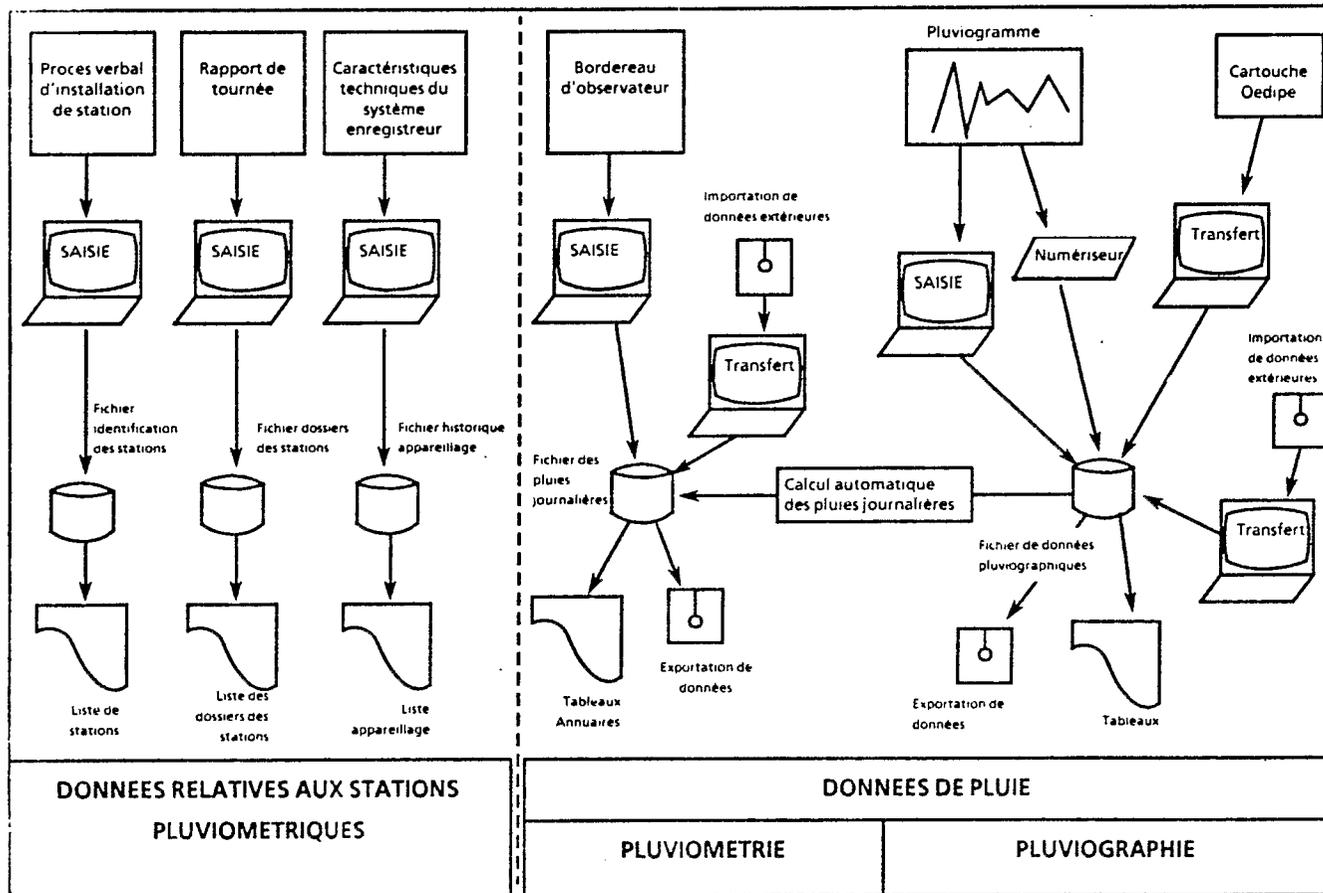


Schéma n°2 : Principales fonctions du Logiciel PLUVIOM

3.1.3. Mise à disposition

En raison de leur format de stockage et du mode de gestion des fichiers, les données sont assez difficilement accessibles par l'utilisateur qui veut procéder à des traitements particuliers. Des extractions de tout ou partie des divers fichiers "identification" ou "données" sont prévues. Elles permettent à l'utilisateur de constituer des fichiers dans lesquels les données sont stockées en clair (code ASCII) dans des fichiers séquentiels.

3.2. Les accès aux fonctions et aux répertoires de données

Les 41 fonctions de PLUVIOM sont accessibles à partir d'un menu principal (figure A) et de douze menus secondaires affectés au traitement des données d'un même type (figure B) ou aux fonctions utilitaires d'un même type pour les diverses données (figure C). La sélection, l'exécution d'une fonction, le retour au menu à partir duquel la fonction a été requise sont entièrement gérés par PLUVIOM.

À l'entrée dans le logiciel, l'utilisateur définit les répertoires contenant les données sur lesquelles il désire travailler durant toute la session.

L'utilisation de PLUVIOM ne nécessite donc aucune connaissance particulière relative aux langages de programmation utilisés. Seules des notions élémentaires sur le système d'exploitation (DOS) des micro-ordinateurs sur lesquels PLUVIOM est installé sont utiles, notamment en ce qui concerne l'organisation arborescente en répertoire des supports de données.

4. DESCRIPTION DE L'ENVIRONNEMENT

L'utilisateur est assisté à divers niveaux au cours d'un travail effectué avec PLUVIOM.

4.1. Contrôles de saisie

Toute saisie manuelle est soumise à un contrôle, ceci dans le but d'éviter les traitements à partir de paramètres incorrects. Ainsi, toutes les données sont soumises à un test de validité, interdisant l'introduction de valeurs aberrantes dans les fichiers (coordonnées, altitude, etc.) ou signalant à l'opérateur des valeurs très fortes (intensité, hauteur journalière, etc.).

Les séquences année-station pour lesquelles l'utilisateur désire effectuer une tâche (extraction, édition, etc.) sont contrôlées (le chevauchement des demandes, par exemple) pour interdire des répétitions dans les fichiers extraits ou les éditions. Toute anomalie est signalée par un message apparaissant à l'écran et précisant à l'opérateur le type d'erreur commise.

4.2. Aide à la saisie

Le recours à des touches programmées pour effectuer certaines tâches simples, comme la duplication ou l'effacement de champ, ou plus complexes, tel que l'affichage d'enregistrement suivant ou précédent le dernier enregistrement vi-

Choisissez le fichier auquel vous voulez accéder ou les fonctions à effectuer par les touches 1 ou ↓ puis validez votre choix par la touche Retour-Chariot

1. Fichier identification des PAYS
2. Fichier identification des STATIONS
3. Fichiers des PLUIES JOURNALIERES
4. Fichier des PLUIES MENSUELLES
5. Fichier des DOSSIERS DE STATIONS
6. Fonctions pour le traitement de la PLUVIOGRAPHIE
7. Fonctions UTILITAIRES

Touches de fonction
SF1 : retour au DOS

Figure A. Menu principal de PLUVIOM.

Choisissez la procédure en déplaçant la flèche par ↑ ou ↓ et validez votre choix par la touche retour-chariot

1. Configurer le numériseur
2. Gérer l'historique de l'appareillage des stations pluviographiques
3. Saisir ou corriger des données pluviographiques
4. Traiter les cartouches à mémoire morte OEDIPE de ELSYDE
5. Inventorier des données pluviographiques
6. Editer des données pluviographiques

Aide aux touches de fonction
F1 : retour au menu précédent
SF1 : retour au DOS

OK (O / N) : .

Figure B. Menu secondaire (traitements des données pluviographiques).

Choisissez la procédure en déplaçant la flèche par ↑ ou ↓ et validez votre choix par la touche retour-chariot

1. Transformer les identifications de STATIONS provenant d'une autre configuration
2. Transformer des PLUIES JOURNALIERES provenant d'une autre configuration
3. Transformer l'HISTORIQUE APPAREILLAGE provenant d'une autre configuration
4. Transformer des DONNEES PLUVIOGRAPHIQUES provenant d'une autre configuration

Aides aux touches de fonction
F1 : retour au menu précédent
SF1 : retour au DOS

OK ? (O / N) : .

Figure C. Menu secondaire (utilitaires de transformation de données extérieures en données PLUVIOM).

sualisé, est généralisé. La signification de l'ensemble des touches utilisables est accessible à l'utilisateur par lecture directe à l'écran chaque fois que cela a été possible ; sinon l'utilisateur peut faire défiler sur l'écran la signification de chaque touche.

4.3. Documentation

Le manuel d'utilisation rassemble toutes les informations nécessaires à l'utilisateur :

- les diverses fonctions et la façon d'y accéder par l'intermédiaire des menus proposés ;
- les opérations à effectuer pour exécuter un traitement ;
- les rappels des touches de fonction utilisables pour une fonction particulière ;
- les messages d'erreur, leur signification et les explications éventuelles sur la façon de corriger l'erreur commise ;
- la description des fichiers de données extraites pour une utilisation en dehors de PLUVIOM.

4.4. Journal

Chaque opération effectuée sur les données gérées par PLUVIOM est consignée dans un fichier dit "journal" comportant la date et l'heure de réalisation de l'opération.

En sortie du logiciel, il est proposé à l'utilisateur de procéder à une édition du journal, et le cas échéant, à sa destruction.

Le but principal de ce journal est de permettre un contrôle des accès effectués dans le fichier de données afin de déterminer à quel moment une altération, éventuellement détectée, s'est produite. Il est alors aisé de reconstituer un fichier à partir d'une sauvegarde antérieure, puis en refaisant les mises à jour effectuées depuis la sauvegarde.

5. CONCLUSION

Dans sa version actuelle, PLUVIOM répond aux besoins des gestionnaires de données pluviométriques : acquisition, contrôle, échange, cession. Il fait l'objet d'une diffusion auprès des unités ou centres ORSTOM dans le but de maintenir la banque de données pluviométriques.

L'expérience acquise au Laboratoire d'Hydrologie par son utilisation en vraie grandeur conduit à prévoir une évolution de PLUVIOM.

Le recours à des logiciels d'aide à la programmation plus répandus et de plus en plus performants permettrait une amélioration sur la forme. Ainsi, l'utilisateur accéderait aux fonctions du logiciel de façon plus attrayante et, surtout, plus rapide (menus déroulants, utilisation possible de la souris, etc.). De plus, il pourrait disposer d'une documentation intégrée (aide en ligne) apparaissant sur demande expresse et associée à chaque fonction, chaque champ des grilles de saisie.

Ces améliorations, qui accroîtraient le confort de l'utilisateur, resteraient toutefois secondaires par rapport à celles qui pourraient être effectuées sur le fond.

L'utilisation d'un langage unique, performant à tous les niveaux (traitements numériques, communications avec les périphériques, gestion de fichiers) et disposant d'un environnement riche (bibliothèque de programmes variés) conduirait à un logiciel homogène dont la maintenance et l'extension seraient facilitées.

Le recours à un outil multilingage de gestion de fichiers permettrait d'augmenter les possibilités, déjà offertes par PLUVIOM, d'utiliser ou extraire les données gérées.

Ces améliorations, qui intégreraient les diverses remarques et requêtes des utilisateurs de la version actuelle, passeraient par une refonte complète du logiciel mais pourraient être un moyen d'assurer sa pérennité.

LA BANQUE DE DONNEES HYDROPLUVIOMETRIQUES DU LABORATOIRE D'HYDROLOGIE DE L'ORSTOM

(Gilbert JACCON & Michel TRAVAGLIO)

RESUME - L'information hydrométrique et pluviométrique, rassemblée par les hydrologues de l'ORSTOM depuis plus de 40 ans, constitue une masse de plus de 100 000 années de données répartie entre 4 400 stations d'observations.

Deux banques de données indépendantes, d'importance comparable et de structure similaire, ont été progressivement constituées à partir de 1967. Leur gestion, confiée au Laboratoire d'Hydrologie, est assurée par les logiciels HYDROM et PLUVIOM.

Cette communication a pour objet de présenter ce système d'information dans son état actuel. Les points suivants sont abordés :

- organisation générale ;
- données (nature, présentation et saisie) ;
- inventaire (répartition chronologique et géographique de l'information) ;
- actualisation des fichiers ;
- cession des données ;
- moyens.

INTRODUCTION

Les données hydrométriques et pluviométriques rassemblées par les hydrologues de l'ORSTOM, soit directement par les observations des réseaux installés sur les cours d'eau d'Afrique francophone à partir de 1948 ou dans le cadre d'études de petits bassins versants dès 1952, soit indirectement par la récupération de données provenant d'autres sources (par exemple des gestionnaires des réseaux pluviométriques nationaux) constituent une masse d'information considérable.

C'est en 1967 que le Bureau Central d'Hydrologie, basé à Paris, a décidé d'informatiser la gestion de ces données. Un long et fastidieux travail de saisie sur cartes perforées a été entrepris et poursuivi durant plus d'une décennie. Parallèlement, des programmes FORTRAN ont été développés par les hydrologues eux-mêmes afin d'exploiter cette information sur les ordinateurs du CNRS (CIRCE).

Plusieurs étapes successives (passage des cartes perforées aux bandes puis disques magnétiques, programmes de gestion en COBOL, utilisation de terminaux, développement progressif de logiciels de gestion) ont conduit à la création

de deux banques de données, l'une pour l'hydrométrie, l'autre pour la pluviométrie, indépendantes dans leur organisation et leur gestion mais toutes deux placées sous la responsabilité du Laboratoire d'Hydrologie, désormais basé au Centre ORSTOM de Montpellier.

Cette communication laisse volontairement de côté les descriptions techniques sur l'organisation interne et les outils de gestion de ces banques, qui sont respectivement HYDROM et PLUVIOM. Le premier a fait l'objet d'une présentation très détaillée par son auteur G. Cochonneau, lors de SEMINFOR 1 (Cochonneau 1987). Le second sera présenté dans ce SEMINFOR 3 par l'un de ses auteurs P. Raous, par une communication et une démonstration.

1. ORGANISATION GENERALE

Elle est schématisée sur la figure 1. Les deux banques, qui fonctionnent de manière centralisée, sont actuellement installées sur les ordinateurs IBM 3090 du Centre National Universitaire Sud de Calcul (CNUSC) de Montpellier, auquel le Centre ORSTOM est relié par une ligne spécialisée.

La gestion et l'exploitation des banques sont assurées, avec les logiciels HYDROM et PLUVIOM dans leurs versions "gros ordinateur", par l'intermédiaire d'un terminal IBM 3278 (consultation des fichiers et extraction de données) et de deux micro-ordinateurs de type IBM-PC pour les échanges.

La consultation des fichiers par terminal est possible par l'intermédiaire des autres centres de calcul en réseau avec le CNUSC : cette procédure est actuellement fonctionnelle pour les hydrologues de l'ORSTOM et de l'EDF basés à Paris, via le CIRCE.

L'existence de versions opérationnelles sur micro-ordinateur des logiciels HYDROM et PLUVIOM, permet une gestion partiellement décentralisée des banques. Le développement de ces versions s'est avéré d'un très grand intérêt logistique car il a permis :

- la saisie des données sur place par les gestionnaires de réseaux ou par les hydrologues dans les différents Centres ou Missions de l'ORSTOM ;
- des échanges d'informations très faciles, soit par disquettes (procédure la plus courante actuellement), soit par la messagerie ORSTOM ou autre réseau de type TRANSPAC ;
- la création de banques nationales (Services Hydrologiques Nationaux) ou régionales (Organismes interétatiques), indépendantes mais associées. Les Etats assument la gestion de leurs réseaux de collecte et sont seuls propriétaires de leurs données, mais la compatibilité des formats et des logiciels rend possible une coopération étroite (copie de sécurité des fichiers, appui informatique pour la maintenance et l'actualisation des logiciels, formation des personnels).

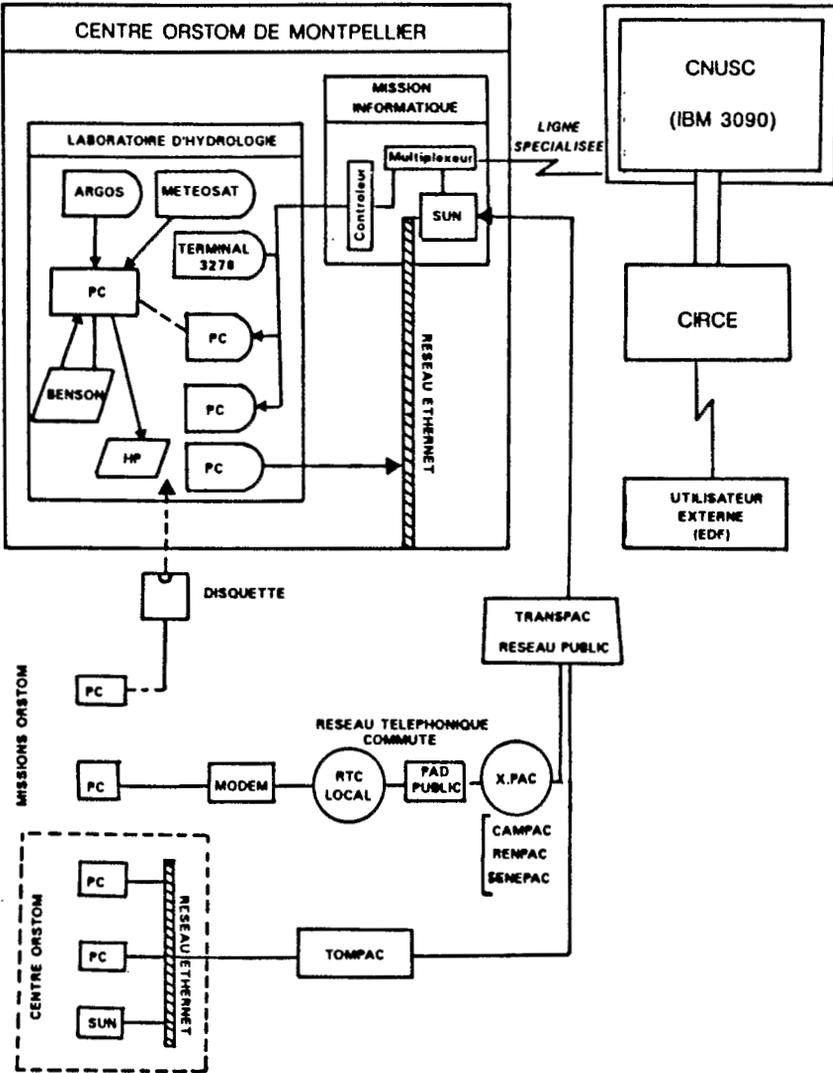


Fig. 1 BANQUE DE DONNEES HYDROPLUVIOMETRIQUES
ORGANISATION GENERALE

2. DONNEES

2.1. Type de données

L'information originale se répartit en deux grandes catégories : les données pluviométriques et les données hydrométriques.

Dans la première, l'information originale est simple : c'est une hauteur de pluie observée durant un intervalle de temps fixe, par exemple 24 h pour les précipitations journalières, ou variable pour les intensités de pluie enregistrées par des pluviographes. La donnée est donc un nombre toujours positif compris entre 0 et 150 000, sachant que l'unité utilisée est le dixième de mm et que la durée maximale de cumul est en pratique d'une année.

L'information hydrométrique est, au contraire, multiple. Elle comprend des hauteurs d'eau lues sur des échelles limnimétriques ou enregistrées par des limni-graphes, des mesures de débit, instantanées (ou considérées comme telles) identifiées par un couple "*débit mesuré - hauteur associée*", les débits instantanés ou moyens sur une période, très rarement mesurés directement, en général calculés à partir des hauteurs d'eau par l'intermédiaire d'un barème de tarage qui est aussi une donnée de la banque.

A chacune des deux banques, pluviométrique et hydrométrique, est associé un fichier des stations, nécessaire pour l'identification et la caractérisation du point de mesure. Les éléments stockés dans ces fichiers sont des informations signalétiques diverses : nom de la station, situation géographique (coordonnées, altitude), équipement, date de mise en service et bref historique du fonctionnement, etc..

2.2. Présentation et Saisie des données originales

Les données originales se présentent sous plusieurs formes différentes, suivant leur nature et le support sur lequel elles se trouvent au moment de la saisie.

On distingue :

- le bulletin ou la fiche mensuelle, original rempli par un observateur une ou plusieurs fois par jour à heures fixes. C'est la source d'information la plus fréquente encore aujourd'hui, autant pour les relevés pluviométriques que pour les hauteurs limnimétriques. La saisie de ces données constitue un travail fastidieux et est une source d'erreurs nombreuses, si une vérification systématique n'est pas faite ;

- l'enregistrement graphique continu, ou diagramme journalier, hebdomadaire, mensuel, obtenu avec des appareils enregistreurs à tambour ou à table déroulante. Le dépouillement des diagrammes est désormais effectué avec une table à numériser, opération longue et délicate. Les hydrologues de l'ORSTOM travaillent encore à la mise au point de logiciels performants permettant une saisie exhaustive de l'information contenue dans le diagramme et sans procédures manuelles intermédiaires ;

- les fiches de jaugeages, bordereaux originaux établis sur le terrain par les hydrométristes, qui doivent être dépouillés avant ou au moment de la saisie ;

- les données mémorisées sur mémoires magnétiques par les pluviographes du type OEDIPE ou les centrales hydrologiques du type CHLOE ;
- les données télétransmises par les stations équipées de balises émettrices ARGOS ou METEOSAT, et enregistrées par des stations de réception.

Pour ces deux dernières catégories, l'information est facilement accessible. Néanmoins un traitement préalable est absolument nécessaire : les données doivent être filtrées, reformatées, quelquefois compactées.

3. INVENTAIRE DES DONNEES

La masse d'information actuellement stockée sur disques et bandes magnétiques au CNUSC est de l'ordre de 400 mégaoctets (400 000 000 octets), avec 160 Mo pour la banque pluviométrique et 240 Mo pour la banque hydrométrique.

Cette information représente :

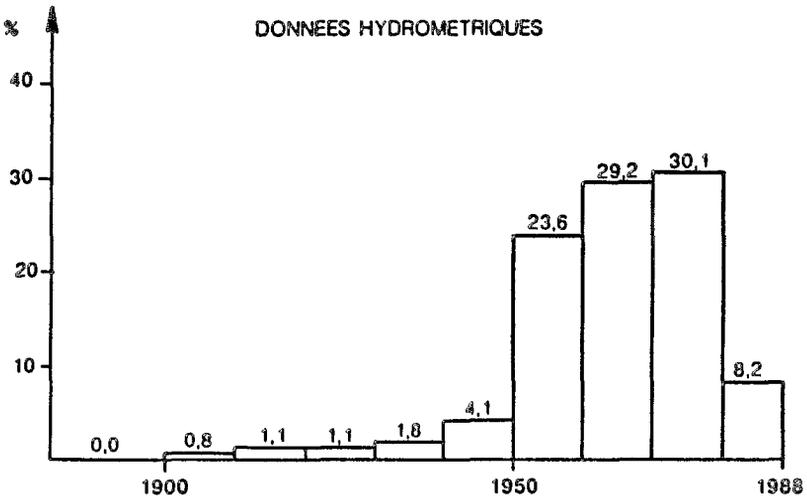
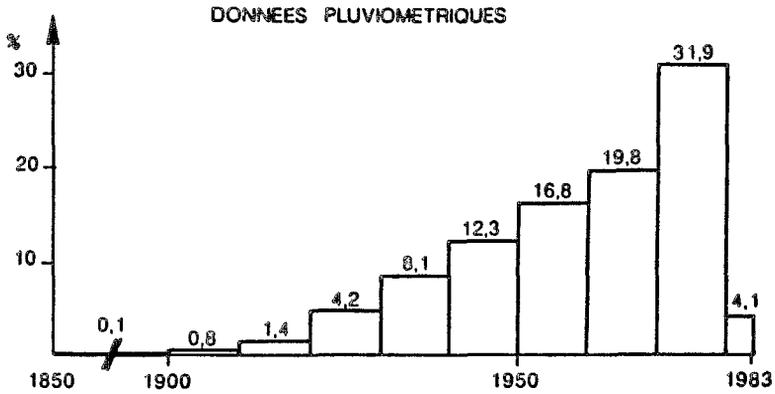
- plus de 57 000 années de données pluviométriques, réparties entre 2811 postes de mesure. La durée moyenne d'observation est d'un peu plus de 20 ans par station, et les relevés les plus anciens datent du milieu du siècle dernier (1852) ;

- plus de 21 300 années de hauteurs limnimétriques observées à 1624 stations. La durée moyenne est supérieure à 13 ans, et les données les plus anciennes datent du début de ce siècle (1902). Rappelons que deux tiers des hauteurs sont traduites en débits instantanés et que la banque hydrométrique contient aussi des fichiers de débits moyens journaliers et des fichiers de débits dits "opérationnels", pour lesquels les données ont été complétées et homogénéisées lors de la préparation de monographies hydrologiques.

La répartition chronologique des données est représentée sur les deux diagrammes de la figure 2, pour lesquels la hauteur de chaque rectangle correspond au pourcentage d'années situées dans la décennie correspondante, et la répartition géographique est illustrée par les cartes schématiques des figures 3 et 4, où sont portés le nombre d'années et le nombre de stations de chaque Etat représenté.

L'analyse de ces figures est très parlante.

Pour la répartition géographique, on constate une grande similitude entre les deux banques : l'information provient à 80% au moins des Etats d'Afrique francophone, secteur d'intervention prioritaire de l'ORSTOM. Le reste vient des DOM-TOM et de certains pays comme Madagascar ou le Cap Vert, où des études ont nécessité la récupération et le traitement des données. L'absence de données pour certains pays ne signifie pas inexistence de données. Par exemple, dans le cas de la Nouvelle-Calédonie, un simple problème de différence entre les formats des fichiers pluviométriques, explique que ces données n'ont pas encore été récupérées.



**Fig. 2 BANQUE DE DONNEES HYDROPLUVIOMETRIQUES
REPARTITION CHRONOLOGIQUE DE L'INFORMATION**

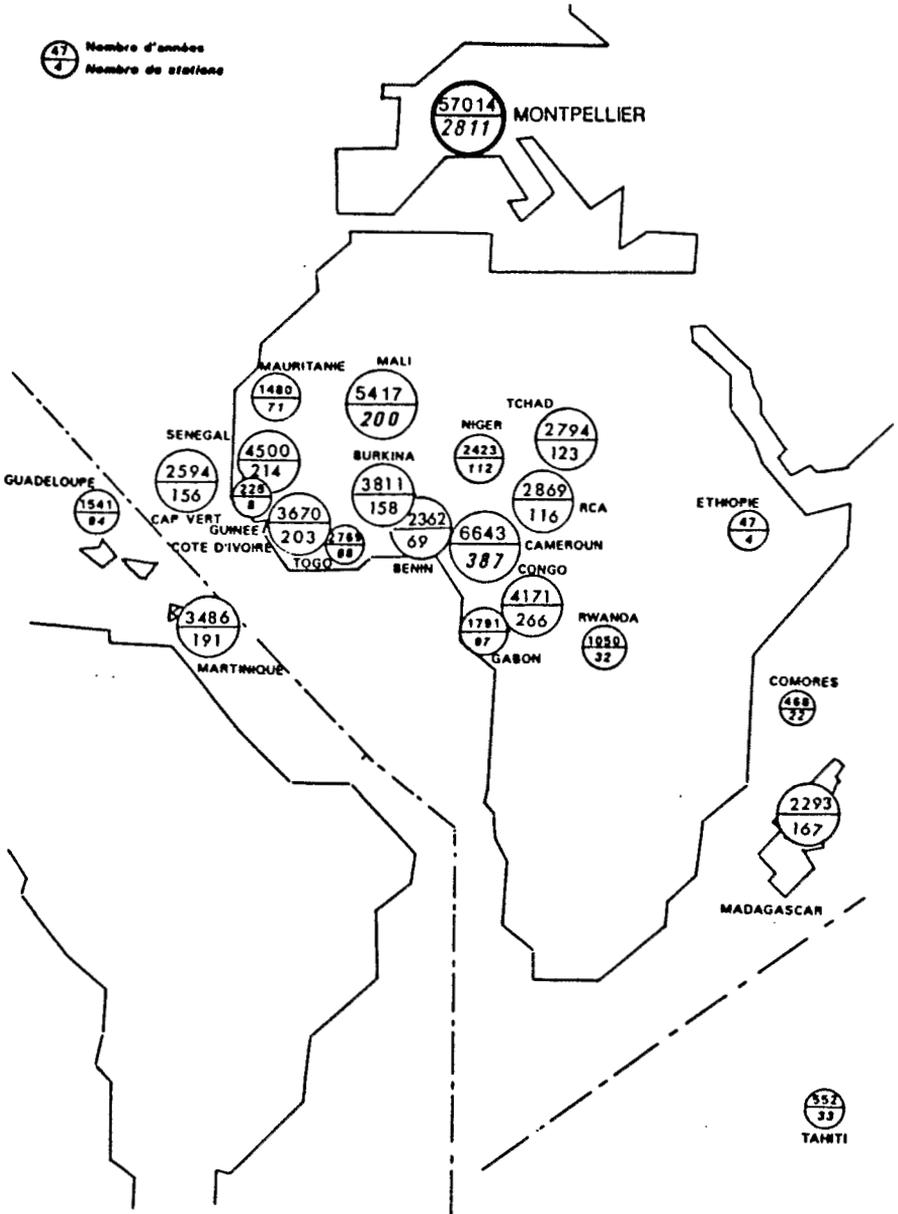


Fig. 3 DONNEES PLUVIOMETRIQUES REPARTITION GEOGRAPHIQUE

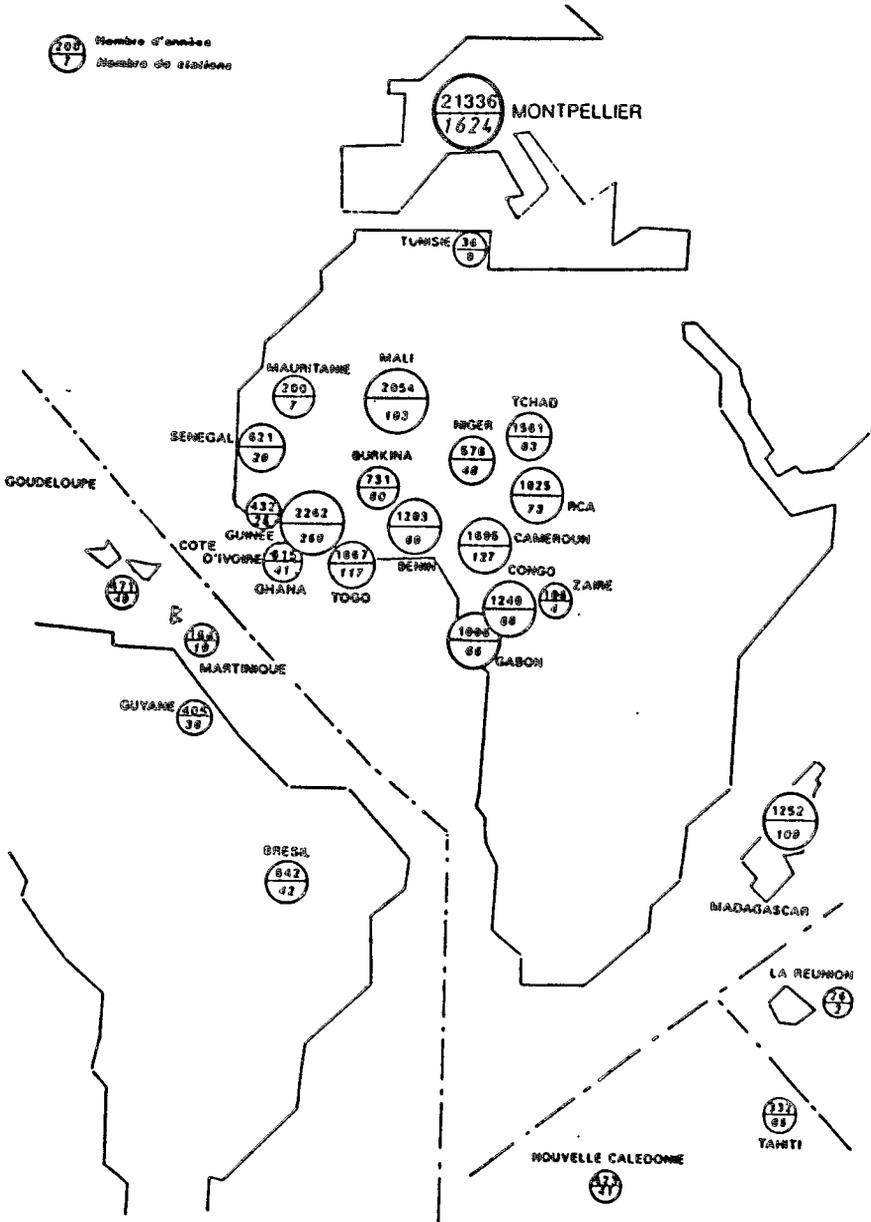


Fig. 4 DONNEES HYDROMETRIQUES REPARTITION GEOGRAPHIQUE

Pour la répartition chronologique, on note l'allure régulièrement croissante du diagramme de pluviométrie, qui s'oppose à la forme concentrée sur la période 1950-1980 du diagramme des données hydrométriques. Ce fait est dû à la forte activité des hydrologues de l'ORSTOM, durant cette période, dans le secteur de la gestion des réseaux et des études de bassins représentatifs.

4. ACTUALISATION DES FICHIERS

La disponibilité d'une banque performante est un objectif pour tout service ou organisme gestionnaire d'un réseau d'acquisition et de traitement de données naturelles. Il est fondamental que cette banque soit régulièrement actualisée.

La création de la banque hydropluviométrique de l'ORSTOM s'est avérée nécessaire pour la mise en forme d'une masse de données (déjà existantes) considérable et rapidement croissante, et pour le traitement d'une information de base, fondement de tous les travaux de synthèse régionale.

Jusqu'au début de la présente décennie, l'ORSTOM produisait ses propres données essentiellement en hydrométrie. Le problème est que l'ORSTOM ne gère plus de réseaux, sauf dans quelques cas isolés, et que les postes d'observation se trouvent, pour plus de 90%, à l'étranger.

L'actualisation de nos fichiers est donc aujourd'hui directement liée aux opportunités qui se présentent, dans le cadre de conventions bilatérales avec les gestionnaires de réseaux ou de contrats d'études nécessitant une mise à jour des séries existantes. La faible densité de données pour la période postérieure à 1979 montre que l'actualisation est loin de présenter un caractère systématique.

En Afrique francophone, la mémorisation des données a été placée sous la responsabilité de l'AGRHYMET (Centre de Formation en AGRonomie, HYdrologie et METéorologie de Niamey), pour l'hydrométrie et de l'ASECNA (Agence pour la SECurité de la Navigation Aérienne de Dakar), pour la pluviométrie.

C'est dans le cadre d'une convention tripartite entre le CIEH (Comité Inter-africain d'Etudes Hydrauliques), l'ASECNA et l'ORSTOM que la banque pluviométrique des 13 Etats membres du CIEH (Benin, Burkina Faso, Cameroun, Congo, Côte d'Ivoire, Gabon, Mali, Mauritanie, Niger, RCA, Sénégal, Tchad et Togo) a été actualisée pour la période 1966-1980. Cette convention, dont les travaux seront achevés en fin d'année, a été financée par le FAC (Fond d'Aide et de Coopération) : les travaux, menés par une équipe composée d'un ingénieur et d'un assistant, a duré quatre années. Même en considérant l'importance et la qualité des résultats (fichiers complets vérifiés et critiqués, annuaires pluviométriques édités pour chacun des pays concernés), cet exemple montre l'ampleur d'un tel travail !

Quelle que soit l'importance que l'on accorde à l'actualisation des fichiers, et les opportunités qui se présentent, cette opération ne peut être faite à n'importe quel prix. Par exemple, la saisie directe de fiches d'observateurs ou de diagrammes implique la constitution d'équipes spécialisées pour ces travaux, aussi exigeants que fastidieux. Ceci a été fait entre 1970 et 1980 pour la

constitution des fichiers hydrométriques à partir des archives du Bureau Central d'Hydrologie. Ce n'est plus le cas maintenant, et l'équipe chargée de la maintenance des fichiers au Laboratoire d'Hydrologie n'a plus les moyens de mener des opérations de saisie, sinon pour des volumes de données très limités afin de compléter des séries.

La démarche actuelle du Laboratoire est de conduire une politique de récupération des données déjà saisies (disquettes ou bandes magnétiques, cartouches EPROM ou données télétransmises). Pour cela, une coopération est systématiquement recherchée avec nos anciens partenaires pour la création de banques hydrologiques locales, à l'aide des logiciels HYDROM et PLUVIOM. Cette politique doit permettre de faciliter les échanges de données et de rendre possible des mises à jour périodiques, en particulier des séries les plus importantes par leur amplitude et leur représentativité régionale.

5. CESSION DES DONNEES

Les seules données cessibles sont les données en l'état, c'est à dire les données originales non critiquées, qui représentent environ 80% du total pour la banque hydrométrique.

La cession de données à des organismes autres que les Services compétents des pays concernés, ne peut être faite qu'avec l'accord de ces derniers.

La seule exception concerne la cession à des étudiants ou Laboratoires universitaires, de données nécessaires pour des travaux de recherches ou de thèses. Ces demandes portent toujours sur de faibles quantités (quelques dizaines d'années).

La fréquence moyenne des demandes de cession, reçues par le Laboratoire d'Hydrologie et d'origine extérieure à l'ORSTOM, est de l'ordre de deux par mois. Ces demandes proviennent pour moitié d'étudiants ou chercheurs, pour moitié de Bureaux d'Etudes français et étrangers. Ces demandes sont toujours satisfaites dans le cadre des limites indiquées ci-dessus et une simple compensation des frais informatiques engagés est exigée.

6. MOYENS

La gestion et l'exploitation des deux banques sont faites par une équipe spécialisée, formée d'un ingénieur et de deux techniciens, alors que le développement et la maintenance des logiciels de gestion HYDROM et PLUVIOM sont assurés par une autre équipe composée d'un ingénieur et de deux informaticiens.

Le coût global de la gestion est de l'ordre de 100 000 FF par an, hors personnel. Cette somme comprend pour moitié les frais d'utilisation du CNUSC et pour l'autre moitié les frais d'amortissement du matériel du Laboratoire.

CONCLUSION

La banque de données hydropluviométriques du Laboratoire d'Hydrologie constitue un système d'information essentiel pour la communauté hydrologique amenée à travailler dans la zone inter-tropicale, car il n'a pas d'équivalent.

Cet outil résulte d'une somme d'efforts considérables, tant humains que financiers. Efforts permanents et sans cesse renouvelés mais justifiés par les résultats obtenus.

REFERENCES BIBLIOGRAPHIQUES

- COCHONNEAU (G) 1987 - *HYDROM, un Logiciel de Gestion des Données d'un Réseau d'Observation Hydrométrique*. In: Séminfor 1, ORSTOM, Paris, 1988.
- RAOUS (P) 1989 - *PLUVIOM : Un Logiciel de Gestion de Données Pluviométriques*. In: Séminfor 3, ORSTOM, Paris, 1990.

SISOND : UNE CONCEPTION DE L'INFORMATIQUE MISE AU SERVICE DE L'USAGER

(Daniel CARDON)

RESUME - SISOND est partie d'un ensemble de deux systèmes créés pour gérer respectivement les données de sondage neutronique et les données biologiques d'une expérimentation menée au Nicaragua. SISOND a été préféré dans cet exposé en raison de sa plus grande généralité mais les deux systèmes présentent une communauté de conception et des facilités identiques. Plus qu'un exposé sur les possibilités de SISOND, ce travail se veut une réflexion sur la conception de l'informatique au service du chercheur. Cette réflexion s'est placée au niveau de l'analyse du système. Elle a abouti à la recherche d'un certain nombre de priorités qui à leur tour ont induit la méthode de réalisation. Le plan de l'exposé suit ce cheminement.

En introduction est expliquée cette conception de l'informatique, en fonction d'elle, dans la première partie, est analysé le problème particulier des mesures neutroniques, dans la seconde on passe à la réalisation pratique, en conclusion le système est examiné de façon critique tant dans sa conception que dans sa réalisation. En s'appuyant sur une description rapide du système, on s'est, dans ce résumé, limité à mentionner les particularités les plus originales de SISOND avec dans chaque cas l'objectif poursuivi, et à exposer brièvement l'insuffisance la plus notoire rencontrée.

Trois éléments interviennent dans les mesures de sonde à neutrons :

- la sonde que l'on ramène à une sonde standard par des calibrages réguliers ;
- le site qui, à humidité égale, conditionne la réponse d'une sonde standard, et exige donc un étalonnage ;
- l'humidité volumique qui est le paramètre recherché.

Ces trois éléments se retrouvent dans SISOND mais seul le traitement réservé aux mesures d'humidité est considéré ici. SISOND prend en charge les mesures d'humidité avec la sonde de la collecte à l'exploitation. Une unité permet la description une fois pour toutes de chaque expérimentation, une seconde la création pour chaque collecte de données d'un registre dans le fichier adéquat et l'impression des feuilles de collecte correspondantes, et une troisième l'informatisation ou la modification des données de terrain (la critique étant effectuée à l'aide d'une unité graphique). Une dernière unité permet alors de créer, pour chaque tube étudié, un fichier d'humidités volumiques à partir des données critiquées. L'exploitation ultérieure de ces données se fait ensuite sous LOTUS (un logiciel du commerce).

Une qualité indispensable à tout travail scientifique est la rigueur, possible seulement avec des données saines. A tous les niveaux SISONND favorise cette rigueur. Le programme de critique graphique est rapide d'emploi et performant mais seule les données absurdes ont le droit d'être supprimées. La possibilité d'associer un code à chaque donnée permet de donner son opinion sans pour autant la supprimer du fichier de données brutes. Un bon programme de critique permet certes de discriminer les erreurs les plus notoires mais il est tellement plus sain de supprimer les causes d'erreur elles-mêmes. C'est pourquoi les feuilles de collecte créées par SISONND ont été voulues claires et agréables à remplir, c'est pourquoi au maximum on a voulu faire de SISONND un système "convivial". Les écrans à documenter sont explicatifs, les réponses aux questions, autant que faire se peut, sans ambiguïté (une croix en face d'une option par exemple), et dans le cas de l'informatisation de données l'écran est l'image même de la feuille de collecte. Ce faisant on peut confier l'informatisation des données à l'observateur de terrain dont le travail se trouve valorisé tout en déchargeant le responsable de l'expérimentation.

Un travail scientifique comporte toujours une certaine part d'inconnu et avant de l'effectuer complètement il est bien difficile de prévoir tout ses développements possibles. Pour cette raison, au contraire des fichiers de données brutes qui sont à accès direct, les fichiers d'humidités volumiques sont séquentiels et en caractères ASCII. Ceci permet leur importation dans des logiciels du commerce, leur utilisation aisée avec tout type de langage informatique et même leur modification dans un traitement de texte.

Une autre qualité remarquable de SISONND est sa sécurité d'emploi. Un système de *BACKUP* obligatoire et *RESTAURATION* éventuelle de tous les fichiers de données brutes modifiés le garantit contre les incidents de fonctionnement. D'autre part les opérations "*dangereuses*" comme création ou suppression de fichiers ou de registres, ne peuvent être effectuées que par les utilisateurs connaissant le mot de passe.

Un facteur important n'a cependant pas été pris suffisamment en considération : celui de la maintenance. On a considéré que celle-ci se limitait à la correction des erreurs résiduelles et à la complémentation du système. En fait, si on veut donner à un travail informatique une certaine pérennité, il faut qu'il puisse bénéficier aux moindres frais des dernières améliorations dans le langage utilisé et des nouveaux outils mis à disposition. SISONND a été écrit en langue espagnole avec la version 3.0 de TURBO-PASCAL et, par conception, nécessitera une refonte importante pour pouvoir bénéficier des avantages de l'environnement lié à la version 4.0 de TURBO-PASCAL. C'est la raison pour laquelle cette nouvelle version prévue en Français ne pourra être présentée ici.

INTRODUCTION

Originellement ce travail a été réalisé pour répondre à un besoin particulier : assurer la gestion des mesures neutroniques d'une expérimentation réalisée dans le cadre de la convention ORSTOM-NICARAGUA. Son ampleur toutefois dépasse largement le problème abordé et la solution proposée essaye de répondre aux besoins de la plupart des utilisateurs de sonde à neutrons. En fait la réalisation de SISOND a été un excellent prétexte pour mettre en pratique une conception de l'utilisation de l'informatique en recherche scientifique. Cette conception existait "en filigrane" dans un travail présenté antérieurement : SISCLIMA (Cardon et al. 1987), mais n'avait pas été dégagée clairement car elle s'était imposée en cours de réalisation. Avec SISOND, au contraire, les "règles du jeu" ont été posées avant.

Tout d'abord qu'est-ce que l'informatique pour le scientifique ? "un instrument de travail, rien de plus qu'un instrument" peut-on répondre. C'est exact, mais cet instrument est d'une puissance exceptionnelle. Par la forme donnée à l'entrée des données, on peut imposer le protocole de mesures. Quant au programme, il domine toute la méthodologie de leur traitement. Alors, pour qui a tâté un peu d'informatique, la tentation est grande, même involontairement, d'imposer ses idées en réalisant une sorte de boîte noire recevant à l'entrée les données de terrain et fournissant à la sortie un travail pratiquement achevé.

Cette façon d'opérer ne nous paraît pas la plus judicieuse. Une secrétaire utilisant un traitement de texte peut apprécier d'être guidée pas à pas. Au contraire, en recherche scientifique, toute contrainte est ennemie de la créativité. D'autre part, au niveau du traitement, le scientifique sait ce qu'il cherche mais pas toujours ce qu'il va trouver. Un système informatique à l'usage de scientifiques devra donc respecter autant que faire se peut la liberté de l'utilisateur, tant au niveau du choix des données que de celui de leur traitement. Enfin, il faut le reconnaître, c'est agréable de programmer son travail et il faut dans une certaine mesure garder à l'utilisateur cette possibilité. La première qualité d'un tel système devra donc être la souplesse d'emploi.

La liberté de l'utilisateur ne saurait cependant être totale. Outre les contraintes techniques, il ne devrait pas être possible de faire n'importe quoi avec n'importe quelles données. La seconde exigence du scientifique à l'égard du système est qu'il doit favoriser les comportements rigoureux. Cette seconde qualité va un peu à l'encontre de la première ce qui, dans certains cas, impose des solutions de compromis.

A ces qualités un peu spécifiques il faut ajouter celles requises de tout logiciel, soit principalement la sécurité et la simplicité, qu'on retrouve dans le confort de l'utilisateur. Mais de quel utilisateur s'agit-il ? Il peut en effet y en avoir plusieurs et leurs points de vue ne sont pas forcément identiques. Ici on en a considéré trois :

- le responsable scientifique qui a imaginé l'expérimentation ;
- le responsable de la collecte qui est la cheville ouvrière de l'expérimentation ;
- l'utilisateur des données

Il s'agit bien sûr d'une distinction un peu artificielle car la même personne peut très bien remplir les trois rôles, mais elle est bien pratique dans l'analyse des problèmes, et nous la retrouverons dans tout ce qui va suivre.

Enfin, sur le plan informatique, aucun travail n'étant parfait, il faut être non seulement prêt à corriger un certain nombre d'erreurs mais même à effectuer des modifications importantes de fond. Ce travail doit éventuellement pouvoir être exécuté par une tierce personne. Deux autres qualités doivent donc également être recherchées la maintenabilité et la lisibilité des programmes.

Le premier chapitre de ce travail a été consacré à l'analyse du problème de l'utilisation de la sonde à neutrons. Dans le second on expose les solutions adoptées pour résoudre ce problème techniquement en conformité avec les exigences exprimées antérieurement. Dans le troisième on montre comment SISONND fonctionne pratiquement.

1. ANALYSE DU PROBLEME

Dans une expérimentation avec sonde à neutrons interviennent trois éléments : l'instrument de mesures, le milieu et le dispositif expérimental. Dans le premier paragraphe on analysera le rôle de chacun d'eux, ensuite on explicitera la conception de l'expérimentation sur laquelle SISONND a été bâti.

1.1. Les éléments intervenant dans une expérimentation neutronique

1.1.1. La sonde à neutrons

La sonde à neutrons permet la mesure non destructive de la quantité d'hydrogène présente dans le sol, soit en pratique la quantité d'eau. Elle se compose principalement d'une source de neutrons rapides, d'un compteur de neutrons thermalisés et d'une électronique associée. La réponse de la sonde à des conditions données dépend de ces trois éléments et de sa géométrie, c'est à dire de la façon dont ils sont assemblés.

Pour tester une sonde on la place dans des conditions de référence et on effectue pendant des périodes de temps déterminées des comptages neutroniques. Ces conditions de référence sont le plus souvent un baril d'eau et parfois la sonde dans sa protection verrouillée.

Par construction la réponse de la sonde dans l'eau est voisine de 1000 ou d'un autre multiple de 10, et, pour exploiter les données, on les ramène généralement à celles d'une sonde idéale comptant 1000 dans l'eau. Cette réponse peut subir des dérives dans le temps, des calibrages périodiques sont donc recommandés et sont indispensables après toute intervention. Comme tout instrument contenant des composants électroniques, la sonde présente une certaine sensibilité aux variations de température. Il est possible de compenser ces variations en utilisant la protection de la sonde comme référence au moment du calibrage, puis sur le terrain au moment des mesures. Cette dernière méthode est controversée à cause de la variabilité des mesures sonde dans la protection et donc peu utilisée.

1.1.2. Le sol

Le sol est considéré ici comme partie fixe ou substrat. La partie variable, celle qu'on mesure, est étudiée avec le dispositif expérimental. Le phénomène de thermalisation des neutrons dans le sol est "perturbé" par un phénomène annexe : l'absorption, variable avec les éléments rencontrés. La plupart joue un rôle tout à fait négligeable, mais certains, même en faible quantité, influent notablement sur la mesure. L'exploitation des résultats d'une expérimentation ne pourra donc se faire qu'après étalonnage du sol.

Deux types d'étalonnage se pratiquent généralement : l'étalonnage empirique et l'étalonnage chimique. Le premier, assez grossier, a l'avantage de pouvoir être réalisé facilement et rapidement. Le second, plus précis, implique une analyse du sol. Une conséquence de cette précision est que l'étalonnage varie avec la géométrie de la sonde. Par conséquent, dans une même expérimentation, à chaque type de sonde doit correspondre un étalonnage.

1.1.3. Le dispositif expérimental

Le dispositif expérimental est composé d'un certain nombre de tubes verticaux destinés à recevoir la sonde (la possibilité de tubes horizontaux n'a pas été retenue dans SISONND). Les mesures sont effectuées à un certain nombre de niveaux qui peut varier d'un tube à l'autre et suivant un nombre de répétitions également variable.

1.2. Conception de l'expérimentation avec sonde à neutrons

Tout d'abord pourquoi se rattacher à une conception particulière alors qu'on voulait donner à SISONND la plus grande généralité possible ?

Un système informatique est une réalisation concrète qui ne peut donc s'appuyer que sur des éléments concrets. Pour garder à SISONND une grande généralité, on a essayé de définir l'expérimentation la plus ample possible de façon à englober la plupart des expérimentations réelles.

Une expérimentation se déroule en trois temps : la phase initiale ou installation, l'expérimentation proprement dite et la phase finale ou destruction du site. Ces trois phases seront successivement examinées, mais auparavant il convient de dire quelques mots des opérations de routine qui se réalisent régulièrement durant toute la conduite de l'expérimentation.

1.2.1. Opérations de routine

Une expérimentation est conduite avec une ou plusieurs sondes à neutrons. La sonde est un instrument délicat. En prévision d'une panne toujours possible, avant tout premier usage ou après toute réparation, il convient, sous peine de perdre tout le travail effectué, de calibrer chaque sonde.

1.2.2. Phase d'installation (voir figure 1)

Au cours de la phase d'installation, on recueille les échantillons de terre prélevés à la tarière lors de la mise en place des tubes et on note avec soin les profondeurs correspondantes. Ensuite on installe les tubes et, immédiatement après, on effectue un ou plusieurs profils avec la sonde.

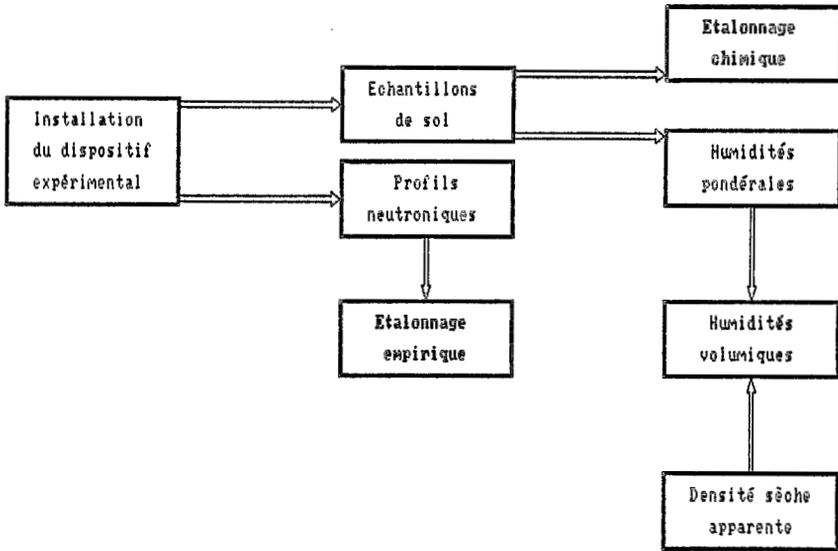


Fig.1 Installation du dispositif expérimental

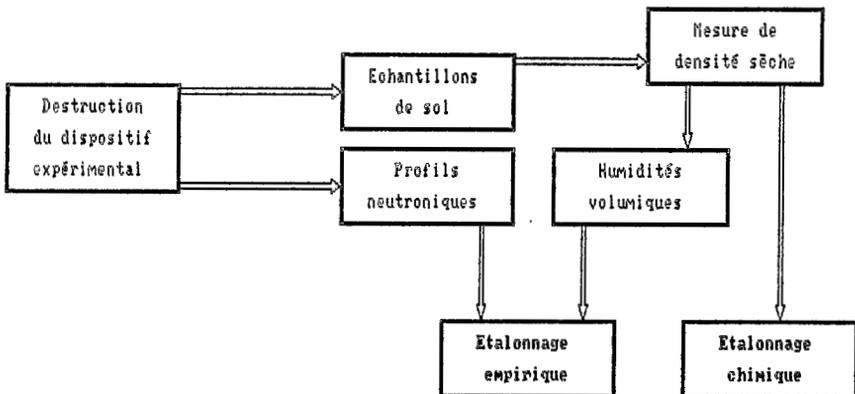


Fig.2 Destruction du dispositif expérimental

Les échantillons de terre pesés, séchés et pesés à nouveau donnent l'humidité pondérale. Par estimation de la densité sèche apparente ou par sa mesure dans un site voisin, il est alors facile de passer aux valeurs d'humidité volumique. La comparaison avec les mesures de sonde fournit un premier étalonnage empirique.

Les échantillons de terre séchés sont ensuite envoyés dans un centre spécialisé, pour être analysés en vue de l'étalonnage chimique de la sonde. Il est à noter que pour utiliser cet étalonnage il faut également disposer des mesures de densité sèche apparente.

1.2.3. *Expérimentation proprement dite*

En routine, et suivant un protocole expérimental invariable, sont effectuées les mesures neutroniques. On les ramène ensuite aux valeurs qui auraient été obtenues avec une sonde comptant 1000 dans l'eau (valeurs normalisées). Pour passer aux humidités volumiques il faut alors disposer de valeurs d'étalonnage. Au début on se sert de l'étalonnage empirique, puis on utilise les valeurs de l'étalonnage chimique.

1.2.4. *Destruction du site (voir figure 2)*

Dans une certaine mesure, c'est une réplique de l'installation mais en inversant l'ordre des opérations et en beaucoup plus précis. On réalise une dernière série de profils neutroniques. Ensuite on profite de la destruction du site pour prélever autour des tubes des échantillons de terre suivant la méthode des cylindres. Ceci permet de mesurer la densité sèche apparente. D'autre part la terre retirée correspond cette fois réellement au volume exploré par la sonde. Le reste des opérations se déroule comme pour l'installation.

2. PRESENTATION DE SISOND EN FONCTION DES OBJECTIFS RECHERCHES

On fera tout d'abord une présentation succincte de SISOND et ensuite on verra les solutions techniques adoptées pour tenter de réaliser les objectifs présentés en introduction.

2.1. Présentation succincte de SISOND

2.1.1. *Les programmes de SISOND*

SISOND, dans sa version actuelle, comprend dix programmes répartis de la façon suivante :

- deux programmes informatiques (BACKUP et installation) ;
- six programmes de gestion de données brutes ;
- deux programmes d'exploitation de données.

Sauf pour le BACKUP, ces programmes nécessitent d'informations pour pouvoir fonctionner. Celles-ci sont communiquées au travers de un ou plusieurs écrans comportant chacun un certain nombre de champs à documenter.

On retrouve dans les données brutes les trois éléments présents dans toute expérimentation neutronique. Un programme est consacré à la gestion des caractéristiques de sonde et au calibrage périodique, et un autre à la gestion des coefficients d'étalonnage du sol de chaque expérimentation. Les quatre derniers gèrent les données relatives aux expérimentations elles-mêmes. Ils remplissent les fonctions suivantes :

- gestion des protocoles expérimentaux ;
- création de registres dans les fichiers de données et impression des feuilles de collecte ;
- informatisation ou correction des données brutes ;
- critique graphique.

Le premier des deux programmes d'exploitation transforme les données brutes en données moyennes normalisées. Autrement dit il calcule les valeurs qu'on aurait obtenues avec une sonde comptant 1000 dans l'eau et, s'il y a plusieurs répétitions, il en fait la moyenne. Le second programme transforme les valeurs précédentes en humidités volumiques à raison d'un fichier par tube de sondage neutronique. Dans l'état actuel de SISOND les études s'achèvent en important ces derniers fichiers dans un logiciel du commerce qui dans notre cas est LOTUS 1-2-3.

2.1.2. Les fichiers de données de SISOND

- LES FICHIERS D'ENTREE

Ce sont tous les fichiers de données brutes auxquels il faut ajouter un fichier de description de sonde et un fichier de description de protocoles expérimentaux. A chaque sonde correspondent un fichier de données de calibrage et un fichier de coefficients d'étalonnage du sol. Pour chaque expérimentation il y a deux fichiers. Le premier est réservé aux éventuelles mesures de référence avant et après expérimentation et aux commentaires. Le second permet de stocker les données brutes de terrain. Tout ces fichiers sont susceptibles de modifications dans n'importe quel registre. Ils sont donc à accès direct.

Pour les fichiers de données de terrain, il fallait trouver un dénominateur commun à toutes les expérimentations afin de l'utiliser comme registre. Ce dénominateur est la répétition-tube. Pour des raisons d'occupation de l'espace-disque, on a limité le nombre de niveaux de mesures par tube à 22. Chaque registre comporte donc l'espace nécessaire au stockage d'une répétition de ces 22 données, et un certain nombre d'informations complémentaires destinées à faciliter la gestion du fichier. Pour des raisons similaires, dans les fichiers de coefficients d'étalonnage, chaque registre correspond aux coefficients relatifs à un seul tube.

Pour tous les fichiers comportant des mesures neutroniques (données de terrain, calibrage), il est possible de donner un avis qualitatif sur la donnée sous forme d'un code de une lettre. Dans ce cas il faudra donc entendre par donnée l'ensemble valeur de la mesure et code associé.

Les fichiers de données normalisées ont été construits comme des fichiers de données de terrain dans une expérimentation ne comportant qu'une seule répétition. Ils sont donc regroupés avec les fichiers d'entrée.

- LES FICHIERS DE SORTIE

Le fichier de sortie ou fichier de données élaborées (par opposition aux données brutes) constitue la première étape de l'exploitation des données. Chaque tube de sondage possède son individualité et doit être étudié séparément. Du point de vue de l'utilisateur des données ce n'est plus l'expérimentation dans son ensemble qui importe, mais les données d'humidité volumique obtenues dans chaque tube pris séparément. Il y a donc un fichier par tube. D'autre part ces fichiers doivent pouvoir être importés par des logiciels du commerce ou être lus dans des programmes conçus par l'utilisateur, indépendamment du langage de programmation. Pour toutes ces raisons les fichiers de sortie sont des fichiers séquentiels de caractères ASCII (type TEXTE).

2.2. La rigueur et la souplesse dans l'utilisation de SISOND

On montre comment ces deux directrices ont été respectées simultanément lors de l'établissement du protocole de mesures, de la conduite de l'expérimentation et de l'exploitation des résultats.

SISOND accepte jusque 20 tubes de mesures par expérimentation. Par tube il peut y avoir jusque 9 répétitions et entre 2 et 22 niveaux de mesure. Les protocoles de mesures sont indépendants d'un tube à l'autre. Par conséquent SISOND permet de suivre à peu près toutes les expérimentations possibles avec sonde à neutrons. Par contre, une fois le protocole expérimental établi, c'est l'ordinateur qui imprime les feuilles de collecte et, à moins de prendre des libertés avec les indications qui y figurent, l'observateur est tenu de respecter de façon rigoureuse le plan de collecte. Une erreur expérimentale est cependant toujours possible et il a été laissé une possibilité de "récupérer" les données collectées à des niveaux incorrects. Cette facilité de changer pour une collecte certains niveaux de mesure est soumise à une restriction importante : seul le responsable de l'expérimentation peut l'utiliser. Une fois les données informatisées, le programme calcule alors par interpolation les valeurs des comptages neutroniques aux niveaux prévus dans le protocole de mesures.

Si l'ordinateur comporte une carte graphique HERCULES ou IBM, un programme graphique très puissant permet de repérer les anomalies dans la valeur des comptages neutroniques. Quand il s'agit d'une erreur de digitation elle est évidemment corrigée. Lorsque la donnée elle-même est en cause plusieurs attitudes sont possibles. Si elle est absurde elle est supprimée, mais si elle est simplement douteuse on n'a plus le droit de le faire. SISOND, dans ce cas, permet néanmoins à l'utilisateur de donner son opinion en utilisant le code associé à chaque donnée (par exemple "D" pour douteuse). Ensuite, lorsqu'on crée un fichier de travail (données normalisées ou données élaborées), ce code permet à volonté de les éliminer ou non, sans pour autant les retirer du fichier de données brutes.

Au niveau de l'exploitation des résultats, la forme donnée aux fichiers d'humidité relative autorise la plus grande souplesse. En particulier elle permet d'envisager des études non-prévues originellement en créant un programme adéquat ou en les traitant dans un logiciel du commerce.

2.3. Sécurité et simplicité dans l'utilisation de SISOND

2.3.1. SISOND et la sécurité

La sécurité a été considérée à trois niveaux : la réduction des causes d'erreurs et possibilité de confusion, le contrôle des données communiquées au système et la sauvegarde du travail effectué.

Chaque fois qu'il était possible, les informations à communiquer au système ont été ramenées à des réponses sans ambiguïté, type oui ou non (O/N) ou encore une croix en face de l'option choisie. Lors de la création de fichiers de données normalisées ou de données élaborées, l'utilisateur a le choix entre un certain nombre d'options : codes des données éliminées, dates des registres initial et terminal de la période d'étude, etc.. Chaque fois qu'un tel fichier est créé, le système permet l'impression d'un rapport comportant tous les détails utiles (noms de fichiers, options, etc...) et en particulier la date de création. Un tel rapport peut également être imprimé lors de l'informatisation de coefficients d'étalonnage, ceux-ci pouvant être amené à évoluer avec l'information disponible.

Lors de la documentation d'un écran, l'utilisateur doit communiquer au système un certain nombre de données pour permettre au programme appelé de fonctionner. Chacune de ces données est vérifiée et en cas de nécessité (appel d'un fichier inexistant, ambiguïté sur la tâche à exécuter etc...) apparaît un message explicatif d'erreur. Tant que l'écran n'a pas été documenté correctement le programme refuse de passer à la tâche suivante. Dans certain cas également le système envoie un message d'avertissement sans pour autant arrêter le processus (création d'un fichier dont le nom existe déjà, etc.).

Si au cours d'une session de travail un ou plusieurs fichiers de données brutes ont été modifiés, le système impose leur sauvegarde dans la disquette BACKUP avant de le quitter. Inversement si le système a été quitté de façon brutale (panne de courant, etc..), les fichiers modifiés sont restaurés dans leur état antérieur à partir de la disquette BACKUP.

2.3.2. Coté pratique et simplicité d'emploi de SISOND

Dans chaque programme l'utilisateur est amené à documenter un ou plusieurs écrans. Pour faciliter cette tâche ceux-ci comportent toutes les indications nécessaires. Parfois même, dans une fenêtre, apparaissent des informations complémentaires (date des dernières collectes de données, code des sondes disponibles, etc.). Enfin, si ces indications se révélaient insuffisantes, il est toujours possible de faire appel au programme d'aide. Le texte en clair correspondant aux données à documenter apparaît alors sur l'écran.

Dans un écran on est amené à documenter des champs de données, passer d'un champ à l'autre ou au précédent etc... Dans un programme on est amené à passer d'un écran au suivant ou au précédent, revenir au menu général, etc.. Toutes ces opérations s'effectuent en utilisant des commandes appropriées. Les indications nécessaires à leur maniement apparaissent automatique-

ment dans des menus de bas de page, au fur et à mesure de leur nécessité. Ces menus ont été normalisés au niveau de SISOND et, autant que faire se peut, utilisent à la fois les commandes de WORDSTAR et les indications des touches fonctions du pavé numérique du micro-ordinateur.

Au niveau de l'informatisation des données, le système ayant généré les feuilles de collecte, c'est l'image exacte de ces feuilles qui apparaît sur l'écran. Les champs à documenter sont parcourus dans le même ordre que lors de la collecte. Moyennant quoi l'informatisation des données peut être confiée à l'observateur de terrain. Son travail s'en trouve valorisé, et cela décharge d'autant le responsable scientifique. Mais comment faire coïncider cette dernière facilité avec les impératifs de sécurité. La curiosité aidant, l'observateur peut être tenté de s'intéresser à d'autres tâches que la sienne. Une fausse manoeuvre est vite faite. Un registre créé de façon intempestive peut faire perdre un temps précieux. Pour cela toutes les tâches présentant un risque sont protégées par un mot de passe connu du seul responsable de l'expérimentation.

3. FONCTIONNEMENT PRATIQUE DE SISOND

On considère d'abord ce fonctionnement vu sous l'angle de chacun des trois utilisateurs possibles, ensuite on montre comment SISOND suit pas à pas les étapes de la réalisation d'une expérimentation neutronique.

3.1. SISOND vu par chacun des utilisateurs possibles

Le point de vue de chaque utilisateur est illustré par un diagramme de flux de données. Chaque diagramme est divisé verticalement en trois zones. La partie gauche est réservée à l'utilisateur, la partie centrale aux programmes et la partie de droite aux produits, tant ceux nécessaires à l'utilisation du programme qu'à ceux résultant de son fonctionnement. Chaque programme s'inscrit dans un rectangle dont la partie supérieure donne le numéro et la partie inférieure son rôle dans le système. Lorsqu'un produit est nécessaire à son fonctionnement, il est relié à la partie supérieure du rectangle. S'il en est issu il est relié à la partie inférieure. Deux types de produit, représentés différemment, sont considérés : les sorties imprimantes et les fichiers de données appelés FO, F1, F2, etc..

Remarque : Pour simplifier l'écriture, on a donné le même nom FO aux fichiers de données de calibrage et de description de sonde, et appelé F2 les deux fichiers de données brutes.

3.1.1. Point de vue du responsable de l'expérimentation (figure 3)

Il a un accès protégé par un mot de passe dans quatre programmes.

Le programme n° 0 est à lui seul un véritable sous-système, chargé de la gestion de tout ce qui concerne les sondes à neutrons, c'est à dire leurs caractéristiques et le calibrage périodique. Chaque fois qu'une nouvelle sonde est utilisée ou qu'un incident se produit avec une sonde en service, le fichier de caractéris-

SISOND

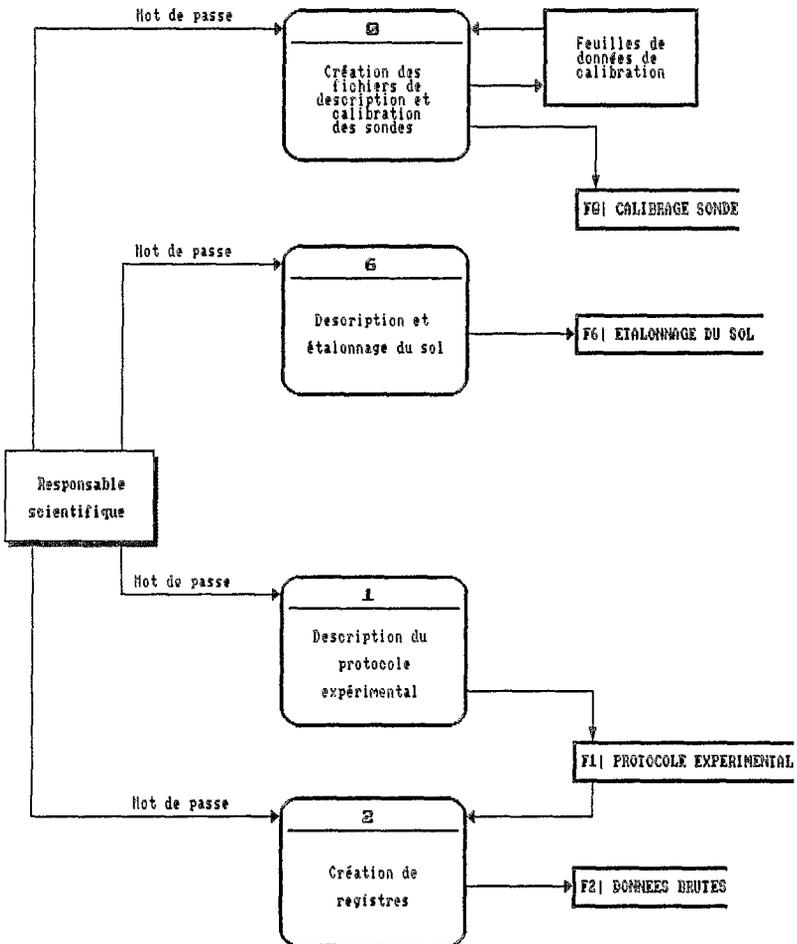


Fig.3 Diagramme de flot de données de SISOND

Point de vue du responsable de l'expérimentation

tiques de sonde est modifié en conséquence. Chaque fois qu'une sonde est calibrée on crée un registre dans le fichier de calibrage adéquat, on imprime la feuille de collecte correspondante et, quand le travail est terminé, on grave les résultats des mesures.

Le programme n° 6 permet, pour une expérimentation donnée, l'écriture des coefficients d'étalonnage du sol (étalonnage empirique ou chimique). L'utilisateur indique le nom de l'expérience, le type d'étalonnage, et s'il s'agit d'un étalonnage unique pour tous les tubes. Il informatise ensuite ces coefficients horizon par horizon en donnant les limites de chacun d'eux. Le programme calcule et grave les coefficients par niveau de mesure.

Dans le programme n° 1 sont décrites les expérimentations, c'est à dire que pour les numéros d'expérimentation choisis sont communiqués : la date initiale, le nombre de tubes, le nom de l'expérimentation, puis, par tube, le nombre de répétitions et les niveaux de mesure. Ces informations sont ensuite stockées dans le fichier de description de protocoles expérimentaux.

En utilisant les données du fichier n° 1, la partie protégée du programme n° 2 permet de créer un registre de données brutes, à une date donnée, avec une sonde donnée, dans le fichier choisi.

3.1.2. Point de vue de l'observateur de terrain (figure 4)

Une fois un registre créé, le responsable des mesures de terrain utilise le programme n° 2 pour imprimer la feuille de collecte. Après documentation de celle-ci, les informations qui y figurent sont informatisées à l'aide du programme n° 3. Les données brutes sont alors soumises au programme graphique de critique. Par comparaison des répétitions on détecte les anomalies. Il s'en suit une correction éventuelle à l'aide du programme n° 3.

3.1.3. Point de vue de l'utilisateur des données (figure 5)

A partir des données brutes, l'utilisateur des données crée avec le programme N° 4 un fichier de données normalisées. Pour cela il dispose d'un certain nombre d'options. En particulier il peut éliminer les données correspondant à certains codes, utiliser ou non les mesures sonde dans la protection comme référence, etc...

Le programme n° 7 lui permet alors de transformer les données normalisées en données d'humidité volumique. Comme précédemment il a le choix entre un certain nombre d'options qui sont ici le type d'étalonnage, le nom des fichiers créés, le *directory* dans lequel ils sont gravés, etc..

Les données élaborées sont ensuite importées dans un logiciel du commerce ou un programme de sa composition pour achever le traitement.

3.2. Utilisation de SISOND dans la réalisation d'une expérimentation

On suppose que la première installation a été effectuée donc que les unités et *directory* de travail et de BACKUP sont documentés, et la disquette BACKUP créée. On considère l'expérimentation-type du paragraphe (1-2) et on suit pas à pas les opérations effectuées dans SISOND pour sa réalisation.

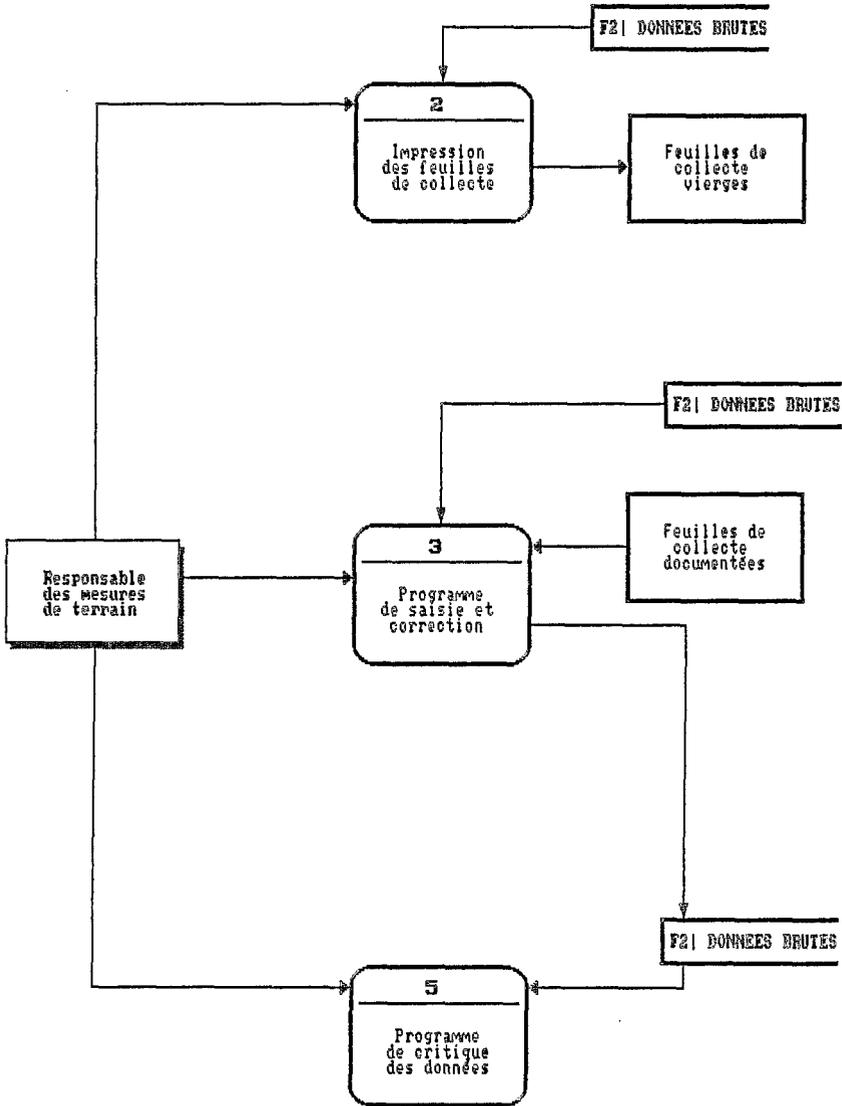


Fig.4 Diagramme de flot de données de SISOND

Point de vue de l'observateur de terrain

SISOND

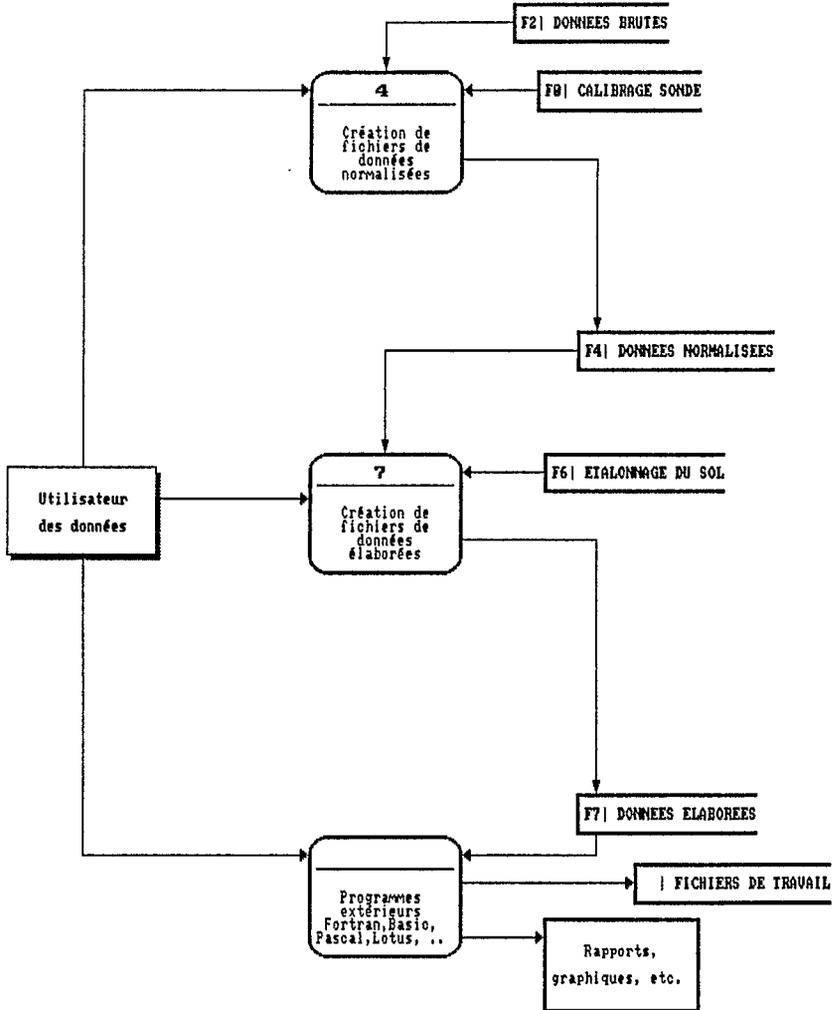


Fig.5 Diagramme de flot de données de SISOND

Point de vue de l'utilisateur des données

3.2.1. Opérations de routine

A l'aide du programme n° 0 on définit les différentes sondes disponibles et on prévoit un calibrage initial. Cette première opération est indispensable car une sécurité dans le programme n° 2 empêche de définir une collecte de données avec une sonde qui n'est pas calibrée.

3.2.2. Phase d'installation

Avec le programme n° 1 on définit le protocole de mesure du premier calibrage empirique.

Avec le programme n° 2 on crée les registres correspondant à une collecte de données et on imprime la feuille de collecte.

Avec le programme n° 3 on informatise les données brutes.

Avec le programme n° 5 ces données sont critiquées en comparant les répétitions.

Avec le programme n° 3 on corrige les données.

Avec le programme n° 4, en spécifiant qu'il s'agit d'un étalonnage empirique, on crée un fichier de données normalisées importables par tout type de programme extérieur au système.

Après calcul des coefficients d'étalonnage, on les informatise avec le programme n° 6.

Il est à noter que, vu de SISOND, la phase de destruction du site est tout à fait identique à la phase d'installation.

3.2.3. Expérimentation proprement dite

Avec le programme n° 1 on définit le protocole de mesure.

Par collecte de données on répète la séquence suivante du paragraphe précédent : programme n° 2, programme n° 3, programme n° 5, programme n° 3.

Pour exploiter les données on réalise les opérations suivantes :

- avec le programme n° 4 on crée un fichier de données normalisées ;
- avec le programme n° 5 on critique les données normalisées en observant l'évolution chronologique de chaque tube ou en les comparant entre eux ;

- avec le programme n° 3 on corrige les données normalisées ;

- avec le programme n° 7 on crée les fichiers d'humidité relative.

Bien sûr quand parviennent les résultats de l'étalonnage chimique ils sont informatisés à l'aide du programme n° 6.

CONCLUSION

SISOND a été testé durant tout un cycle de culture et s'est révélé un instrument de travail particulièrement précieux par le côté complet de sa gestion et par sa souplesse d'utilisation des données. Par exemple, au cours d'une même journée, il permet l'impression des feuilles de collecte et, une fois le travail de ter-

rain effectué, l'informatisation et la vérification des données. Ce traitement en temps presque réel présente de nombreux avantages, entre autres il permet de détecter, alors qu'il en est encore temps, les erreurs de manipulation. Il faut cependant admettre qu'un certain nombre d'imperfections ont dû être éliminées, ce qui paraît normal avec un système de près de 40 000 instructions, mais, par conception de SISONND, ceci n'a pas entraîné de surcroît de travail inadmissible.

Il reste à créer les programmes d'exploitation des données, celle-ci se faisant jusqu'à présent avec un logiciel du commerce : LOTUS 1-2-3. En version originale SISONND est en espagnol et programmé avec la version 3.0 de TURBO-PASCAL. La version complète est prévue en français avec la version 4.0 de TURBO-PASCAL et là nous nous heurtons à un double problème. D'une part il n'avait pas été prévu dès l'origine que SISONND pourrait exister en plusieurs langues et ceci va entraîner une perte de temps non-négligeable, d'autre part le passage en TURBO 4 entraîne une refonte partielle du système. Ceci est sans doute la majeure insuffisance rencontrée dans SISONND : on a donné à la maintenabilité une définition par trop restrictive car c'est non seulement la possibilité de corriger facilement les erreurs, mais aussi et surtout la capacité de s'adapter à l'évolution des techniques et de faire évoluer le système en soi. En attendant de réaliser ce travail, il y aurait une façon élégante de disposer rapidement d'un certain nombre de programmes de traitement. Récemment (1987), Roland Poss a écrit en BASIC un logiciel de traitement des mesures neutroniques (BHYSON 2), très complet sur le plan des applications. Comme pour les données élaborées obtenues avec SISONND, chaque fichier correspond à un seul tube et comprend les mêmes informations. A la différence de SISONND il s'agit de fichiers à accès direct écrits en code. Connaissant la structure des fichiers de BHYSON 2, il suffirait d'écrire dans le langage utilisé par BHYSON 2 une interface entre les fichiers des deux systèmes.

REFERENCES BIBLIOGRAPHIQUES

- CARDON, D.A.; SECHET, P.; DINIZ, T.D. de A.S. & MALTEZ, M.G.M. - 1987 - *SISCLIMA : capitalisation de données agroclimatiques au Brésil*. Hydrol. Continent., vol.2, n° 1, :3-14.
- POSS, R. - 1987 - *BHYSON 1,2. Logiciel intégré pour le traitement des données d'humidités neutroniques*. Editions de l'ORSTOM - Collection LogORSTOM, Paris. 1987.

SISPRAGAS : UNE APPROCHE DE SYSTEME D'INFORMATION POUR LE CONTROLE DES INSECTES PHYTOPHAGES.

(Patrick SECHET, José F. da S. MARTINS, Luiz C. BELARMINO & Alci E. LOECK)

RESUME - SISPRAGAS, le système d'informations phytosanitaires de l'Entreprise brésilienne de recherche agronomique (EMBRAPA), constitue un sous-système de SISGEO, Système d'informations géographiques sur l'environnement rural. Il s'agit d'un travail réalisé conjointement par les entomologistes du Centre de recherche agronomique des basses terres de climat tempéré de l'EMBRAPA (CPATB) et de l'Université fédérale de Pelotas (UFPel), état du Rio Grande du Sud, avec les informaticiens du Département d'informatique (DIN) du siège de l'EMBRAPA à Brasília, et auquel l'ORSTOM apporte son concours.

L'objectif qui a été défini initialement est la caractérisation et le contrôle des insectes prédateurs dans les basses terres de climat tempéré. Il s'agit essentiellement d'un logiciel mis au point pour saisir, stocker et divulguer des données afférentes aux nuisances provoquées par les insectes dans le système productif et les méthodes possibles pour leur contrôle.

Compte tenu du peu d'information dont on dispose sur les aspects les plus importants pour l'établissement de systèmes rationnels de contrôle d'insectes, le développement de la première version du système a été limité aux données qui se réfèrent à la région des basses terres de climat tempéré, où l'activité agricole possède une grande importance socio-économique, ce qui rend d'autant plus significatif l'impact des insectes phytophages.

La version initiale du module local, actuellement en cours de développement, prétend atteindre les objectifs suivants :

- cataloguer les insectes d'importance économique et les caractériser à partir des aspects généraux de leur morphologie ;
- gérer les informations disponibles sur le comportement des insectes, y compris l'influence des facteurs biotiques et abiotiques ;
- administrer les données de variétés culturales qui interviennent dans la relation plante-insecte ;
- gérer les données de symptomatologie de l'attaque des insectes, ainsi que les aspects qui se réfèrent à la tolérance des insectes aux produits chimiques les plus utilisés ;
- stocker les informations sur le contrôle des insectes et automatiser l'indication de méthodes préférentielles de contrôle des insectes prédateurs.

On se propose d'exposer la méthode qui a été adoptée pour la conception de ce système, les raisons qui ont motivé le choix de la structure d'information à laquelle on a abouti, les problèmes rencontrés ainsi que les perspectives à court terme pour le développement du logiciel correspondant.

INTRODUCTION

La connaissance des facteurs de l'environnement qui présentent une influence sur l'agriculture est importante pour le développement de celle-ci. Dans cette optique, l'Empresa Brasileira de Pesquisa Agropecuária, (EMBRAPA), réalise un projet appelé SISGEO, Sistema de Informações Geoambientais, destiné à rendre accessibles sur ordinateur les données de l'environnement rural, comme soutien direct à ses programmes de recherche.

La méthodologie utilisée pour la mise au point d'un tel système est décrite par ailleurs dans le même séminaire (Séchet et al., 1989).

Les problèmes phytosanitaires qui interviennent dans les systèmes productifs agricoles doivent être pris en compte, parmi les divers paramètres considérés par SISGEO dans ses différents modules. Face à la difficulté de structurer un module capable de tenir compte simultanément de tous les aspects inhérents à ces problèmes, on a conçu le système SISPRAGAS, dont les objectifs initiaux concernent l'identification et le contrôle des insectes qui constituent des nuisances pour les principales cultures d'importance économique.

Ce système, en cours de mise au point, suit la stratégie générale mise au point pour les modules de SISGEO. Il se trouve actuellement en phase de spécification, au cours de laquelle un projet physique est élaboré. Cette communication expose les divers bénéfices que l'on espère apporter avec la mise en opération de cette base de données, lesquels bénéfices ayant eux-mêmes conduit à mieux préciser les objectifs du système. Parmi ces derniers, certains se situent bien dans les orientations générales de SISGEO, tandis que d'autres, de caractère plus spécifique, font de SISPRAGAS un projet-clé pour la programmation de recherche dans le domaine de l'entomologie. On décrit également la structure qui a été définie pour le module, et qui constitue le résultat de l'étape de conception, consigné dans le projet logique ou dossier de conception détaillé (EMBRAPA 1988).

1. HISTORIQUE

L'idée de développer un système pour l'identification, la distribution et l'inventaire des insectes prédateurs a surgi initialement en 1986, dans le cadre du programme SISGEO de coopération entre l'EMBRAPA et l'ORSTOM, coordonné par le département d'informatique (DIN) de l'institution. Parmi les centres de ressources de l'EMBRAPA, qui sont les parrains des différents modules de SISGEO, le Centro de Pesquisa Agropecuária de Terras Baixas de Clima Temperado (CPATB), qui dispose de la collaboration des professeurs de l'Université Fédérale de Pelotas (UFPel), a été choisi pour orienter la mise en oeuvre du SISPRAGAS.

Les travaux de la phase de conception ont débuté seulement en mai 1987, avec une première réunion qui s'est tenue à Brasília, et qui a permis de mieux définir les objectifs du système, ceux-ci devenant la "*caractérisation des insectes phytophages dans les cultures de basses terres et l'indication de méthodes*

préférentielles de contrôle". La phase de conception s'est terminée en décembre 1988, en donnant lieu au projet logique du module local de SISPRAGAS, version 1.0, élaboré selon la systématique de SISGEO.

2. BENEFICES ATTENDUS

L'agriculture est une activité de grande importance socio-économique dans la région des basses terres de climat tempéré. Le complexe des insectes phytophages, qui existe dans l'agroécosystème considéré, inclut plusieurs espèces polyphages, qui attaquent les graminées, et en particulier le riz, le maïs, le sorgho ou les pâturages et les légumineuses, comme le soja. Ces insectes ont une gamme très étendue d'ennemis naturels, comprenant diverses espèces de pathogènes, parasitoïdes et prédateurs. Le groupe des prédateurs, plus spécifiquement, est formé d'insectes de divers ordres et familles.

Actuellement, et malgré les connaissances déjà acquises sur le sujet, il est nécessaire de mieux quantifier les indices de préjudices causés par les insectes phytophages sur le système productif, ainsi que les bénéfices obtenus par l'action de leurs ennemis naturels. Ceci est dû au fait qu'il manque encore des informations sur les facteurs essentiels pour l'établissement de systèmes rationnels de contrôle des insectes. En conséquence, il est important de rendre accessibles, par le biais d'un système d'information automatisé, les données qui se réfèrent à la composition entomofaunistique de la région des basses terres, à la bioécologie des insectes, au niveau de leur impact sur le système productif et aux nouvelles alternatives de contrôle.

Un tel système, à l'instar des autres modules de SISGEO, sera postérieurement divulgué et implanté dans les autres régions du Brésil par le biais des unités de l'EMBRAPA, en débutant par les centres de ressources.

SISPRAGAS consistera en un logiciel approprié à la saisie, au stockage et à la divulgation des informations sur l'importance des insectes dans les systèmes productifs de la région des basses terres et sur les méthodes de contrôle correspondantes. Il permettra de pronostiquer l'attaque des insectes et de recommander la meilleure méthode de contrôle, à partir du croisement des paramètres biologiques avec les autres informations de l'environnement, parmi lesquels le climat assume un rôle prépondérant.

Au-delà d'un simple stockage de données, SISPRAGAS introduira le traitement automatisé des données en soutien à une activité de recherche de l'entreprise, l'entomologie économique dans ce cas précis. De plus, il servira également pour indiquer des priorités de recherche en différentes régions, dans le cas où l'on identifie la nécessité d'études pour fournir un complément d'information.

Le système est conçu et structuré de façon à ce qu'il puisse être utilisé par des personnes peu au fait des techniques de l'informatique, laissant ainsi prévoir une utilisation plus générale.

3. OBJECTIFS

Pour pouvoir bénéficier des avantages décrits ci-dessus, il a été nécessaire de prévoir les objectifs suivants pour SISPRAGAS :

- cataloguer les insectes d'importance économique que l'on rencontre dans l'agrosystème des Basses Terres de Climat Tempéré ;
- caractériser ces insectes à partir des aspects généraux de leur morphologie ;
- administrer les informations qui se réfèrent au comportement des insectes, y compris l'influence des facteurs biotiques et abiotiques ;
- gérer les données de variétés culturales, qui interviennent dans la relation plante/insecte ;
- enregistrer les données sur les symptômes de l'attaque des insectes ;
- stocker les informations qui concernent le contrôle des insectes et automatiser l'indication de méthodes de contrôle préférentielles.

4. DESCRIPTION DU SYSTEME

4.1. Modélisation des données

La modélisation des données effectuée pour SISPRAGAS conduit au diagramme d'entités-relations présenté sur la figure 1. Trois entités de base ont été identifiées : insecte, plante et produit chimique. Cette dernière entité a été introduite comme unique entité intervenant parmi les diverses méthodes de contrôle existantes (chimiques, physiques, biologiques, variétales, etc.).

En fait, seulement une partie des caractéristiques de chaque entité interviennent dans le système : il s'agit soit de celles qui interviennent réellement dans la relation plante-insecte, soit encore celles qui correspondent aux attributs morphologiques utilisés pour la reconnaissance (identification) des insectes.

Trois relations ont ensuite été mises en évidence :

- relation de susceptibilité (d'un insecte par rapport à un produit chimique) ;
- relation d'ennemi naturel (d'un insecte par rapport à une autre espèce animale) ;
- relation d'attaque et de contrôle, entre un insecte et une plante.

Cette étape de modélisation s'est avérée la plus difficile pendant la phase de conception, car certains concepts, spécialement ceux d'attaque, de contrôle et de susceptibilité avaient été initialement attribués à l'entité "insecte". En seconde analyse, il a été jugé plus satisfaisant de définir des fichiers spécifiques pour ces attributs, dans la mesure où ceux-ci correspondent à des données croisées plante/insecte ou produit/insecte.

A un niveau de détail plus avancé, des ensembles répétitifs d'attributs apparaissent pour les insectes : ils correspondent à chacune des phases de son cycle biologique (oeuf, larve ou nymphe, chrysalide et adulte). De la même façon, le concept de plante se dédouble en culture et variété culturale, tels que

SISPRAGAS

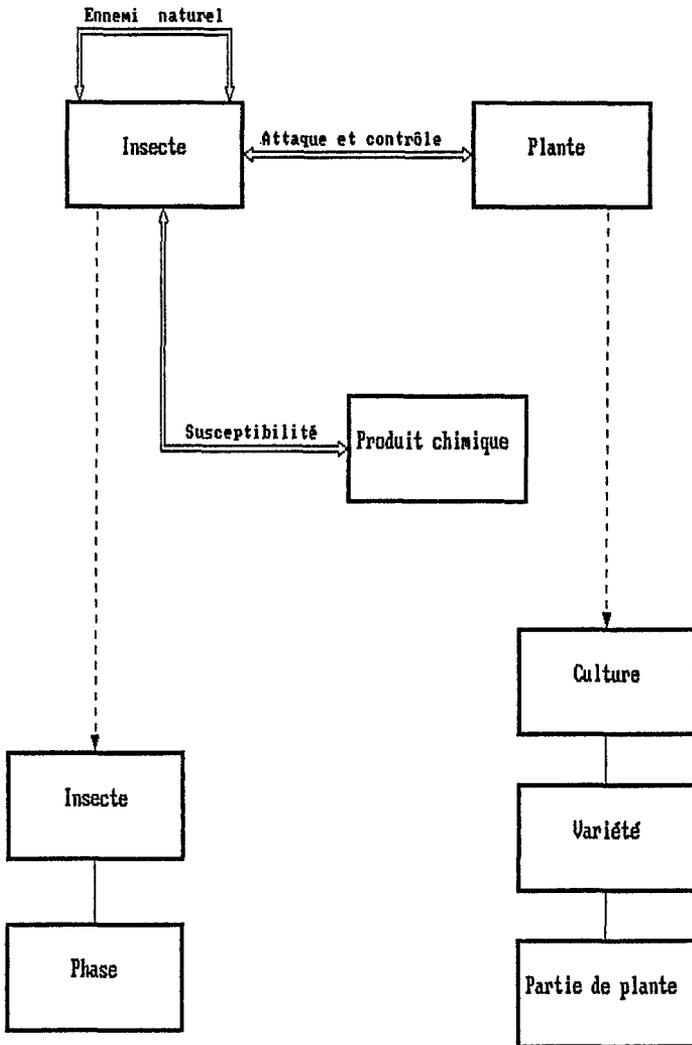


Fig.1 Résultats de la modélisation des données

les variétés culturales d'une même culture se différencient par le biais d'un ou plusieurs attributs morpho-physiologiques. Enfin, le concept de "*partie de plante*", comme racine, feuille, fleur, tige, etc. a été mis en évidence, dans la mesure où il s'agit des divers composants de la variété culturale qui peuvent être attaqués individuellement par un insecte, au cours d'une certaine phase de son développement.

4.2. Analyse fonctionnelle

Conformément à la méthode utilisée par l'équipe du projet SISGEO, et qui a déjà été présentée lors d'un séminaire précédent (SECHET 1987), le travail de conception se poursuit par une analyse fonctionnelle, laquelle s'appuie sur l'élaboration de diagrammes hiérarchisés de flot de données. Le diagramme de niveau le plus général, représenté sur la figure 2, montre l'existence de six fonctions, qui constituent les applications de SISPRAGAS.

La première application, "*Gérer les insectes*" met en évidence trois services : maintenance du dépôt d'insectes, émission de divers rapports sur les insectes et caractérisation de ceux-ci. Le premier service permet d'inclure, d'exclure et de modifier toutes les données d'insectes, tandis que le second permet d'extraire ces mêmes données sous formes d'états opérationnels (inventaire, rapports descriptifs, etc.). Le troisième service offre la possibilité d'identifier les insectes au moyen de certains de leurs attributs morphologiques, comme forme d'aide au diagnostic.

L'application "*Gérer les données d'attaque et de contrôle*" rend possible la saisie des données d'attaque et de contrôle, par le biais des deux premiers services : l'existence de l'insecte et de la variété culturale considérés, dans les fichiers correspondants, est vérifiée à cette occasion. Le troisième service fournit le moyen de croiser les données de l'insecte avec celles de la variété culturale (pour n'importe quelle culture), en utilisant deux types de consultations. En entrant avec une variété culturale déterminée, on peut extraire tous les insectes qui l'attaquent, mais aussi en partant d'un insecte, on peut sélectionner toutes les variétés culturales que celui-ci attaque. Le dernier service émet les rapports les plus complets du système, croisant les attributs de phase d'insecte avec les attributs de partie de plante, en y ajoutant les caractéristiques de l'attaque correspondante, ainsi que les paramètres relatifs au contrôle.

La troisième application, "*Gérer les variétés culturales*", possède une structure semblable à la première, avec ses trois services de maintenance du dépôt de variétés culturales, émission de rapports de variétés culturales et caractérisation de variétés culturales à partir de certains attributs morphologiques de partie de plante. Il est possible d'obtenir divers rapports de variétés culturales, comme inventaire, rapports descriptifs par partie de plante (complet ou simplifié) et rapport de cycle végétatif.

"*Gérer les ennemis naturels*" constitue l'application suivante, composée des deux services de maintenance du dépôt d'ennemis naturels et d'émission de rapports. En partant d'un insecte déterminé, on peut obtenir deux listes distinctes : liste des ennemis naturels et liste des insectes hôtes.

SISPRAGAS

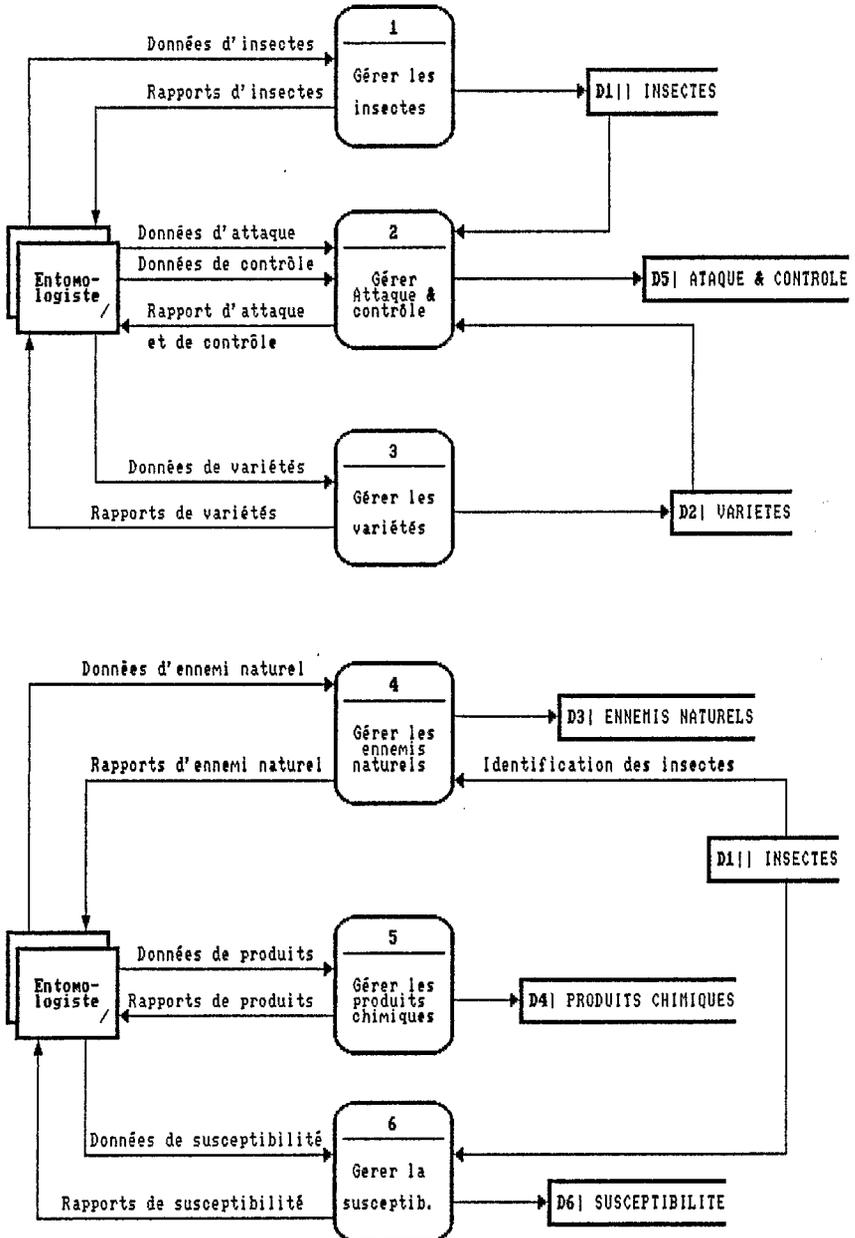


Fig.2 Diagramme de flot de données de niveau 0

La cinquième application, "Gérer les produits chimiques", traite le dépôt de produits chimiques, en fournissant les ressources de saisie et d'actualisation des données correspondantes, en plus de permettre l'émission des rapports opérationnels de ces produits : inventaire et rapport descriptif complet.

Finalement, la sixième application "Gérer les données de susceptibilité", offre un moyen d'intervenir sur le fichier où sont conservés les paramètres de susceptibilité d'un insecte déterminé par rapport à un produit chimique donné. Aussi bien la maintenance de celui-ci, que l'émission des rapports croisés correspondants (complet ou simplifié), font intervenir l'accès aux dépôts d'insectes et de produits chimiques, dans la mesure où l'enregistrement des données de susceptibilité est conditionné à l'existence des données correspondantes de produit chimique et d'insecte.

En plus de ses fonctions intrinsèques, SISPRAGAS possède également des applications qui permettent de réaliser des opérations plus générales, comme l'administration des tables de codification, la gestion de routines complètes de copie et restauration des fichiers, le contrôle d'accès ou encore la "paramétrisation" de l'interface système-utilisateur. Ces fonctions générales sont implémentées par le biais d'utilitaires spécifiques (SISTAB, SISCOP, etc.), mis au point dans le cadre du programme SISGEO.

4.3. Analyse des données

L'étape suivante de la méthodologie est constituée par l'analyse des données, au cours de laquelle les flots de données, identifiés en chacun des diagrammes élaborés durant la phase précédente, sont décrits systématiquement en termes de champs. On détaille de cette façon tous les attributs inhérents à chaque entité qui intervient dans le système, ainsi que tous les paramètres des sollicitations (de rapports, de consultations, etc.). Ensuite, chaque élément de donnée (champ) identifié est spécifié rigoureusement en ce qui concerne ses caractéristiques d'intérêt pour le traitement informatique, comme le type, la taille, le masque d'édition, l'algorithme de validation, etc.. Cette dernière étape conduit à l'élaboration d'un dictionnaire de variables pour le système.

Pour le dépôt d'insectes, les attributs considérés appartiennent à deux catégories distinctes : classification (comprenant dénominations scientifique et populaire, et leurs synonymes) et biologie du développement, avec les caractéristiques de toutes les phases du cycle de vie de l'insecte. Pour chacune des phases, SISPRAGAS enregistre les données de viabilité, certains caractères morphologiques utilisés pour la reconnaissance et les variables qui mesurent l'effet des facteurs physiques de l'environnement sur le développement de la phase en question.

Dans le cas particulier de la phase "oeuf", il existe également un ensemble de données inhérentes à la ponte. Pour les phases larve, nymphe et adulte les informations qui se réfèrent au comportement, telles que l'habitude et la forme de l'attaque, les conditions de l'environnement favorables à l'attaque, etc.,

apparaissent également. Enfin, pour la phase de chrysalide, il y a en plus des données spécifiques, comme le lieu de réalisation (par rapport à la partie de plante) ou le type de protection de chrysalide.

Il est intéressant de noter que SISPRAGAS prévoit d'associer, aux données stockées, un attribut de qualité de l'information (empirique, adaptée ou appropriée).

Pour le dépôt de variétés culturales, les données stockées se divisent en trois groupes : identification (avec la classification de la culture et l'identité de la variété culturale), caractéristiques agronomiques et classes du cycle de développement (avec les limites de chaque stage).

Les caractéristiques agronomiques correspondent à la morphologie des divers éléments de la plante : racine, tige, feuille, fleur et fruit. Pour chaque composant on a essentiellement des caractéristiques de taille, des caractéristiques externes (comme la rugosité, la pilosité et la couleur) et des caractéristiques de distribution, quand elles s'appliquent (racine, feuille et fleur). Dans le cas de la feuille, les composants internes (limbe, gousse, ligule, auricule et pétiole) sont de plus décrits. Il en va de même pour la fleur (pédoncule, sépale, pétales, glumelles supérieure et inférieure, gynécée et androcée), ces deux derniers avec leurs parties également décomposées.

Les champs qui ont été identifiés pour les produits chimiques constituent trois catégories différentes : identification, données du registre et caractérisation. La caractérisation comprend des informations sur la toxicologie, le mode d'action et les classes de produit (insecticide, nématicide, fongicide, acaricide, etc.).

Le dépôt appelé "*données d'attaque et de contrôle*" contient des informations sur l'identification (phase d'insecte et partie de plante), l'attaque et le contrôle. Les caractéristiques de l'attaque sont le local, l'époque (mesurée en terme de stage du développement physiologique de la plante) et les symptômes. Les données de contrôle sont formées de l'unité de mesure pour le niveau de contrôle et des recommandations.

Finalement, dans le dépôt de "*données de susceptibilité*", on trouve les clés identifiant le produit chimique et l'insecte, ainsi que les attributs qui mesurent l'effet de ce produit sur l'insecte, sous la forme d'un dosage nécessaire pour inviabiliser une certaine proportion des individus.

Le dictionnaire des variables élaboré à partir de la structure du SISPRAGAS, telle qu'elle est présentée sur la figure 3, contient, dans cette première version 640 rubriques, chacune d'elles représentant un champ devant être stocké dans l'un des fichiers de la base de données. Plus de 70 tables de codification ont également été identifiées. Une tactique d'implémentation a ainsi été élaborée pour le système, lequel est en cours de spécification pour une implémentation sur micro-ordinateur PC-compatible, équipé d'un disque dur.

4.4. Perspectives quant à l'implémentation

En suivant la méthodologie de travail de SISGEO, le projet logique terminé à l'issue de la phase de conception passe dans les mains de l'analyste de spécification, chargé de la définition physique de la base de données et des programmes qui composent le système applicatif. Compte tenu de la documentation

SISPRAGAS

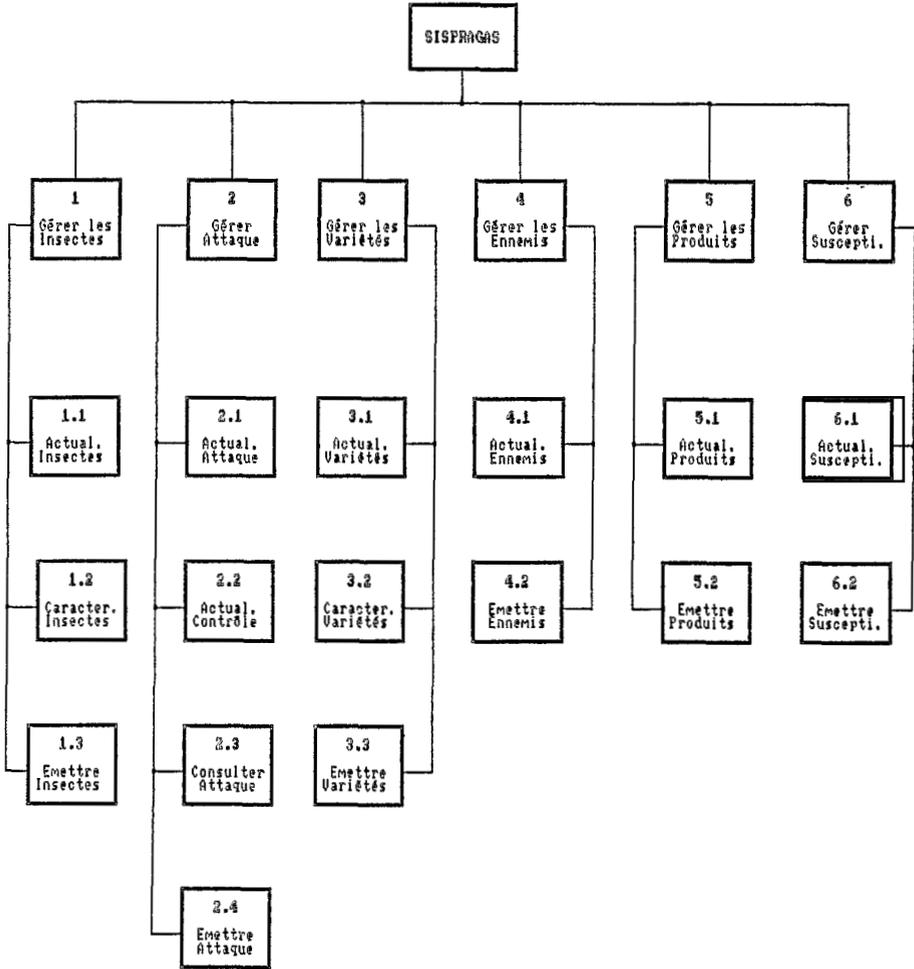


Fig.3 Tactique d'implémentation (structure du système)

disponible dans le projet logique, l'équipe d'implémentation n'a pas besoin de recourir constamment à l'équipe thématique, mais seulement lorsque certaines étapes bien définies du projet sont terminées, pour approbation.

Compte tenu de la taille relativement importante de ce système, il est actuellement envisagé de disposer d'une première version opérationnelle en moins de douze mois, laquelle devra être présentée au prochain congrès brésilien d'entomologie agricole, prévu à Recife, en juillet 1990.

CONCLUSION

L'élaboration du projet logique de SISPRAGAS démontre que l'intégration de spécialistes de disciplines différentes (dans ce cas précis la biologie et l'informatique) est possible au point de structurer des logiciels dotés de caractéristiques propres de soutien de recherche. De tels systèmes présentent une meilleure chance d'être acceptés par la communauté scientifique correspondante, car ils respectent les principes de base de la discipline considérée, lesquels sont fournis par l'équipe thématique. De plus, la participation des professionnels de l'informatique permet de garantir que les fondements essentiels du traitement des données (sécurité et intégrité, normalisation des fichiers, etc.) sont incorporés, conférant ainsi au produit final un bon niveau de crédibilité et une plus grande probabilité de mise en oeuvre effective et de dissémination.

REMERCIEMENTS

Les auteurs remercient Julio José Centeno da Silva, Gláucia Soares da Silva Pinto (chercheurs du CPATB/EMBRAPA), Ida Cristina Lago, Milton de Souza Guerra (Professeurs de l'UFPEl) et Miriam Cristina Galante Gomes (analyste de systèmes du DIN/EMBRAPA), pour leur collaboration lors de la formulation du projet logique de SISPRAGAS.

REFERENCES BIBLIOGRAPHIQUES

- BORROR, D.J. & DELONG, D.M. *Introdução ao estudo dos insetos*. Edição Brasileira, Rio de Janeiro, USAID, Programa de publicações didáticas, 1969. 660p.
- CAVERO, E.S.; GUERRA, M.S. & VIEIRA, V.V. *Inseticidas e Acaricidas - Toxicologia ; Receituário agrônômico*. Piraçicaba, Livroceres, 1985. 424p.
- EMBRAPA (DIN & CPATB). *Projeto lógico de SISPRAGAS*. Brasília, 1988, 230p. (Diffusion restreinte).

- GALLO, D.; NAKANO, O.; SILVEIRA NETO, S.; CARVALHO, R.P.L.; BATISTA, G.C. De; BERTI FILHO, E.; PARRA, J.R.P.; ZUCCHI, R.A. & ALVES, S.B. *Manual de entomologia agrícola*. São Paulo, Ed. Agrônômica Ceres, 1978. 531p.
- GANE, C. & SARSON, T. *Structured systems analysis : tools and techniques*. Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1979. 256p.
- KING, A.B.S. & SAUNDERS, J.L. *The invertebrate pests of annual food crops in Central América*. London, Overseas Development Administration, 1984. 166p.
- KOGAN, M. & HERZOG, D.C. *Sampling methods in Soybean entomology*. New York, Springer-Verlag New York Inc., 1980. 587p.
- LARRA, F.M. *Princípios de resistência de plantas à insetos*. Piraçicaba, Ed. Livroceres, 1978. 207p.
- MARANHÃO, Z.C. *Morfologia geral dos insetos*. São Paulo, Nobel, 1978. 396p.
- MAXWELL, F.G. & JENNINGS, P.R. *Breeding plant resistant to insects*. New York, John Wiley & Sons, 1980. 683p.
- RAWITSCHER, F. *Elementos básicos de botânica*. 8a edição, São Paulo, Editora Nacional, 1979. 382p.
- REISSIG, W.H.; HEINRICH, E.A.; LITSINGER, J.A.; MOODY, K.; FIEDLER, L.; MEW, T.W. & BARRION, A.T. *Illustrated guide to integrated pest management in rice in Tropical Asia*. Los Banos, International Rice Research Institute, 1986. 411p.
- SECHET, P. *Méthode d'analyse structurée pour la mise au point de logiciels*. In : Séminfor 1, Premier séminaire informatique de l'ORSTOM. Paris, ORSTOM, 1987. pp.221-48.
- SECHET, P.; FRESNEDA, P.S.V. & COCHONNEAU, G. *SISGEO : une méthodologie d'administration des données de l'environnement rural*. In : Séminfor 3, Troisième séminaire informatique de l'ORSTOM. Paris, ORSTOM, 1989.
- SILVEIRA NETO, S.; NAKANO, O.; BARBIN, D. & VILLA NOVA, N.A. *Manual de ecologia dos insetos*. São Paulo, Ed. Agrônômica Ceres, 1976. 419p.
- WIGGLESWORTH, V.B. *The principles of insect physiology*, London, Methuen & Co. Ltd., 1965. 741p.

ACQUISITION, TRAITEMENT ET TRANSMISSION DE DONNEES DE VOLCANOLOGIE.

(J. TABBAGH, C. PAMBRUN & M. LARDY)

avec la collaboration de C. ARCHAMBAULT, J.P. EISSEN,
P. MAILLET, M. MONZIER & A. TABBAGH

RESUME - Le volcanisme tient un rôle primordial dans l'évolution de notre planète, on estime qu'il dissipe environ le tiers de l'énergie thermique du globe (y compris le volcanisme des dorsales).

Beaucoup de pays en voie de développement sont soumis aux risques géologiques majeurs et notamment aux risques volcaniques.

Depuis 1985, l'UR 1F du département TOA a entamé avec des partenaires français CNET, PIRPSEV, CNRS, IPG.P, Université de Paris VII, des collaborations pour l'étude de la prévision et de la surveillance des éruptions volcaniques en vue d'apporter sa contribution à la réduction des risques volcaniques dans les pays en voie de développement.

Des matériels simples et fiables ont été implantés sur deux volcans-laboratoires (Matthew 1986, Hunter 1988) situés dans le SO Pacifique. Ces outils entièrement automatiques effectuent des mesures météorologiques, thermiques et sismologiques quotidiennes transmises par voie satellitaire. Ces données sont exploitées de différentes façons après leur mise à disposition au centre CLS de Toulouse (Argos) :

- prélevées automatiquement chaque jour, via le réseau TRANSPAC depuis le centre de Nouméa ; après une rapide analyse, l'observateur commande d'éventuels tracés (variation de température, de pression, histogramme de sismicité, etc.). On envisage à terme la mise en place d'un système expert ;

- interrogé quotidiennement depuis le CTIV (centre de téléobservation informatisé des volcans, CRG GARCHY), le centre Argos de Toulouse fournit les données des dernières 24 heures ; traitées pour partie (élimination des redondances) elles sont transférées vers la banque de données de volcanologie créée au CIRCE (Orsay). Un fichier de données brutes et un fichier de données interprétées sont ainsi accessibles à tous les utilisateurs potentiels, via les moyens télématiques habituels.

Le CIRCE distribue également les données sous formes graphiques et numériques sur un serveur VIDEOTEX, ce qui permet à tous les possesseurs d'un Minitel d'avoir accès à l'essentiel (suivi en cas de crise). Toutes les données enregistrées depuis le début d'une opération peuvent être appelées.

De plus, des fichiers sur disquettes sont constitués au centre de recherches de Garchy, et au centre ORSTOM de Nouméa qui assure également la mise à disposition de certaines données auprès du service météorologique de Nouvelle Calédonie.

L'objectif est une participation à un observatoire mondial, pour l'acquisition de données de volcanologie en temps réel et une collaboration à la préparation d'une décennie internationale (1990-2000) pour la réduction des risques naturels.

Nous proposons donc d'exposer, à partir des outils utilisés, les procédures retenues (tri, représentation, accès, etc.) et de faire la démonstration en temps réel de l'accès aux données sur deux centres serveurs : CLS et CIRCE.

1. SURVEILLANCE DES VOLCANS

Le volcanisme qui témoigne de la dynamique de notre planète est associé à la création de dorsales localisées essentiellement sous la mer, à la destruction de croûte océanique (zone où le plancher océanique s'enfonce sous le manteau terrestre) et, en dehors des frontières de plaques, aux "points chauds" (Açores, Hawaii, Réunion, Canaries, etc.).

Des progrès notables sur la connaissance des volcans ont été enregistrés ces dernières années ; mais chaque édifice à son histoire, ses particularités et ses activités propres, qu'il faut reconstituer et surveiller en permanence.

La reconstitution des éruptions se fait à partir des annales historiques lorsqu'elles existent et par des prélèvements et analyses des matériaux émis au cours des éruptions.

La surveillance se fait par la mesure permanente de paramètres géophysiques et géochimiques. Sur les quelques 500 volcans actifs à la surface de la terre, peu sont surveillés et quelques dizaines sont dotés d'observatoires permanents. Des pays riches déploient souvent des moyens importants autour des volcans (Kilauea, Piton de la Fournaise, etc.), essentiels pour une meilleure compréhension des mécanismes volcaniques et des phénomènes associés à la dynamique de la terre. De tels observatoires réunissant des moyens matériels et humains importants ne peuvent être systématiquement installés sur des volcans aux cycles éruptifs espacés, les plus menaçants et souvent situés dans les pays en développement.

2. MOYENS D'INVESTIGATION

Les transmissions spatiales facilitent la mise en oeuvre à terre d'observatoires automatiques. C'est à l'initiative de PIRPSEV en 1981, et avec la participation du CNET (Service Environnement), qu'a été entreprise une expérience de téléobservation sur l'Etna ; l'expérience acquise a été transposée, et adaptée en 1986 à l'initiative de l'ORSTOM, sur deux îlots volcaniques du S.O. Pacifique, Matthews et Hunter.

Les techniques spatiales permettent d'assurer des mesures automatiques et permanentes pour un ensemble de paramètres significatifs en regard des édifices concernés. Cette surveillance s'apparente à une "sonnette d'alarme" capable de prévenir la communauté scientifique de la persistance de phénomènes anormaux susceptibles d'entraîner une investigation avec des moyens plus importants.

Le système de transmission spatiale utilisé (ARGOS) est associé à des moyens informatiques et télématiques qui facilitent le traitement et la distribution de l'information, à la fois auprès de pays dont les volcans sont sous surveillance (s'ils disposent de moyens de communications associés aux réseaux de transmission de données) et auprès de la communauté scientifique concernée.

3. ACQUISITION

Les mesures de sismologie, températures de fumerolles et de sol, complétées par des informations météorologiques sont transmises toutes les deux minutes et demie par les balises ARGOS sous forme de paramètres numérisés, quinze pour Matthews et dix-huit pour Hunter.

Les données acquises sur les deux volcans sont actualisées toutes les quatre heures et sont recueillies de façon héliosynchrone par les satellites TIROS, huit fois par jour. Après l'émission vers une des trois stations terrestres de réception, les messages sont accessibles moins de trois heures après leur acquisition auprès du centre de distribution CLS de Toulouse ; deux types d'exploitation sont alors réalisés.

4. EXPLOITATION ET TRAITEMENT DES DONNEES

4.1. De Nouméa

Nous interrogeons quotidiennement, via le réseau TRANSPAC, le centre ARGOS de Toulouse afin de contrôler le bon fonctionnement des stations et l'évolution éventuelle des paramètres.

A la vue des dernières données du dernier passage satellite, nous pouvons compléter l'interrogation par la récupération de données supplémentaires. La capture et l'automatisation du traitement de données enregistrées au cours des cent dernières heures, et disponibles au centre de distribution CLS de Toulouse, permet de suivre graphiquement les évolutions les plus récentes des paramètres mesurés sur les volcans MATTHEWS et HUNTER.

Après acquisition et sauvegarde sous "CROSSTALK", le programme "ARGOS", écrit et compilé sous TurboBASIC, permet la mise en forme des données, après suppression des redondances et des valeurs inutiles, et crée un fichier trié exploitable par un logiciel graphique, GRAPH-IN-THE-BOX par exemple. On peut donc tracer à l'écran et imprimer un ensemble de paramètres pour une période maximum de quatre jours (voir exemple 1).

Exemple 1 : Données récupérées auprès de CLS ARGOS à Toulouse, via
TOMPAC (TRANSPAC) depuis Nouméa.

PRV/C,0448,TX,241/10-243/00,09661

Prog 00448

09661 NO LOCATION 241/1332Z

(4) 35 56 69 72

80 72 117 99

164 10 65 44

159 09 10 00

36 55 70 68

80 72 114 100

171 10 66 44

159 09 10 00

09661 NO LOCATION 242/1903Z

(5) 43 59 70 65

80 74 112 96

227 19 64 45

159 09 10 00

42 60 70 67

80 73 114 98

187 11 62 44

159 09 10 00

09661 NO LOCATION 241/1510Z

(3) 35 56 69 72

80 72 117 99

164 10 65 44

159 09 10 00

36 55 70 68

80 72 114 100

171 10 66 44

159 09 10 00

09661 NO LOCATION 242/2041Z

(5) 43 59 70 65

80 74 112 96

227 19 64 45

159 09 10 00

44 62 70 65

80 75 122 98

235 15 64 45

159 09 10 00

... suite passage 37 à 42

ARGOS READY

/LOGOUT

Fichier ci-dessus remis en forme avant tracé

241243.MAT

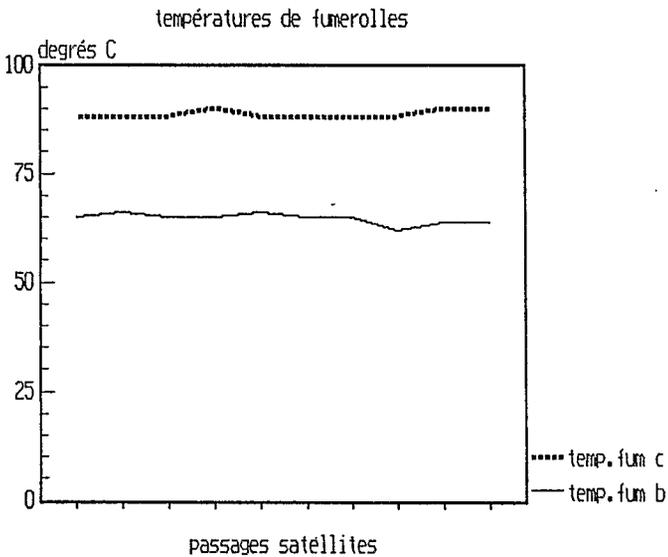
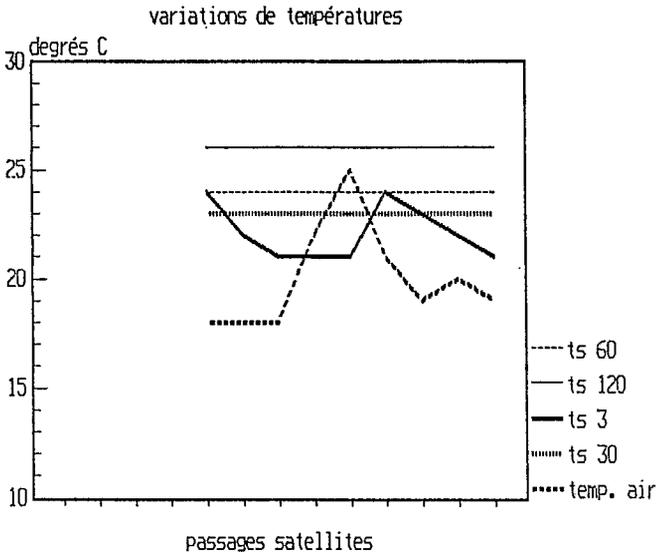
FICHIER 241243.MAT.DONNEES MATTHEW.DATE 241/10-243/00

FORMAT :

N	Tair	Tsol1	Tsol2	Tsol3	Tsol4	Ubat	Patm	Hr	TfumA	TfumB	TfumC			
35	18	23	24	26	24	23	999	82	10	65	88	0	0	0
36	18	23	22	26	24	23	1000	85	10	66	88	0	0	0
37	18	23	21	26	24	22	999	82	9	65	88	0	0	0
38	22	23	21	26	24	23	998	67	1	65	90	0	0	0
39	25	23	21	26	24	24	997	50	1	66	88	0	0	0
40	21	23	24	26	24	23	995	67	1	65	88	0	0	0
41	19	23	23	26	24	23	998	89	9	65	88	0	0	0
42	20	23	22	26	24	23	998	93	11	62	88	0	0	0
43	19	23	21	26	24	22	996	100	19	64	90	0	0	0
44	20	23	21	26	25	23	998	100	15	64	90	0	0	0

Tracés obtenus à partir du fichier 241243.MAT
de la page précédente

Exemple 1



4.2. Depuis le CTIV (CNRS-GARCHY)

On prélève une fois par jour l'ensemble des données (via micro-ordinateur et modem) des dernières vingt quatre heures, auprès du centre serveur ARGOS de Toulouse, à travers le réseau TRANSPAC ; un changement de format assure le compactage des fichiers qui sont stockés sur disquette.

Un ensemble de logiciels permet de trier ces données suivant l'heure de réception des messages, d'éliminer les doubles et de constituer un fichier de données brutes par mois, qui est consultable immédiatement. Un traitement plus complet est effectué une fois par jour ; il consiste à éliminer les erreurs de transmission, à calculer l'heure d'acquisition et la valeur des paramètres physiques à partir des formules de conversion et des constantes d'étalonnage, afin de créer un fichier de données interprétées pour chaque mois.

De plus, des logiciels ont été développées pour permettre des consultations à partir du système Vidéotexte (Minitel) ; à partir du traitement effectué précédemment on peut obtenir sous forme de courbes ou de données la variation quotidienne ou mensuelle d'un paramètre voir exemple 2).

L'ensemble de ces données disponible au CIRCE est stocké dans une banque de volcanologie lancée à l'initiative de PIRPSEV, et maintenant prise en charge par l'INSU. Toutes les données, conservées sous forme d'un fichier de données brutes et d'un fichier de données interprétées, sont accessibles par les moyens télématiques par la communauté d'utilisateurs.

CONCLUSION

Le développement de réseaux téléobservés sur des volcans actifs nécessite un examen méthodique des données ; leur analyse rapide entraînera à terme la mise en place de "systèmes experts" capables de mettre en évidence les variations significatives et de déclencher des alertes.

L'interconnexion des centres de calcul par les réseaux (FNET, EARN, EUNET, etc.), et donc du RIO dont la multiplication des points d'accès est en cours, devrait permettre de "ramener" vers les pays en développement les données précédemment acquises. La mise en place de messageries informatiques (RIO) ne pourra que favoriser les échanges entre les scientifiques qui participent à l'acquisition de données de volcanologie en temps réel pour la réduction des risques naturels.

*Exemple 2 : Fichiers bruts disponibles au CIRCE récupérés par réseau TOMPAC
(TRANSPAC) depuis Nouméa.*

MATTHEWS

448 9661 32 14488 01 47 14 6 51 75 70 66 81 74 142 97 114 01
67 44 160 09 10 00 50 71 70 63 81 73 148 98 130 01 66 45 160 09
10 00

448 9661 32 14488 03 27 39 5 51 75 70 66 81 74 142 97 114 01
67 44 160 09 10 00 50 71 70 63 81 73 148 98 130 01 66 45 160 09
10 00

448 9661 32 14488 01 47 14 6 51 75 70 66 81 74 142 97 114 01
67 44 160 09 10 00 50 71 70 63 81 73 148 98 130 01 66 45 160 09
10 00

448 9661 32 14488 03 27 39 5 51 75 70 66 81 74 142 97 114 01
67 44 160 09 10 00 50 71 70 63 81 73 148 98 130 01 66 45 160 09
10 00

HUNTER

448 9662 32 14488 01 33 32 1 235 98 109 83 83 82 82 158 116 01
255 00 01 00 00 00 08 00 01 255 254 254 127 71 209 42 250 248
248 248 248 248

448 9662 32 14488 01 38 31 2 234 99 110 84 84 84 83 169 132 03
255 00 01 01 01 01 08 00 01 255 254 254 164 70 209 42 224 224
224 250 250 248

448 9662 32 14488 01 43 30 2 235 98 109 83 83 82 82 158 116 01
255 00 01 00 00 00 08 00 01 255 254 254 127 71 209 42 250 248
248 248 248 248

448 9662 32 14488 03 13 14 1 235 98 109 83 83 82 82 158 116 01
255 00 01 00 00 00 08 00 01 255 254 254 127 71 209 42 250 248
248 248 248 248

Fichiers interprétés disponibles au CIRCE récupérés par réseau TOMPAC
(TRANSPAC) depuis Nouméa

DMAT1I89.09

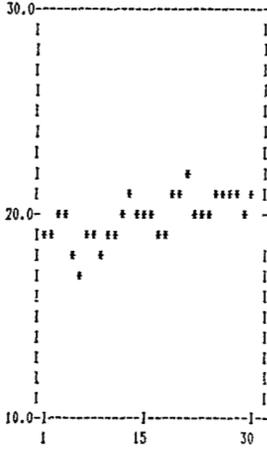
9661	89243	21	12	51	25.0	26.0	23.7	22.3	21.0	24.1	1013.	57.	1.
67.	94.	0	0	0									
9661	89244	1	14	52	20.0	26.0	23.7	22.3	24.3	23.5	1012.	73.	1.
67.	94.	0	0	0									
9661	89244	5	17	53	19.0	26.0	23.7	22.3	22.7	23.1	1015.	78.	3.
67.	94.	0	0	0									
9661	89244	9	19	54	18.3	26.0	23.7	22.7	21.3	23.0	1014.	79.	3.
68.	94.	0	0	0									
9661	89244	13	21	55	18.0	26.0	23.7	22.7	20.7	22.9	1013.	73.	3.
68.	95.	0	0	0									
9661	89244	17	23	56	21.7	26.0	23.7	22.7	20.0	24.0	1015.	59.	1.
68.	94.	0	0	0									
9661	89244	21	25	57	22.3	26.0	23.7	22.3	21.0	24.5	1014.	58.	1.
68.	94.	0	0	0									

DHUN1I89.09

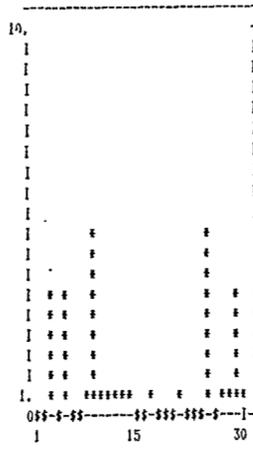
9662	89244	1	18	235	22.44	27.25	20.75	20.75	20.50	20.50	979.0		
58.0	-7.1	55.0	0.0	0.4	0	0	0	0	0	0			
9662	89244	6	38	237	22.36	19.75	20.50	20.75	20.50	20.75	983.5		
79.0	-3.5	89.0	0.0	3.1	0	0	0	0	0	0			
9662	89244	9	18	238	22.20	19.00	20.75	21.00	20.50	21.00	985.5		
80.5	0.0	88.0	0.0	4.6	0	0	0	0	0	0			
9662	89244	11	58	239	22.20	18.75	20.75	21.00	20.75	21.25	986.5		
80.2	0.0	88.0	0.0	5.4	0	0	0	0	0	0			

Tracés obtenus à partir du serveur vidéotexte (MINITEL) du CIRCE

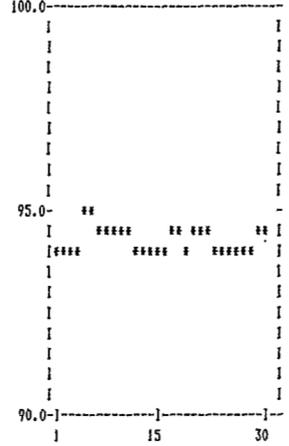
MATTHEW TEMPERATURE SOUS ABR1
SEPTEMBRE 1989



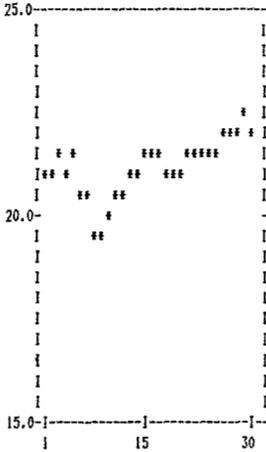
MATTHEW COMPTEUR SISMIQUE NIVEAU 1
SEPTEMBRE 1989



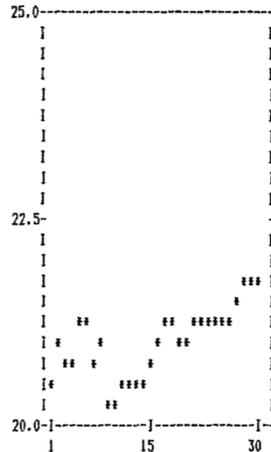
MATTHEW TEMPERATURE FUMEROLLE (3)
SEPTEMBRE 1989



HUNTER TEMPERATURE 30 cm
SEPTEMBRE 1989



HUNTER TEMPERATURE 60 cm
SEPTEMBRE 1989



HUNTER TEMPERATURE 120 cm
SEPTEMBRE 1989



GLOSSAIRE

- CIRCE : Centre inter-régional de calcul électronique (CNRS) ;
CLS : Collecte et localisation de données satellitaires (ARGOS) ;
CNET : Centre national d'étude des télécommunications
CRG : Centre de recherches géophysiques, Garchy (CNRS)
EARN : European academic research network
PIRPSEV : Programme interdisciplinaire de recherche sur la prévision et la surveillance des éruptions volcaniques
TRANSPAC : Transmission de données par paquets
INSU : Institut national des sciences de l'univers
CTIV : Centre de téléobservation informatisée des volcans
FNET : Réseau français des machines Unix
EUNET : Réseau européen des machines Unix.

INTEGRATION D'IMAGERIE SATELLITAIRE DANS UN SYSTEME D'INFORMATION GEOGRAPHIQUE

(Bernard LORTIC)

avec le concours de M. RAKOTO, D. REMY & F. PELLETIER

RESUME - La tentation est grande pour tout opérateur d'un système de bases de données localisées de pouvoir "gérer" toute l'imagerie aérienne ou spatiale disponible sur le domaine géographique de la base.

La nécessité est impérieuse pour tout utilisateur de données satellitaires (d'images quasi-verticales) par voies informatiques de registrer les images (données) de différentes dates sous un même référentiel.

La présente communication montre, sur l'exemple de la ville de Quito, quels sont :

les problèmes à résoudre :

- l'image est constituée de pixels jointifs ;
- sa représentation et son stockage ne peuvent se faire que sous une projection, définie.

Dans le cas contraire il faudrait transcoder chaque pixel en ajoutant les coordonnées explicitement ;

- toute image de la surface de la planète comporte des déformations : distorsion due aux optiques du capteur, roulis, lacets, tangage de la plate-forme (satellite ou avion), relief ;

- les données images sont très encombrantes : une image SPOT représente à peu près 30 Méga-octets pour l'A.I.Q.. Nous disposons en début de programme de sept images ;

- les solutions doivent pouvoir intégrer l'imagerie aérienne sur Quito (la dernière mission de photo aérienne est de) ;

les solutions adoptées par l'A.I.Q. (proposition ATP CNRS/CNES) :

- pointés stéréoscopiques par vidéoplotter ;
- recalage géographique par création des surfaces de déformation par interpolation ;
- corrélation automatique par la méthode SSDA () ;
- itération du processus.

L'intérêt de cette méthode est de pouvoir disposer de données constamment améliorées, alors qu'actuellement l'intégration est chère et unique.

Une démonstration pourra être faite des outils et des résultats sur des images de Quito (prise de vue de quatre dates très différentes) et de Marnes-la-Vallée (prise de vue de quatre dates d'origines très diverses).

INTRODUCTION

On peut rêver de disposer d'un système d'information total. Ce qui impliquerait que l'on dispose en chaque point de l'espace-temps de toutes les informations disponibles. Ce type de système est partiellement réalisé en médecine où l'on imagine actuellement des systèmes de représentation du corps humain en trois dimensions permettant de mieux référencer, visualiser et transmettre les différentes pathologies. On réalise ainsi un écorché visuel dans lequel on peut plonger l'œil sinon la main.

Cependant, ce type de systèmes ne décrit pas vraiment l'objet en trois dimensions mais plutôt une série de surfaces en "pelures d'oignons", bien positionnées dans l'espace à trois dimensions.

Heureusement, l'approche géographique privilégie pour le moment la surface de la planète. On n'envisage pas de coder chaque point de l'espace quelle que soit sa profondeur. Ce qui est en question c'est plutôt la surface terrestre, dans un référent cartographique, avec une série d'attributs comme la pente, l'altitude, mais aussi la population, la végétation, la géologie, les données climatiques, etc..

Ce qui veut dire que l'approche actuelle reste une approche cartographique, avec les présupposés de projections, ellipsoïdes de référence, surfaces et points immatériels y afférents. Nous resterons dans ce cadre traditionnel pour répondre au désir de l'utilisateur d'un système de bases de données localisées qui voudrait pouvoir gérer toute l'imagerie aérienne et spatiale disponible sur un domaine géographique.

D'autre part, l'utilisateur de données satellitaires (images verticales ou obliques) reste tributaire de techniques maintenant habituelles, comme le fait que traitement d'image doit nécessairement s'effectuer sous une forme matricielle, qu'elle soit carrée ou hexagonale. Pour faire se correspondre précisément des images de dates et provenances différentes, cet utilisateur doit donc empiler des matrices de pixels.

Ces deux impératifs, cartographie et matrice de points, ramenés à des prétentions modestes, ou du moins envisageables, ont conduit à définir des outils et un ensemble de procédures utilisables pour intégrer l'imagerie aérienne ou spatiale dans un système d'information géographique. Nous nous situons délibérément dans le cadre d'un institut tel que l'ORSTOM, et dans les applications en cours actuellement avec le système *SAVANE (M. Souris). Cette méthode est destinée à satisfaire aussi bien l'AIQ (Atlas informatisé de Quito), que l'atlas régional de l'état de Veracruz ou des études de terroir au Sénégal.

Nous allons décrire une méthode de recalage d'images, étape indispensable à leur intégration dans une banque de données-images pour extraire une fenêtre, après définition de la zone d'étude et de croiser données-images avec les autres informations de la base. Comment intégrer des images aériennes et spatiales avec une cartographie dans un système de gestion de cette information géographique ?

Nous avons suivi quelques principes :

- le recalage d'images se fera suivant une projection déterminée. Car on ne peut travailler que sur une matrice de points représentant une projection plane. Il n'est pas envisageable, pour des questions d'encombrement, de stocker la données en coordonnées géographiques auxquelles seraient associées des valeurs de radiométrie ;

- les images sont stockées au sein de la base de données images dans leur totalité, avec les fichiers descriptifs. Il est impératif que ces images soient recalées géométriquement ;

- le système de gestion effectuera l'extraction en fonction d'un point bas et d'une taille de pixel définie. On se reportera au fonctionnement d'un système tel que SAVANE : le système permet de déterminer la fenêtre, l'échelle, les données disponibles et leur domaine de validité.

L'utilisation des images peut être une contemplation méditative statique. Elle peut être aussi dynamique par animation d'images. Elle peut être également une simple extraction d'information : statistique descriptive par entité zonale (par exemple, quantité d'eucalyptus dans l'îlot 456).

Elle peut surtout être un croisement des informations. Quels sont les rapports entre les zones actuellement en cours d'aménagement que l'on obtient par analyse des images panchromatiques de SPOT, les autorisations de construire et le plan d'urbanisme ?

Sur un exemple d'une portion de la ville de Quito, nous montrerons les étapes de la superposition de deux images enregistrées en oblique par SPOT, l'une le 23 juin 1986 (22° d'incidence droite); l'autre le 26 novembre 1986 (8° d'incidence gauche).

Nous montrerons ensuite un produit dérivé : ces données peuvent créer un modèle numérique de terrain utilisable, entre autres, pour une restitution en perspective (appelée vision 3D) en noir et blanc et en composition colorée.

Ces données sont enfin recalées sur la cartographie utilisée dans l'AIQ. L'ensemble des images SPOT, "*Thematic Mapper*", mais aussi leurs dérivées (modèles numériques de terrain, classifications thématiques) sont disponibles, visualisables et manipulables par l'atlas informatisé.

1. SUPERPOSITION DE DEUX IMAGES

1.1. Constitution du couple stéréoscopique

La procédure décrite tient compte d'une situation un peu particulière de l'application géographique : la ville de Quito est située dans une région de relief accidenté, avec un fort gradient d'altitude. On passe en 25 km, de 2 000 mètres au Rio Guayalabamba à 4 500 mètres au sommet du volcan du Pichincha.

Si les déformations des images peuvent être dues aux variations d'altitude du satellite (ou de l'avion), roulis ou lacet, les différences les plus importantes sont dues au relief. Si l'on peut parler de déformation en région plane, il vaudrait

mieux parler de géométries différentes en région montagneuse. Or, si l'on veut avoir une chance raisonnable d'obtenir des images sans nuage, il faut pouvoir se contenter d'images du satellite SPOT en prises de vue obliques. Superposer ces images revient à traiter le relief et la démarche décrite est une démarche de photogrammétrie.

Nous utiliserons les deux images SPOT panchromatiques suivantes :

- image SPOT de juin 1986, angle de prise de vue de 22° à droite.
Fichier de 1 000 x 1 200 pixels de dix mètres ;

- image SPOT de novembre 1986, angle de prise de vue de 8° à gauche. Fichier de 800 x 1 000 pixels de dix mètres.

Pour effectuer l'orientation correcte du couple stéréoscopique, on repéré les coordonnées-images de trois points d'appui de même altitude. Nous avons choisi ici trois points d'appui à une altitude sensiblement égale à 2 650 mètres, que nous pouvons repérer sur la cartographie de l'Institut géographique militaire, et sur les images de SPOT.

		1	184
Point numéro 1 :	Nov. x=598, y=033	----->	
	Juin x=782, y=172	:	
		:	139
Point numéro 2 :	Nov. x=427, y=410	2	187
	Juin x=614, y=556	----->	3 184
		:	----->
		:	146
Point numéro 3 :	Nov. x=950, y=672	v	: 148
	Juin x=1134, y=820		v

Vecteurs de déplacement

Les vecteurs de déplacement ainsi obtenus montrent qu'il faut utiliser une rectification par le programme *Degré 1 (M.Rakoto), qui effectue la rotation, la translation et l'homothétie de l'image.

1.2. Prise d'un premier semis de points d'appui

(Stéréogrammétrie par scintillement)

Le programme *COORTOOL (B. Lortic), écrit d'après les normes du logiciel *PLANETES (M. Rakoto), est dérivé du logiciel *VISUTOOL. Son but est d'obtenir la liste des vecteurs de déplacement de chaque point d'appui. Chaque vecteur est noté sous la forme : Ydépart, Xdépart, Yarrivée, Xarrivée.

Ce programme provoque l'alternance rapide de deux images, que nous appellerons "scintillement", et la possibilité de commander un déplacement par le clavier numérique. La comparaison par scintillement est une méthode fréquente en photométrie. Les performances seront nettement meilleures en visualisation noir et blanc.

Pour des raisons de rapidité, il est préférable sur les machines SUN actuelles, de travailler sur des petites fenêtres de 256x256. Il est cependant nécessaire d'avoir une certaine surface de comparaison : une fenêtre de 16x16 par exemple n'est pas assez grande pour une bonne interprétation visuelle. COORTOOL est un programme interactif qui s'apparente aux programmes de jeux vidéo classiques.

Il s'agit pour l'opérateur :

- 1) de pointer avec le curseur sur l'image de référence un objet identifiable sur les deux images ;
- 2) de lancer le clignotement des deux images et d'apprécier le déplacement relatif de l'objet en x et en y ;
- 3) d'arrêter le clignotement (je n'ai pas encore géré les "événements" pendant le processus de clignotement), et de corriger le décalage par les flèches du clavier ;
- 4) de relancer le clignotement, etc. ;
- 5) lorsque l'opérateur constate la superposition parfaite du pixel pointé au départ de l'opération, une touche lui permet de stocker les coordonnées des deux pixels homologues.

On peut archiver, même sans entraînement visuel, une centaine de points d'appui par heure. Notons que cette méthode purement visuelle, permet à l'opérateur d'apprécier un déplacement d'un demi-pixel. C'est ainsi qu'il est plus agréable, parce que plus précis, de travailler sur un pixel doublé.

1.3. Création des surfaces de déplacement

A chaque point d'appui sont donc associés un déplacement en y (dans le sens du déplacement du satellite) et un déplacement en x (parallèlement à la direction oblique de la vision). L'interpolation des vecteurs de déplacement se fait en deux fois : d'abord en x puis en y. Nous utilisons actuellement le programme *FACETT (M. Rakoto), qui interpole les valeurs du semis de points d'appui par "tessellation de Voronoi".

Notons que la surface de déplacement en x est l'interpolation des valeurs de parallaxe. On constate la bonne concordance avec l'altitude lorsque l'on compare les courbes de niveau de la carte IGM, et celles issues de la comparaison des images SPOT.

1.4. Première rectification géométrique

Le programme *RECALAGE crée une nouvelle image dont la géométrie est donnée par les surfaces de déformation. La radiométrie du point de coordonnées (x,y) est la radiométrie du point (x + déplacement en x, y + déplacement en y) de l'image à recaler. La rectification géométrique est donc locale alors que les programmes de rectification géométrique habituels utilisent une fonction générale, les points d'appui inter-agissant les uns sur les autres.

Par analogie, le programme *DEGRE 1 utilise des surfaces de déplacement planes ; le programme *DEGRE 2 (M. Rakoto) utilise des surfaces qui sont des portions de coniques.

On recale donc l'image de droite sur l'image de gauche, et non pas dans ce cas-là les deux images sur une projection verticale. On constate visuellement que le couple d'images ainsi obtenu se superpose bien, excepté pour les accidents de terrain de fréquence élevée. Les grandes vallées sont bien marquées et le scintillement sur écran vidéo ne laisse apparaître que quelques "palpitations" des vallons de petite taille et des ravins de pente forte.

1.5. Affinage du tissu de points d'appui

Le même programme *COORTOOL va effectuer automatiquement, mais sous contrôle visuel, la stéréogrammétrie du nouveau couple d'images.

Soit un indice de dissimilarité calculé par la somme des valeurs absolues des différences des pixels homologues. Lorsque l'on calcule cet indice pour la position relative des matrices de $x-n$ à $x+n$, cet indice décroît pour passer par un minimum lorsque les pixels des deux matrices sont bien superposés. Si les images ont une radiométrie identique, cet indice s'annule totalement lorsque les deux matrices sont superposées correctement. Il suffit donc de noter les valeurs du déplacement n et m pour lequel l'indice est minimum. Le déplacement se calcule en effet en colonnes, mais aussi en lignes de $y-m$ à $y+m$.

Le programme *COORTOOL utilise la mesure d'un indice de dissimilarité sur une maille hexagonale de treize pixels. Pour un déplacement de plus ou moins cinq lignes et plus ou moins dix colonnes.

Il y a malheureusement beaucoup de confusion surtout, comme c'est le cas général, quand les images sont de dates différentes. Pour éviter ces erreurs, on introduit préalablement au contrôle visuel final, des conditions de validité de l'indice. On peut citer simplement le cas des grandes surfaces uniformes pour lesquelles l'indice sera presque toujours très faible : l'indice ne sera pris en compte que si la variance a une valeur forte.

52	53	145	65	35	89	89	89	189	89
	I-----I					I-----I			
34	I 134	34	134	I 34	87	I 89	188	88	I 88
	I			I		I			I
156	I 51	33	52	I 154	87	I 185	89	188	I 84
	I			I		I			I
59	I 26	56	152	I 53	185	I 89	87	84	I 186
	I-----I					I-----I			
125	58	59	57	154	85	86	84	87	89

Exemple de calcul de l'indice de dissimilarité sur une matrice de 3x3 ; l'indice est égal à 733 pour n et m égaux à zéro.

	I-----I												
52	I	53	145	65	I	35	89	89	89	189	89		
	I				I			I-----I					
34	I	134	34	134	I	34	87	I	89	188	88	I	88
	I				I			I				I	
156	I	51	33	52	I	154	87	I	185	89	188	I	84
	I-----I				I			I				I	
59		26	56	152		53	185	I	89	87	84	I	186
								I-----I					
125		58	59	57		154	85		86	84	87		89

L'indice est minimum, égal à 386, pour la position y-1 et x.

Précédemment le programme demandait à l'opérateur de choisir lui-même ses points d'appui. Grâce à cet indice, nous pouvons maintenant faire en sorte que le programme balaye l'image automatiquement. Après chaque écriture de coordonnées, le nouveau point se décale vers la droite avec un pas qui peut être variable. Le résultat du calcul du décalage est soumis au jugement de l'opérateur par scintillement des deux portions d'images autour des points d'appui.

Les valeurs de déplacement ainsi obtenues sont ajoutées à celles des surfaces de déformation précédentes par le programme *CALCVEC, calculant les vecteurs de déformations. L'interpolation des surfaces et le recalage des images sont réitérés. La surface de déformation donnant la parallaxe s'est affinée.

Cette méthode de recalage, économique compte tenu des tarifs pratiqués actuellement par les sociétés de services, présente une grande souplesse et permet à l'utilisateur de pratiquer différents niveaux de qualité de recalage. Elle repose principalement sur les performances exceptionnelles de l'oeil lorsqu'il travaille par scintillement.

Le contrôle visuel permet de mieux suivre les erreurs de détection de pixels homologues par l'indice de dissimilarité.

On constate par exemple que les agencements de pixels du tissu urbain se marient très bien lorsque l'on utilise de petites fenêtres. Les risques d'erreur sont apparemment faibles. A l'inverse, certaines formes de parcellaire agricole posent des problèmes importants. Il peut se produire d'abord une inversion des valeurs de luminance du même champ à cause d'un changement de saison. Il existe un grand nombre de zones de forme incertaine et au contenu mal transcrit par l'image satellite à l'échelle de travail. Les pixels de forêt, par exemple, sont difficilement "variables".

Mais, visuellement, on se rend compte aussi des limites de cette méthode sur un terrain comme celui de Quito., profondément entaillé de ravines abruptes. La pente y est souvent si forte qu'une des deux images ne montre pas le fond de

la vallée (cas des vallées nord/sud). Il peut se produire aussi, même si le fond est géométriquement visible, une confusion entre l'ombre propre (versant lui-même) et l'ombre portée sur l'autre versant qui ne permet pas de marier les pixels de fond de vallée (cas des vallées est/ouest).

2. VISUALISATION 3D

Le fait que la méthode de recalage développée implique la création d'une "maquette numérique de terrain" nous a amené à créer un certain nombre de sous-produits, comme la visualisation en perspective.

Le premier exemple va illustrer la souplesse de cette méthode visuelle : le recalage géométrique d'une image panchromatique avec l'image "éclairage", issue du programme *ECLAIR (D. Rémy), s'effectue très rapidement par la méthode décrite précédemment. En effet, la corrélation visuelle entre une image SPOT et une image de la projection verticale du modèle numérique de terrain (MNT) qui respecte l'azimut du soleil (dans ce cas 140°), est très facile par superposition des ombres.

Le programme *VISUCAVAL (B. Lortic) permet alors de visualiser l'image panchromatique en 3D. La luminance, produit de l'éclairage par le facteur de réflexion de la surface terrestre, nous est donnée par l'image satellitaire, la position de chaque point est donnée par le MNT créé par le logiciel *BABEL (M. Souris).

Il est possible aussi d'utiliser la surface de déformation comme une valeur des altitudes et l'imagerie multispectrale peut donner l'éclairage multispectral. La composition colorée classique en télédétection est alors présentée en relief.

3. INTEGRATION A LA BASE DE DONNEES

Pour intégrer de l'imagerie à la base de données géographiques, il importe de choisir la projection utilisée le plus fréquemment par les divers utilisateurs, car les transformations de projections ne sont pas encore envisagées.

L'opération de recalage effectuée précédemment entre deux images obliques, doit s'effectuer en vue du recalage entre les images et la cartographie disponible.

La digitalisation par l'AIQ d'une cartographie au 1:2 000 de l'IGM représentant les limites des pâtés de maisons dans une projection UTM. Les images SPOT et Thematic Mapper disponibles sont recalées sur cette projection.

Cette technique nouvelle ouvre un vaste domaine d'utilisation. Nous pouvons ainsi obtenir par l'imagerie satellitaire la densité de végétation. L'intégration à la base de données permet d'obtenir cette information pour chacun des îlots. De même, il sera relativement facile d'obtenir, pour chacun des îlots, le nombre de bâtiments construits entre deux dates d'acquisition d'images différentes.

Les fichiers images ont toujours des tailles très importantes. Sur la surface couverte par la ville de Quito, nous avons actuellement suffisamment d'images pour devoir créer une banque de 200 Mégaoctets :

- 96 Mo pour les images SPOT panchromatiques ;
- 24 Mo pour les multispectrales ;
- 70 Mo pour une image de Thematic Mapper.

Toute acquisition d'imagerie fera grossir la banque très rapidement.

Les fichiers images sont donc stockés séparément, éventuellement sur bande magnétique ou dans l'avenir sur disque vidéo. Par contre leurs fichiers de gestion seront constamment intégrés au système.

Ces fichiers de gestion doivent comporter :

- des renseignements propres à l'image : la date et l'heure, le type du capteur, la ou les fenêtres radiométriques, la résolution d'origine, les paramètres d'ensoleillement ;
- des renseignements liés à leur intégration à la base de données et au système de gestion : les coordonnées du point bas-gauche de l'image, la taille du pixel, la projection utilisée, etc.).

Le système de gestion de la banque de données indique à l'utilisateur quelles sont les images disponibles et la taille du pixel à utiliser, en d'autres termes le domaine de validité de l'image.

CONCLUSION

Nous voyons le caractère très empirique du processus de recalage d'images.

Choix d'un référentiel, immuable de préférence, outils de recalage indépendants de la technologie d'acquisition de l'image et des qualités intrinsèques et extrinsèques au système imageur.

Constatons que l'ensemble de la démarche est dictée par une croyance peut-être bien naïve : "la cartographie existante est un référentiel admissible".

Quelques questions restent en suspens :

- que faire en l'absence de cartographie ? D'autres équipes de l'ORSTOM nous donnent la solution : utilisons NOAA, CZCS, Météosat et une méthode de recalage qui tient compte des paramètres intrinsèques à l'outil ;
- comment recaler des données de grande échelle, par exemple une mission de photographie aérienne au 1:5 000 ème ?
- les données images sont-elles indépendantes les unes des autres ?
- doit-on utiliser certaines propriétés pour simplifier le travail de recalage ?

Divers points de la procédure de recalage décrite ci-dessus peuvent et vont être améliorés, par exemple par l'utilisation de l'algorithme de Viterbi, qui peut accélérer la détection des points homologues.

Mais dès maintenant, cette méthode très souple est utilisable par bien des équipes de recherche.

REFERENCES BIBLIOGRAPHIQUES

- W.E.L. GRIMSOM, 1981. *From images to surfaces*. MIT Press, 1981.
- Y. SHIRAI, 1987. *Three dimensional computer vision*. XII, 296p., 313 Fig.,
Spinger-Verlag.
- American society of photogrammetry, 1952. *Manual of photogrammetry*.
Washington DC.
- Bertil HALLERT, 1960. *Photogrammétrie. Principes de base. Levés généraux*.
340p., Mc Graw-Hill. Londres.

Première session : Capitalisation des données.

(Rapport de session, par Jean-Jacques LECHAUVE)

La troisième édition de SEMINFOR (séminaire informatique ORSTOM) a été ouverte par Monsieur Gérard WINTER, Directeur général.

A l'occasion de son exposé introductif, Monsieur WINTER apporte son soutien pour la réalisation de ces séminaires annuels, "*lieu de rencontre privilégié, permettant de favoriser le développement d'une pluridisciplinarité effective à l'Institut*".

Après avoir affirmé la nécessité des systèmes d'information pour le développement, Monsieur WINTER a manifesté quelques préoccupations relativement à chacune des trois sessions de ce séminaire.

A propos de la première, consacrée à la capitalisation des données, les réalisations de systèmes d'information, assurée par des spécialistes du fait de leur complexité sans cesse croissante, font-elles l'objet de toute l'attention nécessaire à garantir leur adéquation à la problématique des utilisateurs finaux ?

Quant à la deuxième session, intitulée '*De la donnée à l'information*', la question centrale est celle de l'accès à l'information. Nos partenaires sont de toute évidence propriétaires des données; les choix conceptuels et techniques qui président à la réalisation de ces systèmes d'information offrent-ils toujours à l'Institut des bases solides de négociation ?

Enfin, au cours de la dernière session traitant des stratégies de spatialisation, la présentation de SIG (système d'information géographique) doit favoriser une meilleure définition de ce concept à l'Institut. D'autre part, Monsieur WINTER suscite, à propos de ces systèmes, une réflexion autour des problèmes d'intégration pluridisciplinaire d'informations élaborées à des niveaux différents d'échelle.

La première session intitulée '*Capitalisation des données*' est introduite par C. MULLON et P. RENAUD.

Leur communication conjointe présente de nouveaux concepts et outils informatiques susceptibles d'apporter une réponse adaptée aux problèmes auxquels sont confrontés les chercheurs de l'Institut lors de la mise en place d'un système d'information. Deux contextes particuliers servent de cadre à cette présentation. Le premier est celui du développement d'une base de données naturalistes où les deux modèles relationnel et hypertexte offrent une alternative intéressante. Le second est celui de l'informatisation d'un projet pluridisciplinaire en l'occurrence HAPEX (Hydrological Atmospheric Pilot EXperimental), où trois démarches (pragmatisme, méthode, objet) sont comparées. Grâce à la mise en oeuvre de concepts novateurs, l'approche objet faciliterait une meilleure intégration de l'acquis informatique des disciplines partenaires et des nouvelles questions émanant de la conjugaison de ces dernières.

Cinq communications ont été présentées au cours de cette première journée devant un public assez large. Certes le planning relativement serré n'a pas toujours permis de longs échanges entre les participants. Cependant bon nombre de questions de fond ont été abordées et les débats confirment la nécessité de poursuivre et même d'intensifier l'effort dans les domaines suivants :

- l'application d'une méthodologie. C'est un impératif pour l'informatisation des grands systèmes de gestion de données ;

- l'intégrité des systèmes de données. L'analyse conceptuelle du système ne peut être réalisée que conjointement par les thématiciens et les informaticiens. La valeur de l'information dépend étroitement des contrôles de validité auxquels sont soumises les données ;

- les choix technologiques. La stratégie retenue doit favoriser la politique de transfert à nos partenaires; cela est favorisé actuellement par une informatique peu coûteuse et où la notion de standard est omniprésente ;

- l'ergonomie. Les thématiciens doivent et peuvent s'approprier facilement les outils développés.

Le débat a été plus particulièrement nourri à propos des questions suivantes :

La circulation de l'information :

A ce propos certains développements, expériences souvent enrichissantes à titre personnel, sont-ils toujours une nécessité ? Le manque d'information est trop souvent à l'origine de réalisations redondantes. Cette constatation suggère que le RIO favorise le rapprochement des concepteurs de logiciels. Le Chef de la MTI propose d'étudier la mise en place d'un système de 'news' et éventuellement d'une banque de programmes accessibles à tous par le réseau. Si techniquement cette réalisation peut être envisageable, elle nécessite cependant de trouver des disponibilités en personnel.

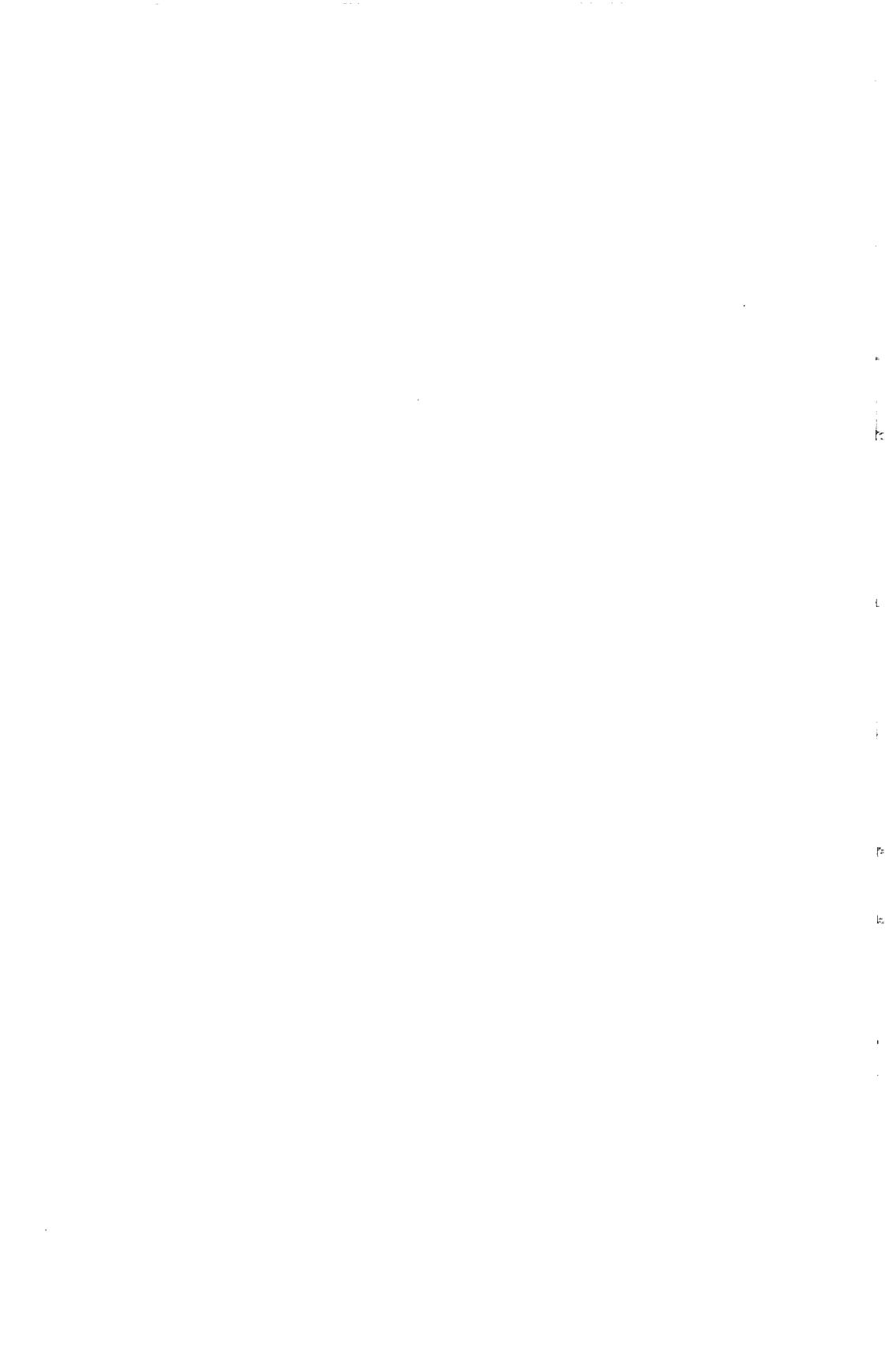
Il faut également évoquer l'opportunité qu'offre le réseau d'ouvrir plus largement les banques de données ORSTOM à la communauté scientifique. Si aucun exposé n'a abordé concrètement cette fonctionnalité, il existe néanmoins des réalisations qui y ont recours.

Le choix d'une méthodologie dans la conduite d'un projet de réalisation d'un système d'information :

La solution proposée par une des communications a fait l'objet d'une controverse. Le grief majeur est certes la rigidité des procédures de mise en oeuvre. Les contraintes de temps fixées par les programmes sont souvent inconciliables avec les délais inhérents au développement de la méthode. Même si les documents issus de la longue phase de conception présentent de l'intérêt, il s'avère que la phase d'implémentation est trop différée dans le temps. Il semble qu'il faille orienter ses choix vers des méthodes plus immédiates et interactives sans négliger pour autant le dialogue entre thématiciens et concepteurs.

Le choix technologique face aux exigences d'une politique efficace de transfert à nos partenaires :

Au cours du débat cette question a été formulée en d'autres termes; doit-on systématiquement développer des 'produits maison' ? Deux communications ont apporté une réponse. Dans les deux cas, le choix d'une réalisation interne a été justifié par les coûts élevés des solutions du commerce. Cependant le débat a permis d'entrevoir des positions nuancées vis-à-vis de cette réponse. Il semble qu'il faille séparer les applications en deux classes, celles qui mettent en oeuvre des techniques pointues qui relèvent encore plus ou moins du domaine de la recherche et les autres. Pour cette dernière, qui inclut la quasi-totalité des systèmes de gestion de données, il semble que le marché offre des solutions abordables financièrement et modulées suivant l'envergure des projets.



Session 2 :

**DE LA DONNEE A
L'INFORMATION,
AIDE A
L'INTERPRETATION**

Deuxième session : De la donnée à l'information, aide à l'interprétation.

(Présentation, par Jean-Yves MARTIN)

J'ai la charge d'introduire la deuxième session de ce Séminfor 3, session consacrée au thème "de la donnée à l'information, aide à l'interprétation". Je vais le faire brièvement. N'étant pas informaticien, je n'ai évidemment pas de commentaires à faire sur les aspects proprement informatiques des cinq communications qui seront présentées tout à l'heure. Je parlerai simplement en tant que Président de la CS7, Commission qui co-organise avec la Mission Technique de l'Informatique la série des Séminfor.

Vous savez que l'une des missions des Commissions Scientifiques est l'animation des disciplines dont elles ont l'apanage. Cela veut dire conforter la légitimité et la vitalité scientifiques de ces disciplines dans notre Institut. C'est un exercice difficile car, à l'ORSTOM, ces disciplines de la CS7 - l'IST, l'informatique et le traitement des données, la physico-chimie analytique - sont des disciplines sans objet. Sans objet parce qu'elles n'ont pas à développer de programmes propres de recherche, mais à intervenir en appui sur les objets des autres disciplines. Pour bien jouer ce rôle d'appui à la recherche, ses représentants - ici les *informaticiens* - doivent entretenir deux qualités : d'une part maintenir un haut niveau de compétence en étant les acteurs d'une veille scientifique et technologique, c'est-à-dire être les analystes du développement de l'informatique et de la technologie concernée au plan national et international et les diffuseurs - et les formateurs - de ces développements à l'intérieur de l'ORSTOM. D'autre part, celle d'être ouverts aux problématiques des autres disciplines pour une meilleure utilisation de l'outil informatique.

De ce point de vue, une manifestation comme celle qui nous réunit me paraît être un outil privilégié d'animation de la discipline informatique : nous y avons une information et un débat entre les divers utilisateurs - venant de diverses disciplines - de l'outil sur le thème choisi. Nous avons aussi des restitutions de la veille scientifique et technologique : l'exposé d'hier de Pascal Renaud et Christian Mullon en témoigne. Instrument d'animation, Séminfor est aussi un lieu de formation réciproque pour ses divers participants. Il est également pour la CS7 un support d'évaluation. Séminfor 2 avait permis de faire le point sur l'activité de modélisation. Nous espérons que Séminfor 3 permettra de faire un point qui rende bien compte de l'activité de l'Institut dans le domaine des systèmes d'information sur l'environnement et qu'il permettra d'alimenter la réflexion stratégique qui constituait l'une des dimensions de l'appel aux communications : dans quel sens et à quelles conditions orienter ces développements à l'avenir ? Cette réflexion trouvera sans doute un écho lors de la session de demain sur les Systèmes d'Information Géographiques. Mais les organisateurs appelaient aussi à une réflexion de nature technique et à une réflexion de nature méthodologique.

Pour revenir à la session d'aujourd'hui on peut dire que si il y a bien un dénominateur commun aux cinq communications que nous allons entendre, l'analyse et le traitement des données, chacune est vraiment singulière, ce qui témoigne aussi de l'inventivité et de la multiplicité des initiatives. Ainsi chacune met l'accent sur un aspect ou une question particulière qui situe bien *la nature et l'objet de la démarche qui a abouti à la création du système d'information présenté*. Ainsi Daniel Cardon a eu, avec SISBIO, le souci de mettre au point un système de gestion de données *fiables* pour des expérimentations. L'utilisation des données étant très ciblée (modèles de croissance et de production), il insiste beaucoup sur la question de la collecte.

La communication de Marc Souris et Dominique Rémy sur le logiciel BABEL insiste moins sur l'utilisation d'un système que sur les modalités de sa création en vue de résoudre des *problèmes de calcul et de représentation de relief* à trois dimensions pour générer des modèles numériques de terrain.

Les deux communications suivantes, celle de Marie-Madeleine Thomassin et celle de Philippe Waniez et Gérard Dandoy, sont parentes dans la mesure où elles traitent toutes deux de systèmes intégrant des variables localisées, mais leur problématique de départ diffère largement. La première manifeste la volonté de parvenir à des interprétations, et ce au prix d'une réelle sophistication de la partie calcul et analyse de correspondances.

Au contraire, peut-on dire, la communication Waniez - Dandoy manifeste le souci de produire, en l'absence de SIG, une informatique légère de recherche, permettant de croiser des données hétérogènes pour identifier des inter-relations en vue d'analyser une dynamique spatiale.

Enfin, la communication de L. Tito de Morais nous parle d'une application de l'intelligence artificielle à l'aide à la détermination de poissons d'eau douce. Tout comme celle de Waniez et Dandoy, mais sur un registre différent, elle débouche sur une réflexion stratégique : quel est l'intérêt et le coût humain, dans une optique de transfert à des partenaires, de la réalisation de bases de connaissances et de systèmes-experts en écologie ? Et pour les précédents : est-il du rôle de la recherche de prendre l'initiative de la constitution du SIG ?

Donc cinq communications à entendre ce matin, cinq démarches originales, et autant de questions.

UTILISATION DE L'INFORMATIQUE DANS LA GESTION ET LE TRAITEMENT DE DONNEES BIOLOGIQUES

(Daniel CARDON)

RESUME - Dans le cadre de la convention ORSTOM-IRENA, une série d'expérimentations sur maïs destinées à vérifier la validité d'un modèle de production est en cours. Trois types de collecte sont opérées, comprenant des mesures de surface foliaire et de pois des différentes parties de la plante. SISBIO est un système informatique créé pour assurer la gestion des données correspondantes.

SISBIO prend en charge toutes les opérations de cette gestion, depuis la collecte jusqu'à l'exploitation des données de l'expérimentation, c'est à dire la création de la feuille de collecte, l'informatisation des données, la critique et la correction de celles-ci, ainsi que leur utilisation.

La façon dont SISBIO s'acquitte de cette tâche dépend de la conception générale du système et des priorités qui lui sont attachées. Pour remplir pleinement son rôle, le programme doit être simple d'utilisation mais, suivant la fonction des utilisateurs, le bien fondé de cette appréciation peut varier notablement. Trois utilisateurs potentiels ont été identifiés : l'opérateur qui saisit les données, le responsable scientifique qui conduit l'expérimentation et l'utilisateur des données. Vu sous l'angle de ces trois utilisateurs, la conception de SISBIO a été guidée par quatre grands principes : la rigueur, la sécurité, la simplicité et la souplesse.

La présente communication expose successivement la conception et les objectifs du système, ainsi que les techniques adoptées pour les atteindre. L'utilisation pratique de SISBIO est également présentée, en suivant pas à pas toutes les étapes d'une expérimentation.

SISBIO a été testé durant un cycle de culture et s'est révélé un instrument de travail précieux, tant par le côté complet de gestion que par sa souplesse d'utilisation, ou encore par l'aspect de sécurité des données. Toutefois, si on veut l'utiliser avec d'autres plantes que le maïs se cultivant en ligne, il sera nécessaire d'intervenir directement dans la programmation du système.

INTRODUCTION

Dans le cadre de la convention IRENA-ORSTOM au NICARAGUA sont en cours de réalisation une série d'expérimentations sur maïs destinées à la vérification d'un modèle de croissance et de production. Ces expérimentations comprennent évidemment des mesures biologiques. L'objet de ce travail est la description de leur gestion et de leur traitement à l'aide de l'ordinateur.

Les données collectées, le mode de collecte et le nombre de répétitions correspondent à un mode opératoire précis mis au point à l'aide de résultats préliminaires (CARDON janvier 1988). Une présentation succincte en sera faite dans le premier chapitre.

L'utilisation de l'informatique s'est effectuée en deux étapes :

- la gestion des données brutes c'est à dire la collecte, l'informatisation, la critique et le pré-traitement de ces données à l'aide d'un logiciel créé à cet effet (SISBIO) ;

- leur traitement proprement dit avec un logiciel du commerce (LOTUS 123).

Ces deux phases seront examinées successivement dans le second chapitre.

Ce système de travail a été testé de juin à octobre 1988, un essai sur maïs ayant été réalisé durant cette période. Le dernier chapitre sera consacré à la façon pratique de l'utiliser et à son analyse critique en fonction des résultats obtenus dans l'essai précité - pour plus de précision voir (CARDON avril 1989) - et des difficultés rencontrées.

1. PRESENTATION DES DONNEES

Avant la mise au point de la méthode de travail présentée, préexistaient un certain nombre de résultats issus de travaux antérieurs. Leur étude a permis de déterminer la variabilité approximative de chaque mesure et, par voie de conséquence, le nombre de répétitions nécessaires à l'obtention d'une précision donnée. Dans certains cas elle a également autorisé une simplification appréciable du protocole expérimental. Enfin la connaissance pédologique du terrain expérimental a dicté l'emplacement des parcelles d'essai et le mode d'échantillonnage.

Suivant les cas, les mesures peuvent être ou ne pas être destructives. On a donc considéré différents types d'expérimentation qui seront exposés successivement après présentation des données.

1.1. Données recherchées

Elles sont de trois types : des mesures de poids, des mesures de longueur et des mesures de surface foliaire. Les mesures de poids concerne les différentes parties de la plante c'est à dire racines, thalle, feuilles et fruits. Les mesures de longueur sont la hauteur de la plante, la profondeur racinienne et les dimensions du thalle et des feuilles, ces dernières permettant d'obtenir la surface foliaire.

1.2. Expérimentation destructive

Elle concerne la mesure du poids sec des différentes parties de la plante à l'exception du poids racinienne. En outre, compte tenu de leur intérêt économique ont été mesurés en fin de cycle, les poids sec et humide de grain. Chaque parcelle est divisée en six blocs et dans chaque bloc sont effectuées deux répétitions en des points tirés au hasard. Le poids racinienne est obtenu en appliquant une fonction à la somme des poids des autres parties de la plante. Cette fonction est un des produits de l'étude mentionnée antérieurement.

1.3. Expérimentation non destructive

C'est la mesure des dimensions de la plante, soit pour chacune d'elles sa hauteur, les dimensions du thalle, et pour les feuilles N° 9, 12, 15 la longueur et la plus grande largeur. Pour chaque feuille un coefficient multiplicatif permet d'obtenir sa surface à partir des dimensions. Pour avoir la surface foliaire de la plante il suffit d'appliquer un second coefficient à la somme des surfaces des trois feuilles considérées (CARDON janvier 1988). Chaque parcelle est divisée en 6 blocs et dans chaque bloc sont effectuées 5 répétitions en des points tirés au hasard.

1.4. Expérimentation de vérification

Elle concerne l'expérimentation non-destructive. Les deux coefficients appliqués dans la mesure de la surface foliaire ont été obtenus dans l'étude ci-dessus référencée. Le premier est très voisin de celui trouvé par d'autres auteurs (MORREY & TOGOLA 1985) et peut donc être considéré comme un résultat acquis. Le second est beaucoup moins classique, il est donc nécessaire de vérifier sa stabilité. Dans chaque parcelle un emplacement est donc tiré au hasard à intervalles réguliers, et la plante qui s'y trouve étudiée de façon complète. La stabilité des coefficients est vérifiée en fin de cycle sur l'échantillon constitué par l'ensemble de toutes les plantes. Cette étude limitée est mise à profit pour effectuer une autre vérification : celle de la profondeur racinienne.

2. PRESENTATION DU SYSTEME DE TRAITEMENT DES DONNEES

Dans une présentation générale on exposera les objectifs poursuivis en créant SISBIO ainsi que la composition de ce système, ensuite on détaillera le rôle de chacun de ses programmes pris en particulier, enfin on considérera la fin du traitement des données dans LOTUS.

2.1. Présentation générale de SISBIO

2.11 Objectifs de SISBIO

Ces objectifs sont de deux ordres : ceux ayant trait aux qualités informatiques recherchées dans le système et le rôle de SISBIO dans le traitement des données biologiques.

Les premiers étaient en "filigrane" dans un travail présenté antérieurement (CARDON & al 1987), et on les retrouve clairement explicités dans un autre système (SISOND) présenté au cours du même séminaire. Quatre grands principes ont guidés la conception de SISBIO :

- la rigueur imposée à l'utilisateur ;
- la sécurité ;
- la simplicité d'emploi ;
- la souplesse.

SISBIO a été créé pour assurer la gestion des données brutes définies dans le premier chapitre. Par gestion on entend l'ensemble des opérations suivantes :

- la création de la feuille de collecte avec mention de toutes les indications nécessaires, particulièrement de l'emplacement des mesures ;
- l'informatisation ;
- la critique ;
- la correction éventuelle ;
- le pré-traitement des données.

Les calculs à effectuer sur les données a été scindé en deux parties : ceux qu'il faudra faire de toute manière et ceux dépendants de l'orientation donnée au travail de recherche. Ce sont les premiers qu'on a appelés le pré-traitement.

Comme le montreront les paragraphes suivants, les directrices précédentes ont souvent induites des solutions en tous points comparables à celles adoptées dans SISONND. Il y a cependant une différence importante due pour part au sujet traité, pour part à la chronologie adoptée dans la réalisation des deux systèmes (SISBIO est antérieur) : au contraire de SISONND utilisable dans la plupart des expérimentations avec sonde à neutrons, SISBIO ne sert que pour les trois types d'expérimentations définis dans le premier chapitre. Certes un paramétrage judicieux permet d'adapter assez facilement les programmes utilisés à une plante ayant par exemple un autre potentiel foliaire que le maïs pourvue qu'elle soit cultivée en ligne comme lui. Mais de toutes façons il faut intervenir directement dans la programmation.

2.12 Présentation générale

A l'appel de SISBIO apparaît un menu général invitant l'utilisateur à choisir l'une des options suivantes :

- programme d'installation ;
- programme de BACKUP ;
- description d'expérimentation ;
- création de registres dans les fichiers de données et impression des feuilles de collecte ;
- informatisation des données ;
- critique des données ;
- pré-traitement des données ;
- fin.

A chaque option correspond un caractère. Après avoir rentré ce caractère au clavier se déroule le programme, choisi puis il y a retour au menu général (sauf pour la dernière option).

Deux types de fichiers sont utilisés par ces programmes :

- les fichiers servant à stocker les données entrées au clavier dans le système, qui sont à accès direct, appelés fichiers de données brutes ;
- les fichiers créés par programme à partir des premiers, en caractères ASCII et à accès séquentiel, ou fichiers de données élaborées.

Les deux premiers programmes sont une copie à peine modifiée de leurs équivalents dans SISONND. Leur rôle est principalement de protection. En particulier il n'est pas possible de sortir du système sans sauver les fichiers de données brutes modifiés au cours de la session de travail.

Les autres programmes constituent le système proprement dit. Comme il y a trois types de données, chacun d'eux est divisé en un tronc commun identique pour toutes, et trois parties, chacune spécifique d'un type de données. Ils fonctionnent suivant le schéma général suivant :

- sur la vidéo apparaît un écran explicatif avec un certain nombre de champs de données à documenter ;
- pour pouvoir poursuivre il faut que ces champs soient remplis de façon acceptable sinon apparaissent des messages d'erreur ;
- on peut passer à l'écran suivant ou revenir à l'écran antérieur à condition bien sûr qu'il y en ait un ;
- dans tous les cas on peut revenir au menu principal.

Certaines opérations sont irréversibles (création de fichiers ou de registres dans les fichiers de données brutes, etc.) et par conséquent comportent un risque. Pour limiter ce risque elles sont protégées par un mot de passe dont la gestion est assurée par le programme d'installation. Si celui-ci n'est pas connu, l'opération est annulée et, suivant les cas, il peut y avoir retour au menu principal.

Autant que faire se peut, les causes d'erreur lors de la documentation des écrans ont été limitées (choix d'option par le caractère "X", apparition de la liste des réponses possibles dans une fenêtre, etc.). Dans tous les cas des explications circonstanciées apparaissent en grisé pour faciliter la documentation des champs de données. Dans le cas de l'informatisation des données brutes, c'est l'image exacte de la feuille de collecte qui apparaît avec indication des noms du fichier et du registre afin d'éviter toute confusion. Dans ce cas et sur option choisie en début de programme, l'image de la feuille de collecte apparaît ensuite une seconde fois pour permettre à l'utilisateur d'ajouter un code à chaque donnée.

2.2. Présentation des programmes de SISBIO

De cette présentation sont exclus les programmes d'installation et de BACKUP. Ils sont examinés un à un dans leur ordre logique d'utilisation.

2.2.1. Description d'expérimentation

Jusque neuf expérimentations peuvent être décrites simultanément. Dans un premier écran, l'utilisateur choisit les numéros des descriptions d'expérimentations à documenter, redocumenter ou effacer. Ensuite, pour chacune d'elles, il documente les données suivantes :

- une description sommaire ;
- le nom en 5 lettres du fichier de données correspondant ;
- le type de ces données ;
- la date du semis.

Pour chaque bloc, il faut enfin indiquer les numéros des rangées extrêmes qui constituent sa partie utile, et la distance en mètres qui sépare les extrémités utiles du bloc du début de chaque rangée.

2.22. Création de registres dans les fichiers de données

Ce programme a un double rôle puisqu'il permet également l'impression des feuilles de collecte. En apposant un "X" dans le champ contigu à leur description sommaire, l'utilisateur sélectionne les expérimentations avec lesquelles il désire travailler. Ensuite il choisit le type de tâche à effectuer (création ou impression), et remplit la date correspondante. Après vérification de cette dernière, le système effectue la tâche désirée. En particulier il tire au hasard l'emplacement des plantes à étudier, et, par voie de conséquence, ces informations sont les premières données informatisées.

2.23. Informatisation des données brutes

L'utilisateur reporte les noms de fichier et de registre des données qu'il désire informatiser, et, en option, indique s'il veut leur associer un code. Sur les écrans à venir apparaissent la ou les feuilles de collecte correspondantes. La documentation s'effectue dans l'ordre de la collecte.

2.24. Critique des données brutes

L'utilisateur communique au système le nom du fichier qu'il désire critiquer et éventuellement les noms des registres de début et de fin de critique (les extrémités de fichier constituent l'option par défaut). Un rapport de critique est alors imprimé. Chaque fois que le programme rencontre dans un bloc une donnée douteuse, il imprime toutes les données du bloc dans l'ordre de collecte et souligne la ou les données douteuses. De plus, au début de chaque trait, apparaît un numéro indiquant le type de contrôle effectué (entre limites extrêmes, contrôle de données entre elles, etc.).

2.25. Création de fichiers de données élaborées

Dans un premier écran, l'utilisateur choisit un nom de fichier à traiter et éventuellement les noms des registres de début et de fin de traitement. Ensuite il indique le nom du ou des fichiers créés ainsi que le *directory* dans lequel ils devront être gravés. Enfin une option permet de décider si les registres de ces fichiers se rapporteront aux données relatives à un bloc ou à une seule plante. Dans tous les cas, chaque registre portera mention des noms du fichier de données brutes et du registre traités, de la date, du numéro du bloc et du nombre de jours écoulés depuis le semis.

Dans le ou les écrans suivants, on a la possibilité de discriminer certaines données individuellement, en utilisant les codes (par exemple les données de code "D" souvent utilisé pour désigner les valeurs douteuses), et/ou géographique. En effet il est possible dans les limites de la parcelle de délimiter jusque 9 rectangles dont la superficie sera exclue ou, au contraire, seule retenue pour étudier les données.

2.3. Traitement des données sous LOTUS

Dans un travail de recherche, il faut bien entendu disposer d'un fil conducteur en fonction duquel on détermine les paramètres à mesurer et les expérimentations pour les obtenir. Néanmoins, au niveau du traitement des données, il y a toujours des surprises et il est très difficile sinon impossible de prévoir tous les développements d'une étude afin de les intégrer à un système informatique. Pour cette raison, la fin du traitement des données biologiques est effectuée dans un logiciel du commerce beaucoup plus souple pour ce genre de travail. On disposait de LOTUS 1-2-3, ce logiciel permet des sorties graphiques particulièrement claires, ce sont les raisons pour lesquelles il a été utilisé.

Le risque d'erreur augmente avec le nombre d'opérations effectuées. En particulier il faut en faire plus d'une trentaine (titres, légendes et unités comprises) pour créer un graphique sous LOTUS. Il est très contraignant de se rappeler tous les détails d'une présentation adoptée, surtout si on réalise des graphiques avec des données de nature différentes. Les instructions utilisées sous LOTUS peuvent être regroupées pour former de véritables programmes, qu'on peut tester, éliminant ainsi le risque d'erreurs et assurant l'homogénéité de la présentation des produits. Ces instructions présentent certes tout les éléments d'un véritable langage informatique (boucles, appel de routines etc...), mais elles n'ont ni la richesse, ni la clarté ni l'environnement de langages évolués tels que PASCAL, BASIC ou FORTRAN. On a donc réservé la création de tels programmes aux cas où la répétitivité et/ou le risque d'erreur le justifiaient.

La vérification de la méthodologie s'effectue avec un nombre réduit de données donc sans programme, et il serait trivial d'expliquer l'usage fait de LOTUS dans ce cas. Pour chacun des deux autres types de données, il existe deux formes de fichier de sortie de SISBIO. Des programmes d'importation adéquats les transforment et, en vue des traitements ultérieurs, les mettent sous une forme unique dans une page LOTUS. Tous les programmes sont préécrits dans une page de travail de LOTUS sans données. Pour les utiliser on importe cette page, on la change de nom, on importe dans la page les données à traiter qu'on utilise avec l'un ou l'autre des deux programmes graphiques. Sur la figure 1 on a représenté le plan sommaire de la feuille de travail avec la localisation approximative des zones réservées. La zone I est une colonne dont chaque cellule occupe 70 caractères. Elle est divisée en trois parties. En I-a on trouve le mode d'emploi des programmes, en I-b la signification de certains choix faits en cours de programme, en I-c- le texte des routines et programmes. Dans la zone II les données importées sont traitées. Dans la zone III sont stockées les bases de données LOTUS. La zone IV contient certaines variables nécessaires à la bonne marche des programmes. La zone V est réservée aux données à représenter graphiquement. Les codes apposés en haut et à gauche de chaque zone, représentent la position de début de la zone dans la page de travail.

Les programmes d'importation de données présentent en commun les caractéristiques suivantes :

- ils permettent d'importer en même temps un nombre quelconque de fichiers créés par SISBIO, à condition qu'ils soient du type requis ;

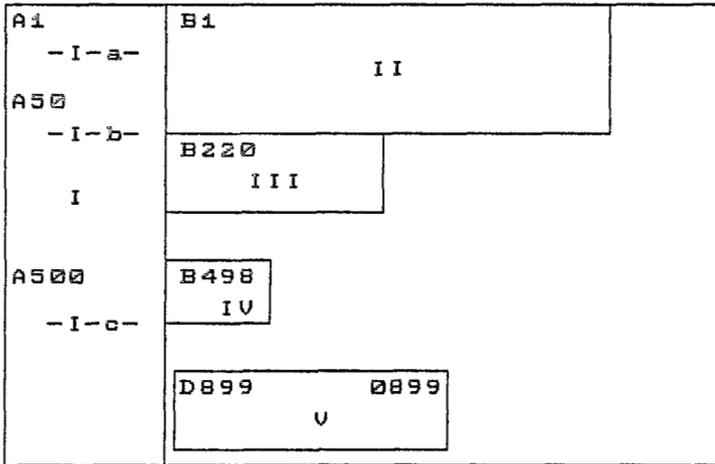


Fig.1 Schéma général de la feuille de travail avec LOTUS

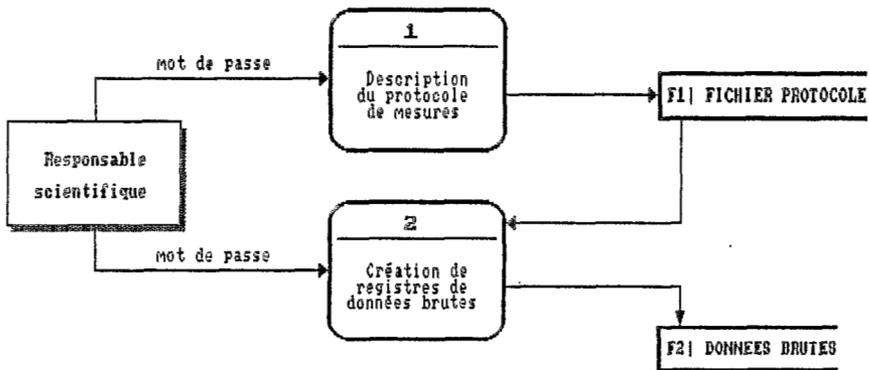


Fig.2 Diagramme de flot de données de SISBIO

Point de vue du responsable de l'expérimentation

- ces fichiers transformés constituent une base de données ;
- on peut créer au maximum quatre bases de données dans une page de travail ;
- les valeurs figurant dans les bases de données sont les moyennes par journée de collecte. En plus figurent des valeurs calculées comme le L.A.I. (surface foliaire) dans le cas des surfaces foliaires, ou des grandeurs repérées (maxima, minima, etc.).

Une difficulté imprévue a dû être surmontée avec tous ces programmes. Lorsqu'on importe avec LOTUS des fichiers numériques externes, le blanc entre deux données est interprété comme une séparation. Pour éviter un décalage dans les données, lorsque l'une d'elles est absente, on l'a remplacée par une valeur absurde. Une fois importée par LOTUS, cette dernière est considérée comme une donnée normale qu'il faut cependant discriminer. De nombreuses facilités de LOTUS (moyennes, écarts-types, etc.) ne peuvent plus alors être utilisées et la programmation évidemment compliquée.

Les programmes graphiques sont de deux types :

- ceux comparant différentes grandeurs de la même base de données ;
- ceux comparant une même grandeur suivant différentes bases de données.

Toutes les représentations se font en fonction du nombre de jours écoulés depuis le semis. Les premiers en particulier permettent de suivre l'évolution simultanée dans le temps pour chaque paramètre : de sa valeur, de son écart-type et de ses valeurs maxima et minima.

3. UTILISATION ET ANALYSE CRITIQUE DU SYSTEME DE TRAVAIL PRESENTE

Le fonctionnement de SISBIO n'est pas simple, pour cela on lui a consacré tout le premier paragraphe. Ensuite on a considéré l'utilisation du système dans son ensemble pour le traitement des données. Dans le dernier paragraphe, on s'appuie sur les données déjà traitées pour porter un jugement.

3.1. Fonctionnement de SISBIO

Pour rendre plus clair le fonctionnement de SISBIO on a considéré l'existence de trois utilisateurs potentiels : le responsable scientifique, le responsable des mesures de terrain et l'utilisateur des données. Il s'agit en fait d'une division un peu arbitraire et la même personne peut très bien remplir les trois fonctions. Les rôles respectifs des deux premiers utilisateurs sont schématisés dans les figures 2 et 3.

Le responsable scientifique conçoit l'expérimentation et la décrit à l'aide du programme N° 1. Ensuite il décide du rythme des mesures et des dates de collecte avec le programme N° 2. Ces deux tâches sont protégées par le mot de passe.

SISBIO

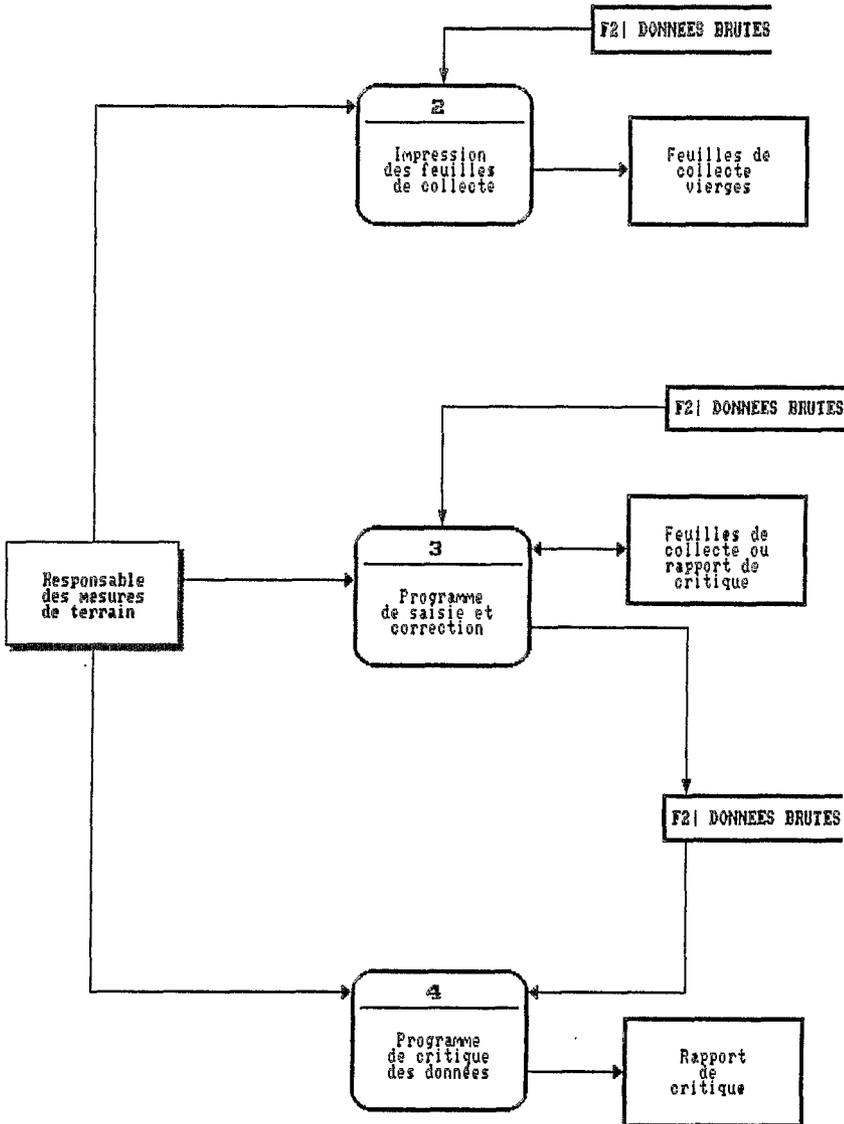


Fig.3 Diagramme de flot de données de SISBIO

Point de vue de l'observateur de terrain

SISBIO

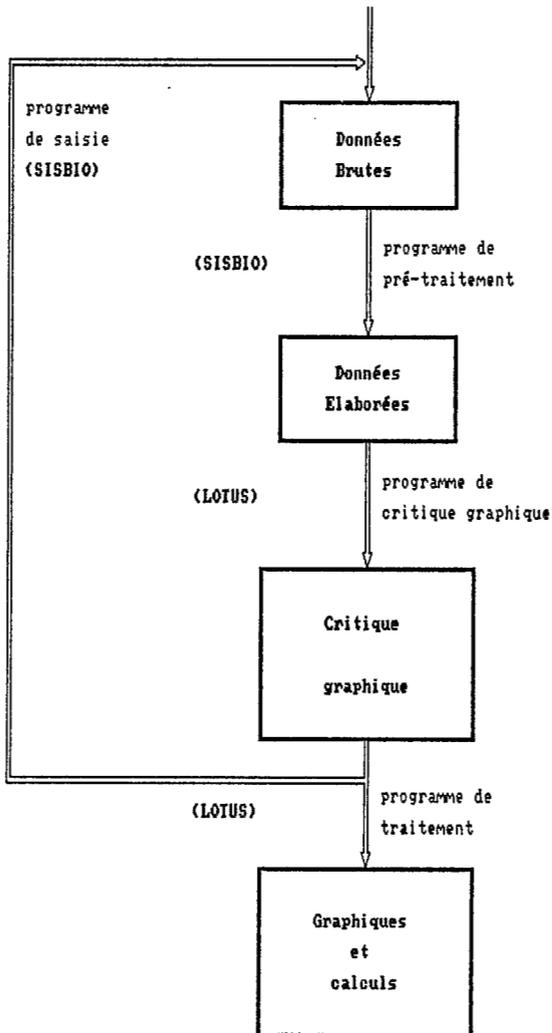


Fig.4 Schéma général de traitement des données

Le responsable des mesures est chargé de toutes les tâches de routines. Chaque fois que le responsable scientifique prévoit une collecte de données il réalise les opérations suivantes :

- avec le programme N° 2 il imprime les feuilles de collecte qui serviront de support à son travail de terrain,
- après chaque collecte il procède à l'informatisation des données à l'aide du programme N° 3.

A intervalles réguliers il procède au contrôle des données en procédant comme suit :

- il utilise le programme N° 4 pour obtenir un rapport de critique de ces données,
- après vérification, il procède à la correction et éventuellement à l'apposition de codes à certaines données, à l'aide du programme N° 3.

Cette dernière opération peut également faire partie des attributions du responsable scientifique.

L'utilisateur des données utilise SISBIO pour créer, à partir des données brutes vérifiées, les fichiers de données élaborées qu'il utilisera ensuite sous LOTUS.

3.2. Fonctionnement général du système

On ne reviendra pas sur l'obtention des données brutes et leur vérification déjà vues antérieurement. Ce paragraphe se confond avec le traitement des données, il est résumé par la figure 4. Celui-ci se passe en deux phases :

- une phase de jugement porté sur les données ;
- la phase d'exploitation proprement dite.

La première phase débute par la création de fichiers de données élaborées utilisant la totalité de fichiers de données brutes. Ces fichiers sont ensuite importés sous LOTUS un par un et/ou par groupes choisis en fonction du dispositif expérimental et des problèmes rencontrés en cours d'expérimentation. Ensuite on les analyse, paramètre par paramètre, en utilisant le programme de critique graphique sous LOTUS, afin de repérer les éventuelles singularités et de tenter de les expliquer. A la différence de la critique effectuée par SISBIO ne faisant intervenir à la fois que les données d'une seule collecte, celle-ci met en relief les anomalies dans l'évolution dans le temps des valeurs prises par les différents paramètres. Elle permet donc une analyse beaucoup plus fine.

La deuxième phase débute par un retour à SISBIO. Eventuellement et en fonction des résultats précédents, on associe à certaines données de nouveaux codes. Ensuite on prend une décision quant aux données à utiliser dans les fichiers de données élaborées. Ces fichiers sont alors importés sous LOTUS. La suite du travail se fait à l'aide des programmes existant lorsqu'il s'agit d'une tâche courante. Dans le cas contraire les opérations sont réalisées directement dans ce logiciel.

3.3. Critique du système de travail

Les données biologiques sur maïs obtenues au cours d'un essai de saison des pluies, réalisé dans le cadre du programme cité en introduction, ont été analysées par cette méthode. Cet essai comprenait trois parcelles, deux en maïs pluvial avec des densités de peuplement différentes, la troisième avec la plus forte des densités précédentes et irrigation d'appoint. L'essai s'est déroulé de juillet à fin octobre 1988. Un rapport interne a été rédigé (CARDON avril 1989), auquel on pourra se reporter pour plus de précision.

Les conditions expérimentales (mauvaises herbes mal contrôlées, maladies, vols mêmes) ont créé une hétérogénéité peu compatible avec une bonne précision. La situation du terrain expérimental, en partie en contrebas d'une route régulièrement inondée, a également été une source de problèmes.

L'utilisation des codes associés aux données brutes a permis de vérifier l'influence de chacun de ces facteurs et éventuellement d'éliminer les plantes leur correspondant. Par exemple, il est arrivé qu'une plante soit sans épis de maïs (vol, etc.). Sur un échantillon de douze données au maximum, le poids moyen de fruits présentait une singularité suffisante pour être détectée par le programme de critique dans LOTUS et un code adéquat permettait d'éliminer ensuite cette plante de l'analyse.

Le problème des inondations a été circonscrit de la manière suivante :

- à chacune d'elles on a repéré son étendue approximative ;
- dans la parcelle touchée on a fait l'analyse des données en considérant tout d'abord leur ensemble puis ensuite seulement les parties non inondées.

Les résultats ont montrés que, pour les mesures biologiques, l'incidence des inondations n'était pas déterminante.

Au niveau de l'utilisation de SISBIO même, celui-ci s'est montré un système sûr, pratique d'emploi même pour des utilisateurs n'ayant aucune notion d'informatique. Le système dans son ensemble s'est révélé très souple et a permis, avec un minimum de travail, de récupérer des données qui, autrement, seraient restées problématiques. On peut donc globalement s'estimer satisfait. Deux points cependant, dans l'utilisation de LOTUS, méritent réflexion :

- la mise au point des programmes sous LOTUS, même simples, exige beaucoup de temps et d'attention ;
- le temps de traitement est considérable, et on peut se demander s'il ne faudrait pas réduire l'utilisation de ce logiciel.

CONCLUSION

Cette conclusion est double car on y sépare les deux phases du traitement informatique soit la gestion des données brutes et le traitement des données qui en sont issues.

SISBIO s'est révélé un outil précieux, notamment par le côté complet de sa gestion des données. Ainsi, au cours d'une même journée, il permet d'imprimer les feuilles de collecte, et, une fois le travail effectué, d'informatiser les données et

de les vérifier. Ce travail, en temps presque réel, simplifie la gestion des dates de collecte et permet, quand il en est encore temps, de détecter les erreurs de manipulation. SISBIO s'est révélé également un système très sûr, tant au niveau des fichiers créés que de l'informatisation et de l'utilisation des données. Enfin sa souplesse d'utilisation lui permet à volonté de supprimer des fichiers de sortie les données correspondant à certaines plantes, à certains blocs ou même à une partie quelconque de la parcelle. Ceci a d'ailleurs eu une application dans l'essai analysé : une inondation inopportune ayant affecté une partie d'une parcelle d'essai, on a pu en mesurer l'effet ou plus exactement l'absence d'incidence réelle. Une insuffisance notoire est cependant à déplorer : son manque de souplesse au niveau de la définition du protocole de travail. SISBIO a été créé pour faciliter un travail déterminé. Certes, par un paramétrage judicieux, on peut à peu de frais modifier le système et l'adapter à une autre expérimentation pourvu qu'il s'agisse de plantes poussant en ligne. Cette tâche cependant ne peut être effectuée sans modifier le "source" du système. SISBIO n'est donc pas un produit "exportable".

L'importation sous LOTUS des fichiers de sortie de SISBIO permet d'achever le traitement des données et d'obtenir des graphiques très clairs directement utilisables dans des rapports ou publications. L'utilisation de programmes écrits sous LOTUS et ensuite testés, a éliminé toute erreur de traitement et permis d'homogénéiser les documents de sortie. Cette utilisation de LOTUS a également permis de mieux cerner les limites de ce type de logiciel. Tout d'abord le traitement est lent (jusque 45 minutes si on utilise le programme d'importation avec toutes les données relatives à une expérimentation). Ensuite l'écriture, la correction et la modification des programmes sous LOTUS s'est avérée particulièrement malaisée. Un programme de plus de 100 instructions devient très vite inextricable alors que des programmes 10 fois plus performants écrits en PASCAL ne présentent aucun problème de cet ordre. En fait LOTUS n'apparaît pas avoir été vraiment conçu pour ce type de travail, le problème du zéro donnée nulle ou donnée absente en est un exemple, et on peut se demander si son utilisation s'impose quand on dispose d'un langage informatique puissant tel que le PASCAL ou le BASIC. C'est pourquoi, pour le futur, les traitements effectués actuellement sous LOTUS dans les programmes d'importation, le seront à l'aide de programmes extérieurs écrits en PASCAL et non intégrés à SISBIO. L'importation des données sous LOTUS se fera lors de la toute dernière phase du traitement, celle de la représentation graphique, pour laquelle ce logiciel reste difficilement remplaçable.

REFERENCES BIBLIOGRAPHIQUES

- CARDON, D. A.; SECHET, P.; DINIZ, T. D. de A. S. & MALTEZ, M. G. M. - 1987 - *SISCLIMA : capitalisation de données agroclimatiques au Brésil*. Hydrol. Continent., vol 2, N° 1, :3-14.
- CARDON, D. A. - 1988 - *Mesures biologiques sur maïs - mise au point de la méthodologie* - Rapport multigraphié, 10p. 6 fig. janvier 1988
- CARDON, D. A. - 1989 - *Essai de saison des pluies - Mesures biologiques*. Rapport multigraphié, 29p. avril 1989.
- CARDON, D. A. - 1989 - *SISOND : Une conception de l'informatique au service de l'utilisateur*. à paraître dans le 3ème Séminaire Informatique de l'ORSTOM, Séminfor 3. Septembre 1989.
- MORRET J. & TOGOLA, D. - 1985 - *Effet et arrière effet de la sécheresse sur la croissance de plusieurs génotypes de maïs*. In : Les besoins en eau des cultures. Conférence internationale, Paris, 11 - 14 sept. 1984. INRA, Paris, 1985 :352-359.

MODELES NUMERIQUES DE TERRAIN : LE LOGICIEL BABEL

(Marc SOURIS & Dominique REMY)

RESUME - Le calcul et la représentation du relief en trois dimensions est un outil important dans beaucoup de domaines, comme la géographie bien sûr, la géomorphologie, la géologie, l'hydrologie, etc.. En dehors de l'aspect purement technique (calcul des pentes, ruptures, écoulements, etc.) la simple visualisation en 3D, sous l'angle et l'incidence voulus, apporte un complément pédagogique important lorsqu'il s'agit de phénomènes spatiaux faisant intervenir le relief. C'est surtout dans ce but que nous avons développé le logiciel Babel, programme permettant la création de modèles numériques de terrain et leur visualisation sous forme de bloc-diagrammes.

1. LA CREATION DU MODELE

Quelles données altimétriques conserver pour rendre compte du relief ? Dès que l'on ne répond pas "*l'altitude de tous les points*", on doit résoudre le problème du "*quoi garder*" (problème de la modélisation) et du "*comment retrouver les points inconnus*" (problème de l'interpolation). Plusieurs réponses ont été apportées, chacune constitue un modèle numérique de terrain (MNT) :

- le support (les points que l'on connaît et que l'on garde) est un canevas régulier, à l'intérieur duquel on applique, pour interpoler, une fonction donnée (qui peut être bilinéaire, biquadratique, bicubique sur maille rectangulaire, linéaire sur triangulation isocèle, etc.) ;

- le support est un canevas irrégulier à l'intérieur duquel on applique une fonction donnée (généralement bilinéaire sur triangulation) ;

- le support est un canevas irrégulier, à partir duquel on applique une fonction de type barycentrique ;

- le support est un canevas irrégulier, à partir duquel on applique un modèle statistique permettant d'évaluer l'altitude la plus probable en un point quelconque (Krigage).

Nous appellerons "*points de référence*" les points du plan qui servent de support au modèle et permettent l'interpolation. Bien sûr, ces points ne sont pas quelconques : ils doivent être choisis en fonction du modèle utilisé, même dans le cas d'un canevas irrégulier.

Nous avons choisi de mettre en oeuvre une méthode barycentrique. Le canevas, pour un tel modèle, peut être irrégulier. La méthode de calcul consiste en une interpolation de type barycentrique sur les huit points de référence les plus proches du point à interpoler, répartis dans les huit portions du plan (figure 1), l'importance d'un point de référence étant d'autant plus grande dans le résultat

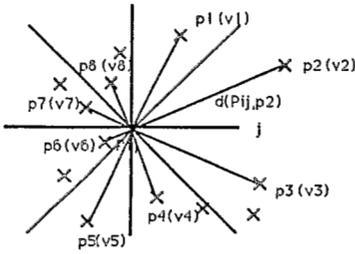


fig. 1 : l'utilisation des points de références

$$v(P_{ij}) = \frac{\sum_{l=1}^8 v(P_{lk}) \prod_{\substack{m=1 \\ m \neq k}}^8 d(P_{ij}, P_m)}{\sum_{k=1}^8 \prod_{\substack{m=1 \\ m \neq k}}^8 d(P_{ij}, P_m)}$$

fig. 2 : la fonction barycentrique pour le calcul d'interpolation

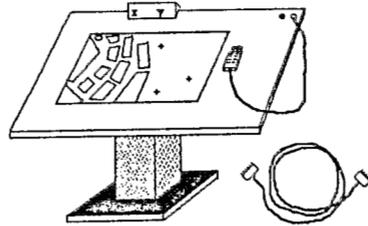
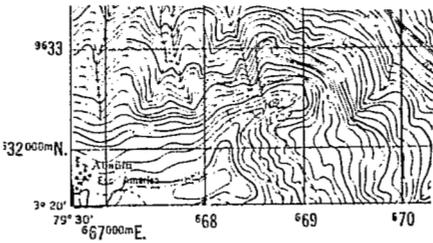


fig. 3 : La saisie des points de références se fait sur une table à digitaliser

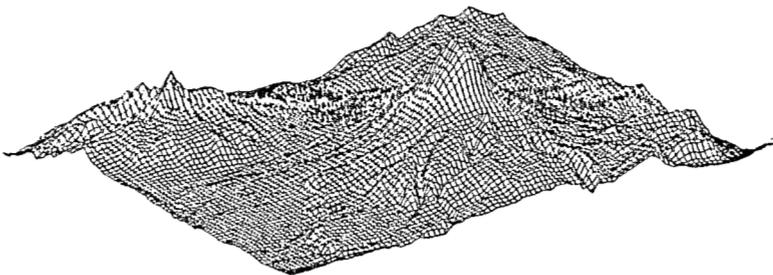


fig. 4 : représentation filaire (volcan Cotopaxi, Equateur)

du calcul que sa distance au point à interpoler est faible (par rapport aux distances des autres points au point à interpoler) (figure 2). Cette méthode ne donne pas une surface différentiable, et, n'utilisant pas de modèles locaux ou de modèle de variance, nécessite une densité de points de référence d'autant plus grande que les variations du relief sont fortes. C'est en fonction de cette variation que les points de référence doivent être choisis si cela est possible : le fonctionnement du modèle est optimal lorsque la densité des points de référence est fonction linéaire de la variation locale. Lorsque l'on dispose de l'altimétrie de base sous forme de courbes de niveaux, ce choix doit être fait lors de la saisie des points par l'opérateur.

A partir de ces points, le modèle permet d'interpoler l'altitude en chaque point du plan. En fait, on va faire ce calcul sur un canevas régulier (maille carrée) suffisamment fin pour pouvoir dire que l'on connaît par interpolation l'altitude en tout point du domaine concerné. Le logiciel Babel crée une grille de 1024×1024 pixels.

Par abus de langage, cette grille de points est également appelée *modèle numérique de terrain*. Elle n'est en fait que le résultat du modèle.

2. LA SAISIE DES POINTS DE REFERENCE

Les points de références sont saisis (position, valeur) à l'aide d'une table à digitaliser reliée à un micro-ordinateur (figure 3). Nous avons utilisé Mygale/PC pour la digitalisation des exemples présentés dans cet article. Les points isolés sont saisis, un par un (cas des cartes bathymétriques, par exemple). Les courbes de niveaux sont saisis en y choisissant suffisamment de points pour respecter à la fois l'aspect de la courbe et la distance entre les courbes adjacentes : ce sont les points saisis sur la courbe qui servent directement de points de référence. Une approche plus efficace sera d'interpréter la forme de la courbe de manière à optimiser le choix des points de références, l'opérateur n'ayant plus alors qu'à digitaliser la courbe d'une manière classique.

Les points sont saisis sur plusieurs feuilles si nécessaire (dans l'exemple présenté ici du Cotopaxi (Equateur) la saisie a été faite en 16 feuilles, représentant plus de 80000 points de référence, pour une surface de 50 km². Echelle de saisie : 1:50 000, équidistance des courbes de niveaux saisis : 40 m). Les points saisis sont ensuite recalés dans leur projection, formant un ensemble unique dans lequel l'utilisateur va définir une fenêtre de travail pour créer le modèle numérique grâce au programme d'interpolation.

3. LA VISUALISATION

Créer un modèle numérique de terrain peut servir à beaucoup de choses. Nous allons nous intéresser ici à la visualisation en trois dimensions, car, avant tout traitement, cette visualisation apporte un complément pédagogique important pour tous les problèmes faisant intervenir le relief.

Le MNT est une grille de points cotés. Représenter cette grille en 3D (choix de la rotation et de l'incidence de l'observateur par rapport au repéré de base) peut se faire de plusieurs manières :

- *représentation filaire*. Seules certaines lignes de la grille sont tracées. L'impression du relief peut être bonne, mais dépend de l'angle de rotation (les lignes sont celles de la grille). C'est néanmoins la seule méthode de représentation sur traceur à plume (figure 4). ;

- *représentation par luminosité*. Une source de lumière est placée suivant un angle de rotation et une incidence choisies. Chaque maille de la grille (espace entre quatre points adjacents) reçoit et reflète une quantité de lumière fonction de la position de sa normale avec les rayons lumineux (on supposera toujours la source de lumière à l'infini, donc des rayons lumineux parallèles). Cette quantité de lumière est donnée par la loi de Lambert (intensité proportionnelle au cosinus de l'angle de la direction de la lumière avec la normale à la surface). La visualisation de chaque facette en fonction de la luminosité réfléchie donne une excellente représentation du relief (figure 5) ;

- *représentation par luminosité avec ombrage*. L'ombre portée peut être rajoutée sur le dessin. Le calcul se fait en affectant une valeur donnée (luminosité ambiante) à toutes les facettes qui ne peuvent être directement atteintes par un rayon lumineux (figure 6) ;

- *représentation par pente, ou par orientation horizontale* (figure 7) ;

- *représentation par luminance*. Les valeurs de réflectance d'une image satellitaire sert à "éclairer" le MNT. Il faut pour cela disposer d'une image recalée géographiquement sur le modèle numérique (figure 8).

Dans toutes ces représentations, les parties cachées sont bien sûr éliminées (algorithme du peintre sur écran, ligne d'horizon pour traceur à plume).

A partir de la représentation sur écran, le tracé sur papier se fait par copie sur imprimante à jet d'encre ou à transfert thermique, rendant par les niveaux de gris l'intensité de chaque pixel

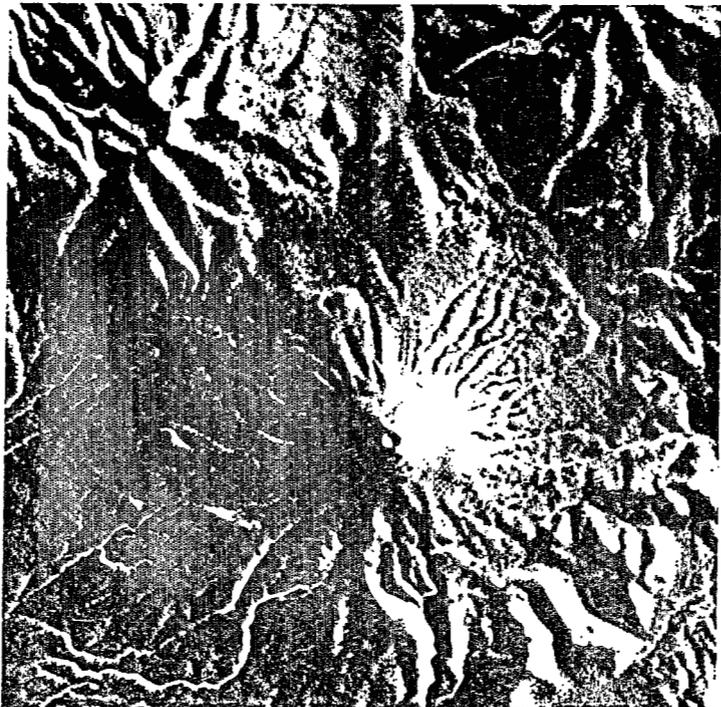
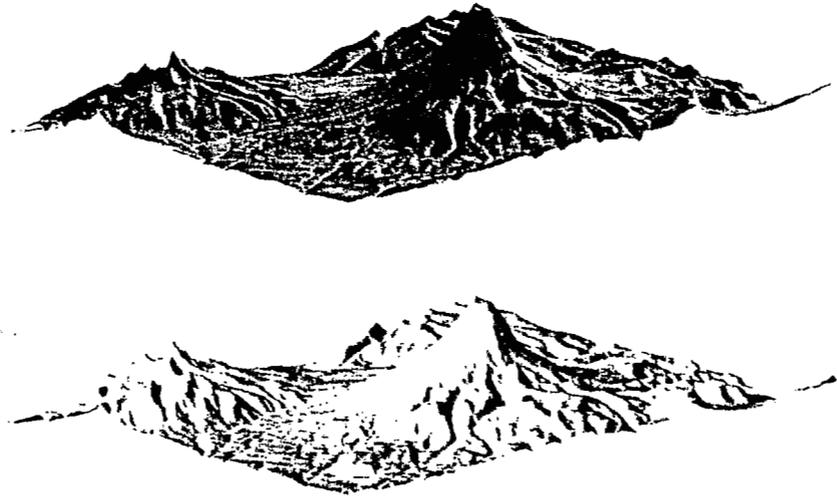


fig. 5 : représentation par luminosité



fig. 6 : représentation avec ombrage



fig. 7 : représentation par valeur de la pente



fig. 8 : représentation par luminance
(image SPOT recalée sur le MNT par visucaval, B. Lortic)

CARTOGRAPHIE SYNTHETIQUE D'UN ENSEMBLE DE VARIABLES REGIONALISEES PAR UN LOGICIEL INTEGRE D'ANALYSE DES DONNEES

(Marie-Madeleine THOMASSIN)

RESUME - L'article décrit l'ensemble des ressources offert par le logiciel MacSAIF (*) pour le traitement de variables régionalisées, et en particulier l'aspect d'aide à l'interprétation qui utilise les méthodes statistiques de l'analyse des données (analyse factorielle de correspondances et classification ascendante hiérarchique).

En s'appuyant d'abord sur un cas simple, pour ensuite montrer le processus de traitement général dans le cas de données complexes réquerant un codage préalable, l'auteur met en évidence les trois étapes principales du travail : choix du découpage en modalités des variables à traiter, choix de la partition et choix des trames utilisées sur la carte.

L'utilisation des différents programmes qui composent le logiciel, et qui permettent de réaliser chacune de ces étapes, est décrit et les principales difficultés d'application sont relevées et illustrées par des exemples.

INTRODUCTION

Le but de cette communication est de montrer aux géographes, aux chercheurs des autres disciplines ainsi qu'aux spécialistes et responsables de la planification, la variété de ressources que présente ce logiciel Mac SAIF. Au chercheur isolé sur le terrain, il permet, à partir d'une base de données complexes, de créer une carte de synthèse intégrant les informations élémentaires apportées par des variables, dont le nombre et le format ne sont soumis à aucune contrainte.

Pour présenter ce logiciel, on a fait très largement appel aux notices des programmes, publiées par le Pr. J.-P. et F. Benzécri dans les Cahiers d'Analyse de Données.

Après avoir brièvement rappelé le traitement par l'analyse des données (c'est-à-dire par l'analyse factorielle des correspondances, AFC et la classification ascendante hiérarchique, CAH) du cas le plus simple : le tableau de contingence, on a abordé le cas des données complexes ou requérant un codage préalable.

(*) Logiciel MacSAIF, "Système d'analyse des informations", créé par le Professeur et F. BENZECRI. Laboratoire de statistique, Université de Paris VI.

Le rôle des programmes de création de tableaux a été illustré à partir de leur application aux données relatives à l'agriculture et à l'élevage de la Sierra équatorienne. La carte de synthèse construite par le programme 'carthage' traduit une typologie clairement explicitée, en délimitant des zonages dont le graphisme exprime les affinités entre classes.

1. ANALYSE DES DONNEES DANS LE CAS D'UN TABLEAU SIMPLE

Partons d'un exemple de géographie économique : le tableau de données croisant l'ensemble des départements de la France, avec un ensemble de variables (les consommations mensuelles du gazole, de 1972 à 1981).

Ce tableau comporte 95 lignes, correspondant aux 95 départements français (la Corse étant considérée comme constituant un seul département), rangés depuis l'Ain jusqu'au Val-d'Oise dans l'ordre de leurs numéros minéralogiques. Le tableau a 12 colonnes, qui sont les mois de Janvier à Décembre. A l'intersection, par exemple, de la ligne Ardèche et de la colonne Juillet, on lit le nombre 236 : ceci signifie que la consommation en gazole dans le département de l'Ardèche au cours du mois de Juillet des années 1972 à 1981 a été au total de 236 centaines de m³.

Ce tableau de données va être soumis tel quel à l'AFC.

L'analyse de ce tableau montrera d'abord une opposition entre les départements caractérisés par une importante consommation hivernale et très faible consommation estivale, et ceux où prédomine la consommation estivale, ou tout au moins les départements où elle s'avère moins réduite qu'il n'est le cas en moyenne. Apparaîtra ensuite le cas des départements betteraviers, à forte consommation automnale.

Afin de mettre en évidence avec précision cette structure, et ultérieurement de produire une carte représentant la régionalisation de ces consommations différenciées selon les saisons, nous devons utiliser l'ensemble des ressources de l'analyse des données :

- l'**analyse factorielle**, qui assigne, d'après le tableau initial, des coordonnées, ou facteurs, à chacun des départements ainsi qu'à chacun des mois. Autrement dit, qui traduit en termes de *proximité* sur un graphique les *ressemblances* existant entre *individus* et entre *variables*, ainsi que les *affinités* des uns avec les autres ; et ensuite,

- la **classification automatique**, qui, au sein de cette représentation géométrique, *détermine des classes* (dans notre cas, principalement des classes de départements à chacune desquelles sera affectée ultérieurement, lors du tracé de la carte, une trame unique, tandis que les classes de mois, déjà en évidence sur les graphiques-plans, nous sont connues a priori).

La première étape d'une analyse de données sera évidemment l'introduction de ce tableau dans l'ordinateur.

Le tableau étant saisi suivant un format convenable (et d'ailleurs très souple), on entre dans le programme '**qori**' d'analyse des correspondances. Outre un fichier de type texte qui contient sous forme de tableaux des résultats numériques dont une partie s'affiche à l'écran, à la fin du déroulement du programme, le programme '**qori**' crée des fichiers numériques, où sont gardés les mêmes résultats d'analyse des correspondances et qui seront utilisés dans toute la suite des programmes de la chaîne : dans les programmes '**planF**' et '**planX**', pour produire les graphiques-plans, puis dans la classification automatique, et enfin dans l'élaboration de la carte proprement dite.

Succinctement, le programme '**planF**' crée, en mode conversationnel, des plans croisant deux axes factoriels qui s'affichent instantanément à l'écran ; tandis que le programme '**planX**', également conversationnel, crée les mêmes graphiques sur un fichier texte, qui peut être consulté grâce à un éditeur de texte et, ensuite, imprimé sur papier.

On entre ensuite dans le programme de '**CAH**' (**Classification Ascendante Hiérarchique**).

Ce programme a de multiples fonctions; notamment le tracé de l'arbre, la création de tableaux d'aides à l'interprétation, d'une grande utilité pour le praticien de la classification automatique, et tout particulièrement pour le géographe.

Le programme '**CAH**' consulte alors le fichier numérique des facteurs, qui a été créé par le programme '**qori**' d'analyse des correspondances sans lequel le programme de classification automatique ne pourrait s'exécuter ; il lit ces facteurs et, par des calculs de distances, il range les uns avec les autres les éléments qui sont le plus proches. Ceci fait, il nous reviendra d'extraire de l'arborescence ainsi construite une partition destinée à être traduite cartographiquement.

En bref, il s'agit de retenir de la classification ascendante hiérarchique, qui comporte dans l'exemple retenu, au-dessus des 95 éléments (départements), 94 noeuds, une sous-hiérarchie à la base de laquelle se trouvera une partition, par exemple en neuf classes, avec, au-dessus de ces neuf classes, des réunions successives en classes supérieures, jusqu'à parvenir au sommet. Pour effectuer un tel choix de façon tout à fait conséquente, il faut avoir une connaissance approfondie des résultats de la classification, et c'est pourquoi l'on prévoit, après une première consultation, de rentrer dans le programme '**CAH**' pour demander à bon escient une sous-hiérarchie.

Quant aux **aides à l'interprétation**, nous nous bornerons à souligner l'importance du listage **Vacor**, qui donne l'interprétation des classes par leurs profils, c'est-à-dire directement en termes des variables du tableau des données.

2. CAS DES DONNEES COMPLEXES OU REQUERANT UN CODAGE PREALABLE

2.1. Tableau de contingence avec un nombre élevé de variables.

Cet exemple de la consommation mensuelle du gazole est particulièrement simple parce que, d'une part, le *profil de chaque ligne* (c'est-à-dire la suite des quotients des éléments de la ligne par leur total) est *directement interprétable* : on dit qu'il s'agit d'un *tableau de contingence* ; et que, d'autre part, le *nombre des variables* (colonnes du tableau) n'est pas très élevé.

Si le *nombre des variables est élevé*, on peut, par le programme 'CAH', faire un *Vacor* expliquant la hiérarchie des départements *i*, non par les variables *j* elles-mêmes (les mois), mais par les *classes j_q de la partition de l'ensemble des variables* qu'il vient d'obtenir par classification des *j* mois. Dans le cas d'un ensemble de douze variables seulement, les douze mois de l'année, l'interprétation de la classification, et donc celle de la cartographie, directement en termes de ces variables est possible, et même facile. Mais prenons dans le même domaine un tableau à 48 colonnes : le *tableau donnant les consommations mensuelles de quatre produits pétroliers*. L'interprétation directe en termes de ces 48 variables devient difficile, sinon impossible. Il faut donc *procéder à l'agrégation de ces variables*, et, donc, considérer en quelque sorte un *tableau intermédiaire* donnant, par exemple, au lieu des consommations en gazole, fuel lourd, fuel domestique, carburant auto, mois par mois, les *consommations de chacun de ces produits suivant des périodes plus longues choisies selon le produit* : saison entière, moitié de l'année, même dans certains cas, trois quarts de l'année opposées à la période d'été etc.. De sorte que l'utilisateur, finalement placé en présence d'un tableau d'une dizaine de colonnes, et de pourcentages calculés sur ce tableau par le programme d'aide à l'interprétation, et disposant de graphiques-plans où s'affichent les sigles de ces agrégats de variables, pourra comprendre sans peine l'interprétation des classes de départements créées.

L'interprétation de la classification se fait donc en termes d'agrégats de variables et partant, la cartographie.

La légende de la carte distingue pour chaque produit, selon des périodes variables, les excédents ou les déficits par rapport à la moyenne, forts ou faibles des consommations mensuelles, de 1972 à 1981. De là, la nécessité d'une recherche au niveau des modes d'expression graphique, l'utilisation des symboles s'étant avérée peu satisfaisante et de lecture peu aisée, comme chacun sait, la création de nouvelles trames par le programme 'carthage' s'est révélée indispensable.

2.2. Tableau des données requérant un codage préalable

Un des cas les plus simples est celui des tableaux de réponses à un questionnaire clos.

Lorsque l'utilisateur non programmeur se trouve confronté, et c'est très souvent le cas du géographe, à un tableau de données *croisant un ensemble d'unités administratives avec un ensemble de variables numériques hétérogènes*, il se trouve dans la même situation que s'il avait à traiter un tableau de réponses à un questionnaire clos, où chaque réponse est initialement notée comme un numéro de modalité. En effet, en découpant en p sous-intervalles successifs l'intervalle de variation d'une variable continue, on peut coder celle-ci comme une variable discrète à p modalités. On peut donc assimiler la variable à une question, et la circonscription au sujet qui y répond. Un tel tableau doit être éclaté, sous forme binaire, en (zéro,un), en attribuant à chaque modalité une colonne.

Le tableau pris pour exemple d'application est extrait du tableau d'ensemble, ayant servi de base à une étude de la régionalisation agricole de la Sierra équatorienne (Cf bibliographie), croisant 238 paroisses avec 48 variables : les 26 premières relatives au milieu physique et à ses potentialités, les 9 suivantes à l'agriculture et à l'élevage, les 13 dernières aux critères socio-économique. Tableau d'ensemble découpé respectivement en 3 blocs (bloc 1, bloc 2, bloc 3). A chacun des blocs ainsi formés et au bloc d'ensemble ont été adjoints des éléments supplémentaires : des variables (en colonnes) et des individus, 42 paroisses (en lignes).

Le tableau de la présente application croise donc 280 paroisses, dont 42 en supplémentaires avec les 9 variables afférentes à l'agriculture et l'élevage.

3. LE ROLE DES PROGRAMMES DE CREATION DE TABLEAUX

Quelques programmes de création de tableaux suffisent à l'utilisateur non programmeur pour dépouiller par lui-même une enquête par questionnaire, et analyser un ensemble de variables hétérogènes recueillies sur le terrain. La présente notice est consacrée à un ensemble de tels programmes.

Le plus important de ces programmes, '**zrang**', sert à ranger, observer et découper les variables d'un tableau :

- *ranger*, c'est-à-dire ordonner par valeurs croissantes ;
- *observer*, en affichant des tableaux de valeurs, des histogrammes, des croisements, (et même des cartes analytiques, servant à établir celle qui rendra compte de la synthèse générale des données effectuée par l'analyse des correspondances et la classification automatique) ;
- *découper*, en fixant un ensemble de modalités et en créant un tableau de numéros de modalités ou un tableau binaire.

Un programme créant des histogrammes doit être conversationnel. Il suffit, en effet, d'un individu isolé très écarté pour contraindre tous les autres à s'enfermer dans un seul créneau. D'autre part, connaître les noms des individus rentrant dans chaque créneau s'avère intéressant. On s'est efforcé, dans le programme '**zrang**', d'utiliser au mieux la place disponible. Ainsi, il est facile de décider du choix des bornes des modalités en lesquelles sont découpées les variables (c'est-à-dire déterminer ce que l'on entend par : très fort, fort, moyen, faible et très faible pour chacune des variables).

Le listage du fichier 'Dcodx' (en annexe) résulte de l'application du programme 'Zrang' (option D) au tableau de données - ici agriculture - élevage. Il indique les bornes du découpage des 9 variables en 34 modalités. Ces bornes correspondent aux bornes initiales, le recodage de ces variables, à la différence de celles afférentes aux autres thèmes, ayant été jugé inutile.

Le tableau à analyser croise donc 280 paroisses avec les 34 modalités issues des 9 variables précitées. Sur ces 280 paroisses, 42 d'entre elles comportant des lacunes de l'information, ont été mises en supplémentaires par le programme 'qori'.

En analyse des correspondances, on appelle *élément supplémentaire* une ligne ou une colonne dont on ne tient pas compte pour déterminer les axes factoriels mais qui est projetée sur ceux-ci. 'Qori' permet de mettre en supplémentaire un sous-ensemble de lignes ou de colonnes du tableau soumis à l'analyse.

Il peut s'agir d'éléments apportant à un axe une contribution telle que l'analyse en est perturbée ; ou encore d'éléments pour lesquels certaines données sont absentes, comme c'est le cas ici, ou imprécises. Souvent, il est préférable de faire totalement disparaître ces éléments, comme permet de le faire le programme 'soustab'.

Parfois, on désire projeter en supplémentaires des éléments qui n'appartiennent pas au tableau de base analysé pour éclairer les résultats de l'analyse et vérifier les hypothèses qu'elle suggère. Si, comme il est d'usage, on place en lignes les individus et en colonnes les variables, on aura, dans le premier cas, un *tableau externe de lignes supplémentaires* ; et, dans le deuxième cas, un *tableau externe de colonnes supplémentaires*. Nous disons *tableau externe* parce qu'il ne s'agit pas d'un sous-tableau du tableau de base soumis à l'analyse. C'est souvent de l'adjonction de tableaux externes appropriés que dépendra l'interprétation des données principales que l'on se propose de représenter sur la carte de synthèse.

Le programme 'qorlsup' sert à créer, pour un ou plusieurs tableaux supplémentaires externes, (de lignes ou de colonnes), un fichier numérique de facteurs et un listage usuel.

Dans l'étude de la Sierra, l'analyse d'un bloc comportait systématiquement la mise en supplémentaires des deux autres, ainsi qu'un certain nombre de variables relatives aux coordonnées géographiques et à la superficie des paroisses (en colonnes) et les 42 paroisses déjà mentionnées (en lignes). L'exploitation des listages relatifs aux éléments supplémentaires à été déterminante pour l'interprétation du tableau principal et la définition des sept facteurs issus de l'analyse, et ce, d'autant plus que les coordonnées géographiques et la taille des paroisses, facilitaient l'association entre certains secteurs de la Sierra et/ou certains types de paroisses (bassins, versants) et les caractéristiques exprimées par les demi-axes considérés.

Enfin, le programme 'discr', partant de deux fichiers numériques de facteurs : le fichier des centres de classes et le fichier des individus permet d'affecter chaque individu au centre dont il est le plus proche. Dans la mesure où chaque individu est censé relever d'une classe, il s'agit de ce qu'on appelle classiquement une *procédure de discrimination* ; d'où le nom du programme. Un exemple

simple est fourni par l'application, où l'ensemble des individus (paroisses) est l'ensemble des usuels ; et les centres sont les i_q , centres de gravité des classes d'une partition des i , obtenues par **CAH**, soit les onze classes de la partition (Cf. figures 3 et 4).

Dans la pratique, '**discri**' servira particulièrement à affecter à des centres, résultant de l'analyse d'un premier ensemble de données, un ensemble d'individus adjoints en supplémentaires à cette analyse (ici les 42 paroisses présentant des lacunes de l'information). D'où le lien de '**discri**' avec '**qorlsup**'.

4. LA CARTOGRAPHIE PROPREMENT DITE

Outre les données statistiques, résultats de l'analyse factorielle et de la C.A.H., le programme '**carthage**' utilise des données numériques représentant le fond de carte, pour construire la carte.

L'interprétation des classes de paroisses en termes de classes de variables (agrégats de modalités), grâce à un Vacor, a permis d'établir une typologie clairement définie (Cf. figures 1 et 2).

Pour cartographier cette typologie relative aux dominantes culturelles et à l'élevage de la Sierra, plusieurs étapes ont été nécessaires. Tout d'abord créer un fond de carte, c'est-à-dire *numériser la carte*. Cette numérisation a été effectuée à partir du 1:750 000 (échelle des cartes de l'étude), aucun document d'échelle plus grande n'étant disponible. Pour ce faire, créer deux fichiers :

- le premier, *fichier des coordonnées géographiques des points*, au nombre de 1217, choisis comme sommets des lignes polygonales délimitant les unités administratives (paroisses), et numérotés séquentiellement ;

- le second concernant *les contours des unités administratives elles-mêmes*, au total 280 (238 principales, 42 supplémentaires), assimilées chacune à un polygone et décrites par son numéro de code, le nombre de sommets placés sur ses contours et les numéros de ces sommets.

Précisons que les mesures de la latitude et de la longitude ont été effectuées en mm, à partir d'une origine située au nord-ouest de la feuille. Il n'a pas été nécessaire de se soucier de l'échelle du document servant de base à la numérisation, le programme de création des fichiers numériques mettant lui-même les nombres à l'échelle commode pour l'exécution du tracé de la carte.

Le programme '**litxcarte**' a permis de compiler les fichiers de coordonnées de points et de contours des paroisses, pour créer un *fond de carte* au format voulu par '**carthage**'. Il convient de souligner qu'il rend possible également la vérification, unité par unité, de la *validité des contours de chacune d'elles*. Plusieurs parcours exploratoires ont permis de corriger les erreurs commises lors de la saisie.

Pour imprimer le fond de carte on se sert de la trame n° 12 (Cf. figure 5)

En outre, '**carthage**' propose à l'utilisateur un choix de trames pratiquement illimité ; car, d'une part, lui sont directement accessibles 32 trames prises dans un tableau de 100, à l'entrée du programme (et il peut modifier le choix

Fig.1
 Consommation mensuelle du gazole en France par département
 (1972 - 1981) par A.Moussaoui

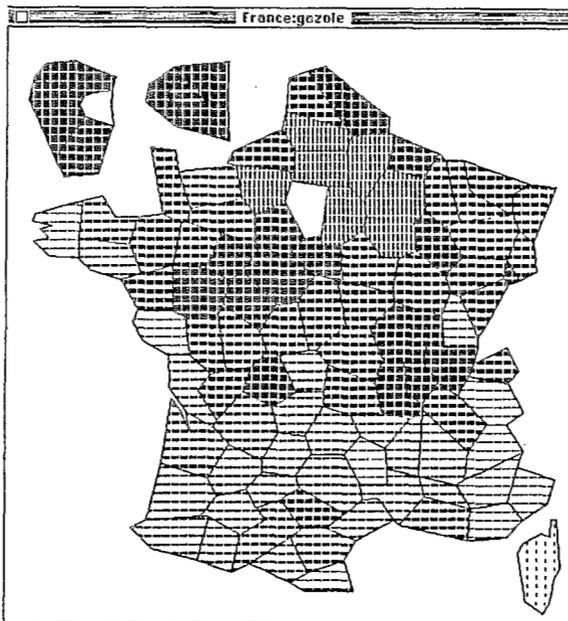
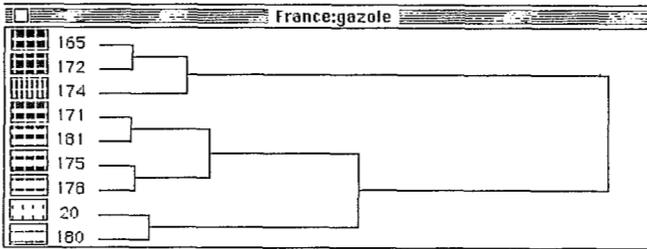
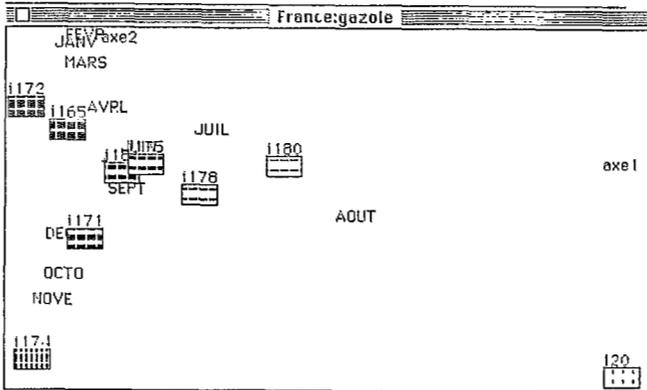


Fig. 2
Classification des variables

c Partition en 17 classes : Sigles des individus de la classe numero c				
361	BV0	CR2	SUPERFICIES CEREALIERES TRES DEVELOPPEES (40 A 70%/sup.tot.utilisee) PEU D ELEVAGE BOVIN(0 A 1 ^o TETES/100 ha d une U.P.A.)	
511	MZ1	LFO	OV1	PAS DE CULTURES MARAICHERES, FRUITIERES ;PEU DE MAIS(0 A 40 %/sup. tot.utilisee) ELEVAGE OVIN MODERE (20 A 39 TETES/100 ha d une U.P.A.)
421	PT1	TR0		PAS DE CULT. TROPICALES , CULT. SPECULATIVE DE LA POMME DE TERRE (0 A 40%/sup.tot. utilisee)
41	BV3			ELEVAGE BOVIN INTENSIF (60 A 99 TETES/100 ha d une U.P.A.)
351	OV2	BV2		ELEVAGES OVIN ET BOVIN DEVELOPPEES(40 A 59 TETES/100 ha d une U.P.A.)
431	CR3	MZ0		PLUS DE 70% en CEREALES/sup.tot.utilisee;pas de MAIS
271	LF2			40 A 70% en CULTURES MARAICHERES ET FRUITIERES
501	OV3	PA2	IR2	ELEVAGE OVIN INTENSIF(60 A99 TETES/100 ha d une U.P.A.) PATURAGES ET SUPERFICIES IRRIGUEES TRES DEVELOPPEES (40 A 70 %/sup.tot. utilisee)
381	IR3	PA3		PLUS DE 70% EN PRAIRIES IRRIGUEES
301	TR2			CULTURES TROPICALES TRES DEVELOPPEES (40 A 70% /sup. tot. utilisee)
471	BV1	OV0	TR1	ELEVAGE BOVIN MODERE(20 A 39 TETES /100 ha d une U. P. A.); PAS D OVIN;PEU DE CULTURES TROPICALES(0 A 40%/sup.tot. utilisee)
491	LF1	CR1	MZ2	FAIBLE % EN CULTURES MARAICHERES ET FRUITIERES(0 A 40%/sup. tot. utilisee);40 A 70% EN MAIS)
401	IR1	PA1		FAIBLE % EN SUPERFICIES IRRIGUEES ET EN PATURAGES(0 A 40%/sup. tot. utilisee
441	OV4	BV4		ELEVAGES BOVIN ET OVIN TRES INTENSIFS(>100 TETES/100 ha d une U.P.A.)
481	MZ3	PA0		CULTURE DU MAIS TRES DEVELOPPEE(>70%/sup.tot. utilisee)
311	IR0			PAS DE CULTURES IRRIGUEES
451	CR0	PT0		PAS DE CEREALES NI DE POMMES DE TERRE

de ces 32 trames sans sortir de 'carthage'), d'autre part il peut écrire des fichiers numériques décrivant des trames nouvelles qui se placent automatiquement à la fin du tableau des 100. De là, une totale liberté pour arrêter les modalités de l'expression cartographique de son choix.

Il importe de signaler que le choix des trames peut être guidé conversationnellement par l'affichage des résultats de l'analyse factorielle et de la classification automatique ; résultats qui ne sont autres que l'expression la plus simple des similitudes existant entre les réalités géographiques correspondant à chaque classe.

On peut en effet demander que s'affichent, sur un plan croisant deux axes factoriels, non seulement les numéros des classes de paroisses, précédés de la lettre i, par exemple i 460 (classe n° de paroisses, dont la composition peut être recherchée sur les listages de classification automatique et apparaîtra sur la carte avec une trame particulière), mais aussi l'affichage de ce n° avec, en-dessous dans un caisson rectangulaire, la trame actuellement choisie pour représenter les paroisses de cette classe (Cf. figure 3).

Il importe de noter qu'avant l'affichage d'une première version de la carte, sont proposées, dans des caissons carrés alignés aux marges gauche et supérieure de l'écran, des trames qu'on peut choisir pour remplacer celles proposées dans de grands caissons rectangulaires associés aux classes (Cf. annexe). La disposition des trames dans les caissons carrés apparaît, dès l'entrée dans le programme, avec un texte offrant de modifier cette palette de 32 trames à partir du tableau de 100.

Le choix des trames est effectué en conformité avec l'interprétation des classes de paroisses en termes de classes de variables, en vue de rendre compte de leurs caractéristiques et traduire les degrés divers de spécialisation de l'agriculture et de l'élevage. Ce choix exprime aussi les affinités et les oppositions entre classes, délimitées sur la carte par des zonages (Cf. figure 4). Pour ce faire, on a joué sur l'orientation, le pas et la texture des trames, la valeur des grisés, la forme des éléments graphiques. La taille de la paroisse a influé grandement sur le choix du graphisme, et il a fallu, en effet, faire plusieurs parcours avant de l'arrêter définitivement.

L'expression cartographique des paroisses principales se différencie des supplémentaires. Sur les cartes de l'étude, imprimées en couleurs, les aplats représentaient les premières, les lignés et tiretés représentaient les secondes, selon l'importance des lacunes de l'information. Dans le cas présent, on a traduit le caractère incertain de l'appartenance à une classe en superposant des lignes obliques lâches (trame n° 13) à la trame individualisant la classe d'appartenance supposée.

Le programme 'carthage' permet en effet de combiner des trames. Il suffit pour cela, dans le fichier de fond de carte, de faire figurer deux fois la même paroisse, la première comme appartenant à une classe donnée, la seconde comme appartenant à la classe des supplémentaires.

Lorsque la carte s'avère pleinement satisfaisante, on peut immédiatement imprimer ce qui s'affiche à l'écran, et également le saisir comme un fichier-image, que l'ordinateur permet de modifier par un logiciel graphique, pour ajouter des

Fig 3
Dominantes culturelles et élevage
Plans graphiques issus de l'AFC

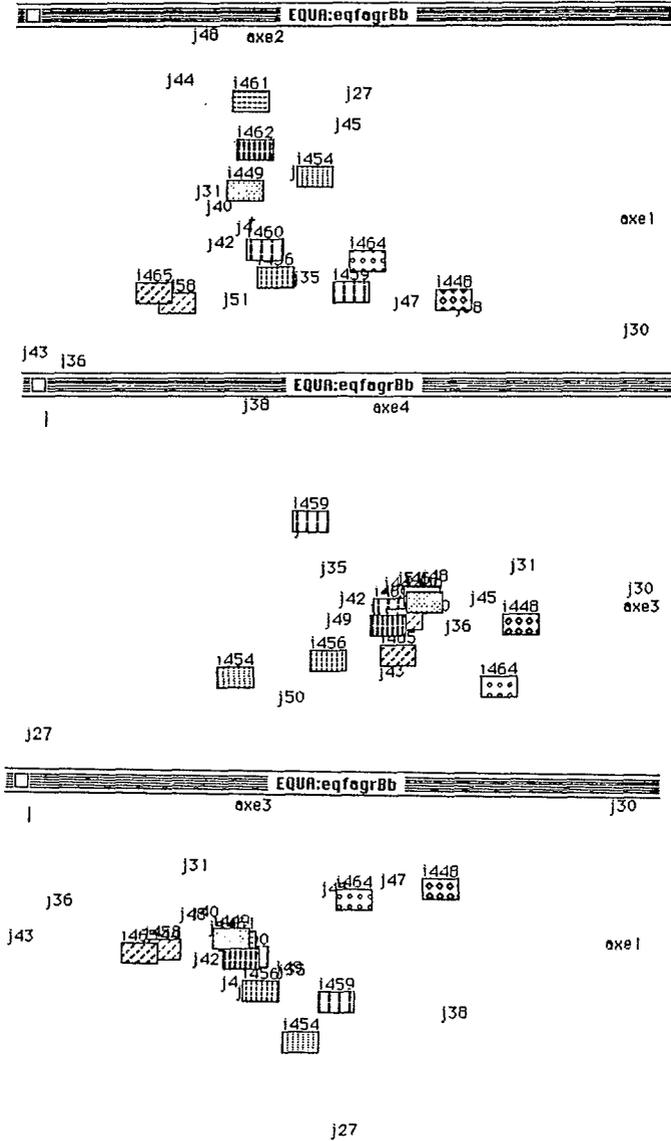
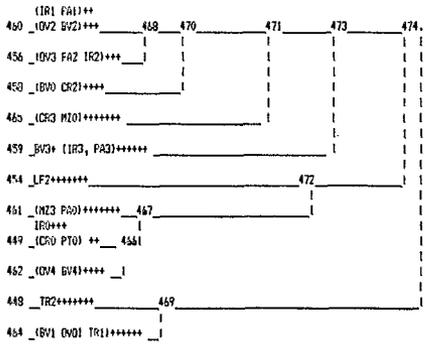


Fig. 4
Interprétation des 11 classes de paroisses de la C.A.H.
en termes de classes de variables

```

c | Partition en 11 classes : Signes des individus de la classe numero c
-----
460| cfc aff osh ehl ogo cci bai bbf oai oia acf oeh ach oqi
   | ada bad oga adb boe bao odf oef oao afh oow oiw (DV2 BV2)+++ (IR1 FA1)++
-----
450| odf aoa oeb aoh bfo cha ahe ogo aoe acc ofo oja bga beo
   | ceo ceh aza bca bpi abe baf bco (OV3 FA2 IR2)+++
-----
456| bed acb aee bcp atc osh cfi acg odf ohe ohe ajc bcb ahh
   | bcc abd ach bdd ehf bci oqg bdb chb add cho bbc
   | l bbe bdd bth adb ace ahi bcf adf (BV0 CR2)++++
-----
451| oac cah cca bdo bga ebe ahg ahj cbo ccf aho cce bbb ade
   | bah adg abg aii beb bbo bca (CR3 M20)+++++++
-----
459| afa ofi afd caa coe cog aaa aci aao aaj abh cib hec b'bh
   | oqi aaw aoi coh aca aag aac cab cac (IR3 FC3)+++++++ BV3+
-----
454| afg ahf bhi aje oig agd bfg agf afo aib aeh aic ael LF2+++++++
-----
461| oio oif tab bti bpo aco abh oed aeg aeg adi ace bof (M3 PA0)+++++++
-----
449| coh bfa bgh bpf bfh bic bib cod acb coc bee bnf bpi bli
   | beo bfe aoc acd cdf bfd bdg IR0+++ (CR0 P10)++
-----
462| aoh aoo aab abi aeb oii bfb bff boc aed ogy aeo aha b'bg
   | bpc oge aih bhe aec bdc b'p oih adc oqc b'go b'ho
   | l bhc bhd b'ie b'ge b'fc bei bio a'fe bih (OV4 BV4)++++
-----
448| oae oag osh cao oaa oec oeb TR2+++++++
-----
444| ofd ofb cdo oeg odg odi cch odf obo off obg o'ed ofe o'fh
   | o'ig o'ic oai a'hc b'fa ofc o'dd c'de c'do (BV1 DV3 TR1)+++++
-----

```



ci dessus l arbre de la partition en 11 classes

titres et d'abord pour assembler les différentes parties qui constituent la Sierra du nord au sud, au nombre de quatre.

Lorsqu'on désire revenir sur le choix des trames, il n'est que d'utiliser la possibilité qu'offre le programme de recommencer depuis le début toutes les opérations que nous avons sommairement décrites :

- choix d'une palette appropriée de 32 trames parmi 100 ;
- consultation des résultats de l'analyse des données ;
- affectation de nouvelles trames à tout ou partie des classes des unités territoriales considérées.

CONCLUSION

Du fait de son caractère conversationnel, le logiciel Mac SAIF permet de remodeler à volonté les trois étapes principales du travail, d'abord choix et découpage en modalités des variables retenues grâce au programme "**zrang**", ensuite choix de la partition adoptée en conjuguant les critères statistiques et critères proprement cartographiques, enfin choix des trames utilisées sur la carte, créées et présentées à l'écran par le programme '**carthage**'.

REFERENCES BIBLIOGRAPHIQUES

- J.-P. & F. BENZECRI : *Notices d'utilisation des programmes du logiciel Mac-SAIF* : [NOT. CORR. CAH], [NOT. CRE TAB.], [NOT. PROG. CART.], in C.A.D. Vol XIV n°1, pp.7-72,, 1989.
- J.-P. BENZECRI : *D'Aristote à l'analyse multidimensionnelle : logique et ontologie dans l'induction statistique* [LOG.STAT.]; in C.A.D. Vol XIII N°3, pp.325-338, 1988 ; (dans cet article est publiée la carte de la consommation du gazole en France, due à A.Moussaoui).
- J.-P. BENZECRI : *Où vit-on le mieux en France ? représentation cartographique de l'analyse de 33 variables* [CARTE VIE FRANCE]; in C.A.D. Vol XIII n°3, pp.349-366, 1988.
- M.-M. THOMASSIN : Cartographie automatique fondée sur l'Analyse des Données ; une application sur micro-ordinateur du programme 'carthage' à la régionalisation de la consommation des produits pétroliers en France (1972-1981); in *Bulletin du Comité Français de Cartographie*, fascicule n°115, (Mars 1988). 8e Assemblée Générale de l'A.C.I., 13e Conférence internationale Morelia (Mexique)-12-21 Octobre 87.
- M.-M. THOMASSIN : *La Sierra équatorienne. Régionalisation agricole et Analyse des Données*. Vol. 1, Texte, 626 p ; Vol. 2, Cartographie : 28 cartes et plans, en couleurs et en noir et blanc. Collection Travaux et Documents. Editions de l'ORSTOM, Paris. 1988.

Fig.5
Fond de carte de la Sierra équatorienne



Fig. 6
Légende de la carte "Dominantes culturales-Elevage"
Typologie des paroisses expliquées par la typologie des modalités, d'après le Vacor

Classe n°	Caractéristique (s) *	Description
460	(O72 B72) +++ (IR1 FA1) ++	: Zones d'élevage intensif bovin et ovin. prairies irriguées modérément développées
456	(O7) FA2 IR2) +++	: Zones d'élevage ovin intensif avec prairies et cultures irriguées
458	(B70 CR2)+++ (MZ1 LFO OVI1)+(LFI CR1 MZ2)+	: Zones céréalières spécialisées avec élevage bovin ; culture du maïs modérée
465	(CR) MZ0) ++++++	: Zones céréalières très spécialisées. absence de maïs.
459	(IR) FA3) ++++++	: Zones d'élevage laitier intensif avec prairies irriguées très développées.
454	LF2 ++++++	: Zones spécialisées dans les cultures maraichères et fruitières.
461	(MZ) FA0) ++++++	: Zones maizicoles spécialisées. pas de pâturages
449	LFO +++ (CR0 FT0) ++	: Très peu de superficies irriguées. de céréales et de pommes de terre
462	(O74 B74) ++++ (MZ1 LFO OVI1)	: Zones d'agriculture maizicole avec élevage bovin et ovin intensif
448	TR2 ++++++	: Zones de cultures fruitières ou maraichères tropicales très spécialisées
464	(B71 G70 TRI) ++++++	: Zones de cultures fruitières ou maraichères moins développées que ci-dessus (448). avec élevage laitier modéré. peu d'ovins

* de la classe de variables d'après le Vacor, plus le nombre de + est élevé. plus les caractéristiques sont marquées.

Arbre de classification des paroisses

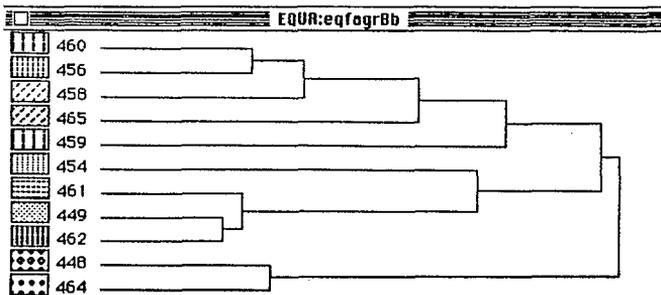
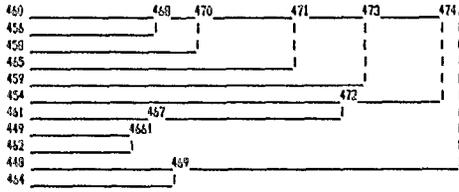
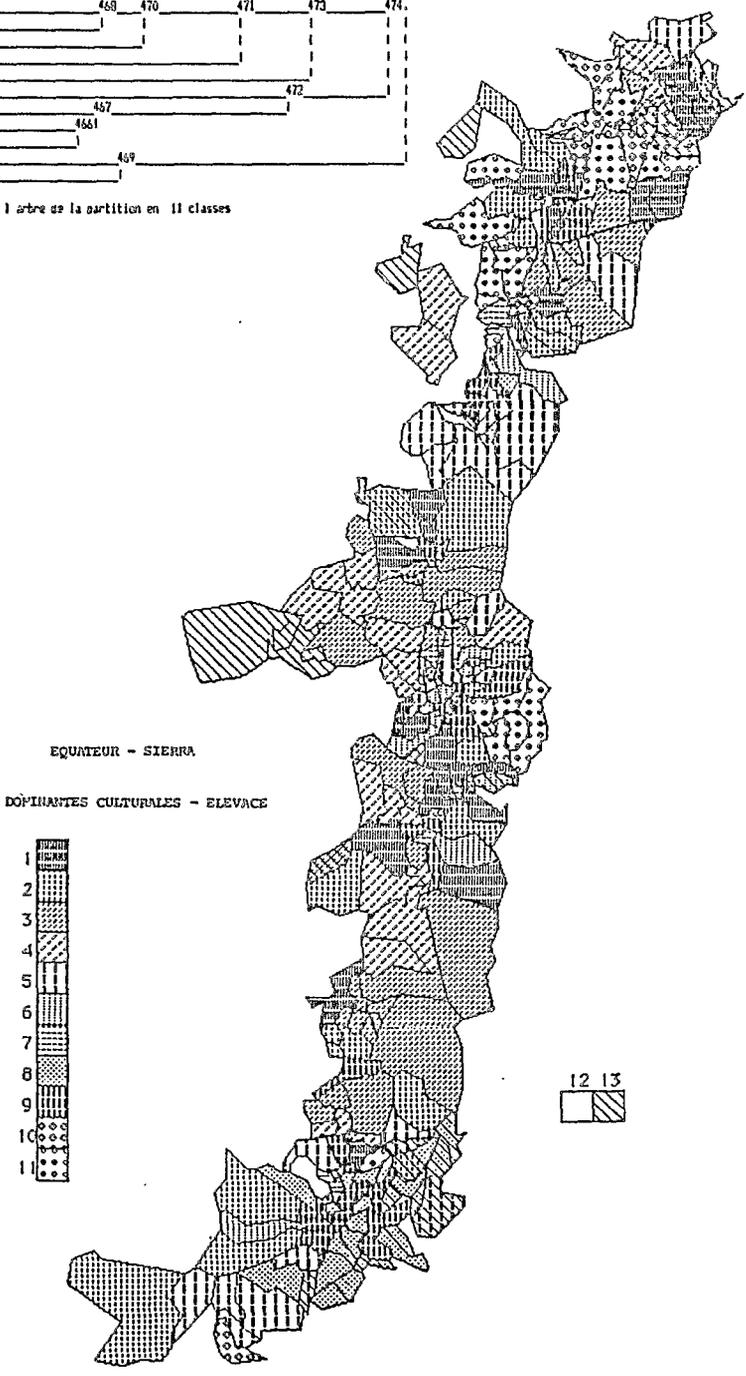


Fig.7

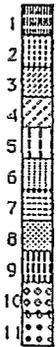


ci dessus l arbre de la partition en 11 classes



EQUINEUR - SIERRA

DOMINANTES CULTURALES - ELEVACE



ANNEXE

Variables principales codées par 'Dcodx'

EQUATEUR agriculture élevage

77:EQagrDcodx: bornes pour le découpage des variables

le nombre des variables est 9

27 a 5 modalités dont les signes et bornes sont

BV0 BV1 BV2 BV3 BV4

0 2 5 8 10

28 a 5 modalités dont les signes et bornes sont

OU0 OU1 OU2 OU3 OU4

0 2 5 8 10

29 a 4 modalités dont les signes et bornes sont

PA0 PA1 PA2 PA3

0 3 7 10

30 a 4 modalités dont les signes et bornes sont

MZ0 MZ1 MZ2 MZ3

0 3 7 10

31 a 4 modalités dont les signes et bornes sont

CR0 CR1 CR2 CR3

0 3 7 10

32 a 2 modalités dont les signes et bornes sont

PT0 PT1

0 3

33 a 3 modalités dont les signes et bornes sont

LF0 LF1 LF2

0 3 7

34 a 3 modalités dont les signes et bornes sont

TR0 TR1 TR2

0 3 7

35 a 4 modalités dont les signes et bornes sont

IR0 IR1 IR2 IR3

0 3 7 10

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

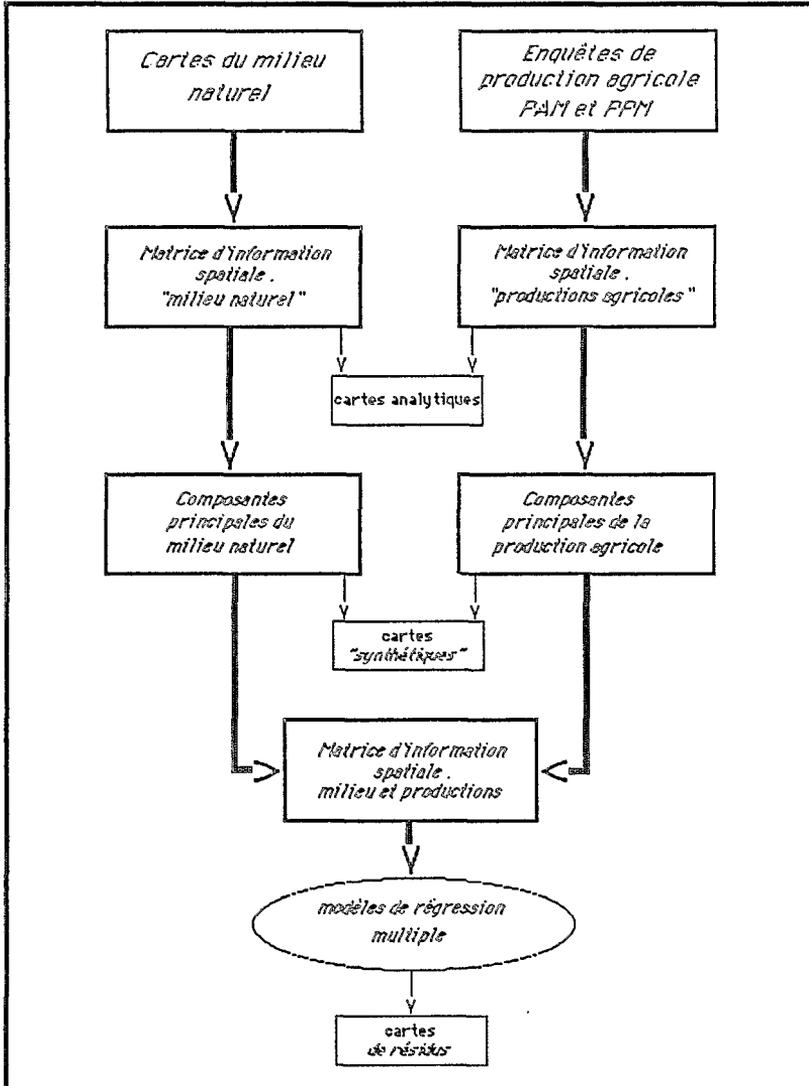
permet de déboucher rapidement sur des résultats tangibles. En effet, ne disposant ni du temps, ni des moyens nécessaires à la réalisation d'un SIG sur les *cerrados*, nous avons conçu une méthode, fondée sur un carroyage constituant le point de rencontre du découpage administratif de la statistique 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.

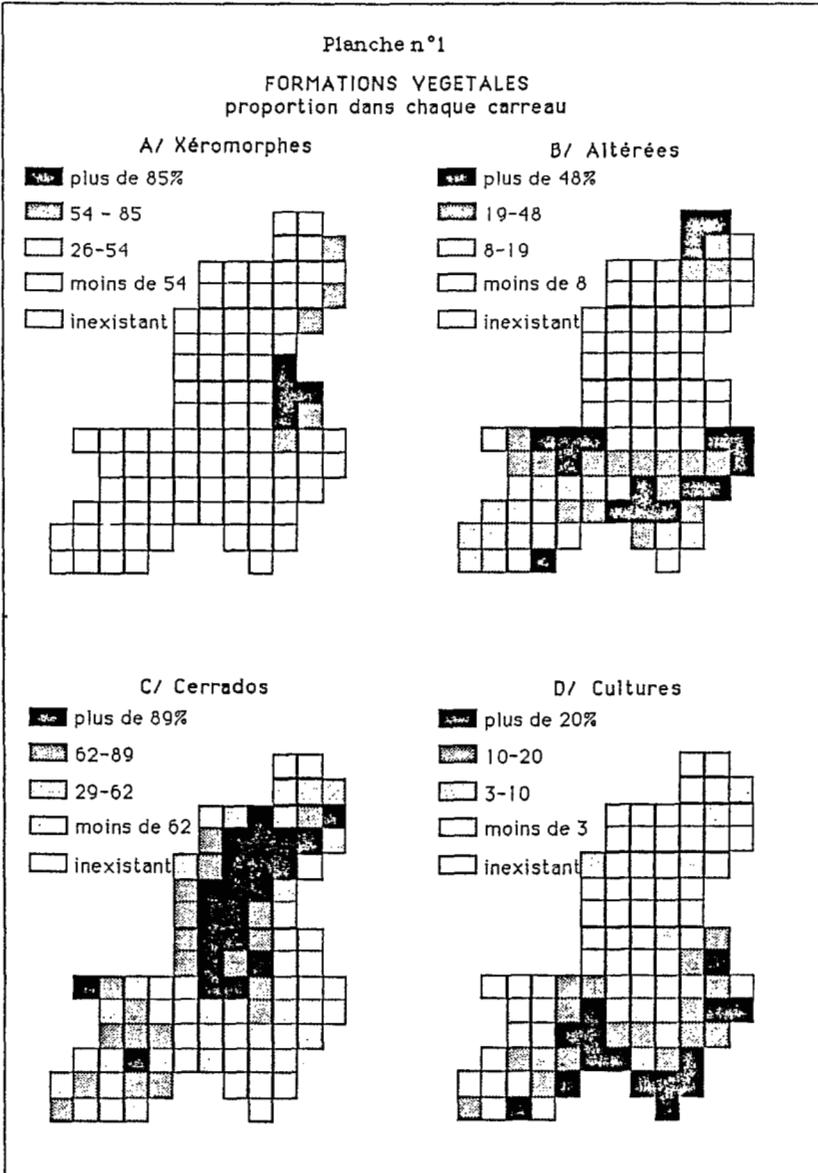
La méthode adoptée privilégie donc la mesure (figure 1). Les stocks d'information se rapportent d'une part au milieu naturel, sous forme de cartes thématiques, et de l'autre à la production agricole, sous forme d'enquêtes annuelles de 1977 à 1984. A partir de ces gisements d'information, on cherche à 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.

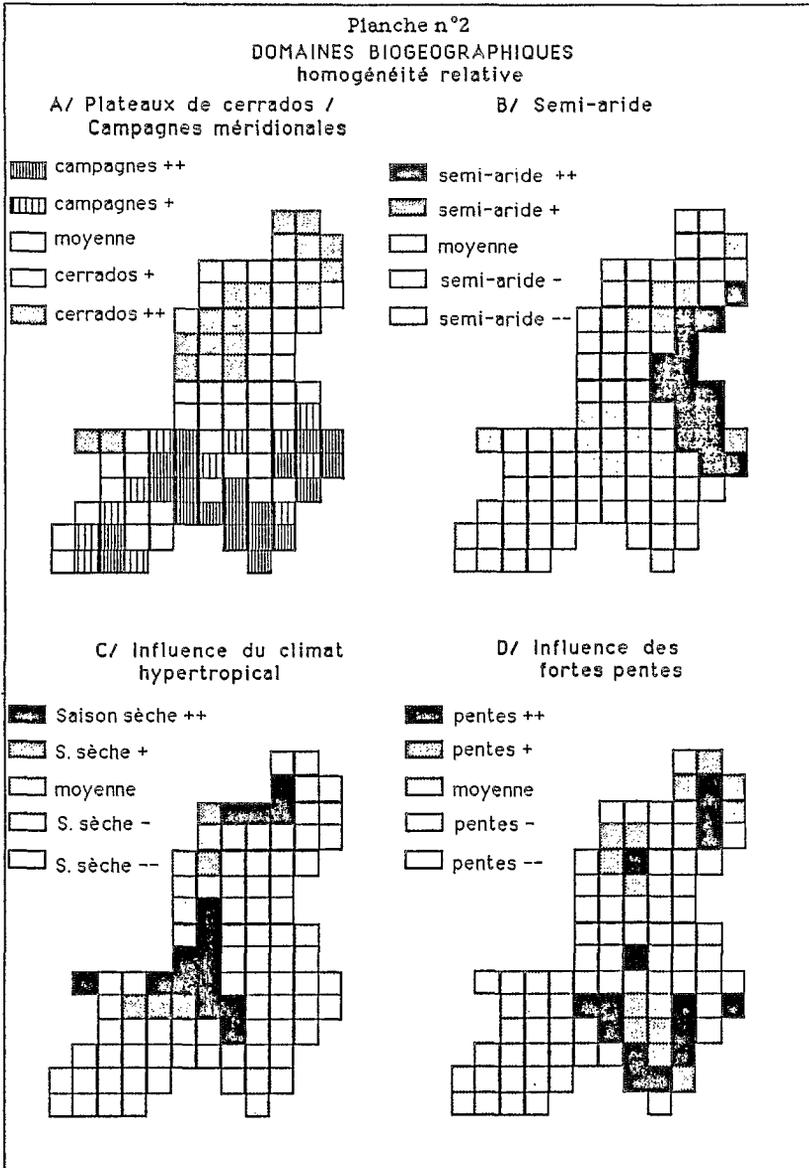
Les principaux thèmes relatifs au milieu naturel sont extraits de cartes thématiques au 1:5 000 000, publiées par L'UNESCO, la FAO et l'EMBRAPA. On procède à la mesure de chaque thème sur un carroyage à double niveau : un niveau de comptage sur plus de 9 500 carreaux et un niveau statistique composé 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







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).

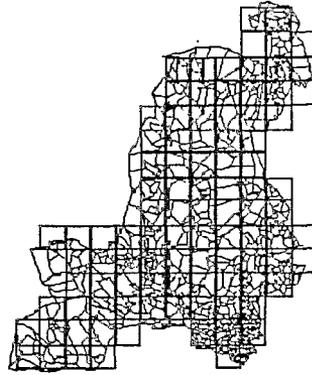
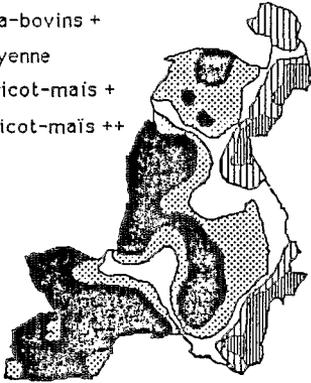
Ces deux ensembles de données (les quatre composantes biogéographiques et les trois composantes de la production agricole) se rapportant à la même maille, le carroyage, sont mises en relation par des modèles de régression multiple. Par exemple, deux composantes du milieu naturel sont corrélées de manière 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 ++

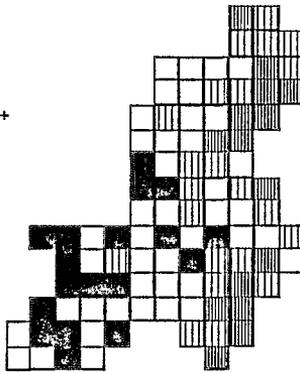
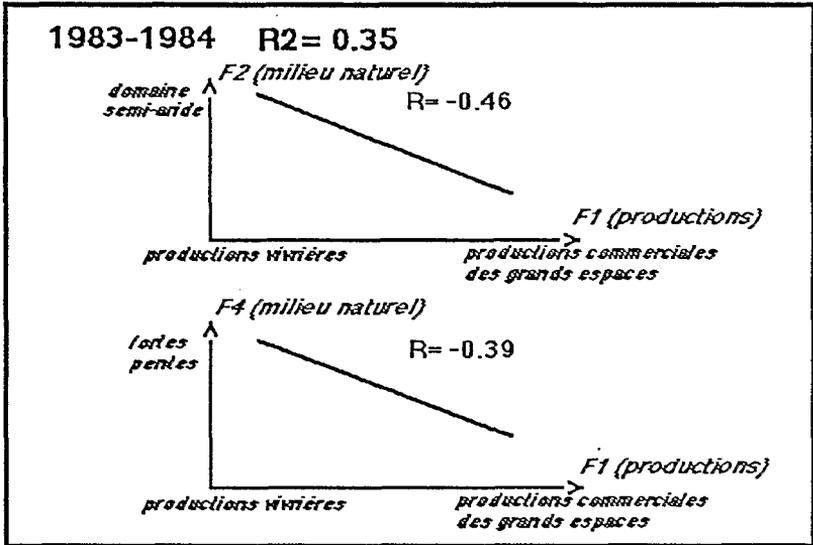


Figure n°2

**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.

DETERMINATION ASSISTEE PAR ORDINATEUR : BASE DE DONNEES, CLE INFORMATISEE OU SYSTEME EXPERT ?

(L. TITO DE MORAIS)

RESUME - Depuis octobre 1988 un groupe informel de travail (1) sur les applications de l'intelligence artificielle à l'aide à la détermination des poissons d'eau douce, s'est réuni régulièrement. L'objectif à court terme est la réalisation d'un système expert d'aide à la détermination des Characidés d'Afrique de l'Ouest. Le travail se déroule en cinq étapes :

- analyse de la démarche de l'expert lors d'une diagnose ;
- "systématisation" de la démarche et réalisation de la base de connaissances ;
- mise en place du système expert (écriture des blocs de contrôle et des règles de production, tests, etc.) ;
- mise en place des routines externes de calcul (traitement numérique des données métriques et méristiques) ;
- amélioration de l'interface utilisateur et tests finaux.

Le travail est actuellement dans la troisième phase. Il a permis de tirer certaines conclusions relatives :

- à la démarche de l'expert. Celui-ci procède par "analyse d'images", i.e. la mémorisation de certaines espèces et par analogie pour les espèces voisines. En cas d'insuccès, ou pour confirmation, il y a utilisation d'une clé de détermination (mémorisée ou consultée). Les données méristiques (nombre de branchiospines, rayons, dents, etc.) ont alors une grande importance ;
- à la réalisation de la base de connaissances et à l'importance des données écologiques et biogéographiques pour l'orientation initiale ou la confirmation de la diagnose ;
- à la mise en place du système expert. Si la réalisation d'une "clé informatisée" permet déjà la mise en évidence d'ambiguïtés, voire d'erreurs, dans les ouvrages de détermination, il importe de dépasser largement ce stade en faisant appel à des routines externes de calcul et à des données bioécologiques régulièrement actualisées (figure 2).

Les mérites comparés, pour ce type d'application, des systèmes à base de règles de production de type ESE d'IBM et "centrés-objet" sont discutés.

En conclusion, et au delà du simple cas d'espèce qui fait l'objet du travail actuel, l'intérêt et les contraintes de la réalisation de telles bases de connaissances et de systèmes experts "ad hoc" sont discutés.

(1) Ce groupe de travail était constitué par J.F. MURAIL du Muséum National d'Histoire Naturelle, MNHN, et de J. CRUETTE, D. PAUGY et L. TITO DE MORAIS de l'ORSTOM.

Dans certains cas précis, avec la baisse des coûts du matériel et leur "convivialité" accrue, la réalisation d'un système expert peut s'avérer intéressante pour fournir à la contrepartie locale un outil de formation et de travail, dans le cadre de programmes censés se poursuivre après le retrait de l'ORSTOM. (Programmes de contrôle biologique à long terme en particulier).

Au niveau des chercheurs, la réalisation de bases de données individuelles est la règle. Une politique de relative standardisation et de mise en place de bases de connaissances régionales, gérées par des systèmes experts, augmenterait les contacts et favoriserait l'échange d'informations au sein des unités de recherches et entre elles. Elle fournirait en outre :

- des outils de formation des jeunes chercheurs et techniciens ;
- des outils de travail pour les chercheurs de spécialités voisines, ou pour des ichtyologues qui changent d'aire géographique d'étude ;
- les maillons d'un réseau de bases de connaissances internationales.

La mise en place de bases de connaissances impliquant plusieurs chercheurs (pour la réalisation comme pour l'alimentation et la mise à jour périodique) demande un engagement des commissions scientifiques autant dans le cadre de leurs attributions relatives à l'animation scientifique et au développement des échanges, que pour la valorisation du travail fourni par les chercheurs pour la réalisation des systèmes.

INTRODUCTION

Dans le cadre d'un groupe de travail ORSTOM-MNHN, nous réalisons une application de l'intelligence artificielle à l'aide à la détermination de quelques poissons d'eau douce : les Characidés de l'Afrique de l'Ouest.

Nous avons délaissé pour l'instant les Characidés nains (15 genres et plus de 60 espèces), dont la systématique pose actuellement un certain nombre de problèmes non résolus au plan fondamental. La famille se décompose donc en 4 genres (Tableau 1) comprenant 42 espèces.

CHARACIDES D'AFRIQUE				
Genres :				
Alestes	Brycinus	Hydrocynus	Lepidarchus	Nains
5 espèces	31 espèces	5 espèces	1 espèce	env. 60 esp.

Tableau 1 : Genres et nombre d'espèces composant les poissons Characidés d'Afrique.

Nous avons utilisé le générateur de systèmes experts ESE d'IBM.

Le travail se déroule en cinq étapes :

(1) Analyse de la démarche de l'expert lors d'une diagnose. (2) "Systématisation" de la démarche et réalisation de la base de connaissances. (3) Mise en place du système expert : écriture des blocs de contrôle et des règles de production, tests, etc. (4) Mise en place des routines externes de calcul (traitement numérique des données métriques et méristiques). (5) Amélioration de l'interface utilisateur et tests finaux. Le travail est actuellement dans la phase (3).

L'expert biologiste procède très souvent par "analyse d'images", i.e. par mémorisation de certaines espèces et par analogie pour les espèces voisines. En cas d'insuccès, ou pour confirmation, il y a utilisation d'une clé de détermination (mémorisée ou consultée). Les données méristiques (nombre de branchiospines, rayons, dents, etc.) ont alors une grande importance. Au cours de la diagnose, l'expert s'éloigne souvent de la stricte phylogénie. Il fait appel à des caractères qui n'ont que peu ou pas de valeur phylogénétique, mais qui à un niveau donné de la détermination peuvent être d'un grand intérêt pratique. Par exemple, des poissons de plusieurs espèces très différentes peuvent avoir le même nombre de branchiospines, cependant, à un niveau bas de la hiérarchie, ce même caractère peut permettre de séparer deux espèces très voisines.

Il y a en effet deux aspects dans la détermination (avec des possibilités d'imbrication multiples) :

- une approche phylogénétique regroupant les espèces suivant des critères systématiques (en genres, familles, ordres, etc.). Elle fait très souvent appel à des caractères difficilement observables (internes, microscopiques, voire ontogénétiques observables seulement à certains stades du développement) ;

- une approche pratique, faisant appel à des critères plus facilement observables (mais pas toujours !). Les regroupements effectués ne correspondent alors pas nécessairement aux groupes zoologiques.

Woolley et Stone ont publié en 1987 un article comportant une brève discussion sur les divers systèmes d'identification taxonomique : tableaux à entrées multiples, clés dichotomiques "papier", clés et tableaux informatisés. Ils présentaient également un système expert fondé sur des règles de production (fonctionnant en chaînage arrière) adapté à la détermination des insectes d'une famille d'Hyménoptères. Depuis cette date, d'autres approches ont été développées. Et notamment en France à l'Université Claude Bernard de Lyon et à l'INRIA à Grenoble où sont étudiés des systèmes "centrés-objet" (Shirka).

Les tableaux à entrées multiples sont d'une gestion extrêmement complexe sur le papier. Dans leur version informatisée, chaque individu est constitué par un ou plusieurs enregistrements comportant les variables descriptives (figure 1). Leur exploitation par un système de gestion de bases de données se fait de manière non optimale et, si le nombre d'individus est élevé, il faut un très grand nombre d'étapes pour parvenir à une diagnose. Le système est très sensible aux erreurs et ne peut pas fournir des résultats pondérés par un coefficient d'incertitude. Il est en revanche facile à mettre à jour et à modifier.

Les clés dichotomiques (papier et "automatiques") sont très complexes à écrire. Toute modification ultérieure conduit très souvent à une refonte globale de

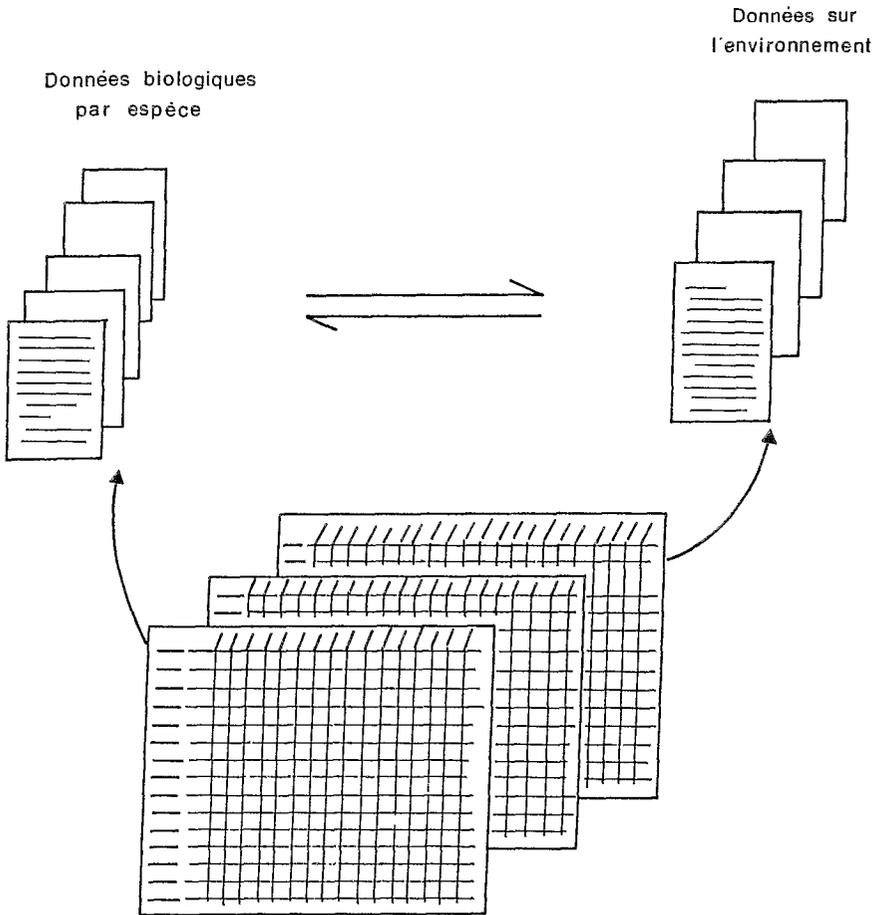


Fig.1: Tableaux à entrées multiples et bases de données
(Voir texte).

la clé. Une version informatisée est d'utilisation plus aisée. Mais dans tous les cas, elles sont extrêmement sensibles aux erreurs de réponse, elles ne permettent pas les retours en arrière (sauf reprise depuis le début) et n'acceptent pas les coefficients d'incertitude ni l'absence de réponse. Leur avantage étant qu'il en existe un grand nombre sur papier, qui ont été éprouvées et améliorées au fil des années.

Les systèmes "centrés-objet" sont ceux dont la "philosophie" et la structure se rapprochent le plus des critères de la systématique, mais pas nécessairement de la démarche de l'expert réalisant une identification. Les objets et leurs regroupements peuvent être reliés aux notions d'espèce, de genre, de famille, etc.. Leur classification hiérarchique est voisine des structures arborescentes de la classification. La création de la base demande en général une réorganisation de la clé de détermination utilisée, mais les modifications et les mises à jour sont aisées.

Les systèmes à base de règles de production sont, dans leurs versions initiales, des clés informatisées optimisées. Déjà dans ces premiers stades, ils apportent cependant d'importantes améliorations par rapport aux "simples" clés (détection d'erreurs, coefficient d'incertitude, meilleure gestion des réponses multiples et des réponses manquantes, historique et justification de la diagnose en fin de parcours, etc.).

Ce type de système est celui dont la conception et l'utilisation se rapproche le plus de la démarche "réelle" d'un expert effectuant une diagnose. Comme lui, il est essentiel que le système puisse faire appel à des connaissances autres que celles d'une clé dichotomique. C'est à dire à des bases de données bio-écologiques, phylogénétiques et à des routines externes de calcul (figure 2). Ces dernières permettent notamment de replacer statistiquement les valeurs métriques et méristiques de l'individu étudié, au sein des intervalles ou des clines de valeurs observées pour les différentes espèces dans différentes régions. *In fine*, le résultat de la consultation doit fournir outre le nom de l'espèce, l'ensemble des informations y afférentes et contenues dans les différentes bases. La modification et la mise à jour des bases annexes sont aisées, celles de la base centrale et des règles de production, proche de la clé dichotomique, reste difficile, car certains changements peuvent altérer la structure même de la base et exiger sa réécriture partielle.

1. DETAIL DU SYSTEME EXPERT DEJA REALISE

Le système est organisé en blocs de contrôle de trois niveaux hiérarchiques (Figure 3).

Le premier bloc est prévu pour tester les intentions de l'utilisateur. Non véritablement opérationnel pour l'instant, il pourra, entre autres, permettre de choisir entre le chaînage avant ou arrière.

Le bloc "clé-groupe" permet de déterminer la "variable-but" intermédiaire GROUPE-POISSON, c'est à dire de choisir entre un des quatre genres ou le

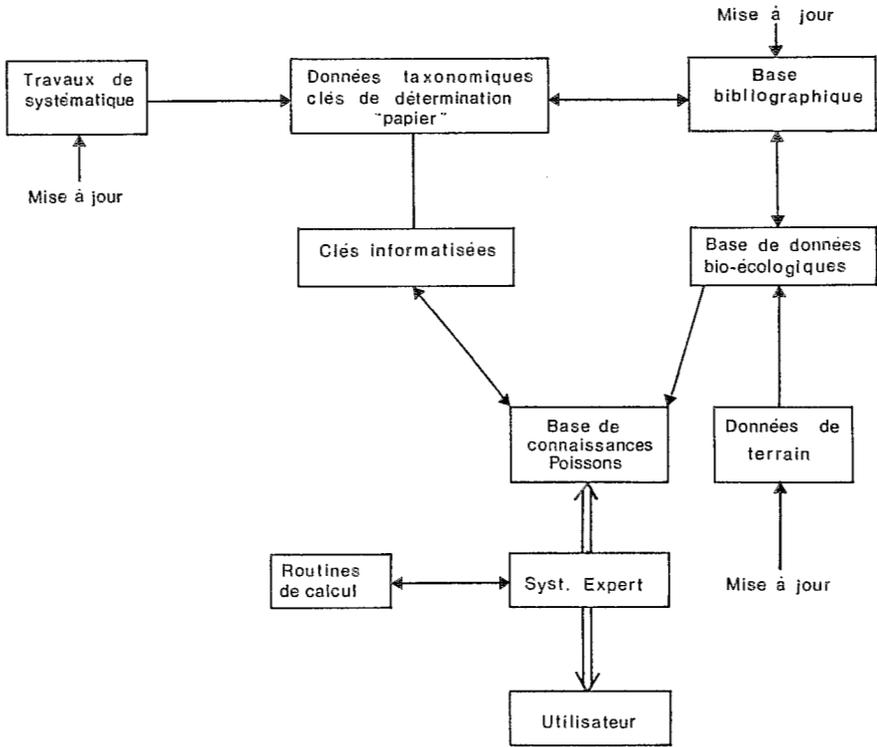


Fig. 2: Schéma d'une base de connaissances sur les poissons d'eau douce couplée à un S.E.

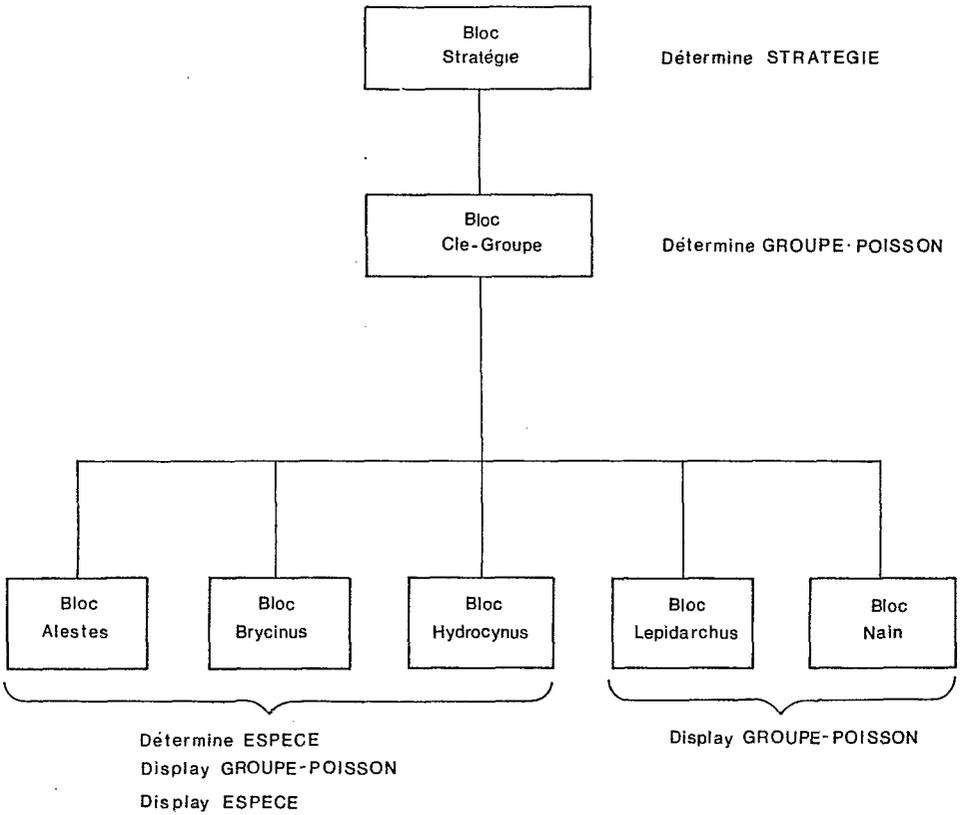


Fig.3 Hiérarchie des Blocs de contrôle et textes de contrôle.
(Voir texte)

groupe des Characidés nains. Ce dernier n'étant pas pris en compte et le genre *Lepidarchus* ne comportant qu'une seule espèce, leurs blocs de contrôle respectifs ne comportent qu'un ordre d'affichage mais pas d'autre ordre "détermine".

A titre d'exemple, je présente le bloc le plus simple : le bloc *Hydrocynus* qui comprend cinq espèces. Le tableau 2 indique pour ce bloc : (1) les paramètres qui y apparaissent pour la première fois, (2) le nom des règles de production (une par espèce), le texte de contrôle du bloc (qui indique le but et les actions à effectuer). S'il existait un bloc "descendant", le dernier ordre dirait de passer la main à un tel bloc ("establish" ..), (3) le nombre de fois où il faut parcourir le bloc, (4) le bloc "parent", (5) un interrupteur pour la remise en ordre permanente des règles (s'utilise en chaînage avant), (6) un interrupteur permettant de fixer un temps d'exécution, (7) la liste des objets auxquels le bloc fait référence, (8) la liste des objets qui font référence au bloc.

La structure d'une règle est la suivante (Tableau 3) : (1) le texte de la règle avec les prémisses "IF...AND..." et une ou des conclusions "THEN..." ("clause d'action") ; (2) le bloc auquel la règle est rattachée ; (3) le type de règle ; (4) la liste des objets auxquels la règle fait référence et (5) la liste des objets qui font référence à la règle.

A tout moment il est possible d'interrompre le déroulement du programme pour faire appel à une base extérieure ou à une routine de calcul.

2. INTERET ET COUT HUMAIN DE LA REALISATION DE BASES DE CONNAISSANCES ET DE SYSTEMES EXPERTS EN ECOLOGIE

Par sa durée, la réalisation d'un système expert d'aide à la détermination "per se" ne se conçoit que dans des cas précis :

- existence d'un plan important de formation de techniciens et de chercheurs locaux. Le rapport intérêt/temps investi peut alors devenir favorable ;

- existence d'un programme à long terme au sein duquel il est prévu que se succèdent plusieurs biologistes dont la connaissance de la faune locale n'est pas acquise. Un système expert peut alors grandement faciliter la tâche des nouveaux arrivants ;

- existence d'une volonté de suivi à long terme, censé se poursuivre après le retrait de l'ORSTOM. La réalisation d'un système expert peut s'avérer intéressante pour fournir à la contrepartie locale un outil de formation et de travail (programmes de contrôle biologique à long terme en particulier, suivi des lacs de barrage, etc.). Dans ces cas, si le produit est bien fini, il est envisageable de vendre le système expert à l'organisme finançant l'opération ou à la société d'exploitation.

En revanche, il serait souhaitable que tout biologiste venant à opérer une révision systématique d'un groupe ou de la faune ou flore d'une région débouchant sur une nouvelle clé de détermination le fasse sous forme de système expert. Avec un bon générateur, ce travail n'implique pratiquement aucune surcharge de travail (hormis l'apprentissage du fonctionnement du générateur) et peut même le faciliter et constituer un excellent test de la cohérence de la clé.

La réalisation de "simples" bases de connaissances bio-écologiques est d'un tout autre ordre. Liées ou non à un système expert, ces bases permettraient de regrouper les informations dispersées sur un grand nombre de supports parfois introuvables, voire non publiées, et destinées à disparaître avec la cessation d'activité du chercheur.

FOCUS CONTROL BLOCKS

CONTROL BLOCK:F_HYDROYNUS

Parameters PARAMETER:P_H_ESPECE
PARAMETER:P_H_DENTS

Rules RULE:R_HTANZANIAE
RULE:R_HBREVIS
RULE:R_HVITTATUS
RULE:R_HFORSKALII
RULE:R_HGOLIATH

Control text display p_groupe_poisson;determine p_h_espece; display p_h_espece;

Max instances 1

Parent F_CLE_GROUP

Dyn Rule Order FALSE

DisposeWhenDone FALSE

If ref it list PARAMETER:P_H_ESPECE-DisplayToParam
PARAMETER:P_H_ESPECE-DetermineToParam
PARAMETER:P_GROUPE_POISSON-DisplayToParam
RULE:R_HGOLIATH-FCBToRules
RULE:R_HFORSKALII-FCBToRules
RULE:R_HVITTATUS-FCBToRules
RULE:R_HBREVIS-FCBToRules
RULE:R_HTANZANIAE-FCBToRules
PARAMETER:P_H_DENTS-FCBToParams
PARAMETER:P_H_ESPECE-FCBToParams
FCB:F_CLE_GROUP-DesctoParent
It ref me list FCB:F_CLE_GROUP-ParenttoDesc

Tableau 2 : Exemple de structure d'un bloc de contrôle dans ESE.

RULE:R_HBREVIS

Rule text IF p_h_dents = '12' and p-h-ecailles = '3'
 THEN p_h_espece = 'H.brevis'

Owning FCBs :FCB:F_STRATEGE
 :FCB:F_CLE_GROUP
 x :FCB:F_HYDROCYNUS
 :FCB:F_NAIN
 :FCB:F_ALESTES
 :FCB:F_BRYCINUS
 :FCB:F_LEPIDARCHUS

Rule type Inference

I ref it list PARAMETER:P_H_ESPECE-LHSPalnThen
 PARAMETER:P_H_ECAILLES-TestPalnPrem
 PARAMETER:P_H_DENTS-TestPalnPrem

It ref me list FCB:F_HYDROCYNUS-FCBToRules

Tableau 3 : Exemple de structure d'une règle dans ESE.

Au niveau de ceux-ci, la réalisation de bases de données individuelles pour usage personnel est courante. Une politique de relative standardisation et de mise en place de bases de connaissances régionales, gérées ou non par des systèmes experts, augmenterait les contacts transversaux et favoriserait l'échange d'informations au sein des UR et entre celles-ci. Elle fournirait en outre :

- des outils de formation des jeunes chercheurs et techniciens ;
- des outils de travail pour les chercheurs de spécialités voisines ou changeant d'aire géographique d'étude ;
- les maillons d'un réseau de bases de connaissances internationales.

La mise en place de bases de connaissances régionales trans-disciplinaires passe par l'existence d'une politique informatique souple qui favorise une relative standardisation sans imposer trop de contraintes rebutantes. C'est là un des "serpents de mer" du développement informatique des grandes institutions ou sociétés. L'expérience a montré que les services informatiques centraux ont une tendance naturelle à développer selon une finalité et une problématique propres, en réduisant les liens avec les utilisateurs extérieurs. Une autre tendance étant le repli sur l'informatique "administrative", par nature plus codifiée et moins "fantasque" que celle des chercheurs.

Par ailleurs, le développement et les prix de la micro-informatique ont habitué les utilisateurs à une très grande variété de configurations matérielles et surtout logicielles. Toute tentative de standardisation sera nécessairement ressentie comme une brimade et tout choix contesté.

Cependant, deux voies qui, à ma connaissance, ne sont pas utilisées systématiquement pourraient être explorées :

- la standardisation matérielle est en fait une réalité (deux options possibles). Néanmoins, lors de l'achat de quantités importantes de matériel informatique (équipement de services au siège, d'un laboratoire, etc.) des contrats sont négociés (ou devraient l'être) pour un achat groupé à prix réduit. Une enquête préalable pourrait permettre à ceux à qui la configuration convient, de s'équiper d'un matériel donné, au meilleur prix sans que cela apparaisse comme une contrainte ;

- la question des logiciels est plus complexe. Cependant, il devrait être possible, après enquête, de négocier auprès des éditeurs de logiciels des licences d'utilisation multisites pour l'ensemble du personnel de l'Institut. Trois conditions essentielles devraient présider à tout choix : (1) être un logiciel répandu internationalement, (2) être d'un prix abordable, (3) pouvoir transmettre facilement les données sous forme utilisable sur un mini-ordinateur.

Au sein de l'Institut, l'augmentation des échanges transversaux et la réalisation d'outils de formation sont des points d'intérêt peu discutables. Mais, et c'est l'un des points de réflexion du présent séminaire, quel est l'intérêt de bases de connaissances au plan international ? Le passage d'un "outil de travail et de formation", à usage interne de quelques dizaines de personnes, vers une base connectée à un réseau international est un saut qualitatif important. Il ne se justifie

à mon avis qu'en liaison avec les efforts d'autres organismes, notamment le MNHN, dans le cadre du programme existant d'informatisation et d'interconnexion entre les Museums européens. Celui-ci devrait aller outre les aspects "conservation" et "catalogue" de spécimens pour devenir une vraie base de données bio-écologiques.

En tout état de cause, de tels projets impliquant plusieurs chercheurs (pour la réalisation comme pour l'alimentation et la mise à jour périodique des bases), demande un engagement des commissions scientifiques, autant dans le cadre de leurs attributions relatives à l'animation scientifique et au développement des échanges, que pour la valorisation du travail fourni par les chercheurs pour la réalisation des systèmes.

CONCLUSIONS

Je reviendrai pour terminer sur le travail que nous avons réalisé. Les systèmes experts ont des avantages certains sur les autres procédés en tant qu'outils d'aide à la détermination. Le choix entre tel ou tel système expert se fait en fonction du critère que l'on souhaite privilégier, l'objet et sa description (systèmes centrés-objet), ou la démarche de l'expert (systèmes à base de règles de production).

Nous avons choisi cette dernière approche parce que :

- les outils informatiques y sont plus développés actuellement et que notre approche est plus biologique et utilitaire qu'informatique. Les systèmes centrés-objet sont encore entre les mains des chercheurs en informatique. Ce que nous ne sommes pas ;

- ces systèmes permettent très facilement l'accès aux bases et routines extérieures ;

- il nous a paru important dans cette première approche de décortiquer le mécanisme du "savoir" de l'expert biologiste, et c'est à cela que nous avons tout d'abord consacré nos efforts.

Les premières conclusions auxquelles nous sommes parvenus, et qui sont brièvement citées au début de ce texte, nous posent cependant un problème fondamental, celui de l'"analyse d'images" réalisée par l'expert biologiste pour orienter sa diagnose. La reconnaissance visuelle d'un individu intègre très rapidement un grand nombre de paramètres sans que l'expert en ait même véritablement conscience. Cela se traduit par une phrase du type "cette individu a un aspect de X, ou ressemble à un X". Or, dans l'état actuel, aucun système informatique n'est en mesure de réaliser une telle opération. Et le seraient-ils un jour, que de tels systèmes resteront longtemps d'un coût élevé et disproportionné pour l'usage envisagé ici.

Si en poussant plus loin l'analyse de la démarche de l'expert, on le prive du contact visuel avec l'échantillon (en passant par un intermédiaire que l'expert interroge par exemple), la démarche change. L'expert privilégie les aspects bio-écologiques (lieu de récolte en particulier,...) pour réduire l'éventail des possibili-

tés. Puis il cherche à connaître l'aspect général de l'individu à déterminer (est-il trapu, élancé, etc.). Mais ces notions sont subjectives, et si l'on cherche à les quantifier l'on s'aperçoit vite que la variabilité peut être importante au sein d'une même espèce. Chez les poissons par exemple, entre des individus bien ou mal nourris, mâles ou femelles, et pour un même sexe en fonction de l'état de maturité. Passé ces stades, l'expert rentre beaucoup plus vite dans une démarche de type "consultation de clé" que lorsqu'il voit l'objet à déterminer. La présentation d'écrans d'images ou de silhouettes schématisant les diverses formes pouvant se rencontrer, pourrait peut-être faciliter la reconnaissance dans les premières étapes de la diagnose. Mais pas dans tous les cas. Dans celui des Characidés par exemple, l'expert discrimine immédiatement des individus dont l'allure générale est très proche.

Devant la difficulté de formaliser le savoir et la démarche "réelle" de l'expert, ma conclusion serait que pour toutes ces approches d'application des systèmes experts à la biologie il faut garder présente la très grande complexité de tout ce qui touche au vivant. Pour en tenir compte, le moyen terme n'est pas admis. Il faut soit adopter une démarche ultra-réductrice et se contenter (dans notre cas) d'une clé dichotomique plus ou moins optimisée, soit faire un grand bond qualitatif et profiter des atouts des systèmes experts, pour gérer des bases de connaissances variées qui cherchent à être aussi exhaustives que possible.

Ceux qui se sont intéressés à la modélisation des interactions biologiques, se sont vite retrouvés face à des phénomènes analogues. Des modèles "simples" peuvent être extrêmement féconds (Lotka-Volterra, modèles de gestion des stocks, etc.), mais dès que l'on veut aller plus loin, il faut aller très vite très loin.

REFERENCES BIBLIOGRAPHIQUES

- WOOLEY, J.B. & STONE, N.D., 1987. *Application of artificial intelligence to systematics : SYSTEX, a prototype expert system for species identification.* Syst. Zool., 36(3) : 248-267.

TRAITEMENT DE DONNEES SATELLITAIRES POUR LA CONNAISSANCE DE L'ENVIRONNEMENT EN MILIEU TROPICAL : APPLICATIONS EN GUYANE.

(Marc LOINTIER)

RESUME - Premier système de traitement d'images satellites installé dans la zone Antilles-Guyane, le laboratoire de télédétection du centre ORSTOM de Cayenne est exclusivement tourné vers les applications en milieu tropical, dans un espace régional défini par chaque discipline de recherche (botanique, sédimentologie, pédologie, hydrologie, etc.).

Après un rappel des caractéristiques de la machine, on présente l'apport de la donnée satellite aux études hydrologiques en zones d'estuaire et en rivière, qui font apparaître la nécessité de connaître la dynamique des bancs de vase se déplaçant d'est en ouest le long du littoral, conditionnant l'extension de la remontée saline vers l'amont. Enfin, une relation radiométrie/turbidité est testée à partir des données Landsat MSS, acquises en simultanéité avec des mesures de la qualité de l'eau.

INTRODUCTION

L'utilisation des données de télédétection aérospatiale pour la connaissance de l'environnement est un domaine qui a reçu une impulsion nouvelle ces deux dernières années, par la mise en orbite avec succès des satellites français SPOT. L'état de l'art en matière de traitement et d'interprétation des images s'est enrichi, depuis une quinzaine d'années, par l'exploitation des multiples informations obtenues avec les satellites LANDSAT, depuis leur lancement. La "banque de données radiométriques" que constitue l'image, ne prend réellement corps, aux yeux des diverses spécialités, qu'après une combinaison judicieuse des canaux et des traitements, dont l'aboutissement est la cartographie spécifique, répondant à un objectif précis. C'est donc devenu un "savoir faire", en relation avec le spécialiste de terrain, que l'évolution des matériels et des logiciels ont rendu opérationnel et perfectible. Ce processus a permis le rapprochement de "l'homme de terrain" avec la vision satellitaire, ce qui assure une plus-value certaine, en regard des démarches habituelles.

Pour la Guyane, riche en spécialistes de terrain, cet outil présente un intérêt évident, connaissant les difficultés d'exploration de certaines zones peu cartographiées, et pour réaliser les études nécessaires à estimer l'impact de nouveaux aménagements sur l'environnement.

Ce faisceau de constats a donné lieu à une série d'actions, engagées dès 1986, auprès de l'Etat et de la Région Guyane. Ces démarches ont été concrétisées par l'implantation, en janvier 88, d'un système de traitement d'image numérique à Cayenne. Celui-ci remplit les quatre fonctions essentielles d'un tel système : lecture-stockage des données radiométriques, visualisation, traitement numérique et production de documents de travail et de documents élaborés.

1. ASPECTS TECHNIQUES DU LABORATOIRE DE TRAITEMENT D'IMAGES

Depuis 1988, le système fonctionne avec l'appui des structures et des moyens de l'ORSTOM. Le matériel est un Péricolor 1500, d'origine MS21, qui possède un éventail de logiciels spécifiques au traitement des données de télédétection. Les raisons qui ont guidé ce choix ont largement tenu compte de l'éloignement du site de la métropole, de l'aspect opérationnel de la machine munie de ses périphériques et des utilisations pédagogiques possibles (stages, accueil de chercheurs étrangers, etc.).

1.1. Description du matériel et caractéristiques techniques

L'unité centrale est un PERICOLOR 1500, ordinateur muni des processeurs suivants :

- 16 bits programmable ;
- arithmétique 32 bits flottants ;
- parallèle 128 bits programmable ;
- matriciel 32 bits flottants.

Il possède 6 Mégaoctets de mémoire vive, et un disque dur de 20 Mégaoctets. Un lecteur de disquettes permet l'échange d'information avec l'extérieur. Le clavier est muni d'une "boule roulante", et le moniteur couleur permet la restitution en haute résolution d'une surface de 512 x 512 pixels.

Les périphériques sont constitués :

- d'une streamer ;
- d'un dérouleur de bandes 1600/6250 BPI ;
- d'une tablette graphique 46 x 30 cm. ;
- d'une imprimante noir et blanc ;
- d'une imprimante couleur à jet d'encre, ACT ;
- d'une carte d'extension pour couplage avec un autre ordinateur

(SUN 4).

Les logiciels comprennent :

- un logiciel de base pour le traitement d'image ;
- un logiciel spécifique Télédétection (Géopéricolor).

1.2. Principales fonctions du logiciel "télé-détection"

Les fonctions de visualisation permettent l'affichage multicanaux, panoramique (sur 16 plans 512*512), composition colorée rapide, opérateur en temps réel sur les quatre opérations arithmétiques, histogrammes, profils, tables de fonction vidéo (LUT), optimisation locale de densité, visualisation d'images 16 bits, animation sur 48 plans.

Les traitements possibles sont :

- des anamorphoses géométriques (angulaires ou par amers) ;
- le filtrage linéaire et l'utilisation des filtres spéciaux (Kirsch, Sobel, Robert, Gradient) ;
- l'élaboration d'index de végétation et brillance ;
- la fusion d'images et masques ;
- la combinaison linéaire de canaux ;
- la conversion RVB en Intensité Teinte et Saturation (ITS) ;
- la corrélation d'images ;
- l'analyse en composantes principales ;
- la classification par minimum de distance ;
- la classification par maximum de vraisemblance ;
- l'apprentissage sur parcelles test ;
- la classification par découpage d'un histogramme bidimensionnel.

1.3. Les données disponibles en 89

Trois scènes Landsat d'octobre 1976 et deux scènes SPOT d'octobre 1986 ont été acquises depuis la mise en oeuvre de l'atelier. Il s'agit des scènes suivantes repérées selon la grille de référence :

- LANDSAT-244 056 (région Cayenne-Iracoubo), du 30 Octobre 1976 ;
- LANDSAT-245 056 (centrée sur estuaire du Maroni), du 31 Octobre 1976 ;
- LANDSAT-245 057 (centrée sur Maripasoula), du 31 Octobre 1976.

Pour ces images, les caractéristiques communes sont les suivantes : élévation solaire de 50 degrés, azimuth de 120 degrés, très peu nuageuse et de bonne définition pour la qualité, capteur MSS et résolution de 80 mètres, prise de vue à 09h02 (heure locale).

- SPOT-691 340 (région Cayenne-Kourou), du 20 Octobre 1986.
- SPOT-689 339 (région Sinnamary), du 20 Octobre 1986.

Ces deux scènes ont été respectivement acquises par HRV2 et HRV1, en même temps, avec un dépointage du capteur de 26 degrés. L'heure locale de prise de vue est 11h16.

Pour ces dernières images, les caractéristiques communes sont les suivantes : mode XS, canaux 1, 2 et 3, résolution 20 mètres, couvert nuageux important sur les zones de forêt, pour la qualité. L'heure de la basse mer, aux îles du Salut, était à 12h10.

Une collaboration avec les organismes de recherche (BRGM, ENGREF, CNES) et d'aménagements (EDF, DAF, DDE) permet de constituer une banque de données "images", sur support magnétique, qui s'étoffe à chaque projet nouveau. Ainsi, trois nouvelles scènes Landsat Thematic Mapper (TM), d'une résolution de trente mètres sur six canaux, ont été acquises récemment.

1.4. Consultation catalogue

Plusieurs consultations des images existantes ont été demandées au cours de l'année, notamment auprès de l'INPE au Brésil, gérant les demandes concernant LANDSAT TM sur l'Amérique Latine.

Pour SPOT, les consultations catalogue sont facilitées par l'accès via le réseau TRANSPAC, au serveur BRAMS de Toulouse. Cette possibilité nous a permis de suivre l'évolution des scènes acquises pendant les périodes correspondantes aux demandes de programmation.

2. ETUDES DES ESTUAIRES ET DES FLEUVES DE GUYANE A L'AIDE DES TRAITEMENTS DE DONNEES LANDSAT MSS ET SPOT.

Le littoral de la Guyane appartient à l'un des plus importants systèmes mondiaux de côtes basses et envasées, entre l'Amazonie et le delta de l'Orénoque (PROS M.T., 1988). De nombreuses observations (YAYER, 1937 ; BOYE M., 1961) et études (DELFT L.H., 1962) ont déjà été réalisées et font référence en la matière. L'interprétation et le traitement des données satellites apportent une information globale, homogène dans l'espace et surtout quantitative. La migration vers l'ouest des bancs de vase et des espaces interbancs, à une vitesse moyenne de l'ordre du kilomètre par an, introduit de profondes modifications morphologiques et sédimentaires le long des rivages et des marais (LOINTIER M. & PROST M.T., 1986), mais aussi dans les estuaires et leurs embouchures (LOINTIER M. & ROCHE M.A., 1988) : les effets océaniques - qui se font sentir en général jusqu'à 30 km vers l'amont - deviennent plus ou moins prépondérants selon la position du banc de vase, ayant pour principale conséquence la transformation des effets dynamiques de la marée et de l'extension de la remontée saline.

Les études antérieures décrivant les systèmes estuariens deviennent la plupart du temps obsolètes et il convient de les situer dans leur nouveau contexte, surtout lorsque l'on raisonne en termes d'aménagement de la plaine côtière.

2.1. Composition colorée standard (Scène LANDSAT MSS du 30 octobre 1976)

L'échelle de travail avec Landsat MSS (pixel de 80 m) est très adaptée à l'étude des processus côtiers globaux et à la classification des grandes unités thématiques de la zone littorale. On observe, en traitement standard, la partie nord du bouclier Guyanais et de la plaine côtière, qui s'étend sur 300 km, entre l'Oyapock et le Maroni. Le bouclier a une altitude moyenne comprise entre 250

et 300 mètres, et il est recouvert de forêt primaire. La plaine côtière est composée de la "plaine côtière ancienne" d'âge Pléistocène terminal et de la "plaine côtière récente" d'âge Holocène. On notera la position particulière des estuaires du Sinnamary, de l'Organabo et de la Mana, dont les cours sont déviés vers le nord-ouest, à la suite du transit côtier. Les panaches turbides à la sortie des embouchures sont également bien visibles. Il s'agit de la remise en suspension des vases, sous l'effet des courants de flot et de jusant. Une estimation des quantités de matière en suspension (LOINTIER M., 1988) montre que la contribution locale des fleuves et des rivières à la sédimentation côtière est très faible, et ne représente que 1 % des apports d'origine Amazonienne. Ces derniers sont estimés entre 11 et 13.10⁸ tonnes par an, dont 10 à 20 % se déplacent le long des côtes de Guyane (EISMA D., 1988).

2.2. Composition colorée standard (Scène SPOT du 20 octobre 1986)

A une échelle plus grande, SPOT (pixel de 20 m), permet de mieux identifier la succession des zones en érosion (Sinnamary), de celles en phase de dépôt (Ile de Cayenne et zone de Kourou). La structure de la plaine côtière récente apparaît nettement, avec des éléments morphologiques importants, comme les anciennes lignes de rivages (ou *cheniers*) qui séparent les différentes phases d'accumulation fine holocène. Cette définition permet aussi d'analyser les différentes masses d'eau des estuaires et des rivières de plus de 80 mètres de largeur.

2.3. Utilisation des effets induits par l'acquisition oblique des données SPOT.

Le dépointage de 26 degrés du radiomètre, lors de la prise de vue du 20 octobre 1986, met en évidence les ondes, comme la houle, se propageant à la surface de l'océan (WADWORTH A. & PETIT M., 1987). Le filtrage par une matrice de réhaussement améliore le document d'origine. Nous avons choisi ici une formule du type :

$$\begin{array}{ccc} -1 & -1 & -1 \\ -1 & 9 & -1 \\ -1 & -1 & -1 \end{array}$$

qui fait apparaître la houle et sa diffraction (figure 1) à son passage près de l'embouchure, ainsi que le panache turbide du fleuve Sinnamary, d'une surface de 10 km².

2.4. Traitement multitemporel n° 1

Données LANDSAT 76, SPOT 86 et TM 88, sur la zone de Sinnamary.

Le traitement multitemporel de trois scènes, prises à dix et douze années d'intervalle, permet de donner, avec une précision de quelques pixels, des éléments quantitatifs sur les transformations du littoral et sur l'étendue des zones intertidales, aux différentes dates (LOINTIER M. & PROST M.T., 1988). La superposition est réalisée par translation à l'aide de plusieurs points d'appui, après rotation des images et réduction à la même échelle (figure 2).

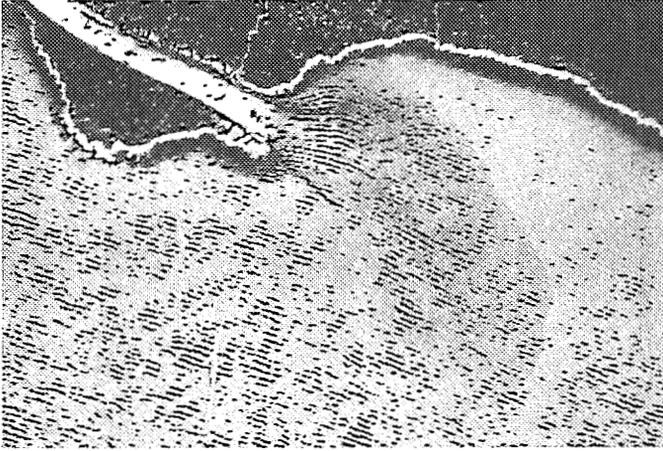


Figure 1

Exemple de réhaussement de dynamique d'une image SPOT prise en visée oblique (26°). La houle et sa diffraction au passage de l'embouchure sont nettement visibles, ainsi que le panache du fleuve Sinnamary, d'une surface de 10 km².

Copyright CNES 1986.

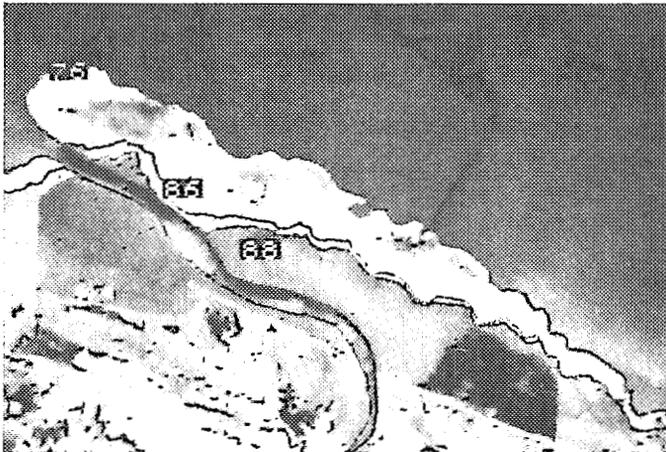


Figure 2

Superposition et recalage de trois fichiers-images, de 1976, 1986 et 1988, sur la région de Sinnamary, en phase d'érosion. Le recul de la côte et les surfaces perdues sont appréciés à l'hectare près ou mieux, selon le type de capteur. Ici, l'image de 76 est d'origine Landsat MSS (pixel de 80 m), celle de 86 est d'origine SPOT (pixel de 20 m) et Landsat Thematic Mapper (pixel de 30 m) pour 1988.

Copyright CNES 1986 INPF 76 88

On peut calculer ainsi que 40 Km² de mangrove ont disparus dans cette zone, où le littoral a reculé en moyenne de 2 km en 10 ans.

La zone la plus étroite de l'estuaire, s'est rompue en mars 87, provoquant un "raccourcissement" de 4 km du cours de l'estuaire du Sinnamary.

2.5. Traitement multitemporel n° 2

Données LANDSAT 76 et SPOT 86 sur la zone de Kourou.

A l'inverse du résultat précédent, le phénomène observé ici correspond à une progradation de la ligne de rivage de 2 km, consécutif au déplacement d'un banc de vase pendant les années 80-85, sur lequel la mangrove s'est installée très rapidement (figure 3).

2.6. Aspect qualitatif de la distribution des eaux de surface de l'estuaire du Sinnamary, en étiage

Données LANDSAT 76 et SPOT 86.

Les mesures de terrain (LOINTIER M. 1984) dans cet estuaire montrent que l'opacité est élevée, avec une disparition du disque de SECCHI entre 0,5 et 1,2 mètre. Les profondeurs mesurées étant supérieures au mètre dans tous les cas, on ne peut pas attribuer les variations de la radiométrie enregistrée aux variations bathymétriques, surtout avec les canaux utilisés dans le rouge et l'infrarouge.

L'index de Brillance ($\sqrt{C_x^2 + C_y^2}$), calculé sur la zone d'estuaire, dans les bandes Landsat 5 et 7 et SPOT 2 et 3, quasiment homologues, permet de classer les pixels selon une échelle de valeurs croissantes qui traduit l'augmentation des valeurs radiométriques. Cet index montre ainsi, qualitativement, la distribution des masses d'eau de surface. Pour un débit proche (64 m³/s le 30 octobre 76 et 80 m³/s le 20 octobre 1986), il est remarquable de constater la similitude de cette distribution, à dix années d'intervalle, avec quelques nuances dues probablement aux heures de marée différentes. On notera que les faibles valeurs sont situées en rive droite, les plus élevées en rive gauche, au niveau de l'élargissement de l'estuaire. Ceci traduit une remise en suspension des vases plus importante par petits fonds (entre 1,5 et 2 mètres).

2.7. Extension des phénomènes océaniques vers l'amont

LANDSAT 76, composition colorée standard.

Les conditions de prise de vue du cliché Landsat, à marée haute et en période d'étiage, correspondent aux conditions les plus favorables de l'extension maximale vers l'amont de la remontée des vases d'embouchure et du coin salé. L'étude hydrologique de l'estuaire du KOUROU (ROCHE M.A., 1976), et du SINNAMARY a permis de constater que les limites vers l'amont de ces deux phénomènes étaient voisines, à quelques centaines de mètres près. La composition colorée standard met parfaitement en évidence le contact entre les eaux fluviales et celles qui ont subi un mélange océanique. Par ce biais, on détermine ainsi la limite de la remontée saline dans les deux estuaires.

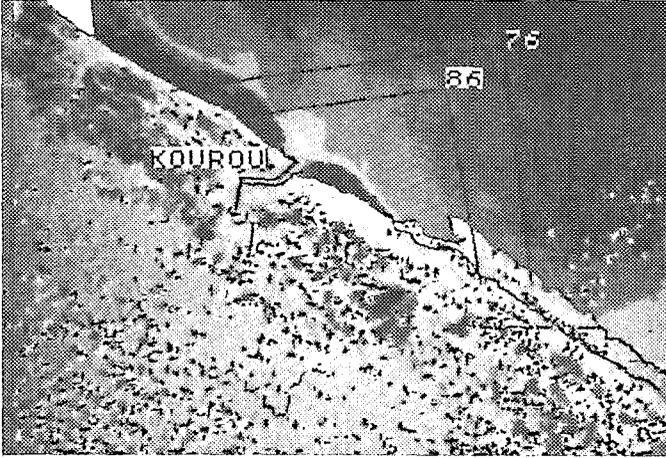


Figure 3

Exemple de côte en progradation, sur la zone de Kourou. Comparaison obtenue par superposition des images de 1976 et 1986. En noir, la position du banc de Kourou en 1986, dans sa partie distale.

Copyright CNES 1986, INPE 76.



Figure 4

Index de Brillance appliqué aux données SPOT, sur le Sinnamary, dont les valeurs sont croissantes du noir vers le blanc. on observe des discontinuités dans la distribution, des radiométries, dans la partie intermédiaire de l'estuaire. Cette première étape d'un traitement montre l'hétérogénéité du mélange des eaux.

Copyright CNES 1986.

Sur la composition colorée, on définit la limite amont du contact entre les eaux fluviales et les eaux mélangées. L'extension est plus importante vers l'amont pour le fleuve Kourou que sur le Sinnamary : ceci a été bien observé dans nos études, et s'explique par la valeur différente du rapport du volume d'eau douce apporté par le fleuve au volume déplacé dans l'estuaire pendant un cycle de flot. Ce rapport, V_{riv}/V_{flot} , est de 20 % pour le Sinnamary, et seulement de 2 % pour le Kourou, qui subit une influence océanique plus grande.

2.9. Traitement des données SPOT (1986) sur la partie médiane de l'estuaire du Sinnamary

Les études hydrologiques (vitesse, débits oscillants, salinité et charge en suspension) ont conduit à penser que la propagation des éléments présents dans l'estuaire (en particulier les vases) s'effectuait vers l'amont par oscillations et dilution successives au cours de plusieurs cycles de marée (ROCHE M.A., 1977). Sans toutefois confirmer une telle hypothèse, un index de Brillance appliqué aux données SPOT, dont les valeurs sont croissantes du noir vers le blanc sur le document (figure 4), montre des discontinuités dans la distribution des radiométries, dans la partie intermédiaire de l'estuaire proprement dit.

Ces observations sont utiles pour juger de la qualité du mélange des eaux et permet d'évaluer ainsi les limites d'un modèle mathématique linéaire de propagation des éléments dissous.

Une vérification de terrain serait très intéressante, mais on sait combien il est illusoire de réaliser des mesures le jour d'un passage satellite, à moins de mettre en place un dispositif permanent d'observation, ce qui ne se justifie pas encore pour nos études.

2.9. Qualité des eaux

Exploitation d'une simultanéité entre des mesures de qualité des eaux sur le Kourou et la vue Landsat : une coïncidence exceptionnelle.

Une partie des prélèvements d'eau réalisés pendant l'étiage de 76 sur le Kourou se sont déroulés la veille et le jour même du passage du satellite à 9h, heure locale, sur cinq sites répartis entre l'embouchure et 30 km en amont.

En dépit du faible nombre de pixels situés en rivière et sans effets de bordure, il a été testé une relation entre la radiométrie et la turbidité. Après plusieurs combinaisons de canaux, (permutations circulaires des canaux dans le calcul des index de végétation et de brillance), c'est dans la seconde composante d'une ACP sur les quatre canaux que l'on trouve une corrélation acceptable, dans laquelle le canal 6 s'est révélé être très important. Ce calcul est effectué à partir des valeurs radiométriques "brutes", comme l'a étudié RITCHIE (1987), qui démontre également l'importance du canal 6 (0,7 - 0,8 μm) pour définir ce type de relation.

La combinaison linéaire retenue sur les quatre canaux est de la forme :

$$P_i = S(i) + K(i, 1) * Q_4 + K(i, 2) * Q_5 + K(i, 3) * Q_6 + K(i, 4) * Q_7$$

avec P_i , valeur du pixel dans la seconde composante ($i = 2$),
 $S(i) = 103,187$,

$K(i,1)=0,7896$, $K(i,2)=0,7244$, $K(i,3)=-0,1787$, $K(i,4)=0,5776$,
 $Q_{4,7}$ valeur radiométrique de chaque canal MSS 4,5,6,7.

Les valeurs radiométriques, repérées sur chaque site, sont comparées aux mesures in situ. La relation obtenue est :

$$\log(T_{\text{urb}}) = 0,0495 * P_{2-4} - 4,6709 \quad r=0,99 \text{ (figure 5)}$$

Cet étalonnage permet ensuite de classer tous les pixels du fleuve (figure 6) et de son estuaire, en cinq groupes de turbidité, avec une incertitude importante (50 %), mais qui ne doit pas surprendre, vu les conditions expérimentales :

Gris 1	T <	10 mg/l
Gris 2	10 < T <	100 mg/l
Gris 3	100 < T <	1000 mg/l
Gris 4	1000 < T <	10000 mg/l
Gris 5	T >	10000 mg/l

La relation exploitée ici s'accorde avec celles qui sont exposées dans la littérature, de type $R = f(\log T)$, avec un résultat "cartographique" en accord avec les observations. A partir de la même combinaison linéaire des quatre canaux, la relation a été appliquée aux eaux de l'estuaire du Sinnamary (figure 7), montrant la même distribution décalée vers l'aval (classes 3, 4, 5), à cause du débit plus élevé.

CONCLUSIONS

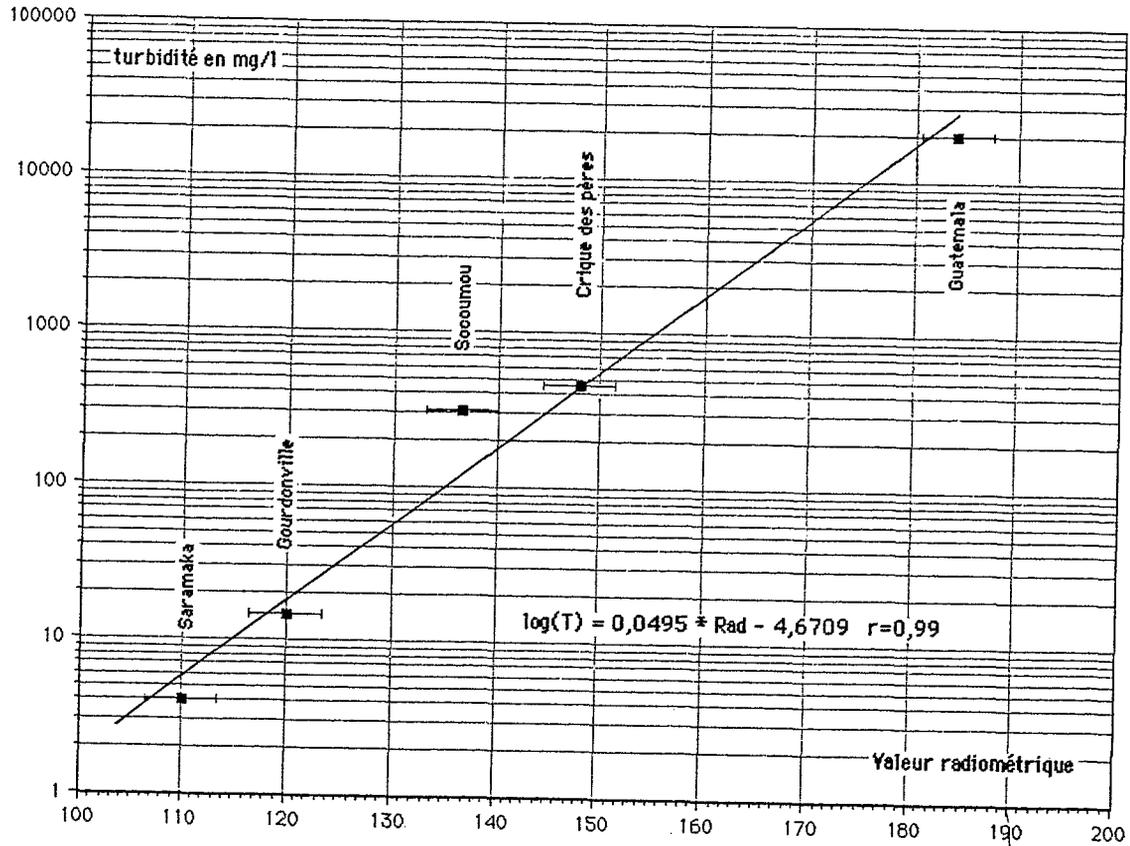
Toutes ces applications ont été développées avec le logiciel implanté dans le Péricolor (Géopéricolor) qui permet de réaliser la plupart des traitements utiles en phase d'analyse des relations radiométriques avec la thématique recherchée. Ainsi, la démarche souhaitée par un utilisateur non informaticien est rapidement confortée par le contact direct avec les données, qui lui permet de définir les principaux objectifs à atteindre, par rapport à sa problématique.

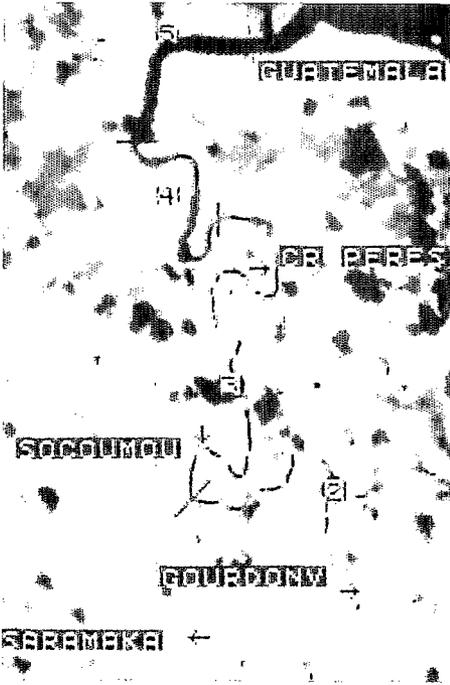
Cette encore une étape de "sensibilisation" et de prospection des possibilités de l'outil qu'il faut poursuivre, d'autant que les données sont en cours de collecte sur le Département. L'utilisation prochaine de données Landsat TM et l'extension des possibilités de traitement par le couplage à un mini-ordinateur, renforceront le potentiel technique et créatif, tout comme l'ouverture "horizontale" déjà engagée vers les autres disciplines.

En développant la banque de données satellitaires, il est constitué une banque de données sur l'environnement par le jeu des spécialistes présents capables de répondre demain aux problèmes de recherche pure, mais aussi aux effets des grands aménagements sur l'environnement : barrage de Petit Saut, base spatiale, développement des zones rurales, pénétrantes forestières.

En maîtrisant ce "savoir faire" propre au milieu, il devient possible de répondre à deux vocations spécifiques de l'Institut, en matière de recherche, d'échange et de transfert de technologie au niveau régional et international.

Fig 5: Relation radiométrie/turbidité





Copyright INPE 76.

Figure 6

Localisation des sites de prélèvement d'eau et limites des cinq zones de turbidité définies à partir de l'étalonnage radiométrique/turbidité, obtenue grâce aux vérités terrain réalisées le même jour que le passage du satellite (30 octobre 1976). En gris clair les faibles turbidités, en gris foncé les plus élevées.



Figure 7

Sur la même image Landsat 76, application de la classification précédente à un estuaire dépourvu de vérités terrain simultanées (Sinnamary).

Copyright INPE 76.

REFERENCES BIBLIOGRAPHIQUES

- BOYE, M., & CRUYS, H. 1961. *New data of the coastal sedimentary formations in French Guiana*. Proceedings of the 5th Interguiana Geol. Conf. Geological Survey Depart., Georgetown. 145-168.
- EISMA D., 1988. *Dispersal of Amazon Supplied particulate matter*. Chapman Conference on the fate of particulate and dissolved components within the Amazon Dispersal System: River and Ocean. AGU. Charleston. Wild Dunes.
- L.H. DELFT. 1962. *Demerara Coastal Investigation*. Report on siltation of Demerara bar channel and coastal erosion in British Guyana. Delft Hydraulics Lab., The Netherlands, 240 p.
- LOINTIER, M. 1984. *Dynamique des eaux et de l'intrusion saline dans l'estuaire du Sinnamary (Guyane Fr.)*. ORSTOM. Cayenne.
- LOINTIER, M. & PROST, M.T., 1986. *Morphology and hydrology of an equatorial coastal swamp: example of the Sarcelle Swamp in French Guiana*. Quaternary of South America and Antarctic Peninsula, vol.4. 59-77. J. Rabassa Ed. A.A. Balkema Publ. Rotterdam.
- LOINTIER, M. & ROCHE, M.A., 1988. *Quatrièmes journées hydrologiques de l'ORSTOM*. Colloques et Séminaires. 328 p.
- LOINTIER, M. & PROST, M.T., 1988. *Environnement côtier des Guyanes*. Rapport intermédiaire, ORSTOM, Nov. 1988, 70 p.
- LOINTIER, M. & PROST, M.T., 1988. *Coastal sedimentation and local rivers supply in French Guiana: comparisons with the Amazon*. Chapman Conference on the fate of particulate and dissolved components within the Amazon Dispersal System: River and Ocean. AGU. Charleston. Wild Dunes.
- PROST, M.T., 1988. *Shoreline changes in French Guiana*. Quaternary of South America and Antarctic Peninsula, vol 5. J.Rabassa Ed. A.A.Balkema Publ. Rotterdam.
- RITCHIE, J.C. et al., 1987. *Using Landsat MSS data to estimate suspended sediments in Moon Lake, Mississippi*. Remote sensing of Environment 23:65-81.
- ROCHE, M.A., 1977. *L'estuaire du Kourou en Guyane Française: possibilités d'alimentation en eau d'une usine de pâte de bois et risques de pollution par rejets industriels dans l'estuaire*. Rapport. ORSTOM. 80 pp.
- WADSWORTH A. & Petit M. 1987. *SPOT, un satellite pour l'océanographie ?* CNES. SPOT 1, utilisation des images, bilan, résultats.
- YAYER. 1937. *Compte-rendu de la mission pour l'établissement de la carte S.H.O.M. n° 5884 (région de Sinnamary)*.

Deuxième session : De la donnée à l'information, aide à l'interprétation.

(Rapport de session, par Patrick SECHET)

Chacun des cinq exposés de cette deuxième session a donné lieu à un rapide débat sur des points soulevés le plus souvent à l'initiative du conférencier. On retiendra essentiellement les quelques points suivants.

L'intérêt d'**utiliser des programmes commerciaux** (tableur par exemple), pour exploiter facilement les données stockées dans une banque et réaliser des *applications non prévues* (graphiques en particulier), et par conséquent non programmées lors de la mise en oeuvre du système. Il convient toutefois de se limiter à une utilisation des fonctions standards de ce type de produit, la programmation se révélant généralement plus lourde, plus restrictive et moins performante qu'avec un langage de programmation classique. Une étude comparative de ces produits commerciaux serait néanmoins nécessaire, d'autant qu'ils s'améliorent rapidement et de manière continue et leurs restrictions disparaissent progressivement.

A propos des **modèles numériques de terrain**, c'est sur ses applications potentielles que la discussion a surtout eu lieu. On peut en effet en déduire une cartographie des pentes, d'orientations, des courbures de la ligne de plus grande pente, mais aussi les réseaux de thalwegs ou de crêtes et les limites de bassins versants, tous paramètres fort utiles pour l'hydrologie de surface, par exemple. D'autres applications peuvent être réalisées dans d'autres disciplines, comme la géologie.

Les travaux en cours à l'ORSTOM ne diffèrent en fait que par les algorithmes choisis pour l'interpolation entre les points cotés : une mise en cohérence est prévue au niveau des fichiers de sortie. Actuellement ces études sont réalisées en utilisant des stations de travail : dans l'immédiat il n'est pas envisageable que ces applications soient opérationnelles sur PC-compatibles, surtout en raison des temps de traitement prohibitifs avec un processeur relativement peu performant.

Deux communications ont permis d'attirer l'attention sur les logiciels existants dans le commerce et permettant de réaliser des **études statistiques avec des données spatialisées** sur Macintosh. L'une et l'autre prouvent qu'il est possible d'obtenir d'excellents résultats à peu de frais, de sorte que la question est posée sur la réelle nécessité des systèmes d'information géographique. Une analogie est d'ailleurs faite avec les projets de réalisation d'atlas, qui constituent également des investissements importants, généralement non justifiés par une seule application. Le plus souvent, une approche plus pragmatique est possible et recommandable.

Le sujet de la **constitution de système expert** offre l'occasion de définir les conditions à satisfaire pour débiter un tel travail, souvent utile dans les domaines naturalistes où il existe une complexité taxonomique particulière : programme scientifique à long terme, établissement d'un plan de formation. Une dé-

cision politique est alors nécessaire et, de toutes façons, il n'est pas clairement établi qu'une systématisation des systèmes experts à l'Institut soit recommandée. Un participant souligne d'ailleurs qu'il existe une alternative nouvelle à la démarche du système expert basée sur des règles de production, celle des recherches actuelles sur la *conservation et la représentation de la connaissance*. On note en effet que l'expertise ne se présente pas à priori sous forme de règles, et qu'il faut réaliser un effort conceptuel important pour la formaliser sous cette forme.

Une discussion débute ensuite sur l'opportunité d'une normalisation des outils choisis pour ce type de travail (choix du logiciel pour le moteur d'inférence) : elle semble conclure dans cette voie, mais après une réflexion menée par discipline.

Session 3 :

**STRATEGIES DE
SPATIALISATION**

Troisième session : Stratégies de spatialisation.

(Présentation, par Gérard DANDOY)

Dans la présentation de ce séminaire sur les "systèmes d'information sur l'environnement", la troisième journée, intitulée "Stratégies de Spatialisation" est consacrée à une réflexion stratégique sur l'éventualité, pour l'Institut, de constituer une base ou de structurer un ensemble de bases sur l'environnement". Au vu de la multiplicité des initiatives dans ce domaine dont témoigne à l'évidence ce séminaire, la question paraît d'une actualité brûlante. Ce d'autant plus que ces nombreuses initiatives ressemblent à un foisonnement dont la cohérence ne paraît pas évidente. La diversité des thèmes abordés, des finalités, des contextes et des moyens mis en oeuvre pour la **collecte**, le **traitement** et la **restitution** des données spatialisées justifie pleinement la tenue de ce troisième séminaire et particulièrement de cette journée de réflexion globale sur nos choix scientifiques, techniques et méthodologiques dans le domaine des données spatialisées.

Or avant de présenter très globalement un point de vue sur ces huit communications, il nous paraît utile de vous faire part succinctement des conclusions d'un rapport récemment rédigé (1988), à la demande du CNES, sur l'état de l'art dans le domaine de l'information géographique, au sens large du terme s'entend, et plus particulièrement sur les systèmes d'information géographique dans le monde.

A. LES SIG : une solution d'avenir, dans quelle mesure ?

Sans nécessairement reprendre à notre compte tous les résultats de cette enquête, menée par le service de consultance en observation de la terre (SCOT) dans une optique bien particulière, à savoir définir le marché potentiel de SPOT par rapport aux SIG en place ou en projet..., il nous paraît nécessaire de vous faire part de certaines remarques propres à nous éclairer par rapport aux grandes tendances actuelles dans le monde du traitement informatique des données spatialisées.

1. Des disponibilités techniques nouvelles

D'abord un constat : "Jusqu'à très récemment les outils informatiques n'apportaient que des réponses partielles aux besoins de traitement et de représentation de l'information géographique ; mais la **conjonction des différentes techniques, au départ indépendantes**, comme les stations de travail graphiques, les bases de données relationnelles, les logiciels "CAO" (conception de dessin assistée par ordinateur) et les systèmes de traitement d'images (STI) permet maintenant de répondre à ce besoin".

En outre, l'originalité de ces outils récemment mis en œuvre de manière opérationnelle, début des années 80, consiste à construire et permettre la **combinaison** logique de **multiples plans d'informations multithématiques et multimédias** (cartes, relevés d'observations, statistiques localisées, images aériennes ou satellitaires...).

2. Des besoins bien cernés

Il faut également remarquer que la mise en œuvre de ce nouveau concept a correspondu initialement aux besoins de gestion opérationnelle de réseaux (gaz, électricité, circulation, etc), de communautés urbaines ou de **gestion** de périmètres forestiers.

Les impératifs de la gestion (précision : échelle de numérisation du 10.000 au 25.000 ; actualisation nécessaire, obligation légale bien souvent...) justifiaient la réalisation d'outils parfois très lourds pour **l'aide à la décision**. On notera que le contexte de réalisation de ces outils est le plus souvent bien différent de celui qui prévaut dans les organismes de recherche.

3. Des applications concrètes

Les réalisations concrètes et fonctionnelles dans ces deux domaines, gestion urbaine et foresterie, sont nombreuses en Amérique du Nord (USA et Canada) où l'on constate le fonctionnement actuel de plus de 100 gros systèmes.

Les exemples de réalisations concrètes dans les autres pays développés sont beaucoup plus rares. Le rapport SCOT relève quelques exemples très intéressants en Europe du Nord et surtout en Hollande (système général pour la planification nationale, INSYRON prenant en compte données du milieu et données sur la population et l'habitat). Les organismes internationaux en particulier la FAO s'intéressent aussi aux SIG. La FAO a en cours (période 85-90) un SIG portant sur 29 cartes thématiques au niveau du continent africain (échelle 1/5.000.000) pour la planification et la synthèse générale. Cet organisme favorise également des projets nationaux ou régionaux en Afrique et semble jouer un rôle pilote dans ce domaine.

La situation en France est évoquée dans ce rapport à travers le cas des SIG urbains (société ICOREM, Toulouse), du SIG expérimental conduit par le BRGM sur le vignoble du Vaucluse. Ce rapport cite enfin les tentatives de l'ORSTOM au Brésil et Mexique. Mais au niveau des réalisations, la **France n'apparaît guère à la pointe** dans ce domaine et à vrai dire elle paraît en retard (Cf. conclusion n°5 du rapport SCOT).

Pour résumer on constate donc :

1°) que de nombreux SIG se trouvent opérationnels comme outils de gestion et d'aide à la décision dans trois types d'applications thématiques :

- gestion des réseaux ;
- gestion urbaine ;
- gestion des périmètres forestiers.

2°) que d'autres applications fonctionnelles sont en cours de réalisation avec cette fois des objectifs de planification globale (ex. hollandais) ou d'études et synthèses (Cf. FAO).

3°) que de nombreux projets se font jour mais demeurent assez flous pour que l'on puisse dire, citant toujours le rapport SCOT, que "cette situation est caractéristique d'une *activité nouvelle qui s'appuie sur des outils informatiques non encore totalement maîtrisés du point de vue à la fois de leur définition et de leur emploi*".

4. Des outils sur le marché

L'offre industrielle, selon ce rapport, est dominée par trois constructeurs principaux : INTERGRAPH et SYNERCOM qui ont en commun d'avoir eu comme point de départ le marché du dessin assisté par ordinateur et de proposer des systèmes complets plus ou moins fermés, et ESRI (environmental systems research institute) qui offre un produit logiciel - ARC INFO - fondé sur les bases de données relationnelles.

Au départ les calculateurs exigés étaient de taille moyenne à grosse (32 bits). On constate de plus en plus une évolution vers des stations de travail sur micro ainsi que sur des systèmes UNIX.

Au niveau technique, le rapport repère encore trois goulets d'étranglement :

- problème posé par les formules de passage entre images en mode raster et données en mode vecteur ;
- le coût élevé de l'acquisition des données, poste le plus onéreux dans les SIG ;
- problème de standardisation des formats d'échange de données numériques et coordination de l'organisation des bases d'information.

B. Quelles leçons pour l'ORSTOM ?

Revenant à l'ORSTOM, où la solution SIG n'est pas inconnue puisqu'elle concerne peu ou prou quatre des huit exposés de cette troisième journée, que pouvons-nous retenir de ces réflexions en dehors de la satisfaction d'être dans le coup et reconnu comme tel ?

Les systèmes du type SIG, pour séduisant qu'ils soient ne peuvent constituer une "panacée", une réponse miracle à tous les problèmes de traitement de données de l'environnement. Il faut en effet user de prudence et bien insister sur la remarque suivante tirée de ce même rapport au sujet des expériences nord-américaines sur le thème environnement : "Les SIG, ainsi mis en place sont relatifs à des problèmes bien cernés et limités dans l'espace. On a affaire ici à des SIG de taille assez modeste".

Par ailleurs la FAO justifie l'emploi des SIG :

1°) lorsqu'il est nécessaire d'assurer un suivi précis et régulier de phénomènes évoluant dans le temps.

2°) lorsque le traitement géographique a pour but de fournir un facteur de décision obtenu par croisement de données spatiales.

Nos objectifs d'inventaire, nos recherches, le plus souvent, ne justifient pas l'emploi de tels outils encore très lourds. Par ailleurs, les moyens à notre disposition, les contextes dans lesquels nous travaillons ne nous permettent pas toujours la mise en oeuvre de tels systèmes.

Par contre, comme nous allons pouvoir le vérifier à travers certaines communications, de telles solutions semblent viables localement, pour certains programmes de recherche avec certains partenaires mais posent aussi de nombreux problèmes. Encore faudrait-il trouver de meilleure cohérence entre ces diverses tentatives similaires.

D'autres perspectives s'ouvrent aussi : les SIG pourraient être un outil d'aide à la synthèse évolutive. En effet, comme la FAO l'a déjà entrepris à une certaine échelle et pour certains thèmes sur le continent africain, l'ORSTOM par le biais des SIG pourrait entreprendre la valorisation de certains de ses inventaires dans une optique de synthèse sur l'Afrique francophone par exemple.

L'effort de standardisation dont certains exposés soulignent d'ailleurs la nécessité est probablement un des impératifs sur le principe duquel tous pourraient s'accorder. Mais qu'en est-il effectivement à l'ORSTOM ?

Sur les exposés eux-mêmes, il me semble que l'on peut y voir trois types de préoccupations dont certaines relèvent plus des thèmes abordés dans les deux journées précédentes :

1er groupe : Avec l'exposé sur le logiciel Neptune autant qu'avec celui sur "Proposition de schéma de collecte de l'information pour le milieu physique", nous sommes dans le monde des problèmes de saisie plus ou moins directe de l'information, de la conception de langage et de terminologie descriptive..., donc à l'amont des phases de traitement et de restitution.

2° groupe : Les communications sur le logiciel TRISKEL, sur l'Atlas informatisé de Quito et le SIG Veracruz ont en commun de traiter de systèmes logiciels intégrés à vocation multithématique et multimédia et à finalité cartographique.

3° groupe : Avec SISGEO et SISIMAGEM que l'on ne peut dissocier, il s'agit d'une importante collection de bases de données thématiques "**autonomes mais intégrées** dans une approche commune" selon les propres termes de P. SECHET.

Enfin, inclassable, quoique posant une question concrète et pertinente l'exposé sur LEICA aborde le problème de la mise à disposition d'une base de fonds de cartes numérisés.

Comme on le constatera donc, ces diverses initiatives paraissent le plus souvent indépendantes les unes des autres d'une part, et d'autre part faire fi des logiciels disponibles sur le marché.

Or, pour revenir à notre perspective stratégique, la question fondamentale de cette journée me paraît être la suivante :

Si l'on est d'accord avec la conclusion de D. DAGORNE selon lequel "*La clé du succès est le logiciel et la clé de la réussite du logiciel est la standardisation*", comment assurer cette standardisation sans pour autant bloquer ce foison-

nement d'initiatives individuelles ou de petits groupes ? Nous ne pourrions sans doute pas répondre à cette question fondamentale dans cette seule journée, mais souhaitons-le, puisse-t-elle contribuer à clarifier le débat à partir d'informations réciproques !

REFERENCES BIBLIOGRAPHIQUES

- CHEYLAN J.P. 1989 - *Projet pour la mise en place d'un réseau de Recherche sur la représentation et le traitement de l'Information Géographique*. R³ IG. Multigraph. MGM - Montpellier.
- DANDOY G. 1986 (Textes réunis par) - *Traitement des Données Localisées - L'Infographie à l'ORSTOM*. Colloques et Séminaires - Ed. ORSTOM.
- DANGERMOND J. 1985 - *Les Systèmes d'Information Géographique*. Bulletin CFC n°96.
- MANGIN A. 1988 - *Enquête sur les Systèmes d'Information Géographique*. Service de Consultance en Observation de la Terre (SCOT) - CNES - TOULOUSE.

SISIMAGEM : UNE PHILOSOPHIE D'UTILISATION DE LA TELEDETECTION A L'EMBRAPA.

(Eduardo D. ASSAD, Jamil MACEDO, Deusdeth T. dos SANTOS & Severiano R. JUNIOR)

RESUME - SISIMAGEM, système de traitement numérique d'images de télédétection, est l'un des sous-systèmes de SIGGEO, système d'information géographique à vocation agricole de l'EMBRAPA. Son objectif est d'offrir à ses utilisateurs des fonctionnalités qui les aident à exploiter les images de satellites (lesquelles fournissent un ensemble de données utilisées dans les études de ressources naturelles), depuis leur acquisition jusqu'à la sortie et l'interprétation des résultats du traitement. Il s'agit d'un ensemble de moyens informatiques mis au point pour rendre opérationnelles et systématiques toutes les étapes de l'utilisation du traitement d'images dans la recherche agronomique, tant pour l'identification de l'utilisation actuelle des terres ou des surfaces productrices qu'en vue de l'aménagement de nouvelles régions.

La première version du premier module, actuellement en cours de développement, se base sur l'équipement central de traitement graphique digital de l'EMBRAPA, un RAMTEK-RM 9300. Il est destiné au contrôle de la disponibilité des images de satellites, dans le but de rationaliser l'obtention des images qui sont utilisées par les diverses unités de l'EMBRAPA. Les fonctionnalités prévues permettront :

- de maintenir une administration adéquate du matériel de télédétection disponible à l'EMBRAPA et dans les autres institutions publiques, en plus des informations sur le stock d'images disponibles à l'Institut de recherches spatiales (INPE). Cette base sera utilisée par un système expert qui orientera les chercheurs ayant peu de connaissances dans le domaine à l'utilisation des fonctions de télédétection ;

- d'utiliser le matériel RAMTEK pour effectuer des pré-traitements de base (lecture et visualisation monospectrale - bande à bande - et multispectrale - plusieurs bandes, au maximum trois) ;

- d'utiliser cet équipement pour des traitements *simples* (histogramme mono et bidimensionnel, calcul de surface, lecture sélective de *pixel*, "navigation" dans l'image - localisation d'un point à partir de sa latitude et de sa longitude, et vice-versa - opération entre canaux, extraction de zones de travail) ;

- d'utiliser imprimante, traceur de courbes et appareil photographique appropriés, pour transformer des images digitales en images imprimées ou photographiées.

Avec ces fonctions disponibles, il sera possible de viabiliser diverses améliorations de ces images, en particulier de mieux les contraster, d'en faire une première interprétation en plus de générer des sous-scènes qui faciliteront l'utilisation de la télédétection dans les diverses unités décentralisées de l'EMBRAPA.

Pour atteindre tous ces objectifs, on envisage d'exécuter les activités suivantes :

- installation du logiciel HLIPS (ancien Hacienda) et du DMAPS sur l'ordinateur central IBM 4341 de l'EMBRAPA et développement d'une interface permettant l'intégration des logiciels en référence ;

- test et évaluation de ces deux nouveaux logiciels de traitement d'images, à l'occasion de projets de recherche actuellement en cours au centre des cerrados (CPAC/EMBRAPA) ;

- si nécessaire, compléter ces nouveaux logiciels par la mise au point de fonctionnalités supplémentaires, telles que la génération de scènes sur disquettes, ce qui permettra la visualisation des images sur micro-ordinateurs compatibles IBM-PC et donc diffusera plus rapidement l'utilisation de la télédétection dans le système coopératif brésilien de recherche agronomique (SCPA), par le biais du module local de SISIMAGEM ;

- implémentation du catalogue d'images, avec une interface "intelligente".

Le logiciel est conçu de façon à pouvoir être utilisé par des personnes n'ayant que peu de connaissances en informatique. De plus, les données stockées avec ce module pourront être utilisées directement dans les versions postérieures. Enfin, toutes les ressources qui doivent être implémentées dans SISIMAGEM le seront sous l'orientation d'un responsable thématique, ce qui est une forme de familiariser les chercheurs avec les outils à développer, dès le début de leur conception.

INTRODUCTION

Depuis le début des années 1980, l'EMBRAPA tente de doter les unités de recherche du Système coopératif de recherche agronomique (SCPA) d'outils informatiques capables d'aider ces équipes (principalement des centres de ressources) dans la réalisation de leurs travaux de recherche, c'est à dire depuis les inventaires de ressources naturelles jusqu'à l'utilisation de ces informations dans les opérations de zonage, par exemple. C'est avec cet objectif que le projet SISGEO, Système d'informations sur l'environnement agricole a été créé.

Pour mener à bien le développement de ce projet au fil du temps, certaines directrices fondamentales ont été établies, lesquelles doivent permettre :

- la participation effective des chercheurs spécialisés lors de l'élaboration des modules correspondant à chaque thème qui intervient dans le système (actuellement : sol, climat, végétation native, données phytosanitaires, socio-économie, hydrologie, cartographie et télédétection) ;

- l'implication croissante des centres de ressources dans le développement du projet, comme forme de garantir un appui institutionnel toujours plus important ;
- l'utilisation immédiate des outils partiels (intermédiaires) mis au point, que sont les modules du système SISGEO ;
- de garantir l'intégration des outils intermédiaires.

1. LE SYSTEME SISGEO

Dans un premier temps, un système d'informations pédologiques (SISSOLOS) a d'abord été développé sur ordinateur central, sous l'orientation du *Service National de Levé et Conservation des Sols (SNLCS/EMBRAPA)*. Avec ce premier résultat, il a été possible de stocker sur support magnétique près de 12 000 profils de sols, et pratiquement tous les inventaires réalisés jusqu'à maintenant par le SNLCS au Brésil.

Postérieurement, compte tenu de l'avancée de la micro-informatique d'une part, des difficultés d'accès que l'utilisation d'un équipement central présente d'autre part (comme il a pu être constaté avec SISSOLOS), on s'est plutôt appliqué à mettre au point des outils orientés vers la réalisation de bases de données thématiques locales. Avec ce type d'outils les travaux d'inventaire peuvent être grandement facilités, et de plus le travail ultérieur de transcription peut être minimisé, voire totalement supprimé, et avec lui les erreurs qu'un travail aussi fastidieux entraîne nécessairement.

De plus, les centres de ressources, mais aussi les autres unités de recherche de l'EMBRAPA, peuvent tirer profit de cette automatisation pour aider à la réalisation des travaux de recherche les plus divers. En effet, la micro-informatique offre désormais une série de logiciels de traitement, qui, combinés avec les systèmes d'informations ainsi mis au point, permettent une énorme gamme de traitement de données susceptibles de fournir de précieuses informations.

Avec cette nouvelle orientation du projet, des logiciels permettant la création, sur des équipements de type PC-compatible, de bases locales de données de sols, de climat, de végétation native, de données phytosanitaires, d'hydrologie, de socio-économie et même de cartographie sont progressivement mis à la disposition de toute la communauté scientifique du SCPA.

La figure 1 ci-après présente un diagramme schématique du projet SISGEO.

2. LE SYSTEME SISIMAGEM

Conformément à cette philosophie de travail, la télédétection a besoin également de disposer d'un outil capable de permettre la création de bases locales d'images tout en autorisant un minimum de traitements primaires, en plus d'une base centrale partagée, disposant de ressources permettant des traitements plus avancés et plus complets.

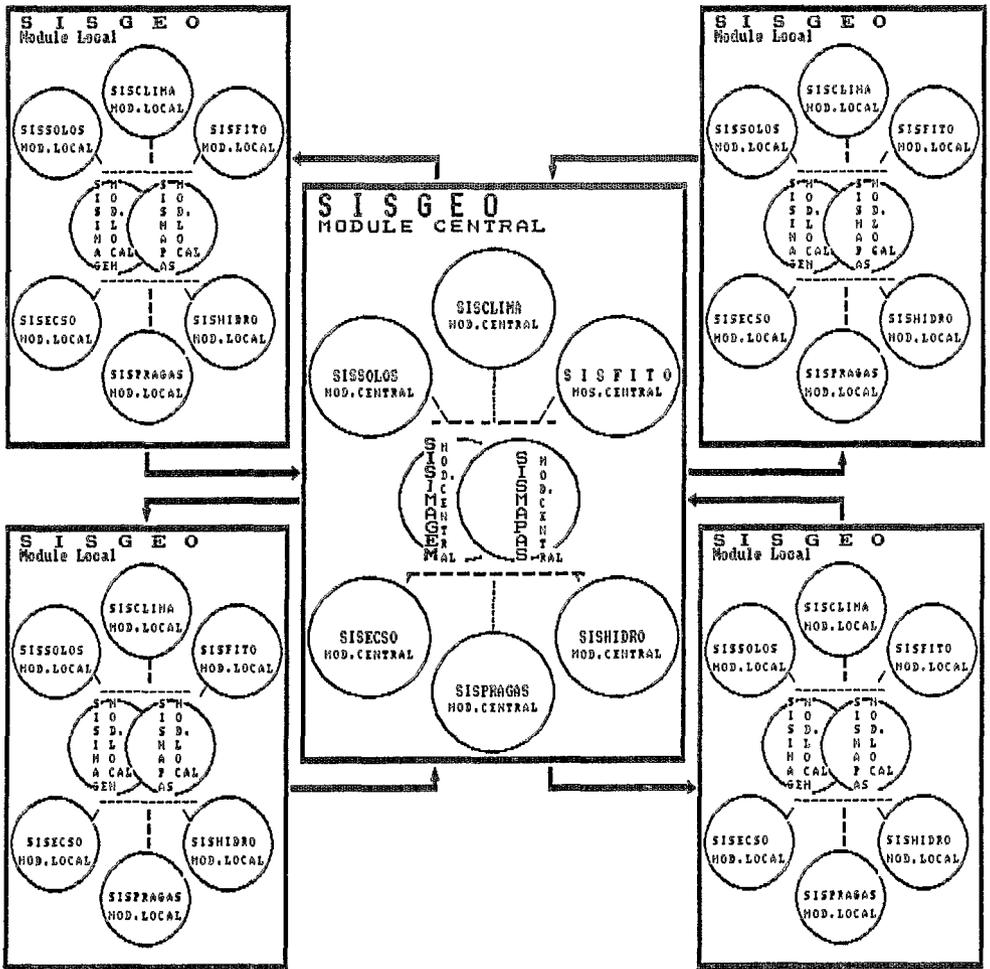


Fig. 1 Structure du système SIGEO

Ainsi, *SISIMAGEM*, *Système de traitement digital d'images de télédétection*, prétend fournir à l'utilisateur un instrument capable de l'aider dans l'exploitation des images de satellites, lesquelles contiennent un ensemble d'informations utiles tant pour l'étude des ressources naturelles que pour l'analyse de l'environnement. Cet instrument est prévu pour l'aider dans toutes les étapes de ce travail, depuis l'acquisition jusqu'à la production et l'interprétation des résultats du traitement.

Il s'agit donc d'un ensemble de routines informatiques mises au point pour rendre opérationnelles et pour systématiser toutes les étapes de sélection des images et de l'utilisation postérieure du traitement d'images dans la recherche agronomique, aussi bien pour l'identification de l'utilisation actuelle des terres et des surfaces productrices que pour la planification de l'occupation de nouvelles surfaces.

On précisera, lors de cet exposé, toutes les directrices qui orientent la mise en oeuvre de *SISIMAGEM* et qui définissent une stratégie de développement pour le projet.

En premier lieu, il est nécessaire de souligner que la mise au point de *SISIMAGEM* ne suppose aucun développement de logiciel de traitement d'images, si ce n'est des ajustements, corrections ou interfaces permettant d'adapter les progiciels existants, et qui ont été choisis pour faire partie de l'ensemble des outils intégrés qui constituera le système appelé *SISIMAGEM*.

La figure 2 ci-après présente un diagramme schématique du projet *SISIMAGEM*.

Les personnes directement impliquées dans la viabilisation de *SISIMAGEM* sont :

- des informaticiens du DIN, *Département d'Informatique de l'EMBRAPA*, ayant une expérience de projet et implantation de logiciels sur ordinateurs centraux et sur micro-ordinateurs, familiarisés avec l'utilisation d'outils modernes de CASE (Computer Aided Software Engineering), de progiciels de bases de données et de langages de programmation modernes. Ceux-ci peuvent de plus s'appuyer sur le groupe d'intelligence artificielle du DIN, qui est déjà l'auteur du progiciel *PSE, Prototypage de Systèmes Experts* et également sur celui du NTIA, *Noyau Technologique d'Informatique pour l'Agronomie*, actuellement couplé avec le CTI, *Centre Technologique d'Informatique*, organisme gouvernemental qui s'occupe de la recherche en informatique dans le but de conquérir l'indépendance technologique nationale dans ce domaine, et qui a déjà mis au point le *SOC, logiciel scientifique* et le *SRI, Logiciel de Récupération d'Information* ;

- des spécialistes en agrométéorologie, pédologie et végétation de l'EMBRAPA, avec une grande expérience de l'application des techniques de télédétection, notamment en utilisant le traitement digital des images de satellites. Ceux-ci comptent également sur l'appui du Département de Géosciences de l'UnB, *Université de Brasília* et du NMA, *Noyau de veille par satellite sur l'environnement*, de l'EMBRAPA.

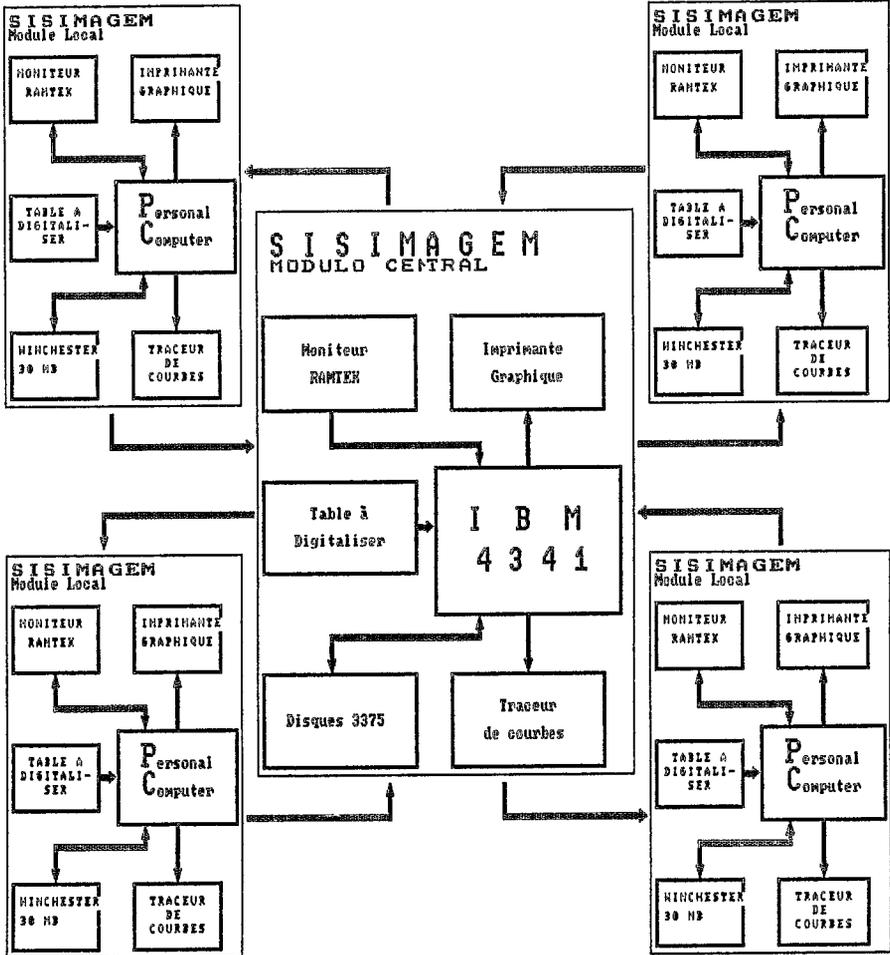


Fig. 2 Structure du système SISIMAGEM

Les moyens matériels actuellement disponibles sont :

- ordinateur central IBM-4381, avec un moniteur couleur RAMTEK, RM 9300, équipement graphique digital d'une résolution de 512 lignes x 640 éléments ;
- deux tables à digitaliser, de format A0 ;
- un traceur de courbe de grande dimension ;
- un traceur de courbe de petite dimension ;
- deux micro-ordinateurs PC-XT de 640 Kbytes de mémoire principal et disque Winchester de 30 Mbytes, dont un possédant une plaque spéciale SITIM ;
- deux imprimantes graphiques.

Les moyens en logiciel actuellement disponibles sont :

- HLIPS, DMAPS, SITIM et TIMOR, comme logiciels de traitement d'images ;
- SQL, SRI, DIALOG, SGI et TurboAccess, comme logiciels gestionnaires de bases de données ;
- PSE et ESE, comme logiciels d'intelligence artificielle ;
- SOC et SAS, comme logiciels de traitement statistique ;
- Cobol, Fortran, TurboPascal, Turbo C, Turbo Prolog et Dialog IV Plus, comme langages de programmation ;
- CSP, MOSAICO-DFD, MOSAICO-DEM et DIALOG/GER, comme outils de génie logiciel.

Etant donné, d'une part, la disponibilité actuelle de progiciels déjà satisfaisants utilisables sur micro-ordinateur, type SITIM ou SGI et, d'autre part, le fait que le RAMTEK, équipement graphique digital de haute résolution, ait été acquis il y a près de dix ans et qu'à ce jour ses potentialités n'ont pratiquement jamais été exploitées au bénéfice de la recherche agronomique, priorité a été donnée à la mise au point du module central de SISIMAGEM. Celui-ci est orienté vers le traitement et le contrôle de disponibilité des images de satellites, avec comme objectif de rationaliser l'obtention de ces images en vue de leur utilisation par les diverses unités de l'EMBRAPA.

Pour le développement de cette première version, on prévoit de mettre en oeuvre l'intégration d'un catalogue de données de télédétection disponibles, avec un prototype de système expert, une base de données d'images et les logiciels HLIPS et DMAPS, de manière à viabiliser l'offre de facilités permettant :

- de gérer des informations sur les données de télédétection disponibles à l'EMBRAPA et dans les autres organismes gouvernementaux, y compris sur le stock maintenu par l'INPE, *Institut de Recherches Spatiales* ;
- de fournir une orientation sur l'utilisation de la télédétection, à l'intention des chercheurs n'ayant que peu d'expérience dans ce domaine, par le biais de la mise en oeuvre de ressources d'intelligence artificielle ;
- d'offrir quelques prétraitements de base, comme la lecture et la visualisation monospectrales (bande à bande), et multispectrales (plusieurs bandes, au maximum 3) ;

- d'offrir des traitements "simples" (histogramme mono et bidimensionnel, calcul d'aires, lecture sélective de "pixel", navigation dans l'image - localisation de point à partir du couple latitude/longitude et vice-versa, opérations entre canaux, extraction de zones de travail) ;

- d'utiliser l'imprimante, le traceur de courbes ou des moyens photographiques appropriés, pour transformer les images digitales en images imprimées ou en photographies.

Avec toutes ces facilités, les chercheurs plus expérimentés pourront réaliser certaines améliorations des images, interprétations élémentaires et extractions de sous-scènes, lesquelles, mises à disposition dans la base de données d'images, faciliteront grandement l'utilisation de la télédétection dans les divers centres de l'EMBRAPA.

Le module local qui sera développé aura une structure semblable à celle du module central (intégration d'un catalogue local avec un prototype de système expert, une base de données locale d'images, SITIM, TIMOR et/ou d'autres logiciels de traitement d'images).

Ce module local, en plus de sa fonction de capteur d'informations pour alimenter le module central, disposera également de moyens de traitement et visualisation relativement puissants.

De tels moyens permettront au chercheur d'utiliser les techniques de télédétection sur son micro-ordinateur individuel, même s'il ne dispose pas de cartes spéciales comme celles du SITIM, puisque certains logiciels comme TIMOR offrent déjà des facilités intéressantes de traitement et restitution d'images, malgré une visualisation de faible qualité.

L'interconnexion du module central avec le module local sera faite dans les deux sens, c'est à dire qu'il existera la possibilité d'alimenter le module central par le module local, mais aussi que les données et ressources de la base centrale seront partagées et accessibles par le biais du module local.

De cette façon, les échanges d'information, l'acquisition de nouvelles images, les traitements et les améliorations de celles-ci seront rationalisés et, par voie de conséquence, il en sera de même de la diffusion de ces données parmi la communauté scientifique brésilienne. On prétend aussi de cette façon atteindre la communauté scientifique internationale, dans la mesure où les réseaux de communication de données se développent de plus en plus.

Une des préoccupations majeures de l'équipe de mise au point du SISIMAGEM concerne la facilité d'utilisation de ses ressources par des personnes peu familiarisées avec l'ordinateur. Dans cette perspective, l'équipe thématique cherchera à transférer la technologie, c'est à dire qu'elle orientera l'équipe informatique dans le but d'automatiser le maximum de connaissances scientifiques et techniques de la télédétection, en même temps que le savoir-faire pratique indispensable, hérité du grand nombre d'expériences d'utilisation de cette technique, vécues par les spécialistes de l'équipe thématique. La figure 3 présente le diagramme fonctionnel de SISIMAGEM, et les figures 4 à 8 montrent le diagramme de chacune de ses applications.

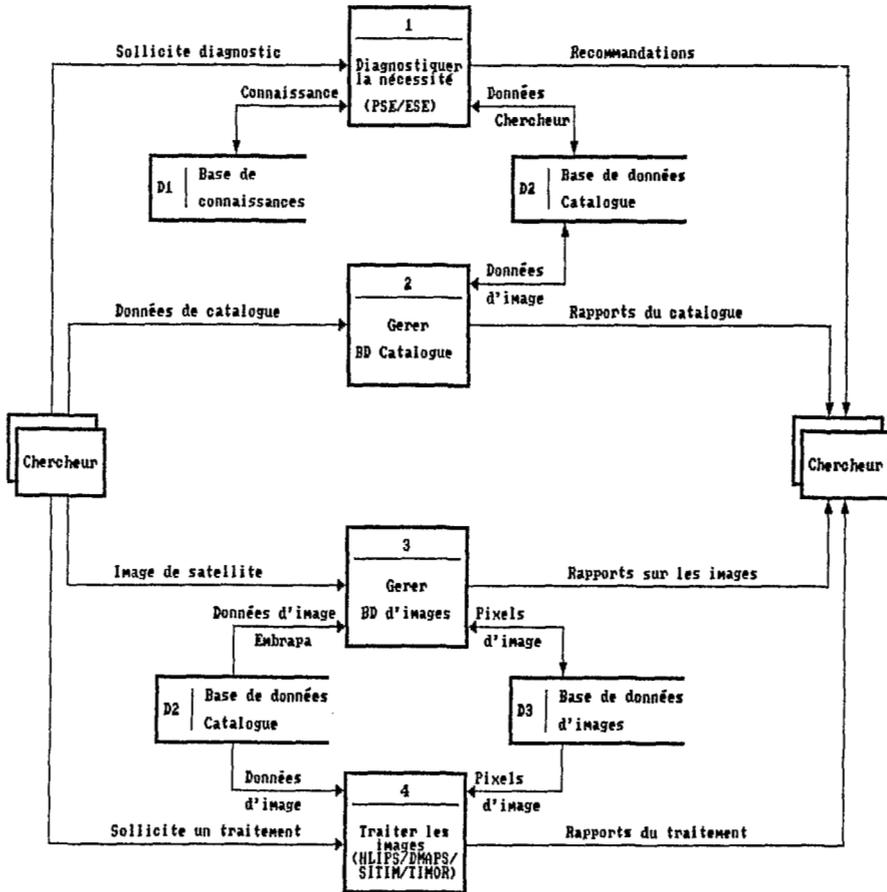


Fig. 3 Diagramme de flot de données de niveau 0

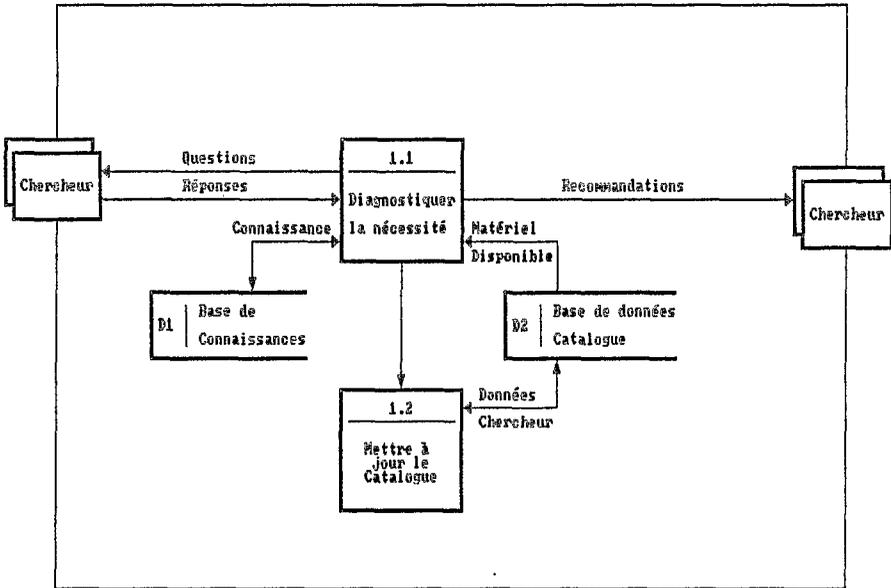


Fig.4 DFD fonction 1: Diagnostiquer la nécessité

SISIMAGEM

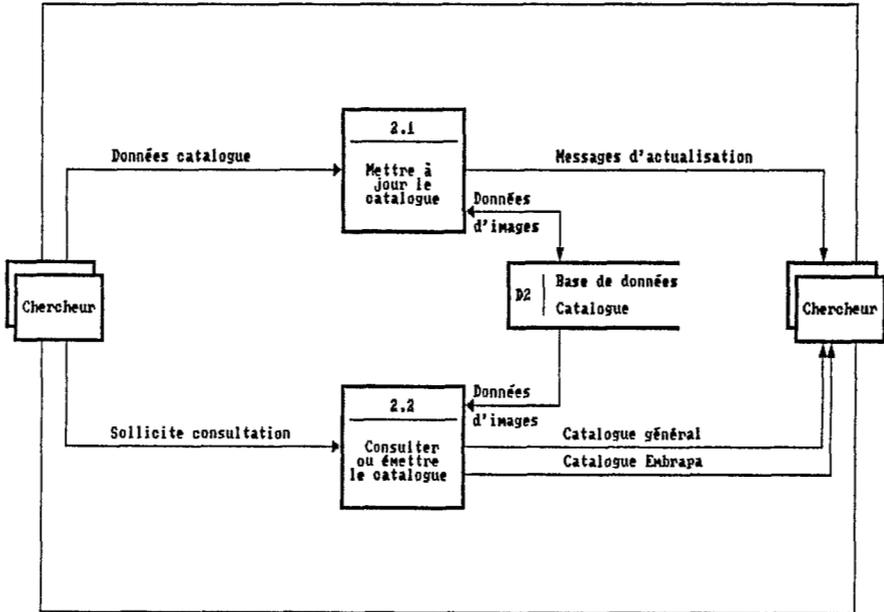


Fig.5 DFD de la fonction 2: Gérer BD catalogue

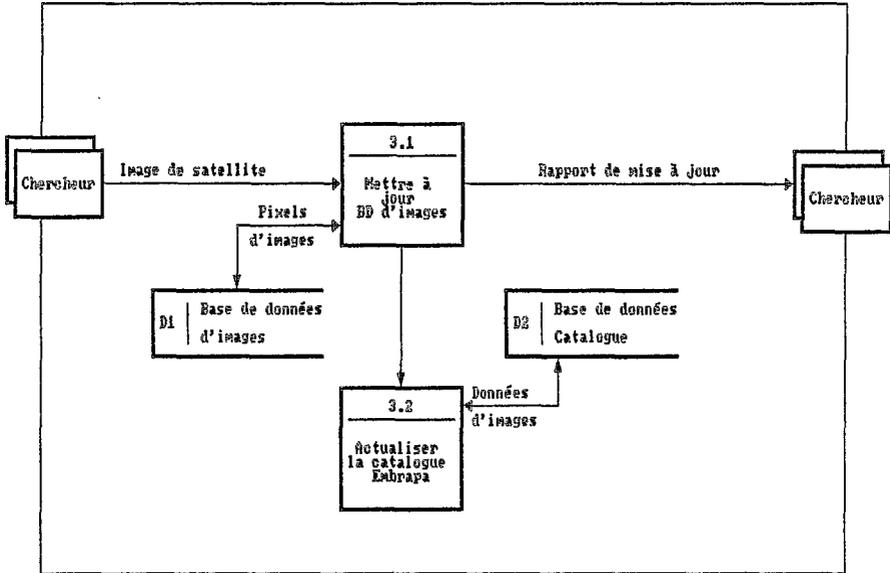


Fig.6 DFD de la fonction 3: Gérer BD d'images

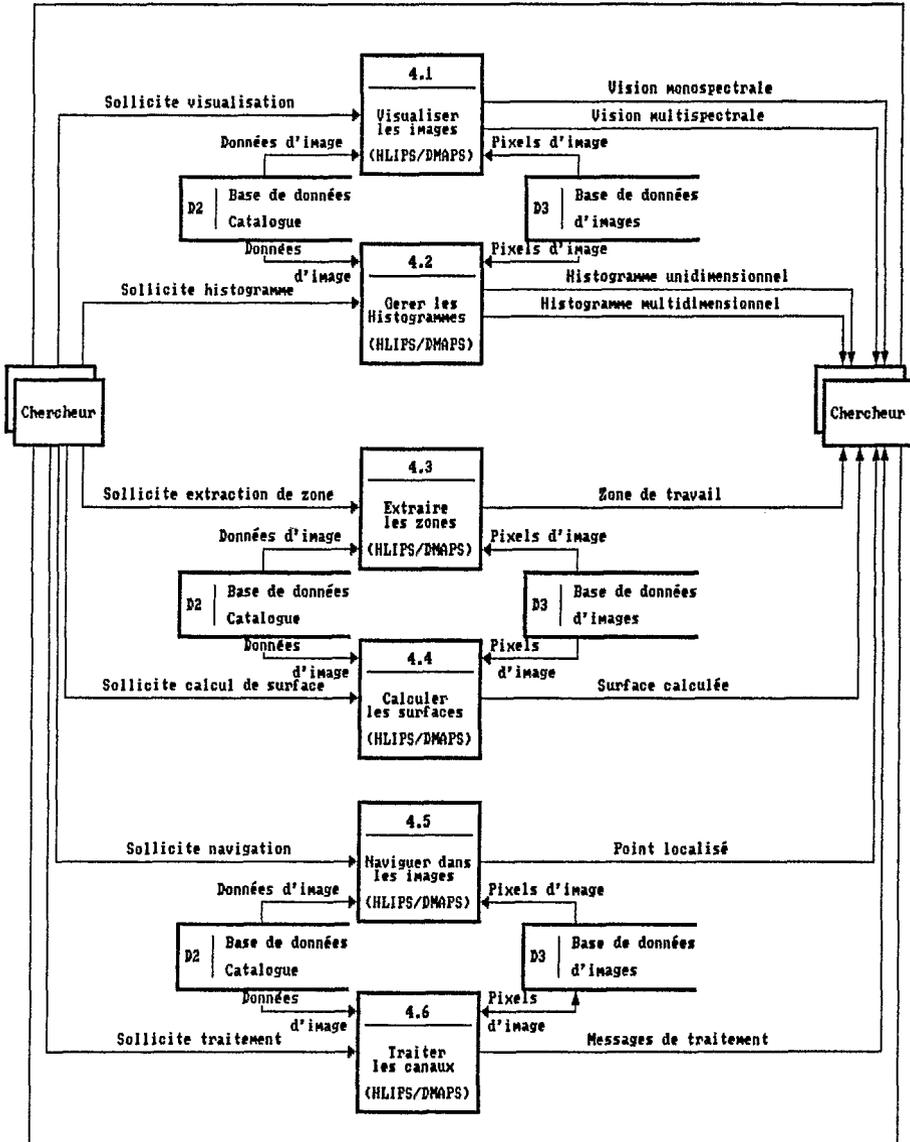


Fig.7 DFD de la fonction 4: Traiter les images

SISIMAGEM

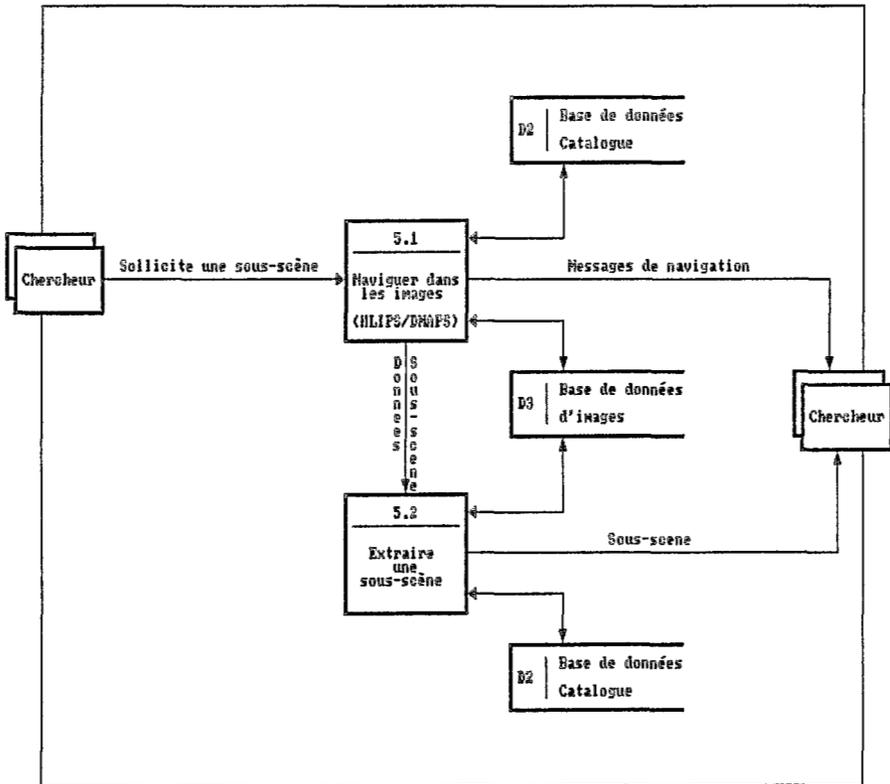


Fig.8 DFD fonction 5: Extraire les sous-scènes

CONCLUSION

Le système présenté a pour but d'offrir à tous les centres du SCPA un accès aux informations et à des orientations sur le traitement numérique d'images : c'est un des objectifs majeurs poursuivis par SISIMAGEM. Les progiciels actuellement disponibles pour le traitement d'images seront utilisés, avec la finalité de mettre en place des facilités qui permettent :

- de gérer les informations sur les images disponibles à l'EMBRAPA, et dans d'autres organismes ;
- d'orienter les chercheurs à propos de l'utilisation des images de satellites, dans des buts bien spécifiques ;
- que ces mêmes chercheurs puissent opérer des traitements d'images simples, tant au niveau du module central (en utilisant le RAMTEK accouplé à l'ordinateur IBM-4381), comme au niveau du module local, opérationnel sur micro-ordinateur PC-compatible disposant d'une carte graphique).

Comme module partiel de SISGEO, SISIMAGEM pourra atteindre près de soixante institutions de recherche agronomique dans le pays, et pourra ainsi rapidement conduire à la formation de chercheurs capables d'utiliser la télédétection comme un outil d'aide aux études sur l'environnement, à l'inventaire et à l'exploitation rationnelle des ressources naturelles brésiliennes.

REFERENCES BIBLIOGRAPHIQUES

- BORLAND International Inc., *Data Base Toolbox*, 1987, 163p.
 BORLAND International Inc., *Turbo Graphix Toolbox*, 1985, 254p.
 BORLAND International Inc. *Turbo Pascal 4.0*, 1987, 629p.
 EMBRAPA/DIN. *Manual do sistema de prototipação de sistemas especialistas (PSE)*, Brasília, 1987, 123p.
 EMBRAPA/NTIA. *Manual do software de recuperação de informações (SRI)*, Campinas, 1989.
 EMBRAPA/NTIA. *Manual do software científico (SOC)*, Campinas, 1987.
 ENGESPAÇO. *Manual do sistema de tratamento de imagens (SITIM)*, São José dos Campos, 1986.
 ENGESPAÇO. *Manual do sistema geográfico de informações (SGI)*, São José dos Campos, 1987.
 GANE, C. & SARSON, T. *Structured systems analysis : tools and techniques*. Prentice-Hall Inc., Englewood Cliffs, New Jersey, 1979. 256p.
 IBM. *Manual do Cross system product (CSP) : application development user's guide*, 1988, 179p.
 IBM. *SQL Data system : terminal user's reference for VM system product*, 1984, 214p.

- IESA/Tecnologia de sistemas. *Mosaico/DFD*, 1988, 148p.
IESA/Tecnologia de sistemas. *Mosaico/DEM*, 1987, 160p.
MAGRI, J.A. *DIALOG Plus Interativo*, 1988, Ed. Atlas.
PARROT, J.F. & CHEVILLOTTE, H. *TIMOR, Traitement d'image par micro-ordinateur*, Paris, 1987. ORSTOM, 56p.
SAS Institute Inc. *SAS User's guide : basics*, 1985, 1290p.
SOFT. *Manual preliminar do DIALOG/GER*, 1989, 203p.

(Traduction libre, Patrick Séchet, août 1989).

TRAITEMENT DE DONNEES METEOROLOGIQUES SATELLITAIRES ET CONVENTIONNELLES POUR LA CONNAISSANCE DE L'ENVIRONNEMENT.

(Dominique DAGORNE)

RESUME - L'utilisation de données satellitaires pour des programmes de recherche en environnement est maintenant une approche courante. Cependant la mise en oeuvre des techniques de télédétection suppose la disponibilité d'un environnement logiciel (et matériel), sans que l'utilisateur ne se transforme en *informaticien*, à chacun son métier !

Pour cela, nous avons développé une chaîne de traitement de télédétection *météorologique*, à partir des données des satellites existants, données utilisées conjointement avec d'autres plus conventionnelles.

Les spécifications de développement du système logiciel ont été :

- la mise à disposition d'un outil général de traitement de télédétection pour des programmes de recherche mis en oeuvre par de non spécialistes ;
- son application à partir des données conventionnelles et satellitaires météorologiques avec les modules permettant la mise au point de diverses corrections ;
- des fonctionnalités informatiques basées sur une approche *qualité logiciel* (utilisation, maintenabilité et portabilité).

Il est important de distinguer la donnée proprement dite, sous forme image ou non, d'origine satellitaire ou non, des représentations possibles. Ces dernières entités peuvent être présentées sous forme d'image (*raster*) ou de graphique (*vecteur*). Il en ressort que le *traitement d'images* n'est qu'une des composantes du logiciel, parmi plusieurs autres possibles dont les fonctionnalités graphiques (en mode GKS) et la manipulation de données spatialisées définies en latitude et longitude. Ces dernières données peuvent se présenter sous forme de *champ ordonné* (climatologie, résultat de modèle, etc.), de points discrets (relevé(s) aux stations) ou de segments (cartographie).

Le logiciel prend en compte la plupart des sources possibles de données satellitaires météorologiques défilants ou géostationnaires, avec les éléments indispensables permettant les corrections radiométriques (calibrages, effets des angles de visée) et géométriques (localisation).

La disponibilité d'outils de *transparence géométrique* permet la superposition et la restitution cartographique de données multisources satellitaires ou non, à diverses résolutions spatiales ou temporelles, et ceci sous forme quelconque.

Les divers modules de traitements, les structures de données et les interfaces de représentation (image ou graphique) sont utilisables sous n'importe quel environnement informatique (aux périphériques près de visualisation et d'impression).

Les outils décrits trouvent un emploi dans un logiciel général d'application en télédétection, dans divers développements avec l'utilisation de bibliothèques de gestion de données et de représentations, ou dans des programmes dédiés à des applications routinières comme c'est le cas dans les recherches menées à Lannion en océanographie et en climatologie (estimation des pluies par satellites).

1. INTRODUCTION

Les communautés scientifique et opérationnelle des sciences de l'environnement (météorologistes, climatologues, océanographes, hydrologues, etc.), doivent faire face à de nouvelles sources d'information, ainsi qu'à de nouvelles technologies, tant pour l'acquisition que pour la manipulation ou la restitution des données. Ceci réclame de plus en plus de compétences techniques, indispensables pour la mise en oeuvre de moyens et de méthodes de traitements grâce à des environnements logiciels et matériels, sans pour cela que le thématicien du domaine concerné ne se transforme en spécialiste du traitement de l'information. A chacun son métier !

Les outils logiciels que nous présentons ont été conçus dans le cadre du support technique à une équipe de recherche pluridisciplinaire utilisant des données de télédétection satellitaires d'environnement (ou météorologiques, par opposition aux données d'observation de la terre), en conjonction avec des données plus classiques, c'est à dire des relevés conventionnels aux stations, des résultats de climatologie, des données de cartographie, etc..

La méthodologie globale mise au point est utilisée en développement, et de façon opérationnelle, dans plusieurs domaines applicatifs. Ce sont des méthodes générales, qui présentent l'intérêt d'avoir une expression informatique sous la forme d'un logiciel répondant à des spécifications préalables (utilisation, définition des données, réalisation). Ce logiciel est réutilisable indépendamment du domaine d'application, vu la généralité des concepts utilisés, et, à un degré moindre, il est également indépendant de l'environnement matériel disponible.

2. SPECIFICATIONS DE REALISATION

Les tendances actuelles des sciences de l'environnement peuvent se résumer en :

- une augmentation des volumes de données à traiter par suite de la mise en service de nouveaux capteurs de télédétection (actifs, passifs) à partir de vecteurs mobiles (satellites défilants et géostationnaires, moyens aéroportés) ou fixes (radar), avec des expressions de données de différentes formes (image ou non), et ceci à des résolutions spatiales, temporelles et spectrales très variées ;
- la disponibilité de bases de données centralisées, accessibles par réseau, et disposant de toutes les informations conventionnelles, véhiculées dans le domaine météorologique par le Système Mondial de Transmission ;

- des moyens de traitements décentralisés toujours plus performants, avec une quasi standardisation autour d'équipements personnels, ou de stations de travail scientifiques ;

- la disponibilité de logiciels *clés en main*, particulièrement nombreux dans divers domaines généraux (statistique, base de données, graphiques, etc.), ce qui évite des développements coûteux de logiciels par leur emploi banalisé.

Compte tenu de ces tendances, nous avons entrepris le développement d'un logiciel (programmes et structures de données) avec la prise en compte de plusieurs objectifs, plus spécialement liés à des aspects spécifiques de notre activité à savoir :

- la mise à disposition pour des utilisateurs scientifiques non informaticiens, de logiciels assurant le traitement de données, avec ses aspects décodage, stockage, manipulation, croisement et restitution. Les programmes développés doivent être facilement utilisables et adaptables aux demandes des thèmes de recherches. Il est alors indispensable de disposer d'une chaîne ouverte et évolutive, utilisable en *boîte à outils* soit l'inverse de certains produits commerciaux de type *boîte noire* ;

- des données plus spécialement liées aux produits de télédétection *météorologique*, données qui nécessitent l'application de divers algorithmes pour l'élaboration de résultats en valeurs géophysiques, ainsi que pour la restitution cartographique de ceux-ci, ceci avec l'addition d'autres sources (conventionnelles, cartographiques, etc.) ;

- une réalisation informatique intégrant certains concepts de *génie logiciel* à l'analyse et à l'écriture avec les aspects d'utilisation du produit, de maintenance (correction, extension, etc.), et celui de la réutilisation sur d'autres sites (portabilité) et avec d'autres logiciels. Ces facteurs conduisent à une approche structurée, modulaire et hiérarchique des programmes, la définition précise des *structures de données*, et l'utilisation de normes de fait en informatique scientifique.

Un système générique de traitement des données d'environnement peut être schématisé sur la figure 1.

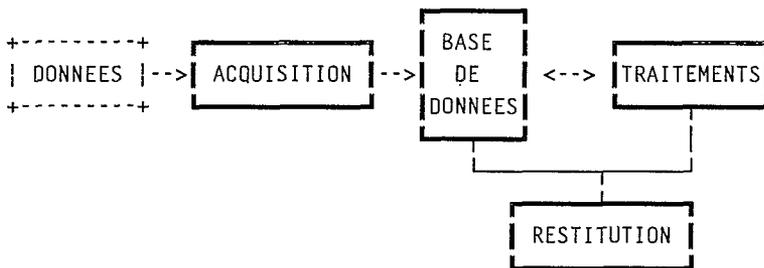


Figure 1 : Système générique pour l'environnement

Il y a quatre fonctions de bases à satisfaire :

- ACQUISITION. Les données doivent être introduites dans le système et décodées ;

- BASE DE DONNEES. Les données doivent être organisées et stockées, prêtes à être restituées ;

- TRAITEMENT. Les données doivent être *traitées* (analysées, manipulées, croisées, synthétisées, etc.) ;

- RESTITUTION. Afin d'être exploitées, les données mises en forme et les résultats des traitements doivent être visualisés et imprimés, et ceci sous forme d'images et de graphiques.

S'il est envisageable de disposer d'une configuration matérielle pour chacune des fonctions précédentes, un simple micro-ordinateur PC-compatible peut en réaliser l'ensemble, à une petite échelle

Le succès des extensions futures, via le passage aux stations de travail et aux environnements spécialisés par l'interconnexion de plusieurs matériels en réseau, est alors lié à la standardisation du logiciel applicatif développé et à la *transparence* des données.

3. DONNEES ET REPRESENTATIONS

3.1. Nature des données

3.1.1. Image

La donnée élémentaire d'une image est la valeur du pixel, celui-ci étant défini par sa position (ligne - colonne), et éventuellement son canal si la mesure est effectuée à travers plusieurs fréquences spectrales.

Les premières données manipulées ont été des images *brutes* issues des capteurs satellitaires météorologiques (radiomètres à balayage), avec la valeur du pixel en Compte Numérique (CN), codée sur plusieurs bits (de 6 à 10). Cette valeur est issue du système de numérisation de la mesure dans le domaine spectral du capteur (essentiellement visible et infra-rouge dans notre cas).

Après *traitements radiométriques* (algorithmes géophysiques, corrections diverses), la valeur du pixel image peut avoir une signification quantitative bien différente de sa valeur initiale, d'où la nécessité de prévoir la mesure de celui-ci, soit directement dans l'image (en valeur physique), soit à travers une table de correspondance CN/valeur. Il est également possible d'avoir des données sous forme d'images *synthétiques*, issues de traitements plus ou moins sophistiqués de reconstitution de pixels, à partir de capteurs non imageurs ou de données externes (*sondeurs atmosphériques* pour les capteurs passifs, radars - imageur ou non - pour les capteurs actifs micro-ondes, grille de *valeurs*, etc.).

Les corrections géométriques (rectification d'images) produisent également des pixels artificiels issus des modèles de reconstitution de l'image *brute* (satellitaire ou non).

3.1.2. Paramètres géométriques de l'image

Toutes les données images sont géoréférencées, c'est à dire qu'à chacun des pixels, il est possible d'associer un point de la surface du globe défini en latitude et longitude. Ceci est réalisé en certains points d'une grille de l'image (matrice de localisation), par les modèles de navigation de l'imagerie satellitaire (modèle d'orbitographie et de capteur), ou des modèles de projection pour les images rectifiées géométriquement suivant une référence cartographique. Des méthodes d'interpolations permettent de connaître la localisation de tous les points d'une image, et inversement de positionner dans l'image tout repérage géographique.

De même, il est possible de calculer en ces points de localisation, diverses caractéristiques géométriques plus spécifiques des données satellitaires, comme les angles de visée ou d'éclairement, etc..

3.1.3. Champs spatialisés

Il est vite apparu que le type de donnée *image* utilisé seul était insuffisant, notamment pour des études où il est nécessaire de disposer de données plus conventionnelles. Ces dernières peuvent prendre deux formes suivant leur origine.

Dans le cas d'une mesure ponctuelle, à une ou plusieurs composantes (relevés d'une station par exemple), le champ produit est du type *discret*. Chaque point de mesure est alors parfaitement repéré au sein de la sphère terrestre par ses coordonnées géographiques, et éventuellement par son altitude.

Pour des données *synthétiques*, comportant éventuellement des valeurs multiples, élaborées par divers calculs : méthodes géostatistiques, résultat de modèle de prévision, analyse climatologique, modèle numérique de terrain (MNT), etc., les valeurs se présentent sous la forme d'un maillage à deux dimensions (ou intégrant l'altitude, 3D), régulièrement réparti et repéré spatialement suivant une représentation cartographique.

La similitude est grande entre les types de données *champ ordonné* et *image* de par leur expression sous forme matricielle. La différence tient à la taille respective des données, et à un mode différent de codage et d'accès lors de l'implantation informatique.

3.1.4. Segments cartographiques

Les images brutes ou redressées, ainsi que les divers champs, doivent pouvoir être superposés avec des données cartographiques classiques, telles que trait de côte, frontières politiques ou réseau hydrographique, par exemple. La donnée est alors une suite de segments. Chacun de ceux-ci, identifié par son type, est une succession de points jointifs repérés en latitude - longitude.

Divers résultats issus des traitements des données *champ ordonné* ont également une expression de cette forme (contourage d'isovaleurs, vecteur), ainsi que ceux qui sont issus de modèles géométriques effectués pour nos travaux (orbitographie, zone d'acquisition, etc.).

3.2. Modes de représentation

Nous ne nous intéresserons qu'aux représentations à deux dimensions des données précédentes, l'altitude (nuages, relief) n'étant pas un facteur prépondérant pour la restitution de nos travaux. Ces représentations sont effectuées suivant les techniques infographiques, avec deux types d'approches liés à la technologie disponible, avec une forte dépendance du matériel.

3.2.1. Mode Image ou raster

L'image est la représentation la plus élaborée d'un phénomène. La valeur du pixel de la donnée image, est affectée d'une représentation liée au périphérique de restitution, pixel défini également en ligne et colonne sur celui-ci.

En mode *fausses couleurs*, sur un écran vidéo, les composantes (additives) rouge, vert et bleu, affectées globalement à un canal, permettent d'attribuer une couleur (combinaison des trois couleurs de base), ou teinte de gris (composantes égales), à chaque compte numérique (ou plage de CN, si le nombre de représentations disponibles du périphérique est inférieur au nombre de CN de la donnée). Des modes plus élaborés de représentation d'images peuvent être mis en oeuvre sur écran monochrome (tramage), et pour des images multicanales avec l'affectation d'un canal par couleur de base (composition colorée, vraies couleurs).

Si les méthodes de traitements *colorimétriques* des images permettent d'avoir des représentations très réalistes et des mises en évidence de phénomènes, le tout accompagné d'une mise en page (légende, texte, etc.) afin d'en faciliter l'exploitation, il est important de se souvenir de la subjectivité de cette représentation, subjectivité liée au manque de sensibilité du récepteur oeil et à la fascination facile par de *belles images*, d'un public non averti.

Il en va de même pour les représentations sur papier en recopie de fichier (*softcopy*) de l'imagerie, où l'on affecte une combinaison de couleur de base (soustractive dans ce cas) à une plage de comptes numériques de l'image, sur un périphérique d'impression couleur, et ce suivant diverses technologies.

Un autre inconvénient de ce type de représentation est son absence actuelle de *normalisation* liée aux contraintes sévères de *temps réel* pour les opérations interactives de *traitement d'images* sur des configurations dédiées. La distinction introduite entre donnée et représentation est alors inexistante, les manipulations envisagées pouvant avoir lieu aussi bien sur la source (en mémoire image du calculateur) que sur son expression graphique (traitement vidéo). Ceci est le cas dans la plupart des domaines de l'imagerie qui ne s'intéressent qu'à l'aspect purement visuel du résultat (robotique, génération d'images synthétiques, médical, etc.). C'est pourquoi, dans notre domaine, il importe de ne pas faire l'amalgame entre *traitement d'images* et télédétection, même si parfois la donnée (satellitaire ou non) peut être traitée directement en tant que représentation sur ce type de matériel.

3.2.2. Mode graphique ou vecteur

On cherche à représenter une donnée par des moyens graphiques issus de la technologie des traceurs. Ce mode est en évolution constante au niveau matériel par l'introduction de la couleur, de l'interactivité et des techniques de rasterisation rendu possible par les progrès (et les coûts) de la technologie (cas des consoles graphiques, des cartes additionnelles sur PC et de certains traceurs).

L'accès à un périphérique s'effectue par des couches de logiciels spécialisés. Ces couches peuvent être disponibles avec les fonctions de base graphiques (interface virtuel) au sein de normes garantissant l'indépendance du matériel (norme GKS dans le domaine 2D).

De plus, il est possible d'accéder dans des bibliothèques spécialisées à des fonctions de hauts niveaux, permettant des représentations de données. Pour les champs ordonnés, l'utilisation de fonctionnalités de contourage, de vecteur flux, ou de représentation à trois dimensions (bloc diagramme) est quasi immédiate avec certains progiciels, qu'il est alors nécessaire d'interfacer avec les structures de données existantes.

L'utilisation d'une norme graphique, si elle permet de rationaliser les développements lourds par l'indépendance matérielle et logicielle, doit cependant être abordée avec précaution, puisqu'il peut exister des solutions ponctuelles, donc moins portables pour les programmes et données, pour bon nombre de problèmes (interface carte PC par exemple, logiciels boîte noire, etc.).

3.2.3. Evolution probable

Il est toujours possible de représenter un graphique sous forme image, en codant le pixel (position, profondeur) de façon adéquate sur le périphérique concerné. L'apparition récente d'écran mémoire (bit-map) tend à réunir, au niveau du matériel au moins, les deux voies distinctes de représentations infographiques. Au niveau du logiciel, et en l'absence de norme unificatrice, la mise en oeuvre de programmes de visualisation d'images est toujours liée à l'environnement existant, malgré des efforts de rationalisation en ce sens (visualisation d'image en mode GKS, interface image standard, normalisation matérielle, etc.).

3.3. Système de traitement

Compte tenu des définitions précédentes, le système logiciel développé pour nos besoins est schématisé sur la figure 2, qui met en évidence les différents liens entre données et représentations, ainsi que les différents traitements possibles.

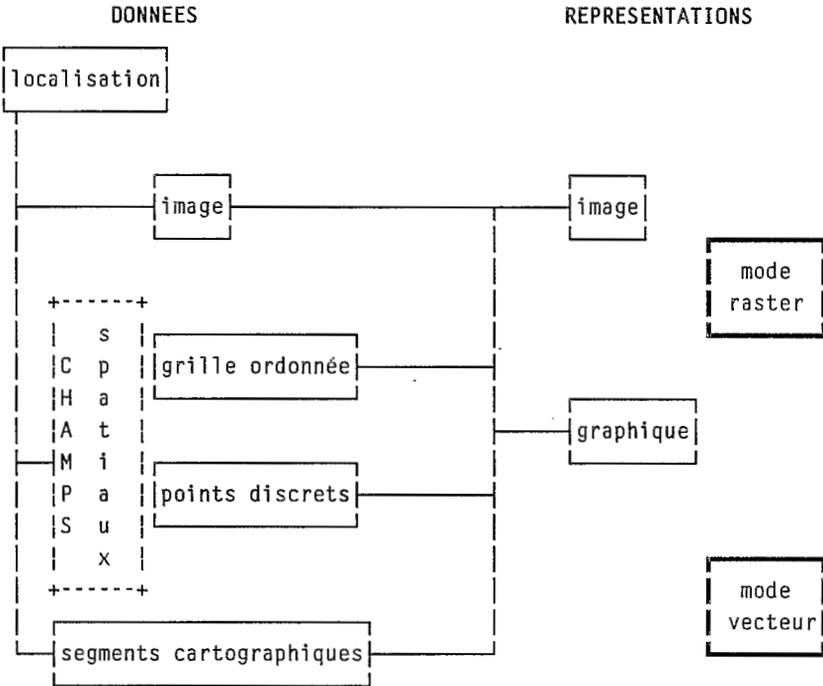


Figure 2 : Données et représentations en télédétection

4. SYSTEME LOGICIEL

Nous décrivons ici la réalisation du logiciel TRISKEL développé au sein de notre laboratoire sur un environnement matériel standard (calculateur scientifique UNIX). On trouvera en annexe une description plus détaillée des diverses fonctionnalités techniques disponibles.

4.1. Structure interne de données

Ce sous-système concerne la définition des formats internes des données, avec les modules élémentaires d'entrées / sorties, ainsi que des fonctions de niveaux supérieurs pour la gestion, l'accès et le stockage. Ces structures correspondent globalement à la nature des données définies précédemment.

Chaque type, dans un fichier unique, est identifié par un en-tête suivi des paramètres de la donnée (type, label, taille, etc.). Les diverses données peuvent ainsi être créées, examinées, éventuellement modifiées par un simple appel. Pendant toutes ces opérations, un contrôle d'erreur approprié est réalisé. L'accès aux données brutes par divers utilitaires (éditeur, *dump*, etc.) est également possible, ainsi que la redéfinition de modules d'interface suivant d'autres principes que ceux développés (optimisation possible, autre langage, etc.).

Les images, de dimensions quelconques en lignes, colonnes et canaux, sont stockées en binaire, un pixel étant codé sur 1 ou 2 octets suivant sa dynamique. L'accès peut être global sur un canal, ou ligne à ligne pour la majorité des traitements. On a ainsi stocké (et traité) une image de 60 Mo issue d'une acquisition AVHRR (NOAA) de 5 000 lignes, 2 048 pixels, 3 canaux (sur 5), un pixel étant codé en température sur deux octets.

Les données cartographiques peuvent avoir deux structures suivant leur taille. Pour des bases de données importantes (> 1 Mo), le stockage est réalisé sur deux fichiers, avec le descriptif des segments et les données proprement dites (en binaire). Pour des données de taille moindre, l'accès est de type formaté avec les outils d'extraction depuis le mode précédent.

Les autres données (localisation, champ discret ou ordonné, *Look-Up Table*, etc.), sont accessibles directement en mode formaté.

4.2. Décodage et mise en forme

En l'absence d'un réseau spécialisé pour l'acquisition de données satellitaires (station d'acquisition et de prétraitement), ces données sont obtenues sur support magnétique. Les multiples sources à traiter doivent alors être décodées, pour être converties dans les structures de données internes du logiciel d'application. Elles comportent généralement, en plus de la donnée image (brute ou prétraitée à des niveaux variables), diverses informations indispensables pour la suite des traitements géophysiques ou géométriques envisagés (géolocation, angles de visée, etc.). Si certains prétraitements ne sont pas inclus dans la fourniture, il peut être intéressant d'y procéder dès ce stade, au moyen de divers modèles externes (localisation, calibrage radiométrique, etc.).

Les données conventionnelles sont saisies manuellement, ou lues sur support magnétique. Elles peuvent être également accessibles sur certains environnements de stockage spécialisés, comme le Système Mondial de Transmission d'informations météorologiques (Météorologie Nationale, Paris), à partir d'un calculateur équipé de moyens de communication, puis transformées en respectant leur format interne.

4.3. Traitements

Ceux ci peuvent être très variés, et correspondent essentiellement à des outils méthodologiques de base. Nous ne traiterons pas des développements spécifiques thématiques écrits pour les besoins de nos programmes de recherche.

4.3.1. Traitement d'imagerie

Ce module assure des manipulations de données image, avec les opérations de transfert général sur support magnétique (restitution, archivage), de traitement classique (arithmétique, dynamique, convolution, statistique, etc.), ainsi que divers autres plus spécifiques développés pour nos applications (analyse d'images multitemporelles, etc.)..

4.3.2. Traitement de champs

Il consiste en manipulations des données de ce type et en génération d'images à partir de champs ordonnés. Un module élémentaire de reconstitution de grille à partir de données discrètes est également disponible, et il est prévu à terme d'y inclure des méthodes évoluées de géostatistique.

4.3.3. Traitements géométriques

Ce module assure la parfaite transparence géométrique de diverses sources. Il concerne les étapes de superposition d'informations spatialisées (cartographie, champs) avec l'image, ainsi que la restitution cartographique de ces résultats (rectification). La maîtrise de ces étapes permet à terme l'utilisation de méthodes *multisource - multitemporelles*, à partir des données disponibles, celles-ci pouvant être quelconques, et à des pas de résolution spatiales, spectrales et temporelles très différents.

Ces traitements intègrent plusieurs modèles : projection cartographique, localisation d'images satellitaire (orbitographie, capteur, etc.), déformation, ré-échantillonnage image.

4.3.4. Accès à d'autres logiciels

Nous avons développé des passerelles de reformatage de données entre notre logiciel et certains autres, aux fonctionnalités complémentaires. A ce sujet, un effort d'*homogénéisation* des données (voire des méthodes) devrait intervenir pour éviter une certaine dispersion technique, au sein d'une même communauté aux finalités proches.

4.4. Sous système image et graphique

Compte tenu de l'environnement matériel disponible, l'aspect restitution d'image a été développé autour de matériels spécifiques (donc non portables), avec des couches d'accès logiciels adaptées. Sont concernés la copie de fichier image sur une imprimante couleur, et les échanges bidirectionnels avec un environnement spécialisé, (performant, fermé et coûteux), de traitement et de visualisation d'images (affichage écran et manipulation mémoire et vidéo, mise en forme, texte, etc., mais aussi récupération de paramètres et de données image, numérisation, etc.).

Cependant, le développement de ces couches a été réalisé afin de permettre une large réutilisation des programmes mis en oeuvre, sur des matériels aux mêmes fonctionnalités. La sortie vecteur sur ces matériels raster a été introduite, permettant la restitution de graphique avec l'image.

Les manipulations et restitutions purement graphiques (interface matériel, logiciel de base et de *haut niveau*) sont réalisées indépendamment du matériel (écran, traceur), grâce à l'utilisation de la norme GKS et de programmes d'utilitaires disponibles.

4.5. Interface utilisateur

Ce système intègre les sous-systèmes précédents, et éventuellement les manipulations *thématiques* concernées par assemblage au sein d'un programme utilisateur adapté à l'environnement.

La figure 3 illustre les possibilités matérielles des stations de travail actuelles, en *traitement* source-destination depuis et vers divers *périphériques* communs (disque dur, mémoire), ou spécifiques en entrées/sorties (systèmes d'acquisition de données, visualisation écran, restitution papier).

L'exemple trivial est le chargement d'une donnée image disque sur l'écran de visualisation, mais on peut en imaginer d'autres avec l'utilisation simultanée de plusieurs *fenêtres*, chacune ayant une fonction différente.

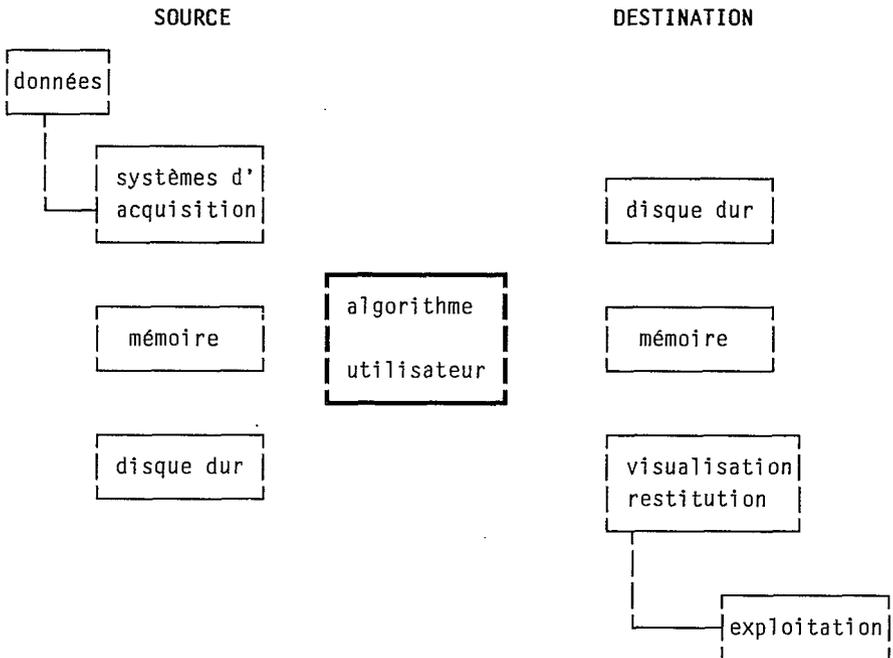


Figure 3 : Environnement utilisateur

Les différents *traitements* ont été conçus comme des entités indépendantes, entités susceptibles de produire un logiciel applicatif général (libre service), ou plus dédié vers un type d'exploitation particulier (recherche - développement, opérationnel), par l'appel des bibliothèques d'utilitaires existantes (gestion, traitement standard, restitution, etc.) et de développements spécifiques.

5. CONCLUSIONS

Le développement du logiciel TRISKEL est une illustration des possibilités pratiques de mise en oeuvre d'une *station de travail*, pour les études d'environnement intégrant diverses données provenant de multiples sources : télé-détection (dont satellitaires), conventionnelles ou base de données. La démarche suivie est de construire des sous-systèmes indépendants, interfacés avec des structures de données puis de réaliser un interface utilisateur à l'aide d'un système d'exploitation puissant et portable. Les divers produits disponibles l'ont été indépendamment des matériels, en mettant l'accent sur les aspects logiciels : structures de données, mode de représentation, analyse et écriture de programmes de *qualité*, utilisation de normes, etc.

C'est un exemple de développement utilisable par une large communauté d'utilisateurs scientifiques, voire opérationnels, en éliminant les duplications coûteuses de certains investissements humains. Les concepts de base en font un système dynamique, qui peut être facilement implanté et modifié pour diverses utilisations. Ce système peut également évoluer vers le haut avec des technologies de pointe, ou être adapté aux nouveaux, toujours plus puissants et peu coûteux matériels *personnel*.

Le défi d'aujourd'hui est de générer du logiciel applicatif à des coûts de développement acceptables. Les logiciels modulaires, extensibles, maintenables, réutilisables, en un mot *standards*, doivent devenir une réalité. La clé du succès de la communauté scientifique faisant appel à l'informatique est le LOGICEL, et la clé de la réussite du logiciel est la STANDARDISATION.

ANNEXE

DESCRIPTIF TECHNIQUE DU LOGICIEL TRISKEL

Développé à l'Antenne ORSTOM de Lannion - Centre de Météorologie Spatiale.

Auteur : DAGORNE D.

Ecrit sur ordinateur scientifique (UNIX) en FORTRAN77 et C, il est utilisé localement avec une console de traitement d'images PERICOLOR 1500 (MSII), et une imprimante couleur à impact COLORPLOT.

DONNEES

- Image de taille quelconque, éventuellement multispectrale, pixel codé sur 1 ou 2 octets. Chaque image peut être géoréférencée par une donnée de localisation ;
- Champs spatiaux ordonné (grille) ou discret (points) ;
- segments cartographiques.

FONCTIONNALITES

- *Gestion des données*
interface vers les structures de données (stockage, accès, etc.).
- *Traitements d'images (disque)*
transfert support magnétique, format *normalisé* (archivage, restitution) ou quelconque ;
transfert disque : extraction canal ou zone ;
opérations classiques : dynamique, arithmétique, logique, statistique, géométrique, locale (filtrage par convolution, par élément structurant, etc.) ;
opérations avancées (appliquées à nos données) : analyse suite d'images, *reconnaitances de forme*, etc.
- *Restitution images*
interface de visualisation et d'impression (adaptée aux matériels) ;
développements en cours et prévus suivant disponibilité matériel pour station de travail (X-window, .), micro (carte EGA, ..) ;
traitement *interactif* de données géoréférencées sur console image (cartographie, localisation image, etc.).
- *Traitements géométriques*
élaboration de localisation brutes (satellitaires), ou rectifiées (cartographique) ;
modèle de localisation directe et inverse ;

modèle de déformation entre images brute et rectifiée, par points d'amers ou entre localisations ;

modèle de réchantillonnage image (rectification), possibilité de mosaïque.

- *Cartographie*

base de données segments (trait de côte, frontière politique, hydrographie) de couverture mondiale à diverses résolutions (origine NCAR ou WDBII/CIA) ;

donnée d'un MNT mondial (résolution $1/6^{\text{ème}}$ de degrés) ;

modèles de projections cartographiques classiques (stéréopolaire, Mercator, Lambert, etc.) ;

génération d'image et de graphique.

- *Localisation et angles satellitaires*

imagerie géostationnaire (Météosat, GOES, etc.) ;

imagerie AVHRR, (radiomètre à balayage des satellites défilants NOAA), à partir d'un modèle d'orbitographie défilant (zone d'acquisition station, paramètres de poursuite, etc.) (origine SATMOS) ;

calculs d'angles satellitaires et solaires.

- *Traitements de champ*

manipulations entre champs (extraction, calcul, etc.) ;

reconstitution d'image ;

génération champ ordonné à partir de données discrètes (en développement).

- *Décodage données satellitaires*

adaptés à chaque source avec, si nécessaire, paramètres de localisation image et de calibrage radiométrique, afin d'entreprendre des traitements géophysiques et géométriques (niveau 1b ou moins) ;

météosat acquisition PDUS (EUMETSAT) ;

météosat archive ESOC (window, B2, etc.) ;

NOAA format master (SATMOS Lannion) ;

imageur AVHRR et sondeur TOVS ;

NOAA format HRPT station (non normalisé, origines diverses) ;

NOAA GAC, LAC (NESDIS) ;

NIMBUS/CZCS (Earthnet).

- *Liaisons avec autres logiciels*

utilisation NCAR/GKS (version 2.0) ;

utilisation GKSbx pour fonctions de base graphique ;

passerelle de données, depuis et vers divers progiciels *micro* ;

passerelle de données vers divers produits ORSTOM (PLANETES, TIMOR, etc.) et CMS (FIS, Vizir, etc.).

NEPTUNE : UN SYSTEME POUR LA CONSTITUTION ET L'EXPLOITATION DE BASES DE DONNEES SUR L'ENVIRONNEMENT

(Anne-Marie AUBRY, Cossi-Jean HOUNDAGBA, Jean-François
RICHARD & Patrick SECHET).

RESUME - Initialement conçu pour être appliqué aux données de sols collectées par les pédologues, le système Neptune est un ensemble de programmes opérationnels depuis 1980, sur gros système (CIRCE, CNRS - ORSAY), développé au centre ORSTOM de Bondy par Anne-Marie AUBRY avec le concours de Raymond VAN DEN DRIESSCHE.

Ce système a été exploité par C.-J. HOUNDAGBA pour l'analyse des paysages du Bénin. Ce travail a été à l'origine de la thèse "Analyse typologique des paysages d'Abomey-Zagnanado (Bénin) : exploitation d'un système de programmes PL1 NEPTUNE". La terminologie utilisée est basée sur la méthode d'appréhension du milieu naturel adoptée et enseignée par des chercheurs de différentes universités africaines et que J.-F. RICHARD, qui en est l'un des principaux concepteurs, expose dans sa thèse "Le paysage, un nouveau langage pour l'étude des milieux tropicaux". Enfin, et c'est encore sur des données factuelles situées géographiquement, mais décrites avec une terminologie n'ayant rien de commun avec les deux précédentes, que le système a été exploité dans un but didactique, de façon à expliquer les différentes phases de mise en place d'une application et les exploitations.

Les fonctionnalités que propose NEPTUNE sont celles liées à la capitalisation de l'information traitée : saisie, validation et mise à jour, d'une part, édition sélective des enregistrements stockés et possibilités d'interrogation, d'autre part. Des possibilités d'inventaire de l'information contenue sont également disponibles sous diverses formes, notamment graphiques, et facilitent une vision synoptique de l'état de la base de données.

Neptune présente la caractéristique de traiter de façon efficace les données descriptives non quantitatives, en utilisant la notion de glossaire (ensemble de tables de codification encore appelé nomenclature) qui permet de valider ce type de donnée à la saisie. Il permet à ses utilisateurs de définir librement les tables à utiliser avant la création de la base, puis de remplir, mettre à jour, consulter et éditer toutes ces tables au cours de l'exploitation du système tant pour la constitution que pour l'exploitation de la base de données.

La réalisation de ce travail repose sur une représentation des données de l'environnement selon un modèle hiérarchique à deux niveaux (hoplexol - élément diagnostique, par exemple). Ce modèle hérité de la conception originale est simple et présente l'avantage de pouvoir s'adapter à de nombreux objets d'étude, en offrant les fonctions de base du traitement des données moyennant seulement une définition appropriée du glossaire.

Pour répondre à un projet proposé par une équipe de géographes physiciens, de "*gestion de l'environnement biophysique ouest - africain*", on se propose de réaliser une nouvelle version de Neptune susceptible d'apporter un outil de traitement décentralisé des données géographiques collectées sur le terrain. Celle-ci sera opérationnelle sur micro-ordinateur PC-compatible, portable et autonome, et implémentera un modèle hiérarchique étendu à plusieurs niveaux, qui correspond à une représentation de l'environnement établie par ces géographes.

INTRODUCTION

Au début des années 1970, une équipe ORSTOM dirigée par Raymond VAN DEN DRIESSCHE développait un ensemble de programmes informatiques destinés à la constitution et à la gestion d'une base de données pédologiques. Un tel système, baptisé Poséidon, était spécialement mis au point pour permettre l'administration des données descriptives de profils de sols et s'appuyait sur l'utilisation des glossaires de pédologie (ORSTOM 1969 ; Informatique et Biosphère 1971).

Opérationnel sur gros système, Poséidon a dû évoluer en fonction de changements de matériels (Univac, IBM), de langages (Fortran, PL/1) et de la réalisation de fonctions complémentaires. La dernière version, appelée Neptune, est implantée sur le CIRCE à Orsay depuis 1980 (AUBRY et alii, 1984).

En 1981, une équipe de géographes en quête d'un outil informatique pour manipuler les résultats de la description des paysages s'est intéressée à ce logiciel. L'adaptation faite a permis d'utiliser le système dans un autre contexte, en montrant son indépendance par rapport au domaine d'application, moyennant la satisfaction de ses hypothèses de base.

Les utilisations qui ont été faites depuis 1984 sur des fichiers en vraie grandeur, par les géographes en Afrique de l'Ouest, conduisent à suggérer la réalisation d'une nouvelle version, opérationnelle sur micro-ordinateur portable.

La première partie de cet exposé décrit brièvement les caractéristiques essentielles du système existant, en insistant sur les aspects qui ont permis de le ré-exploiter dans une autre situation que celle à laquelle il avait été initialement destiné. Le chapitre suivant est consacré à la présentation de ses applications dans le cadre d'études géographiques, et sert d'illustration pour l'utilisation du système. Le texte se termine par la description sommaire du projet de réalisation d'un nouveau logiciel, sur micro-ordinateur, issu de l'expérience acquise grâce à Neptune, et élaboré pour répondre aux demandes des géographes africains.

1. DESCRIPTION DE NEPTUNE

1.1. Généralités

Le système NEPTUNE est destiné à traiter les données obtenues par un observateur lors de la description de l'objet qu'il souhaite représenter sur support magnétique. Il possède la particularité d'autoriser son utilisateur à employer les mots et expressions d'une terminologie proche du langage naturel (terminologie usuelle), qu'il a lui-même définie et qu'il peut enrichir comme il le désire, sans nécessiter l'apprentissage de termes étrangers à sa propre discipline.

Le logiciel n'est pas lié à une discipline particulière, mais il implémente un modèle hiérarchique à deux niveaux, c'est à dire qu'il permet de représenter des données structurées selon une hiérarchie père-fils (par exemple profil-horizon, hoplexol-élément diagnostique, etc.). Dans ce modèle, la sous-unité (deuxième niveau) est, à l'intérieur de l'unité (premier niveau), un sous-ensemble identifiable de variables, dont on rencontre plusieurs occurrences (AUBRY 1987). Le nombre de sous-unités par unité et le nombre de variables de chacun des niveaux varient selon les applications.

1.2 Fonctionnement général

Les données que le système manipule sont soit qualitatives (avec ou sans relation d'ordre), soit encore quantitatives entières.

A partir de la réflexion sur la structure des informations qu'on veut représenter (étape de conception), on constitue d'abord un répertoire (encore appelé nomenclature, glossaire, catalogue ou lexiques) dans un module qui est indépendant de la banque elle-même, et qui contient la *description des données* (au sens SGBD). C'est en se référant à ce répertoire que le système reconnaîtra, vérifiera et stockera les données qui lui sont fournies.

Le système est modulaire, constitué de trois ensembles en interaction :

- le répertoire, ou un des répertoires équivalents ;
- les programmes en binaire exécutable ;
- les données elles-mêmes.

C'est sur les données et sur le répertoire que l'utilisateur agit. D'abord, c'est lui-même qui constitue le répertoire, fichier séquentiel où il décrit ses données et ce qu'il souhaite entrer dans la base. Quant au fichier des données, c'est le système qui se charge de le formater et d'y enregistrer les codes, de façon transparente pour l'utilisateur, en se servant en même temps du répertoire qui lui est désigné et des données saisies par l'utilisateur dans le langage usuel choisi. Cette saisie opère sans ordre imposé ni format : le point est utilisé comme séparateur de données et il n'y a pas d'autres contraintes d'écriture (voir figure 1).

1.3. Entrée des données

Les données sont introduites avec deux clés logiques correspondant à la structure unité et sous-unité. Ces clés sont deux variables entières choisies par l'utilisateur et identifient respectivement l'unité à l'intérieur de la banque et, relativement à cette unité, le numéro de la sous-unité.

Neptune

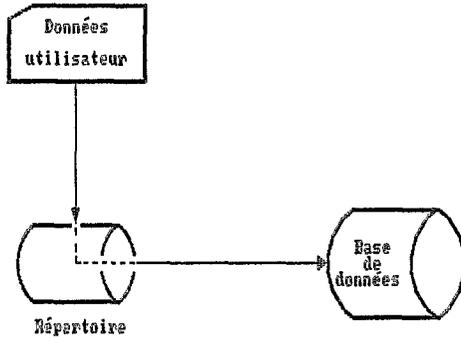


Fig. 1a Schéma d'entrée des données

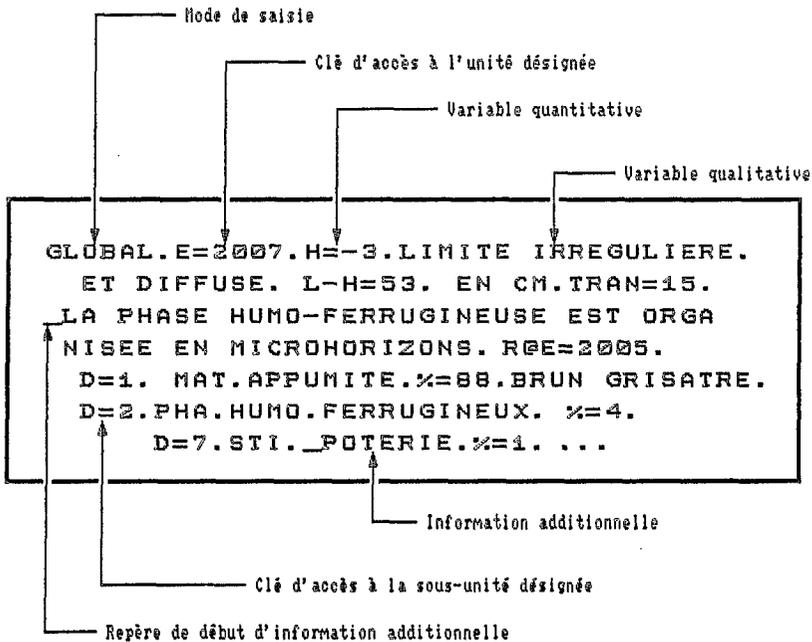


Fig. 1b Exemple de données à l'entrée

Le schéma d'introduction des données est ainsi le suivant :

- clé logique de l'unité ;
- données de l'unité dans un ordre quelconque ;
- information additionnelle éventuelle ;
- ensuite, clé logique de sous-unité, données correspondantes et information additionnelle éventuelle, autant de fois qu'il y a de sous-unités, le nombre maximal en ayant été fixé lors de la définition de la base (étape de spécification).

Il est à noter que si on entre plusieurs données valides pour une variable, c'est la dernière qui est enregistrée. D'autre part, on n'entre que les données effectivement présentes.

Enfin, toutes ces données, clés comprises, sont saisies sans format imposé.

1.4. Validation des données

La validation des données saisies est réalisée automatiquement par le système, par comparaison avec le répertoire auquel on fait référence. Il vérifie en particulier que les données appartiennent bien aux domaines de "valeurs" qui sont précisées dans le répertoire et il *annote* en sortie les données non reconnues.

1.5. Mises à jour

Pour faire les mises à jour, on peut agir à la fois sur le répertoire et sur les données.

Le répertoire est mis à jour en y ajoutant les données qu'ensuite il reconnaîtra. Les variables quantitatives y ont leurs bornes : celles-ci peuvent être modifiées. Les variables qualitatives y ont toutes leurs valeurs admissibles, exprimées sous forme de chaînes de caractères : ces listes peuvent être enrichies de synonymes et de termes supplémentaires.

Pour les données il y a deux modes de correction : un mode "local" qui permet de modifier une seule donnée si l'on veut, un mode "global" qui permet de remplacer toutes les données d'une unité et de ses sous-unités. Dans le mode local, on saisit uniquement la donnée correcte introduite par sa clé d'unité, et de sous-unité quand c'est le cas. On peut utiliser les deux modes conjointement au cours du même traitement.

Il n'y a pas de langage de manipulation à apprendre pour entrer et mettre à jour les données. Il suffit de connaître les mots-clés (global et local) qui introduisent le mode de saisie, et le vocabulaire du répertoire.

1.6. Interrogation

Il n'y a pas non plus à connaître un langage spécial pour interroger la banque. L'expression de sélection ne fait appel qu'à la terminologie usuelle répertoriée, aux opérateurs relationnels stricts (<, >, =) et aux opérateurs logiques *et*, *ou*, *non* (&, !, ^). De plus, on peut utiliser les parenthèses pour condenser son écriture et modifier les règles de priorité des opérateurs, en accord avec les règles de l'Algèbre de Boole. Enfin celle-ci est écrite sans format.

L'expression booléenne élémentaire est de la forme $A > B$, $A < B$ ou $A = B$, où, dans le cas le plus général, A est une variable du répertoire et B est une donnée (constante chaîne) ou une constante entière. L'utilisation des opérateurs booléens (&, | et ^) sur ces opérands (les expressions booléennes élémentaires) permet, avec les règles de priorité classique et les parenthèses, de formuler toutes les questions.

Exemple, on pourrait rencontrer l'expression booléenne élémentaire : "RADICAL PRIMAIRE = STRUCTICHRON", pour sélectionner les hoplexols (unités) dont au moins un des éléments diagnostiques (sous-unités) est un structichron (valeur de la variable *radical primaire*). En rajoutant "& AUTEUR = HOUNDAGBA", on extrait seulement ceux décrits par Cossi-Jean Houndagba.

L'interrogation fournit un double résultat :

- une sortie imprimée contenant, dans l'ordre croissant, les numéros des enregistrements (unités) qui répondent à la question posée ;
- un fichier de ces mêmes enregistrements, présentant les mêmes caractéristiques physiques et logiques que le fichier principal. Celui-ci permet ainsi de poursuivre les opérations de recherche de façon répétée ("nested search").

1.7. Impression/édition

Au cours de tous ces traitements, ce qui est entré au clavier est imprimé tel quel, sauf dans le cas où, pour mettre en évidence l'anomalie détectée, Neptune remplace le point terminal de la donnée par un point d'interrogation. On peut bien sûr éditer le fichier, avec mise en page, totalement ou de manière sélective, en donnant une liste des numéros d'unités.

1.8. Autres exploitations

Le système fournit également un synopsis compact de quantité d'information stockée (voir figure 2) et, d'autre part, les fréquences d'utilisation des données.

Par ailleurs, les fichiers séquentiels obtenus à partir de la banque peuvent être traités par les logiciels statistiques du centre de calcul (Circe) où elle est implantée.

2. EXPERIENCE D'UTILISATION

On a exploité ce logiciel, implanté sur le centre de calcul du CNRS (Circe) à Orsay, pour traiter les descriptions typologiques de paysages dans le cadre de travaux de recherche de plusieurs géographes en Côte d'Ivoire, au Bénin et Sénégal.

La terminologie conçue dans cette méthode d'appréhension du milieu naturel a été aménagée pour l'utilisation de Neptune, ce qui a conduit les géographes de l'équipe à faire des recherches sur le vocabulaire d'analyse typologique des paysages dans le sens d'une rigueur appropriée au traitement informa-

tique, sans rien abandonner de leurs principes. Ainsi, les concepts de diagnostics typologiques développés par J.F. RICHARD dans sa thèse (RICHARD 1985), qui ont été repris lors de l'étude des paysages du Bénin, réalisée par C.J. HOUNDAGBA, sont à la base de la structuration des fichiers.

Pour l'analyse des paysages de la région d'Abomey-Zagnanado (HOUNDAGBA 1984), au Bénin, les descriptions relevées sur le terrain étaient consignées dans des carnets de terrain, tandis que des expériences plus récentes au Sénégal (missions 1986-87) utilisaient une saisie directe sur micro-ordinateur autonome portable. Dans tous les cas ces données étaient ensuite transférées, sans traitement préalable, sur le site central. Les traitements eux-même (corrections des données, mises à jour des répertoires, interrogations et extractions sélectives) se faisaient naturellement sur le centre de calcul.

La structure hiérarchique arborescente à deux niveaux (structure père-enfants) a permis de traiter l'unité HOPLEXOL définie par 300 variables. 800 caractères d'information additionnelle libre, mais restituable, donc pouvant servir de base par exemple à des mises à jour des répertoires, forment une zone purement textuelle qui complète l'enregistrement.

Chaque hoplexol peut contenir jusqu'à 20 ELEMENTS DIAGNOSTIQUES (sous-unités), chacun d'entre eux pouvant être renseigné à l'aide de 35 variables (et une zone de 20 caractères d'information additionnelle libre).

Les deux niveaux permettent de traiter les strates perçues en un point où le chercheur fait un relevé (strates de végétation dans l'atmosphère et couches dans le sol), avec, pour chaque strate, les éléments diagnostiques qu'il y reconnaît.

On peut fournir quelques valeurs indicatives dans le cas concret du fichier Sénégal : le répertoire contient 2 000 valeurs de variables qualitatives et une chaîne de 35 000 caractères, représentant les textes correspondants. Le fichier des données contient 741 enregistrements de 3 800 octets, avec près de 20 000 informations élémentaires, obtenus par transfert de 3 285 images-cartes fournies par le micro-ordinateur de saisie sur le terrain. Le fichier Bénin, quant à lui, occupe actuellement 1 450 enregistrements.

Le répertoire est mis à jour progressivement, au fur et à mesure que se constitue le fichier. Cette facilité permet au chercheur d'insérer des termes nouveaux dans le cadre d'une terminologie susceptible d'évoluer.

Naturellement l'expert ne saisit que les données qu'il observe, hoplexol par hoplexol, et à l'intérieur de l'hoplexol, élément diagnostique par élément diagnostique. Il les introduit respectivement par des clés numériques qu'il donne au système sous forme d'un entier associé au libellé qu'il a choisi, comme pour n'importe quelle variable au moment de la constitution du vocabulaire structuré de référence. De plus l'ordre de saisie n'a pas d'importance et, en outre, si on saisit plusieurs données pour la même variable c'est la dernière valide qui est stockée dans la banque.

Tout ceci offre une certaine souplesse pour la saisie sur le terrain, surtout en utilisant un moyen informatique dès cette étape : la démonstration en a été faite lors de l'expérience la plus récente, notamment au cours des deux missions au Sénégal. En effet, on a noté les avantages suivants liés au système de saisie informatisée sur le terrain :

- gain de temps puisqu'il n'y a pas retranscription d'un carnet de terrain pour la saisie sur support magnétique ;
- pas d'erreurs inhérentes aux transcriptions.

En outre, il n'y a pas d'erreur non plus qui résulte d'un codage quelconque, que la validation soit faite par le logiciel qui identifie les expressions comme admissibles par rapport à un vocabulaire qu'on lui a donné par ailleurs, ou faite à l'aide d'une saisie conversationnelle par menu explicite.

Actuellement l'enseignement de la méthode d'appréhension du milieu naturel, dont J.F.RICHARD est un des principaux concepteurs, se fait dans plusieurs universités africaines francophones et une vingtaine de chercheurs de cette "école" l'enseignent et l'utilisent sur le terrain en Afrique. Cette méthode a pour but une meilleure connaissance du milieu naturel avec toutes les applications que cette connaissance draine avec elle. L'apport de l'utilisation de Neptune a permis dès à présent de cerner quelques problèmes liés à l'utilisation de l'informatique et de mieux définir certaines parties du vocabulaire de description.

3. PROJET MICRONEPTUNE

C'est grâce à cette expérience que les demandes de nos partenaires (y compris celles qui ont déjà été formulées officiellement) de réaliser un système répondant mieux à leurs contraintes matérielles et logistiques - à l'heure de la microinformatisation et de la banalisation de matériels portables et néanmoins puissants - se sont tournées vers notre équipe.

Compte tenu de ce que Neptune était opérationnel sur gros système et malgré la réalisation d'une interface-terrain permettant la saisie *in situ*, l'intérêt de concevoir et mettre au point un système complet et autonome sur micro-ordinateur est apparu évident, ne serait-ce qu'en vue d'une diffusion plus ample et du transfert d'une technologie adéquate pour les pays en développement. Un tel logiciel permettra en effet aux chercheurs intéressés d'être totalement indépendants et de garder la maîtrise de leurs données, tout en maintenant la cohérence scientifique garantie par l'utilisation d'une méthode commune de description.

De plus, l'expérience acquise par l'exploitation de NEPTUNE et de son interface-terrain, a aussi conduit les géographes de l'équipe à mieux identifier les informations à recueillir *in situ*, par rapport à ce qui peut être obtenu par l'étude des documents disponibles (rapports, documents cartographiques, etc.).

On se propose donc de réaliser un logiciel sur micro-ordinateur capable de supporter la saisie, la validation et l'exploitation de l'information que l'on peut recueillir sur l'organisation des milieux et l'état des ressources naturelles, selon l'approche typologique à laquelle ont adhéré les scientifiques participants au "Séminaire sur la dégradation des paysages ouest-africains", qui s'est tenu à Dakar en novembre 1988 (MBOU & RICHARD, 1988).

A l'occasion de la réalisation de ce travail, une nouvelle réflexion approfondie sera menée sur la structuration des informations à représenter, sur la définition des attributs caractéristiques de chaque entité et sur les relations existantes entre ces dernières.

Les fonctionnalités essentielles de ce logiciel seront disponibles sur matériel autonome portable, de façon à permettre la saisie et la validation des données sur le terrain ainsi que des ressources minimales d'interrogation sur un fichier de "campagne". Des fonctionnalités complémentaires seront implémentées pour permettre la constitution et la gestion de bases de données composées de plusieurs fichiers de campagne créés par utilisation de glossaires compatibles.

Le logiciel sera opérationnel sur micro-ordinateur compatible IBM-PC portable autonome, avec au moins 512K de mémoire centrale et deux unités de mémoire externe intégrées. On cherchera à rendre le logiciel aussi convivial et fiable que possible, compte tenu toutefois du public scientifique visé, des conditions particulières d'utilisation et de l'aspect pédagogique que l'on souhaite donner au projet.

Les traitements sur micro-ordinateur donnent l'autonomie et l'indépendance, en l'absence desquelles le chercheur sur le terrain se voit privé du moyen de contribuer efficacement à l'ensemble des recherches sur les paysages. Indépendamment de la seule nécessité d'un matériel peu coûteux, il est essentiel de ne pas avoir à différer les traitements ni à se déplacer dans un centre de calcul avec les problèmes d'accès (à l'endroit où se trouve le matériel, mais aussi au matériel lui-même) que cela comporte, (quand ce n'est pas en Métropole).

Cette décentralisation doit donner des chances égales à chaque chercheur, où qu'il travaille, et quelles que soient ses possibilités financières, d'apporter à des études régionales les connaissances qu'il acquiert au moyen des interrogations qu'il pourra multiplier sur ses données. C'est en effet à partir de ces interrogations que le chercheur découvre, confirme ou infirme des règles d'interaction entre les objets manipulés, qui pourront ultérieurement servir à la construction de "bases de connaissances". Sans cette décentralisation, il serait matériellement et physiquement exclu de ce processus, et ce malgré toute la bonne volonté scientifique et politique des confrères des pays développés.

CONCLUSION

Le logiciel Neptune constitue un outil de recherche présentant divers avantages pour les sciences de l'environnement :

- il s'agit d'abord d'une aide à la constitution d'une terminologie descriptive pour structure arborescente à deux niveaux. Il permet de constituer cette terminologie par approches successives, régies par des choix rigoureux ;
- il aide à l'apprentissage de cette terminologie, dans le cadre de formation et d'enseignement (annotation et rejet de ce qui n'est pas conforme).

Comme outil opérationnel de constitution de bases de données, il présente en outre une certaine convivialité, grâce :

- à sa facilité d'utilisation sur le terrain, due à la souplesse de la saisie, sans codes et sans format ;
- aux mises à jour locales, d'une donnée, pour une unité et sous-unité identifiées, en rentrant simplement la donnée ;

- au fractionnement à volonté de la saisie-stockage, avec les modes de saisie global/local alternés ;
- aux interrogations faciles à exprimer et aux extractions rapides ;
- à l'absence générale de langage de manipulation, de format, de formalisme et à l'emploi d'un minimum de paramètres dans les procédures.

Ce système est de plus particulièrement bien adapté à une recherche en évolution, où l'on peut proposer la création de nouveaux diagnostics qui peuvent venir compléter les listes.

En résumé, on peut rappeler les caractéristiques essentielles du système : entrée des données en un langage proche du langage naturel, défini par les utilisateurs, sans format et sans ordre obligatoire ; mises à jour globales ou locales alternées à volonté, interrogations ne faisant intervenir aucun langage de manipulation, car n'utilisant que les opérateurs relationnels simples et les trois opérateurs logiques appliqués aux variables et données du vocabulaire de référence, avec une écriture elle aussi en format libre.

De plus le système est indépendant de la terminologie choisie. Il décharge l'utilisateur d'une codification fastidieuse et de l'apprentissage d'un langage de commande le plus souvent lié à un SGBD, quel qu'il soit. L'utilisateur doit essentiellement définir la terminologie utilisée dans son domaine scientifique, qui est enregistrée dans le répertoire et sert de référence pour tous les traitements.

REFERENCES BIBLIOGRAPHIQUES :

- AUBRY, A.-M. et al. *From POSEIDON to NEPTUNE - Software for environmental surveys*. In : Burrough P.A. et Bie S.W. (Ed.). *Soil information systems technology*, Bolkesjo, pp. 41-45, 1984.
- AUBRY, A.-M. *NEPTUNE, logiciel de traitement de données descriptives : mise en place et traitement d'une application*. ORSTOM, Bondy, 1987.
- DRME/ORSTOM. *Poseidon : rapport final de convention*. 400p. 1975. (diffusion restreinte).
- HOUNDAGBA, C.J. *Analyse typologique des paysages d'Abomey- Zagnanado (République Populaire du Bénin). Exploitation d'un système PL1 NEPTUNE*. Thèse de 3ème cycle, Université Louis Pasteur, Strasbourg, 280p, 1984.
- Informatique et Biosphère. Glossaire de pédologie : description de l'environnement en vue du traitement informatique*. Informatique & Biosphère ed., 1971.
- ISSS Working group meeting on soil information systems. *Data input and data structure*. pp. 17-21, Wageningen (Hollande) 1975.
- ISSS Working group meeting on soil information systems. *Poseidon*. pp. 68-70, Druzba (Bulgarie) 1977.
- MBOW, L.S. & RICHARD, J.F. *Gestion de l'environnement biophysique ouest-africain. Compte-rendu du Séminaire sur la dégradation des paysages ouest-africains*. Dakar, Nov. 1988.
- ORSTOM. *Glossaire de pédologie : description des horizons en vue du traitement informatique*. ORSTOM, Init. & Doc. Tech. n° 13, 1969.

RICHARD, J.-F. *Le paysage, un nouveau langage pour l'étude des milieux tropicaux*. Thèse de Docteur d'Etat, Université de Paris-VII, 1985. 210p. + fascicule hors-texte (68p.). ORSTOM, Paris 1989.

	0	1	2	3	4	5	
	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	

1 C	I					I	C 50
2 C\$	I					I	C\$ 50
3 S\$	I					I	S\$ 50
4 E	I					I	E 44
5 C%	I					I	C% 29
6 R\$	I					I	R\$ 50
7 P	I					I	P 43
8 P%	I					I	P% 38
9 PKM	I					I	PKM 41
1 F	I	111111111111211111111111112211111222221112112111111				I	F 61
2 EMB	I					I	EMB 0
3 C50	I					I	C50 0
4 C215	I					I	C215 0
5 C600	I					I	C600 0
6 C601	I					I	C601 0
7 DHC5	I		1			I	DHC5 1
8 DHC6	I					I	DHC6 0
9 DASH	I				1 1	I	DASH 2
10 DSH8	I					I	DSH8 0
11 TWIN	I	1	111			I	TWIN 4
...							
65 B707	I	111	1 111	1 1111	1111	1	1 111 B707 20
66 B720	I		1 1	1	1		I B720 4
67 B727	I	11 1111	1 1	1 1 1	11 11 1 1	1 111	B727 19
68 B737	I	111 111111	11 1 1 11	1 11 1 2 1		1111	B737 24
69 B747	I	1 1 1	11 1	22 1 21		1 111	B747 16
70 B757	I					I	B757 0
71 B767	I	1	1			I	B767 2

	0	1	2	3	4	5	
	1234567890	1234567890	1234567890	1234567890	1234567890	1234567890	
	1111	111	111111	1 11111	11 11111	111111	1011111011
	83241801081145109090222053582001452450062700103943						

548

Fig. 2 : Fac-simile d'un extrait de sortie de "synopsis 1".

Chaque ligne correspond à une variable et chaque colonne à un enregistrement. Pour le niveau unité, la présence d'une donnée est représentée par un point à l'intersection de la ligne et de la colonne qui lui correspondent. Pour le niveau de sous-unité c'est la fréquence d'apparition qui est représentée.

L'ATLAS DE QUITO : METHODES INFORMATIQUES

(Marc SOURIS)

RESUME - Nous présentons dans cet article un prototype de système d'information géographique pour la gestion et la planification urbaine, dans ses principes comme dans son utilisation pour le projet d'Atlas informatisé de Quito, actuellement en cours.

Le prototype est conçu, dans son système de gestion de données, comme une extension du modèle relationnel à la gestion des données localisées : les SGBD n'étant pas à même de manipuler des données de dimension deux, nous avons développé des possibilités de gestion étendant l'algèbre relationnel aux données localisées : restriction spatiale, jointure spatiale, projection spatiale. Ces nouvelles opérations nous permettent de manipuler la localisation par des opérations algébriques ou logiques : la carte considérée comme résultat final devient le résultat intermédiaire d'une série d'opérations de gestion et de manipulation de données.

Le système développe un certain nombre de fonctionnalités propres aux données localisées, dont le principe est toujours de créer des attributs en mettant les objets en relation sur un critère d'appartenance ou de distance : sélection des points à une distance donnée d'une zone, critère d'appartenance d'une zone à une autre, etc.. tout en conservant le principe de la gestion relationnelle, le système est également étendu à la manipulation d'objets particuliers, comme les images satellites, les photographies aériennes, objets qui, grâce à des opérations spécifiques développant l'aspect multi-média et la gestion orientée-objet, sont intégrés dans le processus relationnel.

Toutes les fonctionnalités développées autour de ce système trouvent leur application au niveau de la planification urbaine. Nous effectuons la mise en oeuvre concrète du prototype dans le cadre du projet d'Atlas informatisé de Quito (AIQ) mené par l'ORSTOM avec plusieurs institutions équatoriennes (Institut Géographique Militaire, Municipalité de Quito, Institut Panaméricain de Géographie et d'Histoire). L'aire métropolitaine de Quito est confrontée à des problèmes de croissance, de fonctionnement, d'approvisionnement, et un tel outil va permettre en facilitant l'accès à l'information, d'améliorer la connaissance et donc la réflexion sur les problèmes de gestion de la ville.

La base de données est constituée à partir de l'information existante - saisie du découpage en îlots, avec intégration de données sur les équipements, l'infrastructure, le foncier, les constructions, saisie du découpage en unité de recensement avec intégration des données censitaires, saisie des différents secteurs administratifs de la ville, saisie de données du milieu physique sur l'aire métropolitaine (altitude, utilisation du sol, géologie) - et également grâce à l'information récoltée par des enquêtes spécifiques, exhaustives (recensement des activités, par

exemple) comme par sondage (démographie, migration, vulnérabilité aux risques naturels, etc.). Toutes ces données sont analysées, confrontées, de manière à tirer à la fois un bilan thématique de la situation existante et une vision prospective en vue de la planification urbaine (redéfinition du POS, zones d'interventions prioritaires, aménagement des zones d'extension, etc.).

INTRODUCTION

Nous présentons ici les méthodes et moyens informatiques mis en oeuvre pour le projet ATLAS DE QUITO, qu'il s'agisse de matériel ou de logiciel.

Informatiser un atlas exige de savoir faire au moins trois choses : *saisir des données* (dont la plupart sont localisées, donc également saisir des plans), *gérer et traiter des données* (donc posséder un système approprié), et *cartographier des données*, car les résultats attendus pour un atlas sont essentiellement visuels. Mais le projet d'Atlas de Quito n'est pas seulement la constitution d'un ouvrage, conçu ou fabriqué grâce à des machines : c'est également la constitution d'un *outil permettant de faciliter l'accès à l'information, d'améliorer la connaissance et donc la réflexion sur les problèmes de gestion de la ville*. C'est à la création d'un tel outil que nous nous sommes attachés, et c'est dans ce sens que sont conçus les développements informatiques faits autour de ce projet.

1. GESTION DES DONNEES

L'informatique nous sert avant tout de système informatif. Pour une application du type atlas, il est d'abord nécessaire de disposer d'un *système de gestion de données performant*. Ici, nous avons multitude de sources de données, qui sont pour la plupart géographiques, et il s'agit bien sûr de gérer également cette localisation dans l'espace. Il faut pouvoir comparer, juxtaposer, superposer, sélectionner, etc., afin de produire l'information sur laquelle sont effectués les traitements et produits les résultats. La gestion de ces données nécessite à la fois puissance et souplesse : le système que nous développons (nommé TIGRE) essaye de répondre à ces objectifs.

Le *modèle relationnel* est à la base de notre système de gestion de données : laissant à l'utilisateur le choix de la navigation dans l'ensemble des données, c'est le modèle à la fois le plus simple et le plus performant en terme de fonctionnalité. L'interrogation des données avec TIGRE est *interactive* : le résultat d'une opération de gestion sert d'entrée à l'opération suivante, ce qui permet à l'utilisateur de construire sa requête, et d'aboutir, après un cheminement qui peut être complexe, à un résultat ayant fait intervenir un grand nombre de sources de données différentes et reliées entre elles au fur et à mesure de ce cheminement, quand les opérations demandées le requièrent.

2. GESTION DE LA LOCALISATION

Particularité essentielle du système de gestion que nous développons pour gérer les données de l'Atlas de Quito : *devoir gérer des données géographiques*. Prenant exemple sur la gestion relationnelle des données, nous avons étendu les principes de cette gestion au traitement de la localisation. Ce qui signifie que les opérations classiques (sélection, jointure - c'est-à-dire croisement - sur un critère commun) seront également disponibles pour la localisation : sélection sur un espace géographique (par fenestrage, par masquage), croisement de deux objets sur leur localisation (on les met en relation s'ils sont au même endroit, s'ils sont à telle distance l'un de l'autre, le critère de mise en relation étant toujours ici basé sur la distance). La localisation est une donnée intrinsèque : tout objet localisé pourra, sur ce critère, être comparé à tout autre objet localisé de la base de données, ce qui ouvre le champ à une multitude d'opérations de gestion sur des données qui n'ont comme critère commun, comme moyen de liaison, que leur localisation dans l'espace. Ces nouvelles opérations entrent bien sûr dans le processus normal d'interrogation interactive et font partie du cheminement de l'utilisateur dans la base de données : l'algèbre relationnelle est ainsi étendue à la localisation.

Cette méthode de gestion de la localisation a des conséquences importantes dans l'élaboration de l'information géographique de base : chaque donnée peut en effet être conservée dans sa propre implantation spatiale sans réduire les possibilités de traitements ultérieurs, puisque c'est le système qui se charge de gérer la localisation : plus question de carroyage régulier dans lequel on ramène toutes les données pour pouvoir les comparer, en faisant fi de l'implantation spatiale d'origine. Plus question non plus de système dans lequel la localisation n'est traitée que pour cartographie, où les données localisées ne sont comparées que visuellement par la superposition de sorties graphiques. Plus question enfin d'avoir à préciser pour un objet d'un certain type son appartenance géographique à un objet d'un autre type : la localisation seule suffira à établir cette appartenance. Ainsi, le P.O.S. ne sera plus une variable de la parcelle cadastrale : saisi à part comme un découpage thématique de l'espace, c'est lors de l'interrogation des données que l'on pourra établir le P.O.S. pour chaque parcelle. Si l'on veut modifier le P.O.S., on pourra le faire indépendamment des parcelles et trouver ensuite, par une opération de gestion de données, l'impact de la modification sur les parcelles concernées. Par cette méthode de gestion de la localisation, les opérations de simulation sont rendues possibles et faciles à mettre en oeuvre.

Garder chaque objet dans sa propre implantation spatiale, avec la possibilité de comparer et de mettre en relation la spatialisations de différents objets, met bien sûr en évidence les problèmes de validité de la localisation : des objets d'échelles spatiales différentes, ou plus généralement de validité différente pour leur localisation, ne sauraient être joints sur cette localisation sans grande précaution.

3. LES TRAITEMENTS

Le système de gestion nous offre donc, grâce à ces opérations de base, la possibilité de développer, sur un ensemble de données, de nombreux traitements spécifiques. C'est le second aspect du système développé, qui doit offrir des traitements sortant du cadre strict du système de gestion. Ces possibilités sont, sans en faire l'énumération complète :

- l'étude des données par des *moyens statistiques simples* (moyenne, fréquence, écart-type, histogrammes, corrélations, régressions, histogrammes bi-dimensionnels) ;

- la *création de nouvelles variables* (par classification, par calcul numérique, par calcul logique, par agrégation) ;

- en plus, un certain nombre d'opérations propres aux données localisées. Ce sont le calcul de *surface*, de *périmètre*, de *valeur d'appartenance* ou *agrégation* sur un critère géographique (moyenne de points dans une zone, de zones à zones, etc.), la création d'un *masque de sélection* géographique par rapport à certains objets (par exemple, l'espace défini par la condition de se situer à moins de 10 mètres des portions d'un réseau ayant une caractéristique donnée), l'*interpolation géométrique* et la création de bloc-diagrammes.

Ces traitements s'inscrivent encore dans le processus d'interrogation interactive de la base de données : on fait un histogramme pour rechercher les seuils de classification d'une variable, on classe la variable, on fait une sélection sur certaines classes, on calcule un effectif sur le résultat, etc.

D'autres opérations plus spécifiques seront traitées comme des macro-commandes, pour répondre à des problèmes bien précis. Ces modules seront alors développés au-dessus du système de gestion, en se servant de ses possibilités mais sans s'inscrire dans le processus d'interrogation interactive.

Autre possibilité importante : le croisement avec les images numériques provenant des *satellites d'observation de la terre*. En effet, la gestion de la localisation permet, en faisant coïncider information géographique et données issues de la télédétection, à la fois d'*introduire de nouvelles données* pour des objets existants dans la base de données (comme, sur chaque îlot, la valeur étant calculée uniquement sur les pixels appartenant à l'îlot, une moyenne de radiométrie, un indice de densité de végétation, un écart-type local, ou plus généralement la valeur d'une classification de l'image), et également de développer de *nouvelles techniques de traitement en télédétection* (par zone plutôt que par pixel : l'élément étudié n'est plus seulement le pixel ou le groupe de pixels, mais l'ensemble des pixels correspondant au sol à une réalité physique ou thématique connue et corrélée, si possible, à la donnée recherchée sur l'image).

Pour de telles opérations, le recalage de l'imagerie satellitaire sur une projection géographique connue du système est nécessaire. Les opérations sont ensuite rendues simples par la coexistence, dans le système TIGRE, des structures vectorielle et matricielle pour la localisation.

4. LES RESTITUTIONS

Le troisième aspect du système que nous développons concerne bien sûr l'aspect *cartographie* et représentation graphique.

Notons tout d'abord que nous avons choisi de conserver la localisation en coordonnées géographiques globales (longitude, latitude), ce qui nous donne un référentiel commun, quelle que soit l'origine de la donnée géographique (échelle, projection). C'est ce référentiel commun qui permet de comparer la localisation de toutes les données géographiques d'une même base, et ce faisant de *choisir la projection géographique* servant également à la cartographie des résultats obtenus.

La représentation graphique se fait d'abord sur *écran couleur* de haute définition (1100*900 pixels). La capacité d'affichage (256 couleurs sur 16 millions de possibilités) permet à l'utilisateur de composer son image comme il le souhaite. Le système lui offre la possibilité de visualiser les résultats de ses requêtes : images thématiques, données ponctuelles associées à des symboles, contours, réseaux, noms et valeurs. Il a le choix des couleurs et des trames : il peut créer, modifier des palettes, construire sa légende, rajouter des noms et des symboles. Toutes ces opérations sont interactives et rendues très simples par l'utilisation d'une souris.

Toujours directement sur l'écran, le système offre à l'utilisateur des possibilités de *calculs métriques* : distance entre deux points choisis sur l'écran, le long d'une ligne brisée, surface d'un polygone défini point par point. Ainsi par exemple, ayant défini une fenêtre d'étude et fait tracer le contour des îlots dans cette fenêtre, un utilisateur peut savoir immédiatement la largeur d'une voie, la distance séparant tel îlot de tel autre, en passant par la voirie, le long d'un réseau, à vol d'oiseau, etc.. Le système, en fonction de la taille de la fenêtre et de la projection utilisée, donne tous ces résultats en mètres.

Egalement sur l'écran, on peut y *définir directement des objets*, comme des zones thématiques, des données ponctuelles, des réseaux, et ce, soit pour créer un *masque de sélection géographique* (par exemple : on trace une ligne le long d'une portion de voie, puis on demande la zone se trouvant à moins de 10 mètres de cette ligne pour pouvoir ensuite sélectionner les activités se trouvant dans cette zone), soit pour *effectuer des simulations* (ainsi : sur une image thématique représentant par exemple le prix du mètre carré par îlot, image créée à partir des données de la base et visualisée sur l'écran, on y définit directement une zonification que l'on veut tester en vue d'améliorer le découpage des taux d'imposition. Pour chaque îlot, on crée une nouvelle variable indiquant l'appartenance de l'îlot à telle ou telle zone, ce qui permet ensuite, par une opération d'agrégation de variables par îlot sur les zones, de calculer pour chaque zone un indice à partir des données dont on dispose sur les îlots, comme la densité de population, la densité du bâti, le revenu moyen des habitants, etc.).

La sortie cartographique sur papier se fait grâce à un traceur à plume ou un système de recopie d'écran. Les possibilités graphiques d'un traceur sont très différentes de celles d'un écran : meilleure définition graphique, grande taille et choix possible de l'échelle de sortie, mais peu de choix de couleurs ou de trames

efficaces. Il est souvent long et difficile d'obtenir un document d'une bonne qualité graphique, et plusieurs essais sont généralement nécessaires. Pour les cartes de travail, un bon compromis est offert par la recopie d'écran, noir et blanc ou couleur, qui, sur une taille réduite, donne de bons résultats graphiques et une facilité d'utilisation incontestable.

5. CONSTITUTION DE LA BASE

Gestion, traitements, représentation des résultats, pour cela il faut bien sûr commencer par constituer la base de données sur la ville de Quito : c'est également un aspect important de notre travail. Les données doivent être recensées, évaluées (intérêt, coût, volume, fiabilité), avant d'être saisies et intégrées dans la base de données.

L'information géographique pose le problème de la *saisie de la localisation*. Nous avons donc développé un programme spécifique de saisie graphique, MYGALE/PC. Conformément au principe de la gestion relationnelle, chaque thème est saisi séparément, même si plusieurs thèmes sont présents sur une même carte.

La saisie graphique avec MYGALE/PC se fait sur une *table à digitaliser*, munie d'une loupe de grande précision (la saisie se fait à la précision de 0,1 mm) et reliée à un micro-ordinateur. Le programme offre une saisie graphique supervisée (contrôle de la fermeture des zones, de la cohérence graphique) ainsi que de nombreuses fonctions d'édition et de correction graphique. Il permet de saisir des zones, des lignes, des points, ainsi que des éléments non thématiques pour constituer des fonds cartographiques.

Si le domaine à saisir est grand, il peut être *découpé en feuilles* qui seront digitalisées séparément. C'est l'intégration des données graphiques dans la base et le recalage en coordonnées géographiques qui assurera le bon placement de chaque feuille saisie dans l'ensemble géographique.

La base de données sur Quito se compose actuellement des données suivantes :

- données d'occupation du sol par îlot. Le découpage des îlots propre aux données de la Mairie de Quito a été saisi : 6500 îlots, saisis en 110 feuilles distinctes à l'échelle 1:2 000. La saisie a été effectuée d'avril à septembre 1989. Après formation des trois opérateurs, l'ensemble des îlots a été re-dessiné sur support indéformable puis digitalisé, le temps de saisie sur la table étant légèrement inférieur au temps de dessin. Chaque feuille a ensuite été vérifiée (clés, digitalisation des arcs), puis corrigée si nécessaire (durée deux mois) ;

- données de population et logement du recensement de 1982. La saisie des 6200 îlots du recensement de 1982 de l'INEC (Institut national équatorien de statistique) a été effectuée par modification du fond municipal, dessin des modifications d'après les plans de l'INEC, digitalisation de ces modifications et changement des clés, vérification et correction (décembre-avril 1989). Disposant des données par individu, chaque variable a été sommée par îlot pour un certain nombre de ses modalités, le résultat devenant un attribut sur l'îlot ;

- données sur les activités en rez-de-chaussée, à partir du recensement effectué en 1987. Chaque côté d'îlot a été répertorié et un point saisi pour localiser les activités s'y trouvant (35 000 points localisés sur les plans au 1:2 000 et saisis, mai-juillet 1989) ;

- équipements de santé, par point ;
- établissements d'enseignement, par point ;
- banques, entreprises, hôtels, loisirs ;
- géologie et usage du sol sur l'aire métropolitaine (avril 1989) ;
- divers autres découpages de la ville : secteurs, quartiers (juin 1989)

permettant, après des opérations d'agrégation (par exemple la moyenne par secteur de la densité de population de chaque îlot du secteur), de représenter un phénomène dans divers découpages d'échelles différentes ;

- est en cours de saisie, l'ensemble des points cotés, donnant l'altitude dans la ville (à peu près un point par croisement de rues), ainsi que les courbes de niveaux à 10 mètres sur l'aire métropolitaine ;

- le dessin du réseau d'égout est en cours de préparation, ainsi que l'élaboration des données par tronçon. Idem pour le réseau d'adduction d'eau.

L'ensemble SAVANE (MYGALE, TIGRE, ainsi que les différents modules d'introduction et de modification de données) constitue dès lors un système d'information qui devra nous permettre de répondre aux objectifs scientifiques du programme de recherche sur Quito. Les développements informatiques à venir iront dans trois directions :

- *faciliter la gestion urbaine* en produisant information et simulation, en fonction des demandes émanant essentiellement de la Mairie de Quito ;
- *intégrer et gérer de nouvelles sources de données*, comme les photographies aériennes, en utilisant si besoin est de nouveaux moyens de stockage comme les CD-ROM ;
- *rechercher des méthodes d'actualisation* pour les données graphiques.

6. LA CONFIGURATION MATERIELLE

Fonctionnalités des logiciels et possibilités des matériels sont étroitement liées. Pour assurer leur cohérence, le programme nécessite une configuration informatique qui doit répondre à plusieurs critères :

- posséder suffisamment de *puissance* pour gérer un important volume de données, traiter la localisation, manipuler des images ;
- avoir un *environnement graphique* performant (écran couleur de haute résolution, périphériques graphiques classiques) ;
- décentraliser la saisie graphique sur des postes autonomes.

De plus, l'équipement doit être d'un coût total permettant son investissement rapide, aussi bien pour l'Orstom que pour nos partenaires.

Face à ces contraintes, nous avons composé une configuration informatique qui se compose des matériels suivants :

- une *station de travail* MATRA-SUN 3/110C, munie de 4 Mo de RAM, d'un disque dur de 140 Mo, d'un écran graphique 1100*900 pouvant afficher 256 couleurs sur une palette de 16 millions, d'une console alphanumérique, d'une imprimante. La puissance du processeur (2 Mips) , le système UNIX, et le concept de station de travail en font une machine très bien adaptée aux fonctionnalités requises pour la gestion des données et leur manipulation graphique sur écran. Seule la capacité du disque est insuffisante ;

- un *micro-ordinateur compatible AT* muni d'une carte EGA, d'une imprimante, et reliée à un *digitaliseur* BENSON 6301 (surface utile 1m20*0m80) pour la saisie graphique. Le micro est relié à la station de travail par une voie asynchrone pour le transfert des fichiers de saisie ;

- un *traceur à plume* BENSON 1333 ;

- un *système de recopie d'écran* Teta-scan à transfert thermique, relié à la sortie vidéo de la station de travail.

L'ensemble de cette configuration, hormis la recopie d'écran et la console alphanumérique, a été mis en place début mars 1988, l'installation électrique ayant été fournie par l'IGM. Jusqu'à présent, ce matériel nous a donné entière satisfaction, à l'exception d'une panne du digitaliseur, machine pourtant des plus fiables; sa remise en marche n'a pas nécessité d'intervention technique spécialisée. Il va sans dire en effet que le manque de maintenance sur place constitue le principal problème technique auquel nous pouvons être confronté, et qu'aucune réponse satisfaisante ne peut actuellement lui être apportée.

EXEMPLE DE MANIPULATION DE DONNEES INTEGREES A UN SYSTEME D'INFORMATION GEOGRAPHIQUE

(Dominique COURET)

RESUME - La présente communication relate une expérience d'exploitation d'un système d'information géographique, pour la réalisation de cartes thématiques à partir des données qui y sont intégrées. Elle s'appuie sur les expériences faites, dans le cadre du projet d'Atlas informatisé de Quito, avec les données du recensement national de population et logement de la zone "del Rosario" et du "Comité del Pueblo", et en utilisant les moyens disponibles du système "Savane".

L'objectif est de construire une typologie de l'espace urbain en fonction des caractéristiques de l'habitat, en utilisant une démarche initialement basée sur la recherche d'une organisation naturelle des variables disponibles, détectée par une analyse multivariée préalable.

Les deux applications présentées montrent que la complexité augmente rapidement avec le nombre de variables (et de leurs modalités) pris en compte dans l'analyse. Elles mettent également en évidence la nécessité d'un travail alterné entre la représentation cartographique et la simplification du fichier descriptif, représentatif des modalités à visualiser.

1. INTRODUCTION

1.1. Présentation du contexte

Dans le cadre du projet "Atlas informatisé de Quito", il s'agit d'établir comment pourraient être utilisées les données exhaustives fournies par le recensement national de population et logement, afin de servir et aider le travail de la Municipalité, premier utilisateur de la base d'information urbaine.

Dans cette optique et pour un géographe, cela signifie, entre autre, établir une méthode de traitement de ces variables qui permette tout à la fois de réaliser une typologie de l'espace urbain en vastes zones ou strates (échelle du quartier ou du groupe de quartiers) et de retourner ensuite à des échelles plus fines (l'îlot par exemple). Dans un autre sens il s'agit aussi d'établir des postes thématiques généraux et de pouvoir accéder facilement aux différents descriptifs initiaux.

Dans un premier temps, il a été choisi de se limiter à l'étude des données propres au logement et de construire une typologie de l'espace urbain en fonction des caractéristiques de l'habitat.

La première phase du travail a consisté à évaluer comment les outils de manipulation proposés par le système permettaient la mise en oeuvre d'une telle démarche. Ce sont les réflexions attachées à cette première étape que l'on expose ici.

1.2. Abordage du système d'information géographique (SIG)

Les données du recensement de 1982 portant sur le logement ont été intégrées au sein de la base comme variables du plus petit référentiel spatial disponible, c'est-à-dire l'îlot. La méthode de recodage adoptée est la sommation par îlot des valeurs absolues des différentes modalités des variables du logement. De 13 variables au niveau du logement on aboutit à 72 variables au niveau de l'îlot. Face à un tel volume d'information, l'utilisateur peut opter pour différents cheminement.

Une première attitude est d'avoir décidé des différents indicateurs du phénomène que l'on cherche à mettre en évidence, et d'utiliser les possibilités du système pour juger de la répartition spatiale de ce phénomène au travers du comportement géographique d'un de ses indicateurs.

Une seconde attitude possible est de mettre en valeur une organisation "spontanée" des variables disponibles et de chercher à définir le ou les phénomènes qu'elle met en évidence. C'est dans le sens de cette seconde démarche que nous avons abordé l'utilisation du SIG.

2. METHODE D'ORGANISATION DU TRAITEMENT

Le système d'information géographique ne comporte pas de module permettant la réalisation d'analyses de données multivariées. Pour décider de l'ordonnement du traitement conjoint des variables et d'une partition pour chaque variable, une analyse factorielle des correspondances puis une classification ascendante hiérarchique ont été faites à l'extérieur du système. Elles ont donné comme résultats la définition de trois types généraux d'îlots, une partition pour chaque variable et ont conduit à la définition des premières manipulations.

2.1. La définition de trois types généraux d'îlots

Ils correspondent aux regroupements des variables ici présentées dans l'ordre des contributions à la formation des trois axes.

Premier type : "bâti récent et équipement public"

- statut d'occupation "propriété",
- approvisionnement en eau "réseau public branchement individuel",
- équipement en sanitaires "WC particuliers",
- évacuation des eaux usées "réseau public des égouts",
- matériau sol "plancher, vinil, carrelage ou similaires",
- matériau toit "Eternit, Ardex ou similaire",
- type d'habitat "appartement",
- nombre de chambres "3 ou 4 à 6 chambres",
- matériau toit "dalle de béton (toit en terrasse)",
- matériau murs "béton, briques ou parpaings".

Deuxième type: "bâti ancien transformé"

- type d'habitat "location en Residencial (3)",
- matériau toit "tuiles",
- approvisionnement en eau "réseau public branchement collectif",
- statut d'occupation "location",
- matériau sol "parquet traditionnel",
- nombre de chambre " 1 chambre",
- équipement en sanitaires "WC collectifs",
- matériau murs "Adobe ou Tapia" (4),
- nombre de chambres "0 chambre",
- population " 2 personnes",
- matériau toit "matériaux multiples",
- activité pratiquée dans le logement "artisanat",
- statut d'occupation "gratuit",
- population " 1 personne".

Troisième type : "sous-équipement"

- type d'habitat "Mediagua" (5),
- équipement en sanitaires "sans sanitaires",
- évacuation des eaux usées "sans système",
- approvisionnement en eau "puits ou fontaine",
- approvisionnement en eau "réseau public branchement hors édifice",
- approvisionnement en eau "camion citerne",
- évacuation des eaux usées "fosse septique",
- équipement en sanitaires "latrines",
- matériau sol "terre",
- matériau toit "zinc ou similaires",
- matériau sol "brique ou ciment",
- approvisionnement en électricité "sans",
- approvisionnement en eau "ruisseau ou rivière".

2.2. Une partition pour chaque variable en quatre classes

Ces classes permettent de positionner les îlots par rapport à la répartition d'ensemble de la variable. Il s'agit de classes de pourcentages de logements par îlots, généralement 0%, 1 à 33%, 34 à 66% et 67 à 100%, qui définissent les différentes situations possibles par rapport à l'îlot moyen : absence, proportion inférieure, proportion proche, proportion supérieure.

(3) Residencial : maison de location où l'on peut disposer au moins d'une habitation allant de la pièce simple à l'appartement.

(4) Adobe ou Tapia: torchis ou pisé

(5) Mediagua: Habitat d'un seul niveau et ayant un toit composé d'un seul pan.

2.3. La définition des premières manipulations

Les résultats de l'analyse multivariée permettent de dégager l'importance des variables "d'équipement du logement" dans la discrimination de groupes d'îlots. Elles interviennent principalement dans la définition des groupes "bâti récent et équipement public" et "sous-équipement". Un traitement des variables "équipements du logement" a donc été décidé pour mettre en évidence ces deux types.

3. PREMIERE APPLICATION : L'APPORT DE LA REPARTITION SPATIALE DANS LA DEFINITION DE CLASSES PERTINENTES

Type mis en valeur : "bâti récent et équipement public".

Variables utilisées :

- approvisionnement en eau "réseau public branchement individuel" ;
- évacuation des eaux usées "réseau public des égouts".

Pour réaliser le traitement conjoint de ces deux variables, à partir des outils du système, il a été procédé en trois étapes.

1° création de deux nouvelles variables :

- % de logements de l'îlot approvisionnés en eau par "réseau public branchement individuel" ;
- % de logements de l'îlot avec évacuation des eaux usées par "réseau public des égouts".

2° discrétisation de ces deux nouvelles variables en 4 classes.

3° représentation conjointe des deux variables sur une même image.

Sont affectés à la même combinaison l'ensemble des îlots qui correspondent à la présence locale de deux modalités particulières de l'une et l'autre variable. La description de ce croisement est donnée par le fichier descriptif (figure 1) fourni par le système conjointement à l'image. Ce sont ces différentes combinaisons, d'une variable par rapport à l'autre, qu'il décrit (des strates) et non pas les îlots et leurs valeurs.

L'organisation du fichier descriptif et sa traduction thématique permettent de réaliser une légende et une carte (carte 1). Cette visualisation simultanée des deux variables montre le poids de la localisation dans leur répartition. Elles définissent une zone homogène. Le "phénomène "équipement public" que décrivent ces deux variables est de type zonal. Ceci est logique en raison de la nature même de ces deux équipements (réseaux centralisés). la visualisation spatiale de ces variables permet de définir une limite d'extension de ces réseaux.

Par ailleurs, la sortie rapide des résultats d'une classification et sa répétition avec modification permettent de réfléchir et de décider du regroupement de classes en fonction de leur proximité ou éloignement géographique.

FIGURE 1

Relation : EQUIPEMENT LOG

FICHER DESCRIPTIF ORIGINEL

surf./surf. tot.	%EAU1	%EUS1
05.37 %	1 A 33 %	1 A 33 %
59.93 %	0 %	0 %
01.47 %	1 A 33 %	0 %
09.43 %	0 %	1 A 33 %
17.27 %	67 A 100 %	67 A 100 %
02.06 %	34 A 66 %	34 A 66 %
03.91 %	34 A 66 %	67 A 100 %
00.57 %	1 A 33 %	34 A 66 %

%EAU1 : % DE LOGEMENTS DE L'ÎLOT DISPOSANT UN BRANCHEMENT
DIRECT AU RESEAU PUBLIC D'EAU POTABLE

%EUS1 : % DE LOGEMENTS DE L'ÎLOT DISPOSANT D'UN ACCES AU
RESEAU D'EGOUTS MUNICIPAL

FICHER DESCRIPTIF ORDONNE CORRESPONDANT A LA CARTE 1

surf./surf. tot.	%EAU1	%EUS1	THEMATIQUE SOMMAIRE
24.70 %	67 A 100 %	67 A 100 %	PRESENCE CONJOINTE EGALE DES DEUX EQUIPEMENTS
17.27 %	34 A 66 %	34 A 66 %	67 A 100 %
02.06 %	34 A 66 %	34 A 66 %	34 A 67 %
05.37 %	1 A 33 %	1 A 33 %	1 A 33 %
04.48 %			DOMINANCE RACCORDEMENT AU RESEAU D'EGOUT
03.91 %	34 A 66 %	67 A 100 %	
00.57 %	1 A 33 %	34 A 66 %	
01.47 %	1 A 33 %	0 %	PRESENCE DISJOINTE
09.43 %	0 %	1 A 33 %	PRESENCE SEULE RACCORDEMENT AU RESEAU D'EAU
59.93 %	0 %	0 %	PRESENCE SEULE RACCORDEMENT AU RESEAU D'EGOUT
			ABSENCE DES DEUX EQUIPEMENTS

FICHER DESCRIPTIF ORDONNE CORRESPONDANT A LA CARTE 2

surf./surf. tot.	%EAU1	%EUS1	THEMATIQUE SOMMAIRE
29.18 %			PRESENCE CONJOINTE
17.27 %	67 A 100 %	67 A 100 %	PRESENCE CONJOINTE EGALE DES DEUX EQUIPEMENTS
03.91 %	34 A 66 %	67 A 100 %	DOMINANCE RACCORDEMENT AU RESEAU D'EGOUT
02.06 %	34 A 66 %	34 A 66 %	PRESENCE CONJOINTE EGALE DES DEUX EQUIPEMENTS
00.57 %	1 A 33 %	34 A 66 %	DOMINANCE RACCORDEMENT AU RESEAU D'EGOUT
05.37 %	1 A 33 %	1 A 33 %	PRESENCE CONJOINTE EGALE DES DEUX EQUIPEMENTS
70.83 %			ABSENCE OU PRESENCE DISJOINTE
01.47 %	1 A 33 %	0 %	PRESENCE SEULE RACCORDEMENT AU RESEAU D'EAU
09.43 %	0 %	1 A 33 %	PRESENCE SEULE RACCORDEMENT AU RESEAU D'EGOUT
59.93 %	0 %	0 %	ABSENCE DES DEUX EQUIPEMENTS

CARTE 1

CARACTERISTIQUES DU LOGEMENT
SECTEURS ROSARIO/COMITE DEL PUEBLO

500 M



EQUIPEMENT

- APPROVISIONNEMENT EN EAU
- EVACUATION DES EAUX USEES

% DE LOGEMENTS PAR ILOT

APPROVISIONNEMENT EN EAU
RESEAU PUBLIC / BRANCHEMENT INDIVIDUEL

EVACUATION DES EAUX USEES
RESEAU PUBLIC DES EGOUTS

SOURCE: "CENSO DE POBLACION Y VIVIENDA" INEC 1982
AIRE URBAINE DE LA PAROISSE DE QUITO (EQUATEUR)

PRESENCE CONJOINTE EGALE



67 A 100 %

34 A 66 %

1 A 33 %

DOMINANCE RACCORDEMENT AUX EGOUTS



34 A 66% EAU RP / 67 A 100% EGGUT RP

1 A 33 % EAU RP / 34 A 66% EGGUT RP

PRESENCE DISJOINTE



RACCORDEMENT AU RESEAU D'EAU

1 A 33 %



RACCORDEMENT AU RESEAU D'EGGUT

1 A 33 %



ABSENCE DES DEUX EQUIPEMENTS

0 %

Ainsi les catégories "présence disjointe" des deux équipements sont à la fois bien à l'extérieur de la zone principale qui se dégage sur l'image, et se caractérisent par une proportion inférieure de logements équipés par rapport à l'îlot moyen. Cette double analyse amène l'utilisateur à les considérer hors de la zone équipée.

De même les catégories "dominance raccordement aux égouts" apparaissent former sur l'image un bloc homogène avec les catégories "présence égale" des deux équipements. Cette localisation permet d'une part de créer un seul regroupement "présence conjointe" des deux équipements, et d'autre part de mettre en valeur un rapport particulier entre les deux équipements: *en zone équipée l'accès aux égouts est toujours égal ou supérieur à l'accès au réseau d'eau potable.*

La nouvelle carte (carte 2) visualise la zone "équipée" qui se définit comme la zone de présence conjointe des deux variables. Les catégories retenues ne respectent plus la discrétisation antérieure interne à chaque variable en classes disjointes. Ceci est rendu possible par le mode de croisement utilisé qui considère comme entités les combinaisons locales ou strates et non pas les îlots par rapport aux variables.

Le fichier descriptif qui permet de réaliser la définition des classes présente deux utilités annexes.

Il donne les combinaisons réelles par rapport aux combinaisons théoriquement possibles. Ceci peut être un indicateur de la forte ou faible corrélation spatiale (plus le nombre de combinaisons est important plus le comportement d'une variable par rapport à l'autre est diversifié). Dans le cas présent le nombre de combinaisons théoriques est de vingt deux, le nombre de combinaisons réelles de huit.

Il donne le pourcentage de la surface de l'image occupée par chaque combinaison. Cela peut servir à juger de l'importance de telle ou telle combinaison, notamment quand les îlots lui correspondant sont fortement dispersés et de petite taille ou ne forment pas une plage homogène facile à évaluer pour l'oeil.

Cependant ce premier exemple est un cas très simple de manipulation sous le système d'information géographique, nous n'avons là que deux variables qui spatialement se révèlent avoir un comportement proche et dont le croisement ne donne jour qu'à huit combinaisons.

4. SECONDE APPLICATION : LA DISPONIBILITE CONSTANTE DE L'INFORMATION EXACTE, UNE CONTRAINTE OU UN APPORT ?

Type mis en valeur : "sous-équipement".

Variables utilisées :

- approvisionnement en eau

"puits ou fontaine"

"réseau public hors édifice"

CARTE 2

CARACTERISTIQUES DU LOGEMENT
SECTEURS ROSARIO/COMITE DEL PUEBLO

500 M



EQUIPEMENT

- APPROVISIONNEMENT EN EAU
- EVACUATION DES EAUX USEES

% DE LOGEMENTS PAR ILOT

APPROVISIONNEMENT EN EAU
RESEAU PUBLIC / BRANCHEMENT INDIVIDUEL

EVACUATION DES EAUX USEES
RESEAU PUBLIC DES EGOUTS

SOURCE: "CENSO DE POBLACION Y VIVIENDA" INEC 1982
AIRE URBAINE DE LA PAROISSE DE QUITO (EQUATEUR)

PRESENCE CONJOINTE



34 A 100% EAU RP / 67 A 100 % EGOUT RP

1 A 66 % EAU RP / 34 A 66 % EGOUT RP

1 A 33 % EAU RP / 1 A 33 % EGOUT RP

ABSENCE OU PRESENCE DISJOINTE

DES DEUX EQUIPEMENTS



- "camion citerne"
- "ruisseau ou rivière"
- évacuation des eaux usées
 - "sans système"
 - "fosse septique"
- équipements en sanitaires
 - "latrines"
 - "sans sanitaires"

Ce sont donc huit variables qu'il faut ici traiter conjointement. Pour construire la réflexion, il a d'abord été choisi de travailler par équipement (eau, évacuation des eaux usées, équipement en sanitaires). Le processus de traitement est le même que celui exposé pour la première application.

L'analyse spatiale et thématique a abouti à hiérarchiser entre elles les variables descriptives d'un même équipement, à conserver certaines combinaisons et à en éliminer d'autres, jugées non représentatives tant du point de vue des valeurs de modalités auxquelles elles correspondaient, que du point de vue de leur localisation spatiale.

- carte 3 : Représentation simultanée des variables d'évacuation des eaux usées type "sous-équipement", résultat final ;
- carte 4 : Représentation simultanée des variables d'approvisionnement en eau type "sous-équipement", résultat final ;
- carte 5 : Représentation simultanée des variables d'équipement en sanitaires type "sous-équipement", résultat final.

On constate que la zone globalement déterminée varie peu d'une carte à l'autre : plus homogène pour "eau", plus étendue pour "sanitaires", moins étendue mais dans les limites déterminées par "eau" pour "évacuation des eaux usées".

Ensuite, pour aboutir à une vision synthétique des modalités d'équipement type "sous-équipement", une représentation de l'espace commun aux trois répartitions retenues a été demandée. L'organisation du fichier descriptif (figure 2) permet de définir les différentes variations de la présence commune des trois sous-équipements en "approvisionnement en eau", "évacuation des eaux usées" et "équipement en sanitaires", et donne une répartition en cinq postes, allant de la déficience en équipements à la présence d'équipements de substitution.

- carte 6 : Représentation de la présence conjointe des modalités retenues comme descriptives du type "sous-équipement" en "approvisionnement en eau", "évacuation des eaux usées" et "équipement en sanitaires".

La définition des postes "déficience en équipement" ne signifie pas forcément absence de logements bénéficiant d'équipements de substitution. Ainsi au sein du poste "déficience en équipements 1", sont compris des îlots où il y a 1 à 33 % de logements équipés de latrines. De même l'approvisionnement "eau par camion citerne" est fortement présent dans tous les postes "déficience". La présentation de la légende et de la carte oblige à une simplification. Cependant le détail et la nature exacte des combinaisons sont disponibles sous la forme du fichier descriptif (figure 2). *Il permet ainsi de passer rapidement des postes généraux de la typologie aux combinaisons unitaires réelles et détaillées.*

CARTE 3

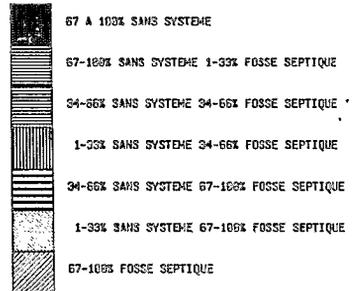
500 M



CARACTERISTIQUES DU LOGEMENT
SECTEURS ROSARIO/COMITE DEL PUEBLO

TYPE "SOUS EQUIPEMENT"
.EVACUATION DES EAUX USEES

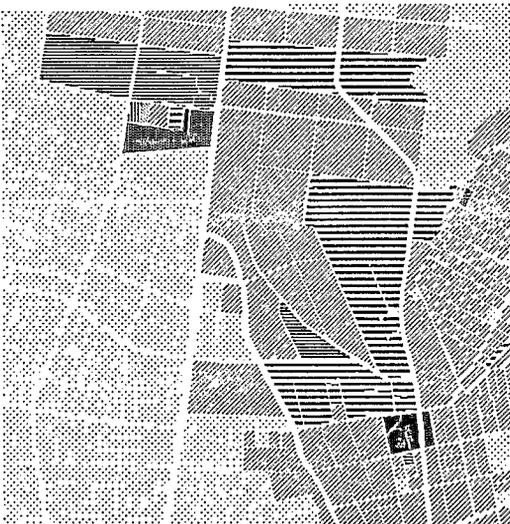
% DE LOGEMENTS DE L' ILOT



SOURCE: "CENSO DE POBLACION Y VIVIENDA" INEC 1982
AIRE URBAINE DE LA PARRISSE DE QUITO (EQUATEUR)

CARTE 4

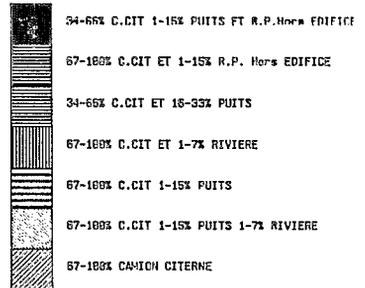
500 M



CARACTERISTIQUES DU LOGEMENT
SECTEURS ROSARIO/COMITE DEL PUEBLO

TYPE "SOUS EQUIPEMENT"
.APPROVISIONNEMENT EN EAU

% DE LOGEMENTS DE L' ILOT

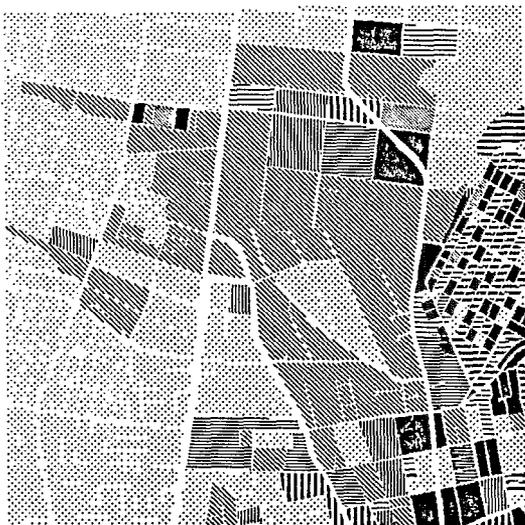


SOURCE: "CENSO DE POBLACION Y VIVIENDA" INEC 1982
AIRE URBAINE DE LA PARRISSE DE QUITO (EQUATEUR)

CARTE 5

CARACTERISTIQUES DU LOGEMENT
SECTEURS ROSARIO/COMITE DEL PUEBLO

500 M



TYPE "SOUS EQUIPEMENT"

.SANITAIRES

% DE LOGEMENTS DE L' ILOT

LATRINES

SANS SANITAIRES

SOURCE: "CENSO DE POBLACION Y VIVIENDA" INEC 1992
AIRE URBAINE DE LA PARISSIE DE QUITO (EQUATEUR)

DOMINANCE SANS SANITAIRES



1-33% LATRINES 67-100% SANS SANITAIRES

1-33% LATRINES 34-66% SANS SANITAIRES

DOMINANCE LATRINES



67-100% LATRINES 1-33% SANS SANITAIRES

34-66% LATRINES 1-33% SANS SANITAIRES

PRESENCE EGALE



34-66%

1-33%

PRESENCE DISJOINTE



67-100% SANS SANITAIRES

34-66% SANS SANITAIRES



67-100% LATRINES

34-66% LATRINES

FIGURE 2

Relation : EQUIPEMENT LOG

suif.anc. tot.	EAU4-4-1	AU02-SHY3	SCH1-SHY4	EAU	EAU USSES	SAINTIER3
59.80 %	PRESENCE DE LA CLASSE "MORS CONDANLSON"			MORS DU TYPE "SUSUS-EQUIPEMENT"		
00.62 %	67-100 0 0 0	1-33 67-100	67-100 0	DEFICIENCE EN EQUIPEMENT 1		
00.90 %	67-100 0 0 0	1-33 67-100	34-66 0	3/4 C.CIT	1/4 F 3/4 SANSI	3/4 LATRINES
00.20 %	67-100 0 0 0	1-33 67-100	67-100 1-33	3/4 C.CIT	1/4 F 3/4 SANSI	2/3 LATRINES
08.23 %	67-100 0 0 0	1-33 67-100	34-66 1-33	3/4 C.CIT	1/4 F 3/4 SANSI	2/3 L 1/3 SANSI
00.07 %	67-100 0 0 0	1-33 67-100	34-66 34-66	3/4 C.CIT	1/4 F 3/4 SANSI	1/2 L 1/2 SANSI
00.03 %	67-100 0 0 0	1-33 67-100	1-33 1-33	3/4 C.CIT	1/4 F 3/4 SANSI	1/3 L 1/3 SANSI
00.87 %	67-100 1-15 0 0	1-33 67-100	67-100 1-33	3/4 C.CIT PUIIIS	1/4 F 3/4 SANSI	3/4 L 1/4 SANSI
02.16 %	67-100 0 0 0	1-33 67-100	1-33 34-66	DEFICIENCE EN EQUIPEMENT 2		
00.32 %	67-100 0 0 0	1-33 67-100	1-33 67-100	3/4 C.CIT	1/4 F 3/4 SANSI	1/3 L 2/3 SANSI
01.20 %	67-100 0 0 0	1-33 67-100	0 34-66	3/4 C.CIT	1/4 F 3/4 SANSI	2/3 SANSI
00.75 %	67-100 0 0 0	34-66 34-66	1-33 34-66	3/4 C.CIT	1/2 F 1/2 SANSI	1/3 L 2/3 SANSI
00.87 %	67-100 0 0 0	34-66 34-66	0 34-66	3/4 C.CIT	1/2 F 1/2 SANSI	2/3 SANSI
00.37 %	67-100 0 0 0	67-100 1-33	67-100 0	EQUIPEMENT DE SUBSTITUTION 1		
00.14 %	67-100 0 0 0	67-100 1-33	34-66 0	3/4 C.CIT	3/4 F 1/4 SANSI	2/3 LATRINES
00.04 %	67-100 0 0 0	67-100 34-66	67-100 0	3/4 C.CIT	2/3 F 1/3 SANSI	3/4 LATRINES
00.72 %	67-100 0 0 0	67-100 1-33	67-100 1-33	3/4 C.CIT	3/4 F 1/4 SANSI	3/4 L 1/4 SANSI
00.52 %	67-100 0 0 0	67-100 1-33	34-66 1-33	3/4 C.CIT	3/4 F 1/4 SANSI	2/3 L 1/3 SANSI
00.45 %	34-66 1-15 0 1-15	67-100 1-33	34-66 1-33	2/3 C.C. PU RPH	3/4 F 1/4 SANSI	2/3 L 1/3 SANSI
02.27 %	67-100 0 0 0	67-100 1-33	1-33 1-33	EQUIPEMENT DE SUBSTITUTION 2		
00.86 %	67-100 0 0 0	67-100 34-66	1-33 34-66	3/4 C.CIT	3/4 F 1/4 SANSI	1/3 L 1/3 SANSI
00.84 %	67-100 0 0 0	67-100 1-33	34-66 34-66	3/4 C.CIT	3/4 F 1/4 SANSI	1/2 L 1/2 SANSI
00.56 %	67-100 0 0 0	34-66 34-66	67-100 0	3/4 C.CIT	1/2 F 1/2 SANSI	3/4 LATRINES
00.19 %	67-100 0 0 0	34-66 34-66	67-100 1-33	3/4 C.CIT	1/2 F 1/2 SANSI	3/4 L 1/4 SANSI
00.38 %	67-100 0 0 0	34-66 34-66	34-66 0	3/4 C.CIT	1/2 F 1/2 SANSI	2/3 LATRINES
02.80 %	67-100 0 0 0	34-66 34-66	34-66 1-33	3/4 C.CIT	1/2 F 1/2 SANSI	2/3 L 1/3 SANSI
00.69 %	67-100 0 0 0	34-66 34-66	34-66 34-66	3/4 C.CIT	1/2 F 1/2 SANSI	1/2 L 1/2 SANSI
00.98 %	67-100 0 0 0	34-66 1-33	1-33 1-33	3/4 C.CIT	2/3 F 1/3 SANSI	1/3 L 1/3 SANSI
05.52 %	67-100 0 0 0	34-66 34-66	1-33 1-33	3/4 C.CIT	1/2 F 1/2 SANSI	1/2 L 1/3 SANSI
00.67 %	67-100 1-15 1-7 0	67-100 1-33	1-33 1-33	3/4 C.C. PUI RPH	3/4 F 1/4 SANSI	1/2 L 1/3 SANSI
12.72 %	67-100 1-15 0 0	67-100 1-33	1-33 1-33	3/4 C.CIT PUIIIS	3/4 F 1/4 SANSI	1/3 L 1/3 SANSI
00.12 %	67-100 1-15 0 0	34-66 34-66	67-100 1-33	3/4 C.CIT PUIIIS	1/2 F 1/2 SANSI	3/4 L 1/4 SANSI
00.63 %	67-100 0 0 1-15	67-100 1-33	1-33 1-33	EQUIPEMENT DE SUBSTITUTION 3		
00.26 %	67-100 0 0 1-15	34-66 34-66	34-66 34-66	14/5 C.C. 1/5 RPH	3/4 F 1/4 SANSI	1/3 L 1/3 SANSI

- EAU4-4-5-3 : VARIABLES "EAU" DESCRIPTIVES DU SOUS-EQUIPEMENT
 - % DE LOGEMENTS DE L'ILET
 - EAU1: CANION CITERNE EAU4: PUIS OU FONTAINE EAU5: RIVIERE ou ROUSSEAU EAU6: RESEAU PUBLIC MORS EDIFICE
- EUS1-EUS2 : VARIABLES "EVACUATION DES EAUX USEES"
 - % DE LOGEMENTS DE L'ILET
 - EUS1: FOSSE SEPTIQUE EUS2: SANS SYSTEME
- SHY3-SHY4 : VARIABLES "EQUIPEMENT EN SANITAIRES"

CARTE 6

500 M



CARACTERISTIQUES DU LOGEMENT
SECTEURS ROSARIO/COMITE DEL PUEBLO

EQUIPEMENT

- . APPROVISIONNEMENT EN EAU
- . EVACUATION DES EAUX USEES
- . SANITAIRES

% DE LOGEMENTS PAR ILOT

TYPE "SOUS-EQUIPEMENT"

APPROVISIONNEMENT EN EAU	EVACUATION DES EAUX USEES	SANITAIRES
R.P. Home Edifice	FOSSE SEPTIQUE	LATRINES
CAMION CITERNE	SANS SYSTEME	SANS SANITAIRES
FUITS-FONTAINE		
RIVIERE-RUISSEAU		

SOURCE: "CENSO DE POBLACION Y VIVIENDA" INEC 1982
AIRE URBAINE DE LA PAROISSE DE QUITO (EQUATEUR)



"DEFICIENCE EN EQUIPEMENTS 1"

"DEFICIENCE EN EQUIPEMENTS 2"



"EQUIPEMENTS DE SUBSTITUTION 1"

"EQUIPEMENTS DE SUBSTITUTION 2"

"EQUIPEMENTS DE SUBSTITUTION 3"



HORS DU TYPE "SOUS-EQUIPEMENT PUBLIC"

Cette répartition apparaît pourtant peu satisfaisante. Elle ne donne pas une zone très homogène en comparaison aux trois cartes partielles. Elle élimine un certain nombre de cas "intermédiaires". Par exemple sont masqués les îlots avec une proportion importante pour les variables "eau par camion citerne" et "fosse septique", sans la présence des variables relevant du sous-équipement en sanitaires. Il apparaît intéressant, une fois mise en valeur cette première zonification qui représente en quelque sorte la partie "pure" du type "sous-équipement", d'essayer de dégager les formes secondaires.

Il a donc été tenté de réaliser directement la combinaison des huit variables de sous-équipement en une seule fois. Le fichier descriptif devient plus complexe, le nombre de combinaisons atteint le chiffre de 104. Il eut été difficile de réaliser une quelconque organisation de ce fichier descriptif sans la connaissance acquise aux étapes antérieures. Les précédentes manipulations ont donc permis de simplifier l'information dans un premier temps, puis de peu à peu complexifier l'analyse.

Des combinaisons de modalités (carte 7), d'abord évacuées, permettent de mieux définir les limites et les postes du type "sous-équipement". Ainsi le coin sud-ouest de l'image se définit avant tout par la présence en faible proportion d'approvisionnement en eau par "réseau public branchement hors édifice".

Par ailleurs la position des îlots dans l'échelle thématique, allant de la déficience la plus importante à la présence d'équipements publics, a elle-même changée. Par exemple les îlots de la catégorie "Équipements de substitution 2" sur la première carte (carte 6), se trouvent répartis en trois catégories différentes de "Substituts" sur la seconde carte (carte 7).

Enfin on a pu avancer dans la définition du sous-équipement. Il se caractérise par quatre formes principales au niveau de l'îlot :

- association "logements sans équipement et logements bénéficiant d'équipements de substitution" (proportion supérieure de sans équipement) ;
- association "logements bénéficiant d'équipements de substitution et logements sans équipement" (proportion supérieure des logements avec équipement de substitution) ;
- association "logements équipés (réseaux publics) et logements sans équipement" ;
- présence de logements "approvisionnés en eau par réseau public branchement hors édifice".

Ce travail met ainsi en valeur des formes intermédiaires qui modifient l'image préalable ville équipée/ ville non équipée et établit le poids des modes d'équipement de substitution.

On voit bien comment le jeu de manipulation entre représentations spatiales et ordonnancements du fichier descriptif a permis la construction de la définition du type "sous-équipement". On comprend aisément comment il est ensuite possible d'étudier les différentes formes dans leurs diversité interne, étant donné la taille du fichier descriptif ordonné correspondant.

CARTE 7

CARACTERISTIQUES DU LOGEMENT
SECTEURS, ROSARIO/COMITE DEL PUEBLO

500 M



EQUIPEMENT

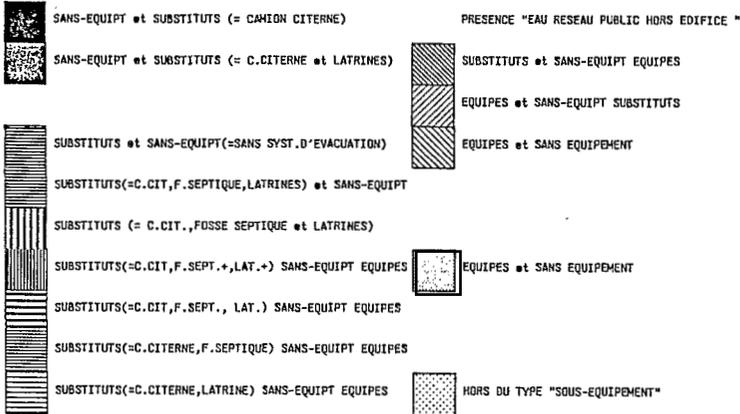
- .APPROVISIONNEMENT EN EAU
- .EVACUATION DES EAUX USEES
- .SANITAIRES

% DE LOGEMENTS PAR ILOT

TYPE "SOUS-EQUIPEMENT"

APPROVISIONNEMENT EN EAU	EVACUATION DES EAUX USEES	SANITAIRES
R.P Hors Edifice CAMION CITERNE PUITS-FONTAINE RIVIERE-RUISSEAU	FOSSE SEPTIQUE SANS SYSTEME	LATRINES SANS SANITAIRES

SOURCE: "CENSO DE POBLACION Y VIVIENDA" INEC 1982
AIRE URBAINA DE LA PAROISSE DE QUITO (EQUATEUR)



Cependant la réorganisation du fichier descriptif est d'une exécution difficile. La méthode d'ordonnement qui consiste à "rapprocher les lignes qui se ressemblent", généralement en s'appuyant sur une ou deux variables jugées principales (ici, camion citerne par exemple), reste une technique peu élaborée. Elle lie trop étroitement analyse des valeurs des variables et interprétation personnelle de l'utilisateur. Pour rendre plus efficace ce jeu carte/fichier descriptif, il est nécessaire d'envisager un outil plus rigoureux. Un traitement intégrant l'usage de la matrice BERTIN comme méthode d'organisation du fichier descriptif est actuellement en cours et donne des résultats satisfaisants. Elle consiste à réaliser une analyse multivariée au même titre que des procédures statistiques d'analyse de données. Elle fait appel à la même technique visuelle d'analyse de l'image que celle utilisée en interprétation cartographique, ce qui facilite le passage fichier descriptif/carte.

EXTRAIRE, VISUALISER ET ARCHIVER UNE CARTE AVEC LEICA

(François DELCLAUX & Claudio LAWIRA)

RESUME - *Au départ* : des fichiers de données cartographiques mondiales (frontières, rivières, lacs, mers, etc.), sans possibilité de les exploiter, ni de les visualiser. *A l'arrivée* : LEICA, logiciel d'extraction interactif pour la cartographie appliquée, logiciel assurant les trois fonctions d'extraction, de visualisation et d'affichage en clair (ASCII) de données cartographiques sélectionnées par l'utilisateur.

Le chemin parcouru : de l'analyse à la programmation, nous avons été amenés à aborder systématiquement les points suivants :

- définition des critères de sélection des données cartographiques par l'utilisateur (choix de la zone et du type de contours, quelle résolution/densité de points ? quel type de projection cartographique ?
- jusqu'où pousser la convivialité du logiciel : menu, aide en ligne, gestion de fichiers "profile" ?
- optimisation de l'exploration des fichiers de la banque en testant différents algorithmes ;
- portabilité et maintenance du logiciel (sur quelle machine ? quel langage ? quelle norme graphique ?)
- intégration de ces fichiers cartographiques ASCII dans l'organisation plus générale des fichiers de travail standard (FTS) mise en place au Laboratoire d'hydrologie.

Toutes ces questions n'ont pas forcément reçu de réponse adéquate, notamment au niveau de la portabilité. Mais leur prise en compte aux différentes étapes de l'élaboration de LEICA a permis la réalisation d'un logiciel relativement "propre" :

- sur le plan de sa conception informatique ;
- sur le plan de son utilisation à des fins cartographiques.

1. INTRODUCTION(S)

Avant de rentrer dans le vif du sujet, il est bon de préciser certains points qui seront autant de points de repéré, notamment pour ceux qui ont une expérience plus ou moins approfondie de la cartographie.

1.1. Une cartographie très, très appliquée

Dans ce qui suit, ce que nous appelons pompeusement "cartographie" correspond en fait à une gestion très simplifiée de données cartographiques :

- pas de gestion au sens "banque/base" de données (aucune fonction d'alimentation ni de modification) ;
- zonage non pris en compte (mais possible par une voie détournée) ;
- fichiers cartographiques "grossiers" (séquentiels ASCII) accessibles sous éditeur.

1.2. Les besoins cartographiques au Laboratoire d'hydrologie

Une part croissante des programmes de recherche en hydrologie fait appel à la représentation spatiale de données telles que la pluie ou le relief. Les outils de traitement et de représentation sont en cours de développement, et le besoin apparaît alors de superposer à ces visualisations de phénomènes physiques (iso-valeurs, par exemple) des fonds de carte géographiques de pays, bassins versants, rivières, villes, stations de mesure. Il est donc nécessaire de mettre à la disposition des utilisateurs des modules de fabrication, de visualisation et d'utilisation de fichiers fond de carte.

Ceci nous a amené à définir des standards de fichiers, à créer différents logiciels de traitement cartographiques et à intégrer le tout dans la structure informatique du Laboratoire. (voir figures 1 et 2)

1.3. Contexte de réalisation de LEICA

Suite à la mise à disposition d'une bande magnétique contenant des données cartographiques mondiales, nous nous sommes trouvés face à des fichiers trop importants (12 Mo pour l'Afrique) et peu pratiques à traiter (codage binaire IBM). Nous avons décidé de mettre en place un outil, LEICA, Logiciel d'Extraction Interactif pour la Cartographie Appliquée, permettant l'extraction de ces données selon certains critères, la visualisation de ces données et leur archivage dans des fichiers ASCII "FTS" (Fichier Travail Standard).

Le logiciel a été écrit en FORTRAN 77 et tourne sur station de travail APOLLO.

2. FONCTIONNALITES DE LEICA

Un certain nombre de problèmes se sont posés avant, pendant et après la réalisation de ce logiciel. Cependant, avant de décrire comment ceux-ci ont été abordés, nous allons expliciter les principales fonctionnalités de LEICA.

2.1. Fonction Extraction

Cette fonction, point de départ de LEICA, permet à l'utilisateur d'extraire un fond de carte, c'est à dire un groupe de segments constitués chacun d'un ensemble de couples de coordonnées (x,y), en sélectionnant les paramètres suivants (voir liste détaillée sur le tableau 3) :

- continent ;
- fenêtre géographique (couple latitude/longitude, minimales et maximales) ;
- nature du contour et du sous-contour ;

L'INFORMATIQUE ET L'HYDROLOGIE

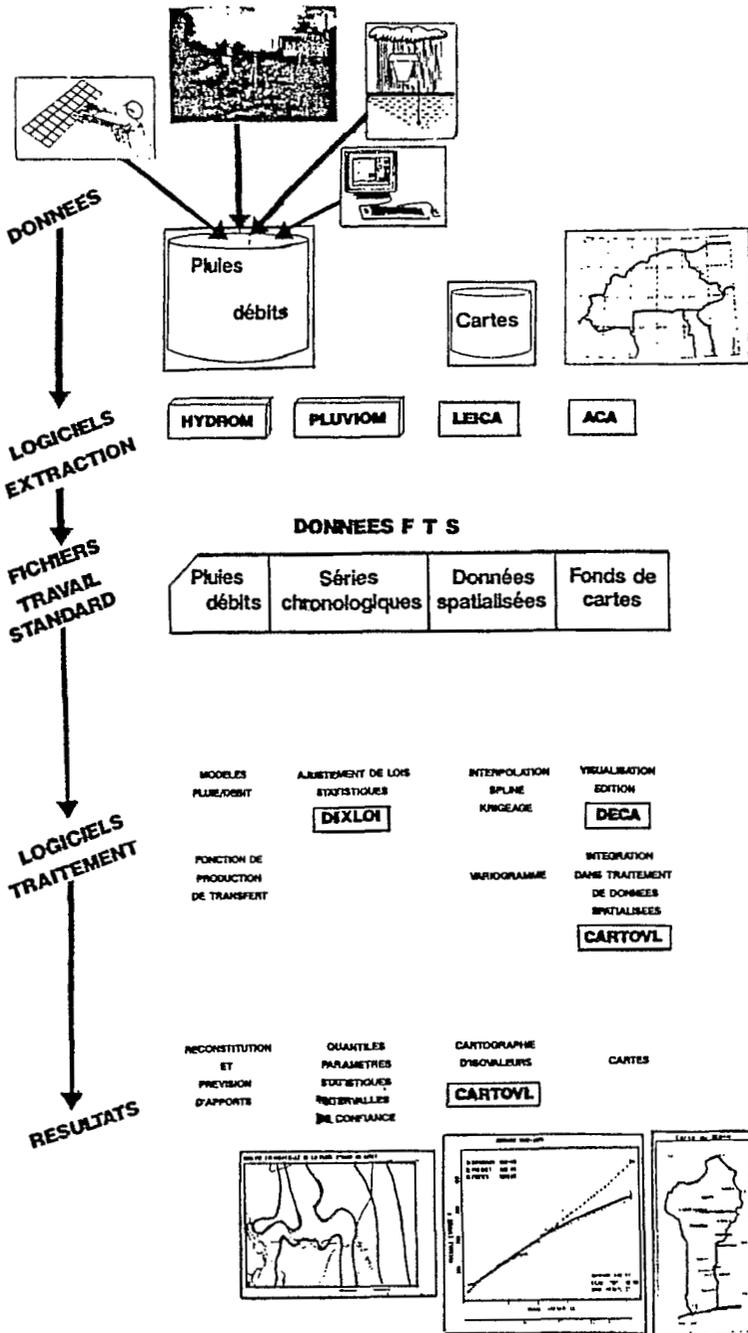


Figure 1 L'Informatique au Laboratoire d'Hydrologie

9101	BURKINA-FASO									0	EN-TÊTES	
26	(10F8.2)									0		
DD	0degre_dec	0.00	0.00	8	MERCATOR	EQUIDISTANTE	DIRECTE			3	0	
	0.1274200000E+09	0.0000000000E+00	0.0000000000E+00								0	
	-0.6000000000E+01	0.3000000000E+01	0.9000000000E+01	0.1600000000E+02							0	
1	4	102	CONTOUR 3, SOUS-TYPE 1								0	COORDONNEES DES POINTS DE CHAQUE CONTOUR
	-5.99	15.70	-5.82	15.70	-5.65	15.69	-5.49	15.69			0	
2	3	102	CONTOUR 3, SOUS-TYPE 1								0	
	-5.49	15.69	-5.45	15.87	-5.43	16.00					0	
3	8	102	CONTOUR 3, SOUS-TYPE 1								0	
	-2.68	9.52	-2.67	9.41	-2.70	9.33	-2.70	9.26	-2.76	9.18	0	
	-2.76	9.08	-2.65	9.05	-2.65	9.01					0	
4	23	102	CONTOUR 3, SOUS-TYPE 1								0	
	-2.69	9.53	-2.74	9.44	-2.82	9.50	-2.89	9.58	-2.93	9.67	0	
	-2.98	9.76	-3.08	9.79	-3.12	9.88	-3.18	9.94	-3.23	9.93	0	
	-3.30	9.92	-3.36	9.96	-3.47	9.98	-3.59	9.97	-3.67	9.99	0	
	-3.77	9.97	-3.90	9.93	-3.98	9.87	-4.08	9.86	-4.16	9.84	0	
	-4.26	9.79	-4.28	9.71	-4.31	9.64					0	
...												
...												
...												
...												
24	2	102	CONTOUR 3, SOUS-TYPE 1								0	
	2.79	9.09	2.78	9.00							0	
25	3	102	CONTOUR 3, SOUS-TYPE 1								0	
	2.79	9.09	2.90	9.11	2.99	9.10					0	
26	3	102	CONTOUR 3, SOUS-TYPE 1								0	
	2.84	12.49	2.92	12.42	3.00	12.37					0	

EXEMPLE DE FICHIER CARTOGRAPHIQUE DE TYPE "CONTOUR" (FTS91)

9299	Stations pluviometriques du Burkina Faso										0	EN-TÊTES
10	(2F8.2,1X,A30,I3)										0	
DD	0degre_dec	9.00	-5.00	8	MERCATOR	EQUIDISTANTE	DIRECTE			3	0	
	0.1274200000E+09	0.0000000000E+00	0.0000000000E+00								0	
	-0.5500000000E+01	0.2800000000E+01	0.9300000000E+01	0.1550000000E+02							0	
	-1.52	12.35	OUAGADOUGOU AERO								111	COORDONNEES ET CARACTERISTIQUES DE CHAQUE POINT
	-0.87	14.23	ARIBINDA								111	
	-3.30	11.75	BAGASSI								111	
	-5.42	10.53	BAGUERA								111	
	-1.50	13.33	BAM (TOURCOING)								111	
	-4.77	10.62	BANFORA AGRICULTURE								111	
	-1.85	14.22	BARBOULE								111	
	-2.92	9.88	BATIE								111	
	-3.68	11.62	BEREBA								111	
	-3.98	13.05	BOMBOROKUY								111	

EXEMPLE DE FICHIER CARTOGRAPHIQUE DE TYPE "POINT" (FTS92)

FIG. 2 : Fichiers Cartographiques
(Fichiers de travail standard)

- résolution associée à chaque sous-contour choisi (pourcentage de points sélectionnés) ;

- projection géographique et paramètres associés.

L'ensemble de ces paramètres est modifiable et visible à chaque instant. De plus, l'utilisateur peut mémoriser ses choix dans un fichier "profil", permettant ainsi un travail plus rapide au cours d'une session ultérieure.

Les données extraites sont stockées dans des fichiers intermédiaires.

2.2. Fonction Visualisation

Ce module, utilisant des routines graphiques propres à Apollo, permet de visualiser les contours extraits. (voir exemple sur figure 4). Par ailleurs, il est possible de superposer à ces contours des données cartographiques de type "point" (villes, stations de mesure, etc.), stockées dans un fichier FTS92. Ceci permet de comparer des données géographiques d'origine différente, ce qui pose assez souvent des problèmes délicats.

Après visualisation, l'utilisateur peut effectuer une nouvelle extraction, ou au contraire, archiver définitivement ses données.

2.3. Fonction Archivage

Le programme crée un fichier ASCII séquentiel (FTS91), selon des caractéristiques définies par l'utilisateur ; puis, il récupère les données stockées dans les fichiers intermédiaires, et les écrit dans le fichier utilisateur.

3. CONCEPTION DE LEICA

Nous présentons ici les problèmes rencontrés au cours de la réalisation de LEICA, sous les aspects techniques et méthodologiques, tout en les situant par rapport à une politique globale de l'informatique au Laboratoire d'hydrologie et à l'ORSTOM.

3.1. La période avant LEICA

Il y a deux ans, les premiers besoins en fonds de carte s'exprimèrent au Laboratoire ;

1^{ère} solution. Bricoler des fichiers de couples (x,y) et tracer des lignes joignant ces points.

2^{ème} solution. Utiliser ce qui pouvait déjà exister ailleurs (CNUSC, Maison de la Géographie, ORSTOM).

La première solution étant, à coup sûr, source de complications ultérieures, la deuxième solution révélant peu de choses, -ce "peu" étant impossible à récupérer pour diverses raisons-, nous avons démarré notre propre "cuisine" cartographique ce qui a abouti à :

- créer nos standards de fichiers cartographiques, FTS91 et FTS92, points de passage obligés de tout logiciel du Laboratoire utilisant des fonds de carte ;

BURKINA-FASO (Resol:10% Stereopolaire Ell.)

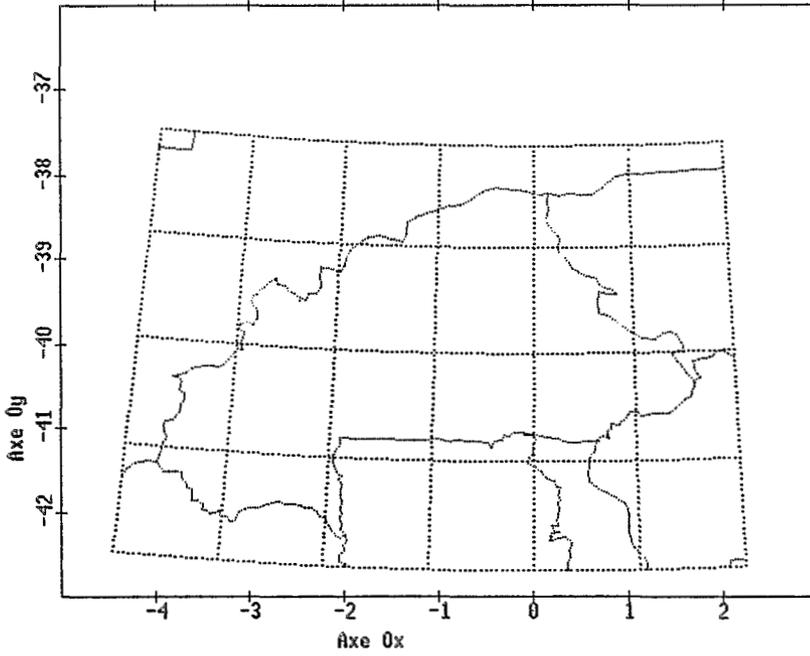


FIG. 4 : VISUALISATION DE DONNEES EXTRAITES

- réaliser un logiciel d'acquisition de carte, ACA, à l'aide d'une table à digitaliser.

Ainsi, à cette époque, après avoir constaté un manque de coordination et d'information à l'ORSTOM face à un besoin qui semblait général, nous avons été amenés à faire des choix techniques sur le mode de représentation des données cartographiques et à les figer pour mieux structurer leur gestion.

3.2. L'analyse de LEICA

L'analyse de LEICA repose en fait sur deux types de choix :

a) *des choix précis et clairs*, qui correspondent à des problèmes posés explicitement, écrits noir sur blanc :

- conversion des fichiers initiaux (codage IBM), définition des fichiers intermédiaires, optimisation de l'accès aux données ;

- choix des critères de sélection d'une carte par l'utilisateur (voir paragraphe 2.1 et tableau 3) ;

- convivialité du logiciel. La gestion des menus est réalisée par un "moteur" qui stocke dans une pile les informations relatives à chaque action de l'utilisateur et qui permet un déplacement rapide dans l'arborescence des choix. Les messages d'erreur sont systématiques. Enfin, l'utilisateur garde une trace de son travail (fichier profil) ;

- maintenance du logiciel. Conception modulaire et bonne documentation technique.

b) *des choix imprécis*, fondus dans un flou artistique et basés sur des réponses à des problèmes, soit trop complexes, soit apparemment trop simples, soit trop longs à résoudre :

- limites du logiciel (on ne peut modifier les fichiers originaux) ;

- utilisation du logiciel (celle-ci est limitée au Laboratoire d'hydrologie) ;

- portabilité (celle-ci est liée au choix de standard de langage et de représentation graphique).

Une partie est résolue par l'adoption du FORTRAN 77, et la partie graphique serait à faire (GKS).

A ce niveau, un manque très net se fait sentir quant à l'utilisation d'un outil d'analyse adapté à la réalisation de programmes scientifiques de moyenne importance. Entre le bricolage individuel de l'utilisateur qui résout une équation différentielle et le cahier des charges (flux et dictionnaire de données, etc.) d'un outil de gestion de l'envergure d'HYDROM/PLUVIOM, il ne reste qu'un ensemble de recommandations plutôt basées sur le bon sens : faire des programmes modulaires, éviter les GOTO, mettre des commentaires.

3.3. La programmation de LEICA

Aucun problème à signaler !

3.4. L'après LEICA

Mis à part quelques petites modifications qu'on pourrait apporter au logiciel, l'utilisation de celui-ci pose immédiatement la question de la qualité des données (on aurait pu s'en douter !).

Celles-ci semblent bonnes, mais :

- certains fleuves n'apparaissent pas !
- lorsqu'on pointe une station pluviométrique à Dakar, celle-ci se retrouve en pleine mer à cause d'une erreur de 15mn sur la longitude du Cap Vert !

Et nous n'avons pas de solution immédiate pour remédier à ce genre de problème.

4. CONCLUSION

En guise de conclusion, il nous semble intéressant de poser deux questions carto-informatiques, tant aux utilisateurs qu'aux animateurs et responsables informaticiens :

- en matière de diffusion de fonds de carte informatisés à l'ORSTOM, peut-on envisager un outil centralisé (ou réparti) permettant d'obtenir des fichiers cartographiques fiables, sous un format standard pour tel ou tel pays ?
- en matière de méthode d'analyse, existe-t-il ou est-il possible d'acquérir, des outils ou une formation permettant de fabriquer du logiciel "intelligemment" sans être obligé de passer par des méthodes complexes nécessitant un investissement trop important ?

REFERENCES BIBLIOGRAPHIQUES

- BERNARDET, X. (1989). La Cartographie et l'informatique. Mémoire BTS, ORSTOM, Montpellier.
- DELCLAUX, F. (1989). Les fichiers FTS. ORSTOM Hydrologie, Montpellier.
- LAWIRA, C. (1988). Notice d'utilisation de LEICA. ORSTOM Hydrologie, Montpellier.

PARAMETRE	CHOIX POSSIBLE
Continent	Afrique
Fenêtre géographique	Couple de longitude/latitude maximum et minimum (exprimées en d°mn sec)
Contour sous-contour	<p>Iles, cotes et lacs</p> <ul style="list-style-type: none"> - Principaux cotes et lacs - Principaux îles, cotes et lacs supplémentaires - Îles et lacs intermédiaires - Îles et lacs mineurs - Principaux lacs intermittents - Lacs intermittents secondaires - Récifs - Principaux marais salants - Marais salants secondaires - Principales calottes glaciaires - Calottes glaciaires secondaires - Glaciers <p>Les cours d'eau</p> <ul style="list-style-type: none"> - Principaux fleuves ou rivières permanents - Principaux fleuves ou rivières supplémentaires - Fleuves ou rivières intermédiaires - Fleuves ou rivières mineurs - Principaux fleuves ou rivières intermittents - Fleuves ou rivières intermittents intermédiaires - Fleuves ou rivières intermittents mineurs <p>Les frontières d'état</p> <ul style="list-style-type: none"> - Frontières stables - Frontières instables - Autres lignes de démarcation <p>Les frontières internes des pays</p> <ul style="list-style-type: none"> - Séparation interne dans un pays (USA)
Résolution	0.1 à 100% pour chaque contour/sous-contour
Projection géographique	<p>Stéréopolaire (sphérique ou elliptique)</p> <p>Lambert sécante</p> <p>Mercator équatoriale</p> <p>Mercator transverse universelle</p> <p>Mercator équidistante directe</p> <p>Lambert conforme</p>

Tableau 3 : Liste des paramètres d'extraction de fonds de carte.

SISGEO : UNE METHODOLOGIE D'ADMINISTRATION DES DONNEES DE L'ENVIRONNEMENT RURAL.

(Patrick SECHET, Paulo S.V. FRESNEDA & Gérard COCHONNEAU)

RESUME - L'EMBRAPA, *Entreprise Brésilienne de Recherche Agronomique* est amenée à collecter des données de sols, climat, végétation, etc., de façon à compléter les informations sur l'environnement dont elle a besoin pour mener à bien ses programmes de recherche, notamment ceux qui visent une exploitation rationnelle des ressources du milieu rural.

Compte tenu du volume important de ces données et de la difficulté de manipulation qui en résulte, il a été décidé de les rendre disponibles sur ordinateur. De cette manière, la composition de toutes ces informations à partir de critères géographiques deviendra possible, et favorisera l'évaluation et l'exploitation des potentialités agricoles de chaque région.

Pour atteindre cet objectif, un programme (SISGEO) a été défini et réunit des spécialistes de chaque thème, rencontrés parmi les équipes de cinq centres de recherches de l'entreprise et du service des sols, avec des informaticiens de la Division de Développement de Systèmes, du Département d'Informatique du siège. Depuis 1984, l'ORSTOM apporte son concours à ce projet, par le biais de l'affectation de personnel et la fourniture de logiciels.

Pour prendre en compte l'état très différent dans lequel les informations peuvent être obtenues pour chaque thème, le programme est réalisé de façon modulaire, thème à thème : un modèle de saisie, stockage, structuration et interrogation des données pour un thème déterminé est conçu et mis au point, par collaboration entre les informaticiens du siège et les spécialistes thématiques de l'un des centres, puis diffusé dans toutes les autres unités. On s'efforce d'ailleurs de faire intervenir le plus grand nombre de spécialistes de l'entreprise, principalement durant la phase de conception.

Ainsi, des systèmes thématiques ont d'abord été définis et couvrent les champs essentiels pour l'activité de recherche de l'entreprise : SISCLIMA pour la gestion des données agroclimatologiques, SISOLOS pour l'exploitation des inventaires pédologiques, SISFITO pour administrer les données des collections botaniques et la phytogéographie, SISECSO pour gérer les informations de socio-économie au niveau de la commune, SISHIDRO pour manipuler les données qui se réfèrent aux ressources hydriques et SISPRAGAS pour cataloguer les informations phytosanitaires.

Ils comprennent un premier module qui utilise les équipements informatiques disponibles dans les centres (micro-ordinateurs PC-compatibles), permettant la constitution de bases de données locales et leur exploitation en vue d'études de portée régionale. Un deuxième module, développé sur les ordinateurs centraux

de l'EMBRAPA, doit garantir le partage, l'intégrité et la sécurité des données, tout en permettant l'utilisation de moyens d'analyse plus complets et la réalisation d'études de portée nationale.

A côté de ces systèmes thématiques, SISGEO prévoit l'élaboration de systèmes transversaux, SISMAPAS et SISIMAGEM, qui permettront d'exploiter la géocodification, ainsi que la mise à jour permanente que la télédétection rend possible. SISMAPAS comprend un premier module de saisie de données cartographiques, tandis que SISIMAGEM prévoit un module de traitement des images de satellites, en vue de leur intégration dans les bases de SISGEO. Ces deux logiciels comprendront un second module de formation de bases de données centrales.

Le partage et l'utilisation composée de toutes ces informations seront garantis par un système d'intégration, qui constituera la dernière étape de SISGEO. Il comportera deux niveaux : le premier fera l'intégration des bases de données implantées sur le site central ; le second mettra en cohérence les modules disponibles dans chaque centre.

Dans l'état actuel de réalisation (début 1989), des logiciels sont opérationnels dans les domaines de climat, sols, végétation et traitement de cartes. On espère terminer cette année un nouveau module de SISOLOS et une première version de SISPRAGAS, en plus de versions plus complètes de SISFITO et de SISMAPAS.

Cette communication présente la méthode qui a été choisie pour l'élaboration de ce travail, compte tenu des traitements envisagés et des problèmes rencontrés dans le contexte particulier de réalisation. Elle fait aussi le point sur l'état d'avancement du programme et ses perspectives à moyen terme.

INTRODUCTION

L'utilisation agricole des ressources naturelles d'un pays doit se faire d'une façon ordonnée et productive, en se fondant sur la connaissance scientifique du monde rural, de manière à produire de bons résultats économiques, tout en protégeant l'environnement d'une exploitation préjudiciable.

Les données de sols, climat, végétation, ressources hydriques, ainsi que les informations phytosanitaires ou encore de nature socio-économique sont nécessaires à l'activité agricole, et, au Brésil, font l'objet d'inventaires de la part de divers organismes. Toutefois, les informations qui en résultent sont hétérogènes, la plupart d'entre elles étant d'accès difficile (souvent non publiées) : leur diversité et leur volume rendent impraticable l'utilisation de moyens conventionnels d'utilisation.

Par ailleurs, l'Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA), dans le but de développer des produits et des techniques adaptés aux conditions agroécologiques de chaque région, entretient des équipes multidisciplinaires de chercheurs dans chaque grande zone climatique. Aussi, ces équipes mirent rapi-

dement en évidence la nécessité d'élaborer un système d'informations géographiques, permettant un stockage adéquat, sur ordinateur, des données de l'environnement agricole et facilitant la récupération et la diffusion de ces informations, sous forme de rapports ou de cartes.

Dès la mi-1982, sous l'impulsion et la coordination des informaticiens du siège de l'EMBRAPA, l'ensemble des chercheurs concernés décidait de collaborer à la réalisation du "Sistema de informações geoambientais (SISGEO)", pour laquelle le concours de l'ORSTOM était sollicité.

Pour l'Institut, l'intérêt d'un tel programme était évident, s'agissant d'une tentative de valorisation des connaissances accumulées sur le milieu rural, domaine de compétences traditionnelles de nos chercheurs. Bien que des expériences analogues étaient déjà en cours ou prévisibles, et particulièrement en Amérique Latine, la perspective de travailler sur le sujet avec un partenaire de la taille de l'EMBRAPA était prometteuse.

La méthodologie envisagée et la stratégie de réalisation adoptée pour ce système ont été définies très tôt et amplement discutées, notamment au cours de réunions scientifiques (SIMPLICIO & SECHET, 1983), (SIMPLICIO & alii., 1984). Cette communication présente les caractéristiques les plus significatives de l'approche adoptée, avec les avantages et problèmes qui en découlent, eu égard au contexte particulier de ce programme. Elle termine en faisant le point de l'état d'avancement et des perspectives actuelles de SISGEO.

1. CONTEXTE INSTITUTIONNEL ET HISTORIQUE

1.1. Le contexte institutionnel

L'EMBRAPA a été créée en avril 1973 pour regrouper et mettre en cohérence les activités de recherche agronomique exercées par un certain nombre d'organismes régionaux rattachés au Ministère de l'Agriculture, comme l'IPEAN à Belém, l'IPSul à Pelotas, le DNPS à Rio de Janeiro, etc..

Aujourd'hui, l'entreprise compte plus de 8 000 employés, dont près de 2 000 chercheurs, répartis dans environ une cinquantaine d'implantations sur tout le territoire national : c'est probablement, après le CNRS en France, le plus important partenaire institutionnel de l'ORSTOM. La carte présentée figure 1 situe géographiquement toutes ces unités.

La structure de l'EMBRAPA est double :

- l'ossature principale est formée par un ensemble de vingt et un "*centres de produits*", implantés dans les régions les plus favorables à la production et chargés de mener une recherche appliquée à un type déterminé de culture végétale (riz, soja, coton, blé, etc.) ou d'espèce animale d'élevage (vache laitière, chèvre, ovin, etc.). Chacun de ces centres est responsable et coordinateur d'un programme national de recherche, auprès duquel tout autre organisme (université, par exemple) peut proposer ses propres programmes de recherche à un financement ;

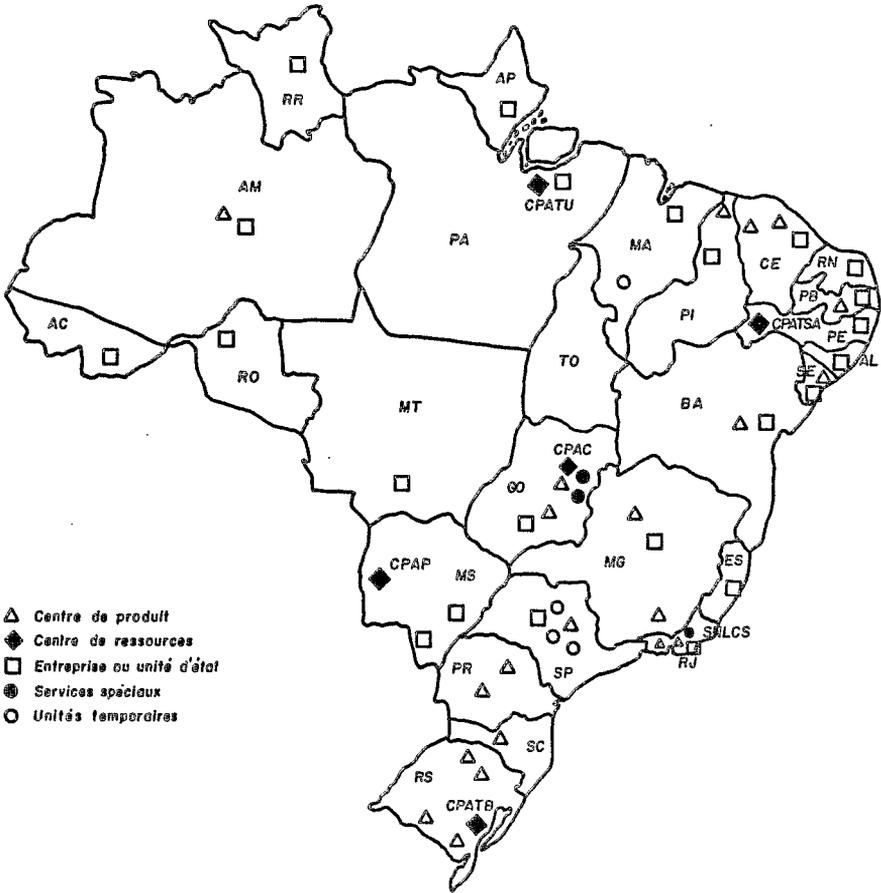


Fig. 1 - Carte de situation des unités de l'EMBRAPA et du SCPA

- une structuration géographique est également assurée par le biais de ses cinq "centres de ressources", responsables pour la recherche d'une meilleure adéquation de l'activité agricole aux conditions particulières offertes par l'environnement dans leur région physiographique d'intervention. Ces derniers centres maintiennent trois équipes de recherche distinctes : une équipe d'évaluation des ressources naturelles de la région (inventaire et caractérisation), un groupe chargé de la recherche sur l'exploitation de celles-ci et une troisième équipe responsable pour la recommandation des systèmes de production les mieux adaptés au milieu.

Quelques unités spéciales ou temporaires, à vocation essentiellement technologique complètent, avec le siège administratif et de soutien technique de l'entreprise, ce dispositif.

La coopération internationale avec les instituts de recherche français (CIRAD et ORSTOM) a débuté fin 1980 : les chercheurs du CIRAD interviennent généralement dans les centres de produits, avec une recherche plus finalisée, tandis que ceux de l'ORSTOM conduisent généralement des programmes communs avec les équipes d'évaluation des centres de ressources.

1.2. Historique

Dans le domaine des systèmes d'information pour l'environnement, un premier travail commun (le projet SISOLOS) a été mené par l'ORSTOM avec le SNLCS, Serviço Nacional de Levantamento e Conservação de Solos, pour conduire à la réalisation d'une base nationale de données de sols (Séchet 1985). Durant la phase de développement (1981-83) ce travail a bénéficié d'un appui technique essentiel du DMQ, Departamento de Métodos Quantitativos, unité centrale du siège de l'EMBRAPA, où l'on a rapidement pris conscience de la nécessité de systématiser et discipliner l'accès aux données de l'environnement, sous la pression de la demande des équipes d'évaluation des trois centres de ressources alors existants.

En effet, les chercheurs du CPAC, Centro de pesquisa agropecuária da região dos Cerrados à Brasília, du CPATSA, Centro de pesquisa agropecuária da região do Trópico Semi-Arido à Petrolina et du CPATU, Centro de pesquisa agropecuária da região do Trópico Umido à Belém, essentiellement engagés dans des opérations de zonification agricole de leurs régions respectives, nécessitaient tous d'un appui informatique pour accéder aux données, généralement existantes mais pratiquement indisponibles.

Un groupe de travail a donc été créé en mai 1982, à l'occasion du Primeiro Encontro de Métodos Quantitativos da EMBRAPA (EMBRAPA 1982), pour mettre sur pied une stratégie de réalisation d'un système d'information géographique à vocation rurale, appelé SIGGEO. Plusieurs réunions réalisées en 1982 et 1983, sous l'impulsion du DMQ/EMBRAPA et de l'ORSTOM, ont permis de définir les modalités d'élaboration du système et de participation de l'ORSTOM, les travaux ne débutant effectivement qu'à l'arrivée (mi-1984) des chercheurs français affectés au programme (Daniel Cardon au CPATU et Patrick Séchet au DMQ).

Durant les cinq années qui se sont écoulées, l'effort de l'ORSTOM consacré à cette coopération s'est maintenu au niveau de deux chercheurs ou ingénieurs affectés : début 1986, Philippe Waniez était affecté au CPAC, peu avant le départ de Daniel Cardon et en mars 1987 Gérard Cochonneau rejoignait Patrick Séchet à Brasília, tandis que Philippe Waniez quittait le CPAC en fin d'année. Bien entendu, l'EMBRAPA allouait dans le même temps des ressources humaines et financières significativement supérieures, avec moins de régularité cependant.

2. LES GRANDES ORIENTATIONS

Quatre principes fondamentaux ont été adoptés pour le projet, dès son origine, et continuent à être respectés :

- réalisation progressive du système, par le biais de modules thématiques autonomes ;
- constitution d'équipes mixtes de développement de modules, comprenant thématiciens et informaticiens ;
- offre de facilités de stockage et traitement de données, à niveau central et à niveau local ;
- recherche systématique de l'appui institutionnel pour le projet.

Les raisons qui ont conduit à l'adoption de ces principes, les avantages que l'on en tire, ainsi que les contraintes qu'ils obligent à respecter, sont détaillés ci-après.

2.1. Modules autonomes

La principale caractéristique originale de SISGEO est son développement par modules autonomes. Six sous-systèmes thématiques ont été définis (figure 2), auxquels viennent s'ajouter deux systèmes transversaux pour la manipulation des documents cartographiques (SISMAPAS) et des images de satellites (SISIMAGEM).

Il s'agit là du principe fondamental, dont l'adoption et le respect conditionnent le développement de SISGEO. Bien que les raisons de ce choix soient multiples, chacune d'elles prise isolément justifierait le principe adopté.

La première raison qui motive le découpage de SISGEO en modules autonomes provient de la volonté d'assurer l'"exhaustivité" du système d'information à élaborer par rapport aux données de l'environnement, du moins à celles effectivement disponibles par les procédures de collecte systématique (inventaires).

En effet, pour prendre un exemple simple, on peut noter que le module d'information sur les sols (SISSOLOS) modélise, stocke et permet l'exploitation de toutes les données collectées en routine durant les inventaires pédologiques. Cela n'aurait pas été possible si l'on s'était satisfait d'une modélisation de la seule information de sols à caractère spatialisé, dans le cadre d'un système d'information géographique plurithématique intégré.

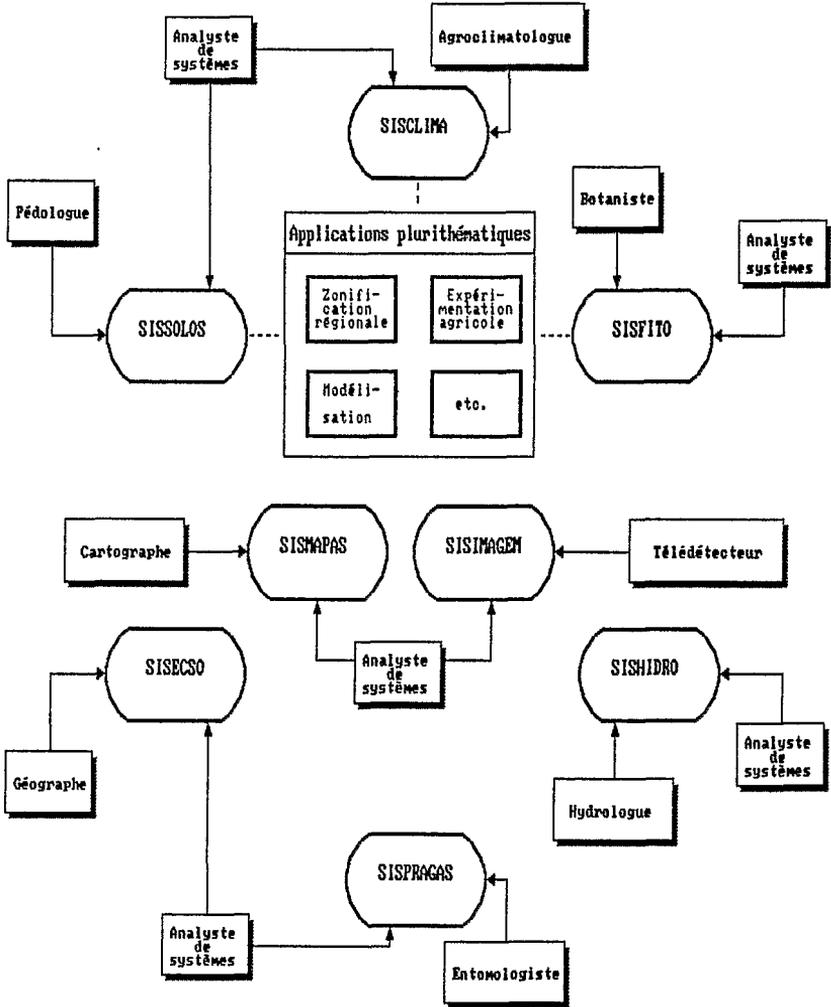


Fig.2 Décomposition de SISGEO en modules autonomes

Les avantages immédiats que l'on déduit de la méthode choisie de réalisation de modules indépendants sont nombreux et évidents, et constituent autant de justifications de ce choix.

Les modules autonomes permettent en effet de satisfaire tous les besoins d'interrogation identifiables à partir des données disponibles, qu'ils soient à caractère géographique ou non. Ainsi, le module d'informations agroclimatologiques (SISCLIMA) permet-il de faire des études d'évapotranspiration potentielle, pour affiner la détermination des paramètres des diverses formules existantes, par exemple.

De plus, chaque module peut bien sûr être utilisé indépendamment des autres, sans avoir à attendre la conclusion du système global, hypothétique dans tous les cas, puisque dépendant d'un effort important de modélisation lequel, à ce jour, reste en grande partie à faire (principalement dans le cas des données socio-économiques ou phytosanitaires, pour ne citer que les plus complexes).

Il faut remarquer par ailleurs que les données de la plupart des thèmes ne sont pas naturellement structurées géographiquement, en raison notamment du mode de collecte. C'est le cas par exemple des données hydrologiques ou climatologiques, rattachées à des lieux d'observation ponctuels (postes, stations). Il existe bien entendu quelques tentatives de spatialisation de ces informations (réalisation d'isolignes, méthode d'"abattement", etc.), mais la mise en oeuvre en est encore très sporadique et une systématisation paraît encore impossible, aucune méthode n'ayant encore fait l'objet d'un consensus.

Ceci revient à dire que les données des différents thèmes ne sont généralement pas toutes disponibles sous forme de cartes, sauf cas particulier, conditionnant par conséquent la possibilité de réalisation d'un système intégré à l'existence de données déjà élaborées sous forme spatiale.

L'approche ainsi adoptée a évidemment pour principal inconvénient de rendre plus lointaines les perspectives d'atteinte de l'objectif global du système proposé, inconvénient partiellement compensé par la fourniture continue des résultats intermédiaires que constituent les modules indépendants. SISGEO apparaît ainsi plus souvent comme une *politique d'informatisation des données de l'environnement* que comme une véritable tentative de réalisation d'un système d'information géographique.

Enfin, le développement en modules autonomes permet d'entrevoir un bénéfice additionnel pour l'ORSTOM, dans la mesure où la plupart des produits obtenus sont susceptibles d'être utilisés par d'autres équipes de l'Institut dans d'autres contextes, soit tels quels, soit le plus souvent après adaptation, soit encore en utilisant simplement les éléments obtenus à l'issue des étapes initiales (dossier de conception détaillé). On peut percevoir d'ores et déjà plusieurs possibilités de ce type, au moins pour SISCLIMA et SISMAPAS, système de traitement automatisé de documents cartographiques.

2.2. Equipes mixtes et responsable thématique

Le développement d'un module thématique nécessite une première réflexion sur les fonctions qu'il est essentiel d'automatiser pour la discipline considérée. Il faut alors concilier deux objectifs différents : celui du spécialiste thématique, soucieux de se voir affranchir des tâches les plus routinières de son activité scientifique et celui de l'informaticien préoccupé par la capitalisation de l'information et son éventuelle spatialisation.

Le système d'informations phytosanitaires (SISPRAGAS) illustre bien la possibilité de conflit entre les deux objectifs (SECHET et alii. 1989).

Il est ensuite indispensable de procéder à la modélisation de l'information qui va intégrer la base de données à créer, ce qui ne peut être fait qu'en additionnant les compétences scientifiques des thématiques aux compétences techniques des informaticiens.

Il existe par conséquent, pour chaque module de SISGEO, une équipe thématique et une équipe informatique, ayant respectivement à leur tête le "responsable thématique" du module et un "analyste de conception", sorte de chef de projet informatique. Le choix des fonctions à automatiser est effectué d'un commun accord entre ces deux spécialistes, tandis que l'équipe thématique assure la fourniture des informations scientifiques et que l'équipe informatique est responsable de l'implémentation du module, selon les normes générales définies pour SISGEO. Par sa propre motivation et par son renom dans la communauté scientifique nationale, le responsable thématique joue un rôle important pour l'approbation du projet de développement du système par les bailleurs de fonds et aussi pour l'acceptation postérieure du produit mis au point par les autres équipes de la même discipline.

Cette forme de travail a été rapidement jugée préférable, pour la réalisation des systèmes de soutien à la recherche tels que les modules de SISGEO, à un travail plus individuel de développement informatique par les propres thématiques, par ailleurs couramment pratiqué (en particulier par les chercheurs de l'ORSTOM), pour les raisons suivantes :

- implication plus forte des utilisateurs potentiels du produit dès le début de sa mise au point ;
- meilleure qualité des produits sur le plan scientifique, due à l'intervention d'un groupe de spécialistes plutôt que d'une seule personne à double compétence ;
- meilleure qualité des produits sur le plan informatique, du fait de sa réalisation par une équipe de professionnels ;
- garantie d'institutionnalisation du produit créé, apportée par la formalisation du projet qui fait intervenir une équipe bidisciplinaire et par le partage de responsabilités au niveau des deux gérances (thématique et informatique).

Il ne paraît cependant pas indispensable que soit appliquée cette règle pour toutes les initiatives informatiques, mais seulement pour des projets de systèmes de soutien de recherche, qui revêtent normalement un caractère plus universel que les autres réalisations dans le domaine.

2.3. Module central et module local

La réalisation d'une base de données suppose normalement un support physique unique, capable d'autoriser une utilisation partagée des facilités mises à disposition, c'est à dire les données et les traitements possibles. Une des ambitions de SISGEO est évidemment de permettre l'accès partagé aux données de l'environnement recueillies pour tout le territoire national, et stockées par les soins, dans un premier temps, des équipes des centres de ressources de l'EMBRAPA.

Le premier système mis au point, SISOLOS, répond à cet objectif et conserve sur les ordinateurs centraux du siège de l'EMBRAPA à Brasília toutes les informations qui correspondent aux levés pédologiques réalisés durant une trentaine d'années au Brésil (totalisant environ 12 000 profils de sols).

Cette seule approche est néanmoins apparue insuffisante, compte tenu en particulier des difficultés que les chercheurs des unités décentralisées doivent surmonter pour accéder à l'ordinateur central, surtout en l'absence d'un véritable réseau de communication de données, encore en cours de mise en place au Brésil.

SISGEO se propose en fait de réaliser systématiquement deux modules pour chacun de ses systèmes thématiques :

- un module dit *local*, destiné à une ample diffusion parmi les équipes productrices d'informations et essentiellement orienté vers la saisie et la validation des données du thème considéré, au niveau d'agrégation le plus détaillé ;

- un module dit *central*, en version unique implantée sur un système serveur muni de facilités d'extraction et de traitement d'information. Celui-ci garantit, en plus de l'accès partagé, la sécurité des données, impossible à assurer avec des modules locaux gérés par les équipes thématiques.

Le module local est prévu pour permettre déjà un certain nombre de traitements sur les données régionales saisies, en favorisant des travaux de recherche méthodologiques ou fondamentaux. L'équipement de toutes les unités de l'EMBRAPA en PC-XT compatibles s'étant généralisé à partir de 1986, ces modules locaux sont opérationnels sur ce type de configuration.

En passant du module local au module central, les données devront subir une agrégation, dans la mesure où les applications, à priori le plus souvent géographiques, qui doivent être réalisées sur un ensemble plus important de données, s'accommodent normalement d'un niveau de synthèse plus élevé.

Il s'agit là d'une manière de résoudre l'épineux problème de propriété des données, puisque ainsi, les "producteurs" de données conservent un droit de propriété exclusive sur celles-ci, du moins au niveau le plus fin. On peut illustrer ce fait avec le module correspondant aux données climatologiques, pour lequel le module local conserve et traite des données à niveau horaire, dites "*instantanées*" (trois observations par jour et registres continus), tandis que le module central ne devra recevoir que les données moyennes journalières, généralement suffisantes pour les applications de type inventaire.

Par ailleurs, la réalisation préalable du module local permet de familiariser les utilisateurs potentiels à la manipulation informatique de leurs données et constituera à l'avenir le moyen d'accès à la base centrale : il sera ainsi possible

d'extraire un ensemble d'informations de la base centrale à partir d'un module local quelconque et, inversement, d'alimenter (ou réalimenter) la base centrale à partir du module local dûment autorisé pour la région considérée. Enfin, pour certains thèmes, l'élaboration ou l'utilisation d'un module local auront été le premier contact des thématiciens avec l'informatique ; le système aide ainsi à introduire un début de culture informatique dans certaines disciplines à l'EMBRAPA.

Avec cette philosophie de travail, SISGEO devient apparemment très complexe, puisque le nombre de ses modules double ! Toutefois, il est important de remarquer que les fonctionnalités implémentées dans l'un et l'autre module sont totalement complémentaires : les seules redondances existent au niveau des modules méthodologiques, en particulier pour celui de traitement des images de satellite, SISIMAGEM, qui doit mettre en oeuvre des opérations semblables aux deux niveaux (ASSAD et alii 1989). De plus, la structure ainsi envisagée pour le système, induite par le contexte institutionnel propre de l'EMBRAPA, est compatible avec sa vocation nationale, dans la mesure où les bases centrales devront progressivement être implantées dans les institutions appropriées (SISHIDRO au DNAEE, SISFITO à l'IBDF, SISCLIMA à l'INEMET, etc.).

2.4. Institutionnalisation de SISGEO

Un programme d'une telle envergure n'a de chances d'aboutir que s'il est connu et devient progressivement un véritable projet de l'entreprise dans lequel il est mené, et non de la seule équipe de réalisation. L'institutionnalisation du programme constitue une préoccupation constante de ses animateurs et sa recherche revêt diverses formes :

- l'utilisation de sigles, associés à celui de l'EMBRAPA, est une forme simple et économique de faire prendre conscience de l'existence du programme. Un écran de présentation normalisé rappelle pour chaque produit le rattachement à SISGEO, et à la collaboration EMBRAPA/ORSTOM (le dépôt de la marque SISGEO est actuellement envisagé) ;

- chaque nouveau développement fait l'objet de l'élaboration d'un avant-projet, dûment cautionné par les autorités des unités concernées de l'équipe thématique (chef du centre de ressources) et de l'équipe informatique (chef du DIN, Département d'Informatique de l'EMBRAPA, qui a pris la suite du DMQ pour la coordination de SISGEO). De cette façon, chaque activité partielle s'inscrit dans la programmation de soutien de recherche des unités considérées, devenant un projet dûment formalisé et suivi ;

- dès février 1986, une équipe permanente d'informaticiens était créée par résolution de la direction exécutive de l'entreprise, avec reconnaissance de la participation effective de l'ORSTOM. Le problème des droits et devoirs respectifs des deux institutions, vis à vis des produits et résultats de SISGEO était alors posé et, compte tenu du caractère pionnier de cette question, aura nécessité près de trois ans avant de trouver une solution satisfaisante pour les deux partenaires ;

- les équipes thématiques de SISGEO sont normalement constituées de chercheurs d'un même centre de ressources, pour des raisons évidentes de facilité de réunion. Ainsi, chacun de ces centres s'est il vu attribuer le parrainage d'au

moins un module thématique, dont il assume la responsabilité du développement face aux autres centres. Le choix du thème n'est pas aveugle, mais dicté par la compétence particulière dont dispose l'équipe d'évaluation, elle-même résultante de l'acuité du problème dans la région d'intervention. Le produit obtenu est ensuite destiné à être distribué à toutes les équipes intéressées. La création par l'EMBRAPA, en 1984-85 de deux nouveaux centres de ressources, le CPAP, Centro de pesquisa agropecuária do Pantanal à Corumbá et le CPATB, Centro de pesquisa agropecuária das terras baixas de clima temperado à Pelotas, a conduit à la distribution des deux modules non encore débutés : SISHIDRO à Corumbá (la région du Pantanal étant confrontée à des problèmes d'inondation périodique) et SISPRAGAS à Pelotas, en raison de l'importance économique de l'aspect phytosanitaire dans la région des rizières et d'élevage du sud du pays. De cette façon, la totalité du réseau de centres de ressources de l'EMBRAPA est impliquée dans l'enjeu de SISGEO.

3. ASPECTS METHODOLOGIQUES

Compte tenu de ce que la coordination du système global est encore actuellement confiée à des informaticiens, les aspects méthodologiques du programme en constituent sans aucun doute un point fort. La préoccupation la plus importante est d'assurer la pérennité des produits développés, ainsi que leur plus large diffusion dans tout le système EMBRAPA. L'institutionnalisation de SISGEO y contribue pour une bonne part, en même temps que les aspects ci-dessous détaillés :

- utilisation d'une méthodologie de développement rigoureuse et formellement documentée. Cet aspect très important, qui touche aussi bien la partie conceptuelle que celle d'implémentation, a déjà fait l'objet d'une description circonstanciée au cours de Séminfor 1, et reste pionnier pour l'Institut (Séchet 1987). En particulier, les modules évoluent progressivement sous la forme de versions successives, de plus en plus perfectionnées et bien sûr compatibles entre elles au niveau des bases de données produites ;

- assimilation progressive d'outils de génie logiciel adaptés au contexte de développement de l'équipe informatique, comme éditeur de diagrammes hiérarchiques (aide à la programmation structurée), outil d'aide à la conception, etc.. Adoption et mise au point d'un environnement de développement approprié à la réalisation des systèmes de capitalisation d'information (gestionnaire de grilles d'écrans, utilitaires de gestion de lexiques en ligne, système normalisé de sécurité, contrôle d'accès, procédures d'installation, etc.). Cette façon de travailler permet d'assurer la meilleure homogénéité possible entre les produits mis au point (facilitant ainsi l'acceptation par les équipes thématiques d'utilisateurs), tout en rationalisant le développement et en facilitant la maintenance pour l'équipe informatique ;

- recherche constante de la convivialité maximale des produits, du moins pour les modules locaux. Cet aspect est tout à fait fondamental étant

donné que l'on souhaite la plus grande diffusion possible, en même temps que l'on entend mettre ces outils à la portée d'un personnel technique intermédiaire. On peut illustrer l'intérêt que représente cet objectif avec le module local de SISCLIMA, pour lequel l'essentiel du traitement (saisie, validation, émission des bulletins mensuels et annuels, etc.) est réalisé par les propres observateurs des stations climatologiques, permettant de libérer le spécialiste en agroclimatologie pour des tâches plus scientifiques d'interprétation, tout en valorisant le travail de l'observateur.

4. SITUATION ACTUELLE ET PERSPECTIVES

A ce jour (juillet 1989) six produits sont actuellement mis à la disposition des utilisateurs de SISGEO, dont trois modules thématiques locaux (SISCLIMA 2, SISFITO 1.1 et SISOLOS 1.0), un module thématique central (SISOLOS), un système transversal (SISMAPAS 2.0) et un utilitaire (SISTAB 1.1). Le processus de production étant continu, il permet d'entrevoir désormais la sortie permanente d'au moins trois produits par an, chiffre qui prend toute sa valeur si on le compare à la production globale de logiciels de l'Institut ou de l'EMBRAPA.

La situation actuelle du programme, ainsi que ses perspectives d'évolution à court terme peut être détaillée par modules thématiques, systèmes transversaux et utilitaires opérationnels.

4.1. SISCLIMA

Le CPATU, comme les autres centres de ressources de l'EMBRAPA, a été amené depuis sa création à installer, maintenir et opérer un réseau significatif de stations agrométéorologiques, permettant une évaluation du climat de la région du "tropicque humide". Le cumul des informations ainsi générées pendant les deux dernières décennies a conduit l'équipe responsable à la réalisation de SISCLIMA, Sistema de informaçoes agrometeorológicas, en collaboration avec l'équipe de SISGEO, dans le but d'en automatiser le traitement.

Les objectifs de ce module, dont le développement a débuté en 1984, est de rendre disponibles plus rapidement les informations collectées, avec un meilleur niveau de confiabilité, tout en allégeant les tâches de gestion de stations, d'élaboration du bulletin agrométéorologique mensuel et de fourniture de données qui incombent aux chercheurs responsables (CARDON et alii. 1988).

Le module local, opérationnel dans sa version 2 depuis avril 1988, est utilisable sur micro-ordinateur PC-compatible et implémente les ressources de saisie, validation et traitement des données les plus routinières. Les applications actuellement disponibles sont illustrées à l'aide du diagramme de flot de données présenté figure 3 et permettent :

- la saisie des données collectées pour les trois horaires d'observation, par le biais d'un programme conversationnel ;
- la critique automatique des informations stockées, à partir d'un ensemble de paramètres (valeurs maximales et minimales, le plus souvent) fournis et modifiables en routine par le responsable de la station ;

SISGED

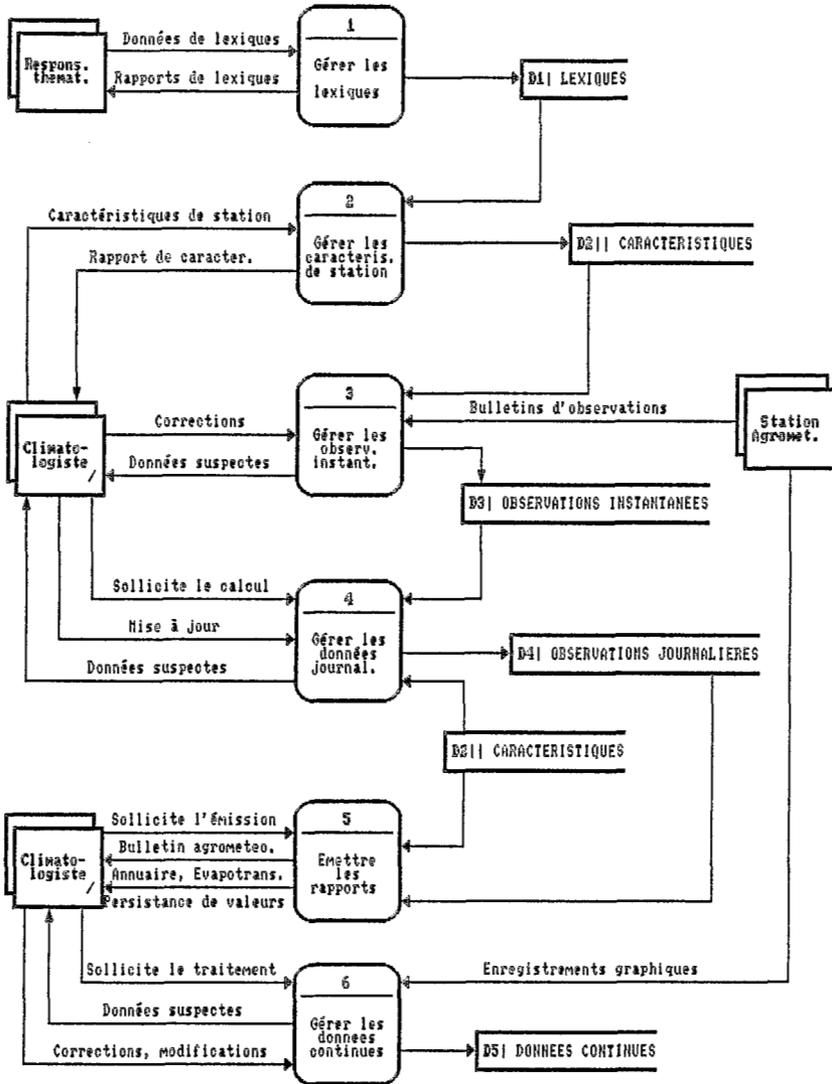


Fig.3 Diagramme de flot de données (niveau 0) de SISCLIMA

- la génération d'un fichier opérationnel de données journalières, sous une forme permettant diverses applications et disposant de moyens pour la reconstitution éventuelle de données (saisie directe des données journalières historiques) ;

- la production automatique de rapports standards, comme un bulletin agrométéorologique mensuel paramétrable ou l'annuaire par variable ;

- le calcul automatique de l'évapotranspiration potentielle, d'après plusieurs formules et l'émission de rapports de persistance de valeurs pour diverses variables, utilisés par les agronomes.

Bien que des pourparlers soient en cours avec l'INEMET, Instituto nacional de meteorologia, pour la constitution d'une base de données nationale (module central), il est actuellement jugé plus urgent de travailler à une nouvelle version de ce module local, encore peu utilisé en raison des difficultés dues à une interface obsolète. La version 3 de SISCLIMA, dont la réalisation est envisagée à court terme, devra compléter les facilités disponibles dans la version actuelle par :

- la gestion (dépouillement, validation et intégration) des diagrammes d'enregistrement continu de certains paramètres ;

- des ressources de visualisation graphique des séries chronologiques obtenues pour n'importe quel paramètre, permettant une critique a posteriori ;

- l'implémentation d'une interface avec un (ou plusieurs) progiciel(s) statistique(s) pour micro-ordinateur, de manière à offrir de nouvelles facilités d'extraction et de traitement de données à niveau local.

A l'occasion de ces nouveaux développements, le module local de SISCLIMA sera actualisé en fonction des standards les plus récents de convivialité, sécurité et intégrité propres à SISGEO.

4.2. SISFITO

Depuis le début des années 1980, le CPATSA s'est engagé dans un programme de constitution de collections botaniques, avec l'objectif d'identifier les ressources floristiques de la région du "tropicque semi-aride". Par ailleurs, le CPATU administre l'un des plus grands herbiers de la région amazonienne, avec plus de 130.000 exsicata : face à un tel volume d'informations, ses responsables ont fait plusieurs tentatives d'automatisation ces dernières années, sans succès.

La réalisation d'un travail en commun a ainsi été proposée fin-1985 par l'équipe de SISGEO, et a donné lieu à la réalisation du module local de SISFITO, Sistema de informaçoes florísticas, co-parrainé par ces deux centres de ressources.

La version 1.1, opérationnelle depuis juin 1989, a été spécifiquement conçue pour automatiser le maximum d'activités qui interviennent dans la gestion des collections botaniques, depuis la saisie des informations de la fiche de terrain jusqu'au traitement des consultations les plus courantes. Le diagramme de flot de données de la fonction "administrer l'herbier" de ce module, et qui fait l'objet de la figure 4, met en évidence les applications essentielles de SISFITO :

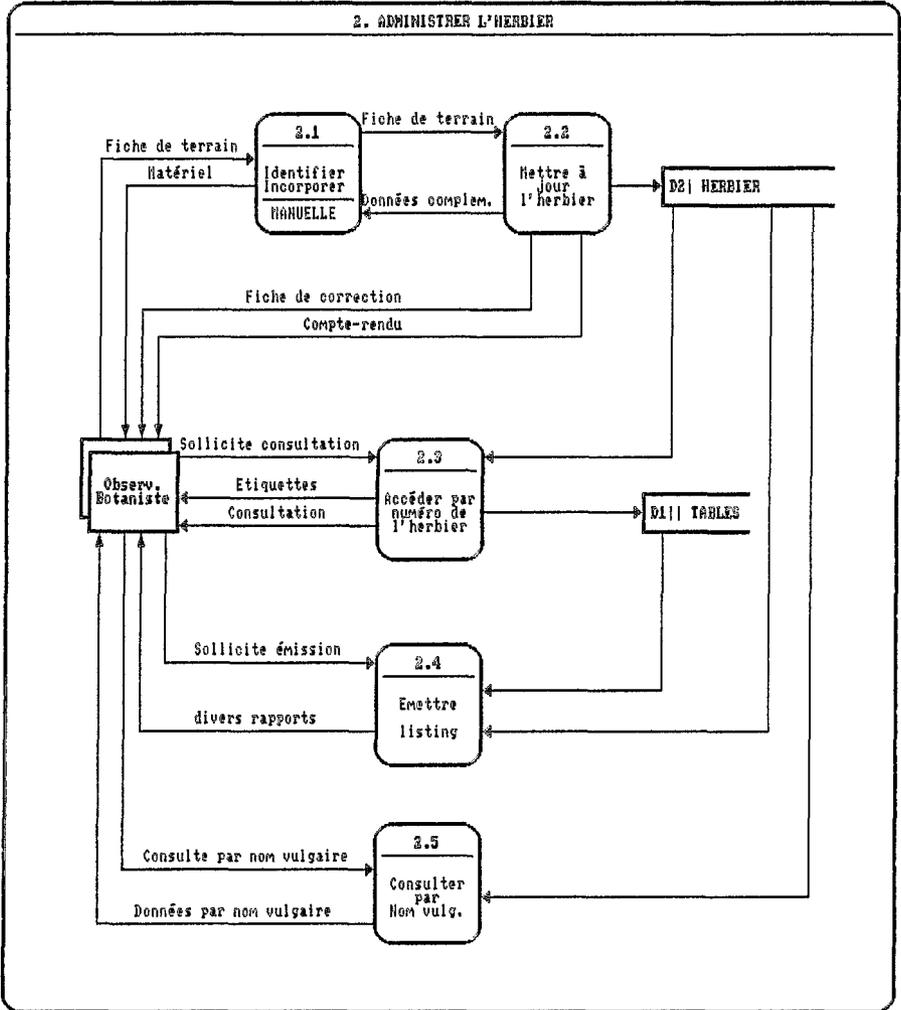


Fig.4 Diagramme de flot de données (niveau 1) de SISFITO

- mise à jour des données de l'herbier, avec possibilité d'inclusion, exclusion et modification des informations correspondant à chaque échantillon, et émission d'une fiche d'actualisation ;

- consultation et émission opérationnelles, c'est à dire par numéro d'échantillon dans l'herbier, permettant en particulier l'impression d'étiquettes ;

- consultations et émissions diverses, offrant la possibilité d'extraire les données de la base par le biais de plusieurs critères de sélection, comme famille, localisation géographique, nom vulgaire, etc..

Bien que réalisé sensiblement après le module local de SISCLIMA, la version actuelle de SISFITO souffre également d'une interface encore peu conviviale. Les premières utilisations de ce système ont aussi montré que la conception présentait quelques insuffisances, notamment en ce qui concerne l'absence de certains attributs, pourtant considérés comme essentiels. Il est donc prévu de réaliser à court terme la version 2 de SISFITO, pour corriger les quelques défauts et lacunes actuellement identifiés, de façon à créer un produit mieux apte à une ample diffusion parmi les unités de l'EMBRAPA.

4.3. SISOLOS

L'élaboration d'un système d'administration des données de sols collectées à l'occasion des divers inventaires pédologiques réalisés depuis les années 50 au Brésil, a fait l'objet d'un travail de coopération entre l'ORSTOM et l'EMBRAPA, par le biais du SNLCS, Serviço Nacional de Levantamento e Conservação de Solo, de 1981 à 1984 (SECHET 1985). La base de données ainsi constituée regroupe actuellement les informations correspondant à plus d'une centaine de levés, à diverses échelles, exécutés par diverses institutions du pays.

Ce système constitue ainsi le module central de SISOLOS, Sistema de informações pedológicas. Il stocke, sur les ordinateurs du site central de l'EMBRAPA à Brasília (deux IBM 4341 sous VM/CMS, en voie de substitution par un IBM 4381 biprocesseur, sous MVS), deux collections de données distinctes :

- les profils de sols, pour lesquels les informations qui se réfèrent aux conditions de son environnement, à la description morphologique et aux résultats des analyses physiques et chimiques de chaque horizon, ainsi qu'à l'analyse minéralogique de chaque fraction granulométrique, sont conservées ;

- les unités cartographiques de la carte des sols, constituées d'associations plus ou moins complexes de sols de différentes classes (les composantes). Pour ces unités, le système administre la description des caractéristiques de leur environnement et leur composition taxinomique.

Ces informations sont accessibles par le biais de routines d'extraction standardisées, dûment documentées, permettant l'édition de sorties normalisées ou le traitement, via le progiciel statistique SAS, Statistical Analysis System, des interrogations simples ou composées faisant intervenir n'importe quel attribut de profil ou d'unité cartographique.

Un certain nombre de difficultés sont néanmoins apparues, durant les quelques premières années d'exploitation de ce système :

- quant à l'accès physique à la base, en l'absence d'un véritable réseau de communication de données au Brésil ;

- quant à l'accès *logique*, dans la mesure où l'utilisateur potentiel doit posséder des connaissances minimales de pédologie, de statistique (utilisation de SAS), d'informatique (utilisation de VM/CMS) voire de communication (utilisation de KERMIT), difficiles à réunir pour une seule personne ;

- quant à la qualité des données stockées, compte tenu de la forme adoptée pour la capitalisation des informations (via bordereaux de codification remplis par des stagiaires), et en l'absence de possibilité conversationnelle de correction et mise à jour (la mise à jour utilisant normalement les procédures originales de traitement par lots).

Aussi, a-t-il été prévu de développer le module local de SISOLOS, dont la conception a débuté à la mi-1988 (avec l'orientation des pédologues de l'antenne régionale "Nordeste" du SNLCS, pour résoudre toutes ces difficultés, tout en facilitant simultanément l'ensemble des activités de bureau qui interviennent au cours de la réalisation d'un nouvel inventaire pédologique. Le diagramme de flot de données de niveau 0 du module local de SISOLOS est présenté figure 5.

Un premier prototype (version 1.0) est opérationnel depuis juillet 1989, et doit être complété d'ici la fin de cette année, en offrant toutes les facilités nécessaires à la gestion de toutes les informations afférentes aux profils de sols. Pour 1990, il est envisagé de mettre en oeuvre deux nouvelles applications :

- "*gérer les critères*", fonction originale de SISOLOS, dont l'objectif est de permettre la création et la mise à jour d'un ensemble de règles intégrées à la base de données, et permettant d'effectuer une critique plus apurée sur les données, mais aussi de les classer, d'élaborer automatiquement les classes de potentialités, etc. ;

- "*transférer les données*" qui permettra de faire passer des informations depuis le module local vers le central (capitalisation), mais aussi d'extraire les données de la base centrale pour les manipuler (correction interactive) avec le système sur micro-ordinateur. Cette fonction est bien sûr très importante pour l'ensemble de SISGEO : en plus du simple transfert physique, dont les modalités restent à définir compte tenu du contexte, l'opération doit s'accompagner d'un contrôle d'accès adéquat, ainsi que de routines de transformation, adaptation et agrégation-désagrégation des informations à transférer entre les deux modules.

4.4. SISPRAGAS

Il s'agit du module thématique le plus récent de SISGEO, mis en évidence tardivement pour intégrer les données phytosanitaires, très importantes pour le système productif agricole. C'est l'équipe d'entomologie agricole du CPATB qui a endossé la responsabilité de ce module, vu la prépondérance de cet aspect dans la région des "*basses terres de climat tempéré*", essentiellement consacrées aux rizières et à l'élevage, du sud du pays.

Une présentation exhaustive de la conception qui a été faite pour le module local de SISPRAGAS, *Sistema de informações fitosanitárias*, fait l'objet d'un article spécifique dans ce même séminaire (Séchet et alii, 1989).

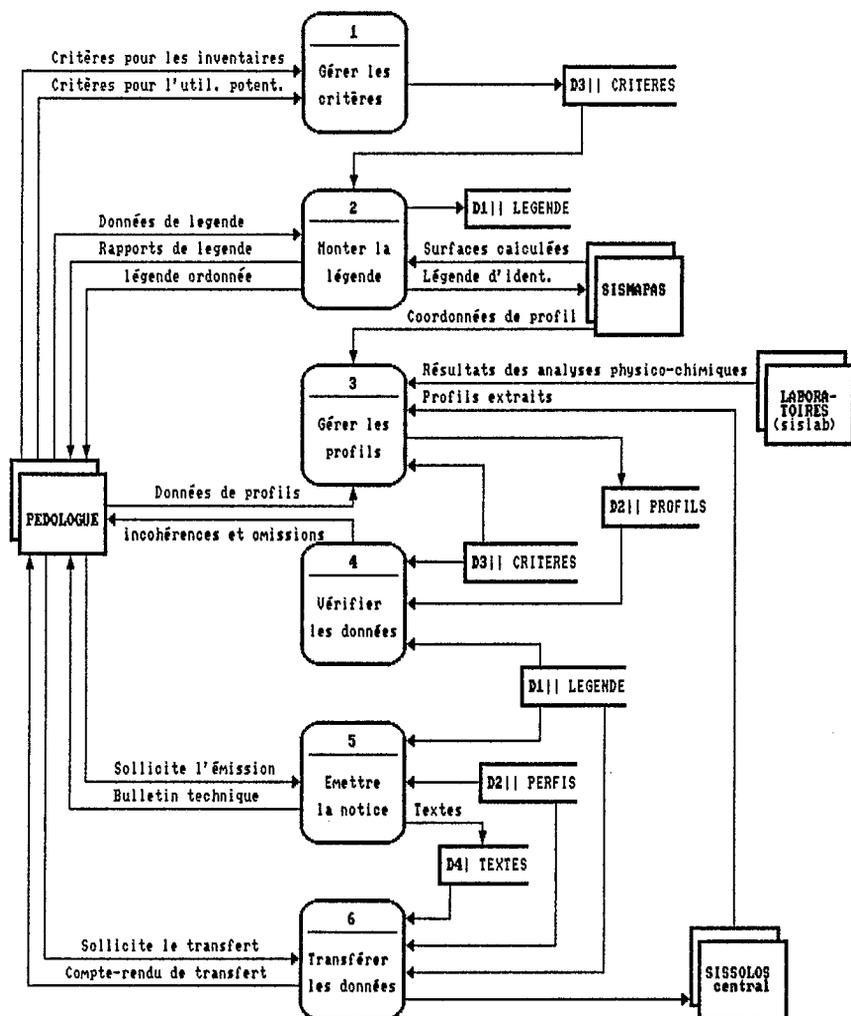


Fig.5 Diagramme de flot de données (niveau 0) de SISOLOS

La phase suivante de spécification, qui a récemment débuté, se limite actuellement à la réalisation d'une première version mettant en oeuvre l'application "Gérer les insectes". Des facilités sophistiquées de récupération d'information seront implémentées à cette occasion pour permettre la sélection des insectes, soit à partir de la dénomination populaire principale, soit encore à partir du nom scientifique abrégé (genre-espèce), de manière à respecter les habitudes traditionnelles des entomologistes. La présentation de cette version est envisagée pour la mi-1990 : si le schéma physique s'avère satisfaisant, des ressources analogues pourront être mises en oeuvre pour les variétés culturales.

4.5. SISECSO

Le module consacré aux données socio-économiques a été pris en charge par l'équipe d'économistes du CPAC, confrontée aux sérieux problèmes posés par la "région des savanes", qui connaît une forte expansion de l'activité agricole. Le travail de conception qui a été réalisé jusqu'ici prévoit le développement d'un logiciel destiné à systématiser l'obtention et faciliter la manipulation des informations économiques et sociales utilisées dans la recherche agronomique.

La version initiale du module central est prévue pour traiter seulement les données de recensement démographique, production agricole et élevage, au niveau de la commune, provenant de l'IBGE, Instituto brasileiro de geografia e estatística. La figure 6 montre le diagramme de flot de données de niveau 0 qui a été élaboré pour ce travail, lequel appelle les commentaires suivants :

- la deuxième application "Etablir les ensembles d'agrégation" doit permettre à l'utilisateur de définir, pour chaque extraction à effectuer parmi les données de l'IBGE, quelle règle d'agrégation doit être appliquée, en raison notamment des changements du découpage municipal intervenant entre chaque recensement ;

- une fonctionnalité appelée "désactivation" a également été conçue pour remplacer un ensemble d'informations par un autre plus récent, le premier étant du coup devenu obsolète.

Bien que tout le détail des DFD et de la description des éléments de données ait été réalisé courant 1988, la phase de spécification de ce module n'a pas encore pu débiter, faute de disponibilité de personnel (analyste de spécification). Une possibilité de reprise de ce travail est néanmoins prévisible d'ici la fin 1989.

4.6. SISHIDRO

SISGEO prévoyait dès son origine un module spécifiquement consacré aux ressources hydriques, baptisé SISHIDRO, Sistema de informações hidrológicas, lequel toutefois n'a jamais été prioritaire, compte tenu du peu de moyens dont dispose l'EMBRAPA dans le domaine hydrologique, malgré l'importance que revêtent ces paramètres pour l'agriculture, notamment dans certaines régions (Tropique semi-aride, par exemple). Les chercheurs du CPAP ont néanmoins manifesté leur intérêt pour la réalisation de ce système, vu leurs préoccupations vis à vis de la limnimétrie du Rio Paraná, qui conditionne l'état d'inondation ou d'exondation de tout le "Pantanal Mato-Grossense".

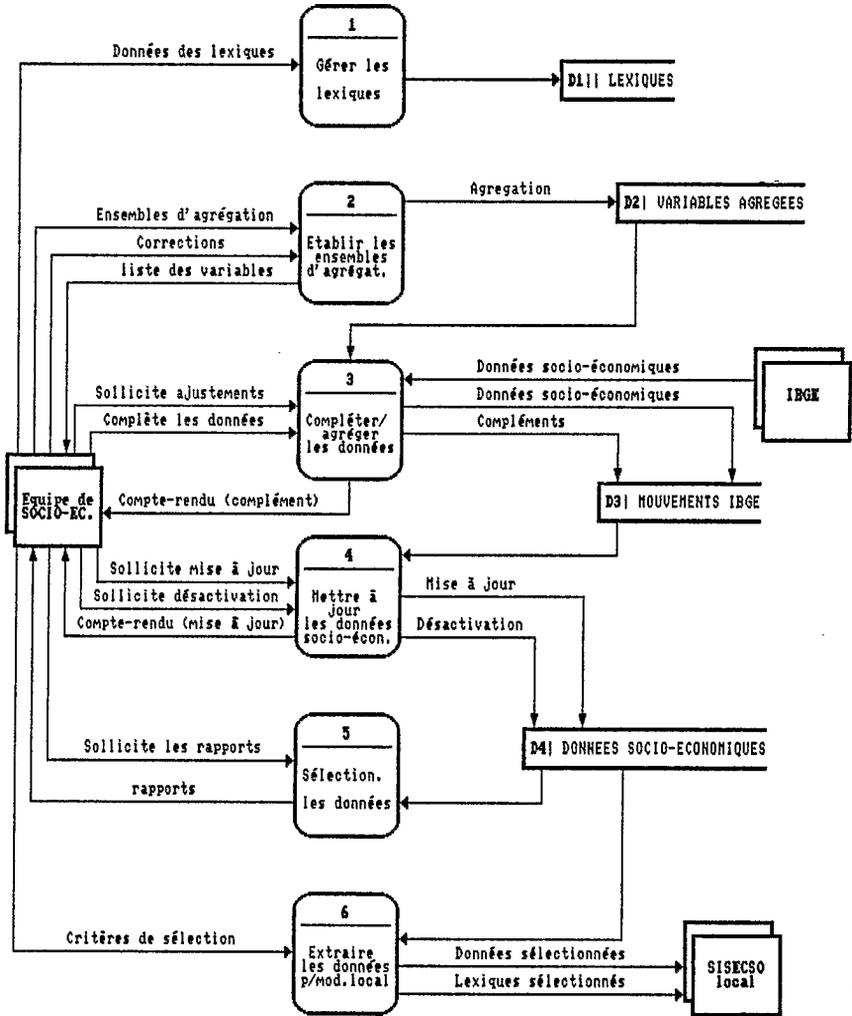


Fig.6 Diagramme de flot de données (niveau 0) de SISECSO

Dans l'état actuel des ressources disponibles pour gérer les informations hydrologiques de la région, essentiellement fournies par un réseau de stations maintenu par le DNOS, il est apparu que le logiciel de gestion des données hydrométriques mis au point par l'ORSTOM, dans son laboratoire d'hydrologie à Montpellier, serait satisfaisant, surtout depuis la parution d'une version en portugais.

Toutefois, la mise à disposition de ce logiciel a été conditionnée à la signature d'un accord particulier entre l'EMBRAPA et l'ORSTOM sur la propriété des logiciels, qu'ils soient développés en commun ou propres à l'une ou à l'autre des deux parties. L'élaboration de cet avenant s'est montrée particulièrement laborieuse et difficile, mais le problème semble avoir trouvé une solution définitive tout récemment, ce qui devrait permettre la libération d'HYDROM pour son utilisation comme prototype de SISHIDRO.

4.7. SISIMAGEM

Le découpage de SISGEO en modules thématiques autonomes a permis de mettre en évidence deux fonctions transversales essentielles, conduisant à la définition de deux systèmes méthodologiques :

- SISMAPAS, Sistema de processamento automático dos dados cartográficos, prévu pour manipuler les informations géographiques qui se présentent sous la forme de cartes polygonales, indépendamment du thème de référence ;

- SISIMAGEM, Sistema de processamento digital das imagens de satélites, pour mettre en oeuvre les ressources de la télédétection et les possibilités de mise à jour des informations géographiques, que celle-ci laisse entrevoir.

C'est à l'équipe du CPAC à Brasília qu'a été confié le soin de mener à bien la réalisation de ce dernier module. Celui-ci fait l'objet d'un projet spécial depuis plusieurs années, compte tenu en particulier de l'existence d'un équipement spécifique de traitement numérique des images (système RAMTEK et logiciel ERMAN II), acquis par l'EMBRAPA en 1983. La méthode d'approche qui a été adoptée pour assister, au moyen de techniques numériques, le télédécteur dans son exploitation des images de satellites depuis leur acquisition jusqu'à la sortie et l'interprétation des résultats du traitement, est largement commentée dans un autre article présenté à ce même séminaire (ASSAD et alii. 1989).

Le module central, actuellement en cours de réalisation, doit permettre l'administration d'un catalogue général d'images, l'utilisation intégrée sur le site central de plusieurs logiciels de traitement d'images, ainsi que l'extraction de sous-scènes susceptibles d'être manipulées sur des configurations de moindre coût, basées sur des micro-ordinateurs PC-compatibles (module local), par le biais des logiciels appropriés existants (SITIM, TIMOR, etc.).

4.8. SISMAPAS

Lors de la réalisation de SISOLOS, et avant que naisse l'idée de SISGEO, il est vite apparu nécessaire de compléter les fichiers de la base de données de sols par une représentation sur ordinateur de l'arrangement géographique des unités cartographiques de la carte des sols, sans laquelle il était impossible de répondre à la plupart des interrogations de nature géographique.

La difficulté particulière que constitue la modélisation d'une information spatialisée a été l'occasion d'une enquête approfondie sur les alternatives possibles, effectuée en France fin 1982 et relatée lors d'une communication à un congrès brésilien de télédétection (SECHET et alii. 1984).

La perspective de SISGEO a permis d'entrevoir une utilisation plus générale d'un module SISMAPAS, spécifiquement mis au point pour permettre la manipulation des documents cartographiques sur ordinateur, dans la mesure où l'on admet que les informations spatialisées de chaque thème peuvent être représentées sous la forme de cartes polygonales. Les possibilités offertes par un système de ce type ont d'ailleurs fait l'objet d'une description exhaustive (SECHET et alii. 1985).

La conception du module local de SISMAPAS a été naturellement confiée à l'équipe du SNLCS à Rio de Janeiro, qui comptait alors sur l'appui d'un ingénieur cartographe. Le projet logique qui a été réalisé au cours de l'année 1986 pour le premier module, correspond aux fonctionnalités présentées sur le diagramme de flot de données de la figure 7. La première version, disponible depuis avril 1988, implémente l'application "Contrôler la cartotheque". La seconde version, opérationnelle depuis le début de cette année, met en oeuvre également la deuxième application "Administrer les coupures", à partir d'une adaptation du logiciel MYGALE, mis au point par l'équipe d'infographie de l'ORSTOM, et que l'équipe de SISGEO avait décidé d'adopter à la suite de l'examen de plusieurs méthodes possibles (SOURIS et alii. 1985 ; SECHET et alii. 1986).

La solution définitive du problème de propriété des logiciels entre l'EMBRAPA et l'ORSTOM, permettra la diffusion de cette nouvelle version à quelques unités dotées des moyens matériels correspondants (table à digitaliser et traceurs de courbes, en particulier), en commençant par le CPAC, de manière à promouvoir au sein de l'entreprise la digitalisation de documents cartographiques, si possible selon un standard unique, celui proposé par SISMAPAS. Il sera alors intéressant d'envisager les alternatives disponibles pour le traitement (application "Traiter les cartes").

4.9. Utilitaires

Dans le contexte de SISGEO, où au moins deux ou trois projets sont menés pratiquement simultanément (dans la même phase du cycle de développement), et avec des perspectives d'interaction de l'ensemble sur le plan géographique au niveau de SISMAPAS, il est facile d'identifier dans chaque système applicatif :

- une partie indépendante de la nature de l'application, correspondant à un ensemble de fonctionnalités qu'il est nécessaire de mettre en oeuvre pour tous les modules développés, comme la gestion des menus, le contrôle d'accès, la gestion des copies de sécurité, l'administration des lexiques, etc. ;
- la partie propre au module en cours de réalisation, correspondant aux fonctionnalités spécifiques.

Pour les modules développés par l'équipe, qui ont tous comme point commun d'être orientés vers la capitalisation de l'information (génération de bases de données), la mise en oeuvre des fonctionnalités communes à l'ensemble des modules ne représente certainement pas moins de 50% de l'effort global nécessaire à la réalisation de chaque système ! Il s'est donc avéré judicieux de consacrer une partie des ressources disponibles à la réalisation des composants les plus couramment rencontrés dans les modules de SISGEO, et non disponibles commercialement, pour élaborer progressivement un "environnement de développement".

Le premier utilitaire qui a fait l'objet d'une implémentation spécifique permet l'administration des lexiques. SISTAB, Sistema de administração de tabelas de codificação, offre les fonctionnalités illustrées sur le DFD présenté figure 8. La version actuelle, opérationnelle depuis janvier 1989, est utilisée selon deux modes différents :

- dans sa version autonome elle permet l'administration des tables de codification de n'importe quel système en temps de conception, permettant ainsi une meilleure implication de l'équipe thématique dès le début du projet. Elle a ainsi pu être utilisée pour SISOLOS et SISPRAGAS ;

- dans sa version hôte elle gère les lexiques en temps d'exploitation du système une fois mis au point, en offrant notamment des ressources de visualisation des entrées disponibles pour tous les champs codifiés. SISTAB implémente ainsi la gestion des lexiques pour les versions actuelles de SISMAPAS et SISOLOS, et est d'ores et déjà utilisé dans d'autres systèmes (SISORC à l'EMBRAPA et MVR à l'ORSTOM).

Le contrôle d'accès, fondé sur l'utilisation de mots de passe conservés dans des lexiques spéciaux, a également été mis en oeuvre par le biais de SISTAB, offrant ainsi finalement un gain de temps appréciable pour la phase d'implémentation des modules de SISGEO (estimé à environ 15%).

Une autre fonction est commune à l'ensemble des systèmes de capitalisation d'information, et mérite une implémentation indépendante : celle de gestion des copies de sécurité. Aussi, un utilitaire spécifique a-t'il été défini dans cette optique : SISCOP, Sistema de gestão das cópias de segurança. Le projet logique qui a été élaboré fin 1988, prévoit la mise en oeuvre des fonctionnalités illustrées sur le DFD de la figure 9.

L'administration complète des aspects liés à la sécurité des fichiers (contrôle permanent de l'état de chacun des fichiers de la base, fonctions de copie et de restauration avec contrôle de la situation des volumes de "back-ups", sur disquette ou disque dur, etc.), est réalisée par le biais de deux fichiers particuliers, dits de "paramètres" et de "contrôle", à inclure dans chaque système-hôte.

Dans ces conditions, il a été jugé opportun d'associer dans SISCOP les fonctions propres à l'installation et à la configuration du système-hôte, permettant également d'entrevoir une économie substantielle de codification pour les modules à mettre au point. SISOLOS bénéficie déjà des facilités partiellement implémentées pour SISCOP, dont la première version complète doit être livrée dans les mois qui viennent.

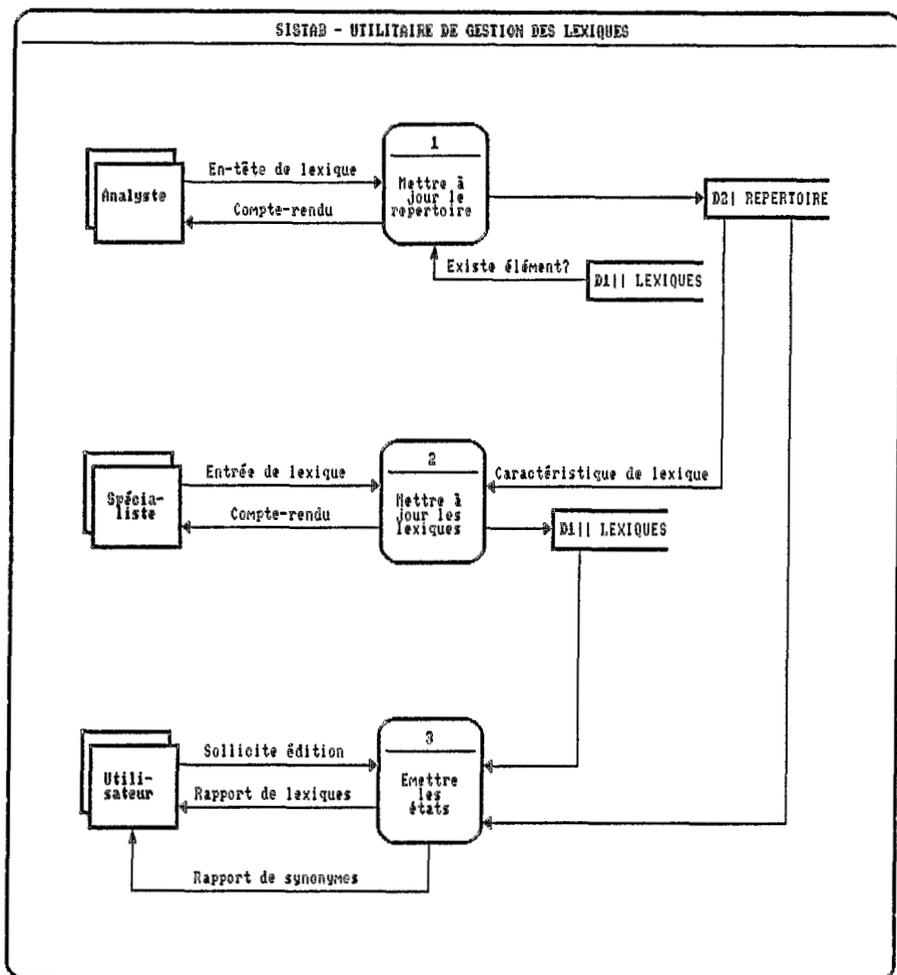


Fig.8 Diagramme de flot de données (niveau 0) de SISTAB

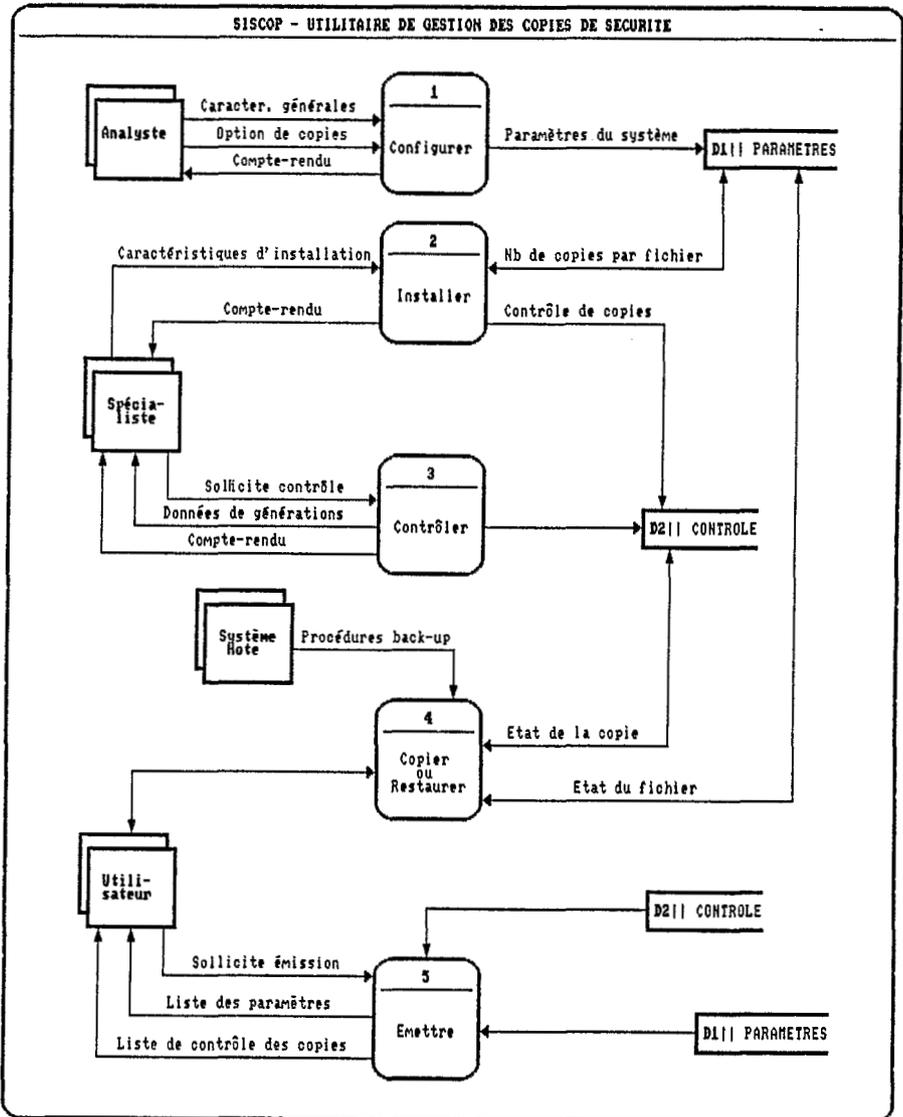


Fig.9 Diagramme de flot de données (niveau 0) de SISGOP

Le prochain élément de l'environnement de développement en cours de constitution pour SISGEO, concerne la réalisation "a posteriori" d'une aide en ligne. Une fonction de ce système, déjà partiellement implémenté (la saisie des textes d'aide) est utilisée pour faciliter la gestion des textes qui accompagnent les relevés sur le terrain, et dont au moins SISOLOS et SISFITO font un usage important.

De plus, d'autres développements ont dû être réalisés pour les besoins spécifiques du projet, lorsqu'il n'existait pas de produit commercial susceptible de les satisfaire. Ce fût notamment le cas en 1985-86, lors de la mise au point d'un gestionnaire d'écrans pour TurboPascal, depuis substitué par un produit commercial (TURBO SCREEN). Ce fût aussi le cas lors de la mise en oeuvre de SISVAR, Sistema de gestão de dicionário de variáveis, encore abondamment utilisé pour l'élaboration de tous les dictionnaires de variables du système, éléments importants des dossiers de conception. C'est enfin dans la même optique qu'il est envisagé de réaliser en 1990 le projet SISTRIBU, Sistema automatizado de distribuição de softwares, pour mettre en oeuvre un système rationnel de distribution des produits de SISGEO, avec un contrôle adéquat des versions et le suivi nécessaire des destinataires et utilisateurs. La figure 10 fournit une idée générale des fonctions qui seront mises en oeuvre par le biais de SISTRIBU.

CONCLUSION & PERSPECTIVES

Dans le domaine des systèmes d'information pour l'environnement, il existe à la fois :

- des bases de données monothématiques, disposant de ressources d'extraction et sélection d'information. Le plus souvent les données ne sont pas spatialisées ou, quand elles le sont, il s'agit de données ponctuelles ;
- diverses applications de traitement de données spatialisées, plus récentes et utilisant des moyens infographiques, certaines d'entre elles fournissant déjà d'excellents résultats.

Toutefois, ces dernières sont généralement réalisées avec des données spécifiquement organisées en fonction de cet objectif et ignorent l'entrée massive de données, ainsi que toutes les précautions qu'elle suppose et qui relèvent de l'informatique traditionnelle.

La caractéristique essentielle du programme SISGEO est de constituer une tentative de rapprochement entre les systèmes d'information constitués à l'initiative des chercheurs de chaque thème du milieu naturel et les nouvelles applications d'analyse géographique de l'environnement, permises par le développement des techniques infographiques.

Pour l'instant le système peut être utilisé comme un ensemble de fichiers de base : seule la totale intégration des données fournies par les modules thématiques dans un système capable de manipuler la spatialisation des informations permettra de tirer profit de l'ensemble des données accumulées, en accord avec les standards définis par les spécialistes de chaque discipline au moment de collecter et analyser les observations qui les concernent.

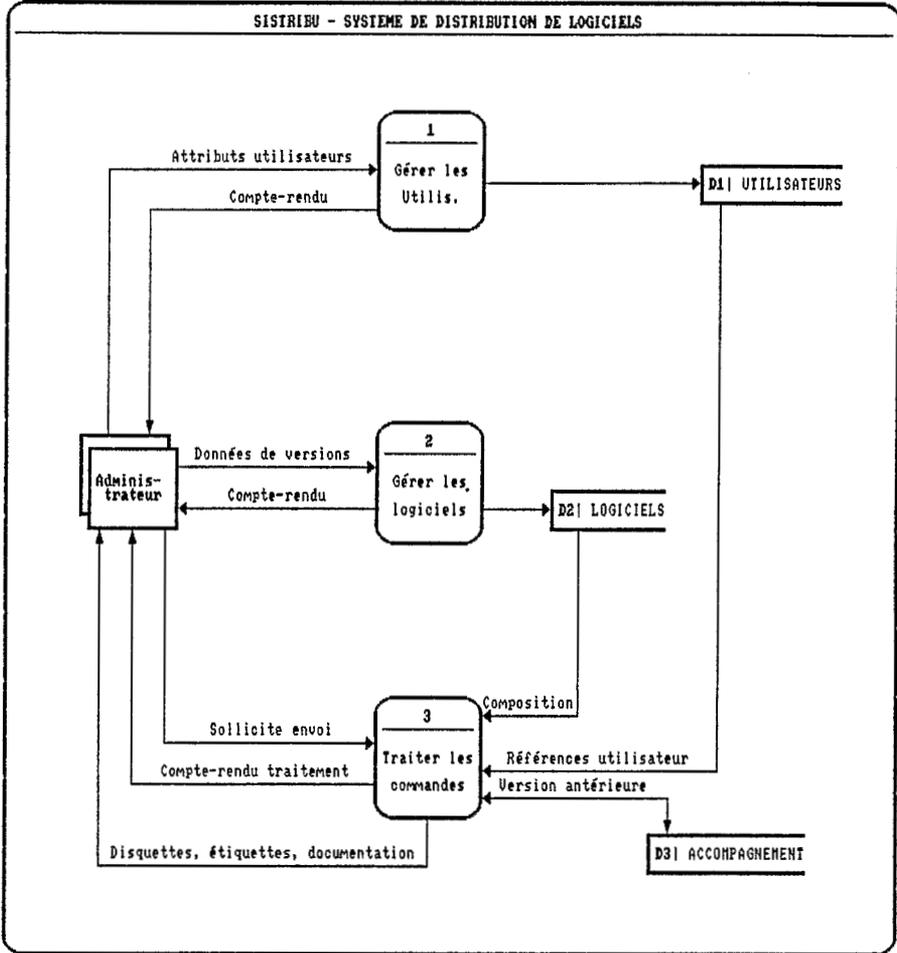


Fig.10 Diagramme de flot de données (niveau 0) de SISTRIBU

Un tel programme est évidemment très ambitieux et suppose qu'un certain nombre de précautions soient prises :

- il a fallu d'abord s'assurer que le cadre institutionnel de réalisation était adéquat à sa mise en oeuvre. Une étude faite en 1984 avait permis de comparer les chances de succès de tentatives semblables, menées par plusieurs institutions au Brésil (IBGE, SERPRO, Projeto RADAMBRASIL, INPE). On avait conclu que l'EMBRAPA, comme entreprise de recherche agronomique très diversifiée et disposant d'une solide compétence informatique était mieux placée pour mener à bien un SIG à niveau national. Les autres tentatives ont d'ailleurs toutes échoué depuis ;

- un effort important doit être fait au niveau de la coordination, pour assurer l'assise institutionnelle, garantir l'homogénéité des modules sur le plan méthodologique et surtout maintenir l'existence continue du programme, compte tenu notamment de sa durée exceptionnelle et de la difficulté d'obtenir les premiers résultats, surtout en regard de l'expectative créée ;

- un certain nombre d'activités secondaires doivent être menées pour contribuer à atteindre les objectifs essentiels de SISGEO. Ainsi le programme aura permis d'adapter la méthodologie de développement de systèmes, de définir des normes de programmation, de mettre au point des utilitaires, de choisir un environnement de développement et d'y former le personnel, d'élaborer un accord contractuel sur la valorisation des produits obtenus par la collaboration des deux organismes, etc.. Certaines de ces activités apparaissent trop souvent comme des déviations inutiles, alors qu'elles sont indispensables pour surmonter les difficultés rencontrées dans le déroulement du programme ;

- un équilibre doit être recherché entre la nécessité de ne pas abandonner certains modules du programme, sous prétexte par exemple du peu d'intérêt d'une équipe thématique pour le travail proposé, et la tentation de consolider les premiers résultats acquis, lesquels constituent généralement une avance significative par rapport à ce qui existe hors SISGEO. L'équipe s'est jusqu'à maintenant efforcée de mener de front l'ensemble des thèmes, en favorisant les modules locaux tout en maintenant une activité minimale au niveau central (SISIMAGEM, SISECSO). Ceci n'est pas toujours facile, compte tenu de l'absence de motivation parfois rencontrée chez les équipes thématiques pour ce type de travail, dont les résultats tardent à venir.

Bien entendu, le programme s'accompagne également d'un certain nombre d'aspects positifs sur le plan de la coopération entre les deux institutions, notamment grâce à la dimension des champs couverts et à sa position stratégique au siège du partenaire brésilien. L'exploitation des liens étroits ainsi créés entre les deux partenaires est susceptible de s'intensifier, pour leur bénéfice mutuel.

Avec SISGEO, l'EMBRAPA s'est résolument engagée dans une politique claire de valorisation des données du milieu rural, par le biais de la constitution de systèmes d'informations thématiques autonomes, mais intégrés dans une approche commune. Ce travail suppose à terme que l'on réussisse à entraîner les autres institutions, chargées de l'administration à niveau national des divers thèmes impliqués, par exemple en fournissant d'emblée le module local. Une telle démarche est en voie d'aboutir pour les données de climatologie avec l'INEMET.

Dans le cas de l'ORSTOM, il apparaît que l'accès aux données des pays partenaires va rapidement devenir impossible, d'autant que l'Institut se dégage volontairement des opérations de type inventaire, où il avait jusqu'à présent excellé. Sa participation à des opérations de constitution de bases de données, comme au Mexique, en Equateur, au Brésil, en Indonésie, etc., si elles sont bien formalisées, apparaissent ainsi comme de nature à inverser cette tendance. De plus, le savoir-faire dans ce domaine est peu répandu et précieux, ce qui suppose une consolidation des connaissances acquises à l'ORSTOM, mais encore trop dispersées : Séminfor devrait y contribuer.

REFERENCES BIBLIOGRAPHIQUES

- ASSAD, E.D.; MACEDO, J.; SANTOS, D.T.dos & JUNIOR, S.R. *SISIMAGEM : une philosophie d'utilisation de la télédétection à l'EMBRAPA*. In : Séminfor, 3., Paris, 1989, à paraître.
- CARDON, D.; SECHET, P.; DINIZ, T.D. & MALTEZ, M. *SISCLIMA : capitalisation de données agroclimatiques au Brésil*. In : Hydrol. continent., vol.2, n°1, 1987 : 3-14.
- EMBRAPA/DMQ *Memória do Primeiro Encontro de Métodos Quantitativos da EMBRAPA*, Brasília, 1982. 390p.
- NAGY, G. & WAGLE, S. *Geographic data processing*, In : Computing Surveys, 11(1), juin 1979, pp.139-181.
- SECHET, P. *SISSOLOS : base de données de sols du Brésil*. In : Cahiers ORSTOM, Série Pédologie, vol. XXI, n°4, 1984-1985 : 285-307.
- SECHET, P. *Le projet SISGEO (EMBRAPA, Brésil) : état d'avancement*. In : Traitement des données localisées : l'infographie à l'ORSTOM. Paris, 1986, Coll. & Sémin., ORSTOM, 1986.
- SECHET, P. *Méthode d'analyse structurée pour le développement de logiciels*. In : Séminfor, 1., Paris, 1987, Coll & Sémin., ORSTOM, 1988, pp.221-48.
- SECHET, P.; MIRANDA, J.I. & ASSIS, D.S. *Numerização de dados cartográficos : relato de uma experiência*. In : Simpósio Brasileiro de Sensoriamento Remoto, 3., Rio de Janeiro, RJ, 1984, Anais.
- SECHET, P.; ASSIS, D.S.; ARAUJO, A.R. & SAROLDI, M.J. *Uso da informática no manuseio e na exploração dos mapas de solos*. In : Congresso Brasileiro de Ciência do Solo, 20., Belém, Pa, 1985. Programa & Resumos. Belém, SBCS, 1985.
- SECHET, P.; LERISSON, D. & SANTOS, D.T. dos. *Implementação da captação de mapas poligonais usando mesa digitalizadora*. In : Congresso Brasileiro de Defesa do Meio Ambiente, 2., Rio de Janeiro, RJ, 1986. Anais. Rio de Janeiro, Clube de Engenharia, 1986, pp.195-212.
- SECHET, P.; MARTINS, J.F.da S.; BELARMINO, L.C. & LOECK, A.E. *SISPRAGAS : une approche de système d'information pour le contrôle des insectes phytophages*. In : Séminfor, 3., Paris, 1989, à paraître.

- SIMPLICIO, F. & SECHET, P. *Metodologia para elaboração de um sistema de informações geoambientais na EMBRAPA*. In : Congresso Brasileiro de geociências e tecnologias aplicadas aos estudos ambientais, 1., Rio de Janeiro, RJ, 1984. (non publié).
- SIMPLICIO, F.; MIRANDA, J.I. & SECHET, P. *Estratégia da EMBRAPA na realização de um sistema geográfico de informações ambientais*. In : Congresso Brasileiro de Defesa do Meio Ambiente, 1., Rio de Janeiro, RJ, 1984, Anais. Rio de Janeiro, Clube de Engenharia, 1984, pp.421-8.
- SOURIS, M.; MIRANDA, J.I. & SECHET, P. *Captação de mapas poligonais na mesa digitalizadora*. In : Congresso regional de informática, 7., Brasília, DF, SUCESU-DF, 1985. (non publié).

LE SYSTEME D'INFORMATION GEOGRAPHIQUE DE L'ETAT DE VERACRUZ (MEXIQUE) : POSITIONS DE RECHERCHE ET OPTIONS METHODOLOGIQUES.

(Luc CAMBREZY)

RESUME - On développera ici les circonstances qui font de ce programme un point de rencontre entre recherche et planification, aux trois niveaux d'approximations retenus : Etat, région, municipale.

- le constat d'une information statistique et cartographique abondante, bien que très inégale, peu utilisée et le plus souvent fort mal. Les caractères socio-économiques d'un espace quelconque (Etat, Région ou Municipale) comme les aspects relatifs au support écologique sont rarement situés dans leur environnement spatio-temporel. Il s'ensuit que la plupart des études ayant pour finalité la planification du développement des activités agricoles, se résument à un catalogue sans intérêt des variables habituelles : hydrographie, végétation, climats, population, etc. ;

- dès lors, le SIG apparaît comme un outil permettant de lever l'habituel goulot d'étranglement que constitue la gestion et le traitement de l'information. Il apporte un plus essentiel, trop peu pris en compte au Mexique (mais les planificateurs ont conscience de cette lacune) : la mise en perspective dans l'espace des phénomènes étudiés ;

- cette perspective, comme celle qui lui est liée, l'analyse en terme de synthèse, est aussi celle des chercheurs - au moins de ceux pour qui l'espace est une variable incontournable. Elle l'est à fortiori lorsque l'étude des organisations régionales est au centre de la problématique scientifique des disciplines impliquées dans le programme : démographie, bio-géographie, géographie, etc..

Mise en place du SIG : aspects techniques et institutionnels.

On évoquera ici l'ensemble des problèmes que posent la mise en place d'un SIG dans un pays du Tiers Monde : depuis la question du transfert de technologie (tous les pays ne sont pas aptes à le recevoir), jusqu'à l'infrastructure (équipement et personnel) qu'exige un tel programme. La nature même de l'outil, comme la liaison Recherche/Planification, contribue à orienter ce type de programme dans une course à l'information, toujours plus fiable et toujours plus actualisée.

Cette spirale, difficilement évitable, impose la mise en place d'un important dispositif de collecte et de correction de l'information. Ceci implique aussi la participation active de nombreuses institutions (agriculture, santé, etc.) ; celle-ci est coûteuse en personnel, mais implique surtout une adhésion totale des instances politiques qui seules peuvent favoriser cette nécessaire coopération.

Problématique générale et thèmes d'étude.

Au delà des thèmes d'étude qui ont été retenus, mais un SIG est par nature dynamique, on insistera surtout sur un aspect essentiel de la recherche. S'agissant d'un nouvel outil, les SIG autorisent la mise au point et l'emploi de nouvelles méthodes, difficilement envisageables sans les techniques liées à l'informatique graphique ; c'est par exemple le cas pour l'étude du milieu physique et de ce que les sociétés agraires en font. Ainsi l'exploitation exhaustive de l'information contenue dans la carte au 1/50.000 (altimétrie, hydrographie) doit permettre un renouvellement des études de modelé et de l'hydrologie de surface. La substitution des MNT - modèles numériques de terrain, aux classiques cartes de pentes en est un exemple.

L'échelle.

Les problèmes d'échelle se posent à toutes les étapes de la mise en place d'un SIG : depuis la saisie des données jusqu'aux produits finals. La question de l'échelle est ici accentuée par l'étendue de la zone d'étude (72000 km²), la nécessaire précision des données (on saisit toujours l'information la plus désagrégée), et la variabilité de la demande. On envisagera :

- nature des données, produits attendus et échelle de saisie ;
- planification régionale/planification nationale : une nécessaire coordination.

Les SIG face au problème de la centralisation.

- le problème de la participation paysanne aux processus de planification de la production agricole ;
- la recherche sous les fourches caudines de la demande à usage politique ?
- l'actualisation permanente : la clef du succès ?

INTRODUCTION

A moins de vouloir créer une base de données dont la seule certitude serait qu'elle ne sera jamais d'aucune utilité, on ne se lance pas dans la mise en place d'un tel outil sans de sérieuses motivations. Qu'elle soit destinée à la planification ou à la recherche plus fondamentale, la création d'une base de données répond en premier lieu aux contraintes posées par le stockage et la gestion d'un volume important de données ; données qu'il faudra certainement actualiser et dont le volume ne peut aller qu'en augmentant.

Le choix du système informatisé permettant la gestion de cette base doit être guidé par l'utilisation qu'on souhaite en faire comme par le mode de représentation des connaissances désiré. On rappellera donc brièvement les objectifs du programme de recherche initié par l'ORSTOM au Mexique.

1. OBJECTIFS ET PREMIERS RESULTATS

1.1. Objectifs

Depuis mai 1989, une équipe de chercheurs et ingénieurs de l'ORSTOM travaille à la mise en place d'un système d'information géographique pour l'état de Veracruz (Mexique). Comme pour l'Atlas informatisé de la ville de Quito, le choix s'est porté sur l'utilisation de l'ensemble de techniques (matériels et logiciels) développées par l'atelier d'infographie de l'ORSTOM : Savane, Tigre et Mygale. Ce programme de recherche fait l'objet d'une convention spécifique entre l'ORSTOM et l'INEGI (*Instituto Nacional de Estadísticas, Geografía e Informática*). Le Mexique, plus de deux millions de km², 80 millions d'habitants, est administrativement divisé en 32 états. L'état de Veracruz comporte 207 municipios, pour une superficie d'environ 71 000 km².

L'INEGI est une institution comparable à l'IGN et l'INSEE, dépendant du ministère du plan et du budget. Cet institut est responsable de la cartographie topographique à différentes échelles, comme de la cartographie d'inventaire classique (géologie, pédologie, végétation, climats, etc.). Par ailleurs, l'INEGI est également maître d'oeuvre des divers recensements régulièrement levés tous les dix ans : population, agriculture, économie. Le choix de l'état de Veracruz se justifie par une bonne connaissance de cette entité de la part de l'ORSTOM (de nombreux chercheurs y ont travaillé), comme par l'intérêt manifesté pour les techniques liées à l'informatique graphique par les diverses instances de planification, fédérales ou de l'état, présentes à Xalapa, capitale de l'état de Veracruz.

S'agissant de tout l'état de Veracruz, on voit que le travail de digitalisation et de capture des seules données produites par l'INEGI est considérable. La convention est d'une durée de trois ans, mais la première année a été essentiellement consacrée au montage de l'opération ; il est donc encore trop tôt pour pouvoir dresser un bilan des résultats.

Le choix du système se justifie évidemment par les résultats attendus. Une bonne partie de l'information disponible étant de nature cartographique, le choix devait s'orienter vers un outil du type système d'information géographique (SIG). Ce système, inspiré des bases de données relationnelles, inclue la localisation et gère celle-ci comme n'importe quel attribut. S'il est possible d'obtenir le résultat d'une interrogation de la base sur imprimante, l'intérêt essentiel d'un SIG consiste à pouvoir visualiser sur carte (sur écran ou traceur) le résultat de la requête ; c'est la principale différence avec les bases de données "traditionnelles". Cette option trouve une justification naturelle en géographie, qui fait de la question suivante l'essentiel de sa recherche : pourquoi là et pas ailleurs ? Pourquoi là plus qu'ailleurs ? On touche là une des questions posées dans ce Séminfor : celle de la représentation des connaissances.

Le système TIGRE permet une production rapide, simple et interactive de cartes thématiques dans la projection souhaitée. On peut augmenter ou réduire à loisir le nombre de paliers graphiques, calculer des surfaces, des rapports, des

pourcentages entre attributs, bref, on s'épargne de longues heures de travail autrefois passées à colorier de nombreux brouillons de cartes "pour voir ce que ça donne", étape préalable à une éventuelle mise au propre.

Le premier intérêt du système, le plus séduisant aussi pour le néophyte, relève donc de ce qu'il est convenu d'appeler la cartographie automatique, encore que ce vocable recouvre une réalité bien différente selon le matériel utilisé.

Mais lorsqu'on travaille sur les localisations et les combinaisons de phénomènes dans l'espace, on se rend bien vite compte qu'il convient de multiplier les découpages de l'espace. La carte a ses vertus, mais il s'agit toujours d'une schématisation de la réalité permettant de rendre celle-ci intelligible. Dans cette perspective, on ne peut avoir de découpage *a priori* de l'espace. C'est en principe connu, mais on oublie pourtant trop souvent que les organisations spatiales, dotées d'un statut propre, n'ont rien à voir avec les délimitations territoriales mises en œuvre et utilisées pour l'élaboration des données.

Le cas le plus évident est celui de la cartographie des données socio-économiques collectées dans le cadre de la division politico-administrative en vigueur. Celle-ci permet sans doute une bonne approche du problème, mais reste forcément réductrice si l'on veut bien admettre que les distributions spatiales rencontrées, notamment dans les pays du Tiers-Monde, seront beaucoup mieux perçues au travers d'autres partitions sans rapport avec la division administrative en vigueur (milieu naturel, systèmes agraires, unités agro-écologiques, systèmes de relation, aires culturelles, etc.).

Cette approche suppose que la capture des données soit, dans la mesure du possible, toujours réalisée au niveau de désagrégation le plus fin. Le système doit ensuite permettre de regrouper ces données selon la partition spatiale souhaitée : des données de population traitées en fonction d'un découpage fondé sur le milieu naturel (bassin versant, unité morpho-pédologique, etc.), des statistiques agricoles ré-agglomérées en fonction d'une division culturelle, ethnique ou linguistique, des données relatives au support écologique interprétées en fonction d'une cartographie des systèmes agraires.

1.2. Premiers résultats

1.2.1. Le cadastre rural

Au Mexique, la tenure de la terre se partage en deux grandes rubriques : la propriété privée d'une part, et la propriété sociale (parfois collective) résultant de la réforme agraire, d'autre part.

Dans le cadre d'une convention entre le "Secretaria de la Reforma Agraria" et l'INEGI, nous avons accès à l'intégralité du cadastre rural au 1:50 000 de l'état de Veracruz. Après de nombreux mois de vérification et de correction de l'information, nous avons pu procéder à la digitalisation de toutes les terres relatives à la propriété sociale (la moitié de la superficie de l'état de Veracruz, quelques 5 000 polygones). La reproduction sur support stable est un préalable indispensable à une digitalisation fiable et sans problème au moment de la jointure des feuilles. Celle-ci a été facilitée par la mise à disposition, au Se-

crétariat de la réforme agraire, d'un système informatisé de cartographie automatique (système HP incompatible). On a donc pu procéder à la cartographie au 1:100 000 de tout l'état - 110 cartes au 1:50 000. La forme très géométrique des polygones autorisait cette réduction d'échelle, et ainsi, limitait le nombre de feuilles (27 au 1:100 000).

Le second préalable consiste en l'attribution des clés de zone qui seront saisies au clavier au moment de la digitalisation. Dans Mygale (logiciel de digitalisation), tout polygone comporte un numéro de zone et une clé ; cette dernière est utilisée au moment de l'intégration des données descriptives. Deux zones peuvent avoir la même clé : deux affleurements géologiques de même nature, ou, dans le cas qui nous occupe ici, un même *ejido* dont les terres sont distribuées en plusieurs parcelles. Dans le cas de la propriété sociale, l'attribution des clés a été dictée par l'existence d'une enquête récente réalisée par l'INEGI : l'*Encuesta nacional agro-pecuaria ejidal*, de novembre 1988. L'intégralité de cette enquête a été saisie sur gros systèmes par les services informatiques centraux de l'INEGI. On a donc repris la même codification : clé du *municipio*, suivie de la clé de l'*ejido*.

Après la digitalisation, les deux étapes importantes sont celles de l'intégration dans une relation, du graphique (les cartes), et des attributs (les variables de l'enquête). Tigre permet l'ouverture d'autant de relations que l'on souhaite (tenure de la terre, démographie, pédologie, etc.). Chaque relation autorise le stockage et le traitement d'un maximum de soixante attributs.

1.2.2. Le traitement du recensement de population de 1980

Dans un pays à la démographie si dynamique, on peut s'interroger sur la pertinence du traitement d'un recensement aussi ancien. celui-ci se justifie pourtant dans la perspective d'une comparaison (graphique, bien sûr) avec les résultats du prochain recensement de population de 1990, auquel on se prépare déjà.

Il reste que le traitement de ce type de données pose quelques problèmes. Pour commencer, ce recensement n'a pas très bonne réputation quant à la fiabilité des données, et bien entendu celles-ci ne sont plus vérifiables. Par ailleurs les résultats ont été publiés au niveau des *municipios*, ce qui ne permet aucune analyse fine de la diversité régionale.

Le fait d'avoir l'INEGI pour partenaire permet bien sûr d'obtenir les données à un niveau plus fin (aire minimale de recensement ou localité), mais le manque de fiabilité des données n'en est que plus flagrant et oblige à de nombreux lissages. Enfin, la division politique elle-même reste sujette à caution ; bien des limites municipales sont encore incertaines et, dans tous les cas, on a pu constater à travers l'analyse de la situation foncière, que la réalité tant sociale qu'économique, n'est pas toujours en rapport, loin s'en faut, à la division politique en vigueur (division fédérale, division municipale).

1.2.3. Les cartes de population

Fidèle à la tradition de l'école géographique française, on considère que la carte de population par points de surface proportionnelle, et plus encore, l'évolution dans le temps de ces densités, est une étape incontournable de toute analyse géographique sérieuse.

L'intervention du thématique se résume à l'analyse de la validité des données, et à la décision qui en résulte. réaliser la carte ou en abandonner l'idée. Nous en sommes à cette étape. Les premiers essais, limités au centre de l'état, montrent l'extrême facilité de la procédure ; il suffit de lire le fichier des localités doté des attributs souhaités (population, coordonnées géographiques, etc.), de l'intégrer dans Tigre comme un fichier de données ponctuelles (ce pourrait être des stations climatologiques ou des profils de sols), et de fixer la taille minimale et maximale des cercles représentant la population.

Ces cartes (1980 et 1989) feront l'objet de développements particulièrement intéressants dans les prochains mois : mise à plat des densités par modèle d'interpolation, cartes de densités rurales, aires théoriques de rayonnement des villes, etc..

1.2.4. La saisie des cartes d'inventaire et l'analyse du milieu naturel

L'INEGI a réalisé une série de cartes thématiques couvrant, à certaines échelles, tout le pays (1:1 000 000 et 1:250 000). La géologie, la pédologie, les climats, précipitations et températures, l'usage des sols et la végétation sont les principaux thèmes de cette cartographie d'inventaire. Cette cartographie est très inégale ; la carte des sols est jugée très convenable, alors que la carte d'usage des sols et de végétation livre une information de qualité sur la végétation naturelle, mais bien peu de choses sur l'usage des sols. La digitalisation de ces cartes (une douzaine par thème) a débuté en août.

On sait que les types de sols, le modelé, les précipitations, bref, les données relatives au milieu naturel, dont on connaît l'importance pour l'agriculteur, interviennent de façon importante dans le niveau d'efficacité des systèmes agraires et des unités de production, comme dans la possibilité, théorique ou réelle, de supporter des charges de population qui vont en augmentant malgré l'exode rural.

Dans cette perspective, c'est moins le support écologique, comme objet d'étude, qui nous intéresse que la façon dont les sociétés agraires l'investissent, l'interprètent et l'exploitent.

C'est moins le substratum géologique et la pédogenèse qui importent, que ce que l'agriculteur retire du sol, en préservant, ou non, la fertilité - c'est moins le total annuel des précipitations, que la répartition des pluies dans l'année - c'est moins une recherche pointue sur les systèmes d'érosion hérités ou en vigueur, que la topographie qui en résulte qui, elle, est quotidiennement parcourue par l'agriculteur.

Par ailleurs, les cartes topographiques de l'INEGI au 1:50 000 livrent une information essentielle pour l'interprétation des faits agraires : la topographie et le réseau hydrographique permettent une lecture quasi-immédiate des problèmes

que peuvent poser les pentes, le modelé, l'accessibilité, les points d'eau, etc.. Il faut donc profiter à la fois de la richesse de cette information et des possibilités de calcul qu'offre l'ordinateur. Une demande spécifique du Ministère de l'agriculture de l'état nous a conduit à réaliser un MNT au 1:50 000 du Cofre de Perote (volcan dominant la ville de Xalapa, capitale de l'état de Veracruz). On souhaite cependant confronter cette démarche, qui pour être séduisante, n'a pas encore fait toutes ses preuves, à la carte classique des pentes. Dans les deux cas, l'idée, qui n'est plus très neuve, est de superposer de l'information thématique (usage des sols, par exemple) à la carte des pentes comme au MNT. ce développement sera sans doute réalisé dans les prochains mois.

1.2.5. La cartographie des systèmes agraires

C'est sur ce thème que nous percevons les meilleures possibilités d'usage de l'image satellite, surtout si celle-ci est intégrée au système d'information géographique ; une manipulation qui est précisément autorisée dans Tigre.

Cependant, un premier travail réalisé dans le cadre d'un PEPS (Programme d'évaluation préliminaire SPOT) a au moins débouché sur le résultat qu'il était illusoire d'espérer en tirer une carte d'usage des sols, sans la mise en place d'une logistique coûteuse (achat de nouvelles scènes, personnel qualifié à plein temps, etc.) qui, au bout du compte, n'assure pas plus d'une cartographie fiable.

Enfin, on peut s'interroger sur la validité d'une carte d'usage des sols qui serait le cliché fidèle d'une situation à un moment donné (la date de prise de vue), mais qui ne dirait rien des successions culturales, des rotations, des jachères lorsqu'il y en a, mais surtout des dynamiques rapides observées dans les campagnes liées au crédit, à l'évolution des cours des produits agricoles où à l'évolution du réseau routier.

Ces premiers travaux nous orientent donc vers la recherche d'une procédure qui serait celle d'une cartographie des paysages et systèmes agraires fondées sur les critères visuels classiques en photo-interprétation. A cet égard, l'image SPOT panchromatique s'avère d'un grand intérêt. Il reste que la couverture complète de l'état de Veracruz signifierait l'acquisition d'une vingtaine d'images SPOT, ce qui ne peut relever des attributions de l'ORSTOM.

2. LES TECHNIQUES

2.1. La description du système

On ne décrira ici que l'architecture et la philosophie générale du système Tigre. Pour de plus amples détails, nous renvoyons le lecteur aux diverses publications de Marc Souris et al.

"L'information géographique est complexe et variée : données multiples, hétérogènes, de sources très diverses (cartes, télédétection spatiale, relevés de terrain, données statistiques, etc.). Ces données n'ont souvent aucun lien a priori

entre elles, autre que leur localisation dans l'espace, qui peut être donnée sous des formes très diverses". L'originalité du système a été "d'étendre le principe de la gestion relationnelle des objets, que ce soit des zones, des lignes, des réseaux, des points, et indépendamment de leur mode de stockage et représentation". Enfin, "si la localisation est utilisée dans le processus de gestion de l'information, elle permet également la représentation cartographique des résultats d'une interrogation au même titre qu'une sortie des résultats sur imprimante. C'est l'attribut *localisation* qui est alors représenté comme résultat du processus de gestion.

2.2. La nature des paramètres à stocker

Les données graphiques peuvent être des points (localisation des stations météorologiques, par exemple), des réseaux (hydrographie, routes, etc.) ou des zones (affleurement géologique, parcelles, etc.). Les localisations ponctuelles peuvent être saisies au clavier, si les coordonnées géographiques sont connues ; dans les autres cas, la saisie se déroule sur la table à digitaliser.

Dans la mesure du possible, l'information doit toujours être saisie au niveau le plus fin. A la digitalisation de la carte des précipitations, on préférera la saisie des données qui ont servi à son élaboration ; à la capture des données censitaires par municipio, on privilégiera la saisie des mêmes données pour chaque localité. Comme on l'a déjà mentionné, c'est cette désagrégation de l'information qui permet de s'affranchir d'un découpage a priori de l'espace.

2.3. La configuration matérielle

La configuration installée au Mexique est la suivante :

- pour la digitalisation : table Benson 6301 + micro PC-AT (disque dur 20 mégas, carte EGA) ;
- pour le traitement : station SUN 3/60, écran couleur 19 pouces, streamer et terminal alphanumérique en complément ;
- pour la restitution : traceur CALCOMP 8 plumes. Pas de restitution couleur, pas d'imprimante laser.

2.4. Problèmes rencontrés

On n'entrera pas dans le détail des problèmes rencontrés, tant ils sont nombreux. On se limitera ici à leur énumération : changements de matériel, communication entre la France, l'Equateur et le Mexique, formation du personnel, accès aux données, validation, actualisation, etc.. Deux questions méritent cependant une attention particulière : la restitution cartographique d'une part, le développement informatique d'autre part.

On a vu en effet que la carte est le résultat privilégié de toute interrogation de la base. S'il s'agit d'une cartographie de travail, dont l'objet est avant tout de produire des documents utiles pour la réflexion, on peut sans doute se satisfaire d'un traceur pour le dessin au trait et d'une impression type *Tetha-Scan*, pour les à-plats et trames en couleur.

Ces impressions sont en revanche de qualité tout à fait insuffisante s'il s'agit d'édition. L'impression sur imprimante laser noir et blanc est très satisfaisante mais ne règle pas le problème de l'édition des cartes en couleurs de grand format. Quant aux périphériques du type Vizir ou Intergraph, leur prix reste encore très prohibitif.

La question du développement informatique constitue un autre goulot d'étranglement. Une des grandes qualités du système Savane (Mygale, Tigre, etc.), est d'être perfectible ; c'est d'ailleurs le souhait des utilisateurs qui, à mesure qu'ils progressent dans la manipulation de l'outil comme dans leur réflexion, se trouvent soudain bloqués dans l'usage qu'ils pensaient faire du système. Etant perfectible, le problème n'est pas insurmontable, mais cela implique la présence sur place d'un membre permanent de l'équipe d'infographie de l'ORSTOM.

ILLUSTRATIONS

L'intégration du graphique constitue une étape essentielle du processus dans la démarche de création du système d'information géographique. Pour éviter les problèmes ultérieurs de recalage, un fond de carte de référence unique est indispensable (tracé des côtes, limites d'états, routes, etc.). En fonction des thèmes et de l'information disponible, la digitalisation s'effectue par feuilles au 1:50 000, 1:100 000 ou 1:250 000. C'est au moment de l'intégration du graphique dans la base qu'on réalise la jointure des feuilles (figure 1).

La figure 2 représente un modèle numérique de terrain du Cofre de Perote, vu de la Sierra (2.1) et vu de l'Altiplano (2.2). L'utilité des MNT est fonction de l'échelle de saisie. Au 1:250 000, celui-ci n'a qu'une valeur illustrative et pédagogique ; à l'inverse, une digitalisation des courbes de niveau au 1:50 000 permet de mieux comprendre la topographie de détail et justifie la superposition d'une information variée.

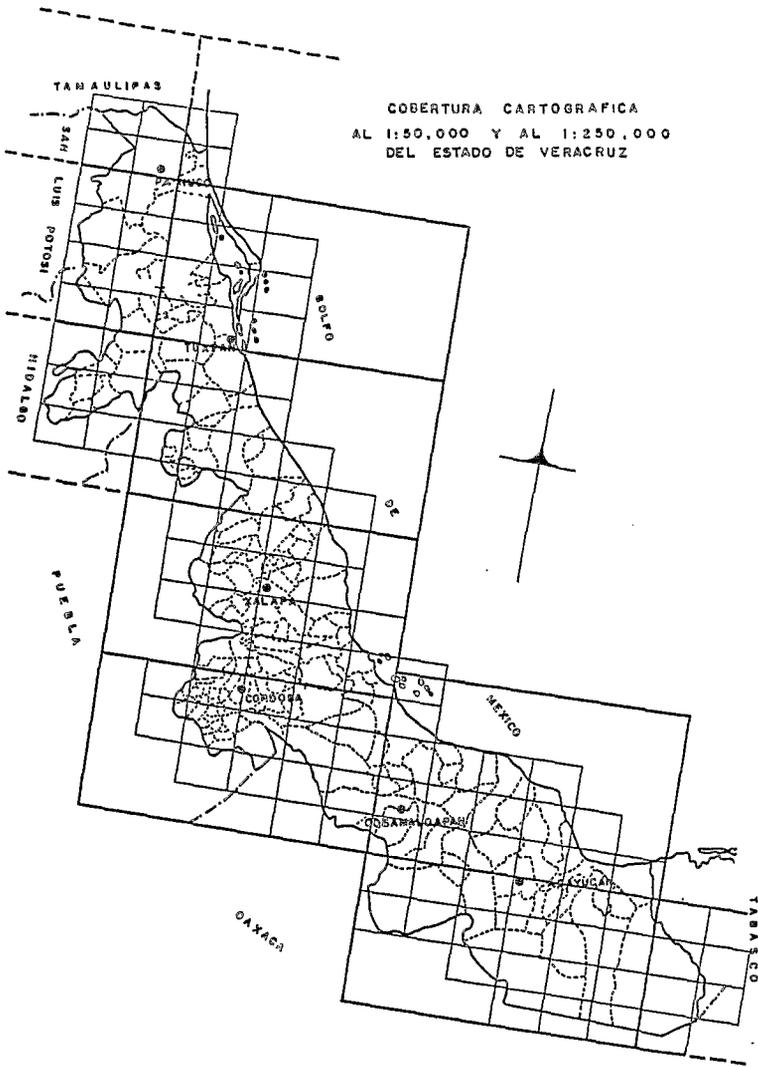


Figure 1

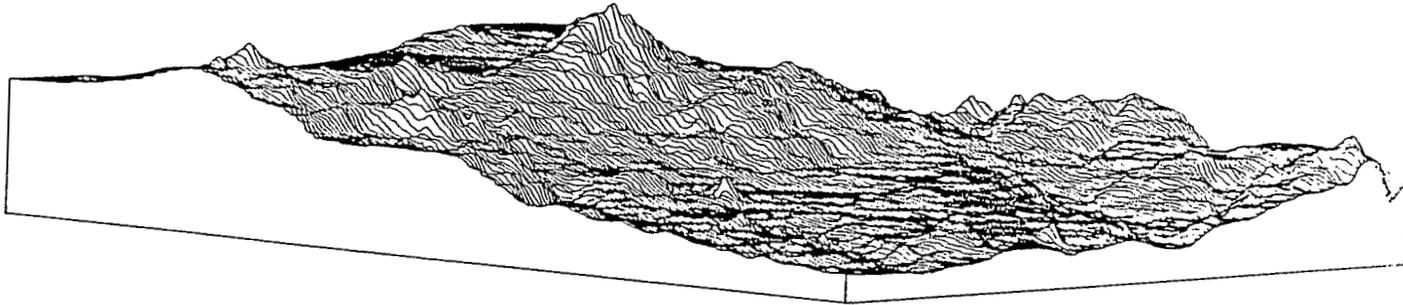


Figure 2.1

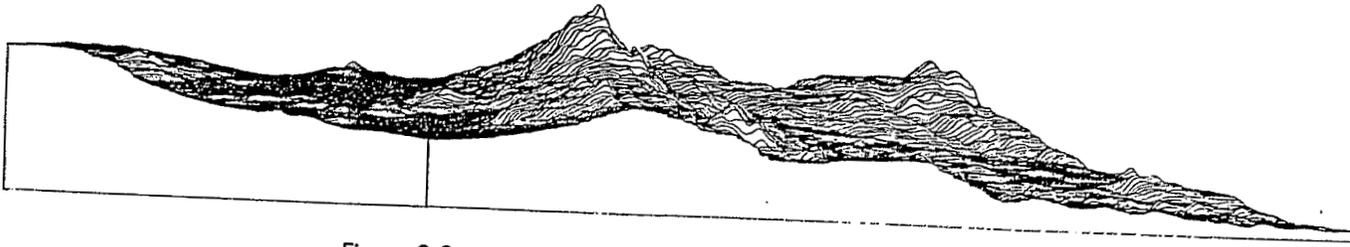


Figure 2.2



PROPOSITION D'UN SCHEMA DE COLLECTE DE L'INFORMATION POUR LE MILIEU PHYSIQUE. APPROCHE EN LANGAGE ORIENTE OBJET.

(Hervé LEMARTRET & Alain BEAUDOU)

RESUME - Le milieu naturel est assez complexe et varié pour qu'il soit difficile de recueillir une information qui représente bien l'objet que l'on veut étudier. Traditionnellement, on utilise des termes du langage courant pour décrire le milieu physique, avec l'ambiguïté que cela peut générer et l'immense quantité d'information non organisée qu'il faut ensuite manipuler. Le travail est celui d'experts et bien souvent la communication des informations se fait difficilement.

Le traitement d'une telle information procède généralement par synthèses successives de textes, de variables partiellement quantifiées ou estimées. Dans ces conditions, la recherche et la mise en place d'une automatisation de la saisie et du traitement des données paraît peu réaliste. Il apparaît qu'il faut utiliser un langage spécialisé pour créer les bases de données ; ce langage doit, dans la mesure du possible, refléter l'organisation du milieu étudié.

A partir de cette hypothèse, nous avons élaboré un schéma d'acquisition des données, qui prend en compte l'existence de toute une suite de structures emboîtées reconnaissables dans le milieu physique. Chaque structure reconnue est identifiée et quantifiée par rapport aux autres structures du même ordre présentes à l'intérieur du volume de référence. Ceci correspond à un langage : vocabulaire et syntaxe. Nous créons ainsi une description arborescente avec spécialisation en descendant la hiérarchie, et synthèse en suivant le chemin inverse (corps naturel, horizon, sol, segments, séquence, etc.).

Les bases de données traditionnelles gèrent difficilement une telle information et il semble intéressant d'étudier les conceptions de bases "orientées-objet" afin de ne pas perdre les données collectées sur le terrain. Il semble que les langages "orientés-objet" présentent une "structure" qui rappelle celles reconnues dans le milieu dit "naturel". Ils offrent en outre la possibilité de manipuler des entités non typées, d'étendre un logiciel sans modification du schéma initial.

INTRODUCTION

La pédologie a pour but de comprendre la genèse des sols, de prévoir au mieux leur évolution, de donner les informations nécessaires à une mise en valeur, etc..

La base du travail consiste, après étude de documents déjà existants (photos aériennes, cartes topographiques, cartes géologiques, données satelli-

taires), à prospecter la zone à étudier, consigner les données de terrain et fournir une carte des sols. Si les contours se déduisent essentiellement des documents précités, le contenu, qui fait l'objet de la légende de la carte, provient d'une exploitation des informations recueillies sur le terrain. Cette phase de traitement des données est, à ce jour, exclusivement manuelle et demande des mois de travail d'un expert.

J'ai volontairement utilisé dans le paragraphe précédent le terme d'expert, car si un personnel qualifié et expérimenté est amené très souvent à produire ce genre de carte, les processus mis en oeuvre sont de type expert (i.e. les mécanismes font appel à l'acquis des connaissances et à l'expérience, fruit de nombreuses années de travail dans ce domaine).

Les documents fournis (cartes pédologiques, cartes d'aptitudes culturelles, etc.) sont, par la suite, seuls utilisables et la masse de données de terrain (cahiers de terrain, fiches de description, etc.) sont inexploités et quasiment inexploitable, il faut bien l'admettre.

1. LA PROBLEMATIQUE PEDOLOGIQUE

Pour pallier cette perte d'information il était nécessaire de constituer des bases de données et, dans cet objectif, un glossaire de pédologie a été publié. De nombreuses fiches de terrain ont été proposées, améliorées, puis en définitive non retenues pour de multiples raisons (souplesse, ergonomie, retranscription, exploitation des données, etc.).

L'essor actuel des micro-ordinateurs portables serait-il la solution à ce problème ?

La cartographie pédologique tend à se marginaliser au profit d'études spécialisées, où le nombre des variables est restreint et limité, et où le traitement fait surtout appel à l'outil statistique. Il n'en demeure pas moins que le sol doit être appréhendé comme un objet complexe et évoluant dans un milieu (climat, végétation, topographie, géologie, géomorphologie, activité humaine, etc.) qui lui est propre, et hors duquel l'interprétation de phénomènes particuliers pourrait être sujet à caution.

Envisageons alors les renseignements concernant le sol comme un tout. Il semble plus judicieux de schématiser ceci par un arbre, où la branche *profil* est seule à être quelque peu détaillée pour donner un aperçu de la complexité de l'information. Au niveau des corps naturels chaque sous-arbre peut être différent, c'est à dire donner lieu à un schéma de description spécifique.

1.1. Rappel de quelques définitions concernant le sol

Profil pédologique : coupe verticale d'un sol allant de la surface à la roche mère (figures 1 & 2) ;

Horizon : dans la coupe verticale d'un sol ou profil, couche généralement horizontale qui se distingue par ses caractéristiques pédologiques. Cette notion permet de situer verticalement (limite supérieure, limite inférieure) les corps naturels et de quantifier ces entités reconnues les unes par rapport aux autres ;

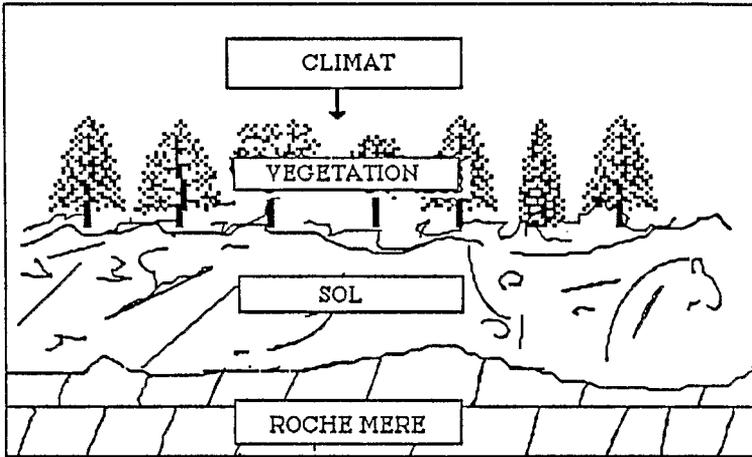


Figure 1 - Un site simplifié

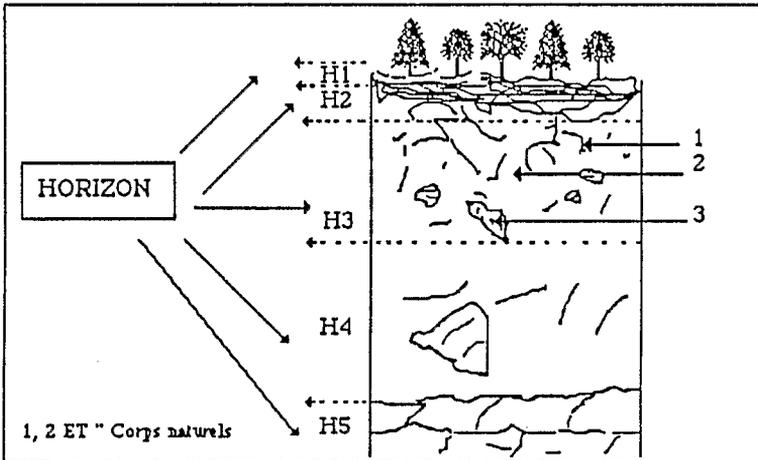


Figure 2 - Le profil horizons et corps naturels

Corps naturel : le plus petit ensemble structural naturel homogène reconnaissable à l'œil nu, qu'il soit possible de délimiter. Le corps naturel élémentaire est caractérisé par un faisceau de traits pédologiques de natures diverses qui lui sont spécifiques, mais qui peuvent varier de façon parfois importante à l'intérieur de certaines limites (A. Beaudou 1988).

La figure 3 fournit, sous la forme d'un tableau schématique, la structure de l'information sur les sols et leur environnement.

1.2. Une première expérience d'informatisation

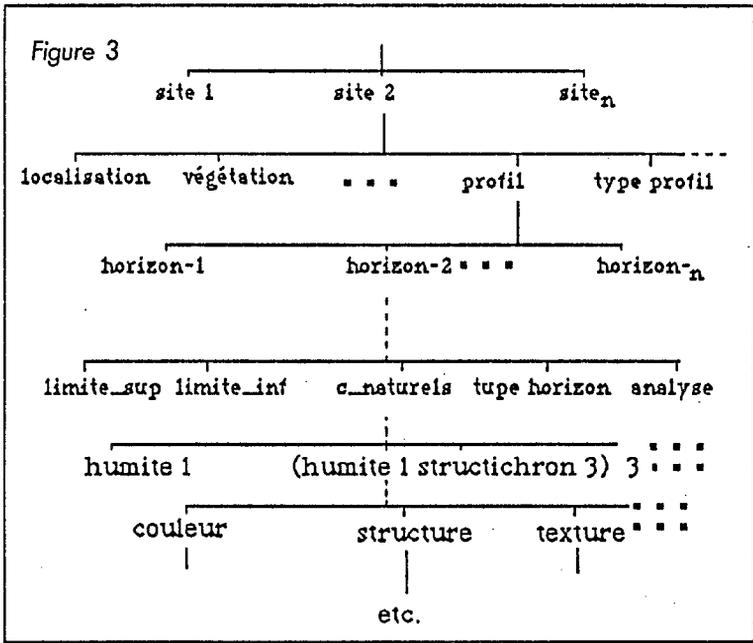
La collecte des données doit être simple et très souple. Le micro ordinateur portable peut être alors un outil très intéressant de saisie directement sur le terrain. Le premier point à développer est alors un logiciel répondant aux critères de souplesse et de simplicité pour l'utilisateur, et dirigeant l'opérateur tout au long de son travail d'acquisition ou de contrôle des données.

Partant de cette idée, un logiciel a été élaboré par un bureau d'étude de Nouvelle Calédonie (1986). Ce programme a été testé lors de missions au Cameroun, et s'est avéré très contraignant quant à la souplesse, et surtout difficile à utiliser lors de modification des données et de leur exploitation. Le point positif qui en ressort est l'aide à la saisie par un système hiérarchisé de questions/réponses.

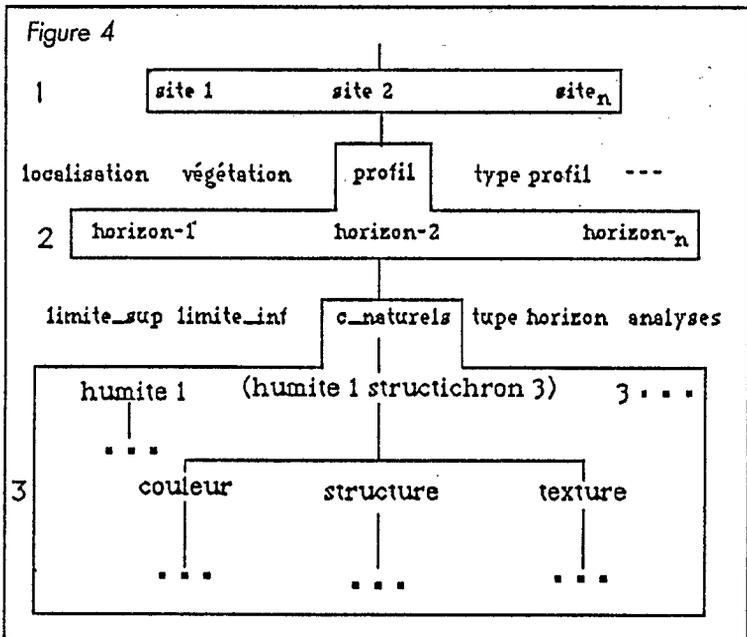
Cette première tentative a demandé beaucoup de temps, une collaboration étroite thématiciens/informaticiens. Les difficultés ont surgi du fait d'une méconnaissance presque totale de "*l'autre discipline*" pour chacune des parties, et ceci accentué par le fait que le seul outil logiciel envisageable alors était DBASE III.

Si l'on regarde la figure 1 on trouve, pour recueillir les informations concernant un site, un nombre 'n' d'horizons, 'x' corps naturels possibles, un sous-arbre spécifique à chaque corps naturel (ou groupe de corps naturels) et la possibilité d'une ramification pour préciser certaines données. Il est alors difficile d'établir un bilan exhaustif des variables et de présenter ceci sous forme de tableau. Cette difficulté pouvait être résolue par un schéma clé-valeur hiérarchisé, mal mis en évidence alors.

S'il était utopique de mettre en place un tableau pré-défini, il semble intéressant de bâtir, *a posteriori*, un tableau des variables effectivement utilisées lors d'une campagne, et de remplir ce tableau pour un traitement plus conventionnel (présence ou absence de tel caractères, statistique, etc.). Il est aussi possible de faire remonter l'information par synthèses successives qui permettront de remplir les champs '*type d'horizon, type de profil*'. C'est cette phase d'analyse qui doit être mis en oeuvre actuellement, ou tout au moins une ébauche, car la mise à plat du raisonnement de l'expert peut s'avérer extrêmement ardu dans ce domaine.



STRUCTURE DE L'INFORMATION
SUR LES SOLS ET SON ENVIRONNEMENT
TABLEAU SCHEMATIQUE



2. LES PRINCIPES DES LOO.

2.1. SmallTalk, un langage orienté objet

La programmation comporte une activité de transfert de connaissances du domaine du problème posé au domaine informatique (F. Detienne). L'acquisition de schéma est fondamentale pour mettre en oeuvre ce transfert (ceci s'acquiert bien par l'exemple, c'est pourquoi le 'tutorial' de SmallTalk m'a été d'une aide précieuse, mais souvent insuffisante).

Rappelons brièvement les concepts de base :

L'OBJET : données et procédures sont regroupées autour de l'idée unique qu'est l'objet. Les procédures étant chargées d'exploiter ces données. L'association des données et des mécanismes qui en contrôlent l'utilisation au sein d'une même entité assure l'intégrité des données, ces données ne pouvant être manipulées que par un mécanisme propre à l'objet dont elles font partie (envoi de messages).

L'ENVOI DE MESSAGES : c'est l'unique moyen d'intervention sur les objets. L'objet receveur du message détermine la méthode adéquate à activer (la méthode est reconnue dans la classe du receveur, si elle a été définie à ce niveau, ou dans l'une des classes ascendantes) le résultat est à nouveau un objet.

LA CLASSE : la classe permet de définir un modèle pour un ensemble de données et de procédures.

HIERARCHIE ET HERITAGE : une nouvelle classe est sous-classe d'une classe déjà définie dans Smalltalk, et hérite donc des structures et méthodes de ses ascendants. L'héritage est le principe par lequel un objet possède toutes les caractéristiques de l'objet à partir duquel il est créé (structure et fonctions). Il possède de plus ses propres champs et méthodes. Ici il n'existe pas d'héritage multiple ce qui résout le problème de conflit éventuels.

Le fait de pouvoir envoyer un message unique à différents objets limite le vocabulaire à mémoriser. La sortie sur imprimante des objets chaîne de caractère, tableau, entier, etc., peut être le résultat d'un message unique 'affichage' que chaque objet comprend, si cette méthode est définie dans la hiérarchie des classes correspondantes, la méthode adéquate est alors activée (notion de polymorphisme).

2.2. La fenêtre principale de smalltalk

Le programme de la fenêtre Transcript est activé par le menu standard et fait apparaître le menu spécifique permettant de sauvegarder les informations saisies en fin de travail ou de lancer le processus de saisie et manipulation des données.

3. PRESENTATION DE LA MAQUETTE

3.1. Identification des objets

La figure 4 représente les entités reconnues comme objets et ne sont qu'une filière de l'arbre de représentation de l'information.

Le *site* : c'est un ensemble d'objets informatiques. Cette liste d'objets est limitée à : LOCALISATION VEGETATION GEOLOGIE PROFIL, etc. L'activation de la procédure initiale dans "system transcript" fait apparaître la première fenêtre : <FENETRE D'INTERFACE 1> (ouverture sur le niveau 1)

DESOLO : description d'un site		
<n°site> lac1 lac2 ...	localisation végétation géologie profil ...	informations préliminaires si la colonne 1 est vide sinon ouverture réservée à l'affichage du site sélectionné en colonne pour la rubrique choisie en colonne 2
Ajout n° Efface n°	Afficher Décrire	

Chaque objet étant défini ultérieurement. La colonne 1 affiche la liste des sites enregistrés. Il est alors possible d'afficher ou de décrire l'une des rubriques de la colonne 2. (menu colonne 2). Le menu de la colonne 1 permet d'enregistrer un nouveau site à décrire ou d'effacer l'un des sites préalablement saisi. L'effacement détruit toute la hiérarchie associée à ce numéro de site, une sécurité est donc nécessaire.

Le *profil* : c'est une organisation par horizon d'un ensemble d'objets corps naturels. Lorsque l'on sélectionné la rubrique profil de la fenêtre 1, et que l'on envoie le message décrire (menu colonne 2), une deuxième fenêtre s'affiche à l'écran :

<FENETRE D'INTERFACE 2> (ouverture sur le niveau 2)

DESOLO : le profil lac 1		
<horizons> H1 H2 ... Ajouter Effacer	limiteSup limiteInf corpsNaturels type d'horizon	informations initiales lorsqu'aucun horizon n'a été saisi

Le bandeau supérieur affiche le numéro de site sur lequel on travaille.

La colonne 1 permet d'afficher les numéros d'horizons du profil courant précisé dans le bandeau supérieur. L'entrée d'un nouvel horizon ou la destruction d'un horizon existant se fait à l'aide du menu colonne 1.

Les rubriques de la colonne 2 sont en principe fixes.

Le *corps naturel* : c'est tout d'abord un 'objet' reconnaissable du milieu naturel ; il sera alors identifié et précisé par un ensemble de descripteurs. Le fait de sélectionner la rubrique CorpsNaturels dans la fenêtre 2, ouvre cette nouvelle fenêtre.

<FENETRE D'INTERFACE 3> (ouverture sur le niveau 3)

LESOL : lacl ---> H1		
<Corps naturels> humite 1 ...	Structure Couleur Texture	grumoclode anguclode sphénoclode amerode Reponse libre
Ajouter Effacer	Ajouter Effacer	Ajouter Effacer
		----- grumoclode 1 anguclode 3 <réponse courante> ou zone de texte pour l'acquisition des ré- ponses libres.
Corps Naturels - exemple : (humite 1 structichron 3)1 LISTE DES CORPS NATURELS : humite arumite structichron réducton rhizagé séméton altérite régolite entaféronto- polite écluton dermilite bioféron ...		

Le bandeau supérieur affiche le site courant et le n° d'horizon sélectionné.

La première colonne permet, comme dans les fenêtres précédentes, de saisir ou de supprimer des corps naturels. La colonne 2 donne la possibilité de modifier le questionnaire de base dont les réponses possibles figurent dans la colonne la plus à droite, réponses qui peuvent être réactualisées (menu colonne 3).

La fenêtre texte au centre et à droite affiche la réponse courante si elle est préalablement saisie. On peut la modifier ou, par le choix du label 'réponse libre', saisir une réponse complexe ou peu commune, qu'il n'est pas nécessaire de voir figurer dans la liste des réponses possibles ; ce choix ne modifiera pas le fichier 'questions/réponses'.

L'acquisition des informations est supportée par un 'objet dictionnaire'. Ce dictionnaire peut être initialisé par chargement d'un fichier, s'il existe déjà des données saisies.

Ces fenêtres sont des ouvertures sur ce dictionnaire. Les méthodes associées servent à corriger, compléter, consulter les informations qui y sont enregistrées, etc..

Les descripteurs et les réponses possibles ne sont pas figées. Un fichier de base peut être constitué et mis à jour en cours de travail sur le terrain ; les fenêtres étant également ouvertes sur des dictionnaires 'questions/réponses'.

La sauvegarde est donc nécessaire si l'un de ces dictionnaires est modifié en cours de session.

La saisie est effectuée en sélectionnant la réponse voulue dans la colonne appropriée. Une réponse libre est toujours possible dans l'ouverture texte (centre droit).

3.2. Structure et méthode des classes définies

3.2.1. La classe SITE :

Dans ce contexte, cette classe est une fenêtre ouverte sur les premiers niveaux d'un arbre Résultat. Les variables d'instance permettent de gérer cette fenêtre et ses trois ouvertures.

Méthodes (figure 5)

initialize : cette procédure se charge d'ouvrir le fichier de travail et de réorganiser le dictionnaire à partir des informations contenues dans le fichier sélectionné. Si le fichier vient d'être créé, le dictionnaire Résultat est vide. La fenêtre est ensuite ouverte pour le dialogue. (open)

affichage : offre la possibilité de visualiser, dans la zone de texte, les informations concernant le site pointé dans la colonne 1 pour la rubrique sélectionnée colonne 2. Cette routine active la procédure **afficheResult:with:** qui se charge de l'affichage effectif formaté.

affiche1 : présente dans la colonne 1 la liste des sites existants. C'est une collection triée par ordre alphabétique des clefs de niveau 1 du dictionnaire Resultat.

affiche2 : liste les rubriques à traiter dans la colonne 2. *afficheText1* écrit dans la zone texte, à droite, des directives éventuelles de saisie.

ajouteMenu1 : prend en compte la nouvelle clé spécifiée si elle n'existe pas déjà dans le dictionnaire Résultat.

description : active la procédure dont le nom est spécifié par la rubrique sélectionnée.

effaceMenu1 : ôte le n° de site sélectionné de la liste des clés du dictionnaire Résultat. Un garde-fou est nécessaire car toute l'information associée à cette clé est détruite.

listch1 et *listch2* : affectent respectivement à *sélection1* et à *sélection2* les valeurs repérées dans les colonnes 1 et 2.

menu1 et *menu2* : construisent respectivement les menus liés aux colonnes 1 et 2.

open : ouvre la fenêtre de dialogue utilisateur. Cette fenêtre est conçue avec un bandeau supérieur affichant le nom du travail en cours, deux ouvertures permettant l'affichage des numéros de site décrits, ou en cours de description, pour la colonne 1, et les rubriques à traiter pour la colonne 2. La partie droite est réservée au texte.

profil : passe les commandes à l'objet profil en construisant la filière du dictionnaire Resultat, si elle n'existe pas.

sauvegarde : sauve le dictionnaire Resultat dans le fichier spécifié ; l'extension du fichier est '.FIL'. Un fichier BAK est créé.

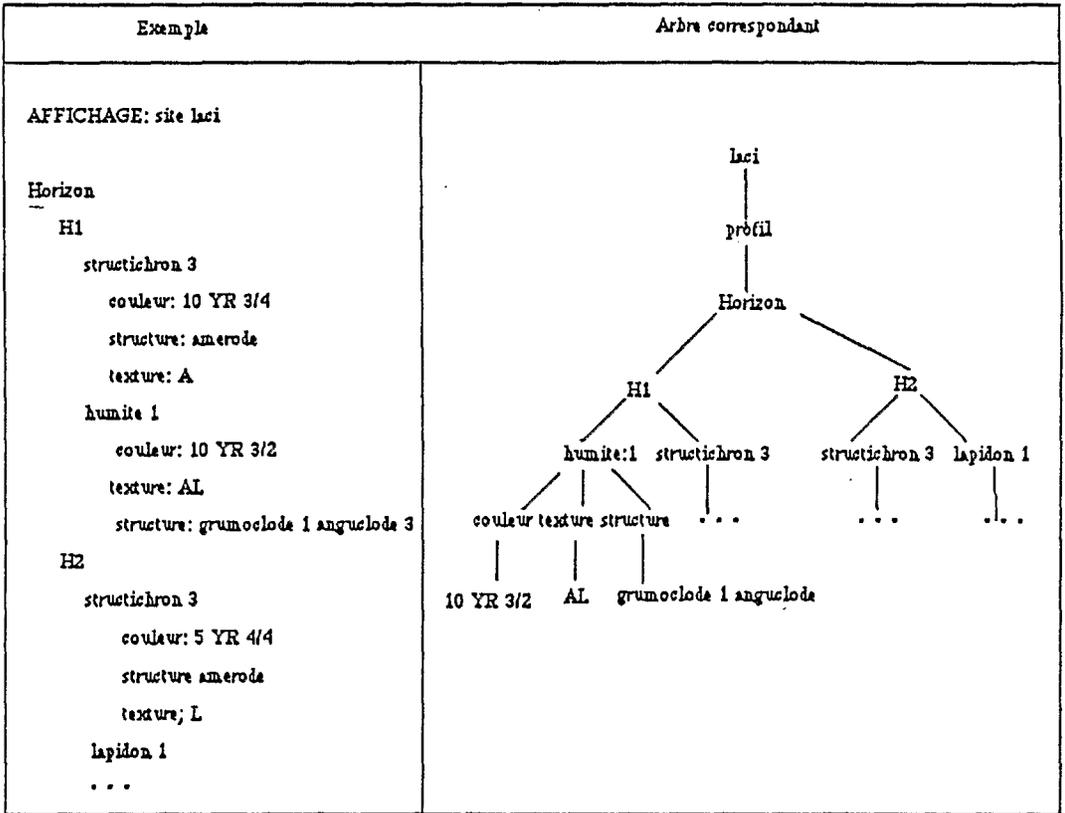


Figure 5

3.22. La classe PROFIL :

Les variables d'instance sont utilisées pour gérer cette fenêtre et ses ouvertures. Le champ réponse est initialisé par un sous-arbre du dictionnaire Resultat.

Méthodes

affiche1 : affiche la liste des horizons existants déjà pour ce profil.

affiche2 : affiche la liste des descripteurs.

afficheText1 : affichage dans *textPanel* des directives éventuelles de saisie.

ajouteMenu1 : routine du menu de la colonne 1. Ajout d'un horizon à décrire.

corpsNaturels : permet d'ouvrir une fenêtre de description des corps naturels.

effaceMenu1 : supprime un horizon existant.

initialise : (un symbole) initialise 'reponse' et 'numeroSite' et procède à l'ouverture de la fenêtre.

limiteInf : "à faire".

limiteSup : "à faire".

listch1 : (aSTRING) initialise 'sélection1' par 'aString'.

listch2 : (aString) initialise 'sélection2' par 'aString' et procède à l'activation de la routine sélectionnée.

menu1 : retourne le menu pour la colonne 1.

menu2 : retourne le menu pour la colonne 2.

open : description de la fenêtre et de son fonctionnement.

typeHorizon "à faire".

3.23. La classe CORPS NATUREL :

C'est une fenêtre ouverte sur le dernier niveau de l'arbre Resultat. Ce dernier niveau est représenté par la variable réponse. La variable *descriptSoil* est initialisé par l'arbre des questions/réponses de ce niveau. Les autres variables servant à gérer cette fenêtre et ses ouvertures.

Méthodes

afficheList1 : affiche la liste des corps naturels déjà entrés.

afficheList2 : affiche la liste des descripteurs correspondant au corps naturel sélectionné.

afficheList3 : affiche la liste des réponses possibles pour le descripteur sélectionné.

afficheText : affiche la liste des corps naturels dans la partie texte en bas de la fenêtre.

afficheText1 : affiche dans la partie texte un commentaire si aucun corps naturel n'a été saisi. Sinon, permet d'afficher la réponse courante.

afficheText1 : (aString from: aDispatcher) accepte 'aString' comme réponse ou remise à jour pour la rubrique sélectionnée et l'insère dans le dictionnaire.

afficheText2 : affiche la liste des corps naturels dans la partie texte.

ajouteMenu1 : permet d'ajouter un corps naturel au travers d'une fenêtre de dialogue.

ajouteMenuL2 : ajout d'un descripteur dans le dictionnaire questions/réponses.

ajouteMenuL3 : ajout d'une réponse possible dans le dictionnaire des questions/réponses.

effaceMenuL1 : enlève un corps naturel du dictionnaire des réponses.

effaceMenuL2 : enlève un descripteur du dictionnaire questions/réponses.

effaceMenuL3 : supprime une réponse possible du dictionnaire des descripteurs de sol.

initialize: (un symbole with: unHorizon) : initialise le dictionnaire des descripteurs à partir du fichier *descript.fil* et fait pointer 'réponse' sur la zone du dictionnaire Reponse appropriée.

list1ch: (aString) : oriente, à partir de la réponse sélectionnée en colonne 1, la description sur la branche des descripteurs appropriés.

list2ch: (aString) : affecte la valeur 'aString' à *selectedName2* ; *selectedName3* est remis à nil.

list3ch: (aString) : affecte la valeur 'aString' à *selectedName3* ; si 'aString' est différent de 'Reponse libre', alors le dictionnaire Resultat est complété avec cette valeur.

menulist1 : retourne le menu de la colonne 1.

menulist2 : retourne le menu de la colonne 2.

menulist3 : retourne le menu de la colonne 3.

open : ouvre une fenêtre de description de corps naturel.

sauveMenuL1 : sauvegarde du dictionnaire questions/réponses.

4. UNE PROPOSITION DE TRAITEMENT DE CETTE INFORMATION.

L'objectif est de synthétiser de proche en proche la masse de données et de créer des ensembles qui permettent de répondre plus rapidement et plus efficacement aux questions posées par des utilisateurs (cultivateurs, services ruraux, etc.) ou développeurs et de faciliter l'élaboration des légendes de cartes pédologiques.

4.1. Cheminement et règles permettant de déterminer le type d'horizon.

Les corps naturels sont tout d'abord regroupés en trois ensembles faisant déjà apparaître des notions de fertilité ou tout au moins des possibilités de contraintes pour une mise en valeur agricole.

ENSEMBLE I	ENSEMBLE A	ENSEMBLE N
topolite lapidon stérite réducton séméton isaltérite allotérite entaféron rudique hydrophyse	<i>sous-ensemble H</i> humite arumite mélanumite nécrumite <i>sous-ensemble M</i> structichron oxydon leuciton vertichron entaféron lutique entaféron arénique	dermilite ecluton téphralite domabion bioféron nécrophytion rhizagé rhizophyse cryptagé

Le volume occupé par un corps naturel au sein d'un horizon est déterminant et les quantifications assignées à chaque objet permet de calculer leurs proportions respectives, données en pourcentage.

Si ce chiffre fait référence à un corps naturel de l'ensemble I, alors le tableau suivant permet de donner un coefficient de contrainte (coefficient x).

Corps naturel élémentaire	Quantification en %	critère de dureté	Nature des éléments	dimension éléments	co. 'x'
stérite - lapidon et entaferon rudique		soluble	-		1
		pauci	non soluble		7
		pauci-fragi			6
		fragi			5
		fragi-duri			4
		duri			3
		duri-pétro			2
		pétro			1
	>80	pauci			7
	50-80	pauci			8
	>70	fragi		<7.5	5
	<50	fragi		>7.5	6
	>50	fragi		>7.5	4
>60	pétro-duri		<2	6	
<60	"		2-7.5	4	
40-60	"		2-7.5	5	
>40			7.5-20	2	
>20			>20	1	
semeton	>50	non soluble			3
	30-50	non soluble			5
	>30	soluble			3
reducton	>80				3
	50-80				4
isaltérite allotérite	>50				1
	>50				2

Les valeurs associées à ce coefficient sont échelonnées de 1 à 8 et comprises de la façon suivante : 1 est le niveau de contrainte le plus élevé et 8 le plus faible. L'épaisseur de l'horizon analysé associé à ce coefficient permet de classer cet horizon en 'H1', ou dans le groupe 'HN', en s'appuyant sur le tableau suivant.

Coefficient "X"	SI Epaisseur de l'horizon supérieur à :	groupe
1	1	alors
2	2	type
3	5	HI
4	10	
5	15	
6	20	
7	30	
8	40	

Les horizons de type 'HH' : ils ne sont pas classés conformément aux critères des horizons 'HI' et respectent les règles suivantes (soit sH et sN les sommes des pourcentages des corps naturels de type H et de type N.) :

Si $sH + sN > 40\%$

si $sH / sN > 3$, alors l'horizon est de type 'HH'.

Les horizons de type 'HM' : ils n'ont pas été classés en type HI ou HH. Les règles suivantes sont alors applicables (soit sA et sN le total des pourcentages des corps naturels de type A et de type N.) :

Si $sA + sN > 50\%$

si $sA / sN > 3$, alors l'horizon est de type 'HM'.

Si l'horizon ne peut être classé dans un des groupes déjà défini, il sera placé dans l'ensemble des horizons de type HN.

Remarque : tous les types d'horizon ou regroupements de corps naturels sont utiles pour organiser de proche en proche l'information, mais n'apparaissent pas à un utilisateur éventuel.

4.2. Les types de profils

Le type de chaque horizon étant connu, soit par traitement des données soit par détermination de l'expert, la succession verticale de ces horizons est le critère de classement en :

ANAPEXOL

HUMOAPEXOL

ORTHOAPEXOL

ANHUMOAPEXOL

CONCLUSION

L'objectif de cette maquette était de montrer à des pédologues l'intérêt des Langages Orientés Objets pour acquérir et organiser les données relatives au sol, données emboîtées hiérarchisées, ayant trait au domaine pédologique proprement dit mais aussi à la botanique, la géologie, la climatologie, etc..

La Programmation Orientée Objet s'est avérée un atout majeur dans l'élaboration de cette maquette pour les raisons suivantes :

- les concepts des Langages Orientés Objets sont bien adaptés à la variété de l'information et sa structuration hiérarchique ;
- l'intégrité des données est assurée par la notion d'encapsulation propre aux LOO (intégration des données et des mécanismes chargés de les manipuler). ;
- la représentation par objet rend aisée l'adjonction d'une nouvelle 'branche d'information' définie comme un nouvel objet.

Un objet étant défini, l'ajout de nouvelles méthodes se conçoit très normalement. Ces procédures peuvent être testées, individuellement par l'envoi de message (à partir du clavier) à l'objet considéré, ce qui rend la mise au point plus facile et plus rapide que dans une programmation classique.

La richesse de l'environnement SmallTalk permet, après s'y être bien familiarisé, de bâtir rapidement un travail et d'y associer du graphisme (grâce à l'interface graphique).

Ces différents arguments font que l'Approche Orientée Objet (hormis la richesse de l'environnement plus spécifique à SmallTalk) est, il me semble, indispensable pour aborder ce problème. Il apparaît toutefois que si SmallTalk est bien adapté à la réalisation d'une maquette, il serait cependant nécessaire, pour la mise en place d'une base de données, de faire appel à des langages alliant l'efficacité de programmation classique et les possibilités de programmation objet, comme par exemple C⁺⁺.

S'il est important de saisir l'information avec souplesse et convivialité, de l'organiser, la stocker, il est aussi indispensable de concevoir un système de caractérisation des horizons, des sols, des paysages. La multiplicité des approches possibles rend impératif l'orientation de la réflexion en fonction d'un objectif bien défini (programme de développement par exemple).

REFERENCES BIBLIOGRAPHIQUES

- BAILLY C., CHALLINE J.F., FERRI H.C., GLOESS, P.Y., & MARCHSIN B. : *Les langages orientés objets*. Cepadues Editions.
- BEAUDOU A.G. : *Recherche d'un système d'information pour le milieu physique*. Mémoire de thèse de doctorat d'Etat en Géographie (Paris 1), décembre 1988.
- DETIENNE F. : *Une revue des études psychologiques sur la compréhension des programmes informatiques*. Technique et science informatiques, 1989, 8(1), pp.5-20.

- DUCOURNEAU R. & HABIB M. : *La multiplicité de l'héritage dans les langages à objets*. Technique et science informatiques, 1989, 8(1), pp.41-62.
- MEVEL A. & GUEGUEN T. : *Smalltalk 80*. Eyrolles Editions.
- PINSON L.J. & WIENER R. : *An introduction to Object Oriented Programming and Smalltalk*. Addison-Wesley publishing company.
- ROCHE C. & LAURENT J.P. : *Les approches orientées objets et le langage LRO2*. Technique et science informatiques, 1989, 8(1), pp.21-40.
- SMALLTALK/V 286 : *Object oriented programming system tutorial and programming handbook*. Digital Inc..

Troisième session : Stratégies de spatialisation.

(Rapport de session, par Dominique RAMBAUD)

La troisième session intitulée "Stratégies de Spatialisation" comportait huit exposés et interventions amenant la réflexion sur la *constitution d'une base ou la structuration d'un ensemble de bases sur l'environnement* dans le domaine des données spatialisées. Les questions posées ou interventions faites après les différents exposés permettent de dégager les quelques points suivants.

Dans la démarche de l'appropriation de l'outil mis en oeuvre dans les pays où nous intervenons, il convient de *développer largement le partenariat* dans ce domaine. L'implication de nos partenaires dans l'acquisition des outils informatiques non encore maîtrisés par eux du point de vue à la fois de leur définition et de leur emploi, est parfois très prononcée soit au contraire peu accentuée. Il faut travailler à ce que le partenariat se concrétise et soit plus actif notamment en Amérique Latine.

Pour ce qui est du développement des logiciels, il serait souhaitable de *faire preuve de plus de professionnalisme* dans la mise au point. Comme cela existe pour la gestion, il importe de définir rapidement des méthodes proposées aux scientifiques : cela suppose une mise en cohérence des outils utilisés et des processus de standardisation qui constituent la clef de la réussite d'un logiciel. Faut-il mieux développer en utilisant un progiciel commercial plutôt qu'une programmation standard ? Cela dépend essentiellement de la *spécificité des objectifs* bien particuliers : mise en oeuvre d'un langage naturel, difficultés des saisies de données sur le terrain, et aspect peu conventionnel de ces données.

A propos de *l'interfaçage entre le logiciel utilisé et l'outil statistique*, il apparaît que son existence est nécessaire. A l'usage, cet interfaçage est aisé à faire suivant un format donné : il suffit de développer des modules à partir du noyau de la matrice existante. Ces divers modules sont réalisés à la demande de géographes ou d'autres utilisateurs, praticiens d'autres disciplines. Il s'agit de mener à ce niveau, une réflexion approfondie sur la meilleure façon d'intégrer dans toutes ces disciplines, les informations élaborées aux différents niveaux d'étude spatiale. Cela rejoint le souci signalé plus haut de favoriser une politique active de transfert de la méthodologie à nos partenaires.

Enfin la question, maintes fois évoquée de la **dépendance informaticiens - thématiciens** (et réciproquement) est apparue dans les débats. Cette dépendance n'est pas toujours bien ressentie par les géographes et elle dépend en outre du degré d'implication de chaque spécialiste par rapport aux techniques de son interlocuteur. Il faut veiller à ce que les thématiciens aient la possibilité de s'appropriier et d'utiliser les outils proposés par les informaticiens qui ont besoin d'utilisateurs de toutes disciplines et de toutes origines. Mais il arrive que dans un programme de recherche, limité dans le temps, l'acquisition et le traitement de l'information ne soient pas compatibles avec les délais imposés. Il y a là un risque d'erreur, car informaticiens et thématiciens doivent mener ensemble et simultanément l'analyse conceptuelle du système de données.

LISTE DE PARTICIPANTS



LISTE DES PARTICIPANTS

Monsieur *ANDRIANASOLO Haja*,
ORSTOM, LIA, 70-74 ROUTE D'AULNAY, 93140 BONDY, FRANCE

Madame *AUBRY Anne-Marie*,
ORSTOM, 70-74 ROUTE D'AULNAY, 93140 BONDY, FRANCE

Monsieur *BEAUDOU Alain*,
ORSTOM C/O AMBASSADE DE FRANCE, JALAN M.H. THAMRIN, 20,
JAKARTA PUSAT, INDONESIE

Monsieur *BOURSIER Patrice*,
UNIVERSITE DE PARIS SUD, LABORATOIRE DE RECHERCHE
INFORMATIQUE, 91405 ORSAY, FRANCE

Monsieur *CAMBREZY Luc*,
Apartado Postal 544, 91000 XALAPA - VER, MEXIQUE

Monsieur *CARDON Daniel*,
ORSTOM/IRENA, CARRETERA NORTE, APARTADO POSTAL 5123,
MANAGUA, NICARAGUA

Madame *CHAUME Régine*,
ORSTOM, 2051 AVENUE DU VAL DE MONTFERRAND, BP 5045,
34032 MONTPELLIER CEDEX, FRANCE

Monsieur *CHEVILLOTTE Hervé*,
ORSTOM, BP 1386, DAKAR-HANN, SENEGAL

Monsieur *CORRE Daniel*,
ORSTOM, IFREMER, BP 70, 29263 PLOUZANE, FRANCE

Madame *COURET Dominique*,
3 RUE DE LA JONQUIERE, 75017 PARIS, FRANCE

Monsieur *CRUETTE Jacques*,
Centre ORSTOM DE BONDY, 70-74 ROUTE D'AULNAY, 93140
BONDY, FRANCE

Monsieur *DAGORNE Dominique*,
ANTENNE ORSTOM AUPRES DU CENTRE DE METEOROLOGIE SPATIALE,
BP 147, 22302 LANNION, FRANCE

Monsieur *DANDOY Gérard*,
ANTENNE ORSTOM AUPRES DU GIP RECLUS, MAISON DE LA
GEOGRAPHIE, 17 RUE ABBE DE L'EPEE, 34000 MONTPELLIER, FRANCE

Monsieur *DELCLAUX François*,
ORSTOM, LABORATOIRE D'HYDROLOGIE, 2051 AVENUE DU VAL DE
MONTFERRAND, BP 5045, 34032 MONTPELLIER CEDEX, FRANCE

Monsieur *DEPRAETERE Christian*,
ORSTOM, LABORATOIRE D'HYDROLOGIE, 2051 AVENUE DU VAL DE
MONTFERRAND, BP 5045, 34032 MONTPELLIER CEDEX, FRANCE

Monsieur *DIZIAIN Roland*,
ORSTOM, 70-74 ROUTE D'AULNAY, 93140 BONDY, FRANCE

Monsieur *DUCHEMIN Jean-Paul*,
ORSTOM, CS6, 213 RUE LA FAYETTE, 75010 PARIS, FRANCE

Madame *DUREAU Françoise*,
ORSTOM, DEPARTEMENT SDU, 70-74 ROUTE D'AULNAY, 93140
BONDY, FRANCE

Madame *EMPERAIRE Laure*,
ORSTOM, DEPARTEMENT MAA, 213 RUE LA FAYETTE, 75010 PARIS,
FRANCE

Monsieur *FAURE Paul*,
ORSTOM, UTGP.LFS, 70-74 ROUTE D'AULNAY, 93140 BONDY,
FRANCE

Monsieur *GIRARD Georges*,
ECOLE DES MINES DE PARIS, CENTRE INFORMATIQUE GEOLOGIQUE,
35 RUE SAINT-HONORE, 77300 FONTAINEBLEAU, FRANCE

Monsieur *GUILLOT Bernard*,
ANTENNE ORSTOM AUPRES DU CENTRE DE METEOROLOGIE SPATIALE,
BP 147, 22302 LANNION, FRANCE

Monsieur *GUISCAFRE Jacques*,
ORSTOM, 2051 AVENUE DU VAL DE MONTFERRAND, BP 5045,
34032 MONTPELLIER CEDEX, FRANCE

Monsieur *HARDY Bernard*,
ORSTOM, UNITE DE CARTOGRAPHIE, 70-74 ROUTE D'AULNAY, 93140
BONDY, FRANCE

Monsieur *JACCON Gilbert*,
ORSTOM, LABORATOIRE D'HYDROLOGIE, 70-74, ROUTE D'AULNAY,
93140 BONDY, FRANCE

Monsieur *LARDY Michel*,
ORSTOM, BP A5, NOUMEA CEDEX, NOUVELLE CALEDONIE, FRANCE

Monsieur *LE BOZEC Jean-François*,
ORSTOM, SIG, 213 RUE LA FAYETTE, 75010 PARIS, FRANCE

Monsieur *LEMARTRET Hervé*,
ORSTOM, LIA, 70-74 ROUTE D'AULNAY, 93140 BONDY, FRANCE

Monsieur *LECHAUVE Jean-Jacques*,
ORSTOM, IFREMER, BP 70, 29263 PLOUZANE, FRANCE

Monsieur *LORTIC Bernard*,
ORSTOM, LIA - UNITE DE TELEDETECTION, 70-74 ROUTE D'AULNAY,
93140 BONDY, FRANCE

Monsieur *MADEIRA José*,
ORSTOM, 70-74 ROUTE D'AULNAY, 93140 BONDY, FRANCE

Monsieur *MALARD Pierre*,
ORSTOM, SIG, 213 RUE LA FAYETTE, 75010 PARIS, FRANCE

Monsieur *MANGIN André*,
SCOT CONSEIL, 18 AVENUE EDOUARD BELIN, 31055 TOULOUSE
CEDEX, FRANCE

Monsieur *MARCHAL Jean-Yves*,
19 RUE LEVERT, 75020 PARIS, FRANCE

Monsieur *MARTIN Jean-Yves*,
ORSTOM, CS7, 213 RUE LA FAYETTE, 75010 PARIS, FRANCE

Monsieur *MEUNIER François*,
ORSTOM, UNITE DE CARTOGRAPHIE, 70-74 ROUTE D'AULNAY, 93140
BONDY, FRANCE

Madame *MICHAUX Monique*,
ORSTOM, 2051 AVENUE DU VAL DE MONTFERRAND, BP 5045,
34032 MONTPELLIER CEDEX, FRANCE

Monsieur *MOUGENOT Bernard*,
ORSTOM, LIA, 70-74 ROUTE D'AULNAY, 93140 BONDY, FRANCE

Monsieur *MULLON Christian*,
ORSTOM, LIA, 70-74 ROUTE D'AULNAY, 93140 BONDY, FRANCE

Monsieur *PAUGY Didier*,
ANTENNE ORSTOM AUPRES DU MUSEUM D'HISTOIRE NATURELLE,
LABORATOIRE D'ICHTYOLOGIE, 43 RUE CUVIER, 75231 PARIS CEDEX
05, FRANCE

Madame *PELLETIER Françoise*,
ORSTOM, UNITE D'INFOGRAPHIE, 70-74 ROUTE D'AULNAY, 93140
BONDY, FRANCE

Monsieur *PELTRE Pierre*,
ORSTOM, LIA, 70-74 ROUTE D'AULNAY, 93140 BONDY, FRANCE

Madame *PENNARUN Jeannine*,
ORSTOM, 70-74 ROUTE D'AULNAY, 93140 BONDY, FRANCE

Madame *PERRIER Edith*,
ORSTOM, LIA, 70-74 ROUTE D'AULNAY, 93140 BONDY, FRANCE

Monsieur *PICHON Gaston*,
ORSTOM, LIA, 70-74 ROUTE D'AULNAY, 93140 BONDY, FRANCE

Monsieur *PION Jean-Claude*,
INSTITUT DE GEOLOGIE, 1 RUE BLESSIG, 67000 STRASBOURG,
FRANCE

Madame *PONCET Yveline*,
ORSTOM, LIA, 70-74 ROUTE D'AULNAY, 93140 BONDY, FRANCE

Monsieur *POUGET Jean-Christophe*,
ORSTOM, BP 1020, 97178 POINTE A PITRE, GUADELOUPE, FRANCE

Monsieur *POUGET Marcel*,
ORSTOM, LIA, 70-74 ROUTE D'AULNAY, 93140 BONDY, FRANCE

Monsieur *POUYAUD Bernard*,
ORSTOM, DEPARTEMENT DEC, 213 RUE LA FAYETTE, 75010 PARIS,
FRANCE

Monsieur PRAT *Christian*,
ORSTOM, LFS, 70-74 ROUTE D'AULNAY, 93140 BONDY, FRANCE

Monsieur RAMBAUD *Dominique*,
ORSTOM, 70-74 ROUTE D'AULNAY, 93140 BONDY, FRANCE

Monsieur RAOUS *Patrick*,
ORSTOM, LABORATOIRE D'HYDROLOGIE, 2051 AVENUE DU VAL DE
MONTFERRAND, BP 5045, 34032 MONTPELLIER CEDEX, FRANCE

Monsieur REMY *Dominique*,
ORSTOM, LIA, 70-74 ROUTE D'AULNAY, 93140 BONDY, FRANCE

Monsieur RENAUD *Pascal*,
ORSTOM, MTI, 213 RUE LA FAYETTE, 75010 PARIS, FRANCE

Monsieur SCHWARTZ *André*,
ORSTOM, LABORATOIRE DES SCIENCES SOCIALES, 70-74, ROUTE
D'AULNAY, 93140 BONDY, FRANCE

Monsieur SECHET *Patrick*,
EMBRAPA-SEDE (DIN/DDS), CAIXA POSTAL 04-0315, 70770 BRASILIA,
DF, BRESIL

Monsieur SONDAG *Francis*,
ORSTOM, 70-74 ROUTE D'AULNAY, 93140 BONDY, FRANCE

Monsieur SOURIS *Marc*,
MISSION ORSTOM, RUSIA NO. 220 Y SHYRIS - AP 6596 CCI, QUITO,
EQUATEUR

Madame THOMASSIN *Marie-Madeleine*,
118 BIS RUE GAMBETTA, 94120 FONTENAY SOUS BOIS, FRANCE

Monsieur TITO DE MORAIS,
ORSTOM, LABORATOIRE D'HYDROBIOLOGIE, BP 165, 97323
CAYENNE, GUYANE, FRANCE

Monsieur TRAVAGLIO *Michel*,
ORSTOM, LABORATOIRE D'HYDROLOGIE, 2051 AVENUE DU VAL DE
MONTFERRAND, BP 5045, 34032 MONTPELLIER CEDEX, FRANCE

Monsieur TREUIL *Jean-Pierre*,
ORSTOM, SIG, 213 RUE LA FAYETTE, 75010 PARIS, FRANCE

Monsieur VAUGELADE *Jacques*,
ORSTOM, BP 182, OUAGADOUGOU, BURKINA FASO

Madame VERNEUIL *Marie-Paule*,
ORSTOM, UNITE DE CARTOGRAPHIE, 70-74 ROUTE D'AULNAY, 93140
BONDY, FRANCE

Monsieur VERNEUIL *Pierre*,
ORSTOM, DEPARTEMENT SDU, 70-74 ROUTE D'AULNAY, 93140
BONDY, FRANCE

Monsieur VIZCONDE *Miguel Angel*,
INEGI, 11 PONIENTE, 1711 PUEBLA - PUE, MEXIQUE

Monsieur WANIEZ *Philippe*,
ANTENNE ORSTOM AUPRES DU GIP RECLUS, MAISON DE LA
GEOGRAPHIE, 17 RUE ABBE DE L'EPEE, 34000 MONTPELLIER, FRANCE

Monsieur WINTER *Gérard*,
ORSTOM, 213 RUE LA FAYETTE, 75010 PARIS, FRANCE

**INDEX
PAR AUTEUR**

INDEX PAR AUTEUR

ASSAD, E. 211 ;
AUBRY, A.M. 241 ;
BEAUDOU, A. 331 ;
BELARMINO, L.C. 79 ;
BOYER, J.F. 37 ;
CAMBREZY, L. 319 ;
CARDON, D. 61,119 ;
COCHONNEAU, G. 287 ;
COURET, D. 261 ;
CRESPY, A. 37 ;
DAGORNE, D. 227 ;
DANDOY, G. 159,205 ;
DELCLAUX, F. 277 ;
FRESNEDA, P.S.V. 287 ;
GUISCAFRE, J. 37 ;
HOUNDAGBA, C.J. 241 ;
JACCON, G. 49 ;
JUNIOR, S.R. 211 ;
LARDY, M. 91 ;
LAWIRA, C. 277 ;
LECHAUVE, J.J. 111 ;
LEMARTRET, H. 331 ;
LOECK, A.E. 79 ;
LOINTIER, M. 187 ;
LORTIC, B. 101 ;
MACEDO, J. 211 ;
MARTIN, J.Y. 117 ;
MARTINS, J.F. da S. 79 ;
MULLON, C. 17 ;
PAMBRUN, C. 91 ;
RAOUS, P. 37 ;
RAMBAUD, D. 349 ;
REMY, D. 135 ;
RENAUD, P. 17 ;
RICHARD, J.F. 241 ;
SANTOS, D.T. dos 211 ;
SECHET, P. 79,201,241,287 ;
SOURIS, M. 135,253 ;
TITO DE MORAIS, L. 173 ;
TABBAGH, J. 91 ;
THOMASSIN, M.M. 141 ;
TRAVAGLIO, M. 49 ;
WANIEZ, P. 159 ;
WINTER, G. 9 ;

ORSTOM Éditeur
Dépôt légal : août 1990
Impression
ORSTOM BONDY

ISSN : 0767-2896
ISBN : 2-7099-1006-3
Editions de l'ORSTOM
72, route d'Aulnay
F-93143 BONDY Cédex

Photographie de couverture
(cliché Patrick Séchet)
*Détail d'un appareil rudimentaire de saisie
de documents cartographiques (1982)*