

Un système d'information géographique en Inde du Sud: Théorie, mise en œuvre et applications thématiques

Christophe Z Guilmoto (IRD/IFP)

Sébastien Oliveau (Paris I/IFP)

Sattia Vingadassamy (IFP)

French Institute PB 33 Pondicherry 605001 India

Fax: 91 413 339534

Email: guilmoto@ird.fr, oliveau@altern.org

Un système d'information géographique en Inde du Sud: Théorie, mise en œuvre et applications thématiques

Résumé:

L'Inde et son milliard d'habitants possèdent un riche recensement plus que centenaire, qui reste une source sous-exploitée pour l'étude des dynamiques socio-spatiales actuelles. L'article résume les étapes d'un projet de géomatique appliqué aux quelque 75.000 villages de l'Inde du sud.

L'apparition de la technologie des SIG dans un pays comme l'Inde bute sur de multiples problèmes (scientifiques, institutionnels et techniques). La localisation géoréférencée des villages a, par exemple, demandé des efforts considérables, en relation directe avec la rareté et la médiocre qualité des fonds cartographiques disponibles sur papier. De même, l'exploitation systématique de la statistique censitaire a mis en évidence ses faiblesses et ses lacunes, nécessitant des corrections de tous types. L'article fait le récit de ces problèmes pratiques et théoriques, ainsi que des solutions trouvées à la question de la restitution de l'information constituée grâce aux médias modernes.

En conclusion, nous présentons des exemples d'applications cartographiques de cette riche base de données et quelques outils géostatistiques qui permettent désormais d'aborder de manière renouvelée l'étude du changement socio-spatial.

Mots-clés: Inde, Inde du sud, Tamil Nadu, Kerala, Karnataka, Andhra Pradesh, SIG, cartographie, géostatistique

A geographical information system in South India: Theory, implementation and thematic applications

Summary:

India whose population has recently reached the one-billion mark enjoys a rich and ancient census apparatus, which remains an under-utilized source for the study of contemporary socio-spatial dynamics. Our paper sums up the stages of a GIS project applied to the 75.000 localities of South India.

The emergence of GIS technology in a country like India comes up against many scientific, technical or institutional problems. Georeferencing of villages, for instance, required considerable effort because of the unavailability of reliable printed cartography. Similarly, systematic examination of the census data has underlined their shortcomings and called for corrections of all kinds. This article chronicles some of these concrete and theoretical difficulties, as well as solutions found to make the collected information available to a large number of potential users.

As a conclusion, we display some of the cartographic applications of this rich database, as well as some geostatistical tools that can fruitfully applied for a renewed approach to the phenomenon of socio-spatial change.

Keywords: India, South India, Tamil Nadu, Kerala, Karnataka, Andhra Pradesh, GIS, cartography, geostatistics.

Un système d'information géographique en Inde du Sud: Théorie, mise en œuvre et applications thématiques *

Christophe Z Guilmoto (IRD/IFP)

Sébastien Oliveau (Paris I/IFP)

Sattia Vingadassamy (IFP)

1. Introduction

Pourquoi ne pas rappeler, en préambule à cet article, ce que la démarche de recherche doit simultanément aux impératifs d'un questionnement problématique et au relatif hasard de la disponibilité de l'information? Car c'est recherchant des matériaux pour un travail nouveau sur la baisse de la natalité en Inde du sud que le premier auteur découvrit de manière presque fortuite l'existence d'une base de données originales qui est à l'origine des projets décrits ici. En effet, l'administration du recensement avait décidé, plusieurs années après les opérations de 1991, de mettre à la disposition du public (averti) quelques résultats sous format numérique. Parmi ceux-ci, des fichiers contenaient les informations relatives aux villages et aux villes, appelées ultérieurement à être publiées dans les volumes par district du recensement.¹

Considérant cette potentielle manne statistique, il fut alors envisagé de constituer une grande banque de données régionale, constituée à partir de l'assemblage des données primaires, afin de disposer d'une base rassemblant tous les villages et les villes des états choisis, en l'occurrence l'Andhra Pradesh, le Karnataka, le Kerala et le Tamil Nadu, auxquels on a adjoint le petit territoire de Pondichéry. Quelques jours plus tard, une discussion tout aussi accidentelle avec quelques collègues partant à l'assaut d'un vaillant système d'information géographique (SIG) indien, conduisit à élargir l'approche à la dimension spatiale: s'il était possible de géoréférencer ces unités censitaires, on disposerait alors d'un formidable outil pour interroger les résultats du dernier recensement indien. En sorte qu'au-delà de la seule étude de la propagation de la baisse de la fécondité en Inde du sud, les données et les techniques désormais disponibles permettraient d'envisager une approche exhaustive de la situation socio-économique de l'Inde du sud, d'un point de vue à la fois statistique et spatial².

Cet article retrace les principales étapes de ce projet et ses divers avatars, illustrant notre propos de plusieurs cartes et documents. Plutôt que de nous limiter aux résultats concrets des analyses, nous avons choisi ici, afin de rendre la leçon plus instructive, d'en présenter avec plus de détails les phases techniques successives. Il témoigne de difficultés que les modélisateurs, forts de bases de données spatialisées anciennes couvrant les pays industrialisés, ont déjà oubliées, mais qui attendent tous les projets concernant les pays en développement. S'il est vrai que les statistiques dans le tiers-monde viennent souvent à manquer ou qu'elles restent sous-utilisées

* Ce travail décrit la mise en place du fond géomatique du *South India Fertility Project* (SIFP), projet coordonné par C.Z.Guilmoto. Le projet a bénéficié du soutien de l'IRD, de l'Institut français de Pondichéry (IFP) et de financements du Wellcome Trust et du FNUAP. Nous remercions R. Amuda de son appui, ainsi que les participants du colloque sur la géographie de la population (Lille, novembre 2001) et Rémy Delage pour leurs commentaires sur une première version de ce texte.

¹ Ces volumes (*District Census Handbook*) sont, à l'occasion de chaque recensement décennal, publiés en anglais ou hindi pour chaque district que compte l'Inde (466 en 1991). Leur publication intervient très tardivement, souvent à la veille du recensement suivant, et le recensement de 1991 n'a pas dérogé à cette règle, en dépit de son informatisation nouvelle.

² Pour des présentations générales de l'application des SIG à l'analyse spatiale des données sociales, voir notamment Martin (1996) et Longley and Batty (1996).

pour de multiples raisons techniques, les fonds géographiques sont encore plus défectueux et seule une infime partie d'entre eux est aujourd'hui disponible en format numérique. A l'heure de l'étude de transformations sociales, plus rapides et parfois plus dramatiques dans les pays en développement qu'ailleurs, l'approche géographique s'impose comme une clé de la lecture de ces mécanismes et il est urgent de canaliser l'effort technologique dans cette direction.

2. La mise en place d'une base de données géoréférencée de l'Inde du sud

Malgré une apparente profusion de sources de données en Inde (recensement, état civil, grandes enquêtes sociodémographiques), les exploitations géographiques précises et les approches cartographiques à une échelle fine restent rares, voire inexistantes. Si les travaux au niveau du district³ sont fréquents, les applications à un niveau administratif inférieur sont plus rares et plus limitées dans l'espace. Les études proprement cartographiques à l'échelon villageois, hormis les approches monographiques, sont, pour ainsi dire, inexistantes.

Cette absence tient notamment au manque cruel de couvertures cartographiques précises et accessibles. Il n'existe que deux sources fournissant des cartes en Inde: le *Survey of India* pour les fonds topographiques et le recensement indien dont les volumes par district fournissent les découpages administratifs, repris d'ailleurs du *Survey of India*. Hormis quelques plans de grandes villes récemment publiés, toutes les cartes en Inde reposent sur les sources du *Survey of India* et de ses dérivés. De plus, une politique datant d'une autre époque interdit la diffusion des cartes topographiques mentionnant des sites considérés comme stratégiques (bases militaires et installations nucléaires, mais également les ouvrages d'art etc.) et exclut pour cette raison une bande côtière de plus de 50 km à l'intérieur des terres. Pour certains états littoraux, comme le Kérala, les cartes topographiques ne sont donc pas officiellement disponibles. La production de nouvelles cartes, comme les cartographies du couvert végétal produites par l'Institut français de Pondichéry, est soumise à de nombreux contrôles administratifs et stratégiques qui ralentissent leur diffusion⁴.

Les cartes publiées sont, par conséquent, incomplètes et peu nombreuses. Il est impossible de situer tous les villages, ni même toutes les infrastructures (les barrages, par exemple, puisqu'ils sont classés dans les lieux stratégiques). Une première solution, apparemment paradoxale, consiste donc à faire appel aux cartes britanniques datant de la colonisation et disponibles plus aisément. La couverture est ainsi complète, mais un certain nombre de cartes datent du début du XX^e siècle: la figure 1 montre par exemple une carte topographique datant de 1914, correspondant à la région de Vellore dans l'état du Tamil Nadu, qui a été utilisée dans notre travail.

Figure 1 : Carte topographique de la région de Vellore (Tamil Nadu, Inde), 1914-15.

La seconde solution repose sur l'utilisation des cartes du recensement (Figure 2) qui présentent l'avantage de fournir une localisation exhaustive des unités de recensement, à savoir les villages fiscaux (*revenues villages*). Ces cartes censitaires demeurent grossières et ne permettent pas de géoréférencement direct. Elles demandent par conséquent un travail minutieux de reconstitution de l'information. Notons enfin que ces cartes sont souvent publiées tardivement après le recensement, les dernières parutions pour le recensement de 1991 datant d'août 2000.

³ Les états indiens ont une maille administrative construite sur différents niveaux: les districts (équivalant aux départements français), les taluks ou tehsils (nom variable, équivalents à des cantons), les *development blocs* (unités de développement), les *revenues villages* (villages fiscaux, ou communes) et les *panchayats* (unités politiques).

⁴ Des voix s'élèvent dans les milieux professionnels indiens pour dénoncer ces politiques restrictives. Voir par exemple la revue *GIS@development* publiée à Delhi.

Figure 2 : Localisation des localités du taluk de Vellore (Tamil Nadu, Inde) d'après la carte du recensement de 1981.

Vu l'ampleur du travail, la mise en place d'une base de données géoréférencée s'avérait donc une tâche de longue haleine. Il a fallu en premier lieu saisir la localisation de tous les villages individuellement. Cette seule entreprise a pris de nombreux mois. En second lieu, on a géoréférencé les cartes des quelque 712 taluks composant l'Inde du sud, en utilisant les cartes topographiques parfois anciennes pour localiser certains points de repère figurant sur les cartes du recensement. Enfin, on a assemblé les cartes des villages de tous les taluks en une carte unique sans chevauchement ni décalage.

Les systèmes d'information géographique (SIG) après géoréférencement (figure 3) des différents fonds autorisent l'assemblage de cartes venues de sources différentes et la mise à jour ultérieure des informations. En effet, à l'intérieur d'un SIG, les objets sont localisés dans un système de projection universel. Ce travail fut long, associant saisie, cartographie et calage des cartes numérisées sur les fonds géoréférencés disponibles. A l'issue de deux ans de préparation, les 75 000 villages du Sud de l'Inde ont finalement été rassemblés sur un SIG intégrant l'ensemble. Cette base originale est sans équivalent ailleurs en Inde. Les informations du recensement de 1991 ont été associées à chacun de ces villages. Les applications dans différents champs scientifiques sont très importantes, et plusieurs travaux les exploitent déjà dans différents domaines comme nous l'illustrons dans les dernières sections de cet article.

Figure 3 : Recalage de la carte des localités (Figure 2) sur la carte topographique (Figure 1).

Quelques étapes de ce travail méritent d'être mentionnées. Ainsi, un premier test a été initialement conduit à l'échelle d'un district (celui de Vellore figurant sur les cartes précédentes), afin de valider la procédure de saisie et la qualité des informations statistiques attachées aux villages du recensement de 1991 (Oliveau, 1998). Le résultat ayant été positif, la saisie de tous les villages a commencé.

Dans le même temps, l'assemblage des fichiers statistiques mis à disposition par le *Census of India* (environ 400 fichiers pour l'ensemble de l'Inde du sud) a fait apparaître de nombreux problèmes statistiques qu'il a fallu corriger. Ensuite, la mise en correspondance des villages du recensement et des localisations a fait apparaître de nouvelles difficultés. Ainsi, de nombreux villages étaient absents des cartes, et cela a été particulièrement vrai dans les zones tribales mal couvertes par l'administration. De plus, la transcription en anglais des noms locaux (quatre langues et alphabets principaux coexistent en Inde du sud) crée de multiples toponymes ambigus. La bonne connaissance du terrain par les techniciens et des vérifications auprès des bureaux régionaux du recensement ont pallié une partie de ces problèmes.

Un des choix majeurs lors de l'élaboration de la base de données a été le mode de représentation des villages. La forme ponctuelle a été choisie contre les polygones pour trois raisons. Tout d'abord, les limites des villages fournies par le census ne paraissaient pas fiables. Des comparaisons entre la surface des villages donnée par le recensement et mesurée par le SIG ont montré la fragilité de cette information. Ensuite, l'évolution future de la base de données a été jugée plus aisée si les entités étaient représentées par des points que si elles ne l'étaient par des surfaces. En effet, l'apparition ou la disparition d'une nouvelle entité administrative se traduit par l'apparition ou la disparition d'un point. La cartographie globale ne varie donc pas. Par contre, dans une représentation polygonale, le moindre changement entraîne la nécessité de modifier les entités jouxtant le polygone créé ou enlevé. Enfin, à l'échelle de toute l'Inde du Sud, la représentation cartographique des villages demande un tel degré de généralisation qu'une tessellation de la base de points donne un résultat très satisfaisant.

Progressivement, de nouvelles couches de données ont été intégrées en provenance d'autres sources disponibles. La base de données, inédite en Inde, qui a vu le jour après deux ans offrait donc une mine incomparable d'informations, couvrant les quatre états de l'Inde du Sud: Andhra Pradesh, Karnataka, Kerala et Tamil Nadu ainsi que le territoire de Pondichéry. Quelques chiffres résument l'étendue du système: la superficie couverte est de l'ordre de 636 000 Km², pour une population de 223 384 786 personnes en 2001, répartie entre 75 523 villages et 843 unités urbaines (en 1991). A chaque unité administrative sont liées 130 attributs concernant des données socio-économiques (alphabétisation, activités professionnelles etc.) et une quarantaine

concernant les infrastructures villageoises et l'utilisation du sol. En plus de ces couches aux échelons villageois et urbains, les données au niveau des taluks et des districts ont été agrégées pour des études plus larges. De même, une couche concernant le réseau routier et ferroviaire a été rajoutée, par croisement des informations disponibles dans le recensement et dans diverses cartes privées. La couverture hydrographique (notamment à l'aide du DCW⁵) et des données concernant l'altitude ont aussi été ajoutées. Un exemple du SIG obtenu est repris sur la figure 4 qui couvre la frontière entre les états du Kerala et du Karnataka. On y relève notamment l'inégale répartition du maillage villageois, très dense dans la région de Mysore au Karnataka et beaucoup plus espacé au Kerala, dont le peuplement se caractérise par des villages de très grande taille démographique.

Figure 4 : Localisation des villages du recensement de 1991 sur la frontière entre Karnataka et Kerala, Inde.

Un point important mérite d'être souligné. Si le SIG ainsi créé est déjà relativement large, il a été conçu pour pouvoir être mis à jour et complété ultérieurement. Une première mise à jour a commencé avec la publication des données du recensement de 2001. Les premiers résultats concernant les districts et les villes sont parus et ont été intégrés. Mais le géoréférencement permet aussi l'intégration d'autres sources comme des fonds vectoriels (comme le DCW), raster ou des images satellites. Nous pouvons toutefois, en conclusion, souligner quelques-unes des limites de l'exercice, qui sont de trois ordres.

L'information ne pouvant pas être produite en temps réel, et les ressources étant tout de même limitées, les données et les fonds de cartes avaient déjà sept ans lors du démarrage du projet. Ils en ont dix aujourd'hui, et cela constitue une première limite, relative, à leur utilisation. Les situations locales ont naturellement évolué, et cela est d'autant plus vrai que les infrastructures et la démographie sont des dimensions particulièrement dynamiques en Inde du sud. Ainsi, par exemple, la route côtière qui relie depuis 1999 Chennai (Madras) à Pondichéry et les nombreuses bretelles qui contournent aujourd'hui les agglomérations manquent aux fonds originaux utilisés.

Le recensement étant la seule source exhaustive à l'échelle villageoise, la deuxième limite réside dans la source unique des données. Les vérifications et corrections possibles restent limitées. Nous avons effectué une série de tests statistiques⁶, des tests logiques sur les données ainsi que des vérifications croisées entre les données numériques et les publications sur papier, quand elles étaient disponibles. La base statistique finale du *South India Fertility Project*, si elle n'est pas parfaite, est bien meilleure que les publications du recensement.

La troisième limite est d'ordre cartographique. Les sources étant diverses et parfois de qualité moyenne, l'erreur de précision finale, basée sur des vérifications faites sur le terrain avec un GPS, est inférieure à 500 mètres, pour un territoire s'étendant sur 1400 kilomètres du nord au sud. Elle est souvent inférieure à 250 mètres, mais des dérapages sont vraisemblables comme dans les régions tribales d'Andhra Pradesh figurant sur la figure 7. La précision générale de la base de données reste satisfaisante et autorise les traitements spatiaux que nous effectuons (cf. infra). Il convient d'ailleurs de rappeler que les sources cartographiques informatisées sur la population indienne sont quasi-inexistantes. Une source comme le DCW est moins précise (sa définition est de 1 : 1 000 000), plus ancienne (les relevés pour l'Inde datent des années 1970) et contient plus d'erreurs: le repérage à partir d'images satellites ne « voit » pas les infrastructures existantes sous les végétaux (routes en forêt ou bordées d'arbres), de même qu'il n'est pas capable de différencier les infrastructures abandonnées de celles encore utilisées⁷.

3. La diffusion par cdrom et Internet

⁵ Le DCW (*Digital Chart of the World*) est une base de données géographiques à l'échelle mondiale. Elle a été développée par l'entreprise ESRI pour l'agence de cartographie de la défense américaine. La source primaire est l'ONC (*Operational Navigation Chart*).

⁶ Ce travail a été notamment réalisé par Bernard Buffière et Véronique Joseph lors d'un stage de DESS à Pondichéry en 1998.

⁷ A propos de la qualité du DCW, voir Kraak and Ormeling (1996 : 207).

Une fois la base de données construite, son exploitation a commencé de façon systématique dans le cadre de projets scientifiques variés que nous décrivons plus bas. Dans le même temps, l'équipe a été rapidement soumise à des requêtes incessantes d'information géographique et statistique de toute nature. Il est donc apparu nécessaire de développer des instruments de communication et de restitution des données pouvant satisfaire les besoins des chercheurs ou d'autres secteurs (ONG, administrations, entreprises privées). Le volume de l'information interdisait une mise à disposition sur Internet de l'ensemble de la base, à moins de limiter les consultations à des échantillons très réduits ou à des utilisateurs très bien équipés.

Il a donc été décidé d'utiliser un cdrom comme mode de diffusion de la base, auquel serait adjoint un outil de consultation cartographique. Le projet SIPIS (*South India Population Information System*), lancé en janvier 2000, a duré moins d'un an, en raison de l'état presque achevé de la base de données. Il a été consacré à l'état du Tamil Nadu et soutenu par un financement du Fond des Nations Unies pour la Population (FNUAP). L'effort a avant tout porté sur le développement d'un outil convivial de consultation, permettant de passer aisément des données aux cartes. Du point de vue de l'information géographique, ce projet nécessitait la production de nouvelles données, calculées à partir des données brutes du recensement, afin que l'utilisateur final n'ait qu'à appuyer sur un bouton pour obtenir une information construite comme un indice d'alphabétisation ou une carte des frontières administratives. Toutes les informations statistiques ont été regroupées en grandes familles (données sociales, économiques, d'emploi, d'infrastructure, etc.). De même, quelques cartes de synthèse sont fournies en format raster (alphabétisation, fécondité...) afin d'aider l'utilisateur à identifier les tendances régionales.

Figure 5 : Exemple d'écran du cdrom SIPIS.

La figure 5 présente une cartographie réalisée à partir du SIPIS. Y figurent une cartographie de la main-d'oeuvre dans le secteur secondaire à l'échelle des taluks, les couches urbaines et routières ainsi que les données démographiques pour une unité particulière. La programmation du logiciel a été faite à partir des outils développés par l'entreprise ESRI et le produit final ressemble au logiciel existant *Arc Explorer*. L'interface a toutefois été simplifiée afin de le rendre utilisable par des personnes non formées. Quelques options ont été ajoutées, notamment une aide en langue tamoule⁸. Le résultat (Guilmoto *et al.*, 2000) a été jugé très favorablement, d'autant que la subvention du FNUAP a permis de le vendre à bas prix. L'importance des commandes de SIPIS auprès de l'Institut français de Pondichéry confirme la demande pressentie.

Suite au succès de cette première valorisation auprès du grand public, la diffusion d'information géographique préalablement traitée a été envisagée. En plus de l'information brute diffusée par cdrom, pourquoi ne pas utiliser Internet comme vecteur gratuit et décentralisé d'une information géographique préparée par des chercheurs et brièvement commentée? De cette question est née l'idée de réaliser un atlas en ligne. La supériorité de l'atlas sur une banque de données pour la diffusion de l'information géographique est de deux ordres: graphique et scientifique. Graphique bien sûr: l'information traitée est directement lisible. Scientifique ensuite car l'information est sélectionnée: des chercheurs ont choisi les cartes à présenter en fonction de leur intérêt et selon un classement thématique. De plus, les cartes sont commentées, remises dans leur contexte.

L'abandon du support papier traditionnel et la réalisation en ligne d'un atlas ont des avantages évidents: moindre coût puisque la diffusion est presque gratuite, grosse capacité (le nombre de cartes est quasiment illimité), comparaison simplifiée (par variation des échelles, un phénomène local peut être replacé dans un contexte régional, national, voire international), et mise à jour rapide au fur et à mesure de l'apparition de nouvelles données. Sur la figure 6, nous illustrons un écran de ce site (www.demographie.net/atlasofindia), reprenant ici une carte de l'urbanisation indienne dérivées du recensement de 2001.

Figure 6 : Exemple d'écran du serveur cartographique du SIFP.

⁸ En revanche, une version entièrement en tamoul s'est avérée impossible, notamment en raison de l'absence d'une liste des localités en langue tamoule. Les listes existantes datent paradoxalement de la période coloniale et ne correspondent plus au maillage villageois actuel.

Des problèmes techniques d'un nouvel ordre se sont posés pour la création du site Web. Problème de structuration d'abord, afin de regrouper thématiquement les 150 cartes disponibles. La structuration régionale (choix par menu déroulant) a été retenue, la sélection de l'entité géographique s'effectuant par simple clic sur le niveau choisi (Inde, Inde du Sud, Andhra Pradesh, Karnataka, Kerala, Tamil Nadu). De plus, il a fallu permettre à des personnes utilisant des accès Internet à faible débit, comme c'est le cas en Inde, de consulter ce site sans obtenir des cartes grossières comme celles actuellement disponibles sur les serveurs cartographiques. Le choix a été fait d'une présentation la plus légère possible, à partir de cartes précises mais de taille informatique très réduite. Le format des images s'est porté sur le GIF, dont l'une des caractéristiques est le codage en 256 couleurs. Celui-ci allège le poids des cartes, mais a entraîné de longues phases d'expérimentation, l'image visualisée sur le logiciel de SIG subissant de nombreuses transformations avant d'être publiée en ligne. Si la qualité des images ne peut égaler celles des atlas usuels, elle est en partie compensée par la rapidité de mise en ligne et le faible coût d'accès.

4. *Lissage et agrégation des données*

Les phases décrites jusqu'à présent feraient presque oublier que le projet initial n'avait pas été conçu pour la restitution et la mise à disposition d'un matériel documentaire abondant. Au contraire, cette valorisation n'est qu'un produit dérivé, puisque l'effort des participants au projet visait avant tout à enrichir des problématiques se construisant autour de la spatialisation du changement social en Inde. Or, dès que notre outil géomatique fut mis en place, le volume de l'information s'est soudain avéré trop important pour permettre une exploitation efficace dans un objectif de recherche. Nous avons en effet une base spatialisée forte de plus de 75.000 villages, dotés de plus d'une centaine d'indicateurs, ne pouvant guère être cartographiée, sinon à une échelle micro-régionale. En effet, d'un point de vue technique, ou même visuel, cette quantité de données ne peut être traitée de manière systématique.

Les unités censitaires correspondaient en outre à un assemblage fort disparate de "villages"; plus de 10 % d'entre eux contenaient moins de 100 habitants, alors que plusieurs centaines des "villages" les plus peuplés, situés avant tout au Kerala, dépassaient au contraire les 10.000 habitants. De même, l'étude de robustesse statistique démontrait que nombre des indicateurs utilisés (taux, rapport ou pourcentage) devenaient dangereusement instables quand la population de référence était d'environ 200 personnes, voire totalement dépourvus de signification pour les unités encore moins peuplées.

Il était par conséquent nécessaire d'adopter une procédure de lissage, qui évitât toutefois d'attribuer à chaque unité un poids identique, comme cela est courant dans les programmes de lissage contenu dans les logiciels SIG commerciaux. En effet, la taille variable des villages (en termes de population, de superficie, etc.) rendrait aberrant le lissage direct des valeurs observées: les zones les plus marginales, comme les nombreux villages inhabités ou très peu peuplés des Ghats méridionaux, viendraient ainsi à peser autant que les villages surpeuplés du Kerala qui les bordent, sabotant ainsi l'effort de synthèse de l'entreprise d'agrégation.

Différentes méthodes de regroupement furent envisagées, mais il fut très vite évident qu'on ne pouvait faire usage du maillage administratif pour des raisons similaires. Les unités les plus petites, à savoir les taluks, n'étaient pas du tout de superficie ou de population comparable d'un état à l'autre. D'ailleurs, en Andhra Pradesh, les taluks avaient été supprimés avant 1991 et remplacés par des unités beaucoup plus réduites (les *mandal*), de dix à vingt fois plus petites que les taluks des régions frontalières.

Nous avons donc décidé de procéder à un groupement des villages exclusivement à partir d'un critère de proximité spatiale, et cela indépendamment des frontières administratives, y compris celles des états. La méthode ordinaire des carroyages ne semblait guère appropriée, car elle aurait pour effet de pénaliser les zones de bordure, le long de la mer ou des frontières septentrionales, qui seraient de taille inférieure au carroyage standard. En introduisant des unités nécessairement plus petites que les autres, dans les confins administratifs ou côtiers, on créerait un biais spatial.

La procédure d'agglomération fut la suivante. On identifia en premier lieu de manière automatique un point d'agglomération autour des villages, après avoir fixé une distance maximale. A partir de ce premier maillage de points d'agrégation, on créa un découpage spatial en polygones de Thiessen (ou Voronoi) recouvrant l'Inde du sud⁹. Ces polygones, sans être identiques, sont de surface très comparable, y compris dans les régions de bordure de notre carte d'Inde du sud. A l'intérieur de chaque aire, les villages sont identifiés et les valeurs observées cumulées. On a défini de cette manière un ensemble de "clusters", c'est-à-dire des regroupements constitués par un nombre variable d'unités primaires villageoises. Les indices sociodémographiques sont ensuite recalculés au sein de chaque cluster pour obtenir une valeur agrégée de l'ensemble. La même opération a été conduite pour quatre jeux distincts d'agrégation, déterminés par des distances respectivement de 2, 5, 10 et 20 kilomètres, créant par conséquent un nombre décroissant de clusters.

Il va sans dire que ces regroupements améliorent considérablement la qualité des statistiques, puisque le nombre d'unités sous-peuplées devient aussitôt modéré, voire négligeable. Ainsi, en passant à l'agrégation avec une distance de deux kilomètres, le nombre d'unités statistiquement vulnérables de moins de 200 habitants passe de 15,8 % à 2,1 % de l'échantillon, et cette proportion devient proprement négligeable aux échelles d'agrégation supérieures. Dans le même temps, le nombre d'unités décroît, passant ainsi respectivement à 24.000 (pour une distance de 2 km), 7.000 (pour 5 km), 2000 (pour 10 km) puis 600 (pour 20 km), ce qui rend possibles toutes les cartographies envisagées. Il suffit de faire varier le niveau d'agrégation utilisé selon l'échelle retenue, choisissant notamment des agrégations plus fortes pour des cartes de l'ensemble du sud de l'Inde. De la même façon, grâce à cet échantillon plus petit et de meilleure qualité, des modélisations statistiques ou géostatistiques (voir plus bas) sont à présent réalisables dans de meilleures conditions de calcul.

Figure 7 : Etapes d'une cartographie du nord-est de l'Andhra Pradesh : agrégation, krigeage et lissage des données de fécondité rurale de 1991.

Sur la figure 7, nous résumons ainsi les différentes étapes de l'étude d'un phénomène, partant des données primaires de notre SIG pour obtenir une cartographie synthétique. En l'occurrence, il s'agit ici de l'indice de fécondité, calculé comme le rapport entre le nombre d'enfants de moins de sept ans et la population féminine de plus de sept ans¹⁰, et la zone représentée recouvre la pointe nord-est de l'état de l'Andhra Pradesh, centrée autour du district de Visakhapatnam. Sur la première carte, l'ensemble des 6 700 villages de cette région est représenté: on soulignera d'emblée que les plus fortes densités des unités de peuplement ne correspondent pas nécessairement à une plus forte densité démographique. Ainsi, une avancée vers le nord (à l'intérieur de l'état d'Orissa) comprend plusieurs centaines de villages, qui ne sont en réalité que des hameaux tribaux isolés, de taille démographique rarement supérieure à 100 habitants. Inversement, les villages sur le littoral orienté vers le sud-est ont des populations moyennes beaucoup plus importantes, mais sont très dispersés.

La seconde carte de la figure 7 réduit cet échantillon à un ensemble d'environ 500 clusters dans cette poche rurale de l'Andhra Pradesh. Les nouvelles unités constituées sont dispersées très régulièrement sur la surface et leur taille démographique moyenne reste importante. Seul un nombre négligeable d'entre elles compte à présent moins de 1000 habitants. De plus, les valeurs de l'indice de fécondité sont nettement moins accidentées, calculées à présent sur des agrégats de plusieurs milliers de personnes. Ils font l'objet d'un traitement par krigeage¹¹ sur la carte suivante, afin de passer d'une représentation vectorielle à une représentation par surface (raster), que nous avons choisi de figurer ici à l'aide de carreaux de 5 km de côté.

La dernière carte est obtenue par contourage des valeurs de fécondité lissées par krigeage ordinaire. La fécondité varie ici du simple au double de la partie centrale de la région à la frontière nord-ouest de l'état où la fécondité est la plus élevée. C'est pour cela que nous avons choisi de représenter ici cette région, car les différentiels y sont les plus forts d'Inde du sud: la partie centrale est une région rizicole, avec un fort développement industriel autour du port de Visakhapatnam (ville de plus d'un million d'habitants), qui entre dans

⁹ Sur les polygones de Thiessen et l'agrégation, voir par exemple Pumain et Saint-Julien (1997) ou Martin (1996).

¹⁰ Il s'agit du "rapport enfants-femmes", indice comparatif très fortement corrélé au niveau local de fécondité et utilisé pour notre analyse de la baisse de la fécondité. L'effet perturbateur de la mortalité infanto-juvénile reste modeste dans la mesure des variations entre régions.

¹¹ La technique du krigeage (ou kriging) est décrite par exemple dans Fotheringham et al. (2000).

la phase finale de sa transition démographique, avec des taux de fécondité parfois proches de 2.5 enfants par femme. Simultanément, les régions forestières du nord longeant le Madhya Pradesh et l'Orissa sont parmi les plus déshéritées de l'Andhra Pradesh, avec un peuplement souvent de majorité tribale et des niveaux d'alphabétisation très faibles: la fécondité y demeure très élevée et l'impact de la proximité des régions riches et urbanisées de la plaine littorale semble invisible, tant la barrière sociale entre les groupes culturels est étanche.

5. *Quelques exemples d'analyse spatiale*

On pourrait, certes, procéder à une analyse plus fine de cette cartographie de la fécondité, soulignant notamment l'impact local des agglomérations urbaines ainsi que celui des axes routiers dans le nord-est de l'Andhra Pradesh. Cependant, nous voulons ici avant tout résumer une démarche que nous avons appliquée à un large choix de phénomènes d'ordre variable au sein du programme Population et Espace en Inde du Sud, conduit par l'IRD et l'IFP à Pondichéry. Ainsi, un travail en cours par Stéphanie Vella utilise la cartographie des rapports de masculinité infantiles pour étudier les discriminations sexuelles au Tamil Nadu.¹² Similairement, un travail sur l'urbanisation est conduit par Sébastien Oliveau, qui étudie l'effet de la proximité urbaine sur les campagnes en comparant les villages selon leur localisation spatiale dans la carte urbaine de l'Inde du sud. Le fond villageois sert également à une recherche doctorale conduite par Virginie Chasles sur l'infrastructure sanitaire d'un district de l'Andhra Pradesh (Chasles 2001). Une autre étude a été consacrée à un milieu forestier des Ghats du Karnataka, menacé par la pression de l'économie de plantation (Guilmoto, 2000).

Un exemple supplémentaire est fourni par la cartographie présentée sur la figure 8, qui reprend des données sur l'irrigation (Guilmoto 2002). Il s'agit ici du pourcentage de terres irriguées dans le terroir cultivé des villages, données qui n'avaient jamais fait l'objet d'un traitement systématique à l'échelon local¹³. Il ne s'agit là que d'une carte de synthèse, dont l'interprétation nécessiterait une comparaison avec les cartes de l'hydrographie, du relief et des sols. Toutefois, le gradient ouest-est de l'irrigation, étroitement relié à la pente allant des Ghats méridionales vers le littoral, se dessine clairement, tout comme l'effet des fortes pluviométries observées sur les côtes. On observe en outre, dans les régions plus arides du plateau du Deccan (dans les zones centrales des états du Karnataka ou de l'Andhra Pradesh), l'effet de grands aménagements hydro-agricoles qui ont permis l'établissement d'importantes poches irriguées. Les valeurs les plus fortes, proches du tout irrigué, sont observées principalement dans les deltas de la Baie du Bengale, à savoir à l'embouchure des fleuves Kaveri, Krishna et Godavari qui prennent leur source dans les Ghats.

Figure 8 : Pourcentage d'irrigation des terres cultivées en zones rurales, Inde du sud, 1991.

Cet exercice, pour l'ensemble de l'Inde du sud, est réalisé à partir de l'agrégation sur 10 km, et dissimule par conséquent les variations micro-locales dans la disponibilité de l'irrigation qui auraient donné à la carte un profil beaucoup plus tourmenté. Certains terroirs disposent en effet de modes d'irrigation très locaux, qu'il s'agisse de puits, de retenues aménagées, voire de canaux dérivés de certains cours d'eau pérennes. On notera en outre que les zones d'irrigation qui apparaissent n'épousent absolument pas les frontières administratives, y compris celles des districts qui constituent le niveau le plus fin auquel ces statistiques sont habituellement publiées en Inde. Ce qui signifie qu'une telle analyse spatiale révèle à bon escient des unités géographiques, parfois de simples bassins versants, que les unités politiques auraient fortement tendance à gommer.

Pour finir, nous fournirons un autre exemple d'application de notre base de données à l'analyse spatiale, ne relevant plus cette fois-ci de la cartographie. Nous avons en effet utilisé les données relatives à différentes variables pour une analyse géostatistique de l'autocorrélation spatiale. Nous nous limiterons ici à la présentation

¹² Les statistiques villageoises peuvent être lissées pour faire apparaître la géographie singulière du déficit féminin au Tamil Nadu. Le déséquilibre entre les sexes parmi les enfants est lié aux avortements sélectifs et à l'infanticide (Vella 2002).

¹³ Olivia Aubriot, qui a dressé des cartes des modes d'irrigation au Tamil Nadu, a produit en 1999 les premières analyses à partir de la base de données villageoises.

de données concernant quatre variables, pour lesquelles nous avons calculé le coefficient de Moran¹⁴. L'indice de Moran, dont la définition est donnée plus bas, permet de mesurer le niveau de corrélation statistique $I(h)$ entre les n paires d'observations localisées (les valeurs z_i ou z_j), classées selon la distance d les séparant.

$$I(d) = \frac{m \sum_{i,j} (z_i - \bar{z})(z_j - \bar{z})}{n \sum_i (z_i - \bar{z})^2}$$

Pour simplifier, disons que l'indice est en numérateur la covariance entre les observations séparées par une distance donnée, rapportée à la variance totale de l'échantillon de taille m en dénominateur. Cet indice prend une valeur proche de 1 quand la corrélation est maximale entre paires d'observations, ce qui peut être le cas quand les localités sont très rapprochées. Il prend une valeur nulle quand il n'y a aucun rapport statistique entre elles, et une valeur de -1 au contraire quand les observations d'une même paire sont diamétralement opposées, cas fort rare. La corrélation spatiale des caractéristiques sociales est rarement nulle pour les observations rapprochées, mais son intensité est très variable et l'examen visuel des cartes ne permet guère d'en rendre compte. Le coefficient de Moran autorise en revanche une analyse systématique des effets de proximité et par conséquent de la cohésion spatiale des phénomènes.

Figure 9 : Coefficients de Moran calculés pour quatre variables, Inde du sud, 1991.

Les résultats présentés sur la figure 9 proviennent des 2151 agrégations (clusters) à 10 km. De taille similaire et de répartition uniforme, ces clusters sont donc dépourvus de tout biais spatial ou administratif. Les valeurs de l'autocorrélation spatiale sont calculées sur des pas de 15 km, en nous limitant aux observations distantes de moins de 300 km. Nous ne montrons ici que les résultats correspondants à certaines variables de la base de données. Ils indiquent que l'alphabétisation (en pourcentage) et la fécondité sont de loin les variables les plus corrélées spatialement: les valeurs s'élèvent à plus de 0,8 pour les localités rapprochées et baisse lentement alors que les distances augmentent, restant significatives à plus de 300 km de distance. De manière plus fine, on observe même sur cet échantillon que la structuration spatiale de la fécondité ainsi mesurée semble légèrement plus accentuée que celle de l'alphabétisation, car plus étalée dans l'espace.

En revanche, le profil du peuplement tribal fait état d'une répartition spatiale très différente, car très concentré à l'échelle locale (moins de 15 km de périmètre), mais beaucoup moins étalé dès que l'on s'éloigne du foyer tribal central. Un dernier indice est également représenté, celui des travailleurs du secteur de la construction, mesuré également en pourcentage de la main-d'oeuvre totale. L'autocorrélation en est très réduite et tend d'ailleurs à disparaître complètement au-delà de 50 km. Ce profil s'explique sans doute par une répartition uniforme de cette main-d'oeuvre, sans lieu de concentration spécifique, car les besoins du secteur de la construction se font sentir partout.

6. Conclusion

Le propos de cet article est resté volontairement descriptif, plutôt qu'analytique, car nous voulions restituer l'ensemble d'une démarche de géomatique. Il est rare aujourd'hui que les géographes soient conduits à parcourir tout cet itinéraire, s'étirant des mesures du géoréférencement à partir de cartes publiées de qualité incertaine jusqu'aux tentatives de modélisation spatiale ou géostatistique. Mais s'il est nécessairement beaucoup plus long que ce bref article ne tendrait à le faire croire, ce parcours de recherche a de nombreuses vertus pour la réflexion disciplinaire.

En premier lieu, il rappelle que de nombreuses régions souffrent d'un manque criant de cartographie moderne, pour un ensemble de raisons qui découlent globalement du sous-équipement technique de l'appareil administratif

¹⁴ Sur le coefficient de Moran, voir par exemple Bailey and Gatrell (1995).

ou scientifique local et que la rapidité des changements, souvent géographiques, contribue paradoxalement à freiner. Dans l'Inde que nous étudions ici, la fréquence et l'ampleur des remaniements de frontières, qu'il s'agisse des territoires urbains ou des unités politiques comme les districts ou les états, tendent très certainement à décourager les gros investissements cartographiques que quelques années de réforme risquent de rendre obsolètes. Les études sont plus fréquemment conduites à partir de l'imagerie satellitaire, qui a l'avantage de rester indifférentes aux redécoupages administratifs continus. L'effort de digitalisation local a, certes, débuté en Inde, mais se concentre plus sur les régions avancées (par exemple, les villes) que sur le monde rural. En Inde, le volume des informations est également réducteur pour une entreprise qui se voudrait exhaustive: nous avons consacré plusieurs années à la cartographie des villages d'Inde méridionale, mais il demeure encore quatre fois plus de villages dans le reste du pays. L'effort à déployer reste incontournable et les investissements consentis dans cette direction par les chercheurs doivent très naturellement participer à l'enrichissement du domaine public.

En second lieu, les résultats que notre article a brièvement présentés confirment pour l'essentiel la logique du projet. Nous pensons en effet que l'approche spatiale apporterait un éclairage nouveau à la connaissance de nombreux phénomènes. Cela semble désormais confirmé. D'une part, la cartographie fine fait parfois apparaître l'ampleur de phénomènes locaux ou sous-régionaux qu'un maillage statistique classique dissimule. D'autre part, comme cela se sait, le SIG permet de croiser les informations de source variable et donc de faire bon usage du large choix des données disponibles. Enfin, de nouvelles questions de recherche trouvent ici de quoi se nourrir, car la très forte structuration spatiale des phénomènes de changement social apporte de nombreux éléments de réponse à l'étude de leur origine et surtout, de leur mode de propagation dans l'espace et le temps. Des questions plus théoriques, comme l'origine de la très forte compacité spatiale de la baisse de la fécondité en Inde, trouvent ainsi de solides bases d'analyse.

Pour finir, notre parcours indique également que le développement de la technologie ne se fait pas qu'au profit des outils de traitement scientifique, que nos calculs géostatistiques illustrent. De puissants outils de restitution de l'information sont à présent disponibles qui visent à rendre plus lisible une information, qui était peut-être jadis existante, mais dont l'inaccessibilité chronique rendait l'usage fort peu fréquent notamment hors des cercles de spécialistes universitaires. A l'époque de la gestion décentralisée indienne (*Panchayati Raj*), le besoin d'information désagrégée se fait de plus en plus sentir, afin de conduire un pilotage local des opérations de développement des ressources. Certaines régions en pointe, comme le Kérala, ont d'ailleurs intégré une cartographie locale aux premiers plans de développement des *Panchayats*. Il est donc important de disposer d'outils à même de livrer l'information à des utilisateurs très variés, allant des administratifs aux étudiants, en passant par les ONG ou les entreprises, qui semblent d'ailleurs avoir été les seules à nous faire concurrence dans ce domaine.

Parmi ces outils, les cdroms ont l'avantage d'offrir des supports fort peu coûteux, se substituant aux atlas que les coûts actuels de production et d'impression rendent difficilement publiables dans les pays en développement. La possibilité d'utiliser des modes d'interrogation interactifs par des logiciels de consultation cartographique rend l'outil encore plus attrayant et l'accès à l'information plus facile. Les publications en ligne, via le Web, représentent des atouts supplémentaires, notamment en termes de mise à jour et d'accessibilité délocalisée. Les retombées non scientifiques d'un projet, doté d'une politique active de valorisation et de diffusion, peuvent donc être importantes et faire taire, temporairement, l'opinion communément admise sur le caractère relativement oiseux des entreprises de recherche fondamentale dans les pays en développement.

Références:

- Bailey, T. C., and Gatrell, A. C., 1995. *Interactive Spatial Data Analysis*, Longman, Harlow.
- Chasles, V., 2001. *L'environnement culturel de la santé maternelle et périnatale en milieu rural en région bordière Andhra Pradesh/Karnataka*, séminaire « les défis de l'Inde : enjeux et perspectives au XXI^e Siècle », Association Jeunes Etudes Indiennes, Paris, 27 novembre.

- Fotheringham, S., Brundson, C, and Charlton, M., 2000. *Quantitative Geography. Perspectives on Spatial Data Analysis*, Sage, London.
- Guilmoto, C. Z., 2002. "Irrigation and the great Indian rural database. Vignettes from south india", *Economic and political weekly*, March 30, 1223-1228.
- Guilmoto, C. Z., 2000 "Demographic and Environmental Changes. Population Growth and Spatial Dynamics", in Ramakrishnan, P.S. , *et al.*, *Mountain Biodiversity, Land Use Dynamics, and Traditional Ecological Knowledge*, Oxford and IBH, New Delhi, 54-70.
- Guilmoto, CZ, and Oliveau S., with S. Vingadassamy and R. Amuda, 2000, *South Indian Population Information System, Volume I: Tamil Nadu and Pondicherry*, CD-ROM published by the French Institute, Pondicherry.
- Kraak, M.-J., Ormeling, F., 1996, *Cartography. Visualisation of spatial data*, Longman, Harlow.
- Longley, P. and Batty, M., eds., 1996. *Spatial Analysis: Modelling in a GIS Environment*, GeoInformation International, Cambridge.
- Martin, D., 1996. *Geographic Information Systems. Socio-economic Applications*, Routledge, London and New York.
- Oliveau, S., 1998, *Atlas du district de Vellore : essai de typologie régionale*, mémoire de maîtrise de l'université de Paris X Nanterre, non publié.
- Pumain, D., et Saint-Julien, T., 1997. *L'analyse spatiale. 1. Localisations dans l'espace*, Armand Colin, Paris.
- Vella, S. 2002. "Trends and differentials of female discrimination in South India: a case study of Salem district", chapitre à paraître in Guilmoto, C.Z., and S. Irudaya Rajan, eds. *Currents of Demographic Change in South India*.

Table des figures

Figure 1 : Carte topographique de la région de Vellore (Tamil Nadu, Inde), 1914-15.

Figure 2 : Localisation des localités du taluk de Vellore (Tamil Nadu, Inde) d'après la carte du recensement de 1981.

Figure 3 : Recalage de la carte des localités (Figure 2) sur la carte topographique (Figure 1).

Figure 4 : Localisation des villages du recensement de 1991 sur la frontière entre Karnataka et Kerala, Inde.

Figure 5 : Exemple d'écran du cdrom SIPIS.

Figure 6 : Exemple d'écran du serveur cartographique du SIFP.

Figure 7 : Etapes d'une cartographie du nord-est de l'Andhra Pradesh : agrégation, krigeage et lissage des données de fécondité rurale de 1991.

Figure 8 : Pourcentage d'irrigation des terres cultivées en zones rurales, Inde du sud, 1991.

Figure 9 : Coefficients de Moran calculés pour quatre variables, Inde du sud, 1991.

Figure 1 : Carte topographique de la région de Vellore (Tamil Nadu, Inde), 1914-15.

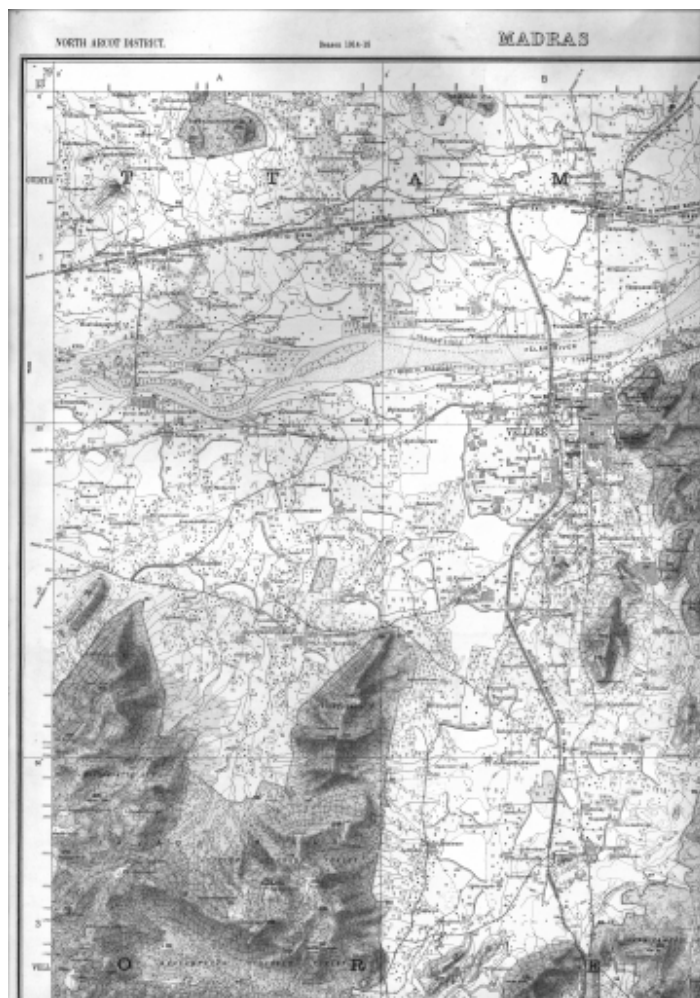


Figure 2 : Localisation des localités du taluk de Vellore (Tamil Nadu, Inde) d'après la carte du recensement de 1981.

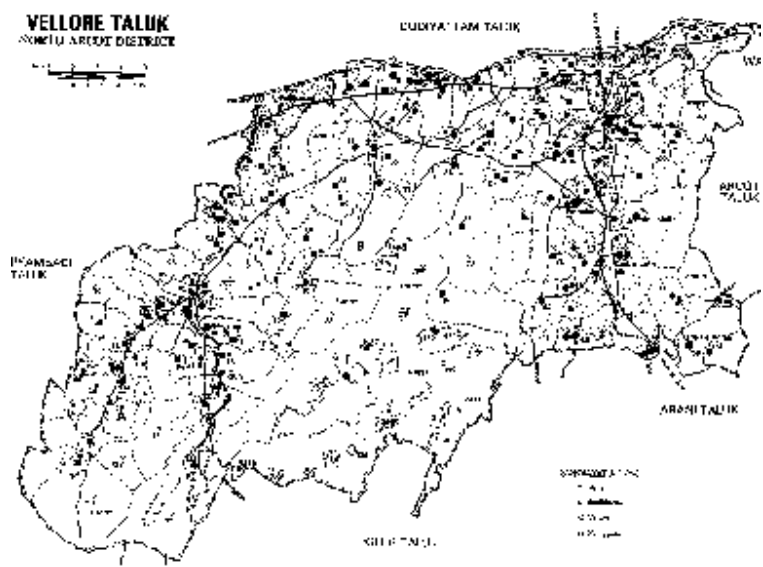


Figure 3 : Recalage de la carte des localités (Figure 2) sur la carte topographique (Figure 1).

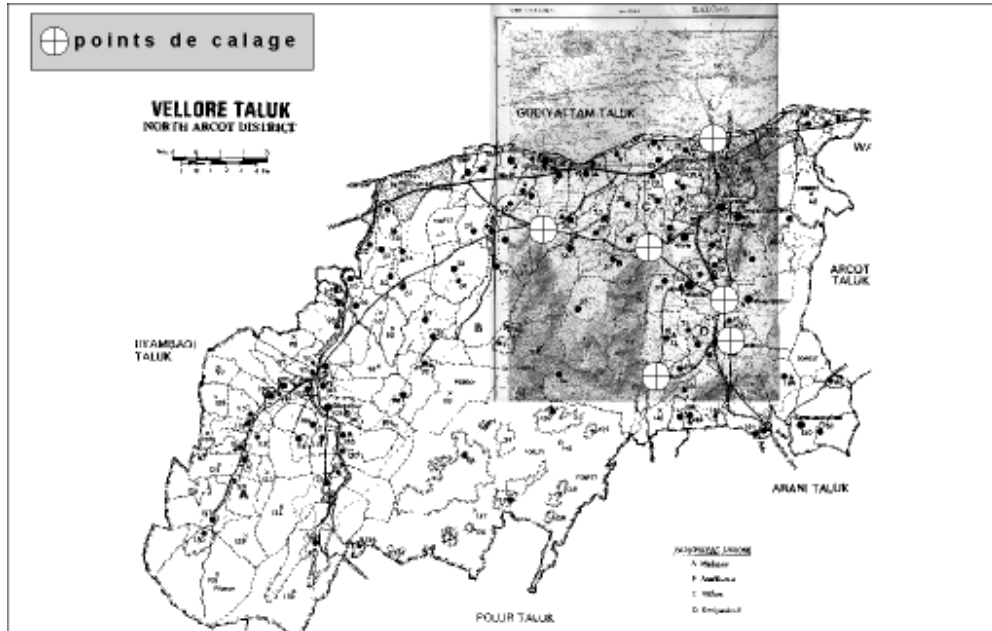


Figure 4 : Localisation des villages du recensement de 1991 sur la frontière entre Karnataka et Kerala, Inde.

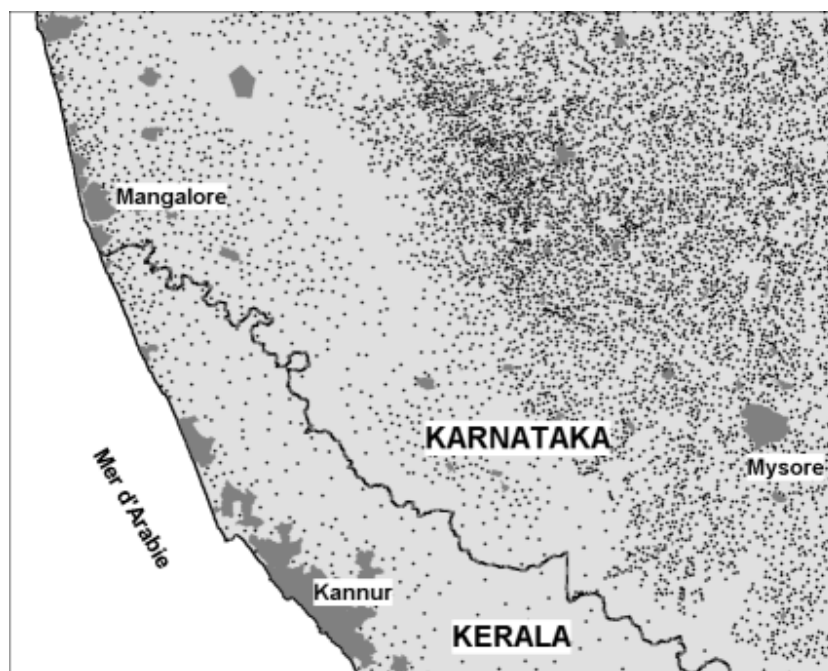


Figure 5 : Exemple d'écran du cd-rom SIPIS.

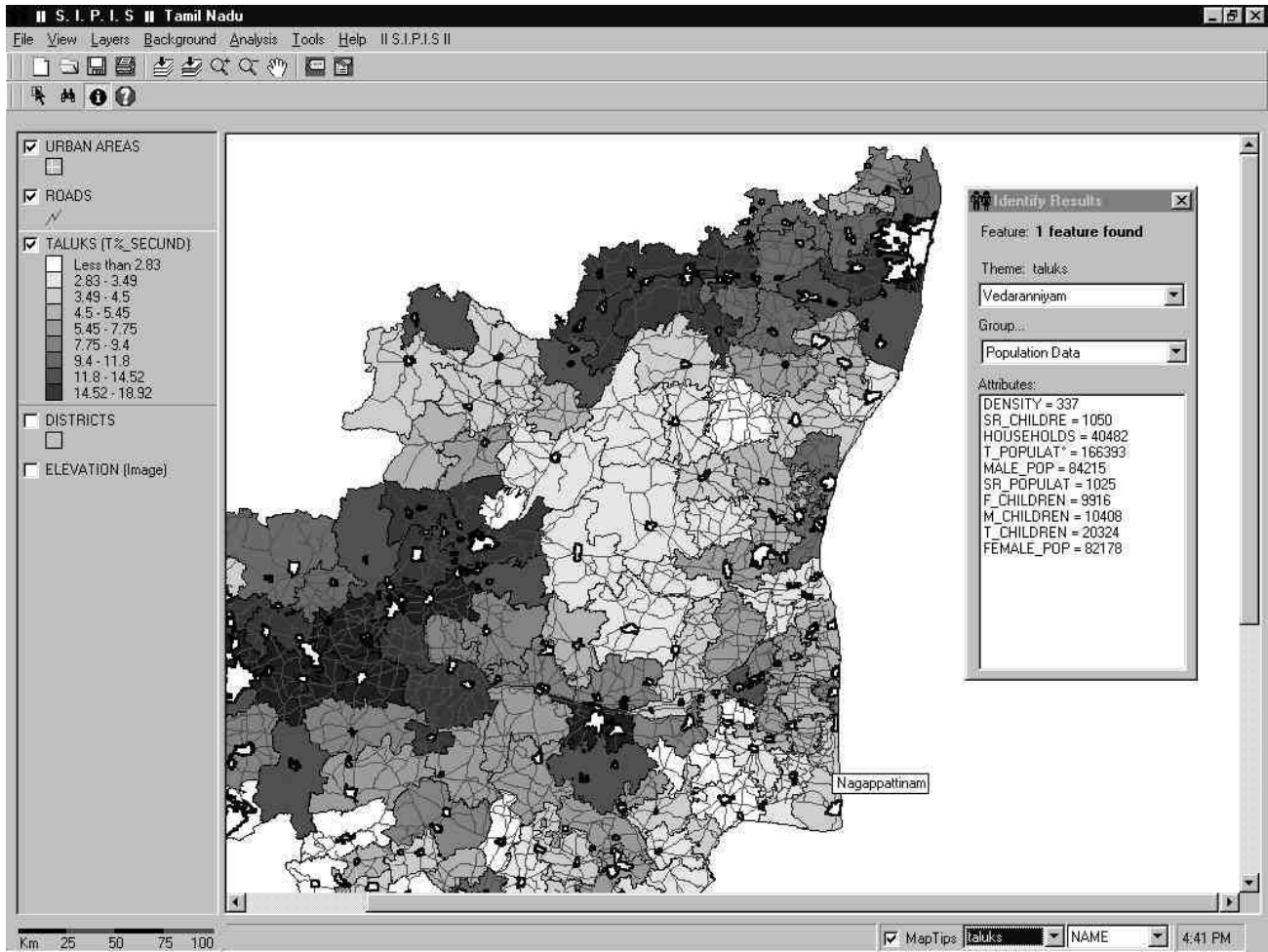


Figure 6 : Exemple d'écran du serveur cartographique du SIFP.

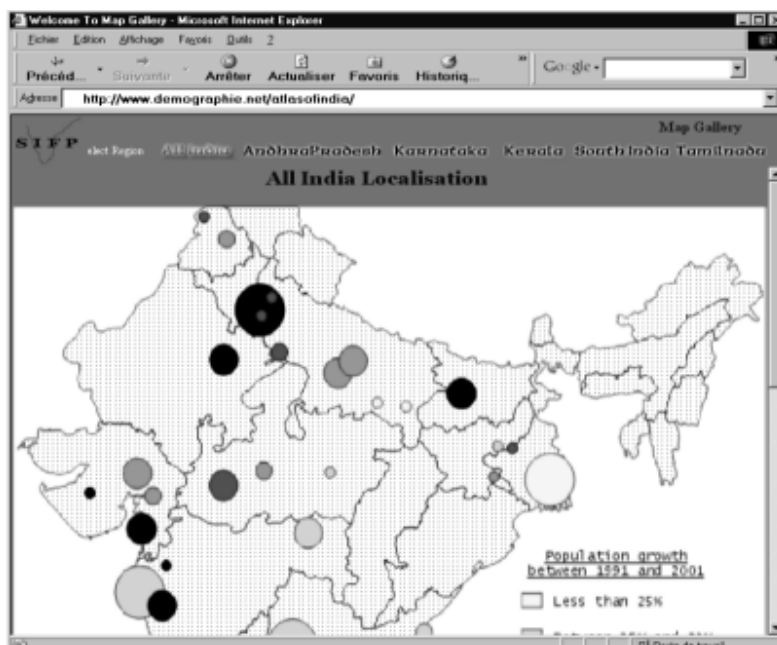


Figure 7 : Étapes d'une cartographie du nord-est de l'Andhra Pradesh, Inde: agrégation, krigeage et lissage des données de fécondité rurale de 1991

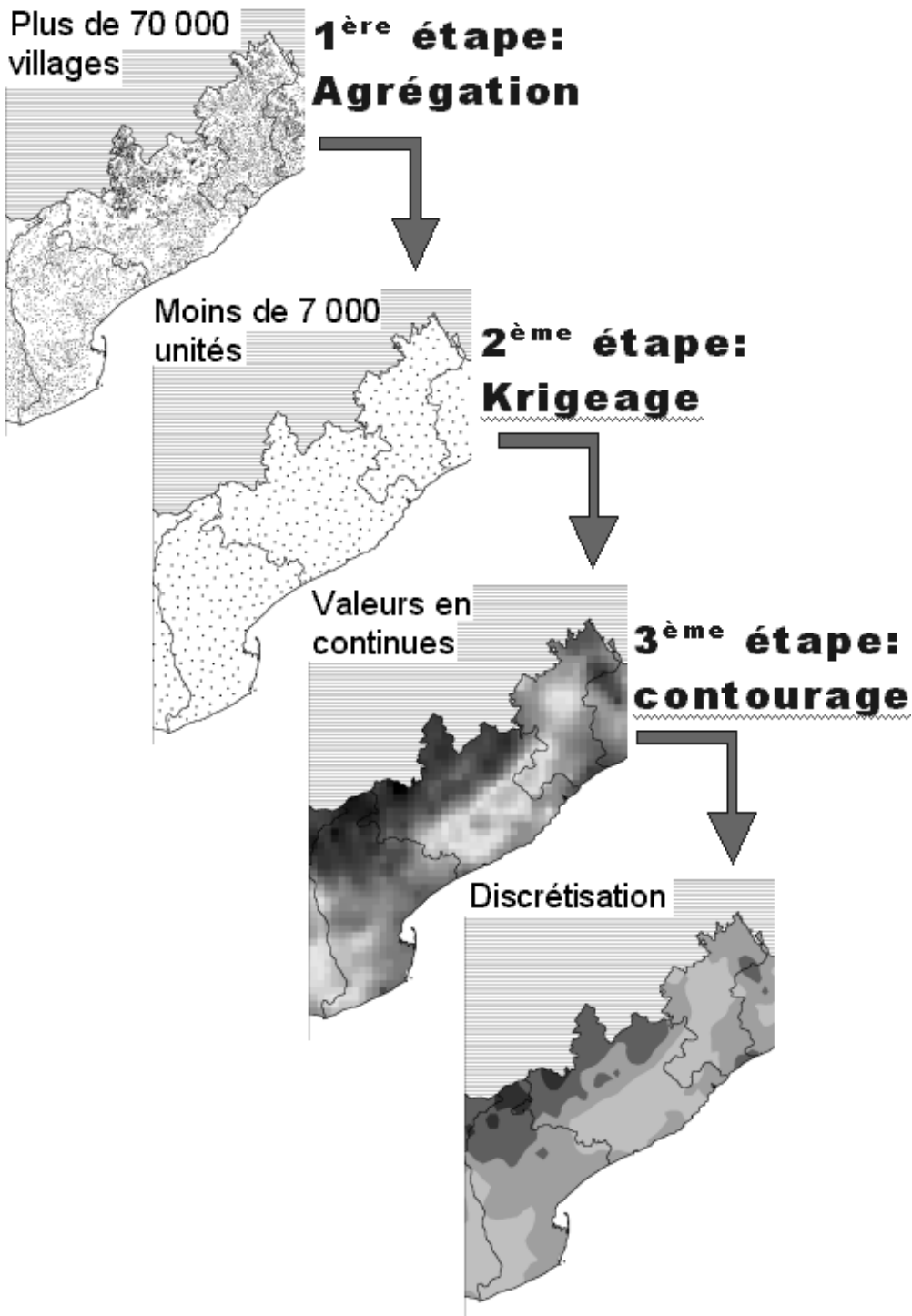


Figure 8 : Pourcentage d'irrigation des terres cultivées en zones rurales, Inde du sud, 1991.

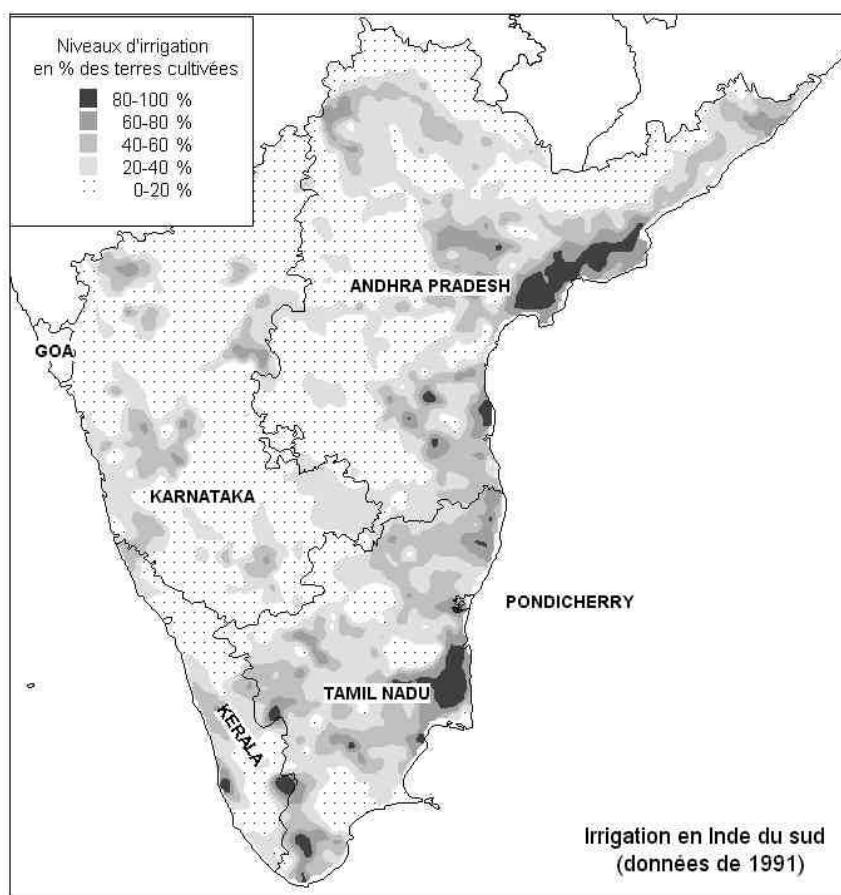


Figure 9 : Coefficients de Moran calculés pour quatre variables, Inde du sud, 1991.

