

Espace et mesure en Inde du Sud (EMIS)

L'espace de la vulnérabilité en Inde du sud

Rapport scientifique

Préparé par Christophe Z Guilmoto et les membres de l'EMIS :

Olivia Aubriot, IFP/CNRS

Virginie Chasles, Un. de Rouen

Christophe Z Guilmoto, LPED/IRD

Sébastien Oliveau, Un. de Provence

Katia Radja, Un. de St. Quentin en Yvelines

Alain Vaguet, Un. de Rouen

Stéphanie Vella, Un. de Bordeaux

Financement de l'appel à propositions GETM
Programme National de Recherche « Société de l'information » CNRS

Paris
Juillet-Septembre 2005

Le projet EMIS a été financé par l'appel à propositions « Géomatique, Espace, Territoires et Mobilités » (GETM), (« visant à renforcer l'interdisciplinarité entre chercheurs en sciences sociales et informaticiens dans le traitement des problématiques liées au traitement de l'espace ») au sein du Programme National de Recherche « Société de l'information » du CNRS.

Le projet utilise des données rassemblées précédemment à l'occasion du South India Fertility Project, programme de recherche de l'Institut français de Pondichéry et de l'Institut de Recherche pour le Développement (UMR LPED), appuyé par le Wellcome Trust.

Ce rapport a été préparé par les membres de l'équipe EMIS et coordonné par Christophe Z Guilmoto (IRD-LPED). L'appui de Donatien Chatillon a également été apprécié. Les sections individuelles de ce rapport ont été préparées plus particulièrement par les personnes suivantes :

- *Partie I, sections 3.4 et 4.4.3, Annexe I : Sébastien Oliveau*
- *Section 4.1 : Katia Radja*
- *Section 4.4.1 : Olivia Aubriot*
- *Section 4.4.4 : Virginie Chasles*

Une version en couleur de ce rapport est disponible sur le site de l'EMIS (www.demographie.net/emis)

Abstract: SPACE AND MEASURE IN SOUTH INDIA (EMIS). THE SPATIAL CONTEXT OF VULNERABILITY IN SOUTH INDIA.

The present report describes the EMIS (www.demographie.net/emis) research project that brought together a team consisting of four geographers, an economist, a demographer and an anthropologist. The report is divided in four parts.

1. The first part sums up the geographical perspective to study social change and describes the spatialized census database (GIS) used for the project (South Indian villages from the 1991 census).
2. The second part describes the original spatial aggregation procedure used by the team to prepare a new database based on spatial clusters in four different formats.
3. The third part is a more systematic study of spatial patterning of social and economic characteristics of South Indian localities. The main tools used here are Moran's index of spatial autocorrelation and Anselin's local indicator of spatial association (LISA).
4. The fourth part of the report is an application of our database (complemented with rainfall information on South India) to devise a local indicator of vulnerability. This section starts with a brief presentation of the economic issues involved in poverty mapping, followed by a discussion of the spatial patterning of vulnerability's main dimensions in South India (with a special section devoted to agriculture and irrigation). This part concludes with the computation of a composite index of vulnerability and presents briefly the resulting map of rural vulnerability in South India.

Appendices to this report present Moran's scatterplot as a tool for spatial analysis and the application of the cokriging technique to micro-level estimation.

Plan du rapport

0.	Introduction : de l'autocorrélation à la pauvreté.....	5
1.	Etudier la spatialité socio-économique en Inde du sud.....	7
1.1.	Pourquoi l'espace ? Diffusion et répartition	7
1.1.1.	A la source des déséquilibres spatiaux : la centralité	8
1.1.2.	La diffusion comme élément de transformation de l'espace.....	9
1.2.	Description générale de la base.....	11
1.2.1.	Spatialiser les données censitaires.....	11
1.2.2.	L'accès aux cartes, toujours un problème en Inde.	12
1.2.3.	Une base de données multidimensionnelle	13
2.	Une agrégation flexible	16
2.1.	Une base trop fine ou trop riche ?.....	16
2.1.1.	Problème du nombre d'unités	16
2.1.2.	Problème de taille des unités.....	17
2.2.	Agréger	19
2.2.1.	Méthodes et résultats	19
2.2.2.	Effets de la clusterisation	22
3.	Structuration spatiale et autocorrélation spatiale.....	24
3.1.	Les principales dimensions : autocorrélation spatiale et distance.....	25
3.2.	Les principales dimensions : autocorrélation spatiale et niveau d'agrégation.....	29
3.3.	Les principales dimensions : autocorrélation spatiale locale (LISA).....	33
3.3.1.	Saisir les structures localement	34
3.3.2.	Introduction aux LISA : l'indice de Moran local et son utilisation.....	35
3.4.	Interpréter les LISA	37
3.4.1.	Significativité de l'indice de Moran local	37
3.4.2.	Interprétation	37
3.4.3.	Une réflexion à propos des échelles	40
4.	Une thématique spatiale : la vulnérabilité.....	42
4.1.	Le concept de la vulnérabilité et sa dimension géographique.....	42
4.2.	Espace et vulnérabilité : les trappes de pauvreté.....	43
4.3.	Une première représentation de la géographie de la pauvreté	44
4.4.	Les dimensions de la pauvreté : définition, estimation et opérationnalisation.....	45
4.4.1.	Pluviométrie et irrigation	46
4.4.2.	Développement social : démographie, éducation.....	55
4.4.3.	Urbanisation et accessibilité.....	56
4.4.4.	Infrastructures	56
4.4.5.	Genre et groupes sociaux	59
4.5.	Synthèse : carte de la vulnérabilité	61
4.5.1.	Opérationnalisation des mesures	61
4.5.2.	Synthétiser les dimensions de la vulnérabilité	62
4.5.3.	Résultat de l'estimation des niveaux de vulnérabilité.....	63
4.6.	Epilogue	67
	Références bibliographiques	69
	Annexe I : Le nuage de points de Moran	73
	Annexe II : Le cokrigeage.....	75

Liste des figures

Figure 1 : les étapes de la diffusion spatiale dans le temps.....	10
Figure 2 : Répartition des villages par taille dans les quatre Etats indiens.....	18
Figure 3 : Villages d'origine, points d'agrégation à 10 km et polygone de Thiessen correspondant.....	21
Figure 4 : Villages d'origine, points d'agrégation à 5 km et polygone de Thiessen correspondant.....	21
Figure 5 : Indice de Moran calculé pour la densité sur des pas de 20.....	26
Figure 6 : Indices de Moran calculés sur des pas de 20 pour les clusters de 10 km.....	27
Figure 7 : Indices de Moran calculés sur des pas de 20 pour les clusters de 10 km (sélection).....	28
Figure 8 : Paires d'observations par pas de 50 km pour différentes familles d'agrégation et les villages d'origine (échantillonnage indiqué entre parenthèses).....	30
Figure 9 : Indice de Moran (REF) calculé sur des pas de 50 km pour différentes familles d'agrégation et les villages d'origine (échantillonnage indiqué entre parenthèses).....	31
Figure 10 : Indice de Moran (REF) pour les villages et les clusters de 5 km.....	32
Figure 11 : 3 en 1, la carte des clusters de LISA. Un exemple avec le sex ratio des enfants.....	38
Figure 12 : autocorrélation spatiale locale de l'alphabétisation.....	40
Figure 13 : autocorrélation spatiale de l'alphabétisation selon trois définitions du voisinage.....	41
Figure 14 : mesure factorielle du développement rural.....	45
Figure 15 : pluviométrie dans le sud de l'Inde.....	48
Figure 16 : carte de l'irrigation : proportion de terres cultivées, irriguées.....	49
Figure 17 : Mise en évidence des zones fortement cultivées, à agriculture pluviale.....	51
Figure 18 : Irrigation et barrages.....	51
Figure 19 : Répartition spatiale de l'irrigation par canal.....	51
Figure 20 : Répartition spatiale de l'irrigation par tank.....	51
Figure 21 : répartition spatiale de l'irrigation par puits et forage.....	52
Figure 22 : Répartition géographique des types d'irrigation majoritaires.....	52
Figure 23 : Type d'irrigation dans les zones à forte irrigation (>50%).....	53
Figure 24 : Irrigation par canal et zones densément irriguées.....	53
Figure 25 : Types d'irrigation dans les zones faiblement irriguées (irrigation <20%).....	54
Figure 26 : enclavement des villages.....	56
Figure 27 : distance aux centres de soins de qualité.....	58
Figure 28 : distance aux centres de soins.....	58
Figure 29 distance à une école primaire.....	59
Figure 30 : distance à l'éducation.....	59
Figure 31 : répartition par niveau de vulnérabilité (clusters 2 km).....	64
Figure 32 : Carte des niveaux de vulnérabilité rurale.....	65
Figure 33 : Carte de vulnérabilité rurale au nord du Tamil Nadu.....	66
Figure 34 : nuage de point de Moran pour l'indice de modernisation (voisinage de 15 km).....	73
Figure 35 : Estimation locale de la mortalité infanto-juvénile selon quatre méthodes : uniformité au sein des districts, krigeage de la valeur district, imputation locale, cokrigeage.....	76
Figure 36 : Estimation locale de la mortalité infanto-juvénile : krigeage et cokrigeage (zoom).....	78

Liste des tableaux

Tableau 1 : Couches de données disponibles.....	13
Tableau 2 : Distribution des rapports enfants/femme selon la taille des villages.....	18
Tableau 3 : Caractéristiques des villages et clusters.....	22
Tableau 4 : Taxonomie des indices d'autocorrélation spatiale (d'après Getis, Ord, 1992 : 263).....	34
Tableau 5 : Répartition de l'irrigation selon les Etats.....	50
Tableau 6 : Dimensions et critères de vulnérabilité.....	61

0. Introduction : de l'autocorrélation à la pauvreté

Les avancées économiques et sociales recomposent le paysage humain des pays asiatiques et l'Inde fournit un exemple remarquable de la rapidité et de l'ampleur des transformations en cours. Mais que l'on adopte le point de vue d'un sociologue, d'un économiste ou d'un démographe, la dimension géographique de ces dynamiques contemporaines reste essentielle pour comprendre leur nature, car le changement social et économique repose en grande partie sur des réseaux d'échange et de communication qui sont spatialement autant que socialement structurés. Les traces spatiales des mutations de la société sont particulièrement visibles en Inde du Sud, région du pays qui a enregistré en plusieurs domaines les changements les plus rapides: baisse de la fécondité, augmentation de l'agriculture irriguée, mais aussi accroissement de la discrimination sexuelle.

Le projet EMIS (*Espace et Mesure en Inde du Sud*) rassemble sept chercheurs de diverses disciplines réunis pour réaliser une analyse comparative de la structuration spatiale de l'Inde du Sud à partir d'un corpus d'informations unique, issu du recensement national, disponible sous forme de SIG. Ce projet explore à la fois des questions théoriques, sur la nature des structurations spatiales mesurées, mais également thématiques, liant les facteurs étudiés à leur mode de propagation locale ou régionale. L'analyse des informations disponibles a conduit dans un premier lieu à des examens des profils cartographiques de variables essentielles, dont les principaux résultats ont été rassemblés dans un atlas (Oliveau 2003). Elle se prolonge par une analyse multiscalaire, rendue possible par une agrégation à géométrie variable, de l'autocorrélation spatiale. Cette analyse mesure la forme des compacités observées, en faisant varier les indicateurs, les échelles, les unités spatiales et les régions de référence.

Le projet procède de trois développements dans le champ scientifique:

A. *La disponibilité de données fines, désagrégées et pouvant être géoréférencées*

L'analyse spatiale du changement social dans les pays en développement est un domaine fort peu exploré, en particulier parce qu'elle requiert des matériaux, notamment géostatistiques, qui font défaut dans la plupart de ces pays. Notre projet prend appui précisément sur l'existence d'une *base de données spatialisée unique* pour l'Inde décrivant les principales caractéristiques du développement social et économique: structure de l'emploi, discrimination de genre, natalité, développement de l'irrigation, développement des équipements ruraux.

B. *L'intérêt pour la dimension spatiale du changement social à différentes échelles*

La dimension spatiale est un thème crucial pour la compréhension générale des dynamiques du changement social et économique, dont le caractère déséquilibré à macro-échelle, et segmenté à micro échelle, est manifeste. La rapidité des transformations en Inde dans les domaines concernés (baisse de la fécondité, développement de l'irrigation, progrès de l'éducation, renforcement des infrastructures collectives, etc.) est remarquable, notamment en regard de l'inertie relative qui caractérise la géographie des pays riches ou de certains pays moins avancés. Cette évolution accélérée s'accompagne d'une *trace spatiale*, évidente à tous les niveaux de l'analyse, en dépit des efforts gouvernementaux pour redistribuer les fruits de la croissance et corriger les déséquilibres socio-spatiaux.

C. *Le développement d'outils de traitement géostatistique applicables à des corpus de données sociales*

L'analyse de la géographie du changement social relève habituellement de l'interprétation cartographique de la distribution des phénomènes à macro-échelle. Ces analyses n'intègrent pas les informations à un niveau plus fin (micro-régional ou local), faute de données ou de moyens de traitement. De plus, elles reposent sur des interprétations cartographiques globales qui privilégient l'intuition (visuelle notamment) sur l'analyse systématique du corpus des données.

Quelques mots enfin pour décrire le présent rapport, car la progression du travail de l'équipe a connu un développement inattendu, autour d'une problématique qui n'a émergé que lors de notre seconde phase de recherche. Le projet EMIS se fondait en effet à l'origine sur un protocole d'analyse systématique d'un ensemble de données pour la mesure et l'analyse approfondie de l'autocorrélation spatiale. On considérait ainsi que la distance offrait le meilleur indicateur indirect (*proxy*) de la densité des interactions entre groupes sociaux et localités, et permettait d'évaluer la nature et le rôle spécifique de la structuration spatiale en Inde du Sud dans les mécanismes de développement sociaux et économiques. Ces aspects de l'analyse spatiale classique sont présentés dans les premières parties de ce rapport, notamment la troisième partie qui est une étude de la structuration spatiale.

Au cours du projet, qui s'est organisé autour de réunions d'équipes régulières (sessions souvent normandes de brain storming en fin de semaine), les participants ont toutefois souhaité mettre en avant des thématiques plus concrètes et conférer ainsi plus de contenu à une démarche formelle jugée parfois bien aride. C'est ainsi que l'intérêt porté aux questions de spatialisation de la pauvreté a fait émerger des complémentarités thématiques originales entre géographes et autres praticiens des sciences sociales, autour d'une question spécifique (*poverty mapping*) qui intéresse avant tout les économistes du développement. La dernière partie du rapport est par conséquent consacrée à l'approche géographique de la vulnérabilité en Inde du sud et représente un segment distinct de notre trajectoire de recherche. Elle passe en revue certaines dimensions majeures de la vulnérabilité des populations, avec un accent particulier sur le monde rural et la géographie de l'irrigation qui ont fait l'objet d'un travail approfondi. Cette recherche offre, grâce aux instruments de modélisation spatiale que l'équipe s'est appropriés, une synthèse efficace sur les contours de la vulnérabilité en Inde du sud et représente d'ailleurs, par son échelle locale d'analyse, un apport tout à fait original dans le champ de la recherche sur l'Inde.

1. Etudier la spatialité socio-économique en Inde du sud

Que ce soit en Afrique, en Asie ou en Amérique du Sud, des croissances (économiques et démographiques) très différenciées ont lieu entraînant des divergences importantes au sein du monde en développement. Ces grands ensembles sont loin d'être homogènes, et les différents pays qui les composent connaissent aussi des fragmentations internes (entre régions, au sein des régions). Les inégalités sociales et économiques internes se structurent autour de différents groupes sociaux (ethnies, tribus, castes, etc.) mais ont aussi une géographie particulière. On opposera volontiers les « régions qui gagnent » (Benko *et al.*, 2000 faisaient référence aux districts industriels Marshalliens) aux « trappes spatiales » de pauvreté. Les relations entre villes et campagnes restent quant à elle un terme majeur de compréhension des inégalités spatiales (Chaléard *et al.*, 1999).

Au sein des sciences sociales, l'approche géographique se caractérise par un accent mis sur ce dernier point. Les dimensions spatiales et les jeux d'échelles des phénomènes étudiés sont la porte d'entrée privilégiée (mais non unique) de la géographie du développement. On cherche alors à décrire les formes observées, pour comprendre l'émergence, la permanence, voire la disparition des phénomènes examinés.

A l'échelle de l'Inde du Sud, le travail entrepris vise à comprendre la distribution et la (trans)formation des inégalités de développement. On a adopté pour cela une approche « spatialisante » à une échelle locale. Nous avons ainsi fait le pari d'aborder les inégalités de développement à l'échelle des villages du Sud de l'Inde, en considérant leurs structures spatiales avant même d'étudier leurs dimensions socio-économiques, nécessairement imbriquées, mais laissées comme hypothèse secondaire¹. Cette posture scientifique est celle de l'analyse spatiale (Haining, 2003, Goodchild, Janelle, 2004) que l'on peut résumer par la formule « space matters ». L'hypothèse centrale soutenant ces travaux est celle du rôle des interactions spatiales comme élément explicatif des inégalités, et celui de l'espace comme « frein » à ces interactions (Pini, 1992, Pumain, St Julien, 2001).

Les approches suivies sont multiples, et les outils à notre disposition nombreux. La démarche la plus classique, préalable au travail sur la dimension spatiale, est celle de la cartographie. Mais, là encore, les possibilités (échelles, données, représentations, etc.) et les contraintes (nombre d'unités, lisibilité, robustesse des résultats...) sont nombreuses, et nous ont amenés à faire des choix pour présenter, à travers un atlas, une première photographie de l'Inde du Sud. La seconde démarche, plus originale, vise à étudier les structures spatiales des phénomènes. Nous avons donc développé un travail sur l'autocorrélation spatiale, la variographie et surtout les analyses locales d'autocorrélation spatiale des phénomènes observés.

1.1. Pourquoi l'espace ? Diffusion et répartition

L'intégration de l'espace dans les sciences sociales est ancienne. Elle a cependant toujours des difficultés à être acceptée, notamment parce que l'intégration de l'espace pose finalement plus de problèmes qu'elle n'en résout (voir Derycke, 1994 à propos de l'intégration de l'espace dans la modélisation économique). Néanmoins, la dimension spatiale des inégalités de développement a déjà été pensée. Ainsi, la notion de « centre-périphérie » vient de

¹ Par exemple, le niveau de d'alphabétisation est fonction de la distance à une structure d'enseignement de qualité (effet spatial) plus que fonction de la présence d'une école primaire (indicateur socio-économique de développement).

l'économie (voir à ce propos Krugman, 2000), et a ensuite été réinterprétée dans l'espace géographique. Les notions de « centre » et de « centralité » sont étudiées aussi bien par les géographes que par les économistes. Les échanges conceptuels entre géographie économique et économie spatiale sont fréquents. Ainsi la diffusion spatiale des innovations, concept géographique ancien, a été reprise pour réinterpréter en termes spatiaux la théorie économique des pôles de croissance. En sociologie, les modalités d'interrogation du concept d'urbanisation sont différentes de celles de la géographie, les géographes s'interrogeant plus sur la dimension spatiale du passage de l'urbain au rural (en termes de conditions et de conséquences notamment).

1.1.1. A la source des déséquilibres spatiaux : la centralité

En 1973, dans son ouvrage « Le développement inégal », Samir Amin utilise le binôme centre-périphérie pour décrire, dans une vision engagée des inégalités mondiales du développement économique, l'opposition que l'on nomme aujourd'hui Nord-Sud. Cette association de termes est alors pensée dans un cadre économique, proche du marxisme, où le centre est exploitant et la périphérie exploitée. La dimension spatiale est secondaire et plutôt utilisée de façon métaphorique. Mais cette approche ouvre la voie à une réflexion plus globale sur les rapports entre sociétés (et espaces) dominantes et dominées, mettant en place un transfert du social au spatial (Di Méo, 1991).

Ce schéma s'applique d'ailleurs à plusieurs échelles, et c'est une de ses richesses. Ainsi, s'il exprime l'antagonisme entre pays développés et pays en développement, il est aussi adapté pour comprendre l'opposition (considérée encore en termes de domination) entre ville (centre) et campagne (périphérie). Cette opposition dialectique a donc été reprise en géographie, car sa puissance évocatrice est forte. De plus, sa capacité à décrire de façon binaire les phénomènes observés, quelle que soit l'échelle, est fort intéressante. Néanmoins cette approche est aujourd'hui dépassée. On sait que le centre peut avoir une partie de sa population dominée (et même une part de son espace), et que la périphérie n'est pas constituée uniquement d'exclus².

Dans un contexte plus large, le phénomène centre-périphérie peut être vu comme un cas particulier et manichéen du concept de centralité. La centralité est une perception subjective de l'espace. On définit à travers elle un point comme étant particulier et se différenciant des points qui l'entourent (Huriot, Perreur, 1994). Ce point est alors appelé centre. Il sert de référent pour structurer l'espace. N'oublions pas cependant que la centralité n'est ni absolue, ni totalisante. Elle est au contraire relative et contextualisée : par le jeu de l'échelle, elle est relative à un espace ; elle est contextualisée car le point central est défini en fonction de l'étude. Dans ce cadre, on peut comprendre la théorie des lieux centraux, énoncée originellement par W. Christaller dans les années 1930, comme un ensemble hiérarchisé de centres, dont la centralité est relative au niveau d'observation (qui définit le niveau de centralité) et dépendante du contexte (c'est-à-dire, dans son étude, des fonctions urbaines).

Pour expliquer la création de centres, on peut reprendre la théorie économique des économies d'agglomération (Catin, 1994). En raccourci, elle considère l'agglomération comme un facteur d'économie, permettant un avantage relatif par rapport aux zones non agglomérées³. L'agglomération, en se différenciant des zones non agglomérées, se définit ainsi comme un

² Le chapitre d'Alain Reynaud dans l'*Encyclopédie de géographie* (Reynaud, 1992) est remarquable par sa précision et ses nuances dans la définition des deux concepts de centre et de périphérie.

³ D'où les quolibets, cités par Krugman (2000 : 50), que l'on a pu entendre à propos de cette vision des choses : « Alors, les économistes croient que les entreprises s'agglomèrent à cause des économies d'agglomérations ? ».

centre. De plus, ce centre se place alors comme un objet concurrent et compétitif des autres objets, ce qui lui permet d'attirer encore plus et de renforcer sa position. C'est ce qu'Huriot et Perreur nomment la « centralité-attractivité » (Huriot, Perreur, 1994: 50). Cette réflexion, issue de l'économie, s'adapte aussi à la compréhension d'autres aspects, qu'ils soient sociaux ou culturels.

Si le centre attire, il ne peut néanmoins tout accueillir, et il se crée alors une hiérarchie des objets accueillis au centre (Thisse, 2002), les autres objets s'organisant autour de lui, suivant une forme non nécessairement concentrique, car d'autres facteurs peuvent influencer. Le centre redistribue alors autour de lui une partie des objets (effet de back wash) et accroît son rôle dans l'organisation géographique de l'espace. Mais le centre, outre sa fonction attractive, peut aussi être un point de départ. C'est la « Centralité-diffusion » définie par Huriot et Perreur (Ibid.). En effet, le centre est aussi au cœur de la diffusion spatiale des innovations (on se réfère évidemment au travail d'Hägerstrand, 1967). Ainsi lorsqu'une innovation apparaît en un point donné, celui-ci se démarque des autres et constitue, selon la définition, un centre. Celui-ci va alors transmettre l'innovation (ou l'information) vers d'autres points (effet de *spill over*).

C'est cette perception du centre diffuseur qui est à la base de la théorie des pôles de croissance. F. Perroux, dans les années 1950, explique que « la croissance n'apparaît pas partout à la fois ; elle se manifeste en certains points, ou pôles de croissance, avec des intensités variables ; elle se répand par divers canaux avec des effets terminaux variables pour l'ensemble de l'économie » (cité par Manzagol, 1992 : 496). Une des idées maîtresses de la théorie des pôles de croissance est celle d'industrie industrialisante : une entreprise innove dans un point de l'espace, son développement entraîne celui des entreprises qui l'entourent (par augmentation de la consommation locale) et celles avec lesquelles elle travaille (par augmentation de la demande de production). Ce modèle théorique a été mis en application dans les années 1970 dans le cadre de plusieurs politiques de décentralisation, en Inde notamment. L'une des critiques de cette théorie se fonde sur l'angle de vue adopté. En effet, à l'inverse de l'idée d'une industrialisation provoquant l'urbanisation par un effet d'agglomération des individus autour d'elle, on peut envisager que ce sont les espaces peuplés qui fabriqueraient de l'industrie (Cassidy, 1997). Cela expliquerait d'ailleurs assez bien l'échec des implantations d'industries industrialisantes dans certaines zones (notamment en Afrique), le pôle de croissance n'ayant pas pu créer d'effet d'entraînement sur son hinterland, trop peu peuplé. Mais finalement, et à la suite de Berry (1973), nous considérons que la théorie des pôles de croissance n'est qu'un cas particulier de diffusion de l'innovation.

1.1.2. La diffusion comme élément de transformation de l'espace

« Etymologiquement, la diffusion correspond à tous les déplacements qui, quelle qu'en soit la force motrice, cherchent à se répandre de manière homogène dans un système, et tendent à faire passer celui-ci d'un état d'équilibre à un autre état d'équilibre » (Saint Julien, 1992 : 559).

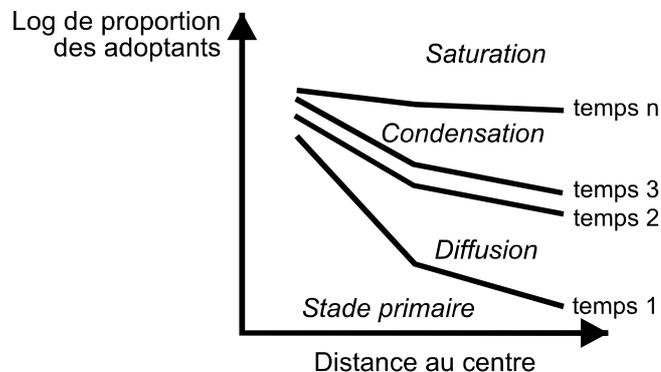
Un phénomène d'innovation se définit comme nouveau, irréversible, et modifiant la dynamique des systèmes dans lesquels il opère (donc sans rapport avec un phénomène de mode). Les sciences sociales interprètent ce type de processus comme endogène (dans une perspective plutôt weberienne) ou exogène (dans la tradition de Durkheim). On peut cependant adopter une position plus nuancée en affirmant qu'un phénomène de diffusion est à la fois le fruit d'un élément exogène, puisqu'il est nouveau, et endogène, puisqu'il est accepté par le système. On rappellera que la diffusion peut prendre plusieurs formes : sociale et

spatiale notamment. Les « déplacements » selon la définition de Thérèse Saint Julien peuvent être à la fois dans l'espace ou, de façon plus imagée, à l'intérieur d'une société. Cela correspond par exemple à la notion de « capillarité sociale » développée par Arsène Dumont (Tabutin, 2000). Les deux formes se complètent, même si le social est d'abord le terrain privilégié du sociologue et le spatial celui du géographe.

Pour bien comprendre comment fonctionne le phénomène de diffusion spatiale, il faut en distinguer les différents éléments et évaluer le rôle de chacun. Nous en distinguerons six, en nous appuyant sur les travaux de Thérèse Saint-Julien (1985) :

- 1) *Les facteurs exogènes* sont à la base de l'innovation, puisque ce sont eux qui l'apportent dans le système.
- 2) *La force de la combinaison innovante* est plus difficile à définir. Il s'agit en fait de l'innovation observée et des phénomènes associés. Par exemple, si l'on observe la baisse de la fécondité comme innovation, la combinaison innovante est l'alphabétisation, le statut de la femme, etc.
- 3) *L'aptitude au déplacement* est la capacité du phénomène à se propager dans l'espace. Elle dépend de deux facteurs : la distance, tout d'abord, qui peut être physique, sociale, économique, culturelle... et les canaux ensuite (qui sont souvent des réseaux : routiers, sociaux, etc.). Ces deux facteurs suffisent souvent à expliquer une bonne part des vitesses différentielles de propagation de l'innovation.
- 4) *La force de propagation* ne doit pas être confondue avec l'aptitude au déplacement. Elle marque surtout le rôle des foyers émetteurs. Plus ce foyer est puissant, plus la diffusion peut aller loin, et vite.
- 5) *Le temps de propagation* est l'élément le plus important pour comprendre la vitesse de diffusion du phénomène. Sa mesure est simple : il s'agit de compter le temps écoulé entre l'amorce de la diffusion et la saturation du système (voir Figure 1). Mais il n'y a pas d'échelle unique, et la comparaison de plusieurs phénomènes reste de l'appréciation du chercheur.
- 6) *Le milieu d'accueil* n'est pas à négliger, puisqu'il marque le rôle des récepteurs, c'est-à-dire ici de la population adoptant ou non les nouvelles normes. Cet élément permet souvent de mettre en valeur la structure du système de peuplement, l'innovation se déplaçant généralement plus vite quand la population est plus dense.

La diffusion spatiale, comme tout processus géographique, reste très liée à la notion d'échelle et de niveau d'agrégation. Elle peut s'analyser à différentes échelles, faisant alors ressortir différents acteurs de la propagation de l'innovation.



**Figure 1 : les étapes de la diffusion spatiale dans le temps.
D'après T. Hägerstrand (1953)**

Une analyse régionale désagrégée, qui prend les communes rurales comme référents, semble la meilleure échelle pour mettre en valeur les objets suscitant la diffusion (les villes, par exemple) ainsi que les structures spatiales du milieu d'accueil (les réseaux). On comprend pourquoi cette échelle a longtemps été privilégiée dans la géographie française, et particulièrement chez les « tropicalistes ». Pierre Gourou, représentant historique de cette école, avait fait de l'échelle régionale son niveau d'analyse privilégié, et c'est sans aucun doute pour la raison évoquée, à savoir la bonne vision des microstructures et la mise en valeur du rôle des encadrements.

L'échelle macro et agrégée ne permet plus, quant à elle, de distinguer le rôle des individus et des contacts directs, mais les grands réseaux de communication et les formes des systèmes de peuplement (grandes aires culturelles par exemple) sont clairement identifiables.

Lorsque l'on envisage d'expliquer la géographie d'un phénomène par la diffusion, il convient alors de rappeler qu'outre la diffusion par contagion, c'est-à-dire de proche en proche, comme nous venons de la décrire, il peut exister un phénomène de diffusion hiérarchique. Cette dernière n'agit pas en fonction de la proximité spatiale comme nous l'avons vu, mais en fonction d'une organisation hiérarchique (système de villes par exemple).

Ces observations doivent maintenant laisser la place aux aspects techniques de notre travail sur un SIG indien permettant précisément la mise en œuvre de ces notions théoriques. Les sections qui suivent décrivent notre base de travail et les modifications qu'on lui apportées.

1.2. Description générale de la base

1.2.1. Spatialiser les données censitaires

Pour un travail exhaustif à l'échelle villageoise, le recensement constitue en Inde la seule source disponible. Son unité de base est la commune rurale (revenue village). Ces villages fiscaux peuvent comprendre plusieurs hameaux, qui seront alors traités comme une seule unité administrative. On notera tout de suite deux biais liés à ce découpage. D'abord, il est impossible de différencier les formes d'habitat groupé et dispersé. Ensuite la comparaison inter-Etats est impossible puisque les découpages sont dépendants de choix politiques locaux. Néanmoins, son utilisation demeure nécessaire, et finalement assez efficace pour ne pas être rejetée (Oliveau, 2004 et 2005). La trame offerte est en effet bien adaptée à l'exercice auquel nous souhaitons nous livrer. De plus, le recensement offre des cartes de localisation des villages, qui, si elles sont de mauvaise qualité, permettent tout de même le géoréférencement nécessaire à la mise en place d'une base de données et au calcul de distance. Par contre, les informations offertes par le recensement restent loin de la qualité et du détail des enquêtes du NFHS⁴ ou du NCAER⁵. Mais il s'agit bien sûr d'un exercice totalement différent. Certaines questions ne sont pas posées, comme celles concernant le revenu, ou les mobilités (même si un effort a été fait dans cette direction pour le recensement de 2001). Les définitions floues, en ce qui concerne l'emploi tertiaire par exemple, et le manque de fiabilité des données infrastructurelles, particulièrement celles concernant l'occupation du sol, en font une source à manier avec prudence. Il ne faut jamais perdre de vue la signification réelle des données et la limite de leur utilisation. Ainsi, les données ont souvent un sens différent de celui qu'on leur donnerait ailleurs. Par exemple, le travail des femmes n'est pas une marque d'émancipation, mais la marque de la pauvreté, car il concerne avant tout les milieux ruraux, et ces travaux

⁴ Nom de l'enquête DHS en Inde. <http://www.nfhsindia.org/anfhs.html>

⁵ Voir leur site : <http://www.ncaer.org/>

sont les plus durs et pas forcément les mieux payés. Le travail des femmes ne signifie pas forcément, pour elles, une plus grande autonomie.

L'exploitation des données du recensement dans une perspective d'analyse spatiale nécessite leur mise en espace (géoréférencement) afin de pouvoir utiliser les SIG (Système d'Information Géographique) et autres logiciels de calculs géostatistiques.

Dans ce cadre, nous avons utilisé la base de données développée au sein du programme « population et espace en Inde du Sud » à l'Institut Français de Pondichéry. Ce SIG a été construit à partir de l'ensemble des localités des quatre Etats de l'Inde du Sud. Nous ne reprendrons donc ici que les grandes lignes du travail, la base de données ayant fait l'objet d'une publication sur CD-Rom (Guilmoto, Oliveau, 2000) et d'une description partielle dans Guilmoto *et al.* (2002).

1.2.2. L'accès aux cartes, toujours un problème en Inde.

Si la population indienne est bien suivie, l'espace indien est encore mal connu, car les cartes constituent toujours un élément considéré comme hautement stratégique par le gouvernement. Ainsi, la diffusion des cartes topographiques mentionnant des sites considérés comme stratégiques (bases militaires et installations nucléaires, mais également les ouvrages d'art ainsi qu'une bande côtière de plus de 50 km à l'intérieur des terres) est interdite. Pour les Etats littoraux, comme les quatre Etats du Sud de l'Inde, une partie des cartes topographiques n'est donc pas officiellement disponible. Par ailleurs, la production de nouvelles cartes est soumise à de nombreux contrôles administratifs et stratégiques qui ralentissent leur diffusion.

Les cartes publiées sont, par conséquent, peu nombreuses et incomplètes. Il est impossible de situer tous les villages, ni même toutes les infrastructures (les barrages, par exemple, puisqu'ils sont classés comme lieux stratégiques). Il n'existe d'ailleurs 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 des fonds 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.

Pour construire notre SIG, nous avons donc dû avoir recours à deux sources : les cartes du recensement d'une part, qui présentent l'avantage de fournir une localisation exhaustive des unités de recensement, à savoir les villages fiscaux, les cartes topographiques britanniques d'autre part. Les cartes du Census sont publiées dans les *District Census Handbook*. Elles ont donc suivi le rythme (lent) de leur parution, les derniers volumes pour le Sud de l'Inde étant sortis en août 2000. Les cartes fournies sont assez grossières, même si elles localisent précisément les villages. Pour les géoréférencer, il a fallu les coupler avec les cartes topographiques anglaises. Celles-ci datent de la colonisation et sont accessibles plus aisément aux chercheurs⁶. Elles datent toutes d'avant la décolonisation, et pour beaucoup du début du 20^{ème} siècle. Malgré leur âge, elles permettent d'identifier des points remarquables autorisant le géoréférencement. Nous avons effectué des tests à cet effet en 1998, et ils s'étaient avérés concluants (Oliveau, 1998).

Le travail de saisie des données a pris plus de deux ans, puisqu'il a fallu saisir les 16 780 villages et les 262 villes tamoules –ainsi que le réseau routier et les limites administratives

⁶ L'Institut Français de Pondichéry possède quasiment toute la couverture du sud de l'Inde.

supérieures (taluks et districts). A l'échelle du programme « Population et espace », ce sont 75 523 villages, 843 unités urbaines et toutes les routes de l'Inde du Sud qui ont été saisis, couvrant 636 000 km² et 197 249 763 habitants.

Il convient, avant d'aller plus loin, de souligner deux faiblesses de cette base de données. D'abord, le recensement étant le seul organisme à fournir la cartographie exhaustive des villages, la source est unique. Les vérifications et corrections possibles restent donc limitées. Nous avons effectué une série de tests statistiques, 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 finalement bien meilleure que les publications du recensement.

Ensuite, les sources cartographiques étant limitées et parfois de qualité moyenne, il existe une erreur de précision dans les localisations. Néanmoins, cette erreur est, d'après nos vérifications sur le terrain effectuées avec un GPS, inférieure à 500 mètres, et parfois inférieure à 250 mètres. A cette qualité de précision, on doit ajouter l'exactitude du fond, elle aussi remarquable. On distingue la précision, c'est-à-dire l'erreur de mesure (qui s'exprime en longueur ou en pourcentage), de l'exactitude qui marque la ressemblance entre la carte et la réalité. Une carte peut être très précise, mais complètement fautive et vice-versa (Didier, 1990). Les cartes que l'on trouve en Inde sont souvent inexactes et imprécises !

On rappellera ainsi que les sources cartographiques informatisées sur la population indienne sont quasi-inexistantes. Une source comme le DCW (*Digital Chart of the World*⁷) 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 utilisées⁸.

1.2.3. Une base de données multidimensionnelle

Pour bien comprendre les informations disponibles, et donc le travail envisageable à cette échelle, il convient de présenter la base de données de façon plus détaillée : huit couches de données sont présentes dans le SIG. La première concerne bien sûr les 75 522 villages du Sud de l'Inde. La seconde regroupe les 843 villes, puis l'on a ensuite les échelons administratifs supérieurs : taluks (équivalent au NUTS 3 européens) et districts (équivalent au NUTS 2 européens). Au niveau des infrastructures, la base possède une couche concernant le réseau routier et une autre le réseau ferroviaire. L'hydrographie constitue une septième couche. Et enfin, la base de données comprend un MNT d'une résolution de 500 m. Dans le cadre du projet EMIS, des informations renseignant sur le climat (températures et précipitations) sont venues compléter ce SIG.

Les données disponibles au niveau villageois pour notre analyse sont celles fournies par le recensement indien à travers deux séries de statistiques, celles du *Village Directory* et celles dites du PCA (pour *Primary Census Abstract*). Ces deux séries de statistiques sont publiées dans un seul et même volume au niveau de chaque district, le *District Census Handbook* (DCH).

⁷ Le *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*).

⁸ propos de la qualité du DCW, voir Kraak, Ormeling (1996 : 207).

Tableau 1 : Couches de données disponibles

Couches	Format	Informations disponibles
Villages	Points	<i>Primary Census Abstract, Village Directory</i>
Villes	Polygones	<i>Primary Census Abstract</i>
Taluks	Polygones	<i>Primary Census Abstract</i>
Districts	Polygones	<i>Primary Census Abstract</i>
Réseau routier	Lignes	Hiérarchie des routes (locale, régionale, nationale)
Réseau ferroviaire	Lignes	-
Hydrographie	Lignes, Polygones	Pérennité
MNT	Grille	Altitude par zone de 500m x 500m
Climat	Points	Température et Précipitation par mois

La publication des DCH a débuté après le recensement de 1951. Elle se place dans la continuité de la tradition des publications britanniques, dans le sens où chaque volume commence par une présentation générale du district, rappelant sur une cinquantaine de pages les traits majeurs qui le caractérisent, à travers des données agrégées (statistiques sociales, culturelles, démographiques et économiques), une carte globale, une description des caractéristiques physiques majeures, des informations sur les principaux lieux touristiques ainsi que d'autres informations de type administratif (changement dans les limites administratives, services existant etc.). On retrouve donc, en condensé, les informations contenues dans les *district gazetteers* de jadis, recueil d'informations de toute nature publiées à l'époque coloniale.

Le *Village Directory* répertorie les infrastructures présentes dans les villages, à savoir :

- Le nombre et le type d'équipements scolaires et médicaux, ou la distance la plus courte à l'unité possédant cette infrastructure.
- Les équipements routiers (type de routes conduisant au village, disponibilité d'arrêt de bus, de station de taxi, etc.)
- La présence d'infrastructures de télécommunication (bureau de poste, téléphone, fax, etc.)
- La disponibilité ou non en eau potable, et le moyen de la distribuer (pompes, puits, etc.) ; La distance à la source d'eau potable la plus proche quand elle n'est pas disponible dans le village.
- La disponibilité en énergie électrique et sa destination (usage domestique, agricole ou industriel)
- L'occupation du sol (surfaces forestières, agricoles -irriguées ou non-, etc.) et, le cas échéant, le type d'irrigation.
- La distance à la ville la plus proche et le nom de celle-ci

Le *Primary Census Abstract* (PCA) contient quant à lui les informations par sexe relatives aux populations, que l'on peut classer en différentes catégories :

- Données démographiques : nombre de logements, nombre de logements occupés, population totale, population masculine et féminine
- Données socioculturelles : population *dalit*, population tribale, niveau d’alphabétisation général.
- Données économiques générales : nombre d’actifs, nombre d’actifs partiels, nombre de personnes sans emploi.
- Données sur l’emploi : le recensement distingue dix catégories d’actifs :
 - Les propriétaires cultivateurs
 - Les employés agricoles
 - Les autres actifs du secteur primaire (pêcheurs, chasseurs, forestiers, employés dans les plantations)
 - Les mineurs et employés dans les carrières
 - Les artisans et employés d’industrie à domicile (*household industry*)
 - Les employés d’entreprises industrielles (*non household industry*)
 - Les employés du bâtiment
 - Les employés dans le commerce
 - Les employés dans le stockage, la manutention et le transport
 - Les autres actifs du tertiaire (services, fonctionnaires, etc.).

2. Une agrégation flexible

Au-delà des problèmes techniques propres à sa mise en place, la création de la base a posé un réel défi dans sa complexité. Elle repose en effet sur une infrastructure administrative qui est loin d'être cohérente de région à région. Nous nous intéresserons ici avant tout aux unités rurales, car le traitement des unités urbaines nous a semblé moins important, notamment dans la perspective de notre étude de la « surface socio-économique » de l'Inde du sud.

2.1. Une base trop fine ou trop riche ?

Nous avons en effet rencontré très vite des difficultés dans la gestion exhaustive de la base de données spatialisée, tenant notamment à nos différents impératifs de :

- Lisibilité cartographique
- Calculabilité
- Hétérogénéité des unités
- Robustesse statistique des indicateurs

Nous traiterons de ces deux premières gammes de problèmes ensemble, alors que les deux dernières sont d'un tout autre ordre.

Notons au passage que les questions propres à la qualité intrinsèque des données recueillies par le recensement indien sont ici hors de notre propos, même s'il est évident que pour les observateurs de la société indienne, elles sont essentielles pour ajuster notre réflexion. Ainsi, la qualité des données ou de l'enregistrement censitaire ont conditionné les choix opérés dans l'identification des indicateurs. Mais ces questions relèvent du domaine « sémantique » que nous n'aborderons pas dans le cadre du projet EMIS.

2.1.1. Problème du nombre d'unités

Il n'est guère réaliste de vouloir cartographier simultanément 70.000 unités. Même des cartes très fines, reproduites sur des supports de qualité (page A4 couleur, diapositives), ne peuvent offrir à l'œil que quelques milliers de points ou de surfaces en contraste. Toute échelle plus fine est illusoire.

En outre, le problème de la lourdeur des calculs devient réel. Si les outils de SIG ou de traitement statistique n'ont guère de difficultés aujourd'hui à traiter 100.000 unités, il en va autrement des calculs géostatistiques. En effet, un grand nombre des protocoles géostatistiques examine le rapport des unités à leur voisinage, procédant notamment par traitement matriciel. Ainsi le krigeage mobilise pour chacune des unités un ensemble de voisins dans un rayon dont la taille doit être raisonnable pour asseoir la qualité des estimateurs. Or, un simple calcul montre qu'un village sud indien compte en moyenne 1020 villages dans un rayon de 50 kilomètres. Ce qui imposerait des krigeages très lourds, si l'on prend en compte un tel rayon, car ils seraient basés sur des matrices de taille 1000 x 1000.

On fera la même observation en matière de calcul d'autocorrélation spatiale : la matrice complète des paires d'enregistrements ($70.000^2 / 2$) est de taille considérable et décourage les traitements exhaustifs.

La première solution à cette difficulté consiste naturellement à se restreindre à des traitements partiels, sur des aires plus limitées contenant par exemple moins de 5000 unités. On se trouve

alors forcé à renoncer à un traitement exhaustif à visée comparative sur l'ensemble du territoire sud-indien. Mais comme on va le voir dans la section suivante, cette solution au-delà des limites qu'elle présente à notre entreprise n'est guère viable pour des raisons d'une autre nature.

2.1.2. Problème de taille des unités

Les unités administratives rassemblées dans notre SIG ont des origines très hétérogènes. S'il n'est pas le lieu ici d'amorcer une histoire administrative de l'Inde, notons toutefois que l'Inde du sud a une histoire politique extrêmement fragmentée. Avant l'indépendance, elle se trouvait divisée en une large partie sous administration britannique, mais également un grand nombre d'Etats princiers autonomes parmi lesquels les plus célèbres sont ceux d'Hyderabad (le Nizam), du Travancore, de Cochin et de Mysore. A l'intérieur de l'Inde britannique (en l'occurrence la Présidence de Madras), l'administration coloniale n'a pu que recomposer une architecture principalement fiscale aux échelles supérieures des districts et des « taluks », que l'on pourrait comparer respectivement à des régions et des départements français.

Cette structure en districts et taluks (alias tehsils, mandals etc.), régulièrement enrichie et complexifiée durant les dernières décennies, reste homogène avec des unités de taille et de peuplement comparables. Mais le semis des milliers d'unités villageoises n'a guère été modifié par le régime colonial et les autorités de l'Inde indépendante. Leur définition repose sur deux principes très distincts : l'histoire politique locale et l'écologie du peuplement. Cette dernière découle elle-même de la rencontre entre des principes sociologiques (distribution et regroupement des castes) et géographiques (organisation du milieu et répartition des terroirs) d'organisation de peuplement.

On prendra par exemple pour critère de comparaison interrégionale des villages la taille de la population totale. En nous limitant ici aux quatre Etats de l'Inde du sud, la figure qui suit montre la répartition des localités selon leur taille. On voit que s'opposent par exemple le Kérala d'une part, doté de gros villages de plus de 10.000 habitants en moyenne, et les autres Etats d'autre part, comme le Karnataka où abondent les hameaux comptant moins de 1000 habitants (ces derniers représentant plus des deux tiers des localités de l'Etat). Une analyse plus fine, à l'intérieur des Etats, montrerait des regroupements géographiques plus accentués encore, notamment des zones de micro-localités caractérisées aussi bien par un habitat dispersé et épars (zones tribales) que concentré, mais administrativement morcelé (région historique de Mysore). Il apparaît donc que de nombreuses régions sont découpées en micro-unités qu'il serait bien risqué de comparer aux villages plus compacts du sud. On ajoutera pour brouiller le tableau que la taille des villages ne correspond que très imparfaitement à la densité démographique moyenne ou même encore au caractère concentré des établissements humains.

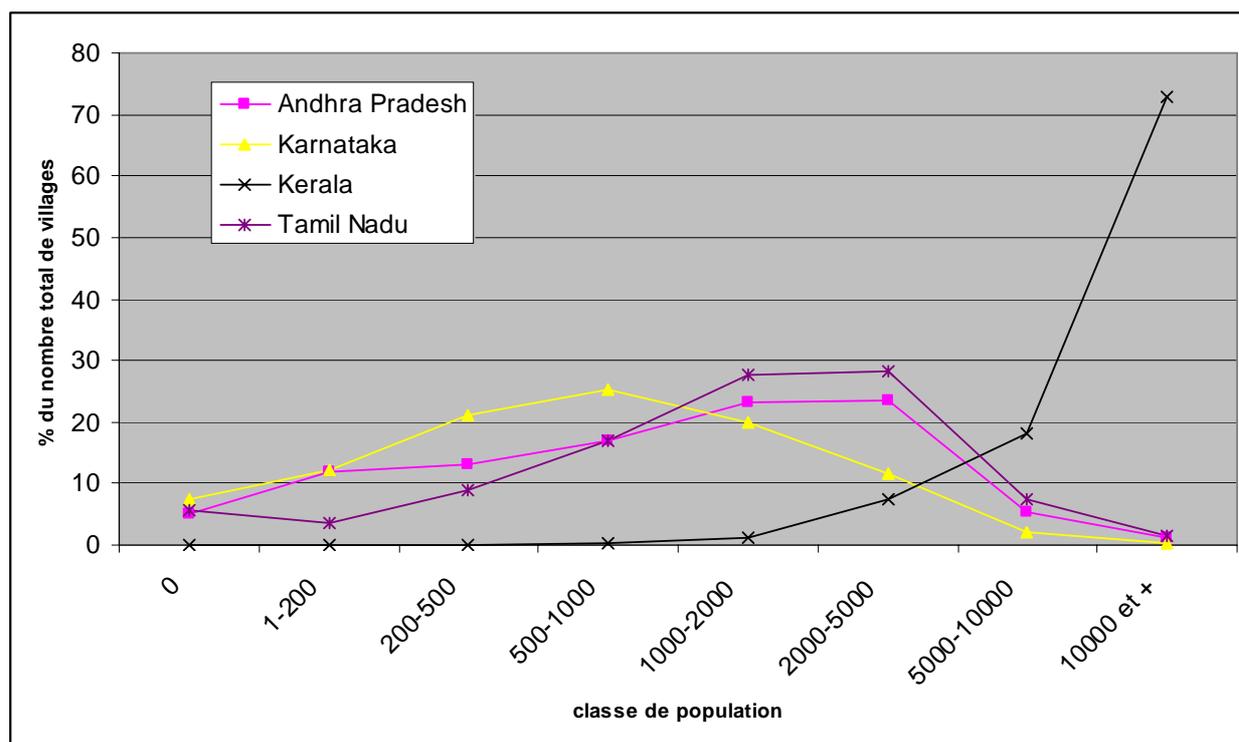


Figure 2 : Répartition des villages par taille dans les quatre Etats indiens

A ces difficultés s'est rajouté le problème proprement statistique lié aux faibles populations. De nombreux indices utilisés pour caractériser les localités s'avèrent instables quand ils sont calculés sur de petits effectifs. On en prendra pour exemple le calcul de l'écart type du rapport enfants/femmes⁹. On voit que cette variabilité de l'indice augmente quand la population des villages diminue, atteignant des valeurs atypiques pour les villages de moins de 500 habitants. Il n'y a sans doute pas de seuil statistique identifiable, d'autant que la moindre variabilité observée dans les villages les plus peuplés peut être due à des facteurs d'un autre ordre (rôle propre au Kérala). Toutefois, il peut sembler raisonnable de se méfier des villages comptant moins de 500 habitants en raison de la présence de valeurs extrêmes, sans lien avec la réalité du phénomène à étudier.

Tableau 2 : Distribution des rapports enfants/femme selon la taille des villages

Classe de taille	moyenne	Écart-type	Coefficient de variation
1-200 hab.	0.458	0.358	78.2%
200-500	0.420	0.137	32.5%
500-1000	0.417	0.116	27.7%
1000-2000	0.406	0.103	25.3%
2000-5000	0.389	0.093	23.8%
5000-10000	0.366	0.086	23.4%
10000 et +	0.328	0.081	24.8%
Total	0.408	0.158	38.6%

⁹ Le rapport enfants/femmes offre un indicateur indirect de la fécondité locale. Il est calculé comme le rapport des enfant de moins de 7 ans aux femmes de 7 ans et +.

2.2. Agréger

Ces dernières observations suggèrent qu'une étude réduite à des seules sous-régions, afin de pallier le nombre parfois prohibitif pour la cartographie et le traitement géostatistique, bute à son tour sur la forte hétérogénéité du peuplement et la présence d'unités villageoises trop peu peuplées. Le découpage administratif n'est toutefois pas de nature à offrir une solution « naturelle » d'agrégation, d'abord parce qu'il est très limité. On passe ainsi de 75000 villages, au niveau 0, à 1513 taluks ruraux au niveau 1, divisant ainsi notre partition par un facteur de 50. De plus, ce découpage est imparfait, en raison du cas particulier de l'Andhra Pradesh qui a, dans les années 1980, opéré un redécoupage de ces taluks originaux en 1100 « mandals » de taille beaucoup plus réduite que les 413 taluks du reste de l'Inde. De ce fait, alors que les taluks sud-indiens sont de taille relativement hétérogènes, la population moyenne des mandals de l'Andhra (44000 habitants) est environ cinq fois plus petite que celle des autres taluks (216000), ce qui rend les confrontations statistiques ou cartographiques malaisées.

On a par conséquent dû opter pour une reclassification automatique, par réagrégation des villages. L'objectif était, rappelons-le, triple : réduire la taille de notre échantillon de référence, réduire l'hétérogénéité dans la dimension des unités de mesure et réduire la part relative des micro-unités statistiquement instables. La décision est naturellement délicate, car le risque est réel en diminuant notre échantillon de réduire d'autant la variance de nos données, effaçant ainsi une part de la richesse de notre base en « moyennant » les valeurs. L'hétérogénéité à l'échelle villageoise est en effet la combinaison de deux composantes : le pur « bruit statistique » à éliminer et les réelles micro-variations locales dont la présence nous intéresse.

2.2.1. Méthodes et résultats

Différentes méthodes s'offraient à nous, privilégiant des critères particuliers pour le protocole d'agrégation des localités censitaires. En résumé, indiquons dès à présent que nous n'avons pas souhaité opérer un regroupement selon un critère proprement démographique, en réagrégeant par exemple les populations par différents blocs de 20.000 personnes. Une telle méthode (appelée parfois méthode des potentiels de population) devrait être appliquée à un paysage marqué par des fortes variations dans la densité du peuplement, variant de moins de 30 habitants au km² (par exemple dans les zones forestières des Ghâts) à plus de 1500 dans les régions rurales très peuplées du Kérala. Elle tendrait donc à dénaturer notre découpage régional, créant des unités de superficie extrêmement disparates. A titre d'exemple, on considérera un « gros village » typique du Kerala, recouvrant 31 km², avec 35.000 habitants. Notons alors que dans les régions septentrionales de l'Andhra Pradesh, il faudrait pour atteindre une telle population assembler 175 localités différentes réparties sur près de 2000 km², soit un espace 70 fois plus vaste que celui du village kéralais. On imagine alors les disparités spatiales d'une partition fondée sur une équi-population.

Nous avons donc choisi de privilégier la dimension spatiale, mesurée par la seule distance euclidienne. Cette dernière reste le meilleur (ou le moins mauvais) « proxy » pour l'intensité des échanges sociaux, qui constitue précisément le moteur de la transmission des comportements économiques ou sociaux dans notre schéma interprétatif¹⁰. De ce fait, nous avons décidé de redécouper l'espace sud-indien en poches de taille identique, appelés ici

¹⁰ Il va sans dire que la distance euclidienne est en outre fortement conditionnée par la densité du réseau de communication et la nature du terrain (forêt, pente, etc.).

« clusters ». Nous nous sommes fixés différentes échelles de regroupement : 2, 5, 10 et 20 km. L'existence de niveaux variables de partition nous permet de choisir la grille appropriée pour chaque analyse, en fonction par exemple du nombre de clusters souhaité.

Décrivons d'abord la méthodologie suivie. En premier lieu, partant du semis villageois exhaustif (y compris les villages inhabités, nombreux dans les zones forestières), nous avons procédé à une première agrégation automatique, réduisant l'échantillon originel à un nouvel ensemble de points d'agrégation en fixant la distance d'agrégation : tous les points à moins de n km les uns des autres sont alors fusionnés. Le résultat est alors un semis beaucoup plus régulier, puisque les localités proches ont été agrégées en leur centre géométrique. A partir de ce nouveau semis, nous créons des polygones de Thiessen/Voronoi qui serviront de base à notre famille de clusters. Nous reprenons enfin la base d'origine et regroupons tous les villages situés à l'intérieur des polygones de Thiessen, additionnant ou moyennant les valeurs individuelles selon la nature de la centaine de variables disponibles. Cette même procédure est conduite quatre fois, afin de réaliser un regroupement selon les échelles spatiales choisies (2, 5, 10 et 20 km).

La carte reproduite ici (nord de l'Andhra Pradesh) montre la répartition des villages d'origine (en noir), les points d'agrégation sur des rayons de 10 km et les polygones de Thiessen correspondant. Ces polygones constituent les limites des clusters et on voit que leur structure, sans être parfaite (notamment sur les bords), offre un découpage beaucoup plus régulier que le semis de points d'origine. On notera également que les emplacements des points d'agrégation sur lesquels sont construits les polygones de Thiessen sont influencés par la distribution spatiale des villages, sans naturellement l'épouser entièrement. La carte qui suit est identique, mais représente les clusters à 5 km qui sont beaucoup plus petits et nombreux.

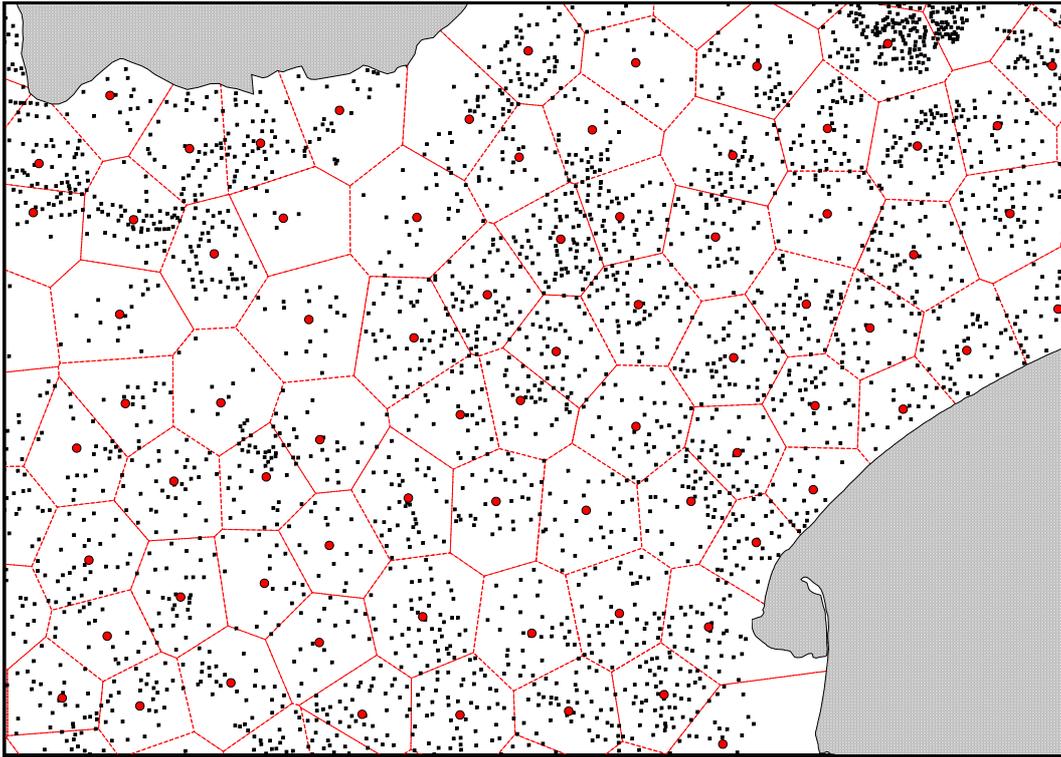


Figure 3 : Villages d'origine, points d'agrégation à 10 km et polygone de Thiessen correspondant.

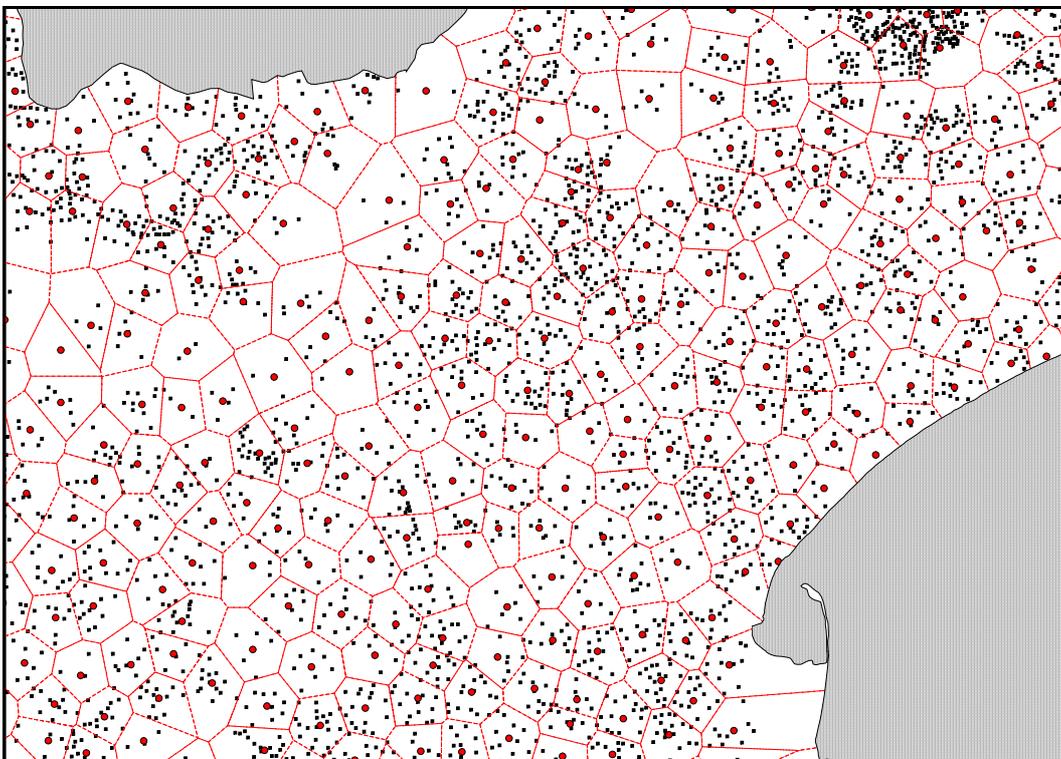


Figure 4 : Villages d'origine, points d'agrégation à 5 km et polygone de Thiessen correspondant.

2.2.2. Effets de la clusterisation

Quoique lourde en termes calculatoires (combinant analyse statistique et réagrégation statistique), cette procédure va donner une nouvelle partition de l'Inde du sud, très régulière en terme spatial et indifférente aux frontières administratives, comme les limites d'Etat, que nous avons choisi volontairement d'ignorer. Le tableau qui suit résume à la fois les résultats obtenus et les avantages de la méthode.

On voit en premier lieu que le nombre d'unités décroît régulièrement, selon l'échelle de regroupement de nos clusters, alors qu'augmente la taille moyenne des aires correspondantes. En second lieu, on observe la forte décroissance de la variance relative (coefficient de variation), à la fois des populations des unités et des aires. La décroissance est plus rapide pour ces dernières, puisque cela était le critère d'agrégation, mais elle n'est pas négligeable pour les populations en dépit des très fortes variations de densité que notre procédure n'entend aucunement gommer. On note en outre la diminution régulière des unités vides ou de population faibles (inférieures à 500 habitants) : elles représentaient 30% de l'échantillon d'origine et n'en font plus que 1% après regroupement à 5 km.

Tableau 3 : Caractéristiques des villages et clusters

		Villages habités	Types de cluster			
			2 km	5km	10 km	20 km
Nombre d'unités		70984	23848	6974	2151	620
Unités de moins 500 habitants		23312	1003	56	1	0
Superficie	Moyenne	784.8	2403	8309	26991	93643
	Ecart-type	1185.9	2046	4153	8854	26123
	Coefficient de variation (%)	151.1%	85.1%	49.9%	32.8%	27.8%
Population	Moyenne	1946.6	5794.3	19814	64241	222877
	Ecart-type	3035.0	6368	17961	53086	166218
	Coefficient de variation (%)	155.9%	109.9%	90.6%	82.6%	74.5%
Ecart-type du rapport enfants/femme		0.16	0.11	0.09	0.09	0.08

A titre d'exemple, nous avons également calculé l'écart type pour une variable donnée, en l'occurrence le rapport enfants/femme (utilisé pour l'estimation locale de la fécondité) ; la forte baisse de la variabilité de l'indice correspond à la diminution du nombre de valeurs extrêmes dans les villages censitaires, c'est-à-dire pour la plupart des valeurs aberrantes liées à des effectifs trop étroits pour un calcul stable de l'indice. La variance ne s'annule naturellement pas, car la fécondité reste très variable en Inde du sud, mais elle enregistre une baisse rapide dès le premier niveau de regroupement spatial.

Un examen plus précis montre que certaines difficultés sont bien résolues, telles que la forme des clusters dans les zones de limites. On sait qu'une méthode de regroupement classique par

carroyage en damier (partition perpendiculaire) provoque précisément de graves dommages sur les zones bordières, créant des zones de taille très inférieure aux cellules perpendiculaires de l'intérieur. Demeurent en revanche quelques « trous » dans l'agrégation d'origine : il s'agit des aires totalement dépourvues de localités, comme c'est le cas dans les zones montagneuses ou de forêt dense. De ce fait, certains clusters dans ces régions seront beaucoup plus grands, sans toutefois acquérir une population bien importante. Il n'y a ici guère de remède, sinon le cas échéant d'exclure *a posteriori* de l'analyse ou de la représentation les zones inhabitées ou sous-peuplées. Une autre méthode consisterait à découper l'espace en polygones de Thiessen, sans passer par une agrégation préalable. Cette méthode résulterait en un découpage encore plus régulier, ne tenant pas compte du regroupement éventuel des villages, mais elle ne réglerait pas la question des zones inhabitées.

3. Structuration spatiale et autocorrélation spatiale

Une des étapes centrales de notre projet a été de parvenir à une base de données homogène permettant de procéder à un examen systématique de la structuration spatiale. Elle se fonde précisément sur les résultats agrégés dont nous disposons, offrant à l'étude des échantillons à la fois plus robustes d'un point de vue statistique, de taille raisonnable en matière de calculabilité et enfin d'une répartition spatiale équilibrée et à ce titre représentative de l'espace sud-indien.

La présente section résume certains des résultats et des conclusions les plus intéressantes de cette étude. Nous souhaitons toutefois souligner en amont notre choix d'outils géostatistiques. Nous allons ici recourir avant tout aux indices de Moran. La statistique de Moran est la plus classique et très proche de l'indice de Geary, utilisé également pour mesurer l'association spatiale. On notera en outre que Moran a l'avantage de se décliner en indice local, dont on tirera parti plus loin dans l'analyse. Il demeure que nos calculs auraient pu être présentés à partir d'indicateurs différents sans que cela ne nuise à notre raisonnement.

L'indice de Moran mesure le degré d'autocorrélation spatiale sous forme d'*autocovariance relative* des observations pour une variable z . La covariance relative n'est autre que le produit moyen des valeurs observées (et centrées sur la moyenne) rapporté à la variance de l'échantillon. Quand la variable est normalisée (*moyenne = 0 et $\sigma^2=1$*), la covariance relative n'est que la moyenne des produits croisés des valeurs de la variable z . On parlera ici d'autocovariance pour tenir compte du fait que les produits sont calculés pour une même variable, mais pour différentes observations. On aura toutefois noté l'analogie avec le coefficient de corrélation, qui rapporte la covariance entre deux variables au produit de leur écart-type.

La définition classique de l'indice de Moran, appliqué à des observations i et j , dépend d'une matrice de proximité notée W_{ij} , prenant la valeur 0 partout sauf pour les i et j voisins.

$$I = \frac{\sum_{i,j} W_{ij} (z_i - \bar{z}) \cdot (z_j - \bar{z})}{n} / \sigma^2(z)$$

ou encore

$$I = \frac{\sum_{i,j} W_{ij} \cdot z_i \cdot z_j}{n} \quad \text{si } \sigma^2 = 1 \quad \text{et } \bar{z} = 0$$

Une extension de cette formule consiste à définir différentes classes de voisinages ; la matrice $W_{ij}(d_k)$ prendra désormais la valeur 1 quand i et j seront distants de d_k . On examinera alors de manière régulière l'évolution de $I(d_k)$, de d_1 à d_m (distance à laquelle l'autocorrélation devient négligeable), au moyen d'un corrélogramme.

$$I(d_k) = \frac{\sum_{i,j} W_{ij} \cdot z_i \cdot z_j}{n} \quad \text{si } \sigma^2 = 1 \quad \text{et } \bar{z} = 0$$

Cette écriture permet d'entrevoir des calculs détaillés de Moran par classe de distance croissante, plutôt que pour les seuls voisins (espaces directement contigus). Le comportement de la variable z est alors décrit sur des rayons croissants.

Une définition supplémentaire reprend la classe des voisinages simples, divisant simplement l'ensemble des paires d'observations en « voisins » et « non-voisins ». Toutefois, le calcul de Moran est déséquilibré, car certaines observations sont plus représentées que d'autres : certaines localités centrales ont en effet plus de voisins que d'autres, situées par exemple sur les limites du territoire ou dans des zones éparées. On choisit alors de corriger ce biais en calculant un indice Moran corrigé du nombre variable de localités prises en compte, imposant donc un poids identique à chaque observation.

La définition classique de l'indice de Moran considère en effet que plus le nombre de voisins est important, plus l'individu aura de poids dans la matrice de pondération. Le nombre de paires de voisins (m) est alors égal à $\sum_i \sum_j w_{ij}$. Cela ne se justifie que rarement. Au contraire, on préfère que chaque individu ait le même poids, c'est-à-dire que sa contribution à l'indice d'autocorrélation spatiale soit la même, qu'il ait un ou plusieurs voisins. Pour cela, il faut standardiser la matrice en ligne (*row standardization*). Cette opération consiste à pondérer le nombre de voisins j de chaque localité i pour que chaque ligne de la matrice (qui décrit les voisins de chaque individu i) soit égale à 1. En d'autres termes, si un point a 5 voisins, chaque voisin comptera pour $1/5^{\text{ème}}$ du total. Dans ce cas, le nombre de paires de voisins (m) est égal au nombre des localités i (n), comme si chaque individu n'avait qu'un voisin.

3.1. Les principales dimensions : autocorrélation spatiale et distance

Pour un premier examen, nous avons choisi d'utiliser l'indice de Moran par classe de distance. Les corrélogrammes obtenus sont alors assez simples à lire et décrivent en détail la structuration spatiale propre à la variable étudiée.

A titre d'exemple, le premier schéma fourni repose sur la variable de densité. Les unités d'observation sont ici les 2151 clusters ruraux de 10 km qui découpent l'Inde du Sud. Leur caractère statistique est relativement robuste en raison de leur taille moyenne, autant en termes de superficie que de population. Sur ce schéma, on lit la décroissance régulière de l'indice d'autocorrélation spatiale en fonction de la distance séparant les paires d'observations. Chaque point est la valeur de l'indice pour un pas de distance donnée.

Ces distances sont ici classées par pas de 20 km, distance suffisante pour le premier intervalle qui contient déjà 3000 paires. La distance exacte représentée est la moyenne des distances dans chaque intervalle, en général identique au milieu de l'intervalle. La distance maximale représentée est ici de 500 km, bien en retrait de la distance maximale entre les extrémités de l'Inde du Sud (distantes de 1500 km). Toutefois, il s'agit d'une limite suffisante, car l'autocorrélation spatiale mesurée par l'indice de Moran est systématiquement nulle ou négative au-delà de ce seuil comme nous le verrons plus bas.

Dans la lecture du corrélogramme, on distinguera notamment:

1. le profil (ou la courbure), en général monotone décroissante
2. la valeur maximale (à distance minimale), ici de 0,63, qui est en général le Moran du premier voisinage souvent utilisé pour résumer cet indice

3. La portée, ici de 430 km, qui est le niveau auquel s'annule l'indice. L'indice de Moran en l'absence d'autocorrélation spatiale est de $-1/(n-1)$, en l'occurrence -0.0004. Sur les gros effectifs, on considérera que l'indice est significatif au-delà de 0.1
4. La courbure des indices correspondant à la vitesse à laquelle décroît l'indice. Cette courbure peut être linéaire, concave (comme ici), voire convexe.

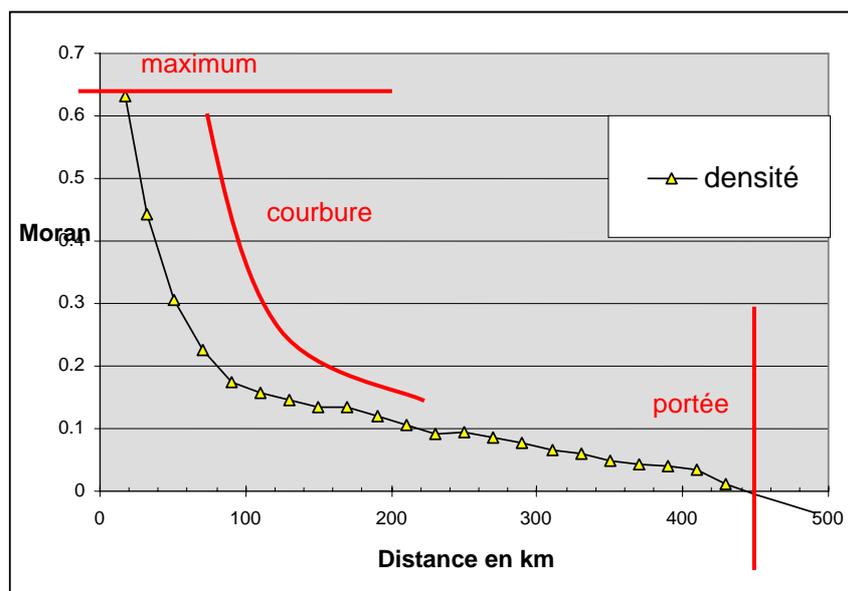


Figure 5 : Indice de Moran calculé pour la densité sur des pas de 20 pour les clusters de 10 km

Le même protocole de calcul a été appliqué à un jeu de variables, choisies pour leur relative robustesse et leur exemplarité analytique. Il s'agit en fait d'indicateurs importants de la configuration socio-économique de l'Inde du Sud rurale, mais nous ne les détaillerons pas ici.

La premier graphique superpose l'indice de Moran par distance pour douze variables différentes. On observera en premier lieu la présence de trois variables qui se distinguent à la fois par des maxima très élevés (supérieurs à 0.8), une courbure linéaire ou faiblement concave et une portée très élevée. Deux d'entre elles (taille du ménage et rapport enfants-femmes) sont en réalité étroitement corrélées, dépendant toutes deux de manière endogène du niveau moyen de fécondité dans les localités. Le niveau d'alphabétisation, lui-même associé à la fécondité de manière exogène, se caractérise notamment par une portée maximale (supérieure à notre seuil de 500 km) qui traduit l'étalement extrêmement régulier du niveau d'instruction sur l'ensemble de la région. Ces trois indicateurs présentent donc la plus forte structuration spatiale, atteignant à faible distance des niveaux proches du maximum de 1. Si la décroissance est très rapide avec la distance, l'autocorrélation spatiale reste accusée à plus de 250 km, même en cas de franchissement de frontières administratives. L'interprétation de cette structuration spatiale propre aux comportements reproductifs et à la progression de l'éducation en zones rurales renvoie directement à des mécanismes de propagation du changement social ; il s'agit en effet de variables dynamiques associés à des phénomènes sociaux qui ont énormément évolué durant les cinquante dernières années (baisse de la natalité, progrès de l'instruction, nucléarisation familiale, etc.). Les effets propres à la

diffusion spatiale de l'éducation ou du malthusianisme sont ici mis en évidence pour une analyse sociodémographique originale.

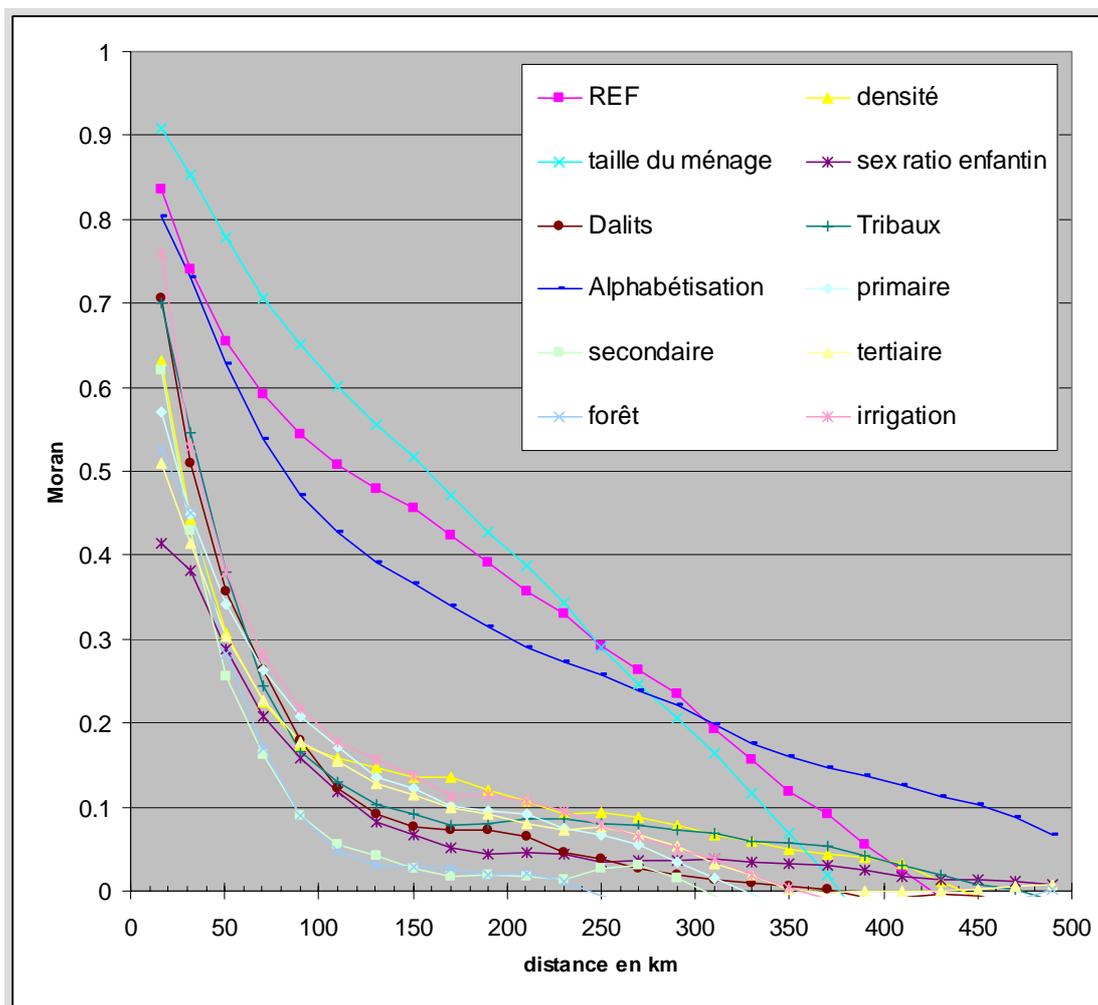


Figure 6 : Indices de Moran calculés sur des pas de 20 pour les clusters de 10 km

Les neuf autres variables présentent une structuration spatiale moins affirmée. On notera que si les maxima peuvent être parfois très élevés parmi ces indicateurs, la décroissance avec la distance est presque verticale durant les premiers cent kilomètres. Nous avons donc ici affaire à des phénomènes qui sont beaucoup plus localisés. L'autocorrélation spatiale devient assez mineure avant 200 km.

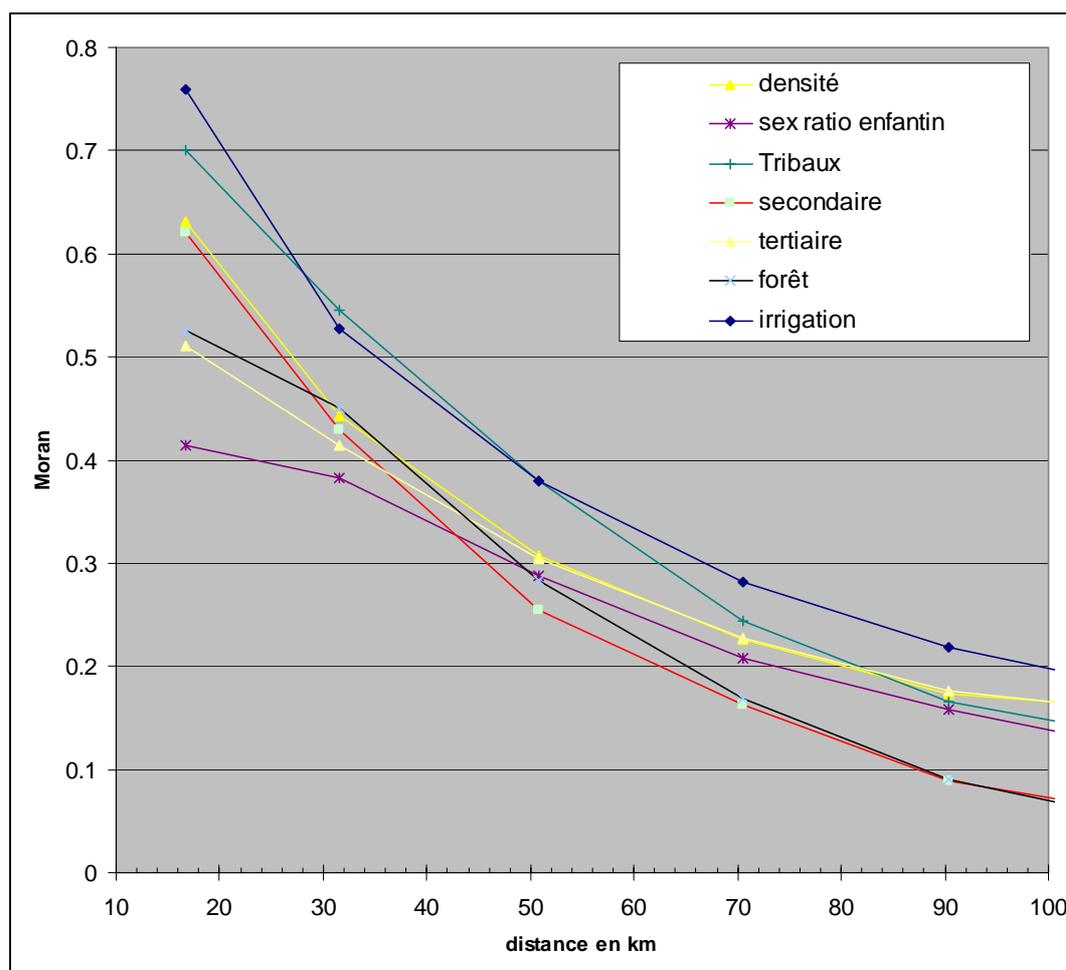


Figure 7 : Indices de Moran calculés sur des pas de 20 pour les clusters de 10 km (sélection)

La figure 7 suivante permet un examen plus fin de certaines d'entre elles. Nous y avons intentionnellement limité la distance à 100 km. On observe ainsi que certaines dimensions, comme la part de la main-d'œuvre dans le secondaire, se caractérise par une forte concentration spatiale à courte distance, mais qui disparaît très vite. Ceci décrit donc les formes d'implantation des activités industrielles ou semi-industrielles en zones rurales. L'étalement du tertiaire est, à titre comparatif, plus large. Parmi les variables du milieu, on notera la forte concentration de l'irrigation, mais une distribution moins concentrée du couvert forestier qui pourrait indiquer indirectement la fragilité des données. L'autocorrélation du couvert végétal devrait en effet être forte et il se pourrait que ce soit plutôt la nature de nos données, relevant de la classification par le Département forestier qui soit problématique.

Notons enfin la particularité du sex-ratio des enfants, avant tout déterminé par les discriminations de genre (foeticide féminin, infanticide ou surmortalité des petites filles). On sait par ailleurs la très forte concentration spatiale à l'échelle de l'Inde des poches de déséquilibres de genre, ainsi que l'existence de foyers de discrimination en Inde du sud, circonscrits en deux régions du Tamil Nadu. Le résultat obtenu est donc contre-intuitif, puisque la dépendance spatiale semble au contraire très faible. En réalité, cette mesure est éclairante puisqu'elle démontre qu'à l'échelle globale de l'Inde du sud, cette autocorrélation spatiale est minimale et que les variations observées dans le sex-ratio enfantin pourraient être en grande partie spatialement aléatoires. Ce qui suggère que la distribution même de ces

valeurs serait aléatoire, comme cela est précisément observé parmi les populations qui ne pratiquent pas de sélection sexuelle active (les variations étant alors liées principalement à des variations, modestes, du sex-ratio à la naissance).

Cette conclusion souligne en même temps les limites de l'indice de Moran global, qui sur l'ensemble régional ne fait état que d'un degré modeste de structuration spatiale alors que la concentration des déséquilibres en faveur des garçons est particulièrement marquée en certaines zones. On verra plus loin comment le « Moran local » permet de porter un regard plus précis sur les variations sous-régionales, indépendamment de la structure spatiale d'ensemble.

3.2. Les principales dimensions : autocorrélation spatiale et niveau d'agrégation

Notre agrégation en famille de clusters de taille croissante permet en outre un examen original. Une question récurrente de l'analyse géographique est en effet la prise en compte de l'effet d'échelle. Au-delà de ce principe général, une forte mise en garde découle des travaux de l'analyse spatiale vise notamment la dépendance scalaire : un même phénomène se lirait différemment à différentes échelles. Les géographes évoquent alors le MAUP (*modifiable areal unit problem*) : les variations d'échelle ou de découpage (zonage) peuvent influencer sur la mesure d'un phénomène. Nous pouvons ici examiner de plus près la dimension proprement scalaire du problème, puisque nous disposons de cinq échelles différentes (auxquelles on pourrait même adjoindre les échelles administratives). On notera en outre que nos échelles successives ont la propriété inhabituelle de ne pas être emboîtées géométriquement: les clusters à 2 km ne correspondent pas aux clusters à 5 km et ainsi de suite (voir les cartes précédentes) Ceci garantit à chaque échelle une complète indépendance de construction vis-à-vis des autres échelles.

La question que nous allons poser est de savoir si l'autocorrélation spatiale mesurée est fonction de l'échelle choisie. Plus précisément, on se demandera si et comment le fait d'avoir agrégé modifie le calcul de l'autocorrélation spatiale. Pour répondre à cette question, nous nous proposons de procéder à un calcul de l'indice de Moran sur une même variable pour différentes familles de clusters. Nous avons choisi ici de résumer nos résultats en utilisant une seule variable, à savoir le rapport enfants-femmes (REF), indicateur indirect de la fécondité déjà utilisé pour nos illustrations et dont la structuration spatiale est très accentuée (voir plus haut). En revanche, nous butons à nouveau sur des problèmes de calculabilité pour les échantillons de taille excessive: villages d'origine (70.000 unités habitées) et cluster à 2 km (24.000 unités).

Comme nous l'avons évoqué précédemment, ces volumes déterminent des matrices de proximité de taille considérable, respectivement de 2 milliards et 280 millions de paires d'observations. Pour résoudre cette difficulté, nous avons choisi de calculer les résultats sur des échantillons aléatoires : 10% des villages d'origine et 25% des clusters de 2 km. Cette méthode souligne naturellement des questions particulières sur l'échantillonnage spatial¹¹ que nous n'aborderons pas ici, faute de place. Les calculs intermédiaires ont toutefois montré que la variation des tirages aléatoires était sans effet sensible sur les résultats. La Figure 8

¹¹ Notons que nous avons procédé à un tirage statistiquement aléatoire et non systématique ou stratifié sur une base spatiale.

représente le nombre de paires d'observations traitées (en unités logarithmiques) et on en dénombre plusieurs dizaines par classe de distance dans la plupart des cas, ce nombre de paires d'observations dépassant le million pour les clusters de petite taille qui sont plus nombreux.

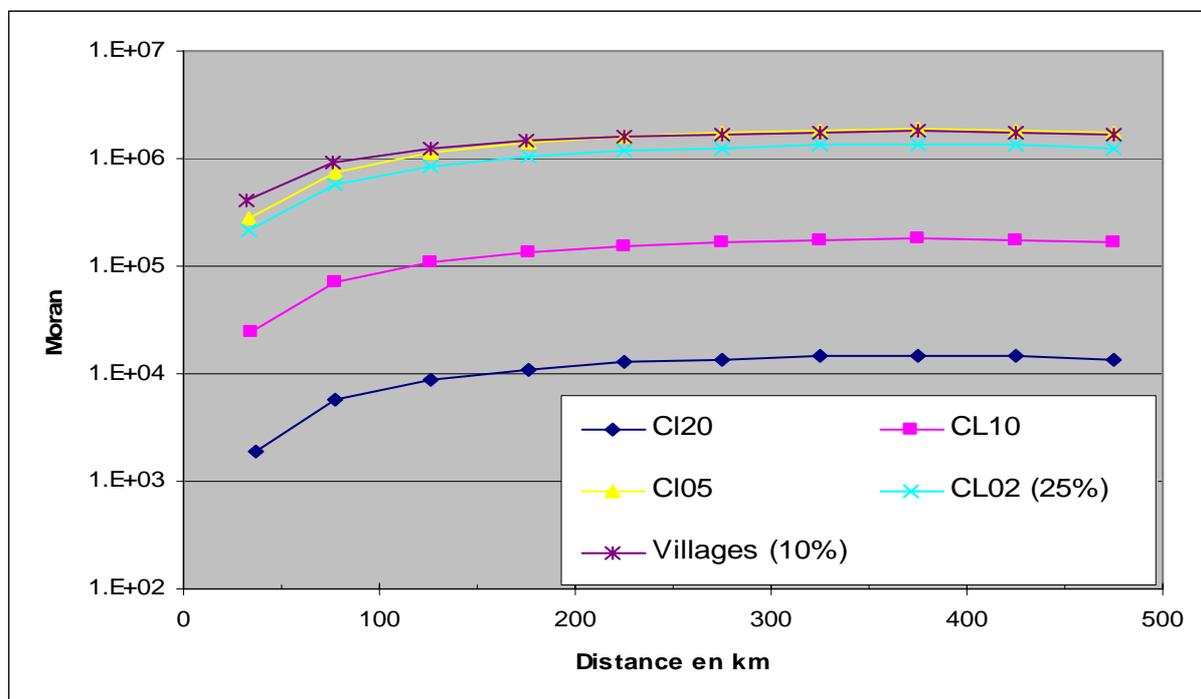


Figure 8 : Paires d'observations par pas de 50 km pour différentes familles d'agrégation et les villages d'origine (échantillonnage indiqué entre parenthèses)

Afin d'homogénéiser, nous avons conservé les mêmes pas de distance pour toutes les échelles de calcul, même s'il aurait été possible d'utiliser des pas de distance beaucoup plus court pour les petits clusters ou les villages d'origine. Les résultats sont présentés sur la Figure 8 qui reprend les indices de Moran pour les différentes familles d'agrégation. Ils indiquent des niveaux très élevés d'autocorrélation spatiale qui vont déclinant avec la distance croissante, de manière linéaire (ce qui suggère un étalement particulièrement régulier des valeurs). On distingue aussi clairement une hiérarchisation évidente de l'autocorrélation spatiale : les différentes courbes se superposent à partir des courtes distances de manière très régulière et sans chevauchement, l'agrégation progressive ayant pour effet d'augmenter graduellement le niveau d'autocorrélation spatiale. Les valeurs de l'indice de Moran sont particulièrement élevées pour les fortes agrégations, atteignant ainsi 0.8 pour les clusters à 20 km qui sont voisins immédiats (moins de 50 km de distance). On observe également que les courbes d'indices de Moran se rejoignent toutes pour les classes de distance 400-450 km à un niveau presque nul : la structuration d'intensité variable selon l'échelle manifeste néanmoins un profil comparable dans son étalement et sa portée.

Un examen plus fin de la courbure des mesures indique en outre un léger décrochement entre la première classe de distance et les suivantes que l'on observe en parallèle sur les différentes courbes. La nature de cette anomalie mériterait un examen spécifique, mais un premier survol de la mesure par classes de distance plus fines (Figure 9) suggère l'existence d'une première courbure concave à faible distance, suivie par une progression plus rectiligne pour des distances plus grandes.

Notons enfin le statut particulier des premiers niveaux d'agrégation, à savoir les données brutes et les clusters à 2 km. On note un décalage marqué avec les niveaux d'autocorrélation spatiale des autres échelles et ceci est notamment vrai des unités villageoises d'origine, dont le niveau d'autocorrélation spatiale est très médiocre. Il plafonne à 0.22 et les tests faits à partir d'autres échantillons à 10% fournissent des résultats de cet ordre.

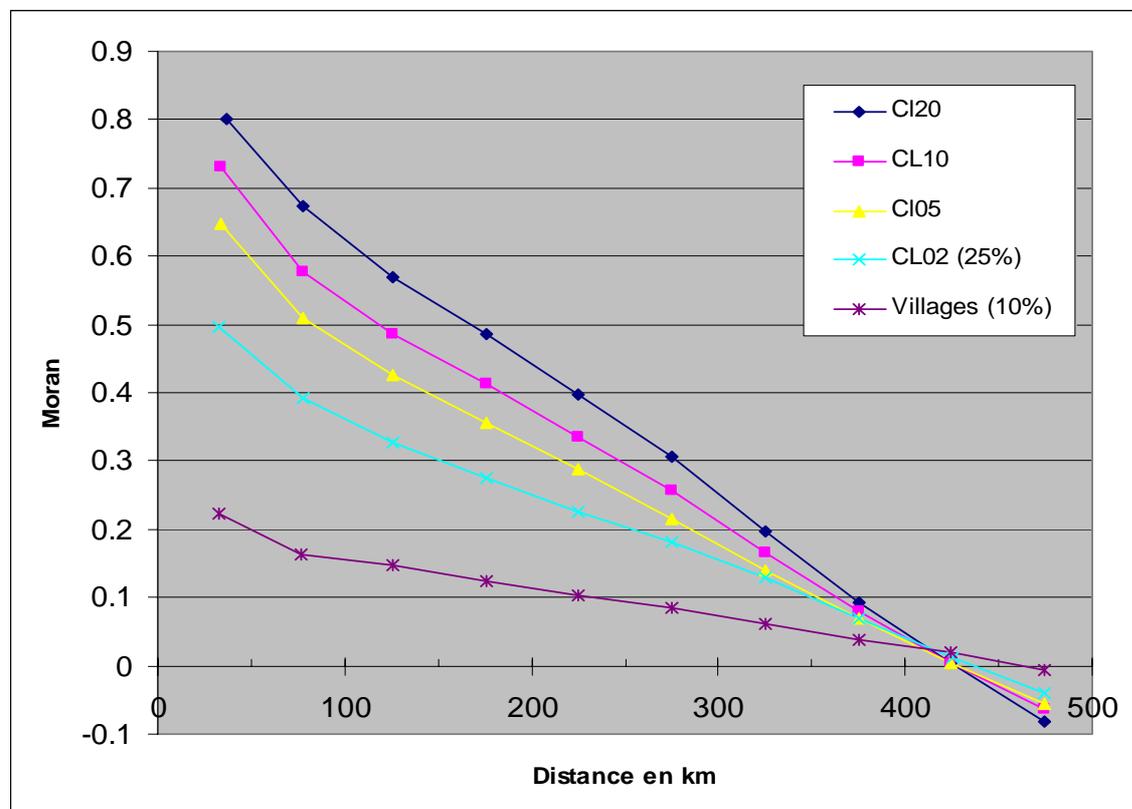


Figure 9 : Indice de Moran (REF) calculé sur des pas de 50 km pour différentes familles d'agrégation et les villages d'origine (échantillonnage indiqué entre parenthèses)

L'interprétation de ces résultats débutera par les données villageoises, dont on relèvera le faible degré d'autocorrélation spatiale. On peut de ce fait avancer comme première hypothèse que la variabilité statistique propre aux données villageoises a un effet considérable sur l'intensité de l'autocorrélation spatiale au point de la neutraliser en grande partie. Ce « bruit statistique » pourrait tenir en partie aux faibles effectifs de nombreuses localités, pour lesquelles le calcul du REF est instable. Mais il est également possible d'y lire la trace d'une micro-variabilité réelle de la fécondité mesurée localement par le REF : les différences entre villages voisins correspondent alors à des variations effectives de fécondité entre localités en dépit du gradient géographique.

Il n'est pas aisé d'établir formellement que les défaillances statistiques des données dans les petites localités jouent un rôle majeur dans la faible autocorrélation mesurée à l'échelle villageoise, mais le graphique qui suit fournit un élément partiel de preuve. On y a en effet superposé les résultats précédents pour les villages et les clusters, ainsi que pour deux autres échantillons de villages : les villages de moins de 1000 habitants et ceux de plus de 1000 habitants, découpage partageant l'échantillon en deux et dont le seuil s'inspire également des observations sur la variabilité des données selon la taille démographique des unités considérées (voir Guilmo et al. 2004). Le tirage aléatoire porte désormais sur 20% de

l'échantillon total. Cette comparaison est éclairante. On constate en effet que l'autocorrélation spatiale des gros villages est très largement supérieure à celle des petits villages, ces derniers se caractérisant par les plus faibles valeurs de Moran. La valeur maximale plafonne désormais à 0.17 pour les villages les moins peuplés contre 0.46 pour les autres villages. Concernant les villages les plus habités, on notera en outre que leur niveau d'autocorrélation spatiale est très comparable à celle des clusters à 2 km, voire plus élevés pour les distances supérieures à 100 km. On rappellera que les clusters à 2 km ont en moyenne une population de 5700 habitants, mais comptent toujours, en dépit de l'agrégation opérée, plus de 8.5% d'unités dont la population totale est strictement inférieure à 1000 habitants : ces clusters pourraient donc contribuer à faire baisser le niveau de l'indice de Moran. Les villages de plus de 1000 habitants ont pour leur part une population moyenne de 3700 habitants, mais par définition une population minimale de 1000 habitants au moins.

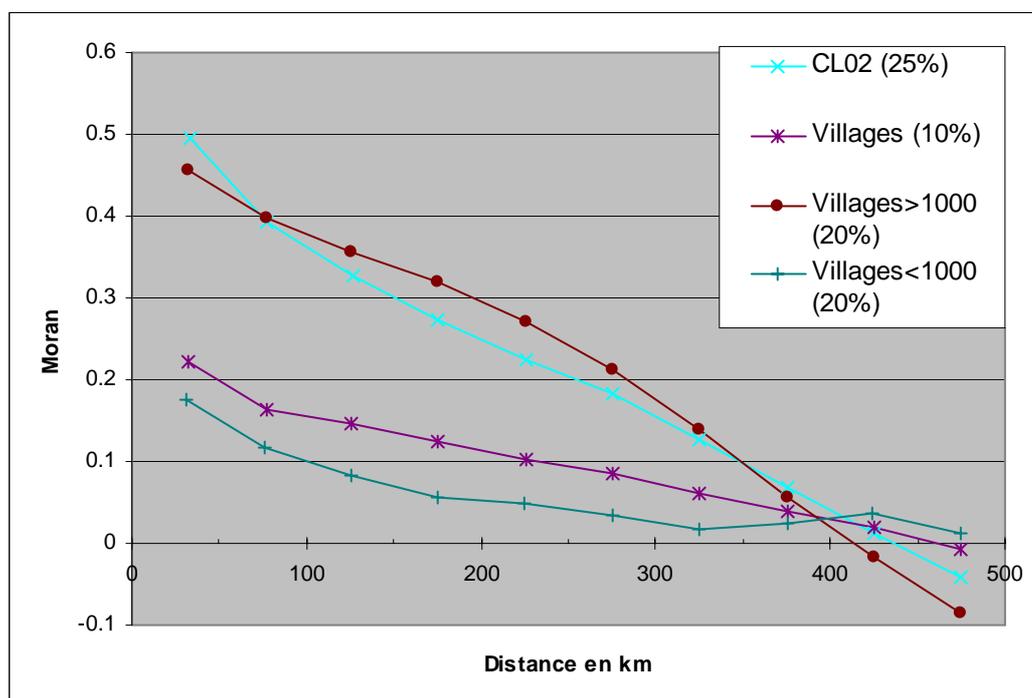


Figure 10 : Indice de Moran (REF) pour les villages et les clusters de 5 km

Si nous ne pouvons tirer de conclusion formelle de cette analyse¹², les enseignements préliminaires semblent indiquer que l'effet propre au bruit statistique dû à la faiblesse des effectifs est considérable sur le calcul de l'autocorrélation spatiale. En l'absence de villages « sous-peuplés », l'autocorrélation spatiale serait vraisemblablement analogue à celle obtenue sur une agrégation à 2 km. Une autre composante potentielle de la faible autocorrélation spatiale à l'échelle villageoise évoquée plus haut, à savoir la micro-variabilité du REF entre localités, serait alors modeste puisqu'on constate que le premier niveau d'agrégation n'améliore pas notablement l'autocorrélation spatiale dès lors que la taille des villages est suffisante. Il s'agit toutefois d'une hypothèse à explorer de manière plus systématique en raison de biais possibles d'échantillonnage dans les différents cas.

Les autres distributions des indices de Moran représentées sur la figure 9 correspondent à présent à des échantillons exhaustifs et à des unités statistiques de taille conséquente ; on ne

¹² Notamment parce que le tirage des villages de plus de 1000 habitants a créé un biais spatial en faveur des régions, comme le Kerala ou le Tamil Nadu, dont les localités sont les plus peuplées.

peut postuler un effet propre aux effectifs ou aux tirages. Il ressort de cette comparaison en premier lieu que le niveau de structuration spatiale est très élevé pour les trois échelles. On observe en outre que l'agrégation a pour effet de rehausser l'autocorrélation spatiale de manière sensible, puisqu'elle passe pour le premier rang de distance de 0.64 (cluster 5 km) à 0.73 (cluster 10 km), puis 0.80 (cluster 20 km). Ceci suggère des tendances micro-régionales de fécondité à distribution spatiale irrégulière que l'agrégation croissante gomme progressivement. Ces irrégularités spatiales détectées pourraient être de différente nature, résultant d'une hétérogénéité du peuplement (par exemple, de la population féminine alphabétisée ou active), responsable de variations de fécondité ou bien encore d'une dimension intrinsèquement hétérogène des comportements de fécondité régionaux. Il demeure que cette dimension est largement inférieure au niveau d'autocorrélation spatiale partagée par les trois familles d'agrégation supérieure, qu'on peut évaluer au niveau d'autocorrélation spatiale minimal mesuré pour les clusters à 5 km.

3.3. Les principales dimensions : autocorrélation spatiale locale (LISA)

La limite de l'exploration des données spatiales par des indices globaux d'autocorrélation spatiale est ici atteinte. L'apport du variogramme nous a permis d'aller plus loin, puisque nous avons pu mesurer le degré de structuration spatiale. Nous ne sommes néanmoins pas en mesure de connaître le détail de la structure spatiale de la variable étudiée (ce point est déjà souligné par Charre, 1995 :82). D'ailleurs, c'est de cette manière que peut être comprise l'objection à l'utilisation du concept d'autocorrélation spatiale formulée par Brunet & Dollfus (1990 : 87). En effet, le point souligné est surtout l'absence de variables totalement autocorrélées spatialement (mais à l'inverse aussi absolument pas autocorrélées). Sans le dire explicitement, les auteurs soulignent l'incapacité des indices de l'époque à aller plus dans le détail.

Sauf à avoir des gradients continus, phénomène assez rare en sciences sociales, comprendre la structure spatiale des phénomènes nécessite ainsi de pouvoir accéder à une information plus locale. En effet, les indices généraux peuvent être « aveugles » face à des phénomènes très fortement structurés, mais sur de petites zones au sein d'un espace absolument pas structuré. C'est par exemple le cas du rapport de féminité des enfants, dont on a montré ailleurs la forte structuration spatiale à l'échelle des deux districts de Salem et Dharmapuri (voir Figure 11), mais qui à l'échelle du Tamil Nadu n'offre aucune structure spécifique. Comparer la carte avec le variogramme (dont les valeurs sont nulles) est, de ce point de vue, instructif sur les limites de cette approche. Une solution consiste alors, comme nous l'avons fait précédemment, à mettre en perspective les informations issues des variogrammes avec la cartographie des variables. Mais l'approche redevient assez empirique, voire *ad hoc*.

Une démarche permettant de dépasser cet état de fait serait de pouvoir inspecter pour chaque point son niveau de ressemblance avec ses voisins. Ainsi, on pourrait mettre en valeur des zones où le niveau de ressemblance local est important, et utiliser cette information pour envisager une étude désagrégée de la structure spatiale des phénomènes. Il s'agirait en quelque sorte d'une mesure locale de l'autocorrélation spatiale, par opposition aux mesures globales.

Arrivés à ces mêmes conclusions concernant l'incapacité des indices globaux à prendre en compte les structures spatiales fines, plusieurs chercheurs anglo-saxons ont travaillé à l'élaboration de nouveaux indices, capables de rendre compte de l'autocorrélation spatiale à une échelle locale. Les résultats les plus notables sont les statistiques G_i et G_i^* (Getis, Ord,

1992), l'indice de Geary local et l'indice de Moran local (Anselin, 1995). Le Tableau 4 propose un récapitulatif des indices permettant de mesurer le niveau d'autocorrélation spatiale distingué en fonction de leur forme statistique (colonnes) et de leur échelle d'application (lignes).

Tableau 4 : Taxonomie des indices d'autocorrélation spatiale (d'après Getis, Ord, 1992 : 263)

Type	Produits croisés	Différences (au carré)
Global, mesure simple	Moran	Geary
Global, différents pas de distance	Corrélogramme	réagrégation
Local, différents pas de distance	G_i, G_i^* , Moran local	Geary local

Les indices locaux mettent en évidence des associations locales de valeur. Le terme d'association locale apporte une légère nuance à celui d'autocorrélation spatiale locale. Quand l'autocorrélation spatiale mesure la plus grande similarité statistique des valeurs de deux individus entre eux que par rapport à la moyenne de l'échantillon, le terme d'association soulignerait plutôt le regroupement spatial des individus de valeurs extrêmes. La nuance est faible, et les deux termes sont généralement employés comme synonymes.

Néanmoins, l'utilisation du terme association indique aussi la finalité de ces indices : mettre en valeur des sous espaces aux caractéristiques similaires, qui se distinguent de la moyenne de l'échantillon observé (on le verra lorsque sera abordée la forme mathématique de ces indices). Le terme décrivant ces zones est celui de « point chaud » (*hot spots*). Le point chaud, nous rappelle *Le Robert*, est un « lieu où il se passe quelque chose ». On retrouve l'idée de concentration locale par rapport à un ensemble général.

C'est donc tout logiquement que les projets qui ont porté ces recherches, depuis la fin des années 1990 jusqu'à aujourd'hui, ont été financés par des institutions publiques de santé ou de justice¹³. En effet, l'épidémiologie et la criminologie, disciplines appliquées et bien financées, sont clientes d'étude à dimension fortement spatiale pour affiner leur compréhension des phénomènes et améliorer leurs prédictions. La diffusion est au centre des travaux de la première, la concentration au cœur de la seconde. Dans ce dernier cas, les points chauds prennent un double sens, celui de concentration, mais aussi celui de zone dangereuse. C'est sans surprise qu'un des premiers logiciels gratuits permettant de calculer des indices locaux d'autocorrélation spatiale se nomme « Crimestat » (Levine, 2002 ; la première édition du logiciel date de 1999).

3.3.1. Saisir les structures localement

Les statistiques gamma (présentées plus loin) sont en quelque sorte une essentialisation des différents indices d'autocorrélation spatiale. Partant de cette figure simplifiée, Anselin développe son idée d'un indicateur local d'association spatiale¹⁴. En effet, puisque cette statistique gamma est une somme pour tous les « i », on peut définir pour chaque individu « i » un gamma local de la forme :

$$\Gamma_i = \sum_j w_{ij} v_{ij}$$

¹³ Ce détail a son importance car il a modelé en partie la réflexion des premiers auteurs sur la façon de développer les outils et orienté l'interprétation de ces indices. De plus, on peut envisager que la diffusion de ces outils ait été restreinte dans un pays comme la France où les géographes s'intéressent peu à ces problématiques.

¹⁴ A l'origine cette forme simplifiée des indices globaux d'autocorrélation spatiale est développée par Cliff et Ord (1981 : chapitre 2) pour mettre au point des tests statistiques de significativité.

On voit bien la relation qui unit un indice gamma local à l'indice global : l'indice global est la somme de tous les indices locaux.

$$\Gamma = \sum_i \Gamma_i = \sum_i \sum_j w_{ij} v_{ij}$$

Un facteur de proportionnalité γ peut éventuellement être introduit, cela ne change pas la relation : $\sum_i \Gamma_i = \gamma \cdot \Gamma$.

Cette particularité permet d'apprécier la valeur d'un gamma local en fonction du gamma global. Ainsi, il est aisé de mettre en évidence des valeurs aberrantes, en comparant la valeur du gamma local à celle de la valeur de la moyenne des gammas locaux (c'est-à-dire le gamma global divisé par le nombre d'individus « i »).

Anselin propose ainsi la notion de *Local Indicators of Spatial Association* (LISA), terme générique regroupant différents indices qui permettent de mesurer l'autocorrélation spatiale locale. Cet acronyme sera rapidement repris par les autres auteurs (Getis, Ord, 1996, Bao, Henry, 1996). Pour Anselin (1995 : 94), « un indicateur local d'association spatiale est n'importe quelle statistique répondant à deux critères :

- a) Le LISA de chaque observation donne une indication sur l'étendue d'un regroupement spatial significatif de valeurs similaires autour de cette observation.
- b) La somme des LISA pour toutes les observations est proportionnelle à un indicateur global d'association spatiale. »

Un LISA est donc un indice qui permet de mesurer localement l'autocorrélation spatiale et dont la somme pour tous les individus donne le niveau d'autocorrélation spatiale global de la variable dans l'espace étudié. La supériorité des LISA sur d'autres méthodes, comme les statistiques G_i et G_i^* , est ici marquée : les premiers permettent un aller-retour entre le local et le global, alors que les secondes ne proposent que la mise en évidence des points chauds. On rappellera enfin que les LISA ont deux fonctions complémentaires. La première est la mise en valeur des regroupements locaux de valeurs similaires (*spatial clustering*), particulièrement dans les cas où l'autocorrélation spatiale globale est nulle. La seconde leur permet de mettre en évidence une instabilité locale lorsque l'autocorrélation spatiale globale existe (les *outliers* dans les grandes tendances).

3.3.2. Introduction aux LISA : l'indice de Moran local et son utilisation

Ces LISA vont nous permettre de mettre en évidence la structure des indices étudiés à l'échelle de l'Inde du Sud. Nous utiliserons pour cela la version locale de l'indice de Moran (indice le plus répandu et le mieux maîtrisé, comme nous l'avons précisé précédemment).

L'Indice de Moran est bien une statistique gamma, puisque si l'on remplace « v_{ij} » par

$$\frac{(z_i - \bar{z})(z_j - \bar{z})}{\sum_i (z_i - \bar{z})^2},$$

on obtient bien la formule générale de l'indice de Moran présentée précédemment (après standardisation en ligne).

L'indicateur local (ou « Moran local ») est donné par la formule :

$$I_i = \frac{\sum_j w_{ij} (z_i - \bar{z})(z_j - \bar{z})}{\sum_i (z_i - \bar{z})^2}$$

Conformément à la règle, la somme des indices de Moran locaux est proportionnelle à l'indice de Moran global : $\sum_i I_i = \gamma.I$, avec $\gamma = n$.

Ce facteur de proportionnalité retenu n'est valable que si les variables sont standardisées et si la matrice de pondération est standardisée en ligne (ce qui est notre cas). Son principal intérêt est de proposer alors une relation simple et pratique de la proportionnalité entre indice global et indice local. En effet, la moyenne des indices locaux est alors égale à l'indice global. Il sera donc possible de proposer pour des sous espaces des indices de Moran globaux qui seront la moyenne pour ces sous espaces des indices locaux, et qui pourront être comparés à l'indice global pour l'ensemble du Tamil Nadu, mais aussi entre eux.

C'est aussi grâce à ce lien de proportionnalité que l'on peut obtenir pour chaque village une estimation de sa ressemblance avec les villages voisins par rapport à sa ressemblance à l'ensemble des villages, et tester la significativité de cette ressemblance. On distingue alors quatre cas de figures¹⁵ :

- Les villages avec un indice fort dans un voisinage qui lui ressemble (autocorrélation spatiale positive et valeur de l'indice élevé). Situation résumée en « plus-plus » (*high-high*).
- Les villages avec un indice fort dans un voisinage qui ne lui ressemble pas (autocorrélation spatiale négative et valeur de l'indice élevé). Situation résumée en « plus-moins » (*high-low*).
- Les villages avec un indice faible dans un voisinage qui lui ressemble (autocorrélation spatiale positive et valeur de l'indice faible). Situation résumée en « moins-moins » (*low-low*).
- Les villages avec un indice faible dans un voisinage qui ne lui ressemble pas (autocorrélation spatiale négative et valeur de l'indice faible). Situation résumée en « moins-plus » (*low-high*).

On a ainsi une indication sur le niveau d'autocorrélation spatiale (comme pour les indices globaux) mais aussi une information sur le type d'association présente.

On comprend mieux aussi pourquoi le niveau d'autocorrélation spatiale augmente lorsque l'on agrège les données. En effet, si l'on considère qu'un indice global est la somme des indices locaux, plus les indices locaux sont autocorrélés, plus le niveau global d'autocorrélation spatiale sera élevé. Or, plus la maille étudiée est fine, plus on a de chance d'avoir des valeurs extrêmes (*outliers*). Ou plutôt, plus on agrège les données, plus les valeurs extrêmes présentes localement sont absorbées dans des mailles de taille supérieure où la tendance domine. Ainsi, c'est la tendance dominante qui l'emporte, et les valeurs aberrantes disparaissent, renforçant l'autocorrélation spatiale globale. Sans approfondir ici le questionnement sur le MAUP¹⁶ (*Modifiable Areal Unit Problem* - problème des surfaces aréolaires variables), on doit néanmoins envisager les LISA comme une nouvelle piste d'exploration de ce problème inhérent à l'utilisation de données géographiques basées sur des mailles de taille inégale. Dans le même ordre d'idée, Amrhein et Reynolds (1996) proposent une exploration des effets de l'agrégation des données à partir des statistiques G_i .

¹⁵ Pour bien comprendre ce paragraphe, un parallèle peut être fait avec le nuage de point de Moran, présenté en 0).

¹⁶ Le terme a été proposé par Openshaw et Taylor (1979). La littérature sur la question est riche, on consultera généralement les travaux d'Openshaw (en particulier Openshaw, 1984) ou plus récemment ceux de Amrhein. Wong et Amrhein (1996) proposent un rapide rappel sur le sujet.

3.4. Interpréter les LISA

Avant d'aller plus loin, il convient de revenir sur un point technique, celui de la significativité des indices calculés localement.

3.4.1. Significativité de l'indice de Moran local

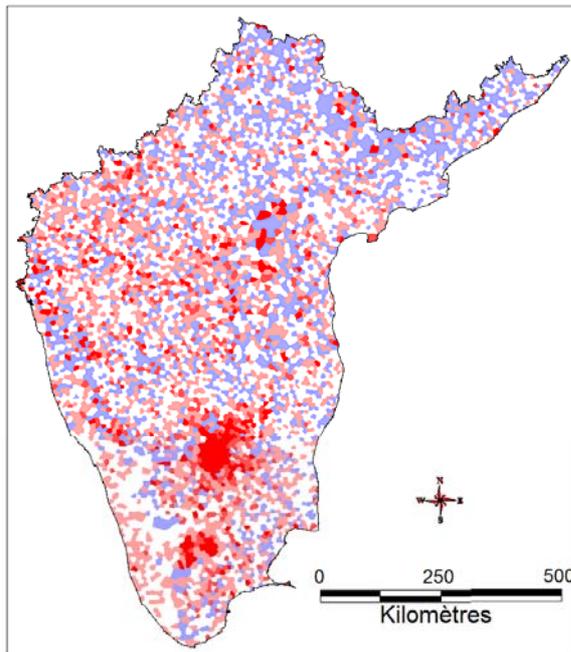
Comme pour des indices généraux de corrélation (spatiale ou non), il est possible de connaître le niveau de fiabilité (significativité) de l'indice de Moran local. La significativité exacte de l'indice calculé peut être approximée, mais la mise en œuvre de ce type de calcul est trop complexe. Anselin (1995 : 99) montre en effet que l'on peut construire un test de significativité en s'appuyant sur le 4^{ème} moment de l'indice local de Moran. Il rappelle cependant que la distribution exacte de cette statistique est encore inconnue. De plus, il rappelle (p. 108) que ces distributions pourraient être connues sous l'hypothèse d'autocorrélation spatiale globale nulle, ce qui est rarement le cas. Pour plus de détails, on consultera les travaux de Michael Tiefelsdorf (plus particulièrement Tiefelsdorf & Boots, 1997 et Tiefelsdorf, 1998).

C'est pourquoi les différents auteurs préfèrent utiliser une randomisation conditionnelle grâce à un test de Monte Carlo pour évaluer la significativité des statistiques locales (Moran local, Geary local, statistiques *G*). Ce test consiste à prendre les valeurs réelles (ce qui constitue la condition) des voisins de chaque point, et de les mélanger aléatoirement (randomisation) pour voir si l'indice calculé est dû ou non au hasard. L'opération est répétée un certain nombre de fois afin de calculer une probabilité. Celle-ci est exprimée en pourcentage. On peut cartographier cette probabilité et la comparer à la distribution des valeurs de l'indice.

Sans s'attarder encore sur la répartition des zones où l'indice est significatif (voir sur la Figure 11), on notera cependant que les espaces où il n'est pas significatif semblent se placer en "tampon" entre des zones d'autocorrélation spatiale positive et des zones d'autocorrélation spatiale négative. On remarquera aussi que le niveau de significativité est organisé autour d'un centre de valeurs fortement significatives. Cette organisation suggère une première lecture en termes de degré d'hétérogénéité. Ainsi les zones blanches (non significatives) recouvrent deux types d'unités, lesquelles peuvent d'ailleurs coexister. Les premières ressemblent au village moyen et ont donc des valeurs proches de la moyenne, ce qui leur confère une autocorrélation spatiale locale faible (puisque leur écart à la moyenne est réduit), alors même qu'il peut s'agir de zones homogènes. Les secondes ont des valeurs d'autocorrélation spatiale locale variée, mais dans un voisinage où les valeurs d'autocorrélation spatiale locales sont hétérogènes. Il en résulte qu'après un test par randomisation, on ne peut pas dégager de tendance locale, positive ou négative (d'où l'absence de significativité). La somme des indices de Moran locaux calculés localement est donc nulle.

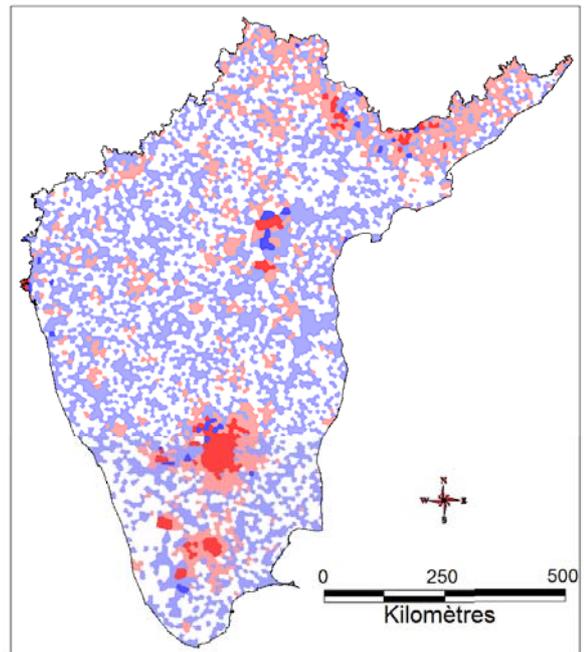
3.4.2. Interprétation

Les informations disponibles pour chaque village sont maintenant de trois ordres. Elles concernent les valeurs de la variable, le niveau local d'association spatiale (la valeur de l'indice local de Moran), et l'intervalle de confiance (le niveau de significativité). En couplant ces trois données, on peut obtenir une vue synthétique de la distribution des valeurs à l'échelle de l'Etat en séparant les villages selon leurs caractéristiques.



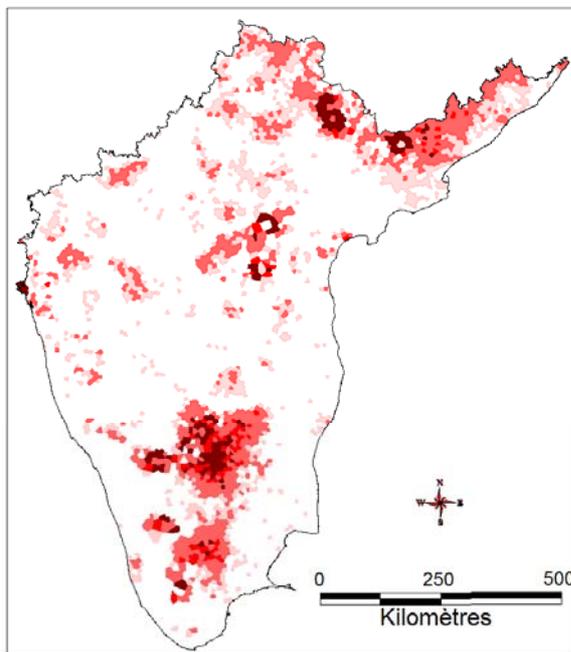
Sex ratio des enfants

- supérieur à 1.15
- 1.1 - 1.15
- 1.05 - 1.1
- 1 - 1.05
- inférieur à 1



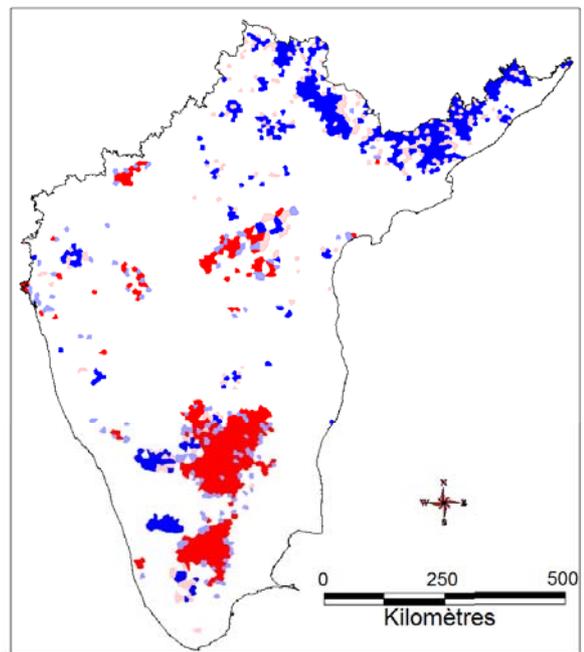
Niveau d'autocorrélation spatiale locale

- ACS locale fortement positive
- ACS locale positive
- ACS locale positive mais inférieure à la moyenne
- ACS locale négative
- ACS locale fortement négative



Niveau de significativité des LISA

- significativité trop faible (50 à 90%)
- significativité faible (90 à 95%)
- bonne significativité (95 à 99%)
- forte significativité (99 à 99,5%)
- très forte significativité (supérieure à 99,5%)



Tri des LISA par cluster

- valeurs non significatives
- valeurs fortes dans un voisinage de valeurs fortes
- valeurs faibles dans un voisinage de valeurs faibles
- valeurs faibles dans un voisinage de valeurs fortes
- valeurs fortes dans un voisinage de valeurs faibles

Figure 11 : 3 en 1, la carte des clusters de LISA. Un exemple avec le sex ratio des enfants

On va ainsi trouver d'abord des villages dont les valeurs de LISA sont non significatives. Ces villages seront exclus du reste du traitement. On trouvera ensuite des valeurs autocorrélées positivement et des valeurs autocorrélées négativement. Chacun de ces deux groupes pourra être divisé selon la valeur originale de l'indice : supérieur ou inférieur à la moyenne. On crée ainsi quatre catégories, qui renvoient au nuage de points de Moran présenté en annexe (voir à la fin) :

- Les valeurs positives (indice supérieur à la moyenne) au sein d'un environnement de points à valeur positive (autocorrélation spatiale positive). Association qualifiée de *high-high* dans la littérature anglo-saxonne.
- Les valeurs négatives (indice inférieur à la moyenne) au sein d'un environnement de points à valeur négative (autocorrélation spatiale positive). Association qualifiée de *low-low*.
- Les valeurs positives au sein d'un environnement de points à valeur négative (autocorrélation spatiale négative). Association qualifiée de *high-low*.
- Les valeurs négatives au sein d'un environnement de points à valeur positive (autocorrélation spatiale négative). Association qualifiée de *low-high*.

La Figure 11 présente la combinaison de trois cartes (valeurs de l'indice, valeurs locale du niveau d'autocorrélation spatiale et niveau de significativité de l'autocorrélation spatiale locale) pour en obtenir une quatrième qui les synthétise.

On met ainsi en valeur des poches d'autocorrélation spatiale positive, qui sont liées à des regroupements de valeurs fortes (ici représentées en rouge foncé) ou de valeurs faibles (en bleu foncé). A contrario, on isole aussi des éléments isolés dans un voisinage homogène. Ce sont des villages dont les valeurs tranchent avec leur voisin, et que l'on peut qualifier de valeurs spatialement atypiques (*spatial outliers*) car elles ne suivent pas la tendance locale. Ils sont représentés en rose pour les valeurs fortes dans des régions de faible modernisation, et en bleu clair pour les valeurs faibles dans les régions de forte modernisation.

Appliqués à des phénomènes qui sont globalement peu autocorrélés, comme c'est le cas du sex-ratio des enfants, les LISA permettent de révéler des sous espaces qui pourraient être autocorrélés (Figure 11, carte des clusters de LISA : au sud, on voit nettement deux poches importantes de sex-ratio déficitaires pour les petites filles).

Face à des phénomènes fortement autocorrélés, la cartographie des LISA permet de mieux comprendre la structure locale des données. Par exemple, si l'on s'intéresse à l'alphabétisation (Figure 12), on voit bien la distribution des zones peu alphabétisées et l'importante homogénéité des zones où elles se situent permettant de souligner l'aspect régional de la faible alphabétisation.

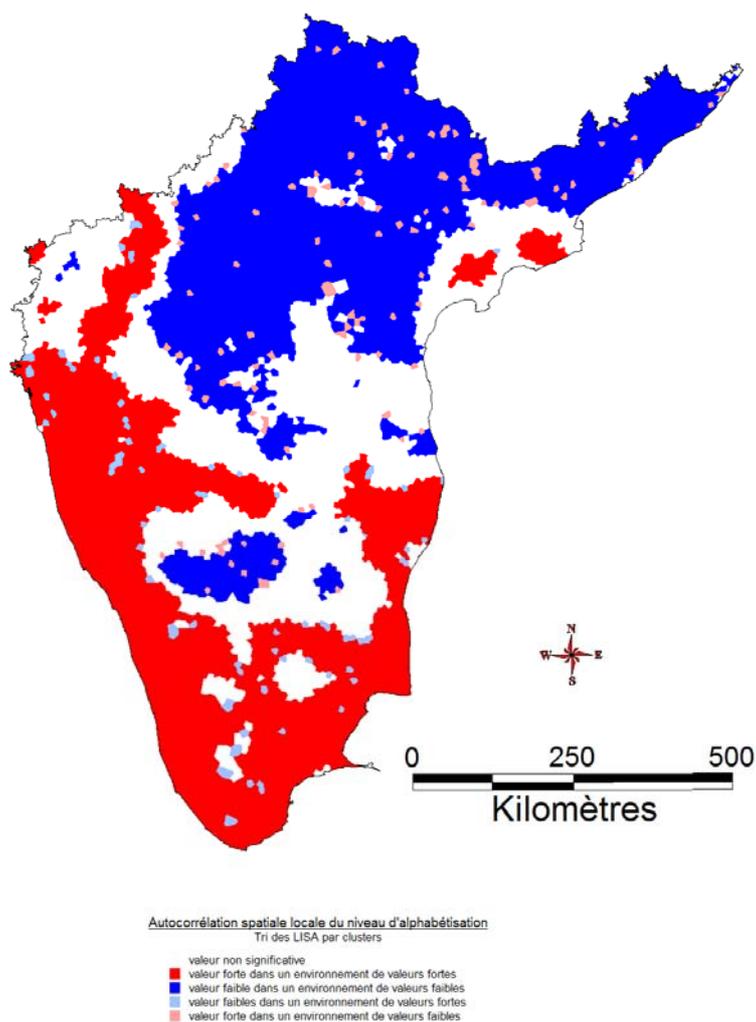


Figure 12 : autocorrélation spatiale locale de l'alphabétisation.

3.4.3. Une réflexion à propos des échelles

Il est intéressant de souligner que la portée du voisinage définie pour mesurer l'autocorrélation spatiale influe directement sur la perception que l'on peut avoir du phénomène. Ainsi, si nous avons remarqué précédemment que le niveau d'autocorrélation spatiale globale variait en fonction de la distance entre les points, on peut mettre en valeur les changements de structure de l'autocorrélation spatiale globale en cartographiant les niveaux d'autocorrélation spatiale locaux pour différents voisinages.

C'est ce que nous montre la Figure 13, qui présente pour l'alphabétisation la mesure de l'autocorrélation spatiale locale en fonction d'un voisinage de 25, 50 puis 100 kilomètres. On comprend mieux ainsi pourquoi l'autocorrélation spatiale globale tend à diminuer, puisqu'elle passe de 0,82 pour un voisinage de 25 km à 0,76 pour un voisinage de 50 km, et enfin à 0,64 pour un voisinage de 100 km¹⁷. En effet, au fur et à mesure que la distance augmente, le

¹⁷ Il s'agit là du niveau de voisinage, c'est-à-dire de la distance entre voisins pris en compte pour le calcul de l'indice de Moran. On ne manquera pas de faire le lien avec l'augmentation du niveau d'autocorrélation spatiale évoquée précédemment (p. 37) qui se produit lorsque le niveau d'agrégation augmente, pour la raison inverse (le nombre de valeurs spatialement aberrantes tendant à diminuer).

nombre de voisins pris en compte augmente, ce qui tend à définir des régions toujours plus grandes, mais dont l'hétérogénéité interne augmente. De facto, l'indice de Moran, qui mesure cette hétérogénéité tend à diminuer. On le voit d'ailleurs sur les figures avec l'augmentation du nombre de clusters ayant des valeurs spatialement atypiques (en bleu clair et en rose) lorsque la distance du voisinage augmente. Nous devons donc souligner ici l'importance à accorder à l'étude préliminaire du choix de la distance de voisinage. Par ailleurs, la mesure locale de l'autocorrélation spatiale se révèle un outil précieux pour comprendre l'organisation des phénomènes sociaux à différentes échelles.

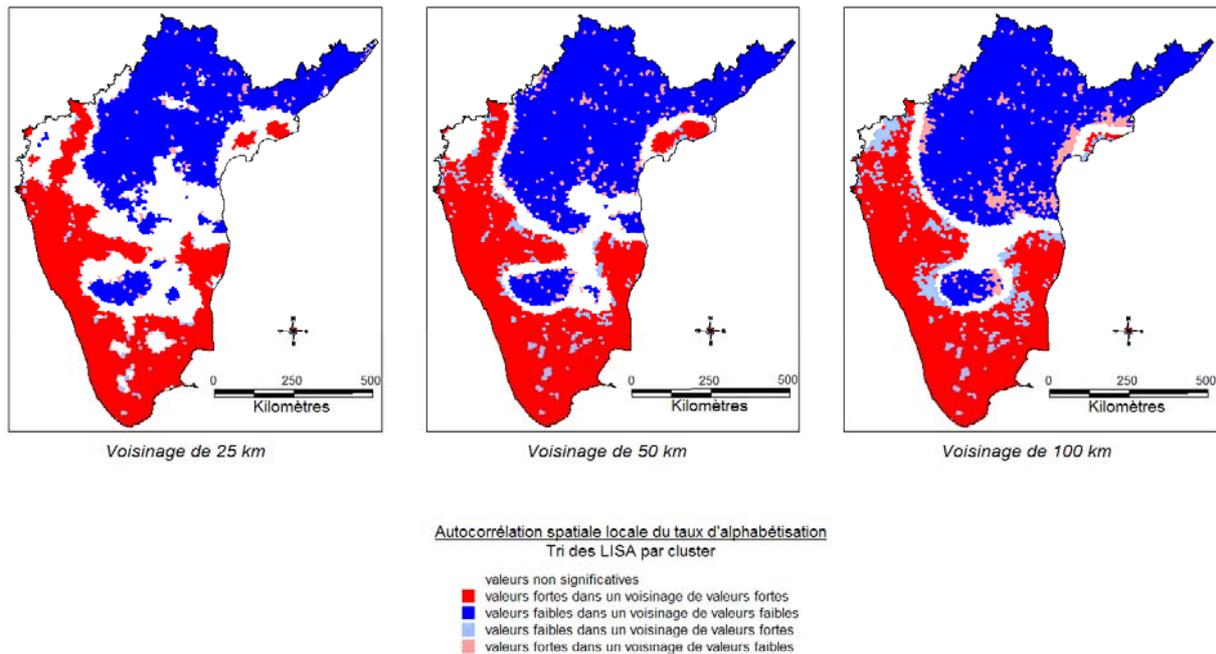


Figure 13 : autocorrélation spatiale de l'alphabétisation selon trois définitions du voisinage.

4. Une thématique spatiale : la vulnérabilité

L'aspect formel des structures spatiales qui a constitué une part importante de notre travail masque l'intérêt des thématiciens qui constituent notre équipe pour différents aspects de la géographie contemporaine de l'Inde, dans ses dimensions sociales, économiques ou encore démographiques. Notre collaboration au sein du projet EMIS a précisément permis d'entrevoir une application conjointe de nos analyses spatiales, autour d'une thématique partagée relative à l'Inde d'aujourd'hui. C'est la raison pour laquelle cette section débute par un bref survol des enjeux de la question abordée, celle de la vulnérabilité et de la pauvreté, et leur traduction spatiale. Il s'agit d'un des enjeux de recherche prioritaires en Inde et nous allons brièvement montrer, avant l'application pratique de notre analyse spatiale et géostatistique, que la déclinaison géographique de la question de la vulnérabilité a des implications à la fois pour la réflexion et pour l'intervention qui sont essentielles.

4.1. Le concept de la vulnérabilité et sa dimension géographique

La vulnérabilité est un concept multidimensionnel s'opérant à tous les niveaux: individuel, communautaire, régional, national. Ce concept est étroitement lié à la notion de risque, susceptible d'être encouru par différentes catégories. Le risque est associé aux chocs susceptibles de se produire dans des domaines très variés comme par exemple ceux du climat (sécheresse, inondation, etc.), des ressources environnementales (production agricole, dégradation de la biosphère, etc.), des catastrophes naturelles, politiques ou sanitaires (tremblements de terre, tsunamis, conflits, épidémies etc.) et des politiques économiques (inflation, politiques d'ajustement structurel, crise, etc.). Les impacts négatifs potentiels de ces risques se situent à plusieurs niveaux: dommages corporels, perte de la vie, pertes de ressources (physiques, monétaires, actifs), et naturellement pauvreté. L'identification et l'analyse de ces risques encourus par les catégories impliquent donc des considérations géographiques locales, renvoient à la question du degré d'exposition de ces populations à ces risques, et par conséquent font appel à la notion de vulnérabilité.

Chaudhuri, Jalan et Suryahadi (2001) définissent la vulnérabilité comme suit: «*We define vulnerability (...) as the ex-ante risk that a household will, if currently non-poor, fall below the poverty line, or if currently poor, will remain in poverty*». Une autre définition de la vulnérabilité est donnée par Glewwe et Hall (1998) et Cunningham et Maloney (2000), qui soulignent spécifiquement le rôle des chocs. Quelque soit le niveau de vie, la vulnérabilité d'un individu est donc la probabilité de voir sa situation ou ses conditions de vie se détériorer et de tomber dans la pauvreté, en dessous de la ligne de pauvreté.

La vulnérabilité est donc associée au degré de risque encouru par les catégories, et à leurs capacités à anticiper, à faire face aux chocs et à remonter la pente. Autrement dit, la probabilité de trouver des catégories vulnérables dans des zones géographiques exposées aux risques est très élevée (Currey, 2002). Le degré d'exposition d'une région donnée à des risques et les capacités de réaction aux chocs des populations déterminent le niveau de vulnérabilité. Ce dernier, étroitement dépendant du contexte local (physique, environnemental, social, politique, économique), varie d'une personne à une autre, d'une catégorie à une autre, d'une région à une autre. La vulnérabilité renvoie à de multiples dimensions qui s'imbriquent entre elles: physique et environnementale, démographique, socio-économique et socioculturelle. Cet ensemble de facteurs (Alwang, Siegel et Jorgensen, 2001) traduit le contexte

géographique et l'organisation économique et sociale dans lesquels le degré de vulnérabilité s'insère. On notera quelques grands domaines de facteurs :

- physiques: la localisation géographique détermine les conditions agroclimatiques (Etat des sols, degré de sécheresse, pluviosité, irrigation, points d'eau, terres cultivables, cultures de céréales, riz, etc.), conditions environnementales (capital naturel: forêt, bois, animaux, troupeaux, etc.)
- infrastructurels : dotations régionales en infrastructures, telles que : électrification, routes, communications, marchés, accès à l'eau, hôpitaux, centres de soins, écoles, proximité d'un centre urbain, etc.
- démographiques : densité de la population, la distribution par âge et sexe, fécondité, mortalité.
- socio-économiques : la vulnérabilité renvoie aux revenus tirés de différentes activités (emplois exercés, production agricole et non agricole), aux transferts et aux actifs physiques et financiers. Il faut tenir compte également des conditions de vie au niveau de l'habitat (type de logement, etc.).
- socioculturelles : les préférences culturelles et le statut social peuvent dans certaines sociétés déterminer la vulnérabilité: les femmes, les enfants, les personnes âgées, les migrants, les malades, les hors-caste et groupes marginalisés sont des groupes vulnérables.

La localisation géographique des populations, la nature et l'imbrication de ces dimensions constituent des facteurs importants affectant le degré, la distribution spatiale et la dynamique de la vulnérabilité

4.2. Espace et vulnérabilité : les trappes de pauvreté

Les analyses de la pauvreté et de la vulnérabilité s'effectuent généralement au niveau national et global, à partir de données agrégées. Ces approches demeurent insuffisantes en termes de politiques de développement local, en particulier lorsque l'on s'attache à la localisation de la pauvreté et aux différences spatiales de pauvreté à un niveau plus désagrégé (dans un pays, une région, etc.). Ceci est d'autant plus pertinent que dans la plupart des pays (Hentschel, Lanjouw J. et al, 1998 ; Lachaud, 1999a et 1999b), il existe de fortes disparités spatiales entre régions en termes de pauvreté et de développement, en raison de spécificités locales associées à chaque région (dimensions physique et environnementale, démographique, socio-économique, socioculturelle).

Dans cette perspective¹⁸, des bases de données spatialisées (SIG) sont conçues à une échelle désagrégée, afin de visualiser la distribution spatiale de la pauvreté et les inégalités spatiales entre régions (Deichmann, 1999 ; Hentschel, Lanjouw¹⁹ et al, 1998). Ces représentations spatiales permettent en outre de préciser la localisation des différentes dimensions (physique et environnementale, démographique, socio-économique, socioculturelle) dans lesquelles s'insère la vulnérabilité ainsi que d'évaluer son degré. Ces cartes de profils de la pauvreté mettent aussi en évidence des zones particulières concentrées localement, appelées «trappes de pauvreté localisées» (*spatial poverty traps*). Jalan et Ravallion (1997) les définissent comme suit: «*Consider two households living in different areas but identical otherwise.*

¹⁸ La question de la représentation spatiale de la pauvreté (outils, bases de données, méthodes, indicateurs) est largement développée par Deichmann [1999] et Bigman et Deichmann [2000].

¹⁹ Lanjow et Hentschel ont mis en œuvre une méthode d'estimation pour construire un indicateur de pauvreté à l'échelle désagrégée, en fonction de facteurs explicatifs, pour l'Equateur. Ils sont amenés à combiner des données issues de bases différentes (LSMS, Census).

Suppose that one of the areas is less well endowed with physical, human and social capital – in short geographic capital- than the other. A spatial poverty trap can be said to exist if the household living in the better endowed area sees its standard of living rising over time, while the others does not».

Cette définition renvoie à la notion de vulnérabilité dans la mesure où la concentration spatiale de la pauvreté traduit des différences entre régions en termes de dotations et d'opportunités. Autrement dit, les caractéristiques de ces «trappes de pauvreté localisées»²⁰ constituent des obstacles structurels de différents ordres, imbriqués entre eux, qui empêchent les individus d'échapper à leur condition de catégories vulnérables et pauvres. En termes de développement, cela implique donc pour les politiques de mettre en œuvre des actions décentralisées et localisées (infrastructures, marchés du crédit et de l'emploi, politique agricole, etc.) pour identifier les catégories vulnérables et les empêcher de tomber en dessous de la ligne de pauvreté. Ces politiques, beaucoup mieux orientées, évitent en outre la dispersion de l'effort en portant précisément l'accent sur des zones particulières. En reconnaissant la nature géographique des poches de pauvreté, elles s'avèrent plus efficaces que les politiques généralistes.

4.3. Une première représentation de la géographie de la pauvreté

L'intérêt théorique pour la cartographie fine de la pauvreté (*poverty mapping*) est porté par l'existence de réelles poches de concentration des groupes vulnérables. Pour s'en convaincre, cette brève section utilise une approche synthétique à titre d'illustration²¹. Nous avons construit un premier indice de vulnérabilité que nous avons cartographié.

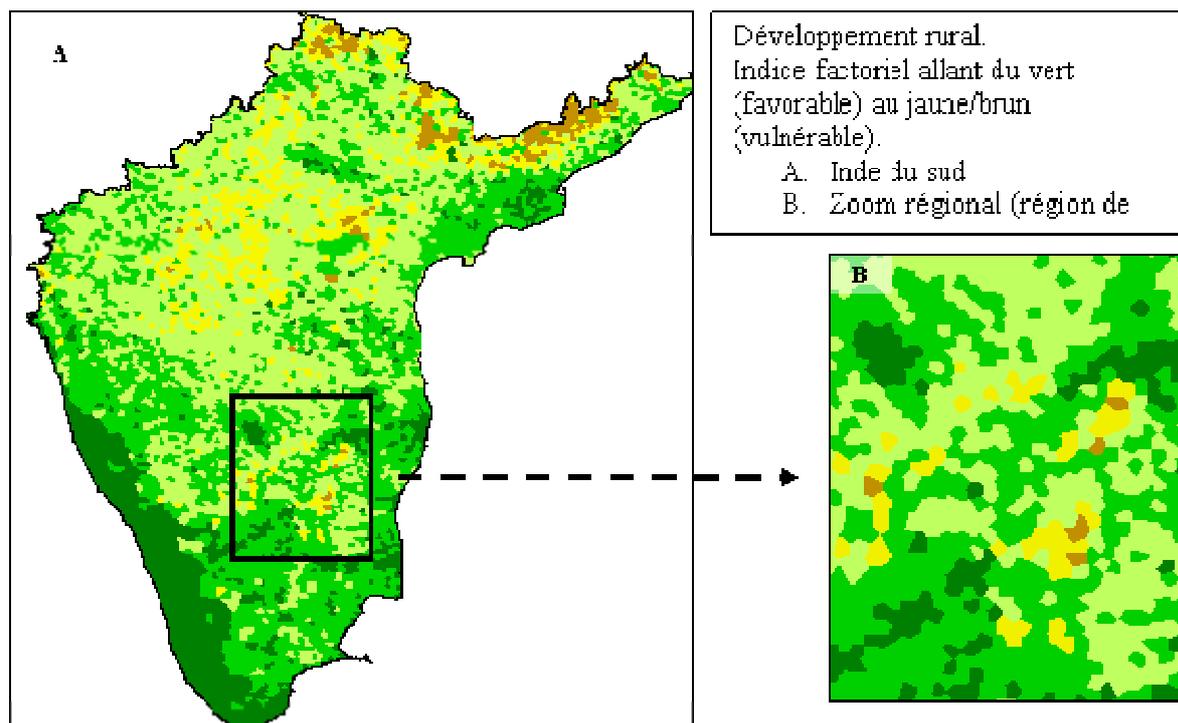
Il s'agit ici d'une mesure relative au développement global des régions rurales. Cet indice a été calculé directement à partir de notre base de données comme produit factoriel de différents indices de sous-développement rural : l'absence d'irrigation, l'analphabétisme, la prépondérance de l'activité agricole, la forte fécondité, la population tribale et le sous-développement du secteur tertiaire. Nous avons extrait ici le premier facteur de l'analyse en composantes principales.

La carte de gauche (Figure 14.A) a été préparée à partir de clusters spatiaux de villages (afin d'éviter les sous-effectifs) autour d'un rayon de 5 km. Chaque cluster a été représenté comme polygone de Voronoi et le retard de développement est figuré en teinte jaune ou brune sur les cartes présentées. Les niveaux favorables de développement rural sont marqués en vert.

La carte de droite (Figure 14.B) est un zoom sur une région frontalière entre Tamil Nadu, Andhra Pradesh et Karnataka, marquée par un relief accidenté, un peuplement à forte composante tribale et des situations socioéconomiques défavorables. On notera le grain de la carte (détail assez fin sur des micro-régions) ainsi que la forte «clusterisation» du phénomène du sous-développement à l'échelle régionale et locale, confirmée par un fort niveau d'autocorrélation spatiale. Si cette mesure du sous-développement est encore provisoire, ne prenant en compte qu'un nombre limité de dimensions de vulnérabilité, elle démontre la forte structuration spatiale du développement rural.

²⁰ Pour les approches théoriques en termes de modèles de croissance, se reporter à Azariadis et Costas [1996]. Jalan et Ravallion [1997, 2002] ont développé et testé économétriquement un modèle de croissance microéconomique pour expliquer les trappes de pauvreté localisées (Cas de l'Indonésie et de la Chine).

²¹ A propos des méthodologies du poverty mapping, voir Davis (2003).



Source : Aubriot et Guilmoto (2003)

Figure 14 : mesure factorielle du développement rural

Les facteurs régionaux à étudier sont potentiellement très divers et doivent inclure en outre les infrastructures locales, l'accessibilité aux marchés extérieurs, les données physiques, etc. Qu'elles soient observables ou non, ces caractéristiques manifestent un fort degré d'autocorrélation spatiale et constituent une « dotation géographique » (*geographic endowment*) spécifique. Le clustering spatial de la pauvreté peut ainsi être considéré comme une dimension aggravante du phénomène et l'existence de poches de concentration au sein de localités défavorisées constitue un piège à pauvreté.

Parmi les questions adjacentes, notons que la capacité à créer des indicateurs *locaux* de fragilité des communautés est cruciale pour identifier les zones à risque. Cette mesure doit se faire à partir des mesures (spatialement) lissées appropriées, voire par imputation macro-micro (combinaison d'indices régionaux et locaux). Une question essentielle concerne les formes de corrélations spatiales des indicateurs obtenus (regroupement en archipel, pôles centrés, etc.). La possibilité d'un indicateur genré peut permettre de relever la dimension propre à la vulnérabilité des femmes (emploi, survie, etc.). Ces questions sont abordées à partir de la section suivante de manière plus systématique.

4.4. Les dimensions de la pauvreté : définition, estimation et opérationnalisation

La dimension spatiale propre aux phénomènes de vulnérabilité incite à mettre notre connaissance du paysage sud-indien au service de l'identification des poches de concentration des populations vulnérables, et plus particulièrement des populations pauvres. Il demeure que l'approche multidimensionnelle désormais adoptée par les économistes en matière de vulnérabilité impose de confronter plusieurs grilles d'analyse et d'identification de la

vulnérabilité. Cela est rarement possible dans les pays en développement, en raison de l'imperfection des sources statistiques disponibles. Rajoutons en outre que ces informations à caractère très disparate (environnementales, démographiques, économiques, etc.) émanent le plus souvent de sources très distinctes et se donnent dans des formats différents et incompatibles (données de ménages, individuelles, agrégées régionalement, satellitaires, etc.). Pour finir, on soulignera que les données ne sont généralement pas désagrégées à micro-échelle et n'existent pas en format cartographiables. Ceci explique que les travaux dans le domaine restent exploratoires et très partiels, en dépit de l'intérêt théorique et politique de forger des nouveaux instruments pour l'identification des concentrations de pauvreté.

Sur ces derniers points, notre base de données présente des avantages notables puisqu'elle offre un SIG déjà construit, doté d'une armature statistique de taille et de diversité satisfaisante, et fondée sur une grille géographique très fine. L'Inde du sud est loin d'être la zone la plus vulnérable du pays, même si elle a été affectée directement par la catastrophe du tsunami en décembre 2004. Son niveau de développement social et économique est avancé et son essor économique récent illustré par ses métropoles (Chennai, Hyderabad, Bangalore) est très rapide. Cet essor reste toutefois inégalement partagé au sein de l'Inde du sud et comme on va le voir, de larges segments régionaux des campagnes demeurent très en retrait de la modernisation urbaine.

Nous ne sommes naturellement qu'à une étape intermédiaire de ce travail, d'autant qu'aucune recherche comparable (en termes d'échelle ou de portée théorique) n'a encore été conduite en Inde²². Les définitions utilisées pour quantifier et cartographier les différentes dimensions de la vulnérabilité sont donc loin d'être définitives²³ et constituent avant tout des propositions pour une consolidation statistique ultérieure. Certaines dimensions font en effet défaut dans notre analyse et nous allons passer en revue la gamme de données déjà disponibles et celles qui ont été acquises pour compléter l'étude de la vulnérabilité. Un accent particulier est mis dans une première section sur les données concernant le milieu et l'irrigation, qui sont cruciales pour définir la base agricole des zones rurales qui nous intéressent. On ne dispose en effet pas de données proprement économiques (production par valeur et récolte, par exemple), même à une échelle plus agrégée du district pour lequel les statistiques sont éparées. En revanche, les données censitaires et climatiques apportent un éclairage tout à fait précieux dont il faut discuter au préalable la validité.

4.4.1. *Pluviométrie et irrigation*

L'Inde est encore fortement agricole, 58% de sa population active travaillant dans le secteur primaire en 2001 (Oliveau 2003) et un bref rappel des systèmes de production ruraux est nécessaire pour marquer l'importance de l'irrigation, qui permet d'augmenter les rendements et d'entreprendre des cultures consommatrices d'eau telles que le riz²⁴ et la canne à sucre. L'irrigation fut d'ailleurs l'un des fers de lance (avec les engrais et les variétés améliorées) de la Révolution Verte qui permit au pays d'atteindre l'autosuffisance alimentaire en une

²² Pour des études à l'échelle du district en Inde, voir par exemple les travaux récents de O'Brien *et al.* (2004), Jayaraman and Srivastava (2003). Le second document explore la cartographie pour des unités administratives inférieures (*mandal*) en Andhra Pradesh.

²³ On prendra pour exemple l'absence dans notre étude de mesure de la vulnérabilité côtière que la catastrophe du tsunami de décembre 2005 (qui a frappé de plein fouet le littoral oriental de l'Inde du sud) est venue illustrer.

²⁴ Nous considérons ici le riz humide qui nécessite une lame d'eau dans la parcelle pendant une bonne partie du cycle cultural. Le riz sec, également cultivé en Inde, se satisfait des apports d'eau par les pluies.

décennie (1965-1976) (Landy, 1997: 272) et de multiplier sa superficie brute irriguée par 2,5 en 35 ans (Dorin et Landy, 2002 : 272). L'Inde détient aujourd'hui la première place mondiale en termes de pourcentage de terres irriguées et de superficies totales irriguées. L'importance de l'irrigation en Inde n'est toutefois pas nouvelle, divers procédés ayant été mis en oeuvre pour pallier au climat semi-aride que subit une bonne partie du pays. Ainsi, le pays est très riche en techniques anciennes, pour certaines toujours utilisées.

L'irrigation ne répond pas aux mêmes besoins ni aux mêmes contraintes selon les endroits de l'Inde : d'une part, les pluies présentent une très forte variabilité spatiale rendant en certains lieux les cultures pluviales très productives et riches ; d'autre part, la disponibilité en eau peut être elle aussi variable, dépendante de divers facteurs environnementaux et anthropiques. Parmi les conséquences les plus marquantes de ces dernières années du manque d'accès à l'eau d'irrigation, on peut noter l'accroissement des migrations saisonnières vers les villes des ouvriers agricoles (des sans-terre) lors de la saison sèche ; ou encore le suicide d'agriculteurs surendettés par l'achat de pompes, l'approfondissement de leur forage etc.

Les données sur l'irrigation sont les principales informations sur l'agriculture disponibles dans le recensement²⁵. On y trouve ainsi la surface irriguée par chacun des onze types d'irrigation distingués par le recensement. Il est ainsi possible de calculer et cartographier :

- la part d'irrigation que nous nommerons densité d'irrigation, c'est-à-dire le pourcentage de terres irriguées parmi les terres cultivées;
- la part de chaque grand type d'irrigation, les onze types du recensement pouvant être ramenés à 3 ou 4 grands types (canal, tank²⁶, puits et/ou forages).

Le raisonnement porte donc sur des superficies irriguées et des proportions pour chaque type d'irrigation. Cela a un double avantage : 1° de considérer tous les types d'irrigation, et non pas uniquement les zones qui sont densément irriguées ou qui dépendent d'un projet d'aménagement; seuls éléments qu'indiquent de nombreux ouvrages traitant de l'irrigation à l'échelle nationale, y compris l'atlas de l'irrigation (cf. plate 22 vol.2 de Kundu 1989) ; 2° de comparer les zones de petite, moyenne et grande hydraulique selon les mêmes critères, quantifiables. A l'échelle régionale, l'atlas de l'irrigation (Kundu, 1989) par exemple cartographie aisément les limites des réseaux aménagés par projets, alors que la place relative de l'irrigation par puits ou par tanks y est difficile à estimer. Chaque symbole représente certes 4000 hectares irrigués par les puits, forage ou tank, mais le figuré est peu visible comparativement à la coloration verte utilisée pour les réseaux irrigués par canaux. De même, Spate *et al* (1967 : 231) localise par des symboles les secteurs où puits et tanks sont très présents, mais la part relative de l'irrigation n'y est pas discernable. D'autres types de cartes sont alors nécessaires, représentant par exemple la part d'irrigation par tank d'une unité administrative par rapport à la superficie totale irriguée par ce type (Vaidyanathan, 2001). Mais elles ne sont disponibles qu'à l'échelle du district tout au plus. Le recensement est la première base de données à pouvoir localiser précisément ces informations, sur une vaste étendue, et à les quantifier selon des critères identiques. Et dans le cadre du présent projet sur la vulnérabilité, s'intéresser plus précisément aux situations de moindre irrigation apparaît primordial.

²⁵ Outre les superficies irriguées, les données concernent les superficies cultivées non irriguées, les superficies non cultivables, les superficies non utilisées, les forêts.

²⁶ Les « tanks » sont des étangs-réservoirs endigués, stockant les eaux de pluie, de ruissellement et parfois alimenté par l'eau de rivière.

La difficulté d'analyse de ces données réside dans la diversité des situations, ainsi que dans le manque de précision – malgré cet effort de distinguer onze types d'irrigation. Ces données pèchent principalement par l'absence d'information sur l'intensité de l'irrigation, c'est-à-dire le nombre de culture (une, deux ou trois par an) effectuée sous irrigation. Or, la fragilité du système de production - ce que nous pouvons appeler la vulnérabilité agricole - d'un agriculteur irriguant ses terres toute l'année à l'aide par exemple d'un tank bien alimenté en eau pendant la mousson puis d'un puits, n'est pas la même que celle de l'agriculteur qui dépend d'un tank pluvial. Pourtant, le recensement peut attribuer à ces deux hommes la même superficie irriguée par tank. On voit à travers cet exemple qu'une même terre peut être irriguée par deux types d'irrigation au cours de l'année ; or, un seul type est pris en compte dans le recensement (celui utilisé pendant la principale saison de culture, la mousson). On y voit également l'absence de distinction entre tanks pluviaux et systèmes de tanks alimentés par une rivière. L'accès à la ressource n'est pas assuré de la même façon dans les deux cas, le premier dépendant exclusivement de la pluviosité, et rendant ainsi l'agriculture plus risquée que dans le second cas²⁷.

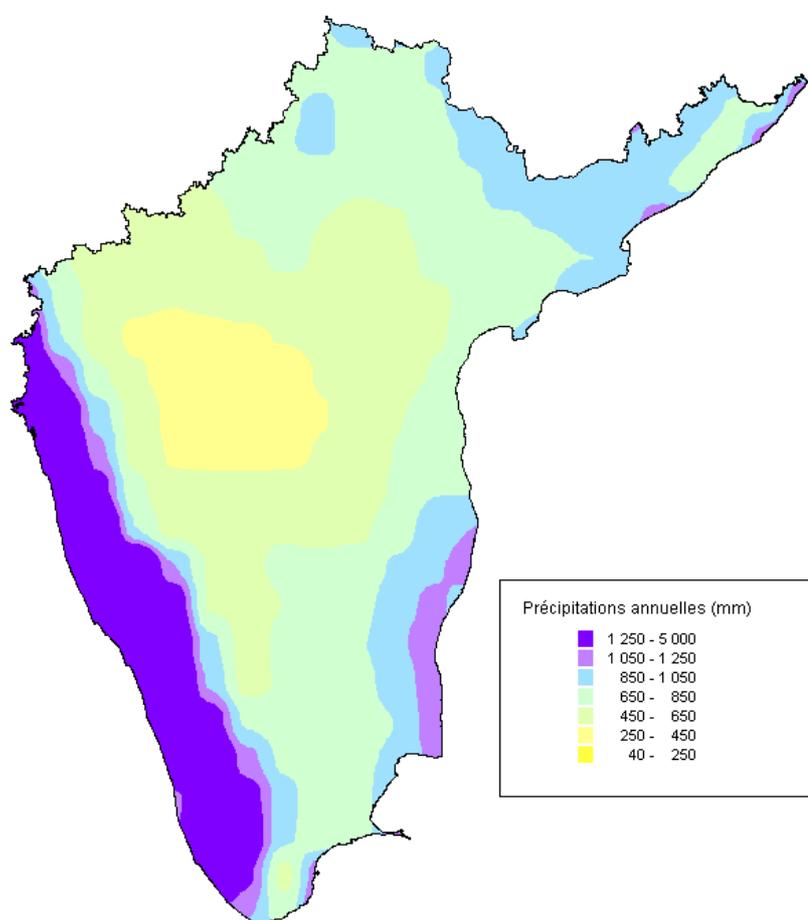


Figure 15 : pluviométrie dans le sud de l'Inde

²⁷ Certes la plupart des rivières sont non pérennes, et donc fortement dépendantes de la pluviosité. Toutefois, certaines prenant leur source à plusieurs dizaines de kilomètres peuvent être alimentées par des pluies localisées en amont ; par ailleurs, nombre d'entre elles ont leur débit régulé par un barrage. L'alimentation des tanks en systèmes s'en trouve nettement sécurisée.

Quant à la qualité des données, elle n'est pas toujours fiable pour des monographies, car souvent les données locales s'avèrent de mauvaise qualité (Chatillon, 2004 et observations de terrain d'O. Aubriot et autres membres de l'EMIS). Elles ne peuvent pas non plus être comparées aux données du PWD (*Public Work Department*), l'institution responsable des infrastructures majeures en irrigation. En revanche, pour une étude régionale s'intéressant aux principales caractéristiques de l'irrigation, à la répartition spatiale des différents types d'irrigation, les données semblent correctes et permettent d'obtenir des résultats d'ensemble intéressants. Le fait de travailler au niveau de clusters 5 km permet en outre de gommer certaines absurdités locales et de faire ressortir les caractéristiques de la dimension considérée à cette échelle.

S'intéresser à la vulnérabilité sociale et économique liée à l'irrigation incite à construire un indice de "vulnérabilité hydraulique", lié à la plus ou moins grande assurance de productivité agricole offerte par l'irrigation. Malheureusement une telle entreprise n'est pas réalisable avec cette base de données. Connaître les types d'irrigation permet certes d'estimer la plus ou moins bonne correction temporelle ou spatiale permise par chaque type (Durand Dastès, 1995). Mais nous venons de le voir, les données sur l'irrigation manquent de précision quant à l'intensité de l'irrigation et la catégorisation des types d'irrigation. Par ailleurs, le risque agricole n'est pas uniquement lié au type d'irrigation : les cultures pratiquées interviennent aussi, certaines étant très dépendantes d'un apport régulier d'eau ; l'accès à la ressource et la disponibilité de celle-ci sont également primordiales dans l'évaluation du risque agricole, données qui dépendent de plusieurs facteurs. On peut mentionner pour les eaux de surface : la climatologie (pluviosité, évapotranspiration), les caractéristiques de ruissellement et les capacités de stockage de l'eau (barrages, tanks), la position dans le réseau d'irrigation ou dans le bassin versant (les agriculteurs en aval dépendant, pour la disponibilité de la ressource, de l'utilisation de l'eau plus en amont) ; et pour les eaux souterraines : la pédologie (capacité de rétention d'eau), la géologie (capacité d'accès à la nappe), la profondeur de la nappe (liée notamment à l'exploitation de celle-ci et donc à la demande en eau).

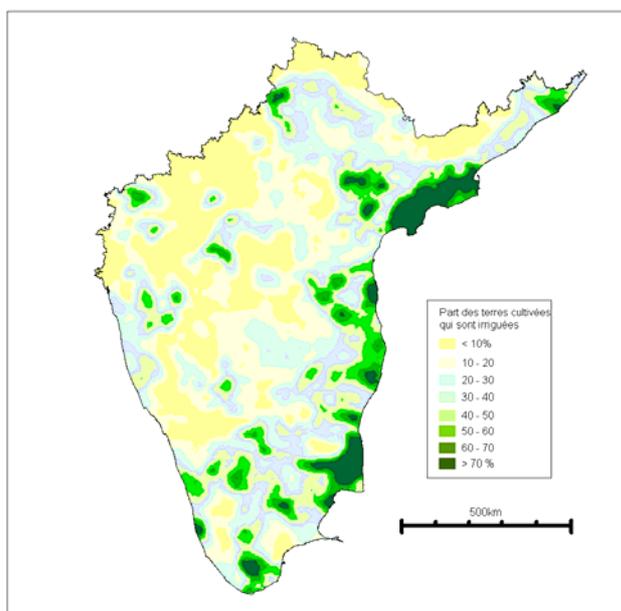


Figure 16 : carte de l'irrigation : proportion de terres cultivées, irriguées

Principales caractéristiques de l'irrigation

La superficie irriguée en Inde du Sud représente, d'après le recensement de 1991, 9 357 600 hectares soit 27 % des terres cultivées. Ce qui signifie que près des trois-quarts des terres cultivées ne sont pas irriguées, alors que la majorité de l'Inde du Sud a un climat semi-aride (Figure 15) d'après la définition de l'ICRISAT dans la mesure où l'évapotranspiration n'est satisfaite par la pluviosité qu'entre 2 à 7 mois par an. Dans ces conditions, l'irrigation est indispensable pour certaines cultures telles que le riz humide ou la canne à sucre (qui requiert de grandes quantités d'eau sur un long cycle cultural), elle sécurise et accroît les rendements des autres cultures et elle est indispensable pour cultiver en saison sèche.

La répartition des terres irriguées est très inégale, comme on peut le voir sur la carte 16 (similaire à la Figure 3 de C.Z. Guilmoto, 2002 et à la carte de l'irrigation de S. Oliveau, 2003) et dans le tableau 5: toute la moitié nord-ouest est plutôt sèche, tandis que la moitié est bénéficie des caractéristiques hydrographiques de cette partie de l'Inde, où les fleuves s'écoulent vers l'est. On y discerne ainsi nettement les deltas de la Godavari, de la Krishna et de la Kaveri.

Tableau 5 : Répartition de l'irrigation selon les Etats

	Superficie irriguée (en hectares)	Part de l'irrigation		Taille de l'Etat au sein de l'Inde du Sud
		par rapport à la superficie cultivée dans chaque Etat	dans l'Etat par rapport à la superficie irriguée de l'Inde du Sud	
Andhra Pradesh	4 255 315	30 %	45 %	40%
Karnataka	1 845 015	16 %	20 %	32%
Kérala	495 046	23 %	5 %	7 %
Tamil Nadu	2 762 224	41 %	29 %	21%

Les secteurs peu ou pas irrigués sont quant à eux localisés dans des régions aux climats forts différents. Certains sont situés au Nord de l'Andhra Pradesh et sur la zone côtière située à l'ouest des Ghâts occidentaux, zones bien arrosées par les pluies en raison du relief et ne nécessitant donc pas d'irrigation. D'autres sont au contraire dans la partie sud du Deccan, très sec. Cette région est étonnante car elle correspond en même temps à l'endroit où la part des superficies cultivées est la plus élevée (carte 17) : il s'agit donc d'un espace entièrement cultivé, mais dépendant exclusivement de la pluie qui est très faible dans le nord du Karnataka (Figure 15). Ces caractéristiques en font une zone d'agriculture pluviale très risquée.

Les quelques poches d'irrigation qui se trouvent dans ce paysage d'agriculture pluviale correspondent à des périmètres irrigués aménagés en aval de barrages. De façon plus générale, les endroits les plus densément irrigués (>40%) sont majoritairement localisés en aval d'un barrage (cf. Figure 18) ou dans les deltas, qui sont exploités et aménagés depuis plusieurs siècles. Peut-on alors dire, comme ces cartes le suggèrent à première vue, que les secteurs densément irrigués sont liés à des investissements dans la grande hydraulique et à une irrigation par canal ? Avant de répondre à cette question, présentons succinctement les différents types d'irrigation.

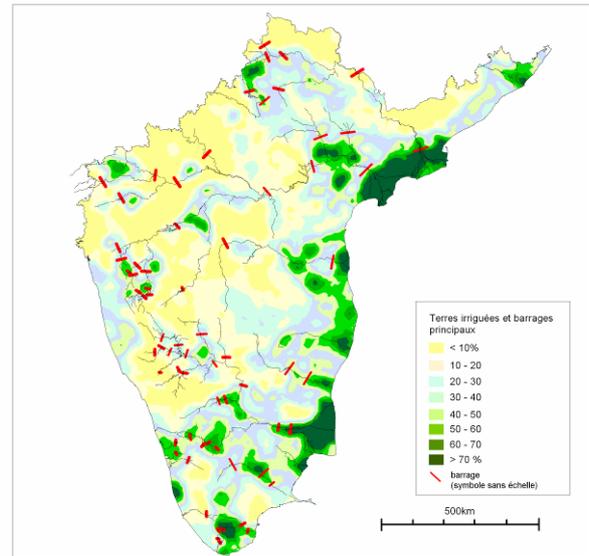
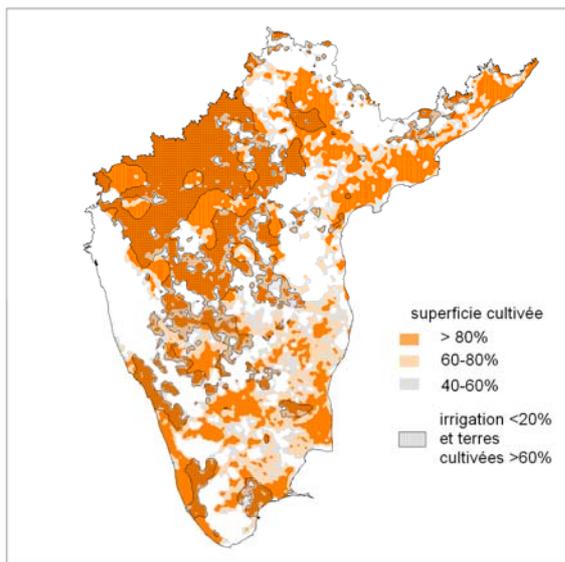


Figure 17 : Mise en évidence des zones fortement cultivées, à agriculture pluviale

Figure 18 : Irrigation et barrages

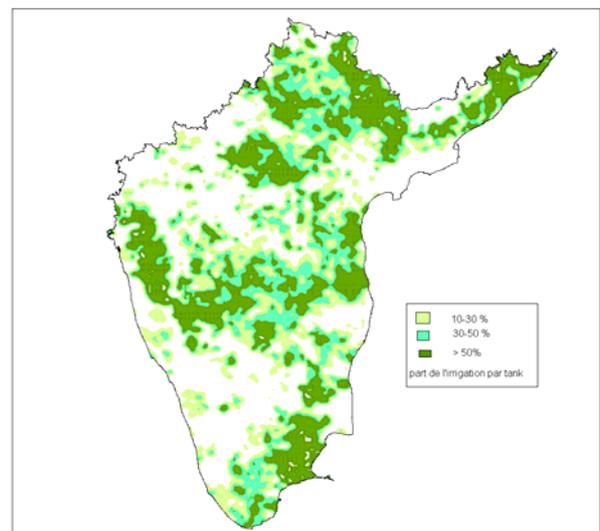
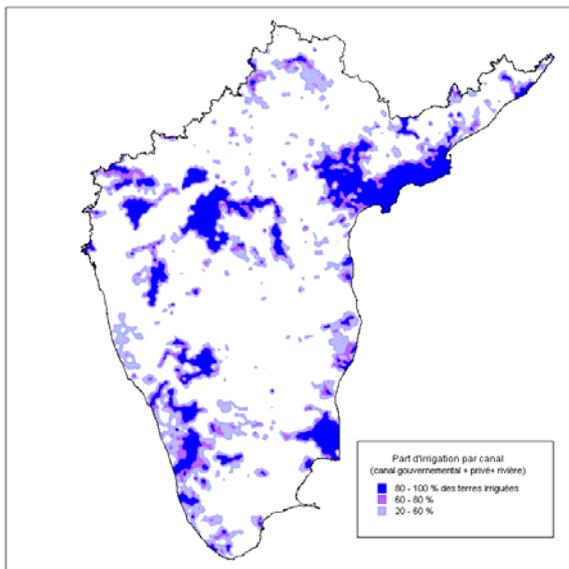


Figure 19 : Répartition spatiale de l'irrigation par canal

Figure 20 : Répartition spatiale de l'irrigation par tank

Différents types d'irrigation

Trois grands types d'irrigation peuvent être distingués en Inde du Sud :

- une irrigation par canal dérivant l'eau de rivière. Ce type est souvent associée aux travaux de grande hydraulique initiés par les Britanniques (barrage et longs canaux de dérivation) ;

- une irrigation par “tank”, ces étangs-réservoirs vieux de plusieurs siècles, très présents en Inde du Sud et remplis d’eau durant 4 à 6 mois de l’année²⁸. Même si certains sont alimentés par une rivière à l’aide des canaux de dérivation, le type d’irrigation retenu est “tank”. En outre, depuis ces tanks, l’eau s’écoule par gravité par des canaux de distribution ;
- une irrigation par pompage dans les nappes phréatiques, à travers des puits ou plus récemment des forages, plus profonds.

L’irrigation par pompage domine aujourd’hui au Tamil Nadu, tandis que l’irrigation par canal est plus importante dans les autres Etats (Fig.1 de Guilmo, 2002), et ces deux types doivent leur essor aux politiques incitatives de la Révolution Verte. Notons que la superficie irriguée par puits et forages est ici sous-estimée puisque le recensement ne mentionne pas les superficies irriguées *nettes*, mais simplement les superficies brutes. Quant à l’irrigation par tank qui a reçu très peu d’attention jusque récemment de la part des politiques de développement, elle a non seulement diminué en part relative mais aussi en superficie nette (-17% entre 1950 et 1986 pour toute l’Inde, Agarwal et al, 2003 : 378).

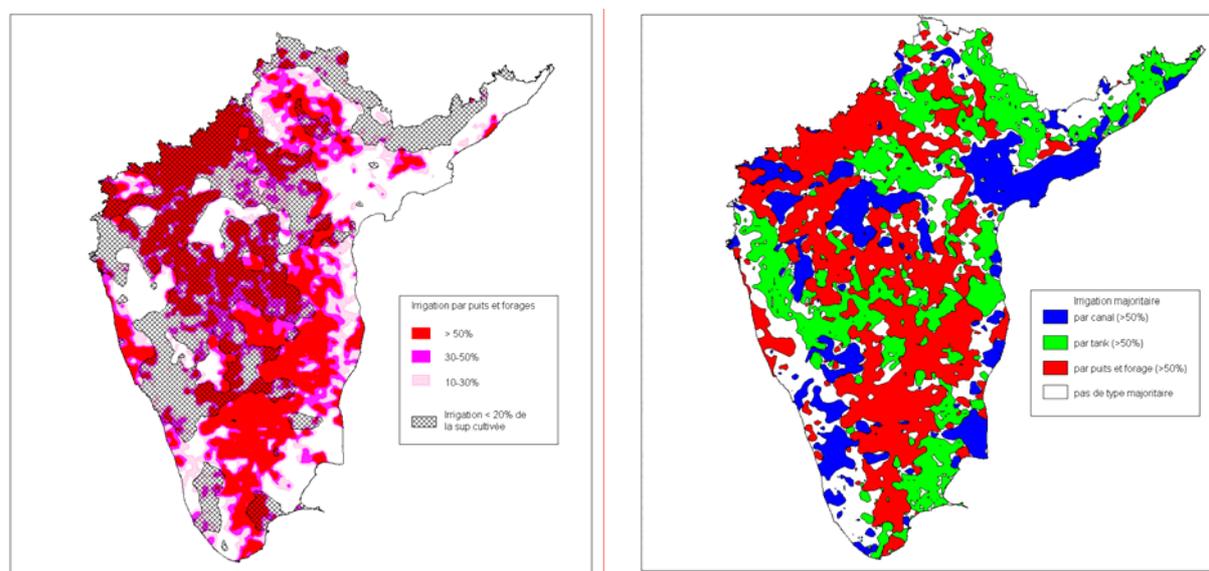


Figure 21 : répartition spatiale de l’irrigation par puits et forage

Figure 22 : Répartition géographique des types d’irrigation majoritaires

Il est hors de propos d’étudier ici les raisons explicatives de la répartition géographique des différents types, plusieurs facteurs d’ordre environnemental (pluviométrie, hydrographie, pédologie, géologie), social et historique devant être intégrés dans une telle analyse. Précisons simplement que cette répartition géographique est relativement concentrée, des zones où domine l’un des types²⁹ se distinguant nettement (cartes 19 à 21). Quelques rares zones de pleine mixité (quand aucun des types n’est supérieur à 50%) se glissent entre les précédentes (carte 22).

²⁸ La superficie qu’ils irriguent est en moyenne de 20 à 50 hectares ; leur propre superficie est à peu près identique et leur nombre varie selon les régions. Ils seraient environ 36 500 au Karnataka par exemple (Agarwal, 2003 : 322).

²⁹ Un type est défini comme majoritaire quand il irrigue plus de 50% de la surface irriguée.

Pour répondre à la question posée sur les secteurs densément irrigués liés à une irrigation par canal, on a décomposé le problème en deux phases. D'une part, on s'est intéressé à la proportion de chaque type d'irrigation dans les zones densément irriguées (> 50%). On a repris la carte 22 sur laquelle les zones fortement irriguées ont été mises en valeur (carte 23). D'autre part, sur la carte des zones irriguées par canal ont été surimprimées les zones densément irriguées (carte 24). On voit sur ces 2 dernières cartes que les zones densément irriguées et les zones où l'irrigation par canal est majoritaire ne coïncident pas exactement : il existe des zones densément irriguées qui ne dépendent pas d'un canal ainsi que des zones où l'irrigation est majoritairement par canal mais elles ne représentent pas une zone fortement irriguée. Les secteurs fortement irrigués ne sont donc pas systématiquement irrigués par canal. Toutefois, une forte corrélation existe (Guilmoto, 2002 :1225) et pour l'ensemble du Karnataka, sur la Godavari au Nord de l'Andhra Pradesh, et dans les deltas de la Kaveri, de la Krishna et de la Godavari, les zones densément irriguées le sont par canal. Ailleurs, la situation est plus complexe : une irrigation ancienne par tank peut dominer; ou une irrigation par puits.

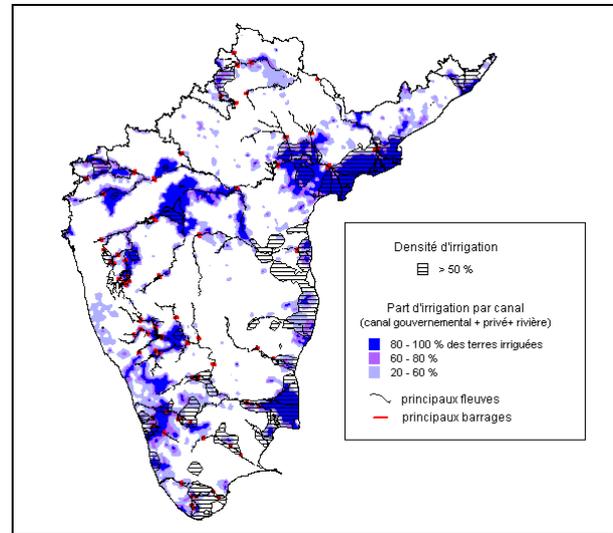
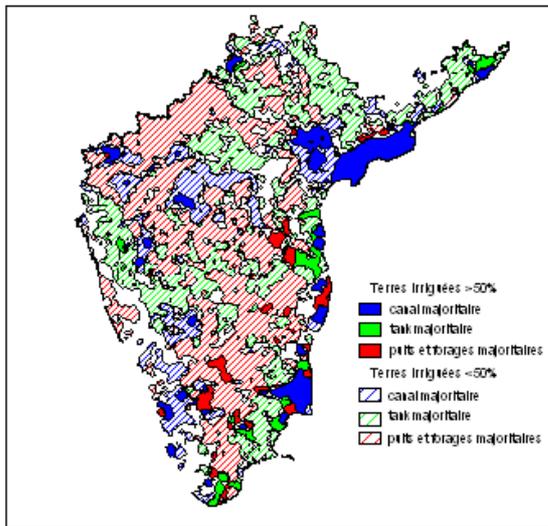


Figure 23 : Type d'irrigation dans les zones à forte irrigation (>50%)

Figure 24 : Irrigation par canal et zones densément irriguées

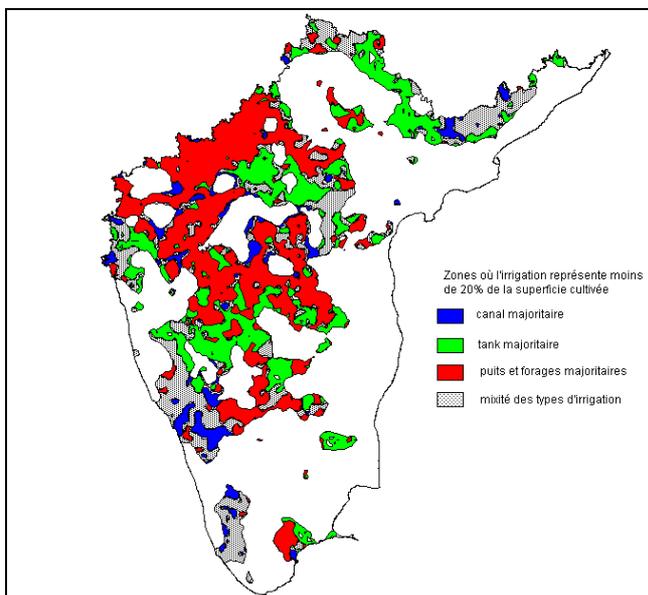


Figure 25 : Types d'irrigation dans les zones faiblement irriguées (irrigation <20%)

La contrepartie d'une forte densité d'irrigation liée majoritairement à une irrigation par canal est que les zones faiblement irriguées concernent peu l'irrigation par canal (carte 25).

Indice du potentiel agricole des terres

L'irrigation permettant d'accroître la production agricole, nous avons choisi de construire un indice du potentiel agricole des terres, afin de caractériser les terres selon un seul critère : la surface, indexée par cet indice. Ainsi une surface irriguée et une surface non irriguée sont ramenées à une même unité, l'indice du potentiel agricole permettant de distinguer les deux.

Pour construire cet indice, il fallait connaître l'impact de l'irrigation sur la productivité des terres. Or, d'après Vaidyanathan (1999: 64-90) qui a effectué une analyse comparée de plusieurs études sur ce thème, cet impact est très variable selon les techniques d'irrigation, les facteurs climatiques, les Etats et les classes de cultivateurs. Globalement, les tanks pluviaux offrent des possibilités limitées en termes de sécurité et de flexibilité de l'ajustement de l'approvisionnement en eau en fonction des besoins des plantes. Les grands barrages, alimentés par une aire d'approvisionnement en eau plus vaste, sont pour cela supérieurs et l'eau souterraine est la meilleure source (Ibid : 69), offrant la meilleure correction temporelle et spatiale pour reprendre les termes de Durand-Dastès. En raison de la complexité et de la diversité des situations, des diverses imprécisions du recensement, nous avons choisi de construire un indice du potentiel agricole très global, tenant simplement compte du fait que la terre soit irriguée ou pas. En outre, généraliser les caractéristiques des différents types est difficile en raison des diversités régionales. Ainsi, le ratio de la productivité des terres irriguées sur celle des terres non irriguées varie de 1,31 (Bihar) à 4,63 (Maharashtra) selon les Etats (Ibid, table 9 p.77). Pour les trois grands Etats du Sud de l'Inde, le ratio est toutefois relativement similaire : 2,97 ; 2,93 et 2,88 respectivement pour le Karnataka, l'Andhra Pradesh et le Tamil Nadu. L'auteur précise que la procédure utilisée sous-estime toutefois la différence de productivité (surestimation pour le sec et sous-estimation pour l'irrigué) pour diverses raisons parmi lesquelles le fait que la productivité des terres a été calculée à partir de la production de céréales, mais certaines cultures très productives sont ainsi exclues ; et une terre irriguée peut accueillir en deuxième culture une céréale non irriguée.

Cette relative homogénéité du ratio entre les trois Etats nous a fait choisir un seul nombre comme multiplicateur global à appliquer à l'irrigation pour notre étude, et en raison de la sous-estimation générale du ratio, nous avons choisi un facteur de 3. On ne prend donc pas en considération la diversité des types d'irrigation, car évaluer les avantages de chaque type et ceux de l'utilisation conjointe de plusieurs techniques se révèle trop difficile à cette échelle d'analyse. On peut éventuellement distinguer les terres irriguées par forages (tubewell) et leur appliquer un coefficient double (soit un facteur de 6 par rapport aux terres non irriguées) en raison de la deuxième culture que ce type d'irrigation assure presque systématiquement (en considérant que l'agriculteur ait accès à la ressource souterraine).

La surface "indexée" obtenue a subi un coefficient multiplicateur variant de 1/3 à un peu plus de 1 : 1/3 pour les terres pluviales ; 1 pour les terres entièrement irriguées par un type d'irrigation autre qu'un forage et un coefficient de 2 pour une superficie entièrement irriguée par forage (ce qui n'existe pas, le coefficient global étant au maximum de 1,1).

Dans le travail qui suit, nous avons choisi de nous limiter dans une première approche à la simple superposition des données pluviométriques et de la répartition de l'irrigation afin de fournir un premier jeu d'indicateurs simples, à la manière des autres indices calculés. Les paragraphes qui précèdent indiquent toutefois la direction de nos travaux actuels visant l'élaboration d'un indice de potentiel qui rapporte directement la disponibilité de la ressource hydrique à la densité de main-d'œuvre agricole dans les zones rurales.

4.4.2. Développement social : démographie, éducation

Les dimensions couvertes par cette rubrique relèvent de la notion globale de développement social, mise en avant par le PNUD depuis de nombreuses années pour compléter l'aspect purement économique du développement stricto sensu. On considère en effet que le renforcement du capital humain est une clé indépendante et indispensable pour l'essor des populations. Ainsi, l'indice de développement humain calculé à l'échelle nationale et parfois régionale dans les pays du monde utilise des indicateurs de bien-être et de « capacités » des populations qui ne relèvent pas de l'économie au sens classique. Il s'agit de la mortalité, la fécondité et du niveau d'instruction.

On rappellera qu'historiquement, l'Inde du sud a joué un rôle particulier dans cette réflexion, car l'anomalie du Kerala a attiré l'attention des économistes depuis les années 1960. Il s'agissait là en effet d'une région connaissant un niveau de développement économique assez médiocre « sur le papier », mais qui bénéficiait pourtant sans conteste des meilleures conditions de vie en Inde. La mortalité y était à son plus bas niveau, frôlant les performances européennes et la grande majorité des femmes y étaient instruites, à la différence du reste de l'Inde. On notait également la bonne position des zones rurales, une baisse de la fécondité accélérée, des avancées très nettes en termes de justice sociale (réforme agraire sans équivalent en Inde) associées autant à la politique éclairée des princes locaux de l'époque coloniale qu'à l'action entreprise par les mouvements communistes régulièrement portés au pouvoir dans l'Etat du Kerala. Cette opposition entre développement social et économique a été reprise plus récemment par Amartya Sen qui, mettant en valeur les différentes voies de développement en Inde, a souligné les particularités de l'Inde du sud face aux Etats du reste de l'Inde, qu'ils soient prospères comme le Punjab ou particulièrement déshérités comme ceux du « Bimaru » (acronyme des Etats pauvres de l'Inde du Nord qui signifie « malade » en hindi).

Les données disponibles dans notre base d'origine permettent un certain nombre de cartographies simples. C'est notamment le cas des dimensions de fécondité et d'alphabétisation qui sont calculables pour chacun de nos clusters. On a choisi de conserver l'indicateur d'alphabétisation pour les deux sexes, réservant l'indicateur de différentiel sexuel pour un calcul ultérieur de discrimination sexuelle.

En revanche, la mesure de la mortalité infantile était plus difficile à obtenir, puisque les estimations les plus fines disponibles concernent les districts d'Inde du sud, unités assez grossières pour procéder à une reconnaissance cartographique détaillée. Nous avons par conséquent une technique de géostatistique relativement complexe, le co-krigeage, combinant à la fois le krigeage ordinaire et l'imputation statistique traditionnelle. Il s'agit d'une technique d'estimation du macro vers le micro, parfaitement adaptée à notre base de données. Les détails de la procédure sont fournis plus bas dans notre section technique.

4.4.3. Urbanisation et accessibilité

Une des dimensions souvent soulignée de la vulnérabilité lorsqu'elle est étudiée dans sa dimension spatiale, est celle de l'enclavement, qui tend à maintenir les habitants d'une zone à l'écart du changement. On sait par ailleurs, que l'innovation est un phénomène avant tout urbain, et que les canaux les plus importants des relations entre villes et campagnes sont les routes (voir Oliveau, 2004 pour un approfondissement sur le Tamil Nadu). Il semblait par conséquent logique de mesurer pour chaque village son degré d'enclavement, c'est-à-dire sa distance à la ville la plus proche et sa distance à l'axe de communication le plus proche. On peut ainsi mesurer pour chaque cluster un enclavement moyen de ses habitants, qui nous permet de proposer une cartographie de l'enclavement à l'échelle du sud de l'Inde (Figure 26). On trouve en bleu les zones les plus proches d'une route et/ou d'une ville et en rouge les zones les plus enclavées.

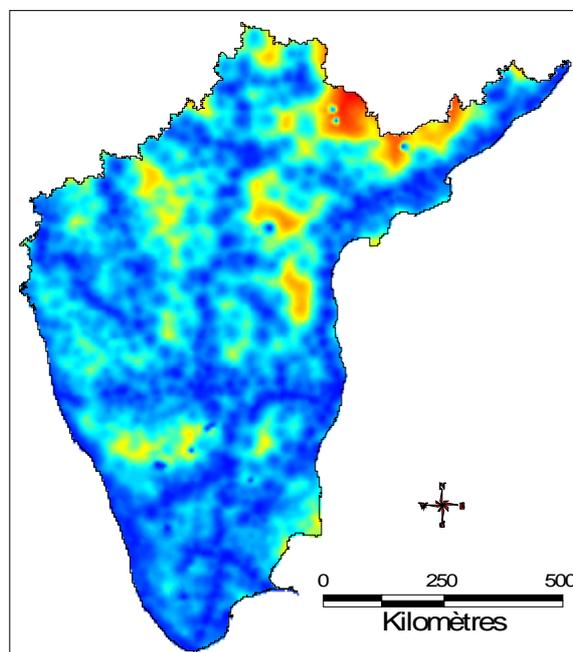


Figure 26 : enclavement des villages

On peut y lire les grands axes qui structurent l'espace sud indien, le long des côtes, mais aussi à l'intérieur, verticalement, le long d'un axe qui relie les villes de l'intérieur (Madurai, Coimbatore, Hyderabad, pour ne citer que les trois plus grandes), et horizontalement : Chennai-Bangalore ou Vishakhapatnam-Hyderabad entre autres.

4.4.4. Infrastructures

Parmi les infrastructures importantes pour comprendre la vulnérabilité, les équipements scolaires et sanitaires sont primordiaux. N'ayant pas d'indicateur à l'échelle individuelle susceptible d'estimer l'offre et le recours à l'éducation et à la santé, nous avons construit deux indicateurs indirects pour chacune de ces dimensions en nous appuyant sur les structures disponibles dans chaque village.

Accès aux soins, oui, mais lequel ?

L'offre de soins en Inde offre une multiplicité de structures, chacune proposant des services proportionnellement développés selon son rang dans l'organisation sanitaire. En outre, la qualité des soins proposés se différencie également selon si le centre de santé relève du secteur public ou du secteur privé. Depuis les années 1970, ce secteur privé a connu une croissance fulgurante et, aujourd'hui, il regroupe plus de 75 % des centres de soins. Deux raisons majeures expliquent qu'il supplante désormais largement le secteur public. Tout d'abord, l'amélioration des conditions de vie s'est accompagnée d'une recherche de soins de qualité. Et, en parallèle, les nombreux dysfonctionnements du secteur public a fait que la proportion de praticiens privés s'est accrue, et ceci aussi bien en ville qu'à la campagne. La tendance générale tend donc à considérer que les praticiens privés sont davantage qualifiés que les médecins du secteur public, du fait notamment de l'efficacité de leurs traitements. Aujourd'hui, malgré l'assise temporelle du secteur public, ses dysfonctionnements le desservent donc sévèrement et c'est particulièrement le cas pour ses structures de proximité (les centres de soins primaires et les centres de soins secondaires). N'étant pas à même de répondre aux besoins des populations, les plaintes à son égard ne cessent d'affluer. Les localisations inadéquates, les difficultés d'accès, le manque de maintenance, le sous-effectif de personnel qualifié, le manque de médicaments, l'absence de système de référence satisfaisant, l'absence de spécialistes ou bien encore l'inexistence de laboratoire d'analyses sont parmi les dysfonctionnements les plus souvent avancés. Structures les plus proches des populations, elles restent donc paradoxalement peu utilisées car elles ne satisfont pas leurs besoins réels.

Ainsi, de manière générale, en ville comme à la campagne, les structures de soins privées sont les plus convoitées, puisqu'elles symbolisent un recours de qualité. Alors, lorsqu'une pathologie est ressentie comme sérieuse, le patient se dirigera de préférence vers ce type de structure ou bien vers les grands centres publics urbains. Ensuite, la qualité des services étant confondue avec le poids du lieu, si le malade l'estime nécessaire, la distance sera acceptée, voire même convoitée, par les ruraux notamment, puisqu'elle se fait garante d'un recours efficace. Par contre, les pathologies ressenties comme anodines, donneront lieu à un recours de proximité, dans des centres globalement de moins bonne qualité et surtout qui implique un coût moindre pour le malade. Les représentations concourant à l'utilisation des différents centres de soins, sont donc peu favorables à la proximité et, à l'inverse, invitent d'abord à la mobilité. On voit donc bien que ce n'est pas la présence d'une structure de soins qui garantit son utilisation mais son efficacité (Chasles 2004).

Le nombre de structures de soins distinguées par le recensement est assez important, puisque l'on dénombre 15 types d'établissements. Malheureusement, rien n'est indiqué sur leur équipement, leur taille et la présence effective de personnels. Nous nous sommes donc appuyés sur nos connaissances locales des structures de soins, pour proposer une typologie des établissements. Au final, on a distingué les établissements de qualité (hôpitaux et centres de soins communautaires), où les recours possibles sont multiples (depuis la consultation jusqu'à l'opération) et effectifs (médecins et personnels médicaux y sont effectivement présents) des autres établissements tels que les centres de soins primaires et secondaires. Nous avons ensuite calculé la distance des villages à ces établissements³⁰ afin d'obtenir une distance moyenne des habitants compris dans un cluster à un centre de soin de qualité. Dans un second temps, nous avons calculé la distance moyenne des habitants à l'ensemble des structures de soins, considérant qu'il fallait envisager les autres structures comme offrant potentiellement une possibilité de soins, ce qui est d'ailleurs le cas. Pour cela, nous avons

³⁰ En l'absence d'information, il a été considéré que les villes étaient toutes équipées de centres de soins de qualité, ce qui est globalement vrai.

distingué 4 catégories de centres, depuis les centres de qualité, jusqu'aux dispensaires, qui constituent les plus petites unités du réseau sanitaire. La cartographie de chacune de ces variables nous permettant donc de mettre en lumière les régions qui, d'un point de vue sanitaire, sont vulnérables du fait non seulement de l'absence de centres de soins mais surtout de leur éloignement aux performances sanitaires.

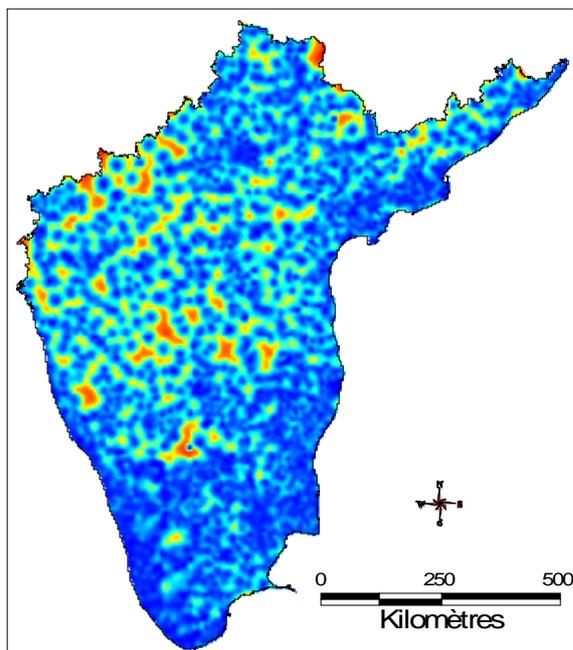


Figure 27 : distance aux centres de soins de qualité

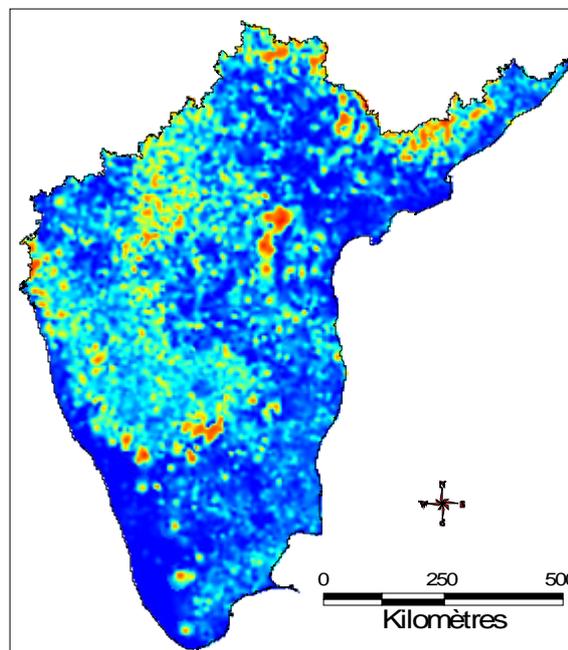


Figure 28 : distance aux centres de soins

Alphabétisation et éducation, une nuance de taille.

Au moment de l'Indépendance, 18 % seulement de la population était alphabétisée, c'est pourquoi la nouvelle Constitution fit de l'éducation l'une de ses priorités majeures. Au cours des cinquante dernières années, le Gouvernement s'attacha alors à faire de l'éducation, primaire tout particulièrement, un service universel et accessible à tous, impulsé notamment en 1986 par le *National Policy on Education*. D'ailleurs, à l'échelle de l'Inde, si l'on prend le seul cas des écoles primaires, en 1991, 95 % de la population rurale avaient accès à ce type d'établissement dans un rayon de 1 km³¹. Mais, malgré des progrès certains³² de l'accès à l'éducation, mesurés à partir des taux d'alphabétisation des plus de 5 ans, il n'en demeure pas moins que ces taux restent encore à améliorer et surtout continuent d'être hétérogènes au sein du territoire, de l'Inde du Sud notamment.

En réalité, dans ce cas également, la présence d'un service ne garantit pas son utilisation et ne suffit pas, si on prend le cas des écoles primaires, à garantir des taux d'alphabétisation optimum. En effet, non seulement, la présence d'une école primaire ne garantit pas que les élèves vont y suivre un cursus complet mais, qui plus est, les taux d'alphabétisation ne sont pas sensibles uniquement aux services de proximité. Plus précisément, il apparaît que la distance aux structures scolaires supérieures se révèle tout à fait discriminante et participe

³¹ Notons simplement qu'il s'agit là de la distance maximale, déterminée par le Gouvernement, qui doit séparer les villages des écoles primaires. Dans le cas des écoles de niveau supérieur (collège, lycée ...), une distance de 3 km des villages comptant au moins 300 habitants est tolérée.

³² Le taux d'alphabétisation des plus de 5 ans était de 18 % en 1950-1951, 28 % en 1960-1961, 34 % en 1970-1971, 43 % en 1980-1981, 52 % en 1990-91 et 65 % en 2000-2001.

également à l'explication des taux d'alphabétisation. Autrement dit, cette variable se fait l'écho de discriminations sociales, mais pas seulement, puisqu'elle comporte des résonances spatiales originales. En effet, il s'avère que la spatialisation des écoles supérieures et, plus précisément de la distance qui les sépare de la population, affiche de la même manière un caractère discriminant majeur. Cela signifie simplement, qu'un village sans écoles enregistrera un taux d'alphabétisation plus faible qu'un village comptant au moins une école, dont une école primaire tout particulièrement. Ensuite, la distance à une structure scolaire de niveau supérieur permet d'affiner ces explications. En effet, un village comptant une école mais qui est loin d'une université, enregistrera un taux d'alphabétisation moindre par rapport à un village sans école mais proche d'une université. Ainsi, contre toute attente, la seule présence d'une école dans le village ne suffit pas à garantir des taux d'alphabétisation optimum.

Pour le système éducatif, le recensement procède avec le même détail que pour le système sanitaire. On peut ainsi distinguer les établissements primaires, des collèges, lycées et premiers cycles universitaires. On peut donc mettre en parallèle l'accès à l'alphabétisation (distance moyenne à l'école primaire) de l'accès à l'éducation (distance moyenne aux quatre types d'établissements).

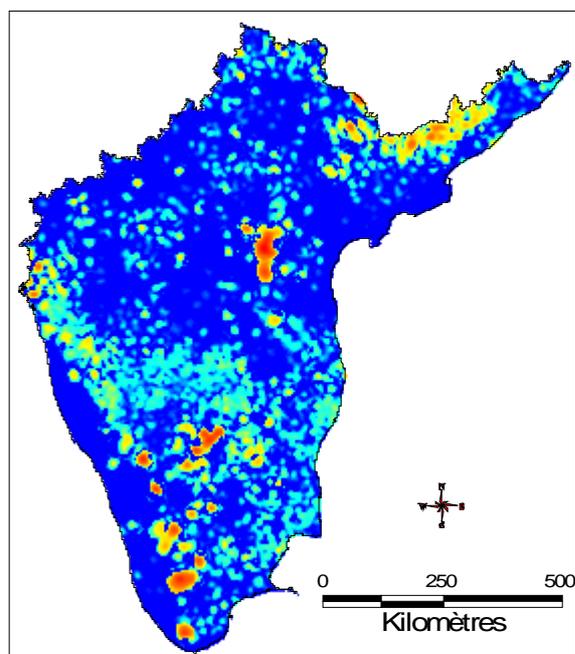


Figure 29 distance à une école primaire

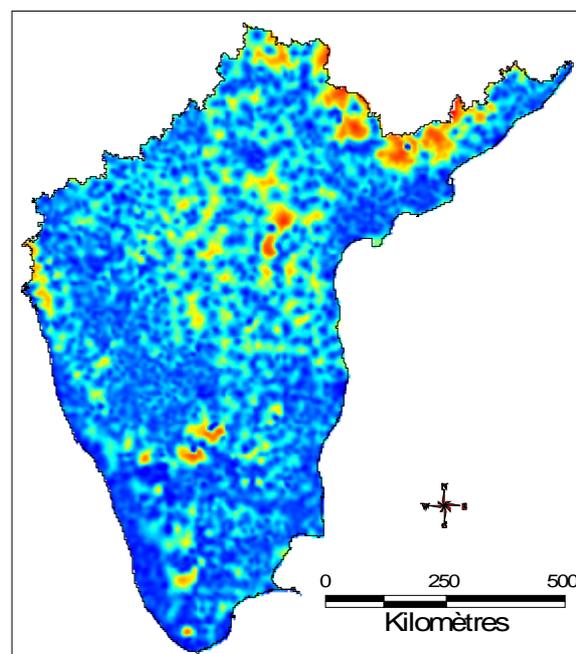


Figure 30 : distance à l'éducation

La cartographie de la distance à l'école primaire fait ressortir les zones les plus vulnérables du point de vue de l'alphabétisation : sans école primaire, ce premier stade éducatif est évidemment plus difficile à réaliser. La carte suivante nuance le propos particulièrement pour les espaces bien équipés en écoles : si l'école primaire marque la possibilité d'accès à l'alphabétisation, l'éducation n'est possible qu'en ayant un accès à l'ensemble du système éducatif.

4.4.5. Genre et groupes sociaux

Les dimensions socioculturelles ont un rôle dans la distribution de la vulnérabilité qu'il est malaisé de circonscrire. Elles sont parfois manifestes, dans la nature des phénomènes

examinés. Mais elles sont la plupart du temps latentes, se traduisant par une association régulière entre pauvreté et caractérisation socioculturelle (par exemple ethnique, religieuse, etc.), sans que l'on ne puisse établir avec précision le mécanisme discriminatoire qui lie la vulnérabilité et l'appartenance de groupe.

En Inde, les déséquilibres de genre dans l'enfance sont un exemple du premier type, car le lien entre les déficits de petites filles et les pratiques discriminatoires est connu : il s'agit de la sélection sexuelle des enfants, phénomène relativement similaire à la maîtrise de la fécondité par ses objectifs (décider de sa descendance) et ses méthodes (l'avortement par exemple). Les déséquilibres de sex-ratio dans l'enfance en sont la trace indubitable et permettent de quantifier précisément l'intensité discriminatoire vis-à-vis des petites filles. Notre cartographie du sex-ratio est donc le reflet direct des discriminations de genre et de leur très inégale distribution dans le paysage sud-indien.

Cet indicateur qui désigne des poches régionales très particulières de l'Inde du sud (et plus précisément du Tamil Nadu) sera complété ultérieurement par d'autres indicateurs de discriminations de genre : discrimination éducative notamment, mais également différentiels sexuels dans les taux d'activité qui marquent la forte marginalisation féminine en de nombreuses régions rurales.

D'autres dimensions sociales étroitement liées aux mécanismes discriminatoires existent qui relèvent plutôt du second type : il s'agit de l'appartenance à des groupes minoritaires dont le profil socioéconomique est systématiquement défavorisé. Le système de caste transcrit en Inde une bonne part des institutions de discrimination traditionnelle et l'outil statistique permet notamment d'identifier deux groupes contre lesquelles les discriminations sont accentuées : les Dalits (alias ex-intouchables, *scheduled castes*, Harijans, hors-caste) et les tribaux (alias *scheduled tribes*, Girijans, Adivasis, populations indigènes, etc.). Sans rentrer dans une description détaillée, notons d'emblée que chacun de ces deux groupes est fort loin d'être homogène, car constitué d'une myriade de communautés distinctes (sous-caste locales, tribus, etc.) qui ne se reconnaissent guère d'identité commune, sinon sous l'effet d'une mobilisation politique relativement récente et encore très segmentaire. Toutefois, la relative homogénéité dans leur marginalisation économique (et spatiale) donne à ces catégories sociologiquement composites une identité particulière dans l'analyse des groupes vulnérables en Inde. D'autres groupes minoritaires, telles que les basses castes ou les groupes non hindous, pourraient être également en position de discrimination systématique. Néanmoins, nous n'examinerons pas leur cas pour deux raisons : ils n'apparaissent pas être les plus défavorisés de l'ensemble (comparativement aux Dalits et tribaux) et surtout, ils sont très imparfaitement identifiables par la statistique locale et ne peuvent donc à présent être intégrés à notre SIG.

S'agissant de dimensions latentes, nous réfléchissons plus loin à la nécessité d'intégrer a priori ces dimensions d'appartenance sociale (aux Dalits ou aux tribaux) comme marqueurs spécifiques de pauvreté, après une réflexion brève sur les caractéristiques de la distribution spatiale propre à ces deux groupes. Il peut sembler en effet difficile de poser a priori l'appartenance à ces groupes comme un signe de vulnérabilité, mais l'universalité de leur relégation socioéconomique en Inde conduit à un réexamen du rapport entre statut et condition socioéconomique. La mise en place depuis l'indépendance d'une politique de discrimination positive systématique en Inde vis-à-vis de ces deux catégories (en langage bureaucratique : les *scheduled castes* -SC- et *scheduled tribes* -ST) dans une gamme très large de domaines – éducation, emploi public, représentation politique- atteste de la validité

de ces notions anthropologiquement floues comme marqueurs sociologiques puissants de la stratification sociale indienne.

4.5. Synthèse : carte de la vulnérabilité

4.5.1. Opérationnalisation des mesures

Les différentes dimensions de la vulnérabilité rassemblées se déclinent en 14 indicateurs. Douze seront directement pris en compte dans notre cartographie. Les indicateurs d'appartenance aux groupes minoritaires évoqués plus haut ne sont pas utilisés dans ce calcul synthétique initial.

Tableau 6 : Dimensions et critères de vulnérabilité

#	variables	Description	Source	Variable	Critère de vulnérabilité
1	Pluviométrie	Précipitations annuelles en mm	Atlas of India	pluvio	< 75 mm
2	Irrigation	% de terres cultivées irriguées	Census	irrig	< 20%
3	Fécondité	Rapport enfants-femme	Census	CWR	>0.5
4	Alphabétisation	% de personnes alphabétisées (7 ans et +)	Census	littot	< 50%
5	Mortalité infanto-juvénile	Taux de mortalité avant 5 ans (rural)	à partir d'estimations par district	q5rt	>100 p. 1000
6	Proximité urbaine	Proximité urbaine	à partir du Census	Distville01	>5
7	Infrastructures scolaires	Indice d'infrastructure scolaire	à partir du Census	D_ecole_moy01	>2.5
8	Infrastructures sanitaires	Indice d'infrastructure sanitaire	à partir du Census	Distsoinsequal01	>7
9	Accès à la route	Accès à la route	à partir du Census	D_acces01	>1
10	Téléphone	% de villages dotés du téléphone	à partir du Census	ph_fac	<50%
11	Secteur agricole	Secteur agricole en % de la main d'œuvre	Census	wagric	>66%
12	Dalit	% de Dalits dans la population totale	Census	dalit	>20%
13	Tribus	% de Tribus dans la population totale	Census	tribes	>20%
14	Discrimination de genre	Sex ratio des garçons pour 100 filles 0-6 ans	Census	sr06	>110 (f/g)

Le tableau qui précède fournit une définition rapide des indicateurs individuels, de leurs mode et unité de calcul. On fournit une définition rapide de leur mode et de leur unité de calcul, ainsi que de leurs sources. On notera qu'ils s'expriment de manière très différente : pourcentages, mesures directes, indices factoriels, etc. Dans la colonne de droite du tableau, nous avons indiqué un seuil de vulnérabilité pour chaque variable, qui sera utilisé dans la mesure synthétique. Certains de ces critères proviennent directement de mesures communément admises (aridité, distance, alphabétisation, etc.), alors que d'autres ont été

déterminés après examen de la distribution des variables en Inde du sud. Comme on l'indique plus bas, ces niveaux peuvent être recalculés selon un autre jeu d'hypothèses.

4.5.2. Synthétiser les dimensions de la vulnérabilité

Les différentes mesures de vulnérabilité doivent être à présent combinées pour présenter un indice récapitulatif des situations locales. L'objectif est en effet de fournir ici un premier indicateur unique, même s'il est possible dans un second temps de procéder à un examen désagrégé de tous les facteurs individuels.

Plusieurs stratégies existent pour le calcul de cet indice synthétique, mais nous décrivons en premier lieu la méthode retenue ici en raison de sa simplicité. Pour chaque dimension de la vulnérabilité, nous avons déterminé un seuil de risque au delà duquel la localité est considérée comme vulnérable (voir le tableau précédent). On a par conséquent transformé un indicateur continu en un indicateur dichotomique. Pour donner un exemple, on a retenu un niveau d'alphabétisation inférieur à 50% comme seuil de vulnérabilité ; toutes les localités dont moins de la moitié de la population sait lire et écrire sont identifiées comme vulnérables selon cette dimension. Le seuil fixé (voir section précédente) est obtenu par des considérations extérieures sur les systèmes de production et les caractéristiques socioéconomiques des zones rurales. Ces n différents indicateurs dichotomiques sont ensuite additionnés pour obtenir l'indicateur global, variant de 0 à n . Sur l'espace considéré, on obtient alors une distribution ordonnée, quasi-continue, de l'indicateur synthétique de vulnérabilité (carte présentée et commentée plus loin).

Cette approche présente différents avantages. Elle est relativement simple à mettre en œuvre, car elle ne nécessite par exemple pas de décisions délicates sur la pondération des indicateurs. Elle considère en outre que tous les handicaps capturés par nos indicateurs ont un poids équivalent, hypothèse en apparence discutable, mais qui conduit à une approche cumulative : chaque handicap joue un rôle indépendant des autres, se rajoutant à la somme des handicaps sans corriger l'effet d'une autre dimension de vulnérabilité. D'un point de vue statistique, elle débouche sur une distribution quasi-continue, alors que le système de pondération peut déboucher sur des répartitions plus accidentées. Toutefois, cette approche repose directement sur le mode de fixation des seuils, à la manière de la mesure traditionnelle de la pauvreté, et doit par conséquent s'appuyer sur une réflexion fine des niveaux à retenir pour chaque indicateur.

Notons que des approches statistiquement plus sophistiquées consistent à faire varier à la fois la pondération et la valeur des indicateurs. Ainsi, on peut regrouper les indices relevant du même domaine afin de mieux pondérer l'ensemble. On considérera par exemple différents domaines suivants :

- Milieu et développement agricole
- Démographie
- Instruction
- Accessibilité
- Infrastructure
- Diversification économique
- Discrimination de genre

Chaque domaine comprend entre un et trois indicateurs distincts. On dispose ainsi de trois mesures apparentées à l'accessibilité : infrastructure téléphonique, proximité urbaine et accès routier. Une moyenne de ces trois mesures peut être calculée, afin de donner un poids unitaire

au domaine « accessibilité » comparable à celui des six autres domaines, décrits eux-mêmes par un nombre variables d'indicateurs. L'indicateur final prendra donc des valeurs plus réduites, allant de 0 à 7, mais avec des valeurs décimales plus dispersées (effet du calcul de moyenne).

La valeur attribuée à chaque indicateur peut également être mieux calculée. Au lieu de s'en tenir à un indicateur dichotomique, on peut partir de l'indicateur d'origine et le standardiser, afin de prendre en compte *l'intensité* de la vulnérabilité, qui a été entièrement négligée dans le calcul dichotomique. On distingue ainsi les localités par degré de vulnérabilité. Cette standardisation pose toutefois de nouveaux problèmes. En premier lieu, le mode de standardisation à des fins comparatives dépend de la distribution de chaque variable, et notamment de la présence de valeurs extrêmes voire atypiques qui pourraient faire basculer l'indice standardisé. Ceci nécessite donc une étude statistique préalable pour chaque variable et des corrections ad hoc. Une méthode automatique de correction par la moyenne et l'écart type ne semble donc pas indiquée. Ajoutons en plus que la moyenne de chaque indice ne désigne pas nécessairement le seuil de vulnérabilité, divisant en deux l'échantillon en zones favorables ou défavorables. Il se peut que pour certaines mesures, le niveau moyen en Inde du sud rurale soit globalement favorable (ou défavorable) et que l'écart à la moyenne ou à la médiane ne soit pas la bonne mesure. Cette difficulté réapparaît plus bas dans notre discussion.

En second lieu, on peut se demander si des conditions favorables pour certaines dimensions doivent compenser les handicaps découlant d'autres dimensions : une forte alphabétisation dans une localité doit-elle compenser l'enclavement ? La combinaison d'indices standardisés tend précisément à créer ces phénomènes de « rattrapage » d'un indice par un autre, qui ne sont pas forcément acceptables d'un point de vue théorique. Plus précisément, on peut en outre considérer qu'il existe de fait un seuil au-delà duquel la mesure de la vulnérabilité est sans intérêt. Un cas typique est fourni par le sex-ratio enfantin. A un niveau « normal », situé en dessous de 105-110 garçons par fille, les différences ne présentent aucun intérêt (ou marqueraient au contraire une discrimination vis-à-vis des garçons !). De la même façon, un seuil d'irrigation supérieur fixé par exemple à 50 % des terres détermine des terroirs très avantagés et les différences au dessus de ce niveau vraisemblablement de moindre intérêt pour notre mesure de vulnérabilité.

Ces considérations suggèrent donc que diverses décisions doivent être prises pour la préparation de l'indice synthétique. Pour chaque indicateur d'origine, on doit éventuellement :

- Réduire l'intervalle de variation (ignorant les valeurs au-delà d'un seuil)
- Corriger arithmétiquement l'indicateur (logarithme, centrage sur la moyenne, etc.)
- Fixer une pondération au sein d'un même domaine

Il s'agit donc de calculs beaucoup plus fins qui sont dépendants toutefois d'un jeu particulièrement complexe d'hypothèses sur l'étendue des variations prises en compte, la pondération des écarts à la moyenne et la pondération des différents indicateurs entre eux. Il nous a semblé raisonnable à cette étape de notre travail de reporter le calibrage de notre mesure synthétique par un jeu de pondération et de standardisation à plus tard et de fournir dès à présent un calcul statistiquement robuste.

4.5.3. Résultat de l'estimation des niveaux de vulnérabilité

La première carte présentée dérive de l'application de la formule cumulative décrite plus haut. Les douze indices retenus sont en premier transformés en valeur dichotomique 0/1, puis additionnés pour obtenir l'indice synthétique. Les valeurs étalées théoriquement entre 0 et 12 sont en réalité plus resserrées en Inde du sud comme le montre la Figure 31.

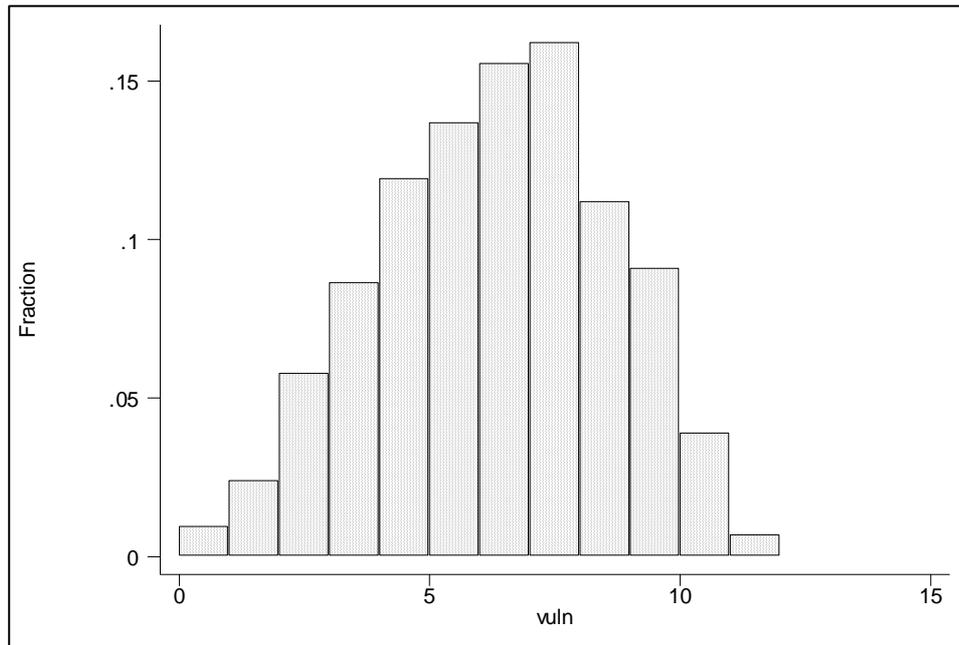


Figure 31 : répartition par niveau de vulnérabilité (clusters 2 km)

La première carte obtenue (Figure 32) résulte donc la composition algébrique de différentes cartes « raster » par simple addition des valeurs de chaque couche. Elle fait apparaître une structuration spatiale globale, orientée notamment vers le centre du Deccan qui apparaît comme la région la plus défavorisée. Une poche relativement compacte s'étend le long de la frontière entre Karnataka et Andhra Pradesh. Des poches plus éparées de forte vulnérabilité apparaissent dans des régions tribales, à commencer par la bordure au Nord-Est des Etats du Chhattisgarh et de l'Orissa. Cette situation résulte de la combinaison de plusieurs facteurs régionaux à caractère très divers. En premier lieu, il s'agit des terroirs les plus arides d'Inde du sud et de ce fait, les moins prospères. L'absence de fleuves ou de dispositif d'irrigation, à l'exception des retenues sur la haute Tambraparni, prive les villages du centre du Deccan d'irrigation efficace et les retenues naturelles (tanks) ne collectent pas assez d'eau pluviale. C'est la région marquée par les effets les plus dramatiques des défaillances de la mousson. Mais le développement social de la région est également à la traîne, avec des taux d'alphabétisation parmi les plus bas du pays et des niveaux de fécondité au contraire plus élevés que partout ailleurs en Inde du sud. La mortalité infanto-juvénile y est également très médiocre comme le montrent les cartes de l'annexe II. Le tissu urbain est très éparé dans le Deccan et entre les grandes métropoles que sont Bangalore au sud-ouest et Hyderabad à l'est (enclave bleue sur notre carte), la région ne possède aucune grande ville. Ces caractéristiques se combinent donc pour donner à notre indice synthétique une valeur souvent supérieure à 9, largement au dessus de la moyenne de la région.

Les autres régions les plus vulnérables d'après notre indice sont les plus isolées et il n'est guère étonnant de constater qu'elles correspondent en général à des paysages boisés à forte densité tribale, même si ces deux caractéristiques ne rentrent pas dans la construction de notre indicateur. Leur situation interstitielle (entre l'Andhra Pradesh et l'Orissa, ou encore entre le Tamil Nadu et le Karnataka) découle avant tout de leur relief et couvert forestier inhospitalier qui en ont les frontières de peuplement depuis des siècles, ménageant ainsi un sanctuaire pour les populations tribales face à la pression des sociétés des plaines.

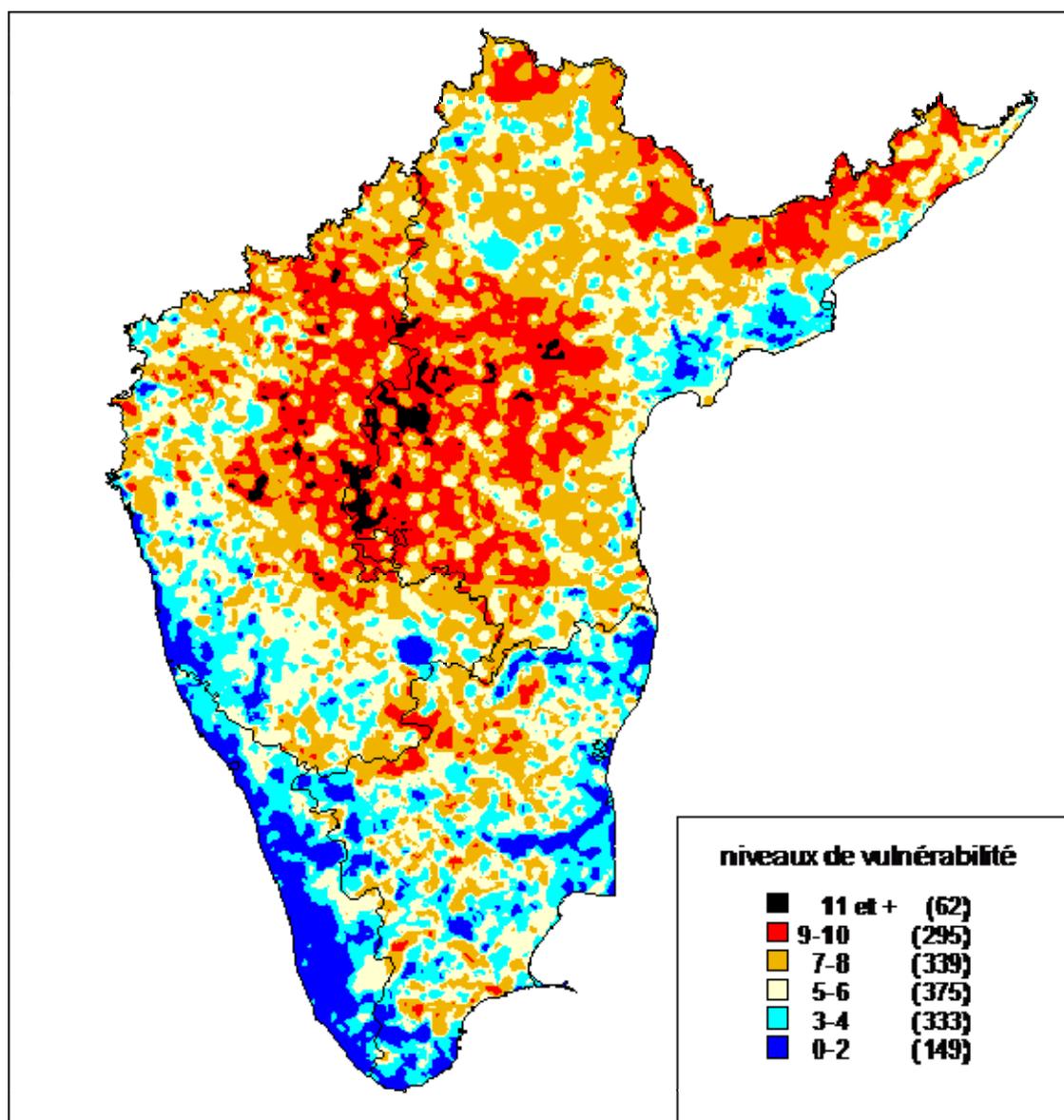


Figure 32 : Carte des niveaux de vulnérabilité rurale

A l'inverse, on notera la position relativement périphérique des régions les plus avantageées, qui appartiennent pour l'essentiel au pourtour littoral de la péninsule sud-indienne. On relève notamment l'ensemble du littoral occidental du Konkan, s'étendant de Goa (qui ne figure dans notre ensemble sud-indien) au nord à la pointe méridionale vers l'extrémité sud qui marque la frontière entre Kérala et Tamil Nadu. Les valeurs estimées au Kérala sont particulièrement élevées, mais tendent à s'infléchir le long des Ghâts quand l'on remonte vers

le nord-ouest. Le lien négatif entre notre indicateur et le développement social (degré d'instruction, comportements démographiques) est particulièrement manifeste sur la côté occidentale.

Le reste du littoral favorisé fait face à la Baie du Bengale, mais il constitue plus une succession de poches distinctes qu'un ensemble homogène. On y lit les effets de l'urbanisation (comme autour de Chennai) ou la présence de zones deltaïques à très forte productivité agricole (embouchures de la Kaveri, de la Krishna et de la Godavari). Quelques régions urbaines intérieures comme autour de Bangalore ou de Coimbatore déterminent également des zones « rurales » (mais en fait assez urbanisées et industrialisées) privilégiées qui se détachent en bleu sur notre carte.

On notera également que le niveau de variations micro-régionales reste accentué à échelle réduite, en dépit de la présence des grandes formations régionales que nous venons de décrire. A une échelle plus fine, les variations restent en effet prononcées. Certains éléments de notre indicateur de vulnérabilité sont en effet très variables à micro-échelle et c'est par exemple le cas des proximités urbaines (ou routières) et des éléments de développement économique qu'elles induisent via par exemple les activités du secteur non agricole. Ces variations donnent à notre carte cet aspect tacheté et expliquent la présence de petite oasis de développement, souvent périurbaines, dans des régions par ailleurs défavorisées.

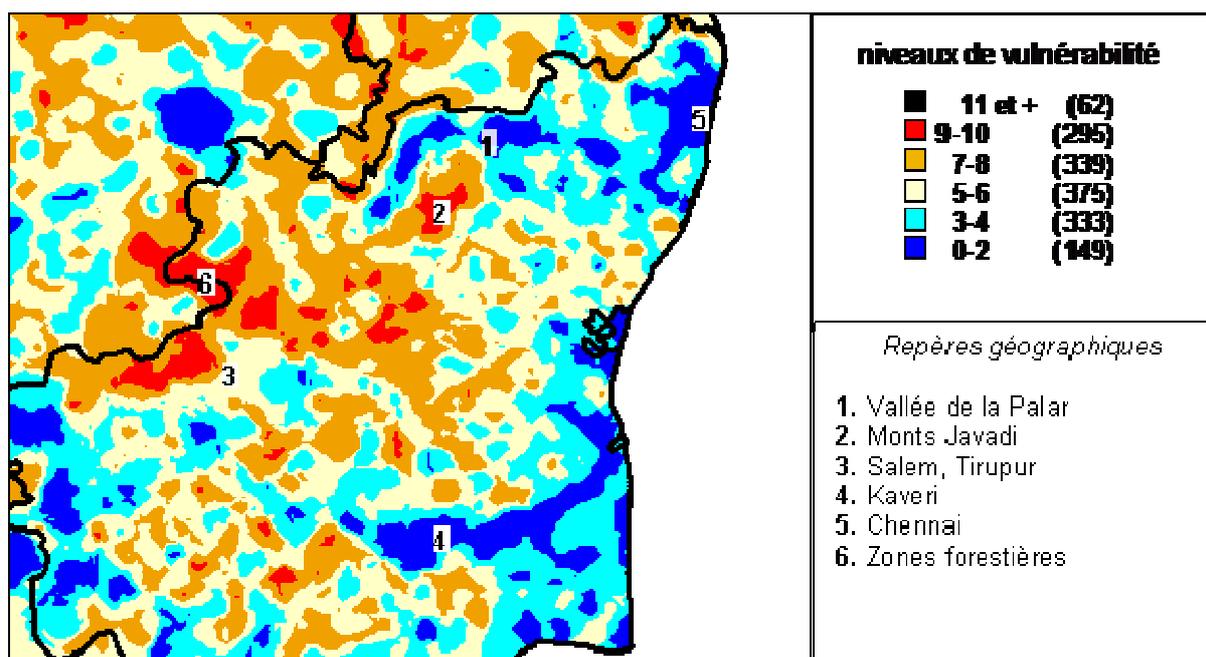


Figure 33 : Carte de vulnérabilité rurale au nord du Tamil Nadu

C'est sans doute un des avantages du mode de construction de notre indice synthétique que d'être capable de restituer le « grain » des variations micro de pauvreté. A la différence de l'indice factorielle présenté plus haut, qui tend à lisser les variations locales, l'addition de facteurs distincts conserve à chaque dimension de la vulnérabilité son rôle dans l'indice final. La qualité de la carte permet de surcroît, en des régions données, de se risquer à des analyses géographiquement plus fines et de tester ainsi la vraisemblance de cet indice global.

Ainsi, dans la moitié nord du Tamil Nadu³³ représentée sur la Figure 32, la trace de la Vallée de la Palar qui s'étend vers l'est de Chennai est très visible : elle oppose les villes industrielles situées le long du fleuve aux zones périphériques beaucoup moins prospères, telle que les monts Javadi au sud-est (Oliveau 1998). On distingue aussi plus au sud de cette région la vallée de la Kaveri, le long de laquelle le développement agricole est favorisé par l'irrigation fluviale et dont l'infrastructure urbaine est plus dense qu'ailleurs.

Ailleurs, les terres sont souvent moins fertiles et l'irrigation plus rare et les pôles de développement m'émergent que dans les environs de certaines villes (comme Salem ou Tirupur au sud-ouest) dont l'essor a facilité le développement industriel local avec de nombreuses répercussions sur le développement villageois. On notera en revanche que la trace de la discrimination féminine, particulièrement accentué aux alentours de Salem en raison des pratiques anciennes d'infanticide féminin, a disparu de la carte présentée ici, car la discrimination représente une dimension singulière dans l'indice synthétique, irréductibles aux autres composantes de la mesure de la vulnérabilité.

4.6. Epilogue

Le dernier exemple commenté tend à confirmer la qualité de l'indicateur synthétique que nous avons utilisé. Les résultats semblent en cohérence avec une connaissance plus fine du terrain. Ils permettent en outre de dépasser les mesures intuitives dérivées d'une connaissance qui serait avant tout locale tout en échappant aux généralisations des statistiques agrégées à des échelles supérieures. Il s'agit sans nul doute un instrument d'évaluation plus fiable et transposable de région à région. Mais il ne s'agit ici sans doute qu'une première exploration et les derniers paragraphes se doivent de rappeler les limites de notre exercice.

Le domaine concerné (celui du poverty mapping) est encore peu exploré en Inde. Les travaux disponibles pour les exercices d'identification des zones vulnérables reposent au mieux sur l'échelle des districts, voire des états, et n'aboutissent qu'à des indicateurs plutôt frustrés, par rapport des variations micro-régionales des inégalités que nous voulions saisir. En outre, les indicateurs proprement économiques de la pauvreté, mesurée par exemple à partir du revenu moyen des ménages ou de leurs biens, qui complètent d'ordinaire la mesure statistique dérivée d'un recensement sont rarement disponibles à une échelle inférieure à celle de l'Etat ; cela empêche la validation des indicateurs indirects que nous utilisons ici. Il a été question à de nombreuses reprises de l'insuffisance de notre base statistique. Le problème est récurrent et l'intégration de données environnementales et sociales est encore très rare (voir toutefois O'Brien *et al.* 2004). Le problème complémentaire est la difficulté de joindre analyse globale et données de terrain par ménage, comme cela est courant dans le domaine, ou d'évaluer la perte de qualité des données due à l'utilisation de données agrégées quand elles proviennent du recensement (Minot and Baulch, 2005)

Ajoutons qu'en l'absence d'étude comparable dans le pays, il est difficile d'interpréter de manière finale les variations d'indice et notre travail nécessite une confrontation avec des mesures dérivées d'autres régions. On sait qu'en Inde, la partie méridionale est loin d'être la moins favorisée et il sera donc important de disposer dans le futur d'étalon de comparaison pour d'autres régions rurales. Mais l'expérience du poverty mapping à échelle désagrégée montre que la diversité entre localités rurales en matière de vulnérabilité reste très forte à l'intérieur des unités administratives régionales et que les estimations par village donnent des

³³ Région familière à plusieurs membres de l'équipe qui ont conduit des travaux de terrain dans diverses localités de cette région.

tableaux beaucoup plus contrastés que les moyennes globales, ce qui a été mis en évidence en d'autres pays en développement quand les données le permettent³⁴. C'est précisément l'essor économique régional, avec ces fortes polarités urbaines et ses concentrations géographiques (zones littorale, districts industriels, etc.) qui risquent d'accentuer les écarts de revenu entre groupes sociaux et localités ainsi que la fragilité des économies rurales.

L'effort d'EMIS butte donc ici contre le manque de données comparatives et l'examen des cartes de pauvreté illustre cette situation propre à la géostatistique ou à l'analyse spatiale quand elle s'applique à des pays en développement. En dépit des enjeux que les mesures peuvent revêtir, et le cas des indicateurs désagrégés de pauvreté l'illustre clairement, les pays en développement ne disposent pas encore de savoir-faire ou de données *ad hoc* pour ce type d'analyse spatiale. Il y a certes un grand nombre de raisons à cette situation, mais la progression des SIG et de la maîtrise de l'information spatiale assure que dans quelques années, ces mesures locales seront disponibles et permettront, au-delà des intérêts avant tout scientifiques que représente notre travail, de se former une image plus claire des paysages de la pauvreté en Inde et de mettre en évidence en particulier les clusters qu'on distingue dans plusieurs régions rurales. Il appartiendra alors à l'ensemble des experts concernés, des géographes aux socio-économistes, de se saisir de ces outils pour interpréter finement ces variations régionales et en tirer les conclusions pour la lutte contre la précarité et la vulnérabilité. Il faut donc espérer que les travaux pionniers comme ceux de notre équipe serviront des points de repère pour les efforts à mobiliser dans les années à venir.

³⁴ Pour une illustration à partir de l'exemple indonésien, voir le travail de Sumarto (2003).

Références bibliographiques

- Agarwal Anil; Narain Sunita, 2003 (1st ed.1997), *Dying wisdom. Rise, fall and potential of India's traditional water harvesting systems*. New Delhi : Centre for science and environment (a citizens' report 4).
- Alwang J., Siegel P.B. et Jorgensen S.L. (2001), «Vulnerability: A View from Different Disciplines», *Social Protection Discussion Paper n°0015, World Bank*, Washington DC.
- Amrhein, Carl G., Reynolds, Harold, 1996, « Using spatial statistics to assess aggregation effects », *Geographical systems*, Vol. 3, pp. 143-158.
- Anselin, Luc, 1995, « Local indicators of spatial association - LISA », *Geographical Analysis*, Vol. 27, n°2, pp. 93-115.
- Anselin, Luc, 1996, « The Moran scatterplot as an ESDA tool to assess local instability in spatial association », in Fischer, Manfred, Scholten, Henk K., Unwin, David, *Spatial analytical perspectives on GIS*, Taylor & Francis, London, pp. 111-125.
- Anselin, Luc, 2003, *GeoDa 0.9 User's Guide*, Spatial Analysis Laboratory, Department of Agricultural and Consumer Economics and CSISS, University of Illinois, 125 p.
- Aubriot, O. et Guilmoto, C.Z., 2003. «Espace et mesure en Inde du sud», Atelier «Regards des sciences sociales françaises sur l'agriculture indienne», INRA/CIRAD, Mars, Nogent sur Marne.
- Azariadis S. et Costas (1996), «The Economics of Poverty Traps», *Journal of Economic Growth*, 1:449-486.
- Bao, Shuming, Henry, Mark S., 1996, « Heterogeneity issues in local measurements of spatial association », *Geographical Systems*, Vol. 3, pp. 1-13.
- Benko, Georges, Lipietz, Alain, Dir.) (2000, *La richesse des régions. La nouvelle géographie économique*, PUF, Paris, 564 p.
- Berry, Brian J. L., 1973, *Growth Centers in the American Urban System*, Ballinger Publishing Company, Cambridge, 195 p.
- Brunet, Roger, Dollfus, Olivier, Ed., 1990, *Mondes nouveaux*, Géographie Universelle, Belin – GIP RECLUS, Paris - Montpellier, 551 p.
- Cassidy, John F., 1997, « The Use of Growth Poles as a Strategy for Regional Economic Development in China », *Political & Economic Review*, Department of Government & Society, University of Limerick, Limerick (<http://www.ul.ie/~govsoc/ulper/1997/ARTICLES/Cassidy.htm>).
- Catin, Maurice, 1994, « Economies d'agglomérations », in Auray, Jean Paul, Bailly, Antoine, Derycke, Pierre-Henry, Huriot, Jean-Marie, *Encyclopédie d'économie spatiale*, Economica, Paris, pp. 105-109.
- Chaléard, Jean Louis, Dubresson, Alain, Ed., 1999, *Villes et campagnes dans les pays du Sud, géographie des relations*, Karthala, Paris, 258 p.
- Chasles, Virginie, 2004, *Entre genre et société : les espaces de la maternité en Inde rurale*. Thèse de doctorat en géographie, Université de Rouen.
- Chatillon Donatien, 2004, *Cartographie de l'irrigation en Inde du sud*, Mémoire de fin d'étude de l'Institut National Agronomique.
- Chaudhuri S., Jalan J. et Suryahadi A. (2001), «Assessing Household Vulnerability to Poverty: a Methodology and Estimates for Indonesia», *Discussion Paper n°0102.52, Columbia University*.
- Cunningham W. et Maloney W.F. (2000), «Measuring Vulnerability: Who Suffer in the 1995 Mexican Crisis?», *Mimeo, World Bank*, Washington DC.

- Currey B. (2002), «Strategic Directions for Vulnerability Analysis and Mapping in the 21th Century: Examples from Bangladesh», *Discussion Paper n°16, Food Policy and Applied Nutrition Program*.
- Davis, Benjamin, 2003, *Choosing a method for poverty mapping*, FAO, Rome.
- Deichmann U. (1999), «Geographical Aspects of Inequality and Poverty», *Working Paper, World Bank*, Washington DC.
- Derycke, Pierre-Henri, 1994, « L'intégration de l'espace dans le champ économique », in Auray, Jean Paul, Bailly, Antoine, Derycke, Pierre-Henry, Huriot, Jean-Marie, *Encyclopédie d'économie spatiale*, Economica, Paris, pp. 19-24.
- Di Méo, Guy, 1991, *L'homme, la société, l'espace*, Anthropos, Economica, Paris, 319 p.
- Didier, Michel, 1990, *Utilité et valeur de l'information géographique*, Paris, Economica, 255 p.
- Dorin Bruno, Landy Frédéric, 2002, *Agriculture et alimentation de l'Inde. Les vertes années (1947-2001)*, Paris : INRA éditions.
- Durand-Dastès, F., Mutin, G., 1995, *Afrique du Nord. Moyen-Orient. Monde indien*, Paris : Belin-Reclus (Géographie Universelle), 480p.
- Getis, Arthur, 1991, « Spatial Interaction and Spatial Autocorrelation: A Cross-Product Approach », *Environment and Planning A*, Vol. 23, 1269-77.
- Getis, Arthur, Ord, J.Keith, 1992, « The analysis of spatial association by use of distance statistics », *Geographical analysis*, Vol. 24, No. 3, pp. 189-206.
- Getis, Arthur, Ord, J.Keith, 1996, « Local spatial statistics: an overview », in Longley, Paul, Batty, Michael, éd., *Spatial Analysis : Modelling in a GIS environment*, Geoinformation International, Cambridge, pp. 261-277.
- Glewwe P. et Hall G. (1998): «Are Some Groups more Vulnerable to Macroeconomic Shocks than Others? Hypothesis Tests Based n Panel Data from Peru», *Journal of development Economics*, 56:181-206.
- Goodchild, Michael F., Janelle, Donald G., 2004, *Spatially Integrated Social Science*, coll. Spatial Information Systems, Oxford University Press, New York, 456 p.
- Goovaerts, P., 2005. "Detection of spatial clusters and outliers in cancer rates using geostatistical filters and spatial neutral models" in Ph. Renard, H. Demougeot-Renard, and R. Froidevaux, eds, *geoENV V - Geostatistics for Environmental Applications*. Kluwer Academic Publishers, Dordrecht.
- Gould, Peter, 1981, « Letting the Data Speak for Themselves », *Annals of the Association of American Geographers*, Vol. 71, n°2, pp. 166-176.
- Guilmoto Christophe Z., 2002, "Irrigation and the Great Indian Rural Database. Vignettes from South India", *Economic and Political Weekly*, vol. 37, n°13, March 30 2002 : 1223-1228
- Guilmoto, C.Z., Oliveau S., Chasles V., Delage R. and S. Vella, 2004. *Mapping out social change in South India: a geographic information system and its applications*, Pondy Papers in Social Sciences 31, Institut français de Pondichéry, Pondicherry, 117 p.
- Guilmoto, Christophe Z., Oliveau, Sébastien, 2000, *South Indian Population Information System, Volume 1 : Tamil Nadu & Pondicherry*, Institut Français de Pondichéry, CD-ROM.
- Guilmoto, Christophe Z., Oliveau, Sébastien, Vingadassamy, Sattianarayanan (2002, « Un système d'information géographique en Inde du Sud : Théorie, mise en œuvre et applications thématiques », *Espace, Populations et sociétés*, Lille, pp. 147-163
- Hägerstrand, Torsten, 1967, *Innovation diffusion as a spatial process*, University of Chicago press, Chicago, 334 p., traduit du suédois par Allan Pred. Edition originale: (1953, *innovations förloppet ur korologisk synpunkt*, C.W.K. Gleerup, Lund (Suède).

- Haining, Robert, 2003, *Spatial data analysis : Theory and practice*, Cambridge University Press, Cambridge, 452 p.
- Hentschel J., Lanjouw J. et al. (1998), «Combining Census and Survey Data to Study Spatial Dimensions of Poverty», *Policy Research Working Paper 1928*, World Bank , Washington DC.
- Huriot, Jean Marie, Perreur, Jacky, 1994, « La centralité », in Auray, Jean Paul, Bailly, Antoine, Derycke, Pierre-Henry, Huriot, Jean-Marie, *Encyclopédie d'économie spatiale*, Economica, Paris, pp. 35-46.
- Isaaks, H.E., and R.M. Srivastava, 1989, *An Introduction to Applied Geostatistics*. Oxford Univ. Press, New York.
- Jalan J. et Ravallion M. (1997), «Spatial Poverty Traps?», *Working Paper*, World Bank , Washington DC.
- Jalan J. et Ravallion M. (2002), «Geographic Poverty Traps? A Micro Model of Consumption Growth in Rural China», *Journal of Applied Econometrics*, 17:329-346.
- Jayaraman V., and Sanjay K Srivastava, 2003, *Poverty Mapping and Monitoring using Information Technology: Learning and Perspectives From India*, Consultancy Report, Ad Hoc Expert Group Meeting (EGM) on Poverty Mapping and Monitoring using Information Technology - Economic and Social Commission for Asia and the Pacific (ESCAP), United Nations, Bangkok, 18-20 August 2003.
- Krugman, Paul, 2000, « Where in the world is the "new economic geography" ? », in Gordon, L. Clark, Feldman, Maryann P., Gertler, Meric S., Ed., *The Oxford handbook of economic geography*, Oxford University Press, Oxford, pp. 49-60.
- Kundu A.K. (ed.), 1989 (revised ed.), *Irrigation atlas of India*. Vol.1 & 2. Calcutta : National atlas and thematic mapping organization.
- Lachaud J-P. (1999a), «Les différences spatiales de pauvreté en Mauritanie: un test de dominance», *Document de travail n°35 du CED*, Université Bordeaux IV-Montesquieu.
- Lachaud J-P. (1999b), «Le différentiel spatial de pauvreté au Burkina Faso: «capabilités» versus dépenses», *Document de travail n°36 du CED*, Université Bordeaux IV-Montesquieu.
- Levine, Ned, 2002, *CrimeStat II: A Spatial Statistics Program for the Analysis of Crime Incident Locations (version 2.0)*, Ned Levine & Associates: Houston, TX/ National Institute of Justice: Washington, DC.
- Mangazol, Claude, 1992, « La localisation des activités spécifiques », in Bailly, Antoine, Ferras, Robert, Pumain, Denise, *Encyclopédie de géographie*, Economica, Paris, pp. 471-496.
- Minot, Nicholas, and Bob Baulch, 2005, "Poverty Mapping with Aggregate Census Data: What is the Loss in Precision?", *Review of Development Economics*, 9, 1, pp 5-24.
- Monmonier, Mark, 1989, « Geographic Brushing enhancing exploratory analysis of the scatterplot matrix », *Geographical analysis*, Vol. 21, n°1, pp. 81-84.
- O'Brien K, et al. "Mapping vulnerability to multiple stressors: climate change and globalization in India" *Global Environmental Change* 14 (2004) 303-313
- Oliveau S. (Ed.), (2003), « Digital Atlas of South India 1991 », *Cybergeogeo*, <http://www.cybergeogeo.presse.fr/AtlasSIndia/html/> (Commentaire des cartes économiques par S. Oliveau, cartes agricoles par F. Landy et sur l'irrigation par O. Aubriot.)
- Oliveau, Sébastien, 1998, *Atlas du district de Vellore : essai de typologie régionale*, mémoire de maîtrise de l'université de Paris X Nanterre, non publié.
- Oliveau, Sébastien, 2004, *Modernisation villageoise et distance à la ville en Inde du Sud*, thèse de doctorat de l'université de Paris 1 Panthéon-Sorbonne, non publiée. <http://tel.ccsd.cnrs.fr/documents/archives0/00/00/76/81/>

- Oliveau, Sébastien, 2005, « Mapping out fertility in South India: Methodology and results », in Guilmo, Christophe Z., Rajan, S. Irudaya, Ed., *Fertility Transition in South India*, Sage, Delhi.
- Ord, J. Keith, Getis, Arthur, 1995, « Local spatial autocorrelation statistics: distributional issues and an application », *Geographical analysis*, Vol. 27, 4, pp. 286-306.
- Pini, Giuseppe, 1992, « L'interaction spatiale », in Bailly, Antoine, Ferras, Robert, Pumain, Denise, *Encyclopédie de géographie*, Economica, Paris, pp. 557-576.
- Pumain, Denise, saint Julien, Thérèse, 2001, *Les interactions spatiales*, coll. Cursus, Armand Colin, Paris, 191 p.
- Saint Julien, Thérèse, 1985, *La diffusion spatiale des innovations*, coll. Reclus modes d'emploi, GIP RECLUS, Montpellier, 40 p.
- Saint Julien, Thérèse, 1992, « Diffusion Spatiale », in Bailly, Antoine, Ferras, Robert, Pumain, Denise, *Encyclopédie de géographie*, Economica, Paris, pp. 577-598.
- Spate O.H.K., Learmonth A.T.A, 1967 (1st ed: 1954), *India & Pakistan. Land, people and economy*. London : Methuen & co, 439p.
- Sumarto, ASS, 2003, *Poverty Mapping Efforts in Indonesia* Ad Hoc Expert Group Meeting (EGM) on Poverty Mapping and Monitoring using Information Technology - Economic and Social Commission for Asia and the Pacific (ESCAP), United Nations, Bangkok, 18-20 August 2003.
- Tabutin, Dominique, 2000, *La ville et l'urbanisation dans les théories du changement démographique*, Université Catholique de Louvain, Département des Sciences de la population et du développement, Document de travail n°6, 38 p.
- Thisse, Jean François, 2002, « Dispersion, agglomération et re-dispersion », *RERU*, n°1, pp. 13-24.
- Tiefelsdorf, Michael, 1998, « Some Practical Applications of Moran's I's Exact Distribution », *Papers in Regional Science*, Vol. 77, pp. 101-129.
- Tiefelsdorf, Michael, Boots, Barry N., 1997, « A Note on the Extremities of Local Moran's I_i and their Impact on Global Moran's I », *Geographical Analysis*, Vol. 29, pp. 248-57.
- Tukey, John Wilder, 1977, *Exploratory Data Analysis*, Reading MA: Addison Wesley publishing Company, 688 p.
- Vaidyanathan; 2001, *Tanks of South India*, Centre for Sciences and Environment, New Delhi, 178 p.

Annexe I : Le nuage de points de Moran

Le nuage de points de Moran est décrit en détail dans Anselin (1996), il est aussi évoqué dans Anselin (1995 & 2003). Il s'agit d'un mode de visualisation des données statistiques spatiales. Anselin fait le constat en 1996 que si les techniques consistant à « laisser les données parler d'elles-mêmes » (Gould, 1981) s'étaient bien développées (voir par exemple Monmonnier, 1989) depuis les travaux de Tukey (1977) et la mise en valeur des techniques d'EDA (analyse exploratoire des données), il n'en était pas de même pour l'analyse exploratoire des données spatiales (ESDA).

Fort de la mise en place des indices locaux d'association spatiale (LISA), Anselin propose une autre innovation, qui leur est liée, sans toute fois être identique. Comme il le rappelle lui-même : « le nuage de points de Moran n'est pas un LISA » (Anselin, 1995 : 105).

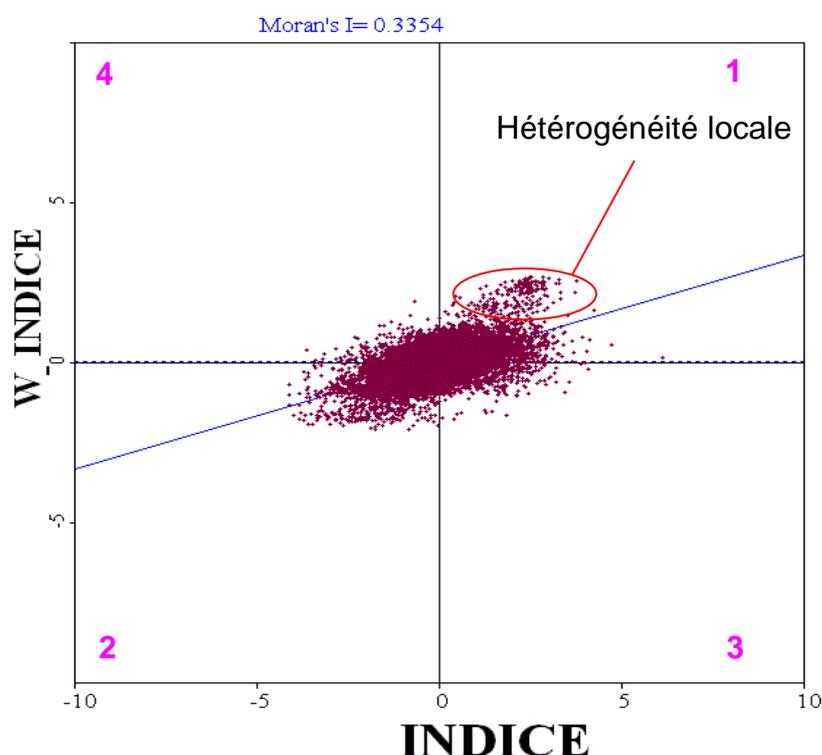


Figure 34 : nuage de point de Moran pour l'indice de modernisation (voisinage de 15 km)

Le principe de ce nuage de points est assez simple. Il consiste à mettre en opposition la valeur d'une variable en un point « i » et sa valeur dans son voisinage « j » (voisinage défini par une matrice w_{ij}), après avoir normalisé et standardisé les variables. En répétant l'opération pour tous les points « i » et leurs voisins « j », on obtient un nuage de points qui nous permet de visualiser à la fois l'autocorrélation spatiale de la variable et son hétérogénéité.

L'autocorrélation spatiale est déduite de la pente de la droite de régression (en bleu sur la figure) entre les valeurs en « i » (INDICE) et en « j » (W_INDICE), soit 0,3354 dans l'exemple présenté. Les zones d'hétérogénéité locale (poches où l'autocorrélation spatiale est plus forte ou plus faible que dans l'ensemble de l'échantillon) sont mises en valeur par des regroupements visibles sur le nuage de points. On voit, entourées en rouge sur la figure, des

valeurs de voisinages plus fortes que la moyenne des valeurs des voisinages, qui correspondent à la poche au Sud du Tamil Nadu.

On peut aussi distinguer quatre types d'associations qui correspondent aux quatre quadrants du nuage de points de la Figure 34:

- Les valeurs positives (indice supérieur à la moyenne) au sein d'un environnement de points à valeur positive (autocorrélation spatiale positive). Association qualifiée de *high-high* dans la littérature anglo-saxonne sont visibles dans le quadrant n°1 (chiffre en rose).
- Les valeurs négatives (indice inférieur à la moyenne) au sein d'un environnement de points à valeur négative (autocorrélation spatiale positive). Association qualifiée de *low-low*. Ils sont visibles dans le quadrant n°2.
- Les valeurs positives au sein d'un environnement de points à valeur négative (autocorrélation spatiale négative). Association qualifiée de *high-low*. Ils sont visibles dans le quadrant n°3.
- Les valeurs négatives au sein d'un environnement de points à valeur positive (autocorrélation spatiale négative). Association qualifiée de *low-high*. Ils sont visibles dans le quadrant n°4.

On comprend ainsi mieux que les points situés dans les quadrants 1 et 2 participent à la tendance rendant l'autocorrélation spatiale positive, alors que les points situés dans les quadrants 3 et 4 tendent à rendre l'autocorrélation spatiale négative. Les points situés dans ces deux quadrants sont qualifiés de « *spatial outliers* » (valeurs spatialement atypiques), puisqu'ils ont des valeurs opposés à celles de leurs voisins.

Annexe II : Le cokrigeage

La méthode du cokrigeage n'a guère été utilisée dans l'analyse de la pauvreté. Pour la décrire en quelques mots, on dira que cette technique consiste à compléter un krigeage pour estimer des données locales à partir d'estimations éparses de données complémentaires, disponibles à l'échelle locale. Le cokrigeage associe en effet un krigeage d'une variable z (disponible à une échelle donnée) à une modélisation statistique de la même variable z grâce à un vecteur x de covariables différentes de z , mais disponibles à une échelle locale plus fine. La modélisation statistique permet de mieux estimer les micro-variations de z par imputation à partir du vecteur x , tout en conservant les propriétés du krigeage comme estimateur spatiale³⁵.

Cette méthode nous a semblé pourtant cruciale pour importer localement des données disponibles à une échelle supérieure et mériterait d'être systématisée à ce genre d'exercice. Elle n'est à notre connaissance que rarement utilisé en science sociale.³⁶ Nous allons en premier lieu résumer notre problème d'estimation et examiner les différentes solutions disponibles et leurs avantages respectifs.

Nous disposons ici d'une estimation de la mortalité infantile rurale à l'échelle des 80 districts de l'Inde du sud (en 1991). Nous aimerions disposer d'estimation géographiquement plus fines, applicables par exemple à nos clusters (c105 ou c110). Il s'agit d'un cas relativement classique d'imputation macro->micro. Les districts sont les unités macro et les clusters sont les unités micro.

- Sources macro : mortalité infantile, covariables du recensement
- Sources micro : covariables du recensement
- Informations spatiales : localisations micro et macro
- Objectif : estimation micro de la mortalité infantile

Différentes méthodes se présentent que nous résumons dans la liste qui suit :

1. Imputation uniforme : toutes les unités micro prennent la valeur de l'unité macro
2. Krigeage : krigeage sur les unités micro des valeurs macro, attribuées au centre de l'unité macro
3. Imputation statistique : modélisation macro du lien entre la valeur observée et d'autres covariables. Application micro du modèle obtenu aux mêmes covariables disponibles, pour obtenir une estimation micro.
4. Cokrigeage : krigeage (cf. plus haut) en tenant compte des mêmes covariables macro (modélisation) et micro (imputation)

Les avantages et désavantages respectifs sont également résumés :

1. Simplicité absolue de la méthode. Mais résultat uniforme sans distinction micro.
2. régularité spatiale de l'estimation. Mais centralité de l'estimation et lissage en partie artificielle entre centroïdes
3. Imputation micro détaillée grâce aux covariables. Mais résultat aspatial et forte sensibilité à la qualité de la donnée micro

³⁵ Sur le cokrigeage (cokriging), voir par exemple Isaaks and Srivastava (1989).

³⁶ Pour un exemple en géographie épidémiologique, voir Goovaerts (2005).

4. Combinaison des avantages des deux précédents : estimation spatialement régulière et maximisation de l'usage des covariables micro. Mais effet de centralité du krigeage et sensibilité à la qualité de l'information micro.

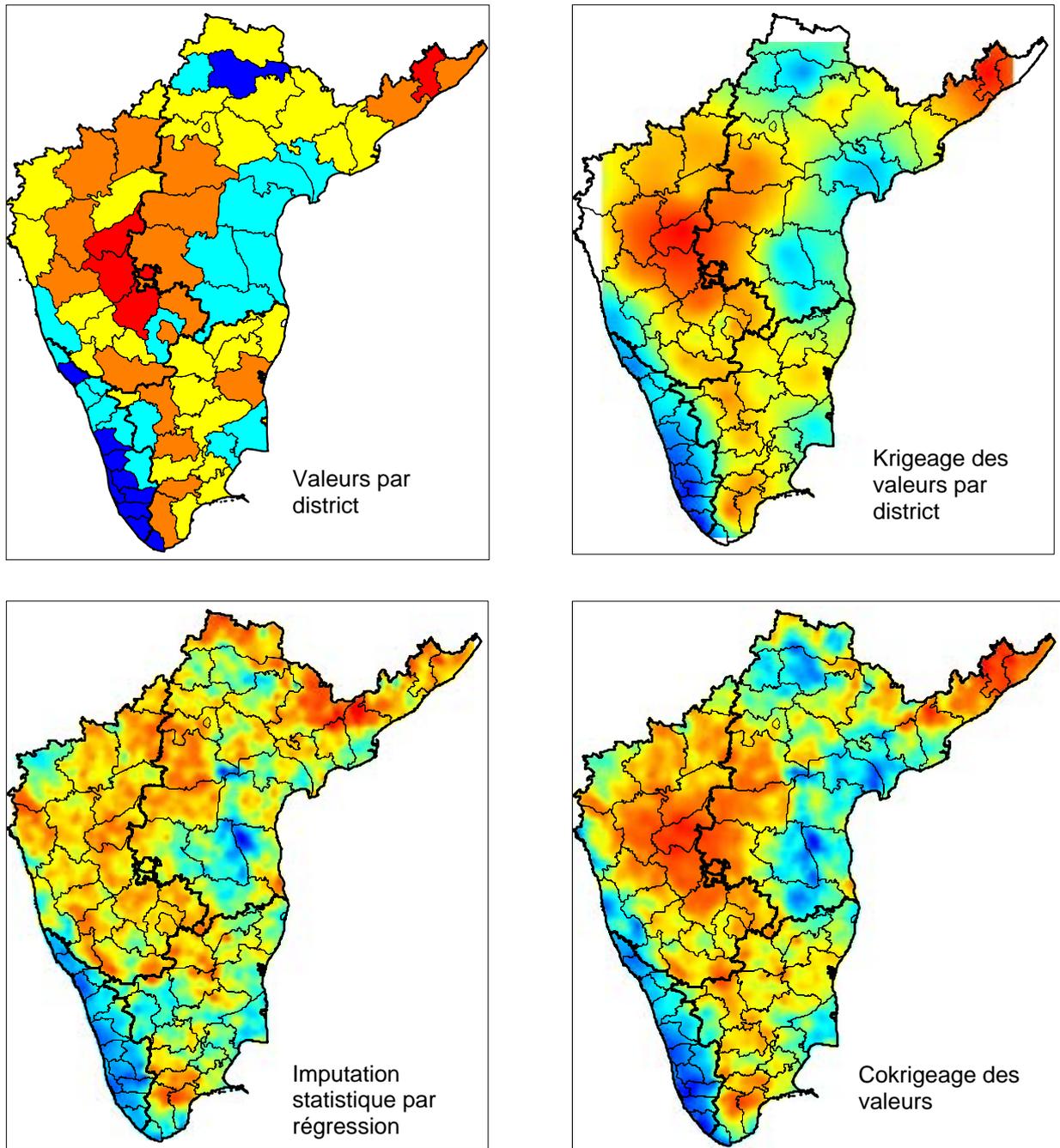


Figure 35 : Estimation locale de la mortalité infanto-juvénile selon quatre méthodes : uniformité au sein des districts, krigeage de la valeur district, imputation locale, cokrigeage

Notons que pour rendre opérationnelle les deux dernières méthodes 3 et 4, nous avons choisi une batterie de variables liées à la mortalité infantile. Ces dernières ont été examinées par ajustements successifs afin d'éliminer l'effet de variations « aberrantes » à l'échelle micro.

Considérons par exemple l'effet d'une variable à faible variation macro, comme la présence de tribaux, variant de 0 à 15% dans les districts. Cette variable peut avoir un lien statistique, faisant par exemple monter la mortalité. L'équation liant la mortalité à ses covariables, parmi lesquels la proportion de tribaux, a toutefois été calibrée à partir d'observations macro à faible écart-type. Or, si au sein d'un cluster on trouve un pourcentage atypique de tribaux (disons 100%), cela peut faire entièrement basculer l'imputation et décupler la mortalité infantile. Le chiffre local devient proprement absurde et provient simplement de la présence d'une valeur extrême de la covariable, valeur non prise en compte dans la modélisation faite à l'échelle macro.

Nous avons donc dû éliminer les variables dont les variations locales étaient sans rapport avec les variations macro. A partir des covariables conservées, nous avons estimé la mortalité théorique macro et micro selon la méthode 3. Puis nous avons utilisé ces estimations comme base pour le cokrigeage. Sur la Figure 35, nous avons rassemblé les quatre méthodes. Les niveaux les plus élevés de mortalité infantile sont représentés ici en rouge, alors que les plus bas sont en bleu. Les points représentés correspondent aux centres des districts.

On reconnaît sur la première carte la distribution des valeurs par district. Aucune distinction à l'intérieur des districts n'est faite. La carte suivante est l'estimation par krigeage, dérivée de la précédente. Elle est centrée sur les centres de district qui servent de support à l'estimation surfacique. Les différences intra-district sont régulières.

Sur la troisième carte, l'imputation micro est faite à partir des covariables disponibles. La carte est alors beaucoup plus accidentée, car les variations locales affectent directement l'estimation. Elle ne correspond pas exactement avec la carte par district, en raison de l'imprécision de l'estimation.

La dernière carte de la Figure 35 est dérivée du cokrigeage. Elle correspond par conséquent à une reprise de la première carte obtenue par krigeage, mais avec une correction pour tenir compte des variations micro de la covariable et de l'intensité de la corrélation. Sans être très éloignée de la précédente, elle rétablit donc une plus forte régularité spatiale, à une échelle désormais plus fine que le krigeage originel, et les valeurs locales concordent mieux avec les moyennes de district de la première carte.

Pour donner une vue plus fine des particularités du cokrigeage, on a procédé à un zoom sur les cartes précédentes sur une partie de l'Inde du sud. Les deux cartes de la Figure 36 comparent donc les résultats par krigeage (à gauche) et cokrigeage (à droite) pour la région de Bangalore-Chennai, en utilisant toutefois une gamme chromatique légèrement différente pour faire ressortir les contrastes locaux. On distingue plus clairement le gain relatif au cokrigeage dans l'estimation des valeurs locales de la mortalité infanto-juvénile. La carte de gauche obtenue par krigeage fait en effet apparaître un profil très lisse, correspondant à une estimation presque linéaire de la mortalité entre les points d'observations, à savoir les valeurs par district.

La méthode du cokrigeage donne des résultats plus nuancés. A l'intérieur d'un district donné, les écarts entre localités sont marqués et irréguliers. L'estimation de ces écarts locaux repose naturellement sur les variations localement observées dans les covariables de la mortalité de notre modèle. On voit par exemple apparaître au sud de l'Andhra Pradesh côtier une petite poche à plus forte mortalité (en jaune) et les zones à faible mortalité (en bleu) ont un découpage plus tourmenté.

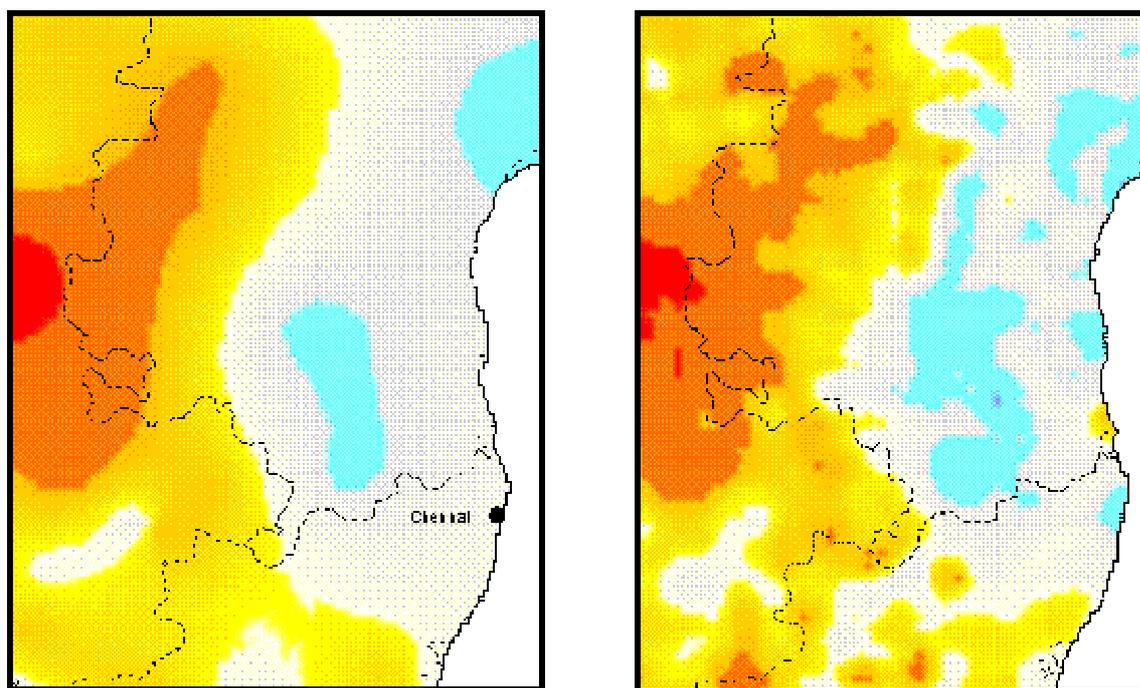


Figure 36 : Estimation locale de la mortalité infanto-juvénile : krigeage et cokrigeage (zoom)

L'absence de données qui a justifié le cokrigeage initial ne nous permet naturellement de confirmer l'existence de ces micro-variations géographiques, mais les covariables de la mortalité (telle que le degré d'alphabétisation) fournissent sans doute des prédictions raisonnables. On voit ainsi apparaître sur la carte de cokrigeage une zone à faible mortalité (en bleu) sur le littoral du Tamil Nadu, absente de l'estimation initiale opérée par simple krigeage. Cette région n'est autre que la zone métropolitaine de Chennai, relativement prospère, où la mortalité est effectivement significativement inférieure à la moyenne locale. Inversement, le cokrigeage fait émerger de nouvelles poches à forte mortalité dans des régions longeant la frontière entre Tamil Nadu et Karnataka, qui correspondent effectivement à des régions semi-forestières de colline particulièrement déshéritées (région du fameux brigand Veerapan), où les conditions sanitaires sont très certainement plus médiocres que dans les plaines. On pourrait d'autres exemples qui semblent intuitivement plausibles, mais l'absence de données mesurées nous interdit d'aller plus loin.