



HAL
open science

L'influence des contraintes environnementales sur la distribution de la population française entre 1793 et 1999

Mattia Bunel

► **To cite this version:**

Mattia Bunel. L'influence des contraintes environnementales sur la distribution de la population française entre 1793 et 1999. Géographie. 2015. dumas-01242929

HAL Id: dumas-01242929

<https://dumas.ccsd.cnrs.fr/dumas-01242929v1>

Submitted on 14 Dec 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - ShareAlike 4.0 International License

UNIVERSITÉ D'AIX-MARSEILLE
MASTER 2 DE GÉOGRAPHIE

Mémoire de recherche
Septembre 2015

L'influence des contraintes environnementales sur la distribution de la
population française entre 1793 et 1999

Mattia Bunel

Mémoire réalisé sous la direction de M. SÉBASTIEN OLIVEAU

Remerciements

Je tiens à remercier mon directeur de recherche, M. SÉBASTIEN OLIVEAU pour m'avoir donné l'opportunité de réaliser ce sujet de recherche et pour m'avoir aidé à le réaliser tout au long de cette année.

Je remercie également M^{me} ISABELLE BLÖSS-WIDMER pour son aide lors de la réalisation de la base de données, ainsi que l'ensemble des membres du projet DEMOMED pour leurs aides et enseignements.

Enfin je remercie les membres du jury, SÉBASTIEN OLIVEAU et FRÉDÉRIC AUDARD, pour leur présence en ce jour.

Table des matières

Introduction générale	4
1 La distribution de la population française et ses variations	7
Introduction	7
1.1 Les sources et leur traitement	7
1.1.1 De la révolution à la création de la SGF	9
1.1.2 De la première guerre mondiale à la fin de la périodicité quinquennale	14
1.1.3 Le recensement après 1946	16
1.2 Échelle nationale	22
1.2.1 Évolution de la population et de la densité	23
1.2.2 Dispersion statistique.	26
1.2.3 . . . Concentration géographique	31
1.3 Spécificités locales	37
1.3.1 Cartographie thématique	37
1.3.2 Indicateurs locaux d'association spatiale	40
1.3.3 Typologie des communes	47
Conclusion	50
2 L'influence de l'environnement	52
Introduction	52
2.1 Théorie de la distribution de la population	52
2.1.1 Distribution de la population et environnement	52
2.1.2 L'importance déterminante de la composante historique	55
2.1.3 Sources, données et traitements	59
2.2 Approche cartographique	60
2.2.1 Évolution de la superficie	60
2.2.2 Cartes	67
2.3 Approche quantitative	72
2.3.1 Régression linéaire	72
2.3.2 Analyses de la variance	80
2.3.3 Typologie	82
Conclusion	88
Conclusion générale	90
Annexes	91
A Documents relatifs à la base de données	91
A.1 Logiciels utilisés	91
A.2 Un exemple de notice communale	92
A.3 Évolution de la définition de la population	92
B Lois relatives au recensement	95
B.1 Extrait de la loi relative à la démocratie de proximité	95
B.2 Ordonnance du 16 janvier 1822	96
Bibliographie	98
Table des figures	104
Liste des tableaux	105

Introduction

« Je crois que je n'ai jamais autant entendu parler de géographie, [...] depuis le lycée [...]. »¹

C'est par cette constatation, légèrement amère pour les géographes, qu'Éric NAULLEAU introduisit, le 13 avril 2012, l'invité du *talk-show* portant son nom. Ce jour là l'invité n'est autre que le géographe Christophe GUILLUY, venu présenter sa théorie socio-spatiale du *glissement vers la droite* de l'électorat populaire. Cette dernière connaîtra une médiatisation conséquente suite à la publication en septembre 2014 de l'essai *La France périphérique* du même auteur. Il s'agit de la troisième publication (fortement médiatisée) de l'auteur sur les questions d'inégalités socio-spatiales².

L'effervescence médiatique de septembre 2014 autour de cet ouvrage ayant cessé il nous a semblé intéressant de revenir sur son contenu qui a fait de son auteur (tout du moins à la fin de l'année 2014) un des *géographes le plus médiatisés*, au même titre que Sylvie BRUNEL, ou Jean-Robert PITTE.

L'auteur explique dans son ouvrage un phénomène observé à l'échelle nationale (à savoir la migration des votes de l'électorat populaire de l'extrême gauche à l'extrême droite) par le biais du prisme spatial. La réflexion est basée sur une opposition binaire entre la France métropolitaine (« *belle vitrine de la mondialisation heureuse* ») et la *France périphérique* uniformément délaissée. Bien entendu nous ne nions pas qu'il existe une rupture socio-spatiale, ni même que la mondialisation a pu conduire à une explosion de cette (ces) rupture(s). Mais l'auteur considère que l'opposition binaire qu'il défend suffit à expliquer le phénomène étudié, par conséquent il considère que la population métropolitaine est uniformément gagnante (quel que soit son statut) et que la *France périphérique* est uniformément perdante. Ce qui est sans aucun doute une affirmation trop simpliste.

Mais alors comment expliquer la grande réussite médiatique de cet ouvrage ? Il est probable qu'elle est avant tout due à la très faible médiatisation des géographes. Peut accoutumée à ce prisme de lecture, la doxa médiatique a découvert dans l'ouvrage de GUILLUY l'apport de la prise en compte de l'espace dans une réflexion qu'elle soit spatiale ou politique. Si l'on y réfléchit, rares sont les concepts géographiques ayant durablement marqué le grand public³. En ce mois de septembre 2014 la crainte d'une énorme percée du *Front National* aux élections départementales est dans tous les esprits. Aujourd'hui, alors que le score obtenu par ce parti est plus faible qu'escompté, l'accueil du livre aurait sans doute été tout autre.

Pour autant la rupture spatiale présentée dans cet ouvrage n'est clairement pas anecdotique et est le fruit d'une longue histoire. Et si notre travail n'aborde pas la question sociale étudiée par GUILLUY l'étude cette rupture spatiale sera au cœur de notre sujet de recherche, visiblement dans l'air du temps.

Présentation du sujet

Ce sujet est né avec notre lecture, en début d'année, de la *planète au village*, un ouvrage rédigé par LE BRAS et LABBE [50] en 1993. Dans un des premiers chapitres on retrouve une affirmation interpellante. Les zones plus difficiles à vivre sont peu à peu dépeuplées, même à une époque où les progrès technologiques laisseraient supposer l'inverse. Selon les auteurs le rôle des progrès technologiques est avant tout à chercher dans l'augmentation des rendements agricoles, permettant une production plus importante sur les terres les plus fertiles et ne rendant plus la culture de certaines terres (notamment de montagne) inutiles. Cette affirmation nous a interpellé, pensant, sans doute naïvement, que les progrès technologiques faciliteraient la vie humaine dans les zones les plus contraignantes. Il nous a donc semblé intéressant de chercher à valider cette hypothèse et de nous questionner sur l'évolution de la distribution de la population et du lien qu'entretient cette évolution avec différentes composantes environnementales. L'affirmation de LE BRAS et LABBE [50] est basée sur une observation, celle d'une tendance à la concentration de la population

1. Eric Naulleau, *Zemmour et Naulleau* [Vidéo en ligne], Paris Première, le 13 avril 2012. Repéré à <https://youtu.be/IV7FCyGF8fk?t=2m10s>, 9 min 31.

2. Les deux précédentes étant l'*Atlas des fractures françaises*, publié en septembre 2000 aux éditions l'Harmattan et l'essai *Fractures françaises* publié en octobre 2010 chez Bourin.

3. La *diagonale du vide* et la *banane bleue* en sont deux exemples.

française au cours du temps. Nous sommes partis du principe que l’invalidation de cette observation nous permettrait d’invalider l’ensemble de l’affirmation, et inversement qu’il nous était indispensable de montrer l’existence d’une tendance à la concentration de la population pour valider cette observation. L’organisation générale de ce travail a donc été organisée suivant cette logique. Le premier chapitre de ce travail porte sur l’étude de la distribution de la population française, son évolution et vise à démontrer l’existence d’une tendance globale à la concentration de la population. Ce n’est que dans le second chapitre que nous introduirons la dimension environnementale à notre travail. Cette étape sera l’occasion de nous questionner sur l’effet de l’environnement sur la distribution de la population et de montrer l’existence d’un lien entre caractéristiques physiques des communes et l’évolution de leur densité.

Présentation des données

Le travail ici présenté n’aurait pas été réalisable sans l’obtention de la base de données mise en ligne par le *Laboratoire de Démographie et d’Histoire Sociale (LaDéHiS)*⁴ sur le site *Des villages de Cassini aux communes d’aujourd’hui*. Ce dernier est conçu comme la mise en relation de trois types de données, une base de données administratives et démographiques, basée sur les travaux précédents du laboratoire, notamment la collection de dictionnaires *Paroisses et communes de France : Dictionnaires d’histoire administrative et démographique* (Première publication en 1974 dont les données ont été ressaisies pour les 47 départements présents dans la collection et complétée pour les départements restants [58]. Il en résulte une base de données démographique nous permettant d’obtenir la population de chaque commune pour trente-trois dates (de 1793 à 1999, soit une amplitude de 206 ans). Ce corpus de données est couplé à un *fichier national des modifications de limites administratives* (réalisé conjointement avec l’INED) constituant la partie administrative précédemment mentionnée [48].

Les deux composantes restantes sont un fond vectoriel des limites administratives de la France à l’échelle communale (basée sur la base de données de l’IGN BD-carto datée de 1999, désormais remplacée par le fichier GéoFla) couplée à la carte de Cassini (1757-1790). Seule la première composante présentait un intérêt pour notre travail, il s’agit par conséquent de la seule qui ait été utilisée.

La base de donnée se présente comme un ensemble de fiches communales⁵ (41 475) accessibles par le biais d’une interface cartographique ou d’un index alphabétique. Quelque soit la commune étudiée la notice est composée des mêmes informations, organisées de manière identique⁶. Le grand nombre de fiches communales est expliqué par la présence de notices concernant toutes les communes, y compris celles disparues ou fusionnées. Ces dernières se sont vues attribuer le code Insee (actuel) de la commune à laquelle elles furent rattachées.

La construction de notre base de données nécessitait en premier lieu l’obtention des données précédemment mentionnées. Étant donné que ce site est le seul à proposer une si grande amplitude temporelle il était impensable de travailler sur notre problématique sans ces données. Dans un premier temps nous avons contacté à de multiples reprises les auteurs de ce travail⁷. Mais nous n’avons obtenu aucune réponse. La solution retenue a donc été d’automatiser le téléchargement à l’aide d’un script Python *ad hoc*. Sa réalisation a nécessité plusieurs semaines de travail, mais il nous a permis de télécharger les données en moins de cinq heures⁸. Notons que le script ne se contentait pas de télécharger les notices, il récupérait uniquement les données voulues avant de les inscrire dans un fichier *csv*. Nous n’avons pas récupéré l’intégralité des données présentes sur l’ensemble des fiches communales et ce pour deux raisons. Tout d’abord nous n’avons sélectionné que les données présentant un intérêt dans le cadre de notre travail⁹, ainsi les données portant sur l’administration, les différents noms et les modifications n’ont pas été retenues. De plus, ces informations auraient été nettement plus délicates à traiter notamment du fait des relations (avec d’autres notices) qu’elles induisent. Par conséquent nous avons préféré ignorer ces informations, quitte à modifier ultérieurement le script réalisé si la nécessité d’obtenir ces données se faisait sentir.

C’est suite au téléchargement de ces données que notre travail est passé de *l’utilisation d’une base de données existante*, à la *constitution d’une base de données*. Face à la taille de la base de données, mais

4. Anciennement *Laboratoire de démographie historique* (ldh).

5. Pour un exemple on consultera la figure A.1 page 92.

6. Cette caractéristique était une condition à l’automatisation du téléchargement des données, et donc à leur obtention.

7. Je tiens particulièrement à remercier ISABELLE BLÖSS-WIDMER, responsable du projet DEMOMED, pour son aide lors de cette étape.

8. La longue durée nécessaire à l’obtention de toutes les données s’explique à la fois par la faible performance de la connections internet du serveur hébergeant la base de données, mais également à la lenteur (connue) du *parseur* utilisé lors de cette opération (BeautifullySoup). Toutefois il convient de relativiser cette longueur, dans le cas d’un téléchargement manuel cette étape aurait nécessité entre trois cent quarante-cinq heures (16 jours) et mille trois cent quatre-vingt-trois heures (58 jours) de travail cumulé (estimation basée sur la durée moyenne observée lors du traitement manuel d’un ensemble de notices)

9. Les données sélectionnées sont les suivantes : *nom, superficie, altitude, coordonnées, code insee, statut, nombre d’habitants* (34 variables), mais toutes n’ont pas été traitées. Certains champs n’ayant été utilisés que dans un but vérificateur

surtout à la complexité des opérations nécessaires nous avons travaillé à l'aide d'un *système de gestion de bases de données relationnelles* (SGBDR), SQLite. Aucun traitement des données n'ayant été réalisé par les créateurs de la base de données téléchargée il s'est avéré nécessaire d'agrèger les données. Ainsi, étant donné que le code insee donné sur la notice est celui de la commune (actuelle) de rattachement nous n'avons fait que calculer la somme de la population (pour chaque date) de toutes les communes présentant le même code Insee). Mais, l'attribution des codes Insee à été faite à partir de la base de données BD-Cartho de 1999. Or cette dernière n'est plus disponible. Il a donc été nécessaire de travailler à partir de son équivalent actuel, la base de données géofla (V2.0 2014). Évidemment entre ces deux dates il y eut quelques modification du maillage communal. Il a donc été nécessaire de procéder manuellement à quelques ajustements¹⁰.

Quatre cas de figure se sont présentés. Le premier, (le plus simple à traiter) est celui des arrondissements municipaux, si la base de données Géofla contient le détail des arrondissements des villes de Lyon, Marseille, et Paris ce n'est pas le cas de la base *Des villages de Cassini aux communes d'aujourd'hui* – désormais mentionnée sous le nom *cassini.ehess* (en référence au nom de domaine) – qui ne donne la population qu'à l'échelle de la ville. Dans ce cas le traitement ne consiste qu'à fusionner les polygones du fond de carte et leur attribuer le bon code Insee¹¹.

La seconde situation est celle des communes qui ont été ultérieurement fusionnées. Dans ce cas la base de données *cassini.ehess* contient des communes qui n'existent plus (et donc qui n'apparaissent pas dans la base *Géofla*. Le *COG* contenant un historique des communes ayant existé une simple recherche à permis de connaître, pour chaque commune fusionnée, la commune fusionnante. Nous avons ensuite attribué le code de la commune fusionnante à celle de la commune fusionnée (en prenant soin de ne pas détruire l'information originelle).

La troisième situation est plus délicate. Il s'agit des communes n'existant pas (plus) en 1999 mais recrées avant 2015. Il s'agit exclusivement de communes précédemment fusionnées puis scindées. Il est par conséquent impossible de faire correspondre les données de la base *cassini.ehess* avec celles du fond de carte. Nous avons donc fusionné les communes du fond de carte, afin qu'elles correspondent à la base de données. Par conséquent nous ne travaillons pas exactement sur le maillage communal de (fin) 2014, mais sur un maillage *ad hoc*, modifié par contrainte. Toutefois ces communes sont relativement rares, la modification du maillage originel est donc limitée.

La dernière situation ne concerne que trois communes. Ces dernières existent à la fois dans les bases *cassini.ehess* et *Géofla* mais les données ne sont pas présentes sur la fiche communale. Il s'agit vraisemblablement d'une erreur de traitement ou d'un oubli de ces trois communes. Par conséquent elles n'ont pas été traitées dans le cadre de ce travail. Fort heureusement ce cas est très rare.

Une fois ces corrections réalisées (en nombre limité si on le rapport au nombre total de communes) la base de donnée *cassini.ehess* a été jointe au fond de carte. Cet ensemble données constitue le cœur de ce travail. Si dans la seconde partie du travail nous couplons ces données à un corpus de variables environnementales (présentées ultérieurement), une grande partie de ce travail en traite exclusivement (notamment la première partie).

Malgré la qualité variable des données (données absentes, lacunes, qualité variable du recensement (dénombrement) *etc.*) il faut noter que cette dernière est de très bonne facture. Dans le pire des cas (début du XVIII^e siècle) la part des communes manquantes (ou non utilisables) est aux alentours de 5 % ce qui nous permet de travailler avec un nombre d'individus conséquent. Toutefois, bien que le nombre d'individus manquants soit relativement faible leur absence n'est pas sans conséquences sur la significativité de l'échantillon. L'utilisation de la base de données *cassini.ehess* apporte une réelle plus value à ce travail, car elle nous permet de travailler sur une durée (et à une échelle géographique) vraisemblablement inédite. Bien sur il faut être conscient des limites de cette base de données, c'est pourquoi nous reviendrons longuement sur la qualité des données recueillies par le *LaDéHiS* pour construire la base de données *cassini.ehess*. Cette partie n'a pour vocation que d'introduire les données sur lesquelles nous avons travaillé. Sa présence dans l'introduction est justifiée par l'importance de ces données dans le cadre de notre sujet, qui, si elles n'avaient pas été disponibles l'auraient rendu irréalisable.

10. L'intégralité de ce travail à été réalisée à l'aide du code géographique mis en ligne sur le site de l'INSEE (abrégé COG). (« *Le code officiel géographique rassemble les codes et libellés des communes, [...] au 1^{er} janvier 2015* » [42].)

11. Pour éviter les confusions le code Insee de l'ensemble de la commune ne fait pas doublon avec celui d'un arrondissement, par exemple le code insee des communes de Lyon, Marseille, et Paris est respectivement 69 123, 13 055 et 75 056. Alors que les codes Insee des arrondissements sont respectivement situés dans les pages 69 381-89, 13 001-17, 75 101-20.

Chapitre 1

La distribution de la population française et ses variations

Introduction

Comme expliqué dans l'introduction générale, l'objectif de ce premier chapitre est d'étudier l'évolution de la distribution de la population française entre 1793 et 1999. C'est en confrontant les affirmations de LE BRAS et LABBE [50] avec nos observations que nous pourrions valider une partie de observations de ce dernier (à savoir qu'il existe une tendance à la concentration de la population). Ce chapitre, comme l'ensemble du mémoire est organisé suivant un gradient de complexité. Dans le premier chapitre nous n'étudions que la distribution de la population, que nous couplerons dans le second chapitre à un ensemble de composantes environnementales. De la même manière la seconde partie du premier chapitre aborde la question de la distribution de la population en ajoutant une composante volontairement omise en première approche, la dimension spatiale. Ainsi après nous être questionnés sur l'évolution de la population française dans son ensemble à l'aide d'une approche *aspatiale* (cf. partie 1.2 : *Échelle nationale*) nous introduirons cette dimension dans la partie *Spécificités locales*. L'objectif de cette division et de faciliter la compréhension du phénomène en simplifiant la première étude au maximum, puis en ajoutant un degré de complexité supérieur à chaque nouvelle étude. Mais avant de traiter les données il nous a semblé nécessaire de revenir sur certains points, notamment sur la base de données utilisée. Celle-ci a déjà été sommairement présentée en introduction, mais nous avons décidé de compléter sa présentation au sein d'une première partie. L'amplitude temporelle rend la qualité des recensements utilisés très variable, et il est nécessaire de revenir sur ce point. Comme nous l'avons déjà dit, le principal défaut de cette base de données est sa très faible documentation. Il nous était impossible d'utiliser ces données sans en apprendre plus sur l'origine de ces données, leur qualité, l'évolution de la pratique du recensement et autres spécificités. La première partie de ce chapitre synthétise donc les informations recueillies à partir de nos recherches bibliographiques, dans le but de dresser un portrait, que l'on souhaite le plus précis possible, des données et plus particulièrement de leur qualité.

1.1 Histoire de la statistique française et création de la base de données

C'est par le prisme de l'histoire de la statistique française que nous avons abordé la question de la qualité des données issues des trente-trois dénombremens et recensements composant la base de données *des villages de Cassini aux communes d'aujourd'hui*. Il va sans dire qu'entre 1793 et 1999 la conception et la pratique de la statistique a profondément changé. Différentes institutions et méthodologies se sont succédées. Notre propos est organisé en plusieurs parties chronologiques, mettant en exergue les ruptures historiques. Bien que le premier dénombrement contenu dans notre base de données soit celui de 1793, il convient de faire remonter l'histoire à l'année de la révolution française (1789), date marquant une véritable rupture dans la conception (française) de la statistique [31].

Bien que la pratique du recensement soit bien antérieure à la révolution française, (la pratique du recensement remonte à l'antiquité, le titre de *censeur* étant même, sous la république romaine, la plus

prestigieuse des charges¹) aucun *dénombrement national*² n'a été effectué en France sous l'ancien régime [30, 32, 31]. Il y eut certes plusieurs tentatives, dont *l'état des paroisses et des feux de 1328*, opération quasi-nationale, ou encore les tentatives réalisées sous Charles VIII (1492), Louis XII (1503) et François I^{er} [29]. Mais ces dernières n'ont probablement pas abouti⁴, elles font toutefois partie de rares opérations de dénombrement (antérieures au XVII^e siècle) pour lesquelles nous avons quelques informations. La plupart des opérations réalisées sous l'ancien régime ne portaient que sur une aire limitée (qui peut aller de la ville, à la province, en passant par la paroisse), et n'étaient pas à l'initiative du pouvoir royal⁵. Ces données (à la qualité de conservation très variable [30, 32]) ne peuvent être logiquement exploitées que dans le cadre de travaux portant sur une unique région à une date précise. Mais dans le cadre de travaux portant sur tout un pays, l'exercice d'harmonisation des sources paraît extrêmement délicat (voir impossible) et conduirait inévitablement à un résultat fortement lacunaire.

Bien que Vauban ait proposé en 1686 une *Méthode générale et facile pour faire le dénombrement des peuples*, définissant une méthodologie pour le dénombrement⁶, et que Colbert soit à l'origine d'un dénombrement annuel des colonies⁷, l'exercice (déjà rare) est totalement abandonné sous Louis XV, au profit d'une estimation de la population par la méthode des multiplicateurs (la population est estimée en multipliant le nombre de naissances par un coefficient, dont les valeurs choisies variaient entre 20 et 30 [29, 30]), proposée par Leibneiz en 1700 [29, 30, 31]. Mais l'absence de dénombrement sous l'ancien régime n'est pas le corollaire d'une absence de données. La France est dotée d'un registre enregistrant les baptêmes, les décès et les mariages. Le *registre des baptêmes* a été instauré par l'ordonnance de Villers-Cotterêts (1539) [29, 30]. Complétée par l'ordonnance de Blois (1579) imposant l'enregistrement des mariages et des décès [29, 30]. Dans certaines paroisses les registres remontent même jusqu'au XV^e siècle [29, 30]. Ces registres constituent la principale source de données démographiques avant le XIX^e siècle [31]. C'est à partir de ces documents que furent réalisées les différentes estimations de la population antérieures à la révolution. Selon DUPÂQUIER et al. [30, 31] cette source de données n'a été exploitée à des fins scientifiques qu'à partir de la seconde moitié du XVIII^e siècle. Et ce n'est qu'à partir de 1772 que furent constitués des *tableaux annuels du mouvement de la population*, à la demande de l'abbé Terray, contrôleur général des Finances [30], dont l'intégralité n'a pas été publiée (voir retrouvée). L'objectif de ces documents était d'estimer la population totale du pays en appliquant au nombre des baptêmes ou à celui des décès un multiplicateur (toujours suivant la méthode de Leibneiz) *ad hoc*.⁸ Ce dernier était dans la majorité des cas *universel*, (*i.e.* la même valeur s'appliquant sur tout le territoire) ce qui a pu conduire à de nombreuses erreurs d'estimations⁹. L'un des buts de l'enquête de l'abbé Tardy était la construction d'un multiplicateur local. Mais ne dernier n'a visiblement jamais été employé [30].

C'est donc la révolution française qui marquera le point de départ d'une profonde modification de la conception de la statistique. Ainsi, si le choix de la période d'étude a été imposé par la disponibilité des données¹⁰, on peut également comprendre le choix des créateurs de la base de données *cassini.ihess.fr*, qui ont choisi la révolution française, marquant une rupture dans la pratique de la statistique nationale, comme point de départ de leur base de données.

C'est à partir de cette date que l'on donnera la préférence aux dénombrements plutôt qu'à la statistique du mouvement naturel, et que ces derniers seront réalisés de manière fréquente, puis régulière. Mais la nouveauté de cet exercice implique une période trouble, où le dénombrement est encore balbutiant, et que l'on peut aisément qualifier de période *proto-statistique*.¹¹

1. Les censeurs n'étaient élus que tous les cinq ans et ce pour une durée de dix-huit mois, sans réélection possible. De plus une telle charge nécessitait d'avoir été consul. Il s'agissait par conséquent de l'apogée du *cursum honorum*.

2. Le terme de recensement n'a été adopté qu'en 1954 [48]. Pour le reste de ce travail nous utiliserons les deux termes en fonction de l'époque. Toutefois, pour éviter une certaine lourdeur syntaxique, nous avons quelques fois utilisé le terme de recensement pour décrire un ensemble d'opérations, quelle que soit leur date de réalisation.

3. Par le terme de *dénombrement national* nous entendons une opération ayant porté sur l'ensemble du territoire (dans ses limites d'alors).

4. Il existe peu d'informations sur ces opérations. Si il existe des preuves de leur existence [29, 30], aucune donnée n'a été retrouvée.

5. Le dénombrement du Diocèse de Laon en 1684, de la généralité de Limoges en 1686, de la Franche-Comté en 1688, ou de la Bourgogne en 1786, sont quelques exemples (à différentes échelles) des nombreuses opérations effectuées [23, 29, 30]

6. Vauban insiste (par exemple) sur la nécessité d'utiliser un *formulaire* standardisé (qui ne sera adopté qu'en 1876) [29, 30, 31].

7. Un tel exercice n'a pas été mené en métropole. Le pouvoir craignant une réaction négative du peuple, les opérations de dénombrement préfigurant généralement une hausse des impôts [30, 32]. Seul le *dénombrement des feux*, (qui conduira à la publication de *l'atlas des gabelles*) a été effectué en France, ce dernier visant à rééquilibrer les impôts [30, 32].

8. Ce qui nécessite l'hypothèse d'une relative stabilité des taux de natalité et de mortalité [29, 30].

9. En 1784 NECKER [59, P. 221] estime la population française à 24 millions et 676 mille habitants (méthode des multiplicateurs), on l'estime aujourd'hui à 27 millions [30].

10. Il est probable que même dans le cas où nous aurions eut accès à des données antérieures à la révolution française, la quantité de travail nécessaire à leur harmonisation avec les données de *cassini.ihess* aurait pu nous conduire à la non exploitation de ces dernières.

11. Cette dénomination est reprise de l'ouvrage DUPÂQUIER et al. [31].

1.1.1 De la révolution à la création de la SGF

Nous aborderons ici l'histoire de la statistique française et de ces institutions au cours du XIX^e siècle (et au début du XX^e). Cette période verra l'instauration du recensement exhaustif de la population. Suite à une période de transition, différentes institutions en charge de cet exercice seront mises en place. C'est également durant cette période que la méthodologie du recensement sera progressivement mise en place. Contrairement au XX^e siècle, le XIX^e siècle est une période où la statistique est plus un « artisanat » qu'une science. La base de données que nous avons constituée est composée de 21 dénombremments pour cette période¹².

La *proto-statistique* (1789 à 1848)

La période s'étendant de 1789 à 1848, verra la mise en place des premiers dénombremments et le développement de leur méthodologie. Si l'exercice du dénombrement n'est pas nouveau, la mise en place d'une opération exhaustive et régulière elle l'est. C'est durant cette phase que sera mise en place la périodicité quinquennale (à partir de 1822 [31]).

Bien que les avis sur la qualité absolue des dénombremments soient variables, la faible qualité (relative) des huit dénombremments sélectionnés entre 1789 et 1848 fait l'objet d'un consensus [32, 31, 57, 61]. Les informations quantifiées sur les omissions d'une frange de la population sont assez rares, nous contraignant à nous contenter de l'avis qualitatif de différents auteurs. Mais décompte des communes sans données peut compléter ces avis. En effet dans de nombreux cas certaines données sont absentes des fiches communales mises en ligne sur le site *cassini.ehess.fr*. Ces absences, principalement situées dans le premier quart de la série temporelle, sont la résultante de diverses causes, malheureusement très peu documentées par les créateurs de la base de données. Ces derniers ont réalisé une typologie des communes sans données en cinq classes. Toutefois seules trois classes concernent une véritable absence de données, dans les deux autres cas il s'agit de communes recensées avec une autre commune, la population de ces dernières a donc été comptabilisée, mais elle n'est pas associée à la commune en question. Les trois autres catégories concernent les communes pour lesquelles nous n'avons aucune donnée à une date *t*. La première catégorie, la plus fréquente, est celles des communes *absentes*. Dans ce cas les documents contenant la valeur recherchée n'ont pas été retrouvés, voir publiés lors du recensement [48]. La seconde catégorie est celle des communes *lacunaires*. Ces dernières ont été oubliées lors du recensement [48]. La troisième catégorie, qui concerne très peu de communes¹³, et celle des communes *illisibles*. La valeur de la population figurant sur les documents (généralement manuscrits) est tout simplement illisible, les informations présentes sur les documents, n'ont donc pas pu être exploitées [48].

Les huit dénombremments de la période 1793–1848 totalisent 59 % des *communes absentes*, 45 % des *lacunes*, et 70 % des *valeurs illisibles*. Le dénombrement de 1793 est considéré comme celui dont la qualité est la plus discutable, suivi par celui de 1800 [31, 57]. Toutefois il ne s'agit pas des opérations pour lesquelles le nombre de communes sans données est le plus important. C'est surtout du point de vue de leur méthodologie que ces derniers posent problème. Il est avéré que le dénombrement de 1800 est caractérisé par de fortes approximations de la population. Dans certaines régions la population n'a pas été dénombrée, mais estimée de manière non rigoureuse [31]. La qualité des données issues de cette opération varie donc en fonction des régions. Mais, et ce fait est valable pour la majorité des dénombremments / recensements, aucune information ne nous permet d'identifier les régions pour lesquelles la qualité des données et la plus discutable. Les auteurs traitant de l'histoire du recensement traitent généralement le sujet à l'échelle nationale, sans prendre en compte les variations locales de qualité.

Le dénombrement de 1793 (également connu sous le nom de dénombrement de l'an II [48, 55]) à été effectué dans des conditions *quasi-chaotiques*. Sa tenue précède toute tentative d'organisation de l'administration statistique, dont les prémices remontent à 1798. Cette opération a été menée sous la responsabilité du comité de direction qui a mis en place l'opération et traité les données communiquées par les communes [55]. Ce dernier à été mené dans une période où la jeune administration révolutionnaire était difficilement à même de réaliser une telle opération dans les contraintes de temps qu'elle s'était imposée. La période 1789-1799 est extrêmement riche en terme (de tentatives) de dénombrement. M. REINHARD, comptabilise seize décrets imposants la tenue d'un dénombrement de la population. Le résultat de cette grande hâte est « [...] *un immense gâchis* [...], [dont] *il ne subsiste que des épaves* » [31, P. 17]. De toutes les opérations menées durant cette période le dénombrement de 1793 reste celui dont la qualité est la plus élevée et dont les résultats (aujourd'hui archivés aux *Archives Nationales*) sont les mieux conservés. 576 communes sont absentes et 135 sont lacunaires. Il est difficile d'expliquer l'absence de certaines communes dans la base de données, d'autant plus que les auteurs de cette dernière n'ont que très peu documenté

12. 1793, 1800, 1806, 1821, 1831, 1836, 1841, 1846, 1851, 1856, 1861, 1866, 1872, 1876, 1881, 1886, 1891, 1896, 1901, 1906 et 1911.

13. Quarante-trois communes sur l'ensemble de la période.

sa construction. Toutefois MEURIOT [55] nous donne quelques pistes dans son article. Notamment que pour 6 des 561 districts nous n'avons aucune trace d'expédition des données. La carte 2.5a nous montre également qu'une grande partie de la Vendée est composée de communes sans données, absence qui s'explique par les révoltes royalistes secouant la région au lendemain de la révolution française.

Le dénombrement de 1800 est également connu sous le nom de dénombrement de l'an VIII. Il s'agit de l'opération pour laquelle nous avons le moins d'informations. Mais cette dernière est réalisée après un semblant d'organisation de l'appareil statistique. En 1792 le ministre de l'intérieur FRANÇOIS DE NEUFCHÂTEAU a procédé à une réorganisation du ministère et a confié à un bureau de la première division l'ensemble des tâches statistiques [31]. Lorsque LUCIEN BONAPARTE devient ministre de l'intérieur (le 25 décembre 1799) il procède à une nouvelle réorganisation du ministère et un *bureau particulier* en charge du courrier et des archives du ministère est créé [31]. Ce dernier appartient toujours à la première division du ministère de l'intérieur. Le *bureau particulier* est scindé en deux entités à partir de 1800, dont la seconde section dirigée par ALEXANDRE DE FERRIÈRE prend le nom de *bureau de statistique* en 1801 [31]. Ce petit service, alors composé de huit membres [31] est indirectement l'ancêtre de l'INSEE d'aujourd'hui.

Le dénombrement de 1806 est d'une qualité bien supérieure aux deux précédents [31]. Il s'agit de la seule opération réalisée sous le premier empire. Seules 23 lacunes sont identifiées. Les données n'ayant été publiées qu'à l'échelle des arrondissements [57], les données de ce recensement ont été récupérées à partir de documents manuscrits dont les copies sont stockées aux archives nationales, ou au service historique de l'armée de terre (en fonction des départements)¹⁴. Selon MOTTE, SÉGUY et THÉRÉ [57], les données peuvent différer suivant leur source (Archives départementales ou nationales). Les données ici utilisées proviennent des archives nationales (et du service historique de l'armée de terre), contrairement à celles publiées dans les dictionnaires des *Paroisses et Communes de France* [48, 57].

Il n'y eut aucun dénombrement national entre 1806 et 1821. Toutefois la population fut estimée en 1811 à l'échelle départementale par une actualisation du dénombrement de 1806 à l'aide de données sur le mouvement de la population [31]. Pour cette raison, les chiffres de 1811 n'ont pas été sélectionnés lors de la construction de la base de données [48]. Il est toutefois nécessaire de préciser que certains préfets prirent l'initiative d'un véritable dénombrement [57]. En 1815 le bureau de la statistique fut supprimé et la publication des statistiques fut pendant un temps (jusqu'en 1835) à la charge des différents ministères [31]. THIERS, alors ministre du commerce, proposa en 1833 un « *projet de diffusion de documents statistiques* » [32, P. 34] dont l'agrégation sera réalisée au sein du ministère du commerce. Cette tâche fut confiée au deuxième bureau du conseil supérieur du commerce et la première publication du bureau, l'« *Essai des travaux de statistiques qui doivent être publiés par le ministère du Commerce* », a été réalisée en 1835 [31]. C'est par un arrêté ministériel de 1840 que le deuxième bureau fut transformé en bureau de la statistique générale de la France SGF qui fut dans un premier temps cantonné à la centralisation des documents en vue de leur publication [31].

Le dénombrement de 1821 est réalisé suivant les mêmes modalités que celui de 1806. En effet, malgré le changement de régime de 1815¹⁵ la statistique de la population reste à la charge de la direction de l'administration générale des départements. Il est avéré que de nombreux préfets se sont contentés de reprendre les chiffres issus d'opérations menées en 1817 et 1818¹⁶ [31]. Les résultats seront authentifiés par l'ordonnance du 16 janvier 1822 (cf. annexe B.2, page 96), celle-là même qui prescrira la périodicité quinquennale¹⁷ [32, 31, 57]. Les dénombrements ultérieurs seront tous réalisés (sauf cas particulier) lors de la première et de la sixième année de la décennie. Comme pour la précédente opération (et pour les mêmes raisons, à savoir l'absence de publication), les données à l'échelle communale sont (une fois encore) conservés sous la forme de documents manuscrits conservés aux archives nationales et à la bibliothèque municipale de Rouen [48, 57]. La rubrique des *sources* du site *cassini.ehess* ne précise pas si l'une des sources a été privilégiée. Toutefois, il est précisé dans le dictionnaire d'histoire administrative, *Communes d'hier, communes d'aujourd'hui : les communes de la France métropolitaine, 1801-2001*, (réalisé par la même équipe) que ce dernier n'utilise que les données de la bibliothèque de Rouen, ce qui laisse supposer qu'il en va de même pour la base de données exploitée.

Suivant la prescription de l'ordonnance précédemment citée il y eut un dénombrement en 1826. Mais ce dernier n'est en réalité qu'une mise à jour des données de 1821 [48]. La méthode de rectification employée consistait à ajouter à la population recensée en 1821 le solde du mouvement naturel de la période intercensitaire (1821-1826) [31]. C'est dans une circulaire de 1826 que le ministre de l'intérieur

14. Les originaux sont quant à eux conservés aux archives départementales [57].

15. Seconde abdication de Napoléon I^{er} et le retour au pouvoir de Louis XVIII [19].

16. Suite à la mise en place de la loi de finances du 28 avril 1816, fixant, entre autre, les droits d'entrée des boissons dans les villes au prorata de la population, de nombreux préfets ont, de leur propre initiative, procédé à des opérations de dénombrement, les chiffres de 1811 n'étant, comme signalé précédemment, qu'une réactualisation de ceux de 1806 [31].

17. Pour être plus précis cette ordonnance ne prescrit pas textuellement la *périodicité quinquennale*. Mais c'est son interprétation, et plus particulièrement celle de l'article premier qui conduira à la mise en place de la périodicité quinquennale.

jugea que le dénombrement de 1821 était « assez récent pour que l'administration puisse, par les moyens qu'elle a à sa disposition, rectifier [...] les tableaux dressés à cette époque, et suppléer ainsi une opération générale qui entraînerait des lenteurs, des embarras, et qui pourrait être fatigante si elle devait désormais se reproduire tous les cinq ans » [31, P. 31]. Comme pour le dénombrement de 1811, la nature contestable de l'opération de 1826 a conduit les créateurs de la base de données *cassini.ehess* à l'ignorer [48].

Le changement de dynastie royale (1830) conduit à certaines modifications de l'appareil statistique français. Le bureau de la statistique, auparavant rattachées au ministère de l'intérieur, passe sous l'égide du ministère du commerce et des travaux publics [31]. Un nouveau dénombrement est rapidement ordonné par le ministre du commerce (Antoine d'Argout) qui relève dans la circulaire du 25 mars 1831, l'insuffisance des résultats estimés en 1826 et la nécessité de procéder à une nouvelle mesure de la population. Selon NOIN [61], le dénombrement de 1831 est la première opération de bonne qualité. Toutefois on recense 1 631 communes absentes (cf. carte 1.2b), ce qui en fait une des opérations où le nombre de communes sans données est le plus important. À partir de cette date (et jusqu'en 1872), les données proviennent des archives départementales. Seules les communes « ayant une population totale de trois mille habitants et au-dessus, ou une population agglomérée de quinze cent habitants et au dessus » (le seuil sera porté à deux mille habitants en 1846) voient leur population publiée au *bulletin des lois* [57]. La période 1831–1872 correspond de plus à la période où le nombre de communes sans données est maximal. Il est fort probable qu'une partie des absences de cette période soient explicables une variation de la qualité de conservation des données en fonction du département, mais, une fois encore il ne s'agit que d'une hypothèse, les auteurs de la base de données ne donnant aucune information permettant de la valider.

En 1836 est réalisé un nouveau dénombrement de la population française. Ce dernier a lieu cinq ans après le dénombrement de 1831, comme préconisé par l'ordonnance du 16 janvier 1822. La plus grande originalité de cette opération est le recours à une nouvelle définition de la population (ce qui, contrairement à d'autres transformations, impacte probablement notre travail). La circulaire du 10 avril 1836 (prescrivant la tenue d'un nouveau dénombrement) indique que la population recensée sera la *population de droit*¹⁸ et non plus *de fait*, utilisée jusqu'à présent. Cette dernière est définie de la manière suivante :

« Les individus de tout âge et de tout sexe, habitant ou domiciliés dans les communes, même ceux qui en seraient temporairement absents pour raison de service militaire ou tout autre motif ». [31, P. 31]

Mais on revient à la population de fait dès le dénombrement de 1841 comme le ministre de l'intérieur (Tanneguy Duchâtel [20]) l'indique dans la circulaire du 23 avril 1841 :

« De nombreuses réclamations, des omissions innombrables et des difficultés de diverses natures m'ont démontré qu'il était indispensable de revenir au domicile de fait [...] » [31, P. 31]

Malgré cette nouvelle définition, le dénombrement de 1836 reste une opération menée avec une grande rigueur. C'est à cette date que fut employé pour la première fois une ébauche de bulletin individuel (seulement à Paris). Les opérations de vérification ont également été soigneusement réglementés, ce qui se traduit par une bonne qualité des données [31].

La qualité du dénombrement de 1841 est quant à elle beaucoup plus discutable. Le retour à la population de fait et la volonté des services fiscaux d'améliorer l'exactitude de l'estimation de la population (et l'ajout des garnisons à la population communale) a fait craindre un fort accroissement de la valeur de la population, conduisant nécessairement à une augmentation des différents impôts. Les agents des services fiscaux ont par conséquent reçu la mission de vérifier les données transmises par les maires et le cas échéant à renouveler l'opération. « La méfiance populaire en fut exacerbée, et des troubles éclatèrent dans le Midi, notamment à Toulouse » [31, P. 33]. L'opération fut par conséquent plus longue qu'à l'accoutumée. Il est par conséquent préférable, selon DUPÂQUIER et al. [31] d'utiliser les résultats « avec précaution », d'autant plus qu'une partie des listes nominatives (50 %), permettant une vérification des résultats, a été perdue.

Pour éviter les heurs du dénombrement de 1841, la population des établissements publics (garnisons, prisons, asiles, etc.) estimée lors du dénombrement de 1846 n'est plus prise en compte lors du calcul des impôts. La définition de la résidence est modifiée, et les personnes qui ne sont pas légalement domiciliées dans la commune, mais qui y résident la majeure partie du temps sont considérées comme domiciliées dans leur commune de résidence [31].

Le dénombrement sous la seconde république et le second empire (1851 à 1870)

Le début des années 1850 marque un profond changement dans l'organisation administrative du dénombrement. Là où précédemment, un embryon de SGF n'avait à sa charge que la centralisation et la

18. Personnes ayant leur domicile dans la commune, qu'elles y soient présentes ou non le jour du dénombrement [44].

publication des statistiques produites par différents ministères, le dénombrement de 1851 est entièrement organisé par le ministère de l'intérieur, et la jeune SGF est entièrement réorganisée (1852) [31].

Malgré ces changements administratifs, la méthodologie du dénombrement de 1851 ne diffère que peu avec celui de 1846. Quelques précisions méthodologiques sont apportées, notamment pour le dénombrement des ouvriers en charge de la construction des chemins de fer (dénombrés avec la population flottante de la communes, à jour fixe) [29]. Un des éléments notables, et que, selon les estimations de BERTILLION [9], un grand nombre d'enfants en bas âge auraient été omis (environ 400 000 enfants entre 0 et 5 ans). Toutefois, selon DUPÂQUIER et al. [31, P. 37] : « *Le recensement de 1851 fait date dans l'histoire de la statistique, non seulement par sa qualité, mais surtout parce-qu'il fait l'objet d'une exploitation et d'une publication beaucoup plus complètes que les précédentes* ». Mais cet avis élogieux contraste fortement avec nos propres observations. Le nombre de communes *absentes* s'élève à 1 759 ce qui fait du dénombrement de 1851 l'opération pour laquelle nous avons le moins d'individus à notre disposition. Nous n'avons pas de données pour les actuels départements de Savoie et de la Haute-Savoie ainsi que pour le nord des départements de la Meurthe et Moselle, de la Moselle, ou la partie orientale des Alpes-Maritimes.

En 1852 la SGF est entièrement réorganisée et placée sous la direction d'ALFRED LEGOYT. Durant sa direction (1853–1870) la SGF a joui d'une situation d'autonomie vis à vis de son ministère de tutelle, le ministère de l'agriculture et du commerce [31].

En 1856 a lieu le premier dénombrement faisant suite à la réorganisation de la SGF de 1852, et par conséquent de la première opération menée sous la tutelle d'ALFRED LEGOYT. Suite à la demande du ministre de l'agriculture, du commerce et des travaux publics on préconise l'utilisation des premiers *bulletins de ménage*, dont il ne subsiste aucune archive. Concernant le décompte de la population, on note un sous-enregistrement des enfants en bas âge [31] comme lors du dénombrement de 1851, auxquels vient s'ajouter l'oubli des hommes de l'armée de Crimée [31]. Selon DUPÂQUIER et al. [31, P. 39] l'oubli concerne 165 000 hommes, là où FIGES [36] estime que l'armée de Crimée était composée de 310 000 français, avec des pertes humaines s'élevant à environ 100 000 hommes, ce qui laisse supposer un oubli plus important que celui estimé par DUPÂQUIER et al. [31] en 1988. Si le nombre de communes sans données est inférieur à celui du dénombrement de 1851, un grand nombre de communes manquent à l'appel. On recense 467 communes absentes et 486 communes lacunaires.

Le deux dénombrements suivants (1861 et 1866) possèdent les mêmes caractéristiques que le dénombrement de 1856. Ces derniers sont également caractérisés par l'oubli d'une partie de l'armée française (non quantifiée dans les différents ouvrages consultés). En 1861 l'armée française en Algérie n'est pas comptabilisée, et en 1866 viennent s'y ajouter les troupes présentes en Italie et au Mexique [31]. On a comptabilisé respectivement 442 lacunes en 1861 et 37 en 1866.

Quatre dénombrements auront été réalisés sous le second empire. Bien que le dénombrement de 1851 soit considéré d'excellente qualité selon DUPÂQUIER et al. [31] il s'avère que ce dernier, comme l'ensemble des dénombrements de la période, est riche en lacunes et en communes absentes (cf. tableau 1.2 et carte 2.5), ce n'est qu'à partir du dénombrement de 1876 que l'on observe une réelle diminution du nombre de communes sans données. C'est également à la fin du second empire que la SGF est à nouveau réorganisée, elle perd son autonomie et son directeur en 1870 [31].

Le dénombrement de 1870 à 1914

La fin du XIX^e siècle, est une période charnière dans l'histoire du dénombrement en France (et plus généralement de la statistique). C'est durant cette période que sera adopté le bulletin individuel, mais également la mécanisation ainsi que la centralisation du traitement des résultats. Au vu des avancées effectuées durant ces quarante ans, on peut considérer que c'est cette période, qui marque l'entrée de la SGF dans le XX^e siècle. Cette mutation est avant tout l'oeuvre d'un homme, LUCIEN MARCH, qui sera chef (1907 à 1910), puis directeur (1910 à 1920) de la SGF (il a intégré l'institution en 1896¹⁹) [4, 26]. Si l'adoption du bulletin individuel a précédé sa direction, il sera l'instigateur de la centralisation du traitement des résultats, et le chantre de la mécanisation de ce même traitement. Sous sa direction les effectifs de la SGF seront multipliés par 5 (passant d'une vingtaine à plus de 100) [26].

Le premier dénombrement de cette période (sur un total de 9) est celui de 1872, initialement prévu pour l'année 1871, mais décalé à cause du conflit franco-allemand de 1870 [31]. C'est lors de cette opération que certains préfets conseillent aux maires l'utilisation d'un bulletin individuel, toutefois son utilisation n'est pas encore imposée. Il faudra attendre le dénombrement de 1876 pour que le bulletin individuel devienne le document de base pour l'ensemble du territoire. Contrairement aux dénombrements ultérieurs, ce dernier est alors rempli par l'agent recenseur en présence de l'intéressé [31]. On notera également, bien que ce point n'ait aucune influence sur les données que nous traiterons, que l'on ne demande plus l'âge des individus, mais leur année de naissance, ce qui permet de limiter le phénomène d'*attraction des nombres*

19. Notons que LUCIEN MARCH était chef des services techniques de la SGF lors du dénombrement de 1896 [4].

ronds [33]. De plus, dans une volonté de minimiser l'effet des migrations saisonnières sur le dénombrement de la population, l'opération est effectuée à la fin de l'année. C'est également à partir de cette date que le noms des communes est défini officiellement et qu'est changée de logique de publication. La SGF publie désormais les résultats par communes [57], dont les archives sont intégralement conservées à l'Insee et ce pour tous les dénombrements (et recensements) suivants. On notera que le recensement de 1876 marque une diminution (définitive) du nombre de communes *absentes, lacunaires ou illisibles*, (ce dernier passe de 467 communes *absentes*, et 233 communes *lacunaires* en 1872, à 13 communes *illisibles* en 1876) (cf. tableau 1.2). Ce qui est très probablement dû à la mise en place d'une toponymie rigoureuse (à cause du peu d'information portant sur la construction de la base de données *cassini.ehess*, nous ne pouvons que supposer ce lien.) Selon DUPÂQUIER et al. [31] le dénombrement de 1876 est l'un des meilleurs du XIX^e siècle, la grande rigueur de la méthodologie ayant permis une importante diminution des erreurs.

Le remplissage des bulletins individuels par l'agent recenseur est abandonné dès le dénombrement de 1881. Ce dernier se contente désormais de vérifier et de préciser les réponses. Selon DUPÂQUIER et al. [31] La qualité de l'ensemble du dénombrement s'en ressent. Le rôle de l'agent recenseur est réduit à la récupération des bulletins, précédemment remplis [31]. Mais ce sacrifice à la qualité est nécessaire pour procéder à l'intégralité du dénombrement à date fixe (toute la population est concerné, contrairement aux dénombrements précédents où seule la population flottante était dénombrée à date fixe [31, 37]). L'intégralité du recensement est effectuée le dimanche 18 décembre 1881 [31]. Le dénombrement de 1881 voit également la mise en place d'un nouveau système de définition de la population. Deux populations sont désormais comptabilisées, la *population légale* (personnes présentes et absentes résidant habituellement dans la commune) et la *population présente*, ce système sera conservé jusqu'en 1946 [44]. Comme pour l'opération précédente, en 1886 toute la population est recensée le même jour, à savoir le dimanche 30 mai [31]. La synthèse d'une majorité des résultats n'est effectuée qu'à l'échelle départementale, la population légale est la seule variable disponible à l'échelle communale. Si ce changement n'influe pas notre variable de travail, il en résulte tout de même un appauvrissement global de la statistique [31]. 1891, le recensement a lieu le 12 Avril. Les individus absents de la commune le jour du recensement le sont plus portés sur la liste nominative de la commune où réside leur famille, mais de celle où ils se trouvent [31].

En 1896, le recensement est mené le 29 mars. Il est couplé à une enquête portant sur les professions [31]. Celle-ci sera traitée sous l'égide du ministère du commerce (traitement centralisé à Paris, sous la direction de LUCIEN MARCH (alors ingénieur) avec l'aide de machines électromécaniques construites par HOLLERITH (notamment utilisées pour traiter le recensement américain de 1896). Cette étape ouvre la voie à la mécanisation intégrale du dénombrement de 1901 [26].

Après 1896, le personnel employé par le ministère du commerce est rattaché (d'abord provisoirement, puis définitivement en 1905 [31]) à la SGF, et la décision est prise d'appliquer la même méthodologie (dépouillement centralisé et mécanisé) lors du recensement de 1901 [31] (cf. tableau 1.2). Ainsi est mise en place une exploitation mécanographique intégrale à l'aide des classi-compteurs-imprimeurs de Lucien March [4]. Ces derniers sont encore plus efficaces pour traiter le dénombrement, il ont en effet été conçus à cet effet et permettent l'impression directe de tableaux statistiques contrairement aux machines de HOLLERITH. Cette mécanisation permet une augmentation conséquente du nombre de questions posées [31]. Le dénombrement est effectué le 24 mars et est entièrement traité à Paris. Le premier recensement du XX^e siècle marque un tournant capital, il s'agit de la première opération dont le dépouillement est intégralement centralisé et mécanisé, et les avancées de cette opération seront exploitées pour toutes les opérations ultérieures.

Le dénombrement de 1906 est le premier à être réalisé sous la direction du LUCIEN MARCH [31]. Toute la population est recensée à une date fixe (le dimanche 4 mars), comme toutes les opérations depuis celle de 1881 [31]. Une nouvelle méthodologie rigoureuse est mise en place [31]. Le maire divise la ville en secteurs de recensements qui sont à la charge d'un agent recenseur. De plus on introduit un contrôleur pour chaque groupe de 2000 habitants. Deux jours avant la date du recensement l'agent recenseur dépose les documents dans chaque maison, il les récupère ensuite le jour du recensement et il doit avoir effectué cette opération sous les trois jours. Les feuilles de ménages et les bulletins individuels sont ensuite groupés dans « le bordereau de maison » lequel est rempli par l'agent recenseur [31]. Ces documents sont ensuite transmis à la mairie là où les contrôleurs vérifient l'absence de lacunes. Les bulletins individuels sont ensuite séparés en quatre groupes (sans profession, population comptée à part, chômeurs, actifs) les bulletins concernant les actifs travaillant dans une autre communes sont transmis à cette dernière. Les documents sont ensuite transmis à la préfecture, qui les transmet, après vérification au ministère de l'intérieur. Une méthodologie identique sera appliquée pour le dénombrement de 1911, mais du fait de la première guerre mondiale, l'exploitation des données et les publications qui en résultent seront plus limitées [31]. Notons également que le dénombrement de 1911 est le seul de la période 1881–1914 pour lequel une commune est absente (cf. tableau 1.2).

La fin du XIX^e siècle et le début du XX^e siècle est une période d'importantes modifications de la méthodologie du recensement en France. Avec l'introduction successive des bulletins individuels, de la

mécanisation et la mise en place d'une nouvelle méthode rigoureuse la qualité du dénombrement à considérablement évolué durant cette période [31]. Et cette augmentation de la qualité est notamment visible avec la diminution du nombre de communes sans données (cf. tableau 1.2). À partir de 1876 le nombre de ces omissions devient nul, à l'exception des dénombrements de 1876, 1911, 1926 et 1946, totalisant 9 lacunes et 13 valeurs illisibles.

1.1.2 De la première guerre mondiale à la fin de la périodicité quinquennale

Cette seconde phase de l'histoire du recensement français verra la réalisation de 12 dénombrements. Durant cette période trois institutions statistiques se sont succédées et l'appareil statistique français à connu de nombreuses mutations, notamment technique. Si la SGF est entrée dans le XX^e siècle avec MARCH, une nouvelle phase commence lorsque ce dernier quitte l'institution en 1922.

De la SGF à L'INSEE

Là où la statistique du XIX^e siècle était organisée autour d'une seule institution, trois institutions ont été chargées du dénombrement de la population entre 1914 et 1999. La première de ces institutions est la SGF créée en 1833 et qui perdurera jusqu'au 11 octobre 1941 [32]. Celle-ci se retrouve absorbée par une entité plus large, le *service national des statistiques* (SNS) dirigé par RENÉ CARMILLE. Cette nouvelle institution sera remplacée par l'INSEE à partir de 1946, la SNS étant une institution de la France de Vichy [32]. Les changements opérés durant cette période sont très importants, et il nous est nécessaire d'en parler. Comme dit précédemment la première partie du XX^e siècle (avant la première guerre mondiale) a été capitale en terme d'améliorations de la pratique du dénombrement. Sous la direction de LUCIEN MARCH une méthodologie rigoureuse a été mise en place, et la mécanisation du dépouillement permise par le classificateur-imprimeur de son invention a permis de centraliser le dépouillement qui était précédemment à la charge des communes (cf. tableau 1.1) [26, 32, 31]. L'apport de MARCH, qui prend sa retraite en 1920 [32] est capital, et sous sa direction l'effectif de la SGF sera multiplié par cinq (de 20 à 100) [32], et il faudra attendre 1941 et la direction de RENÉ CARMILLE pour voir des mutations similaires [26].

Le 14 novembre 1940 est créé un service de la démographie directement rattaché au ministère des finances [32]. L'objectif de cette institution, dont la direction est confiée à RENÉ CARMILLE, est de « *préparer les conditions d'une nouvelle mobilisation, au cas où la France reprendrait la lutte contre l'Allemagne.* » [32, P. 16]. Ce service est doté d'un effectif de 925 emplois (la SGF possédait quant à elle 200 employés au début de la guerre [32]), et cherche à effectuer « *un recensement continu de la population [...]* » [32, P. 16]. Très vite, (juillet 1941) des conflits administratifs apparaissent entre la SGF et le service de la démographie (rattachés au même ministère), notamment lorsque ce dernier cherche à obtenir auprès des mairies les bulletins d'état-civil des soixante dernières années (pour constituer la base de ce qui sera le *répertoire national d'identification des personnes physiques* [32]). Pour résoudre ce « *conflit d'attribution* » la décision est prise (le 11 octobre 1941) de fusionner la SGF et le service de la démographie dans une nouvelle institution, le *service national des statistiques* dont la direction est confiée à RENÉ CARMILLE [32]. De façon similaire à ce qu'avait réalisé LUCIEN MARCH en 1922 en participant à la création de l'ISUP en 1922, RENÉ CARMILLE sera à l'origine de la création d'une école chargée de la formation des fonctionnaires de la SNS, la future ENSAE [26, 32]. Les activités de CARMILLE ne s'arrêtent pas avec l'invasion de la zone libre en 1942, même si cette dernière rend toute mobilisation impossible [32]. CARMILLE continue à perfectionner la SNS qui devient une énorme institution. Toutefois CARMILLE est arrêté en 1944 pour ces activités de résistant et est déporté à Dachau où il décédera le 25 janvier 1945 [26, 32]. La courte direction de CARMILLE a été capitale dans l'histoire de la statistique française, MARCH avait fait passer l'effectif de la SGF de 20 à une centaine de personnes [26, 32], sous sa direction l'institution avait perdu son aspect quasi-artisanal, pour devenir une véritable institution scientifique. Sous la direction de CARMILLE les effectifs sont passés de 200 à 7 000 hommes, et les classificateurs-imprimeurs de MARCH, vieux de quarante ans ont été remplacés par du matériel plus moderne et plus efficace [26, 32]. Selon DESROSIÈRES, MAIRESSE et VOLLE [26, P. 6], CARMILLE a « *puissamment contribué à créer le système statistique moderne* » à l'aide de sa compréhension des possibilités techniques de l'outil mécanographique. DUPÂQUIER et al. [32, P. 18] est encore plus éloquent, en modernisant le matériel et les méthodes, en améliorant la formation des fonctionnaires et leur effectif « *RENÉ CARMILLE [se place comme] le véritable fondateur de l'INSEE.* »

Mais il faudra attendre 1946 pour que l'INSEE soit fondée. Elle l'est officiellement par la loi du 2 avril 1946. La question de la suppression de la SNS avait été abordée à la libération, mais elle est sauvée par « *l'idéologie planificatrice et la nécessité de mettre enfin sur pied une comptabilité nationale* » [32, P. 18]. Mais l'énorme effectif de la SNS est peu à peu réduit, de 7 000, en 1945, il passe à 3 000 en 1950 et à 2 700 en 1960 [26]. Comme la SGF et la SNS avant elle l'INSEE subira une importante modernisation notamment au cours des années soixante, où sous la direction de CLAUDE GRUSON seront employés les premiers

ordinateurs (notamment pour le recensement de 1962) [26]. C'est également sous la direction de GRUSON (1961-1967) [26] que les effectifs de l'INSEE (qui avaient atteint leur minimum depuis la libération en 1960) commencent à augmenter, il passent à 4000 hommes en 1970 et à 7 000 en 1976 [26].

Au cours de cette partie nous avons cherché à montrer à quel point le XX^e siècle est riche en transformations de l'appareil statistique français. Pour DESROSIÈRES, MAIRESSE et VOLLE [26] il est possible de résumer cette période, et plus généralement l'histoire de l'INSEE par les trois directeurs que nous avons cités. Il est vrai que leur travail est remarquable, et que tous ont œuvré dans le même sens. DESROSIÈRES, MAIRESSE et VOLLE [26] remarque par ailleurs qu'ils présentent de fortes similitudes, aussi bien dans leur formation (tous sont polytechniciens sans formation particulière pour les statistiques), que dans leur direction. Chacun à son époque c'est fait chantre de la modernisation des outils de travail (introduction de la mécanisation pour MARCH, remplacement des classi-compteurs-imprimeurs pour CARMILLE et informatisation pour GRUSON), de l'amélioration de la formation (ISUP, ENSAE) et c'est également sous ces directions qu'on eut lieu les plus grandes phases de recrutement [26]. Ainsi, comme on peut s'y attendre l'organisation des recensements tout au long de cette période sera elle aussi grandement modifiée.

Les dénombremments de 1921 à 1946

Le dénombrement de 1916 est la première opération à avoir été annulée depuis la mise en place de la périodicité quinquennale. La première guerre mondiale a rendu la tenue de cette opération impossible. Après la fin de la première guerre mondiale, le rythme précédemment fixé est suivi à nouveau, et le premier dénombrement d'après guerre a lieu en 1921 [32]. La durée qui sépare ce dénombrement du précédent est l'une des plus longues de la série statistique. Peu de faits notables sont à retenir de cette opération, celle-ci reste similaire aux opérations d'avant-guerre. L'organisation de ce dénombrement est similaire à celle des premiers dénombremments du XX^e siècle, telle qu'elle avait été mise au point par LUCIEN MARCH en 1901 [32]. Le dépouillement reste centralisé par la SGF qui continue à traiter les formulaires à l'aide des classi-compteurs-imprimeurs de MARCH. Si nous disposons de l'intégralité des communes pour ce dénombrement, ce dernier semble touché par un certain nombre de sous-enregistrements frauduleux [32]. La qualité des dénombremments de 1901 à 1936 a été estimée par HENRY [40] en 1948 à l'aide d'une méthodologie basée sur la comparaison de la masculinité des trois premières classes d'âge avec la masculinité théorique, laquelle est calculée « à l'aide des taux de survie, à partir de masculinité à la naissance. » [40, P. 7]. Le constat HENRY [40] est éloquent, si entre 1901 et 1911 aucun département ne présente d'anomalie, il obtient des résultats aberrants pour 4 départements en 1921, 3 en 1936, 6 en 1931 et 9 en 1936 [32, 40]. Cette méthode montre clairement une dégradation de la qualité des recensements pendant l'entre deux guerres, les raisons sont probablement multiples, il peut aussi bien s'agir d'erreurs que de fraudes. DUPÂQUIER et al. [32, P. 34] note que « la répartition des communes en fonction de leur population présente des anomalies : les fréquences observées juste au-dessus de certains seuils, liés à des avantages administratifs ou financiers sont trop fortes [...] ». C'est notamment le cas des villes de Lyon²⁰, de Marseille²¹, ou encore de l'ensemble de la Corse²². Si ces sur-enregistrements frauduleux ont pour la majorité cessé à partir du dénombrement de 1946, la Corse connaîtra un sur-enregistrement endémique jusqu'en 1975 [32]. Il est délicat d'expliquer les raisons de cette baisse globale de qualité durant la période 1921-1936, mais une composante importante de cette baisse de qualité est la multiplication des sous-enregistrements frauduleux. Les motivations de ces derniers sont multiples, il peut aussi bien s'agir de raisons financières, que d'une question de prestige, notamment pour les villes de Marseille et de Lyon qui se sont disputés le titre de deuxième ville de France [32]. Mais ces fraudes ne sont rendues possibles que par l'organisation des dénombremments de cette période [32]. Contrairement à des opérations ultérieures, que l'on peut qualifier de *monolithiques*, (notamment le recensement de 1999) pour lesquelles l'INSEE avait un contrôle total des opérations, la SGF n'avait (et ce depuis 1901) à sa charge que l'organisation du recensement, le dépouillement, le traitement, et la publication des résultats, la collecte et la vérification des résultats n'étant pas à sa charge (cf. tableau 1.1). Les maires avaient donc la possibilité de manipuler aisément les données, et les importants sur-enregistrements frauduleux montrent que ces derniers ne s'en sont pas privés [32].

Conformément à la périodicité quinquennale le dénombrement suivant est réalisé en 1926. Comme précédemment cette opération est entachée des mêmes fraudes que le dénombrement de 1921. Mais la qualité de cette opération est vraisemblablement plus faible que celle du recensement de 1921. Si l'on compare la population légale (*i.e.* résultant du dénombrement) et la population légale rectifiée par l'INSEE

20. La population de Lyon a été volontairement surestimée de 58 800 habitants en 1911, de 93 500 habitants en 1921, de 97 800 habitants en 1926, de 120 800 habitants en 1931 et de 112 600 habitants en 1936 [32].

21. La population de Marseille aurait été surévaluée de 52 000 habitants en 1926, de 191 000 habitants en 1931, et de 294 000 habitants en 1936 [32].

22. La population de la Corse était officiellement de 281 940 habitants en 1921 et de 322 854 habitants en 1936, alors que l'excédent naturel sur la même période est de 8 500 [32]. Selon DUPÂQUIER et al. [32] la population de l'île n'aurait probablement pas dépassé les 200 000 habitants durant cette période.

dans les annuaires rétrospectifs [32] on observe que l'écart mesuré en 1926 (population surestimée de 174 mille habitants) et plus important qu'en 1921 (102 mille habitants) ce qui indique une baisse de la qualité du dénombrement entre 1921 et 1926 (*n.b.* la rectification opérée par l'INSEE ne prend pas en compte les fraudes corses [32]). À ces fraudes viennent s'ajouter la présence de sept communes lacunaires, seules absences de données pendant l'entre-deux guerres.

De la même façon qu'entre 1921 et 1926, la qualité du dénombrement de 1931 est-elle inférieure à celle du dénombrement de 1926 ? Tout semble l'indiquer. Si, comme précédemment on compare la population initiale et celle corrigée ultérieurement on remarque que l'écart est désormais de 311 mille habitants, c'est à dire presque le double qu'entre 1921 et 1926. De plus HENRY [40] trouve des résultats du taux de masculinité aberrants pour six départements. Ces deux indicateurs confirment la baisse de la qualité. Le même constat est valable pour le dénombrement de 1936, l'écart entre population corrigée et population non corrigée est de 405 mille habitants, HENRY [40] estime quant à lui que neufs départements présentent des valeurs aberrantes.

Un nouveau dénombrement était prévu en 1941, mais ce dernier n'aura pas eut lieu à cause de la seconde guerre mondiale [44]. Toutefois il y eut une opération de dénombrement en 1941 menée par la SNS sous la direction de RENÉ CARMILLE mais cette dernière ne portait que sur les activités professionnelles et était restreinte à la zone libre, mais aucune publication des résultats n'a été effectuée [32]. Le prochain dénombrement aura lieu après-guerre, en 1946.

Durant l'entre deux guerres quatre dénombremments ont été effectués. Cette période est intéressante car tous les dénombremments réalisés sont sujets aux mêmes défauts. Toutes les opérations sont sujettes à des sur-enregistrements frauduleux massifs et de plus en plus présents. Le travail de HENRY [40] montre par ailleurs que cette situation commence avec le dénombrement de 1921, aucune anomalie (à l'échelle départementale) n'étant relevée pour la période 1901–1911. Nous avons expliqué que cette situation était permise par l'organisation des dénombremments, qui laisse aux maires la possibilité de manipuler les résultats. Mais, comme l'indique le tableau 1.1, tous les dénombremments du début du XX^e siècle sont organisés d'une manière similaire, les fraudes étant donc tout à fait possibles. Ainsi il nous est impossible d'expliquer l'apparition soudaine de fraudes massives, et c'est par ailleurs un point non traité dans tous les ouvrages que nous avons consultés sur ce sujet. La seule conclusion que l'on peut retenir est que la période de l'entre deux-guerres est caractérisée par d'importantes fraudes. Cette situation contraste avec les recensements ultérieurs (1946–1999) qui sont quant à eux sujets à de nombreux sous-enregistrements [32].

1.1.3 Le recensement après 1946

De nombreuses modifications de la pratique du recensement eurent lieu entre 1946 et 1999. Il est notamment intéressant de constater qu'aucun des recensements effectués durant cette période ne respecte la périodicité quinquennale, en vigueur depuis 1822. Certes, il existe des précédents, à savoir les dénombremments de 1872 (déplacé), de 1916 (non effectué) et de 1941 (non publié), mais pour ces trois opérations le décalage à été justifié par une situation de guerre ou d'occupation [31, 32]. Alors qu'à partir de 1946 le non respect de la périodicité quinquennale concerne toutes les opérations et ce retrouve justifié par de nouveaux arguments, généralement financiers [32]. Au final seul le dénombrement de 1946 sera effectué l'année prévue par le rythme préalablement respecté [32]. Au final huit opérations de recensement seront réalisées durant cette période de cinquante-trois ans, soit trois de moins qu'en cas de respect de la périodicité quinquennale.

L'abandon de la périodicité quinquennale

Le dénombrement de 1946 est la première opération réalisée après la seconde guerre mondiale. Cette opération est particulière, puisqu'elle à été réalisée par deux institutions successives, la SNS pour la production des données, et l'INSEE pour leur analyse. Cette opération a en effet été réalisée le 10 mars 1946 [12, 37], alors que l'INSEE à été crée par une loi du 2 avril 1946 [32]. Selon DUPÂQUIER et al. [32] ce changement est avant tout un changement de sigle visant à recrédibiliser une institution « *qui apparaît trop lié[e] à la lourde gestion bureaucratique du temps de Vichy* » [26, P. 23]. l'institution reste liée à la SGF et à la SNS dont elle hérite les méthodes et les effectifs [26, 32]. Le calcul de la population est légèrement modifié pour cette opération, comme pour les opérations précédentes (depuis 1881) de multiples définitions sont employées et l'on continue à distinguer la *population légale* (celle présentée dans la base de données *cassini.ehess.fr*) de la *population présente* (cf. tableaux A.1 et A.2) [44]. Mais le calcul de la *population légale* est modifié par l'inclusion de la population *momentanément hors métropole*, principalement des militaires en Allemagne, dont l'effectif s'élève à 312 000 individus [44]. Convenant la qualité des données, il s'agit du dernier dénombrement pour lequel une commune est lacunaire (cf. tableau 1.2). Toutefois cela n'indique pas pour autant que l'estimation de la population est optimale. Selon DUPÂQUIER et al. [32], le

dénombrement de 1946 est caractérisé par un certain nombre de sous-enregistrements. Ce qui contraste avec la situation d'avant-guerre, où les sur-enregistrements frauduleux étaient légion (cf. tableau 1.2).

La situation de sous-enregistrement est encore plus notable lors du recensement de 1954 [32]. En effet les résultats du premier recensement effectué par l'INSEE [44, 41, 32] ne sont compatibles avec le dénombrement de 1946 qu'en « *supposant un solde migratoire négatif de 100 000 au cours de la période intercensitaire* » [32, P. 37]. Cette observation conduira l'INSEE à mettre en place un programme de contrôle des résultats à partir de 1962, permettant d'estimer l'importance des sous-enregistrements [32]. Le système d'estimation de la population basé sur la dichotomie *population légale, population présente* à été abandonné pour ce recensement. Comme avant 1881 la définition utilisée est celle de la *population de fait* [44]. Les créateurs de la base de données *cassini.ehess.fr* ne font pas mention de ce changement de définition dans les différents documents traitant de la constitution de la base de données. C'est pourquoi on ne peut que supposer que la population présente dans la base de données est la *population de fait* (cf. A.2). Le recensement de 1954 est la première opération à avoir été déplacée pour d'autres raisons qu'une guerre. L'opération qui était prévue en 1951 à été décalée pour des raisons budgétaires [44], à partir de cette opération la périodicité quinquennale ne sera plus respectée jusqu'à la fin de la série statistique (1999). Le non respect de la périodicité quinquennale n'est pas remplacé par un nouveau rythme, la tenue des recensements est désormais variable, et la durée des périodes intercensitaires varie de sept à neuf ans.

Le recensement de 1962 (correspondant au report, pour des raisons financières de celui de 1956 [44]) est caractérisé par deux modifications majeures. La première est la mise en place d'une nouvelle définition de la population [44, 32]. Ce nouveau système inédit fait appel à des multiples catégories de population (d'où le choix du terme de *définitions multiples* pour le tableau A.1) plus étoffées que durant la période 1881–1946 [44, 32]. Une part des personnes autrefois comptabilisées dans la population *comptée à part* (notamment les militaires ou les internes) sont désormais inclus dans la *population légale* [44, 32, 45]. La mise en place d'un « [...] *enregistrement aussi précis que possible des résidences temporaires* [...] » [44, P. 5] permet d'effectuer une distinction entre les populations *sans* et *avec doubles comptes* [44, 32]. L'objectif de la population sans doubles comptes est de ne comptabiliser chaque personne qu'une seule fois, ce qui permet d'éviter les doubles comptes présents lors des précédentes opérations [44]. Cette définition de la population sera utilisée jusqu'en 1999 [44, 32, 45]. La population retenue par les créateurs de la base de donnée est la *population sans doubles comptes* [48] (cf. tableau A.1) et ce jusqu'en 1999. Le second point est la mise en place d'un programme de contrôle des données, suite aux sous-enregistrements de 1954 [32, 43]. Les conclusions de cette enquête furent sans appel, environ 800 000 personnes furent omises et 200 000 furent comptées deux fois [32]. Omissions auxquelles il faut ajouter la surestimation de la population Corse, et les militaires du contingent non comptabilisés [32]. Toutefois ce recensement « *est/ considéré, par ceux qui y avaient participé, comme étant de bonne qualité.* » [18, P. 27]. Le choix de l'INSEE à été de se compter de corrections mineures, conduisant à admettre un total fictif, sous-estimant la population totale d'environ 500 000 habitants [32]. La population totale issue de notre base de données correspond par ailleurs au total fictif (42 422 000 habitants) de l'INSEE, ce qui nous laisse supposer que les auteurs de la base de données n'ont réalisé aucun traitement complémentaire.

Contrairement aux autres opérations, le recensement de 1968 à été avancé, et non reculé. Ce dernier devait à l'origine avoir en 1970, mais ce dernier à été avancé pour étudier le rapatriement massif des pieds-noirs suite aux accords d'Évian (18 mars 1962)²³ [44].

Aucune opération de contrôle ne fut réalisée pour le recensement de 1975 [32]. Nous n'avons par conséquent aucune information sur l'importance des sous-enregistrements. Cependant DUPÂQUIER et al. [32, P. 38] estime que les sous-enregistrements ont probablement augmenté en 1975, étant donnée que : « [...] *le principe même de l'opération était alors contesté dans les milieux libertaires.* » Il a également été constaté que les omissions concernaient en premier lieu les étrangers [32]. Un groupe de travail interministériel fut mis en place pour estimer l'importance des omissions. Ce dernier estime dans son rapport de 1979 que la sous-estimation des étrangers est de l'ordre de 10 % [32].

Malgré les forts soupçons de sous-enregistrement de la population planant sur le recensement de 1975, aucune enquête post-censitaire sur la qualité des données ne fut opérée en 1982. Un contrôle de l'exhaustivité du recensement avait été souhaité par l'INSEE, celle-ci proposait de coupler les résultats du recensement au fichier des côtes mobilières. Mais ce couplage a été refusé par la CNIL [32]. C'est peut-être en réponse à la contestation du recensement de 1975 que le recensement de 1982 fut le premier à être doté d'une importante campagne de communication [44], celle-ci pris plusieurs formes, notamment l'édition d'un timbre dédié à l'opération (cf. figure 1.1).

Pour le recensement de 1990 une enquête de contrôle post-censitaire à été réalisée [18, 44]. Il s'agit de la première enquête réalisée depuis 28 ans (1962) [18]. Pour rappel l'enquête de 1962 avait quantifié 200 000 doubles-comptes et 800 000 omissions (1,7 %) [32]. Les résultats de l'enquête de 1990 indiquent

23. Le recensement de 1962 a eut lieu le 7 mars 1962 [32].



FIGURE 1.1 – Timbres à l’effigie du recensement de la population. Source : www.timbres-de-france.com.

qu’il y aurait 2 % d’omissions et 1 % de doubles comptes [18]. Comme ce qui avait été observé en 1962 les omissions et doubles comptes « *ne sont pas régulièrement [répartis] entre toutes les catégories de la population [..]* » [32, P. 38]. Les populations les plus touchés en 1990 sont les populations mobiles, principalement les jeunes adultes [18]. Le pourcentage d’omissions est comparable à celui du recensement de 1962. Mais selon COEFFIC [18] il convient de prendre plusieurs précautions. Selon lui les efforts consentis lors de l’organisation du recensement pour réduire les omissions et les doubles comptes (meilleure formation des agents recenseurs, meilleur suivi de leur travail, contrôle de l’exhaustivité de la collecte des logements) ne sont pas traduits par la comparaison directe des chiffres de 1999 avec ceux de 1962,

« *En effet, on peut considérer que, depuis 1962, les difficultés de recensement se sont aggravées en raison de l’urbanisation et de l’évolution des conditions de vie [..] Aussi, pour un travail des agents recenseurs de qualité égale, aurait-on dû avoir une augmentation sensible du taux d’erreur. Il n’en est rien, ce qui est un résultat plutôt satisfaisant.* » [18, P. 27]

Le bilan du recensement de 1990 est donc « *largement positif* » [18, résumé], et ce dernier est (tout du moins localement) « *meilleur que celui de 1982* » [18, P. 27].

Le dernier recensement de notre série statistique à été effectué en 1999. Ce dernier était initialement prévu pour 1997, mais il à été reporté, comme la majorité des recensements de la période, pour des raisons financières [44]. De ce report découle l’une des plus longues périodes inter-censitaires de la série statistique.

Le recensement après 1999

Le recensement effectué en 1999 est le dernier recensement exhaustif de la population. L’article 156 de la loi n° 2002-276 du 27 février 2002 relative à la démocratie de proximité (cf. annexe B.1, page 95), définit la nouvelle organisation du recensement. Ce dernier est désormais basé sur cinq enquêtes annuelles. Dont les premières ont été réalisées entre 2004 et 2008, leur traitement a conduit à la réalisation d’un recensement millésimé 2006²⁴ [16, 39], que nous n’avons pas retenu (pour les raisons énoncées précédemment). La nouvelle législation du recensement aura également été l’occasion d’abolir la périodicité quinquennale, qui bien que non respectée, était toujours en vigueur (jusqu’en Janvier 2004) [48, 52].

L’objectif de cette nouvelle méthodologie est de réduire fortement les coûts financiers et humains d’une telle opération. Malgré les réticences d’une partie de la communauté scientifique [5, 11] l’Insee et ses membres (ou tout du moins une partie) considèrent que la nouvelle organisation du recensement répond parfaitement aux critères de qualité définis par l’ONU²⁵ [43, 17, 16, 52]. Notamment sur les critères d’universalité et de régularité. Toutefois la question du respect de simultanéité est plus délicate. Si chaque

24. Soit l’année médiane.

25. La division statistique du département des affaires économiques et sociales de l’ONU a publié à de multiples reprises (entre autre en 1958, 1980, 1998, et 2004) sous le titre de *Principes et recommandations concernant les recensements de la population et des logements* un ensemble de recommandations visant à garantir une qualité minimale des différents recensements [25]. Ces recommandations sont organisées autour de quatre points principaux :

- Significativité : le recours à des techniques d’échantillonnage ne doit pas se faire au détriment de la qualité des données.
- Exhaustivité : tout le territoire et toutes les populations doivent être recensées.
- Simultanéité : le recensement doit être effectué dans un laps de temps réduit, et ce pour éviter les effets saisonniers.
- Périodicité : un recensement doit être effectué au moins tous les dix ans.

Date	Collecte	Dépouillement	Vérification	Publication	Sources
1856	Communes		Préfectures	SGF	[31, 44, 41, 57]
<hr/>					
	Collecte	Vérification	Dépouillement	Publication	Sources
1901	Communes	Préfectures	SGF		[31, 44, 41, 57]
<hr/>					
	Collecte	Dépouillement	Vérification	Publication	Sources
1999	INSEE				[57, 44, 32]

TABLE 1.1 – Quelques exemples d’organisation du recensement.

enquête annuelle est réalisée à la même période de l’année ces dernières restent séparées d’une à plusieurs années.

Dans le cadre de ce travail nous avons décidé de nous concentrer sur le recensement exhaustif de la population, et donc d’arrêter notre étude en 1999. Différentes raisons, aussi bien méthodologiques que pratiques, nous ont conduit à prendre cette décision. Le recensement de 1999 étant à la fois le dernier recensement exhaustif de la population et le dernier recensement du XX^e siècle, cette date nous a semblé être un point d’arrêt tout indiqué. De plus, si nous avions décidé d’introduire le dernier recensement de la population (2006), nous aurions été confronté à des problèmes, purement techniques, de jointure des données. Sachant que de nombreuses étapes ont été nécessaires pour joindre les données des autres recensements (*i.e.* provenant de la même base de données) au fond de carte, le travail nécessaire pour joindre des données en provenance d’une autre source aurait été trop contraignant relativement à son apport. Les trente-trois recensement en notre possession étant déjà une quantité de données plus que conséquente.

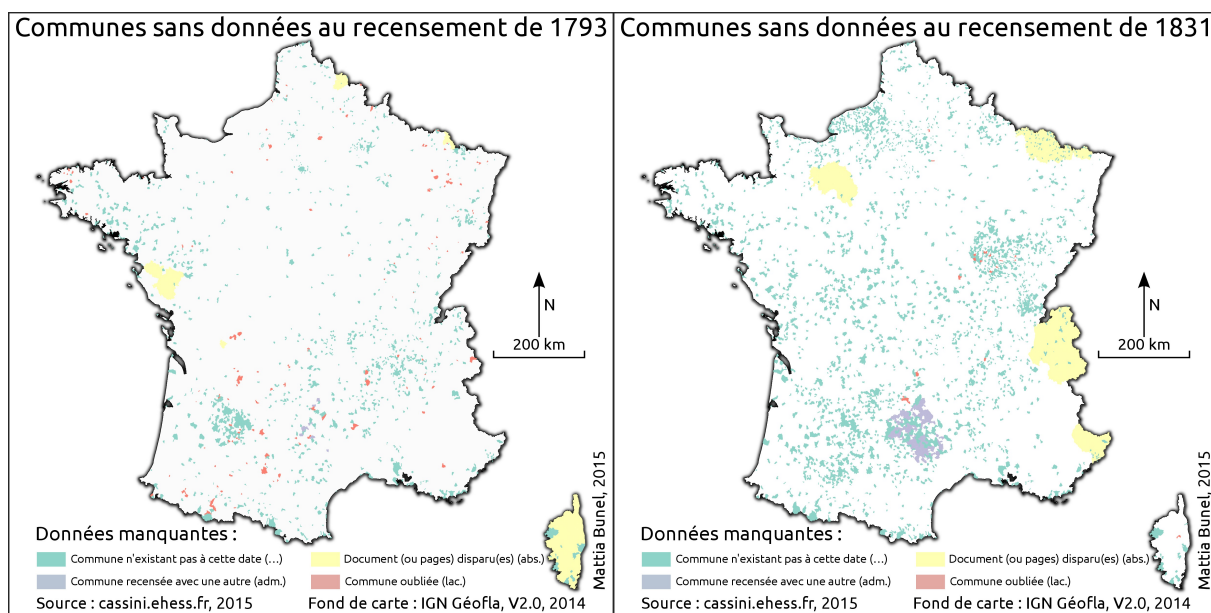
a. *I.e.* communes figurant sur un document perdu ou non publié [48].

b. *I.e.* communes oubliées lors du recensement [48].

c. *I.e.* communes dont la modalité est illisible [48].

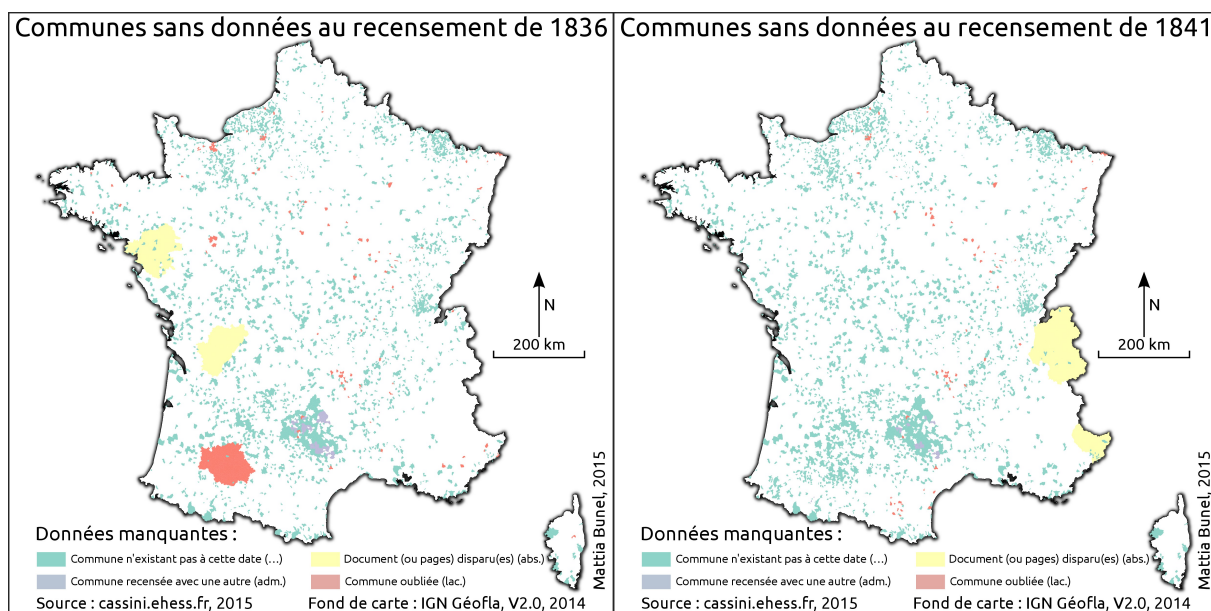
Date	Abs. ^a	Lac. ^b	Ill. ^c	Observations
1793	576	135	<i>Na.</i>	
1800	<i>Na.</i>	71	<i>Na.</i>	Fortes approximations dans l'évaluation de la population
1806	<i>Na.</i>	23	<i>Na.</i>	Certaines données auraient été manipulées
<i>1811</i>	<i>Na.</i>	<i>Na.</i>	<i>Na.</i>	<i>Le dénombrement se limite à une mise à jour des données de 1806</i>
1821	<i>Na.</i>	26	30	Certains préfets se sont contentés de réutiliser les chiffres de 1817 et 1818
<i>1826</i>	<i>Na.</i>	<i>Na.</i>	<i>Na.</i>	<i>Le dénombrement se limite à une mise à jour des données de 1821</i>
1831	1613	20	<i>Na.</i>	Premier dénombrement de qualité selon NOIN [61]
1836	604	587	<i>Na.</i>	Dénombrement de la population de droit
1841	680	73	<i>Na.</i>	Retour à la population de fait
1846	589	51	<i>Na.</i>	
1851	1759	2	<i>Na.</i>	Les enfants en bas âge ne sont pas recensés (environ 400 000)
1856	591	486	<i>Na.</i>	Omission de l'armée de Crimée (165 000 hommes)
1861	<i>Na.</i>	442	<i>Na.</i>	Omission de l'armée française en Algérie
1866	<i>Na.</i>	37	<i>Na.</i>	Omission des troupes à l'étranger (Algérie, Mexique, Rome)
1872	467	233	<i>Na.</i>	Recommandation du bulletin individuel (adoption en 1876)
1876	<i>Na.</i>	<i>Na.</i>	13	Grande diminution des omissions et des imprécisions
1881	<i>Na.</i>	<i>Na.</i>	<i>Na.</i>	Bulletins remplis par les intéressés (diminution de la qualité)
1886	<i>Na.</i>	<i>Na.</i>	<i>Na.</i>	Seule la population totale est synthétisée à l'échelle communale
1891	<i>Na.</i>	<i>Na.</i>	<i>Na.</i>	Les individus absents sont recensés dans la commune où ils se trouvent
1896	<i>Na.</i>	<i>Na.</i>	<i>Na.</i>	Mécanisation et centralisation du traitement de l'enquête professionnelle
1901	<i>Na.</i>	<i>Na.</i>	<i>Na.</i>	Mécanisation et centralisation du traitement du dénombrement
1906	<i>Na.</i>	<i>Na.</i>	<i>Na.</i>	Première mise en place d'une réelle méthodologie rigoureuse
1911	<i>Na.</i>	1	<i>Na.</i>	Exploitation des résultats réduite à cause du conflit
<i>1916</i>	<i>Na.</i>	<i>Na.</i>	<i>Na.</i>	<i>Dénombrement non réalisé (première guerre mondiale)</i>
1921	<i>Na.</i>	<i>Na.</i>	<i>Na.</i>	Possibles sur-enregistrements frauduleux
1926	<i>Na.</i>	7	<i>Na.</i>	Possibles sur-enregistrements frauduleux
1931	<i>Na.</i>	<i>Na.</i>	<i>Na.</i>	Possibles sur-enregistrements frauduleux
1936	<i>Na.</i>	<i>Na.</i>	<i>Na.</i>	Possibles sur-enregistrements frauduleux
<i>1941</i>	<i>Na.</i>	<i>Na.</i>	<i>Na.</i>	<i>Dénombrement professionnel limité à la zone libre (aucune publication)</i>
1946	<i>Na.</i>	1	<i>Na.</i>	Possible sous-enregistrement
1954	<i>Na.</i>	<i>Na.</i>	<i>Na.</i>	Déplacé pour des raisons budgétaires. Possible sous-enregistrement (notable)
1962	<i>Na.</i>	<i>Na.</i>	<i>Na.</i>	Enquête post-censitaire : sous-enregistrements
1968	<i>Na.</i>	<i>Na.</i>	<i>Na.</i>	Possible sous-enregistrement
1975	<i>Na.</i>	<i>Na.</i>	<i>Na.</i>	Possible sous-enregistrement
1982	<i>Na.</i>	<i>Na.</i>	<i>Na.</i>	Mise en place d'une grande campagne de communication
1990	<i>Na.</i>	<i>Na.</i>	<i>Na.</i>	Enquête post-censitaire : résultats comparables à ceux de 1962
1999	<i>Na.</i>	<i>Na.</i>	<i>Na.</i>	Dernier recensement exhaustif

TABLE 1.2 – Essai de synthèse de la qualité des recensements (et dénombrements) français, D'après INSEE [44], LDH [48], MEURIOT [55], MOTTE, SÉGUY et THÉRÉ [57], MOTTE et VOULOIR [58] et DUPÂQUIER et al. [30, 31, 32, 29]. *Les dates en italique correspondent à des dénombrements non réalisés ou non retenus.*



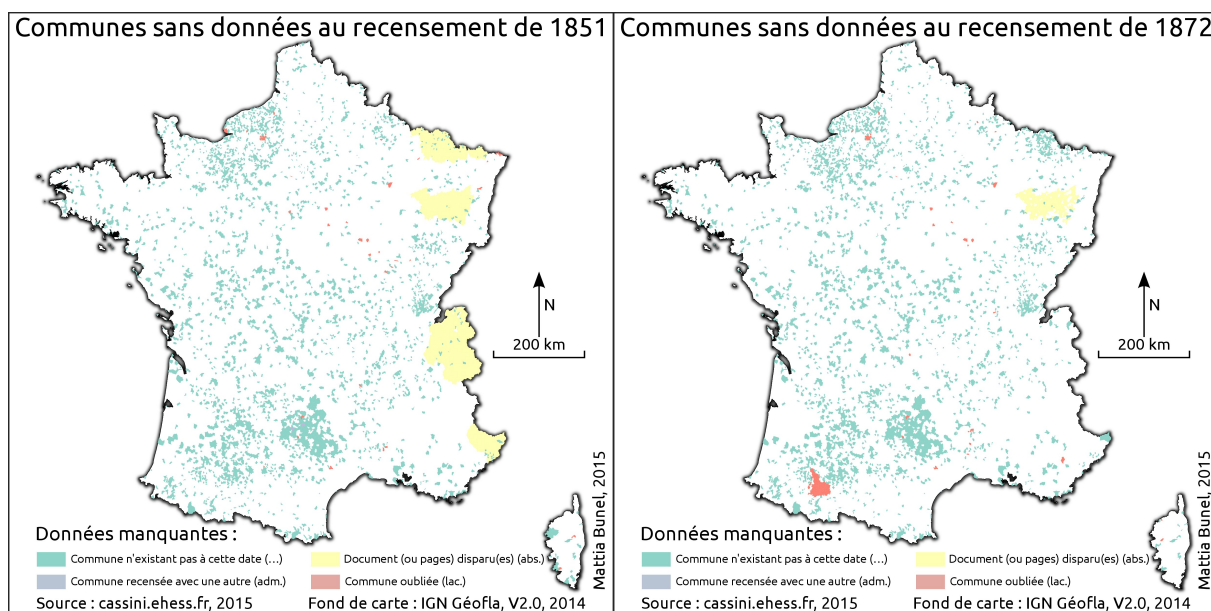
(a) Écart à la densité moyenne en 1793.

(b) Écart à la densité moyenne en 1901.



(c) Écart à la densité moyenne en 1921.

(d) Écart à la densité moyenne en 1946.



(e) Écart à la densité moyenne en 1975.

(f) Écart à la densité moyenne en 1999.

FIGURE 1.2 – Communes sans données en fonction du recensement.

Conclusion

L'objectif de cette première sous-partie était de revenir sur l'origine, l'histoire et la qualité de nos données. Face à l'amplitude temporelle de ces dernières (33 recensements / dénombremments s'étendant sur plus de deux siècles) il nous a semblé indispensable d'aborder l'histoire des données et plus généralement du recensement par le prisme de l'histoire, plus générale, de la statistique française. Au terme de cette présentation des données, nous sommes à même de juger de la qualité des différents recensements / dénombremments. Bien que cela puisse sembler évident, la qualité de ces opérations n'a (globalement) fait que croître. Mais malgré cette tendance à la hausse, la qualité des recensements fluctue fortement. De nombreux dénombremments ne sont, par exemple, que des ersatz d'un précédent dénombrement d'une qualité bien supérieure (les deux meilleurs exemples restent les dénombremments de 1851 et de 1856, et ce malgré l'oubli de l'armée de Crimée en 1856).

Mais, il convient de le rappeler, notre travail ne se base que sur une seule variable, qui est à la base de toute opération de dénombrement. Du fait de sa simplicité il est fort probable que sa qualité soit bien plus homogène que celle d'autres variables nettement plus sujettes à des changements de définition²⁶. Mais en l'absence de plus de précisions nous sommes contraints de faire l'hypothèse que la qualité de la variable étudiée (*population communale*) est fortement corrélée à la qualité du recensement dans son ensemble.

Il est également assez courant de devoir travailler sur un dénombrement non complet. C'est principalement le cas pour les huit premiers dénombremments (1793 à 1846) pour lesquels certaines communes sont manquantes. Les raisons peuvent être multiples. Le cas le plus rare est celui des communes illisibles. Seules quarante-trois communes sont concernées durant toute la période étudiée. Il s'agit du cas où les modalités sont tout simplement illisibles. Puis l'on retrouve les communes lacunaires, au nombre de 2 188. Il s'agit de communes dont les résultats n'ont pas été publiés. Cette situation ne concerne généralement que quelques communes isolées. Sauf lors des dénombremments de 1856 et de 1861 durant lesquels l'intégralité du département des Ardennes n'est pas dénombré, ou encore lors du dénombrement de 1836 où l'intégralité du département du Gers est lacunaire. Mais le cas le plus fréquent reste celui des communes dites *absentes*, au nombre de 6 879. Toutes ces omissions résultent de la disparition (ou de la non publication d'un document). Ces pertes sont l'illustration de la grande variation qualitative de la conservation des archives. La période 1826-1872 concentre 92 % de l'effectif des communes absentes. Or il s'agit de la période durant laquelle les données ont été récupérées auprès des archives départementales [48]. Il est donc fort probable que cette décentralisation du stockage des résultats soit à l'origine des multiples absences recensées. Mais, une fois encore, le peu d'informations disponibles sur la création de la base de données ne nous permet pas de confirmer cette hypothèse.

La qualité des données utilisées lors de ce travail est donc variable, mais il convient de la relativiser. Dans un premier temps il faut rappeler que le site *cassini.ehess* est le seul moyen d'obtenir des données à l'échelle communale sur une si grande durée. On peut certes trouver des séries statistiques historiques sur le site de l'Insee, mais ces dernières ne sont au mieux qu'à l'échelle départementale. Il est également indispensable de rappeler le travail effectué en amont par les constructeurs de la base de données, visant à unifier les résultats des dénombremments/recensements avec le maillage (plus ou moins) actuel. Certes cette opération n'est pas parfaite. Le principal problème étant l'existence de communes fusionnées à un temps t puis redivisées. Face à leur très grand nombre (plus de dix mille sur la période étudiée) et à l'absence d'informations facilement exploitables par un *script* informatique nous avons pris la décision de laisser la situation telle quelle, ce qui nous contraint à devoir travailler avec de nombreuses communes dénombrées avec une autre, leur population est donc comptabilisées avec celle d'une autre commune sans qu'elles ne soient fusionnées au niveau du fond de carte. Ce qui conduit à l'introduction d'un biais pouvant se traduire par des fluctuations de la densité calculée. Avec le recul désormais acquis il nous semble que ces divers biais n'influent que peu sur la qualité de notre travail. La pire des situations est sans aucun doute le dénombrement de 1851, pour lequel 1764 communes sont sans données. Rapporté à l'effectif total 95 % des communes (au minimum) ont des données exploitables pour un dénombrement donné.

1.2 Échelle nationale

Comme nous l'énoncions dans l'introduction de ce travail, notre objectif est de confronter l'hypothèse formulée par LE BRAS dans *la planète au village* en 1993 [50], précédemment formulée. Face à la complexité de cette dernière, notre méthodologie a consisté à étudier les différentes composantes indépendamment.

La présentation des résultats qui sera effectuée ici, suit parfaitement le déroulement de nos recherches.

²⁶. Ce que nous avançons n'est qu'une hypothèse. Les différentes publications traitant de la qualité du recensement (et allégrement citées dans cette sous-partie) ne traitent que de la qualité des dénombremments/recensements dans leur ensemble, et ne se concentrent pas sur une variable précise.

1.2.1 Évolution de la population et de la densité

Une fois les données précédemment décrites téléchargées une des premières opérations réalisées a été d'étudier l'évolution de la population et de la densité de population à l'échelle de la France métropolitaine. Les figures 1.3a et 1.3b ont été réalisées à partir des données issues de la base de données *cassini.ehess.fr*, couplées à des données issues d'autres sources, permettant une comparaison. La valeur de la population française représentée sur la figure 1.3a a simplement été calculée à partir de la somme de la population de toutes les communes pour chaque recensement. Étant donnée l'absence de certaines communes lors d'une partie des dénombrements du début du XIX^e siècle la valeur de la population présentée sur la figure 1.3a est indubitablement sous-estimée à certaines dates, notamment lors des dénombrements de la période 1793-1876 (se référer à la partie précédente). Toutefois rappelons que le nombre de communes sans données n'est jamais supérieur à 5 % de l'effectif total. La sous-estimation de la population est donc toute relative. La figure 1.3a nous montre par ailleurs que la valeur et l'évolution de la population française obtenue en agrégeant les données de la base *cassini.ehess.fr* correspondent aux estimations provenant d'autres sources, notamment la série la plus complète provenant de DUPÂQUIER et al. [32, 31]. Avant de décrire cette figure il est nécessaire de préciser que l'estimation de la population par NECKER en 1784 a été réalisée suivant la méthode des multiplicateurs précédemment décrite. Cette estimation sous-estime la population française de plus de deux millions d'habitants selon DUPÂQUIER et al. [30]. Mais étant donné qu'il s'agit d'une des rares sources antérieure à la série que nous avons trouvée, il nous a semblé intéressant de la confronter aux autres données, même si son analyse n'est pas pertinente d'un point de vue scientifique.

On peut repérer trois grandes phases lors de l'évolution de la population française entre 1793 et 1999. La première phase est la plus longue d'entre toutes, elle se déroule du début de la série statistique (1793) au début du XX^e siècle (dénombrement de 1911). Durant cette période la population croît de façon linéaire et augmente de 10 millions d'habitants en quatre-vingt-dix ans. Si l'évolution des données issues de notre base correspondent à celles estimées par DUPÂQUIER et al. [30, 32, 31], certains dénombrements s'éloignent de manière notable. Si l'on recoupe la figure 1.3a avec les informations sur la qualité des dénombrements précédemment présentées, on constate que les deux dénombrements s'éloignant le plus de l'estimation de DUPÂQUIER et al. [30, 32, 31] sont ceux de 1831 et 1851, soit les opérations pour lesquelles le nombre de communes sans données est maximal. On notera également que la majorité des écarts sont observés entre 1831 et 1872, période concentrant la quasi-intégralité des données manquantes. Certes la régularité des estimations réalisées par DUPÂQUIER et al. [30, 32, 31] s'estompe après 1986, mais on constate une concordance quasi-parfaite entre 1876 et 1886, périodes à partir desquelles le nombre de communes sans données est proche de zéro (seuls les dénombrements de 1876, 1911, 1926 et 1946 ont des communes sans données, le total s'élève à 22 communes manquantes). On peut donc supposer que l'écart constaté entre nos données et celles de DUPÂQUIER et al. [30, 32, 31] seraient, en partie ou totalement, dues aux communes sans données. Cette période s'inscrit durant la phase de *transition démographique* européenne s'étendant de 1740 à 1940 [79]. Ce phénomène, induit par plusieurs mutations dans les sociétés européennes a conduit, pour la majorité des pays européens à une explosion démographique. Cette transition combine deux phénomènes, à savoir une diminution de la mortalité suivie d'une diminution de la natalité. Ces deux phases étant décalées, il en résulte un important accroissement de la population. Le cas français est quant à lui particulier. Contrairement à d'autres pays européens, la baisse de la fécondité a suivi de près la baisse de la mortalité, il en résulte un accroissement de la population plus limité [79]. Une des façons de quantifier cette transformation est d'appliquer le *multiplicateur transitionnel* proposé par JEAN-CLAUDE CHESNAIS en 1986 [79], définit comme le coefficient multiplicateur de la population entre le début et la fin de la période de transition [15]. La valeur du multiplicateur transitionnel est estimée à 1,62 pour la France, alors qu'elle est de 2,26 pour l'Italie et de 3,83 pour la Suède [79]. Ainsi la France n'a pas véritablement subi une explosion démographique comme d'autres pays, dont le Royaume-Uni est un bon exemple [80]. Mais les effets de la transition démographique ne se limitent pas à une phase de croissance de la population plus ou moins soutenue. Cette dernière a également eut des conséquences structurelles, notamment des migrations [79]. Si la question des migrations internationales ne nous intéresse pas ici ce n'est pas le cas des migrations internes à la France métropolitaine. Ces dernières auraient profondément modifié le rapport villes / campagnes en France métropolitaine [79]. Toutefois les documents présentés jusqu'ici ne nous permettent pas d'observer ce phénomène, nous explorerons en détail ce point plus en aval.

La seconde phase commence à partir du dénombrement de 1911. Durant cette période (1911-1954) l'augmentation de la population est quasi-nulle, on passe d'une population de 41 426 304 à une population de 42 771 730 en quarante-trois ans. Cette faible croissance s'explique par la succession de deux phases de diminution de la population respectivement entre les 1911-1921 et entre 1946-1946. Durant ces périodes inter-censitaires la population française diminue respectivement de 2,2 millions d'habitants entre 1911 et 1921, et de 1,4 millions d'habitants entre 1936 et 1946. Cette diminution de la population est évidemment la

manifestation du grand nombre de morts induits par les deux guerres mondiales. Durant la période 1914-1919 l'excédent des décès sur les naissances a été de 2 868 000 [32]. Compte tenu des sur-enregistrements frauduleux réalisés par les mairies de Marseille et de Lyon²⁷ il faudra attendre 1950 pour que la population retrouve son niveau de 1914, la population comptabilisée entre 1921 et 1936 étant probablement supérieure à sa valeur réelle DUPÂQUIER et al. [32]. Après une nouvelle augmentation de la population entre 1921 et 1936, on assiste à une nouvelle période de diminution de population causée par la seconde guerre mondiale. La diminution de la population entre 1936 est estimée à 1,3 millions d'habitants selon les données corrigées de DUPÂQUIER et al. [31] (*i.e.* prenant en compte les manipulations des chiffres, notamment par les communes de Lyon et de Marseille), et à 1,4 millions selon les données de la base *cassini.ehess.fr*. La diminution de la population est inférieure à celle observée précédemment lors de la première guerre mondiale, ce qui est avant tout dû à une diminution du nombre de pertes de guerre, là où l'armée française a perdu 1 322 000 hommes durant la première guerre mondiale, les pertes militaires sont de 600 000 hommes lors de la seconde guerre mondiale [31]. La diminution de la population résultant de ce second conflit mondial est donc plus faible que celle résultant de la première guerre mondiale.

Après le dénombrement de 1946, on assiste à une nouvelle période de croissance de la population, qui dure jusqu'aux derniers recensements (2011). La croissance de la population durant cette période est plus importante que sur la période 1793-1911, là où il avait fallu quatre-vingt-dix ans pour que la population passe de 30 millions à 40 millions d'habitants, elle passe de 40 à 50 millions en 22 ans (1946-68), et de 50 à plus de 60 millions en 38 ans (1968-2006). Cette rapide augmentation de la population est évidemment due au phénomène démographique connu sous le nom de « *baby-boom* » dont l'année 1939 est considérée comme le point de départ [31, P. 297]. Ce phénomène est caractérisé par une importante hausse de la fécondité, après la seconde guerre mondiale, facilitée par la mise en place d'une politique familiale mise en place dès les années 30 et renforcée, notamment par une hausse considérable des prestations sociales, en 1938 [31]. Ainsi le nombre moyen d'enfants par femmes qui était de 1,82 en 1941 atteint la valeur de 3 en 1946 [31], puis la fécondité diminue à partir de 1950, à environ 2,6 enfants par femme, et augmente à nouveau de 1957 à 1963 où il atteint les 2,9 enfants par femme.

C'est cette augmentation de la fécondité qui conduit à une hausse de la natalité, et par conséquent à la reprise de la croissance de la population après la période 1911-1946.

Même si l'augmentation de la population continue, on observe une diminution de la croissance à partir du recensement de 1968 qui s'amplifie à partir de 1975. On pourrait supposer que cette diminution de la fécondité est imputable à la légalisation de l'avortement (loi VEIL du 18 janvier 1975) et la diffusion de contraceptifs modernes (pilule et stérilet) [53] (la contraception orale ayant été autorisée par la loi NEUWIRTH du 19 décembre 1967, mise en application à partir de 1972). Toutefois HENRI LERIDON réfute cette idée, selon lui la diffusion de la contraception et la légalisation de l'avortement ont conduit à la baisse du nombre d'enfants non désirés et non à la baisse du désir d'enfant [79]. Cette baisse de la fécondité, observée dans toute l'Europe depuis la fin des années soixante [79], obéit à de nouveaux facteurs sociaux, tels que le recul du mariage, la diffusion du divorce et des familles recomposées, la reconnaissance des couples homosexuels et l'augmentation de la fécondité hors mariage [79]. Ce phénomène est qualifié de seconde transition démographique [79, 81].

L'évolution de la densité (cf. figure 1.3b) suit logiquement la même évolution que la population. Comme précisé dans la légende la densité de population a été calculée pour le territoire actuel, à l'exception de l'estimation de NECKER [59], que l'on sait biaisée [30]. Par conséquent la valeur de la superficie utilisée dans le calcul de la densité est constante, l'évolution de la densité de population est donc identique à l'évolution de la population. Les trois phases précédemment décrites sont par conséquent toujours visibles, une première phase de croissance linéaire de la densité, suivie d'une période de fluctuation entre 1911 et 1946, pour finir par une période de forte croissance, dont la vitesse diminue légèrement à partir de 1975.

Cette augmentation de la densité a dû être prise en compte lors de la création de la variable à étudier (que nous nommerons également *variable de travail*). En effet, notre travail porte sur l'étude de la distribution de la population française, son évolution, et sur le rôle de l'environnement dans cette distribution. Il est donc nécessaire de concevoir une variable permettant d'étudier la distribution de la population. Une première solution aurait été de traiter directement les données téléchargées, sans aucune transformation. Toutefois on préfère généralement utiliser la densité de population [6, 24, 68, 69], l'utilisation de cette variable permet de s'affranchir de l'effet de masse [14], inhérent aux variables issues d'un comptage, en rapportant la population à la superficie qu'elle occupe, on s'affranchit du biais causé par la variabilité des superficies communales. Cette solution aurait été parfaitement adaptée dans le cas où nous aurions travaillé sur une seule date. Or comme le rappellent notamment PINCHEMEL et PINCHEMEL [65] et PUMAIN [68] l'interprétation d'une même valeur de densité et fortement dépendante du contexte spatial [27, 68], deux sociétés « comparables » peuvent avoir des densités très différentes (c'est notamment

27. Après correction du recensement DUPÂQUIER et al. [31] estime que la population en 1936 était de 41 194 000 habitants, là où les résultats de l'agrégation de la base de données donnent un total de 41 900 407 habitants [31].

le cas du Royaume-Uni et de la Nouvelle-Zélande) [6, 10], et inversement deux sociétés différentes peuvent avoir des densités proches [24]. C'est pourquoi la densité ne peut être considérée comme « *une mesure de l'efficacité géographique d'une société* » [27], cette dernière dépend avant tout du contexte spatial. Or il existe un second contexte, tout aussi important, qui est toujours éclipsé par le contexte spatial. Il s'agit du contexte historique. En effet notre travail porte sur une région somme toute assez réduite à l'échelle du globe, et présentant une unité rendant les comparaisons de densité au sein du territoire beaucoup moins aléatoires qu'une comparaison entre la densité de la Lorraine et du pays Dogon. Mais notre travail porte sur une période de deux cent-six ans, et comme le montre la figure 1.3b, la densité n'a cessé de croître, par conséquent la signification d'une même valeur de densité varie aux court du temps, là où une commune de 70 hab./km² était plus dense que la moyenne en 1850, cette même densité correspond à des communes moins peuplées que la moyenne en 1975. Ainsi si il demeure possible de comparer les densités des communes à une date t il n'est pas possible de comparer des densités de différentes communes à différentes dates, il est par conséquent impossible de travailler à partir de la densité. Nous avons donc cherché une solution permettant de coupler les avantages de la densité (limitation de *l'effet de masse* [14], permettant une comparaison entre les différentes communes) et nos contraintes. Celles-ci provenaient exclusivement de l'inflation de la densité, il était donc pertinent de travailler à partir d'un point de référence constant, la densité moyenne ne pouvant jouer ce rôle. Il a donc été décidé de travailler à partir des écarts à la densité moyenne. Ainsi, on ne se réfère plus directement à la densité d'une commune, mais à l'écart de cette commune par rapport à la densité moyenne, ce qui correspond à centrer la variable [28]. L'inflation de la densité est donc prise en compte lors du calcul. Il est désormais possible de suivre la trajectoire des communes par rapport à la moyenne et ainsi savoir si ces dernières gagnent ou perdent en densité, sans le biais de l'inflation. Ainsi nous ne travaillons pas directement sur la densité, mais sur l'écart à la densité moyenne. Par construction la moyenne de la série est nulle. Il faut également préciser qu'avant de calculer les écarts à la densité moyenne nous avons calculé le logarithme de la densité, la variable de travail est donc calculée de la manière suivante :

$$Ed_{it} = D_{it} - \bar{D}_t \quad (1.1)$$

Avec Ed_{it} , l'écart à la densité moyenne pour l'individu i à la date t , D_{it} le logarithme en base dix de la densité pour l'individu i à la date t , $D = \log(\text{densité})$ et \bar{D}_t le logarithme de la densité moyenne à la date t . La création de cette variable nous permet donc de comparer la densité relative de différentes communes à différentes dates. Après avoir calculé cette nouvelle variable pour chaque commune et chaque date nous nous sommes penchés sur l'étude de la distribution statistique et l'évolution de cette variable, comme on va le voir dans la partie suivante cette simple étude statistique va nous permettre de valider certaines hypothèses, et d'en formulé d'autres.

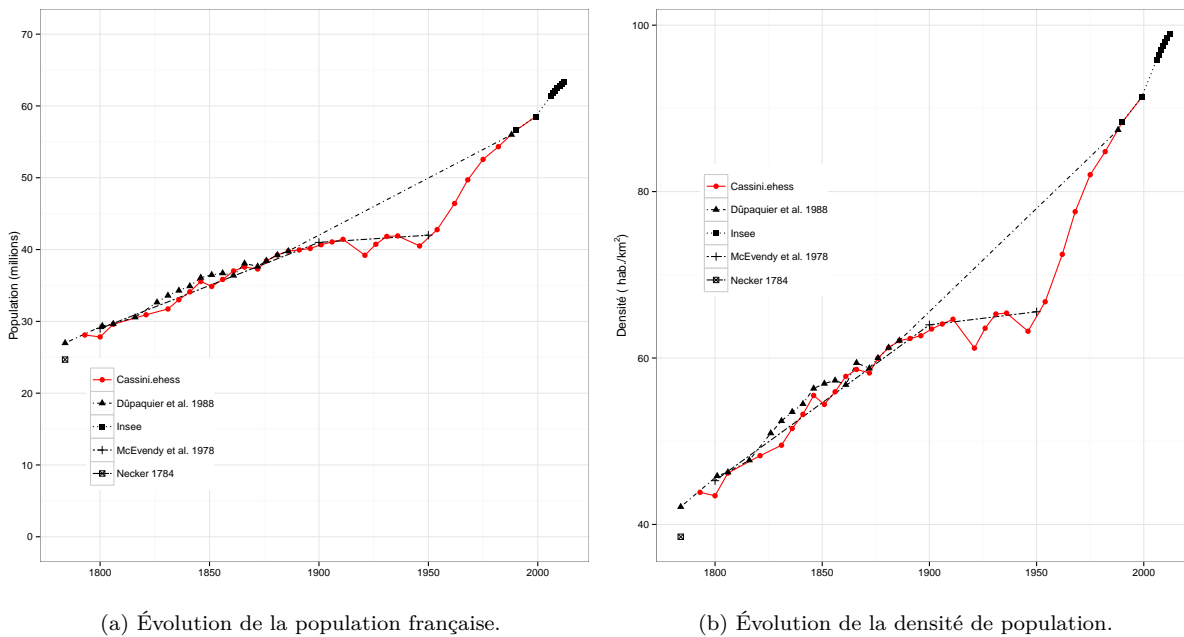


FIGURE 1.3 – Évolution de la population et de la densité de population en France entre 1793 et 2011. Sources : *Cassini.ehess.fr*, INSEE, ainsi que les ouvrages suivants : DUPÂQUIER et al. [30, 32, 31], MCEVEDY et JONES [54] et NECKER [59] pour la figure 1.3b. Les valeurs de la densité et de la population sont estimés pour le territoire actuel (sauf pour NECKER [59]).

1.2.2 Dispersion statistique...

Nous cherchons donc à savoir si il existe une tendance à la concentration de la population au cours de la période étudiée. Pour ce faire nous avons décidé de traiter le problème par une approche globale et spatiale, nous introduirons la dimension spatiale lors de la prochaine partie. Lors de l'obtention de notre base de données, nous avons commencé par adopter une approche exploratoire, en analysant la distribution, et les indicateurs de centralité et de dispersion des variables traitées. Cette étape n'a pas été réalisée dans le but de valider une hypothèse précédemment formulée. Toutefois nos observations ont constitué un socle pour la suite de notre travail. Nous avons traité notre corpus en considérant l'écart à la densité moyenne à la date t comme une variable, ce qui porte leur nombre à trente-trois. Il en résulte le graphique 1.4a, lequel met en évidence la distribution de chacune de trente-trois séries statistiques. On peut déjà faire plusieurs observations. Dans un premier temps on notera certaines caractéristiques liées à la construction de la variable. Par construction chaque série est centrée sur zéro (ce qui est dû au fait que l'on retient l'écart à la moyenne, dont la moyenne est par définition nulle) et toutes ont une forme que l'on peut apparenter à une distribution de LAPLACE-GAUSS. Il convient de rappeler que nous utilisé le logarithme de la densité de population pour calculer l'écart à la densité moyenne, l'objectif de cette transformation était de « redresser » une variable que l'on sait d'expérience dissymétrique. Toutefois la réalisation d'un ensemble de test de KOLMOGROV-SMIRNOV ainsi que JACQUES BERA²⁸ ne nous a pas permis de valider l'hypothèse d'une distribution normale, quelle que soit la date (cf. tableau 1.3). Bien que ce fait soit gênant d'un point de vue statistique, de nombreux modèles étant basés sur l'hypothèse d'une distribution normale, il est parfaitement logique dans le cadre d'une problématique géographique. Dans *statistique et territoire* [14, P. 46] JOËL CHARRE se questionne sur la signification géographique d'une distribution normale et en vient à la conclusion que cette distribution caractérise généralement une distribution spatiale aléatoire [14], la concentration spatiale correspondant à une distribution dissymétrique. Le rejet des hypothèses de normalité nous permet donc de supposer que la variable étudié est distribuée suivant une logique spatiale, laquelle sera longuement étudiée une fois cette hypothèse validée.

Malgré le fait que l'étude de l'écart-type prenne tout son sens lors de l'analyse d'une distribution normale [14], dans notre cas son étude n'est pas dénuée d'intérêt. On observe sur le graphique 1.4a, une tendance à « l'affaissement » de la série statistique. Plus l'année du recensement augmente, plus la base des courbes s'élargit et leur sommet s'affaisse. En clair, ce que l'on observe est une augmentation de la dispersion de la distribution [14]. Cette dernière peut-être quantifiée à l'aide des habituels indicateurs de

28. Ce dernier étant plus performant que le test de KOLMOGROV-SMIRNOV qui « n'utilis[e] pas les propriétés spécifiques de la loi de GAUSS [...] » [74, P. 369]. Le test de BERA fait appel aux coefficients d'assymétrie et d'aplanissement (respectivement *skewness* et *kurtosis*) dont il « évalue les écarts simultanés [...] avec les valeurs de référence de la loi normale » [71, P. 24].

dispersion. Nous utiliserons principalement l'écart-type. Lorsque la distribution étudiée suit une loi de LAPLACE-GAUSS, l'écart-type correspond à la distance séparant les deux points d'inflexion de la fonction de masse [74]. Cette propriété n'est pas valable dans notre cas, mais elle peut nous servir à estimer visuellement l'évolution de l'écart-type. En effet, bien que la distribution étudiée ne soit pas normale son *affaissement* et son *élargissement* au cours du temps ne peut que correspondre à une augmentation de l'écart-type [74]. Mais la superposition de ces multiples courbes ne permet au lecteur qu'une vision globale de l'évolution de la distribution et le détail des variations intercensitaires ne peut que lui échapper. C'est pourquoi nous avons construit les graphiques 1.4b et 1.5b permettant une meilleure compréhension du phénomène étudié. Ces derniers valident la précédente supposition – à savoir l'augmentation de l'écart-type au cours du temps – mais nous apportent également de nouveaux éléments d'analyse. Le principal apport des ces deux figures est de nous renseigner sur la variation de l'évolution de l'écart-type. Les courbes de distribution (cf. figure 1.4a) ne permettaient pas d'observer ce phénomène. Il est extrêmement intéressant pour notre problématique de constater que les ruptures et les points d'inflexion observables sur la figure 1.4b correspondent à des ruptures historiques déjà identifiées dans de nombreux travaux de recherche Compléter. Toutefois il est préférable d'éviter de *plaquer* un commentaire historique banal à un phénomène qui nous est encore, à ce stade, presque inconnu. On risquera de modifier inconsciemment les faits pour qu'ils correspondent au mieux à notre volonté. Plusieurs ruptures sont visibles sur la figure 1.4b, la première est sans doute la moins visible. Le terme de rupture est sans aucun doute inapproprié pour cet événement qu'il est préférable de qualifier de *transition*. Cette dernière s'opère entre le dénombrement de 1831 et celui de 1856. On passe d'une phase où la variation de l'écart-type est quasi-nulle (entre 1793 et 1831) à une phase où l'écart-type augmente de façon non linéaire. La figure 1.4c a été réalisée pour mettre en exergue les ruptures observables sur les figures 1.5b et surtout 1.4b. Étant donné que sa lecture peut être délicate il est nécessaire de revenir sur sa construction. Cette figure représente la pente locale pour chaque recensement / dénombrement. Cette valeur est obtenue à l'aide de la formule canonique :

$$m = \frac{\Delta y}{\Delta x} \quad (1.2)$$

Où y et x sont respectivement les ordonnées et les abscisses. Dans notre cas on utilisera la formulation suivante :

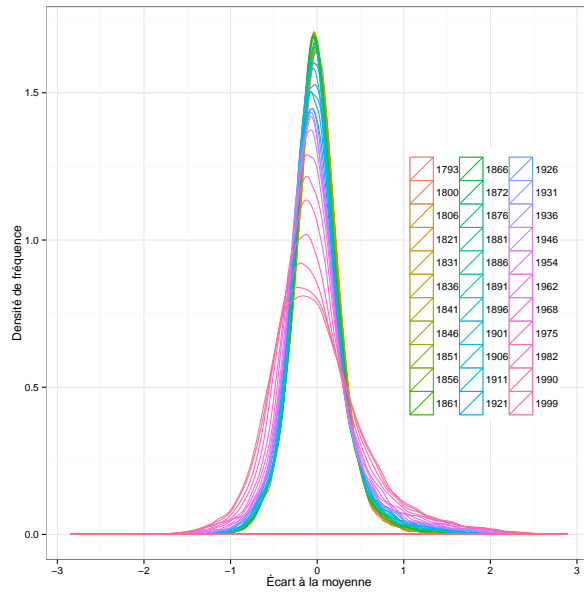
$$m = \frac{\Delta y}{\Delta t} = \frac{y_{t+1} - y_t}{t_n - t_{n+1}} \quad (1.3)$$

Où y est la valeur de l'écart-type, t la date du recensement. Par conséquent Δy est la différence des écarts-types et Δt à l'écart inter-censitaire. Par conséquent la valeur observable à la date t correspond à la pente entre la date t et $t + 1$, par conséquent la pente entre 1990 et 1999 est rapportée au point d'abscisse 1990, et le graphique est constitué de 32 points pour 33 recensements. Cette figure confirme les précédentes observations, l'augmentation progressive de l'écart-type visible à partir du milieu du XIX^e siècle se manifeste sur ce second graphique par une valeur positive du coefficient directeur de la tangente, et l'augmentation de la valeur de ce même coefficient et une manifestation d'une augmentation de plus en plus rapide de l'écart-type. Cette tendance à la croissance de la pente locale est stoppée entre 1936 et 1946. Évidemment l'écart-type continue à croître, comme l'indique la valeur *positive* de la pente locale, mais la vitesse de cette croissance est fortement diminuée. Ces observations sont parfaitement visibles sur la figure 1.4b, où l'on observe une diminution de la croissance entre 1931 et 1936, mais surtout entre 1936 et 1946. C'est cette période de transition (1936-1954) qui donne tout son intérêt à la figure 1.4c. On remarque qu'à partir du recensement de 1946 l'écart-type continue à augmenter, mais à une vitesse qui n'a jamais été atteinte jusqu'ici. Et la figure 1.4c illustre parfaitement cette croissance, dont l'apogée est l'évolution de l'écart-type entre les recensements de 1962 et de 1968. À partir de 1975 on observe une diminution de la vitesse de croissance de l'écart-type, qui atteint son plus haut niveau de la série en 1999. Au vu de la transition qui s'opère entre 1968 et 1975 il apparaît clairement que le recensement de 1975 marque une transition, au même titre que celui de 1946. Deux autres ruptures, minimes sont identifiables. Mais contrairement aux deux précédentes ruptures elles le sont uniquement sur la figure 1.4c, ce qui nous laisse supposer que leur importance est bien moindre. La première est située entre le dénombrement de 1836 et celui de 1841. C'est en effet à partir du dénombrement de 1841 que la croissance du coefficient directeur de la tangente semble suivre un modèle linéaire, c'est donc cette probablement cette date qui marque la distinction entre la phase de relative stabilité l'écart-type décrite précédemment et la phase de croissance polynomiale de l'écart-type visible jusqu'en 1946. La seconde rupture est observée entre le dénombrement de 1911 et celui de 1921. Bien qu'invisible sur la figure 1.4b il s'agit de la plus importante augmentation de l'écart-type jusqu'en 1946.

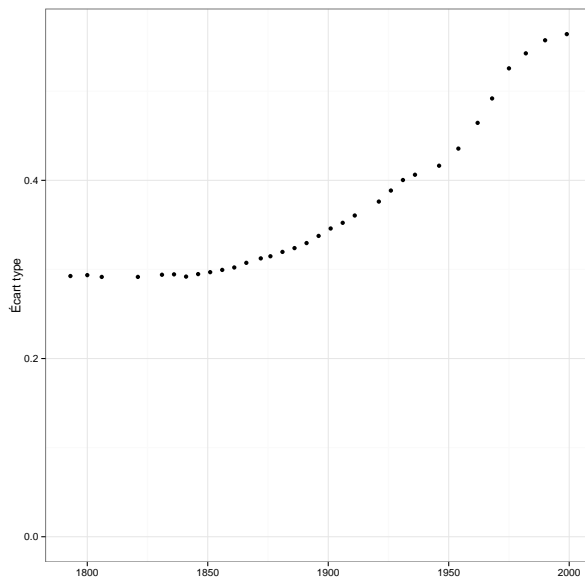
On peut coupler cette observation à celle de la figure 1.5a, qui, contrairement à la figure 1.5b, est une « *boite de Tuckey* » [51, 78] (on parlera désormais de *boxplot*) des plus classiques. L'immobilité de la barre centrale indique que la valeur médiane est constante (toujours égale à zéro) quelle que soit l'année du recensement. Mais l'allongement des « *bords* » et des « *moustaches* », indique respectivement une

augmentation des intervalles inter-quartiles et inter-déciles. La simple comparaison visuelle des figures 1.5a et 1.5b permet de remarquer la proximité entre ces deux productions graphiques. Ainsi quel que soit l'indicateur de dispersion utilisé (écart-type ou écart inter-quartile) ces derniers augmentent au cours du temps. Il est, bien entendu, parfaitement logique que les deux indicateurs de dispersion utilisés évoluent de manière similaire. Mais il semblait indispensable de présenter conjointement les deux *boxplot* réalisées afin que le lecteur puisse les comparer, le recours d'une *boxplot* semblable à la figure 1.5b n'étant pas commun dans notre discipline. D'un point de vue statistique l'augmentation des indicateurs de dispersion est l'illustration d'une plus grande dispersion des individus. Ainsi l'écart-interquartile n'est que l'amplitude entre le deuxième quartile et le troisième quartile, *i.e.* la différence entre la valeur minimum du second quartile et la valeur maximum du troisième quartile. De la même manière l'écart-type mesure l'écart moyen à la moyenne (ici zéro par construction), plus ce dernier est important, plus les valeurs sont dispersées autour de la moyenne [74], et plus la distribution de la variable est hétérogène [14]. Il convient de rappeler que le terme d'hétérogénéité n'a pas la même signification en statistiques et en géographie [14]. Une distribution hétérogène renvoie à une dispersion statistique maximale (et donc un des valeurs élevées des indicateurs de dispersion). Alors que ce même terme renvoie à une dispersion spatiale minimale en géographie, *i.e.* que « *quelques unités spatiales tendent à concentrer l'essentiel du total des valeurs de la variable* » [14, P.43]. Une distribution statistique fortement dispersée correspond par conséquent à une faible dispersion géographique et inversement. Ainsi, l'augmentation de la dispersion statistique illustre un phénomène de concentration géographique [14]. Étant donné qu'une forte hétérogénéité statistique implique une grande différence entre les individus, ce qui appliqué à notre problématique revient à dire que l'écart à la densité moyenne augmente, et que par conséquent l'importance des communes fortement (et faiblement) denses augmente, ce qui implique une concentration de la population dans certaines communes, et ce au prix d'une diminution de l'importance relative de certaines communes. Les travaux de JOËL CHARRE nous permettent donc de valider, élégamment, la première hypothèse que nous avons formulée et de laquelle dépend la validation des hypothèses suivantes. Nous sommes donc en mesure d'affirmer que l'augmentation de l'écart-type, et plus généralement des indicateurs de dispersion, n'est que la manifestation statistique d'une concentration géographique de la population. De même l'évolution de l'écart-type est l'illustration que cette tendance à la concentration n'est pas linéaire et son évolution est probablement marquée par des ruptures historiques.

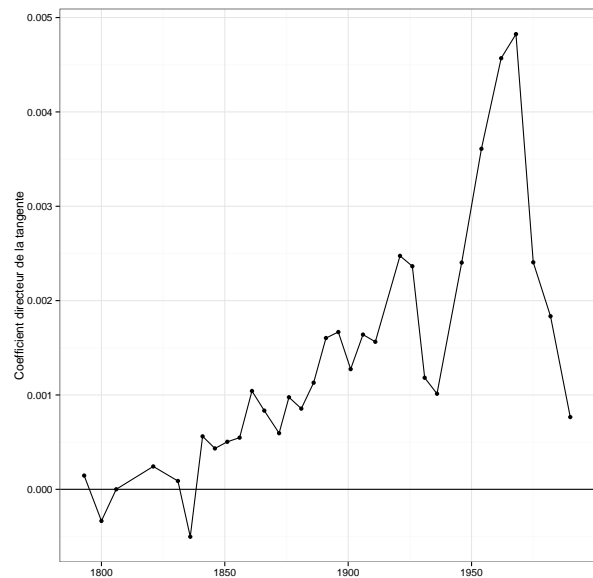
Mais face à l'importance que revêt cette hypothèse dans le cadre de notre travail (elle peut invalider les hypothèses suivantes) nous avons choisi de confronter les résultats obtenus à ceux d'une seconde méthode que nous avons adapté à notre problématique : les courbes de LORENZ et le coefficient de GINI. Cette seconde méthode sera également utile pour confirmer les variations temporelles de l'évolution de la concentration de la population.



(a) Distribution de la variable *écart à la densité moyenne* pour chaque recensement.



(b) Évolution de l'*écart-type*.



(c) Pente locale pour la figure 1.4b.

FIGURE 1.4 – Évolution de l'*écart-type* et pente locale associée pour la variable *écart à la densité moyenne* en fonction de l'année du recensement.

Date	KOLMOGROV-SMIRNOV		JACQUES BERA		Hypothèses retenues
	Valeur	p-value	Valeur	p-value	
1793	0.2750723	< 2.2e-16	14260.245	< 2.2e-16	H1 / H1
1800	0.2738502	< 2.2e-16	10904.659	< 2.2e-16	H1 / H1
1806	0.2749775	< 2.2e-16	11036.787	< 2.2e-16	H1 / H1
1821	0.2743499	< 2.2e-16	10603.224	< 2.2e-16	H1 / H1
1831	0.2742663	< 2.2e-16	11082.378	< 2.2e-16	H1 / H1
1836	0.2743860	< 2.2e-16	12420.983	< 2.2e-16	H1 / H1
1841	0.2775808	< 2.2e-16	15439.409	< 2.2e-16	H1 / H1
1846	0.2756981	< 2.2e-16	16089.146	< 2.2e-16	H1 / H1
1851	0.2770837	< 2.2e-16	17610.906	< 2.2e-16	H1 / H1
1856	0.2762356	< 2.2e-16	21319.517	< 2.2e-16	H1 / H1
1861	0.2752498	< 2.2e-16	22140.990	< 2.2e-16	H1 / H1
1866	0.2736429	< 2.2e-16	23438.871	< 2.2e-16	H1 / H1
1872	0.2724477	< 2.2e-16	25992.817	< 2.2e-16	H1 / H1
1876	0.2730604	< 2.2e-16	28179.802	< 2.2e-16	H1 / H1
1881	0.2728010	< 2.2e-16	32551.141	< 2.2e-16	H1 / H1
1886	0.2726004	< 2.2e-16	35421.252	< 2.2e-16	H1 / H1
1891	0.2709416	< 2.2e-16	36911.866	< 2.2e-16	H1 / H1
1896	0.2689949	< 2.2e-16	38747.982	< 2.2e-16	H1 / H1
1901	0.2663988	< 2.2e-16	40686.587	< 2.2e-16	H1 / H1
1906	0.2646195	< 2.2e-16	42286.616	< 2.2e-16	H1 / H1
1911	0.2629997	< 2.2e-16	45042.405	< 2.2e-16	H1 / H1
1921	0.2555919	< 2.2e-16	45460.680	< 2.2e-16	H1 / H1
1926	0.2525993	< 2.2e-16	45787.090	< 2.2e-16	H1 / H1
1931	0.2483898	< 2.2e-16	45782.173	< 2.2e-16	H1 / H1
1936	0.2463170	< 2.2e-16	44552.373	< 2.2e-16	H1 / H1
1946	0.2411028	< 2.2e-16	38985.261	< 2.2e-16	H1 / H1
1954	0.2335250	< 2.2e-16	34315.609	< 2.2e-16	H1 / H1
1962	0.2235705	< 2.2e-16	30317.394	< 2.2e-16	H1 / H1
1968	0.2133040	< 2.2e-16	25359.251	< 2.2e-16	H1 / H1
1975	0.1988429	< 2.2e-16	17389.730	< 2.2e-16	H1 / H1
1985	0.1873496	< 2.2e-16	11651.502	< 2.2e-16	H1 / H1
1990	0.1756197	< 2.2e-16	8406.870	< 2.2e-16	H1 / H1
1999	0.1712079	< 2.2e-16	7100.646	< 2.2e-16	H1 / H1

TABLE 1.3 – Valeurs et p -values des tests de KOLMOGROV-SMIRNOV et de JACQUES BERA pour la variable *écart à la densité moyenne à la date t* . Quelle que soit la date et le test, l’hypothèse alternative (H1) est systématiquement retenue.

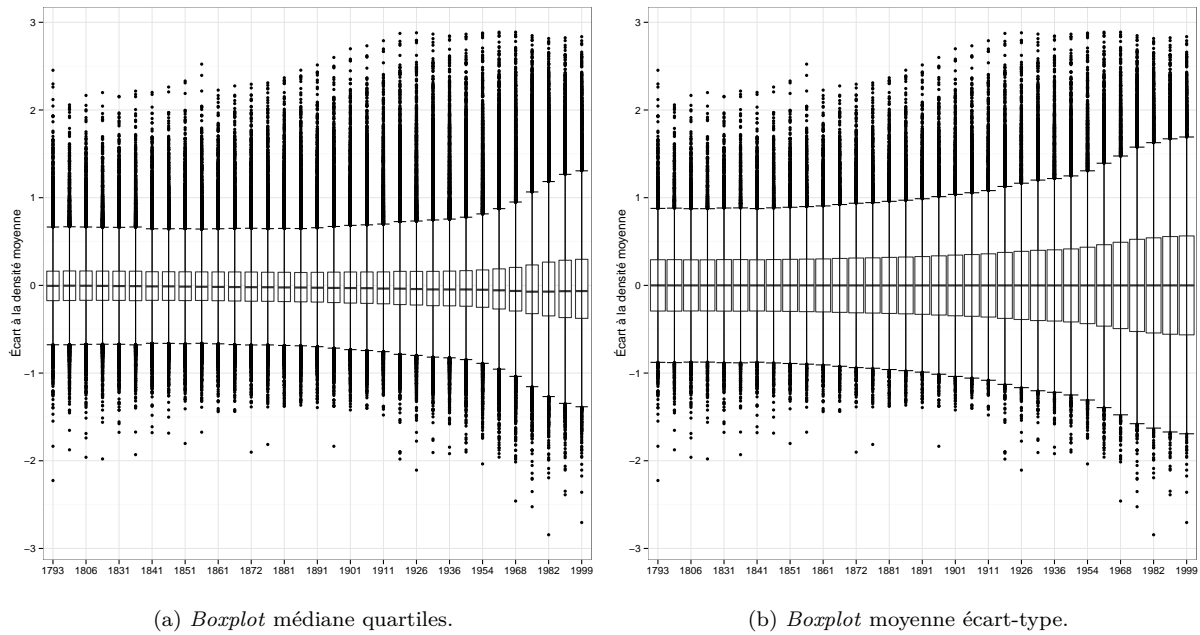


FIGURE 1.5 – *Boxplot* de la variable *écart à la densité moyenne* en fonction de l'année du recensement. La figure 1.5a correspond à une *boxplot* classique telle que proposée par TUCKEY [78], la figure 1.5b correspond à une *boxplot* où la valeur centrale est la moyenne, le sommet de boîte est situé à un écart-type de la moyenne, les outliers sont au-delà de trois écarts-types.

1.2.3 ... Concentration géographique

Pour valider l'hypothèse, précédemment formulée, d'une tendance à la concentration de la population durant la période étudiée nous proposons de recourir à une seconde méthode, habituellement utilisée pour mesurer (et comparer) les inégalités économiques. Il est tout à fait possible d'utiliser les courbes de LORENZ et le coefficient de GINI pour mesurer la concentration de la population, PUMAIN et SAINT-JULIEN [70] proposent d'ailleurs d'appliquer cette méthode à des problématiques spatiales. L'idée est ici de représenter sur un graphique l'effectif cumulé des communes en fonction de leur population cumulée. Le graphique permet donc de représenter la part de la population représentée par une proportion donnée de communes, et inversement. Dans le cas d'une distribution parfaitement homogène (statistiquement), chaque commune possède la même part de la population totale. La courbe de LORENZ devient une droite, traditionnellement représentée sur les graphiques (comme c'est le cas ici). Plus la courbe obtenue est « creuse », plus le phénomène étudié est concentré, ce qui dans notre cas implique qu'une part réduite des communes concentre une part importante de la population totale.

Le but du coefficient de GINI (G) est de quantifier cette concentration. Il est défini comme étant le double de l'aire comprise entre la courbe de LORENZ et la droite d'« équipartition » [P.73 70, 74], définie telle que $f(x) = y$. Sa valeur théorique est donc égale aux double de la différence de l'intégrale de la fonction d'équipartition ($\int_0^1 f(x) dx$) et celle de la courbe de LORENZ ($\int_0^1 L(x) dx$). L'intégrale de la fonction $f(x) = y$ sur l'intervalle $[0; 1]$ étant égale à un, l'indice de GINI est donc égal à :

$$G = 2 \cdot (1 - \int_0^1 L(x) dx) \quad (1.4)$$

Dans sa formulation théorique. Mais en pratique on ne dispose pas nécessairement de la fonction $L(x)$, il est donc nécessaire de calculer le coefficient de GINI en estimant l'intégrale de la fonction $L(x)$ à l'aide de la formule suivante :

$$B = \sum_{k=0}^{n-1} (X_{k+1} - X_k)(Y_{k+1} + Y_k) \quad (1.5)$$

Où B est l'intégrale de la fonction $L(x)$, X est l'effectif cumulé, et Y la population cumulée (dans notre cas). Cette méthode revient donc à faire la somme de l'aire des rectangles de coordonnées $(X_k, X_{k+1}, Y_k, Y_{k+1})$, dont la longueur L est égale à $(Y_{k+1} + Y_k)$ la largeur à $(X_{k+1} - X_k)$ et l'aire à $L \cdot l$. C'est cette seconde approche qui est utilisée par le package *ineq*, ce dernier calcul le coefficient de GINI à l'aide

d'une des fonctions suivantes (la première fonction étant celle utilisée dans notre cas) :

$$G = \frac{2 \sum_{i=1}^n i y_i}{n \sum_{i=1}^n y_i} - \frac{n+1}{n} \quad \text{ou} \quad G = \frac{2 \sum_{i=1}^n i y_i}{(n-1) \sum_{i=1}^n y_i} - \frac{n+1}{n-1} \quad (1.6)$$

Avec (dans notre cas) n l'effectif, y_i la population communale préalablement tirée ($\forall i \quad y_i \leq y_{i+1}$). G est borné entre 0 et 1, valeurs qui ne sont que théoriques puisqu'elles supposent respectivement une distribution parfaitement égalitaire, ou la concentration en un seul individu [47, P.13], ce qui est peu probable, voir impossible dans notre cas²⁹

L'application de cette méthodologie à notre sujet nous a conduit à réaliser les graphiques 1.6a et 1.6b. Le graphique représente 1.6a la courbe de concentration (courbe de LORENZ) de la population à l'échelle communale pour chaque recensement, et la figure 1.6b l'évolution du coefficient de GINI associé à ces mêmes courbes. La multiplication des courbes sur la figure 1.6a rend sa lecture détaillée délicate. Dans l'optique d'une meilleure lisibilité nous avons rendu les courbes bornant la série statistique (1793 et 1999), plus opaques que les courbes intermédiaires. Notez que nous avons suivi une démarche similaire à celle de l'étude de l'écart-type. À une première figure introductive, s'ajoute une seconde production graphique, permettant une meilleure compréhension et une quantification du phénomène.

La confrontation des trente-trois courbes de LORENZ (figure 1.6a) nous montre une tendance à l'éloignement successif de la droite d'équipartition. Les différentes courbes ne se chevauchent que rarement, à l'exception notable des premiers et des derniers recensements. Ce phénomène est plus visible pour la fin de la série. Ainsi on observe que le dénombrement de 1793 ne correspond pas à la situation où la concentration est la plus faible, et 1999 n'est pas la date où la concentration est maximale. Si l'on se réfère à la figure 1.6b on constate que la concentration minimum a été atteinte en 1812 et que la concentration maximale correspond au recensement de 1975 (respectivement les valeurs minimales et maximales du coefficient de GINI). Ces nouvelles informations n'étaient pas identifiables sur la figure 1.4b, on observait alors qu'une phase de relative stabilité (début de la série statistique) et une phase de ralentissement de la croissance de l'écart-type (fin de la série). Plus le temps passe, plus l'on s'éloigne du modèle d'équipartition, et donc d'une répartition uniforme de la population [70]. La figure 1.6b, représente la variation du coefficient de GINI au cours du temps. La ressemblance entre l'évolution du coefficient de Gini et celle de l'écart-type est remarquable. Les deux courbes sont loin d'être identiques, mais elles croissent de façon similaire. Les phases précédemment observées sur la figure 1.4b sont toujours présentes. Une phase de « stabilité », ou tout du moins de faible fluctuation, s'étendant de 1793 à 1831 (la rupture est plus marquée que sur la figure 1.4b), suivie d'une phase de croissance linéaire, contrairement à la croissance de l'écart-type, suivie d'une nouvelle période de stabilité, voir de décroissance, à partir de 1975. On notera toutefois que le seuil de 1945 est toujours présent, mais ce dernier est nettement moins visible que sur la figure 1.4b.

L'augmentation du coefficient de GINI et l'éloignement de courbes de LORENZ de la courbe d'équipartition indique une tendance à la concentration de la population. L'évolution du coefficient de GINI est similaire à celle de l'écart-type, ce qui confirme l'hypothèse précédemment formulée d'une tendance à la concentration de la population. Certes il existe certaines différences entre ces deux courbes, notamment entre 1831 et 1936, sur la figure 1.4b l'augmentation n'est pas linéaire (ce qui est confirmé par la figure 1.4c), et l'on peut faire la même observation pour la période 1946–1975, sur la figure 1.6b la croissance du coefficient de GINI est identique à celle de la période 1831–1836, là où la croissance de l'écart-type est nettement plus importante que celle de la période précédente. De la même manière le recensement de 1975 constitue l'apogée du coefficient de GINI, alors que cette date n'était que l'apogée de la vitesse de croissance de l'écart-type. Malgré ces différences, les deux courbes possèdent une structure notable. Notamment au niveau des ruptures que nous avons précédemment identifiées. On note, une fois encore que la croissance du coefficient de GINI commence entre 1831 et 1856, et que cette dernière continue jusqu'en 1946, là où commence une nouvelle phase de croissance, qui continue jusqu'en 1975, et qui est suivie d'une phase de stabilisation. On remarquera aisément que ce commentaire est quasiment identique à celui que nous avons précédemment réalisé. Une fois encore, il semble que les ruptures visibles sur cette seconde courbe ne soient que la manifestation d'importants événements historiques ayant conduit à une modification de l'inscription spatiale de la société française. Bien qu'il ne s'agisse que d'une supposition, certains résultats du projet *Archæomedes* confortent cette observation.

La figure 1.6c est inspirée d'une réalisation de ce projet de recherche [3, P. 149]. Nous avons choisi de reprendre cette figure qui complète efficacement la figure 1.6a en facilitant la lecture. L'ordonnée figure, comme pour la figure 1.6a, l'effectif cumulé, mais les abscisses représentent cette fois l'année du recensement. Le graphique permet donc de savoir l'évolution de l'effectif cumulé correspondant à une part fixe de la population cumulée. Ce graphique présente de manière plus lisible les résultats que l'on obtiendrait si on lisait sur la figure 1.6a, l'effectif cumulé pour chaque recensement, correspondant à

29. Dans notre cas $G = 0$ implique que la population de toutes les communes est égale, et $G = 1$ implique que toute la population de la France métropolitaine est concentrée dans une seule commune.

une population cumulée fixe (ici 25, 50, 65, 75, et 80 %), *i.e.* les multiples ordonnées correspondant à une même abscisse. Le choix de ces seuils est motivé par la volonté de comparer notre figure à celle du projet *Archæomedes*. Nous avons choisi les mêmes seuils, auxquels nous avons ajouté les seuils (25 et 75 %) pour permettre de visualiser le comportement de tous les quartiles. La figure présentée dans l'ouvrage couvrait (quasiment) la même période³⁰, mais ne portait que sur les départements de la Drôme et de l'Ardèche, contrairement à nos productions qui portent sur la France métropolitaine (dans ses limites actuelles). Mais on observe toutefois une similarité très intéressante entre ces deux figures. Cette proximité s'explique probablement par la très forte ressemblance du taux d'urbanisation des ces deux départements avec l'ensemble de la France métropolitaine [3]. Mais il est nécessaire de montrer que le moteur de la concentration de la population est la transition urbaine.

Notez que les lignes verticales en pointillés (1820, 1920, 1975), illustrant des ruptures sont reprises de la figure précédemment citée [3, P. 149]. Elles ont été conservées car elles s'adaptent parfaitement à notre figure. On remarque que les phases relevées par le projet *Archæomedes* correspondent parfaitement aux seuils que nous avons identifiés lors de l'étude de la figure 1.4b. De plus, on arrive à identifier le dénombrement de 1821 comme le point de rupture entre une période de stabilité, et une période de concentration, là où l'étude de l'écart-type ne nous avait pas permis d'identifier une date précise, mais seulement une plage. Les quarts phases identifiées par le projet *Archæomedes* [3, 70] sont une première phase de *stabilité*, suivie d'une phase de *concentration*, de *concentration accélérée* et pour finir une phase de *deserrement*. Ces phases correspondent aux observations faites à partir des figure 1.4b, 1.6b et 1.6c. De 1793 à 1821 on n'observe pas une tendance particulière, l'écart-type fluctue et le coefficient de GINI diminue légèrement. Ce qui traduit « [...] un relatif équilibre de la répartition de la population à la période pré-industrielle [...] » [3, P.149]. Puis, à partir de 1821, commence une période de concentration de la population qui s'accélère au cours du XX^e siècle. Les auteurs du projet *Archæomedes* considèrent que cette accélération débute aux alentours de 1920 [3], mais aux vu des figures 1.6b et 1.6c il semble que ce seuil serait plutôt situé aux alentours de 1946 pour la France métropolitaine. Durant cette phase de concentration le coefficient de GINI passe de 0,49 (1821) à 0,79 (1975). Là où en 1821 environ 45 % des communes concentrent 80 % de la population, en 1975 moins de 20 % de ces dernières concentrent la même proportion de la population. Le recensement de 1975 marque une rupture nette dans l'évolution de la répartition de la population. À partir de cette date on observe une stabilisation, et une légère diminution du coefficient de GINI. Cette phase est qualifiée de *deserrement* dans l'ouvrage ARCHÆOMEDES [3]. Le recensement de 1999 (qui n'a pas été étudié par *Archæomedes*) s'inscrit dans la lignée des précédents recensements et dans la dynamique de *deserrement*. Cette phase est caractérisée par une déconcentration de la population, correspondant probablement à la dynamique de *péri-urbanisation* observée en France à partir des années 1970 [3].

La concentration de la population mise en évidence par les figures 1.6a, 1.6b et 1.6c illustre une importante modification de la distribution de la population française durant cette période. Cette transformation est qualifiée de *transition urbaine* [3]. Cette dénomination, proposée par ZELINSKI en 1969 [3], rappelle le phénomène de *transition démographique*. Ces deux termes décrivent de façon analogue le passage d'un régime à un autre. Le modèle de la *transition démographique* décrit le passage d'un régime avec une forte natalité et une forte mortalité, à un régime où la natalité et la mortalité sont faibles [79, 3, 80]. De la même manière, la *transition urbaine* décrit le passage d'un système de peuplement *pré-industriel*, défini par une faible urbanisation, à un système de peuplement *post-industriel*, fortement urbanisé [3]. Durant la *transition urbaine* le taux d'urbanisation croît suivant une courbe logistique (cf. figure 1.7), équivalente à la fonction illustrant la diffusion des innovations [3]. La première phase est caractérisée par la stabilité du taux d'urbanisation, ce dernier n'évolue quasiment pas, ou alors il le fait en suivant une croissance très lente [3]. Puis vient la phase de transition, dont « le début [...] peut être situé au moment où le taux d'urbanisation non seulement augmente plus vite que durant les périodes précédentes, mais surtout lorsque cette augmentation se maintient durant plusieurs périodes successives [...] » [3, P. 124], et qui est caractérisée par une augmentation très importante du taux d'urbanisation, ce dernier pouvant subir une hausse de 60 à 70 points. Cette phase est elle-même subdivisible en deux parties, séparées par le point d'inflexion de la courbe logistique. Ce dernier marque l'apogée de la vitesse de croissance, passé ce point le taux d'urbanisation croît de plus en plus lentement jusqu'à atteindre la dernière phase, celle de la stabilisation. Cette phase correspond à « [...] un certain niveau de maturation du système de peuplement industriel » [3, P. 124]. Système de peuplement qui tend alors à une situation d'équilibre dynamique [3].

La figure 1.8 illustre l'évolution du taux d'urbanisation³¹ entre 1700 et 2007, comme pour les figures 1.3a et 1.3b les données proviennent de multiples sources (précisées en légende). Les deux séries attribuées à l'INSEE diffèrent par leur définition de l'urbain. La classique définition en vigueur depuis 1856 a été

30. Pour être exact l'amplitude temporelle de la figure issue de l'ouvrage ARCHÆOMEDES [3] et de cent quatre-vingt-dix-sept ans, alors qu'elle est de deux cent six ans dans notre cas. La publication de l'ouvrage [3] étant antérieure au recensement de 1999.

31. *I.e.* le rapport entre la population urbaine et la population rurale [66].

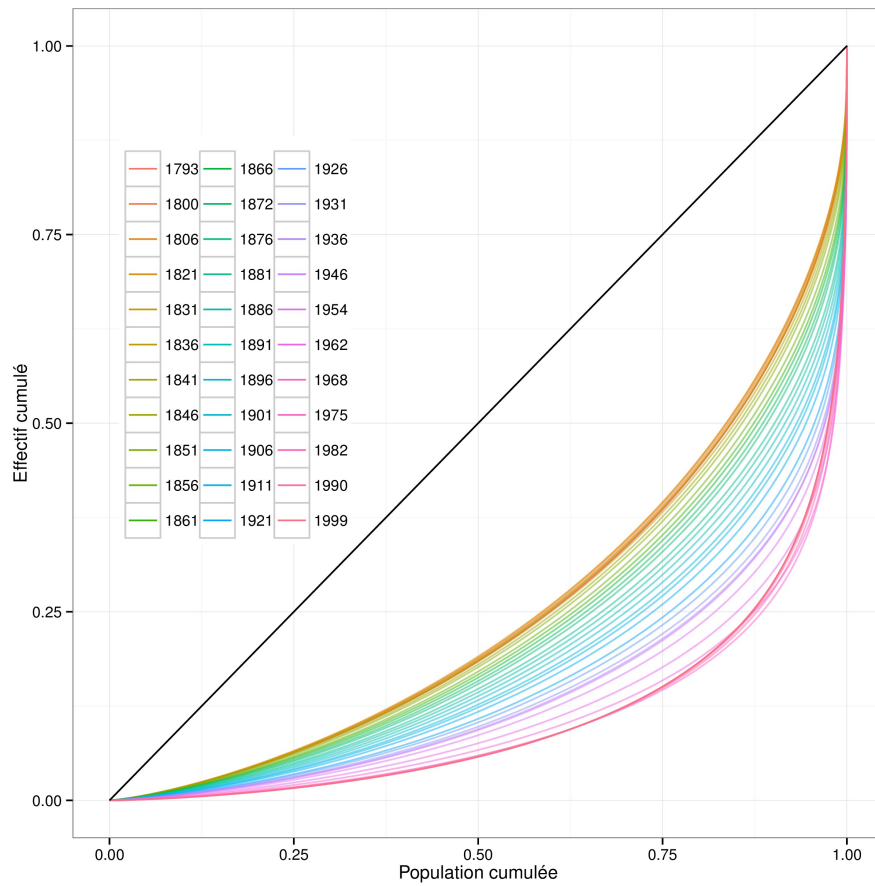
remplacée en 1962 par une nouvelle définition, plus complexe, le *zonage en aires urbaines*.

Cette figure permet de mettre en évidence le processus de *transition urbaine* décrit précédemment. On retrouve peu ou prou les trois phases précédemment décrites. de 1700 à 1800 le *taux d'urbanisation* n'évolue que très lentement, il progresse de 0,6 points sur cette période [79]. Nous sommes donc, jusqu'en 1800 en présence d'un *système de peuplement pré-industriel*, faiblement urbanisé et relativement stable. C'est à partir de 1800 que le *taux d'urbanisation* commence à croître de manière notable. Entre 1800 et 1830 ce dernier augmente de 2,8 points, puis il augmente de 3,8 points entre 1830 et 1850. Au vue de la figure 1.8, la croissance du taux d'urbanisation entre 1830 et 1850 correspond à celle observée entre 1850 et 1975. La croissance du taux d'urbanisation durant la période de transition ne suit donc pas une croissance exponentielle, comme présentée dans le modèle, mais linéaire [3]. Cette observation n'est pas valable sur la fin de la série puisque l'on observe une très forte croissance du taux d'urbanisation entre 1950 et 1975, précédant une phase de stabilisation du *taux d'urbanisation*. À partir de 1975 commence la dernière phase, celle de la stabilisation du *taux d'urbanisation* [3]. La vitesse de croissance du taux d'urbanisation diminue, même si ce dernier continue à croître à un rythme plus faible au vu des dernières données.

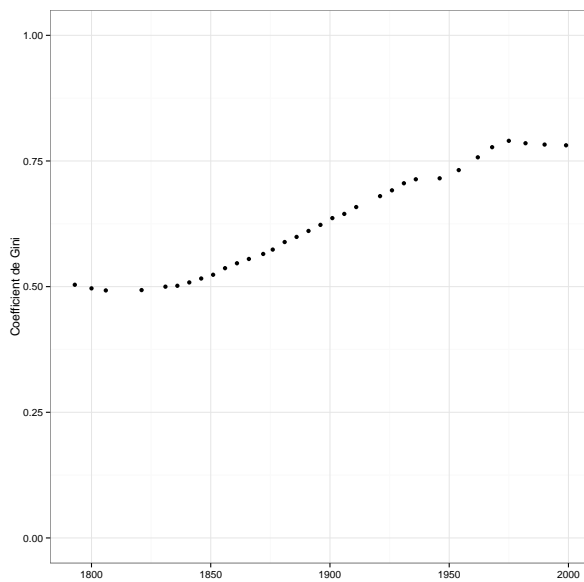
Le recours à la méthode présentée ici n'a pas été vain. En analysant l'évolution du coefficient de GINI il nous a été possible de mettre en évidence la tendance à concentration de la population en France métropolitaine entre 1793 et 1999, ce qui valide les observations réalisées à partir de l'écart-type. Mais nous ne nous sommes pas contentés de valider l'hypothèse de la concentration, l'utilisation de cette nouvelle méthode à permis de raffiner les résultats précédents, notamment en identifiant une phase de déconcentration à partir de 1975. Nous avons également put étudier l'évolution temporelle de cette concentration. Et il est intéressant de noter que cette dernière n'est pas linéaire, mais qu'au contraire, elle fluctue et son évolution est marquée par des ruptures. Ces mêmes ruptures ont été identifiées pour les départements de la Drôme et de l'Ardèche (dont la population cumulée varie entre 506 951 en 1793 et 723 801 en 1999, source : *cassini.ehess.fr*) par travaux du projet *Archæomedes* [3] en 1997 et correspondent à celles que nous observons à partir de nos données et qui portent sur la France métropolitaine. Le comportement de ces deux départements est donc similaire à celui de l'ensemble du pays sur la même période.

L'identification de ces quatre phases³² nous impose désormais de nous questionner sur les causes *socio-historiques* cette profonde modification de la répartition du peuplement. Nos différentes lectures [3, 32, 31] montrent que cette dernière est due au phénomène de *transition urbaine*, suivit d'une dynamique de *péri-urbanisation*. Toutefois la nature *aspatiale* de notre démarche ne nous permet pas d'identifier les communes concernées par ces mutations. De plus le phénomène de *transition urbaine* implique d'autres phénomènes tels que *l'exode rural* ou la *métropolisation*, qu'il est nécessaire d'étudier. Pour approfondir les répondre à ces questions il est nécessaire d'adopter une méthodologie prenant en compte la dimension spatiale. Et c'est en combinant une approche exploratoire (cartographie) et des outils statistiques *ad hoc* que nous traiterons ces nouvelles questions.

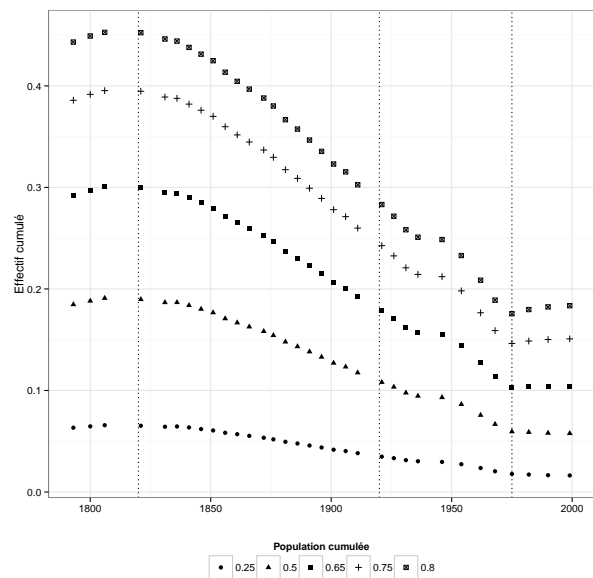
32. Stabilité, concentration, concentration accélérée et desserrement.



(a) Mesure de la concentration de la population française pour chaque dénombrement



(b) Évolution du coefficient de GINI



(c) Concentration de la population française.

FIGURE 1.6 – Concentration de la population française entre 1793 et 1999.

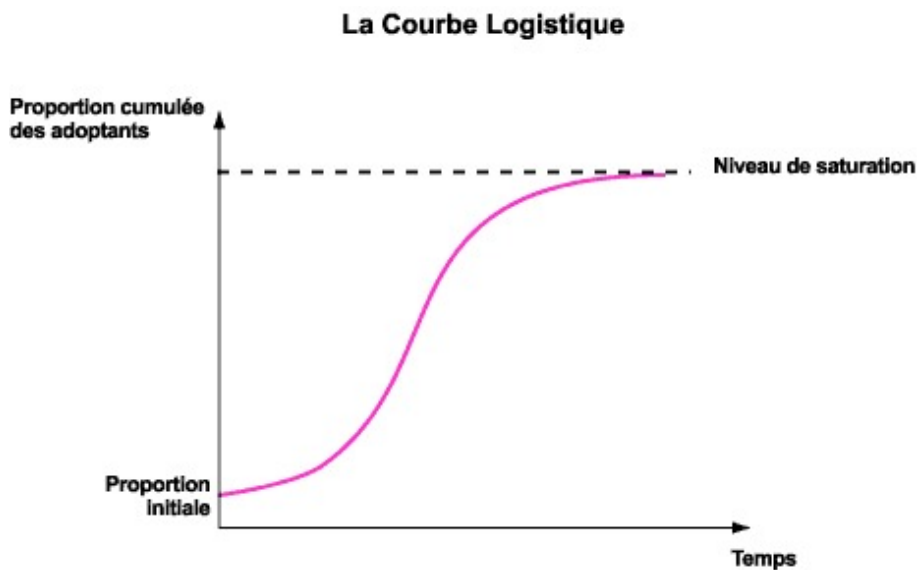


FIGURE 1.7 – Courbe logistique illustrant le processus de diffusion des innovations d'HÄGERSTRAND et pouvant être appliqué au modèle de la *transition urbaine* [3, 73], extrait de SAINT-JULIEN [73].

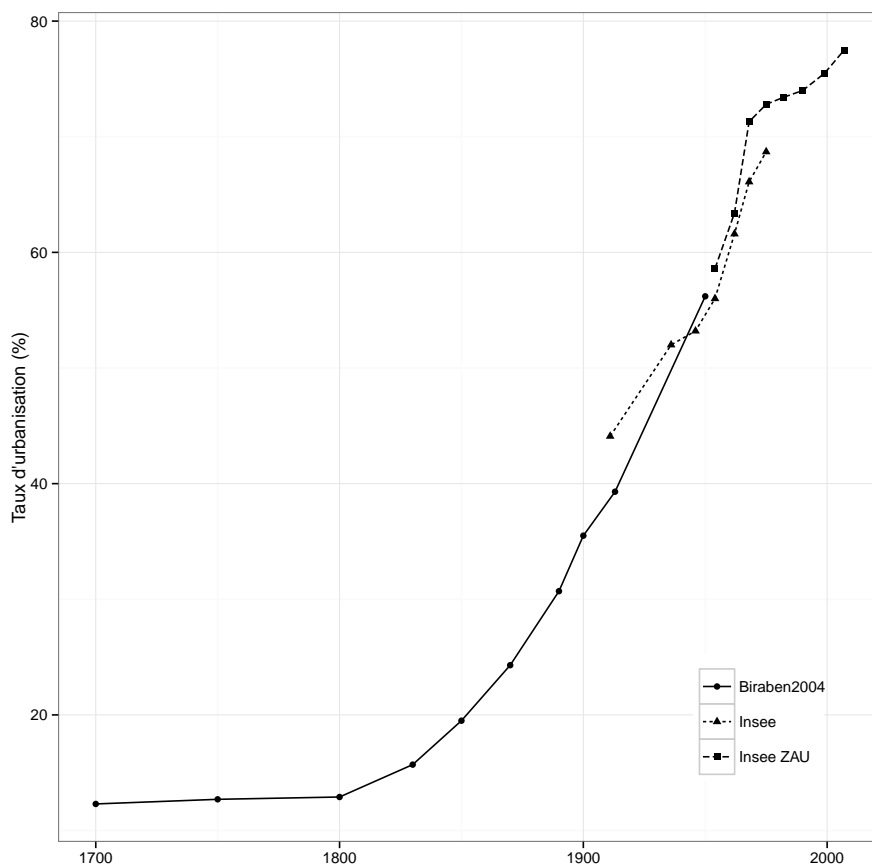


FIGURE 1.8 – Taux d'urbanisation en France entre 1700 et 2007. Sources : INSEE, d'après RONCAYOLO [72] de 1911 à 1982, *Insee.fr* de 1990 à 2007, BIRABEN [10] et VALLIN [79]. Pour les données issues de BIRABEN [10] et VALLIN [79] le taux d'urbanisation correspond à la part de la population des villes de plus de 5 000 habitants [79, P. 68].

1.3 Spécificités locales

Nos précédents travaux constituent une base solide de travail pour les étapes suivantes. Nous avons pu montrer l'existence d'une tendance successive à la concentration puis à la déconcentration de la population en France métropolitaine, mais également étudier les variations de cette tendance. Certes la recherche bibliographique nous informe sur les raisons et les espaces à l'origine de cette concentration, mais nos explorations statistiques ne nous permettent pas d'observer les phénomènes de *métropolisation*, ou *d'exode rural* mais seulement leurs effets à l'échelle nationale. La raison est que les méthodes employées sont *aspatiales*, en aucun cas la dimension géographique du phénomène n'a été étudiée. Nous nous sommes contentés de traiter les communes comme des individus statistiques, sans prendre en compte leur dimension spatiale. Ainsi si nous savons que la tendance à la concentration observée statistiquement entre 1793 et 1999 implique une densification relative de certaines communes et, inversement, une perte d'importance relative d'autres communes, nous n'avons, à ce stade aucune idée de la répartition spatiales des communes concernées. Or pour valider l'hypothèse que la concentration de la population observée précédemment est la résultante du phénomène de *transition urbaine* il nous incombe de montrer que nous sommes en présence d'« [...] une modification majeure de la répartition du peuplement dans l'espace, avec la concentration de la plus grande partie de la population dans les villes. » [3, P.117]

L'objectif de cette sous-partie est donc d'approfondir nos précédentes observations à l'aide de différentes méthodes (cartographie thématique, *lisa*, *K-means*) permettant d'une manière ou d'une autre de prendre en compte la dimension spatiale du phénomène. Avec cette nouvelle approche nous pourrions répondre à la question du *où ?* Il sera ainsi possible de savoir dans quelles communes la population se concentre et au détriment de quelles communes. Ainsi nous serons en mesure de savoir si nos données valident les hypothèses socio-géographiques précédemment formulées.

1.3.1 Cartographie thématique

Une première façon d'aborder cette nouvelle problématique consiste à cartographier notre variable de travail, à savoir *l'écart à la densité moyenne*. Il en résulte deux séries de trente-trois cartes (une pour chaque recensement) présentées dans leur intégralité dans les annexes de ce mémoire. Pour des raisons évidentes de place, nous ne commenterons pas l'intégralité de ces cartes, mais seulement six d'entre elles (le cinquième de la série), permettant la meilleure visualisation de l'évolution de la distribution de la population française au cours de ces deux-cent-six ans. Il en résulte la planche 1.9, où la variable *écart à la densité moyenne* est cartographiée selon une discrétisation standard portant sur l'ensemble des recensements (la discrétisation est donc identique quelle que soit la carte). Les six dates retenues sont le début et la fin de la série (1793 et 1999), le premier dénombrement du XX^e siècle (1901), les dénombrements suivants les deux guerres mondiales (1921 et 1946) et enfin le recensement de 1975. Les transformations du XIX^e siècle sont si minimes (relativement) qu'elles sont figurables en deux cartes, là où les changements de l'après deuxième guerre mondiale sont si importants que la moitié des cartes présentes sur cette planche concernent les cinquante-trois dernières années de la série. Cette période est au vu des cartes (mais également des précédentes productions) la période où les transformations observables sont les plus importantes.

La carte de 1793 est la première de notre corpus. C'est également la seule carte de la planche 1.9 où un grand nombre de communes sont absentes. Les communes dont la densité est très supérieure à la moyenne ($e \geq 0,94$) sont assez rares et correspondent aux plus grandes villes françaises. Sans rentrer dans une longue énumération on retrouve notamment les communes de Paris, Marseille et Lyon, ou encore Toulouse, Bordeaux ou Strasbourg, mais également des villes moins peuplées telles que Toulon, Brest ou le Havre. On notera que la rapide décroissance de l'écart à la densité moyenne au fur et à mesure que l'on s'éloigne des grandes villes. Il est fréquent que ces dernières aient pour voisines des communes dont la densité est inférieure à la densité moyenne. Les autres communes dont la densité est supérieure à la densité moyenne ne sont pas distribuées ponctuellement, mais au contraire elle forment des étendues compactes. On repère notamment le cours de la Garonne, ou celui de la Loire (à partir d'Orléan). Mais la structure la plus visible est le cordon littoral s'étendant du nord de la Bretagne à la limite sud de la *côte d'opale*. Une autre structure notable est la présence de communes plus denses que la moyenne le long de la moitié la plus septentrionale de la frontière Belge. Inversement on note également la présence d'importantes zones de faible densité, notamment les Pyrénées, les Alpes et leur piémont, l'Aquitaine ou encore la Sologne. On distingue également la fameuse *diagonale du vide*, qui se manifeste par une large bande de communes moins denses que la moyenne s'étendant du sud de la frontière Belge aux Pyrénées.

Si l'on compare la carte de 1901 avec celle de 1793 on constate qu'en plus d'un siècle la distribution de la population n'a guère évolué. La structure reste fortement similaire, et les oppositions précédemment révélées restent d'actualité. Ainsi, comme l'indiquent MAURICE GARDEN et HERVÉ LE BRAS « *De 1801 à 1866, en deux tiers de siècles, et malgré une importante croissance démographique, presque rien n'a*

changé dans la concentration spatiale de ces 38 millions de Français » [31, P. 139]. Toutefois malgré la stabilité de la structure spatiale du phénomène on observe des mutations, et plus particulièrement de deux type, une extension des zones les plus denses et, plus généralement, un *renforcement des contrastes*. Le premier point est particulièrement visible pour quelques agglomérations, notamment Lyon, Strasbourg, Rouen, mais, surtout Paris. Pour lesquelles on observe une extension de l'aire appartenant à la classe des communes les plus denses. On observe donc une augmentation de l'aire densément peuplée et par conséquent une augmentation de la taille de certaines villes. Le cas de la commune de Lille est singulier. La densification ne concerne par uniquement les communes adjacentes à la ville de Lille, mais tout une bande parallèle à la frontière Belge. Cette structure unique contraste fortement avec le reste du territoire, et même si ces prémisses étaient visibles sur la carte 1.9a il s'agit de la région qui a connu la transformation la plus visible entre 1793 et 1901 (avec la région parisienne). Le second point notable est le *renforcement des contrastes* qui s'opère entre ces deux dates. L'écart entre les zones les moins denses et les plus denses est plus visible en 1901 qu'en 1793, et les étendues de communes plus denses que la moyenne apparaissent plus clairement. C'est notamment les cas de la vallée du Rhône, frontalière au massif alpin dont la densité relative à clairement diminué entre les cartes 1.9a et 1.9b. C'est également le cas de la vallée du Rhin dont la densification est notable. Concernant la Corse qui fait ici son apparition dans la série statistique elle est majoritairement composée de communes dont la densité est inférieure à la moyenne à l'exception notable des communes d'Ajaccio et de Bastia.

En vingt ans les transformations sont assez minimes. Les cartes 1.9b et 1.9c se ressemblent fortement. Les principales différences sont visibles au niveau de la région parisienne, et du département du Nord. Ces deux zones ce sont visiblement densifiées, mais de deux manières. Pour la région parisienne cette évolution est comparable à celle observée entre 1793 et 1901, le nombre de communes présentant une densité nettement supérieure à la moyenne augmente, contrairement au département du nord pour lequel on observe pas une augmentation du nombre de communes les plus denses, mais une densification des communes déjà plus denses que la moyenne. On observe donc un renforcement de la structure, et non une extension spatiale comme c'est le cas pour la région parisienne.

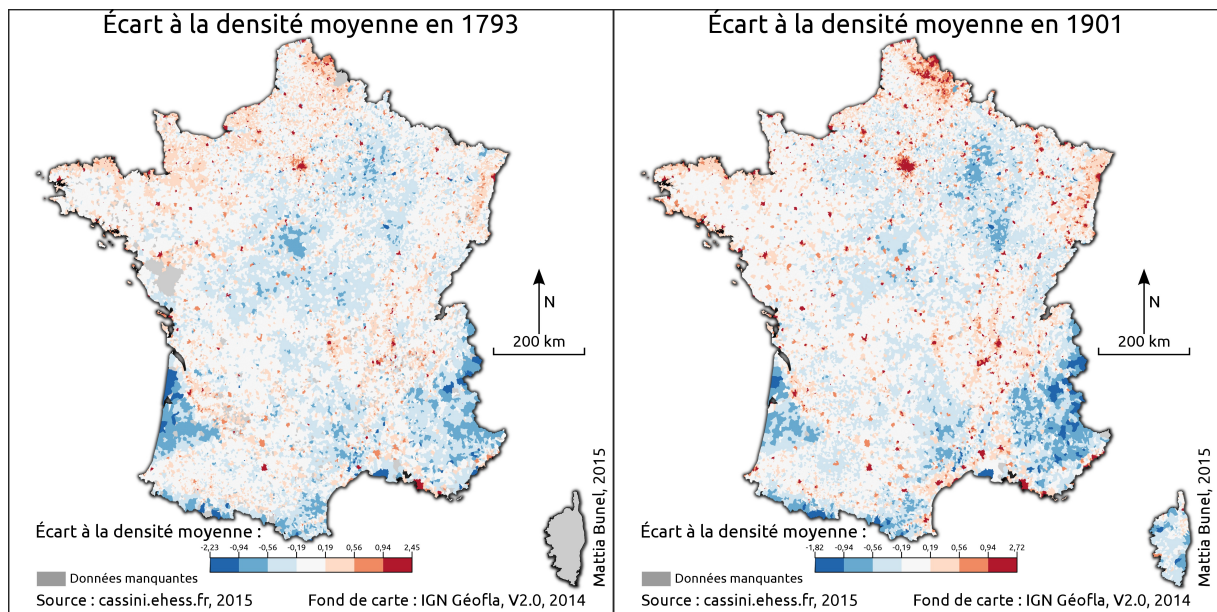
Les transformations observées entre 1946 et 1921 sont comparables à celles observées entre 1901 et 1921. Il n'y a pas de transformation profonde de la distribution de la population, mais les grandes agglomérations sont de plus en plus importantes. Cette prise d'importance ce fait au détriment des zones les moins denses. C'est au niveau de l'agglomération parisienne que cette augmentation est, une fois encore, la plus visible. À l'opposé on continue d'observer une *dé-densification* des zones montagneuses et de la *diagonale du vide*.

La comparaison des cartes 1.9d et 1.9e nous montre à quel point la distribution spatiale de la population à été modifiée durant cette période. Là où les principales transformations étaient avant tout visibles au niveau de la région parisienne, elles concernent désormais toutes les grandes agglomérations. Comme pour la région parisienne entre 1793 et 1975, les communes voisines aux grandes villes ont vu leur densité augmenter, de la même manière on observe clairement autour des agglomérations l'existence d'un gradient densité/distance.

La carte 1.9f marque l'apogée d'une structure dont on devinait les prémices dès le premier dénombrement. Les principales villes, ainsi que leur proche banlieue se sont densifiées contrairement au reste du territoire qui n'a cessé de perdre en densité. Les vastes zones de faible densité sont plus marquées. C'est notamment le cas de la *diagonale du vide*, plus visible que jamais.

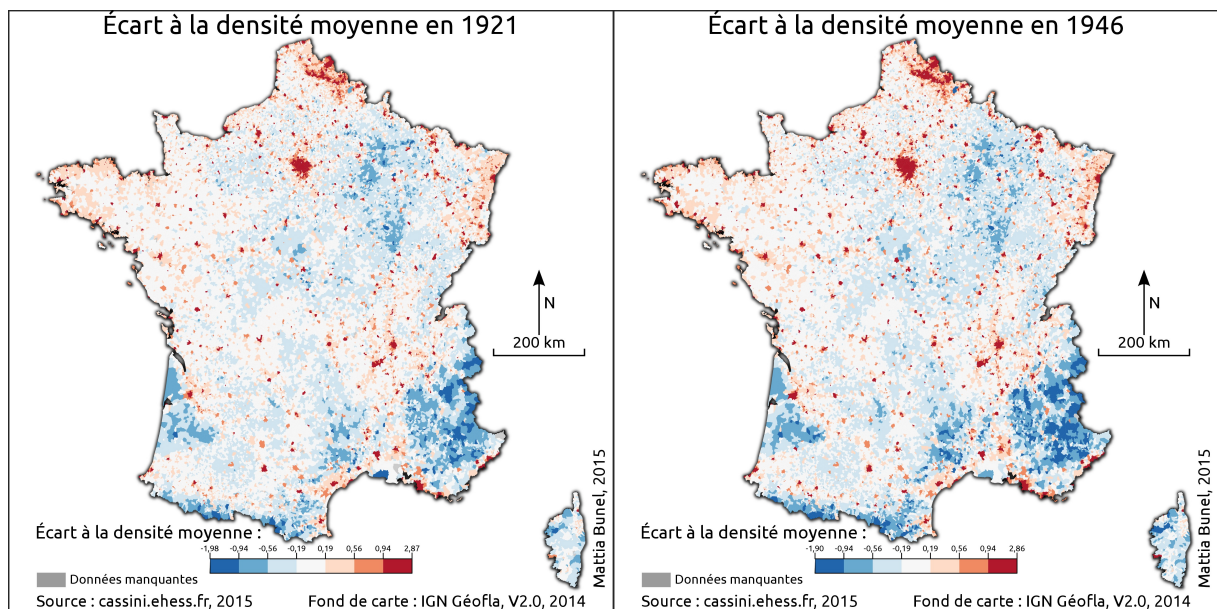
Après ce commentaire point par point il est nécessaire de revenir sur l'ensemble des transformations. On remarquera que les précédents commentaires traitent systématiquement des mêmes espaces. Les agglomérations puis leurs banlieues ne cessent de prendre de l'importance et ce au prix de la perte d'importance du reste du territoire, *i.e.* des communes montagnardes et rurales. Si l'on compare les cartes 1.9a et 1.9f on constate que les structures visibles à la fin de la série statistique existent dès 1793. Ces cartes nous permettent de faire deux observations que l'on pourrait supposer incompatibles, dans un premier temps on remarque une profonde modification de la distribution de la population, cette dernière ce concentre dans les agglomérations et leurs banlieues, au prix d'un abandon des zones rurales et montagneuses. En regardant la série cartographique dans son ensemble, on observe une augmentation des contrastes entre urbain et rural, induite par l'augmentation de l'amplitude entre faibles densités et fortes densités. Ce phénomène est d'ailleurs observable sur la figure 1.5a, où l'on observe clairement l'augmentation de l'amplitude entre 1793 et 2006. Mais ce « creusement des écarts » n'est pas incompatible avec la conservation d'une trame de peuplement observable à chaque date. Ainsi bien que les écarts augmentent les zones les plus denses en 1999 correspondent aux zones les plus denses de 1793, et inversement. De plus la hiérarchie entre les différentes agglomérations reste stable [32]. Sur ces deux siècles on observe une permanence de la trame du réseaux urbain, et ce malgré les profondes observations observées [32].

Ce que l'on observe entre les deux extrêmes de la série statistique n'est pas l'apparition *ex-nihilo* d'une nouvelle structure mais au contraire l'évolution d'une précédente structure. Pour approfondir et confirmer ces observations nous allons recourir à une méthode, les indicateurs locaux d'association spatiale (*lisa*).



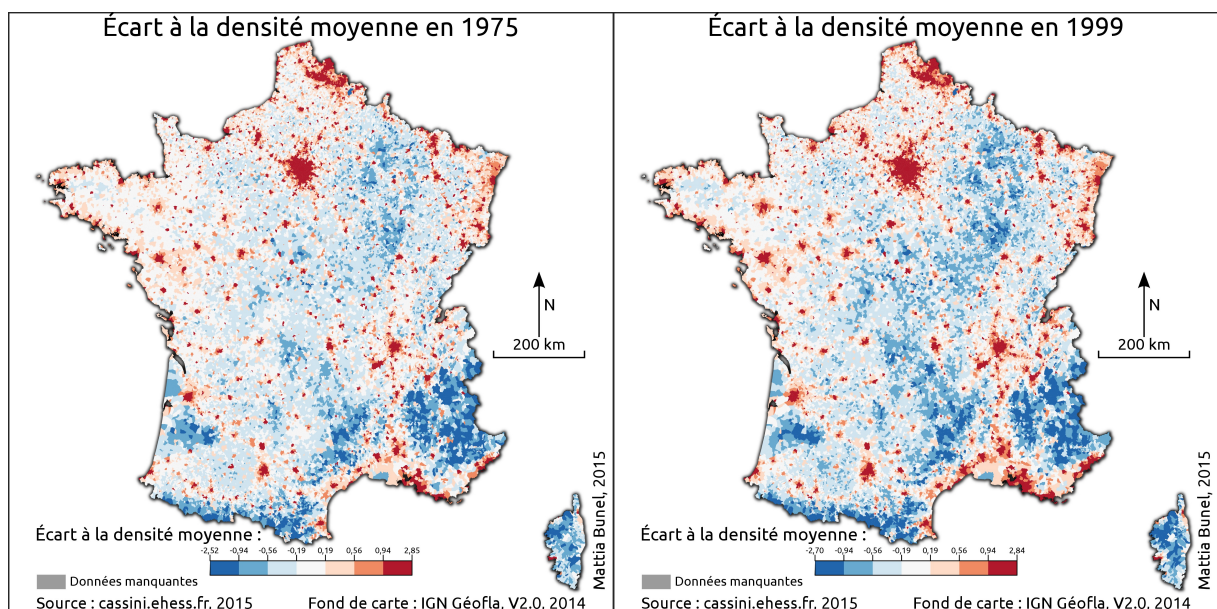
(a) Écart à la densité moyenne en 1793.

(b) Écart à la densité moyenne en 1901.



(c) Écart à la densité moyenne en 1921.

(d) Écart à la densité moyenne en 1946.



(e) Écart à la densité moyenne en 1975.

(f) Écart à la densité moyenne en 1999.

FIGURE 1.9 – Cartographie de la variable écart à la densité moyenne.

1.3.2 Indicateurs locaux d'association spatiale

Une des manières de quantifier l'importance des structures spatiales observées sur les figures 1.9a à 1.9f et de « mesur[er] l'intensité de la relation entre proximité des lieux et leur degré de ressemblance » [70, P. 133], *i.e.* l'autocorrélation spatiale. Les deux coefficients régulièrement utilisés sont les coefficients de GERS et de MORAN. Ce dernier est généralement préféré du fait de sa plus grande stabilité [70]. Le coefficient de MORAN mesure, de façon analogue au r de BRAVAIS-PEARSON, le rapport entre la covariance et le produit des écarts-types, ce qui revient à mesurer la corrélation entre deux variables, où la variable dépendante est expliquée en fonction de la valeur moyenne des individus voisins (plus de détails ci-dessous).

Il en résulte les figures 1.10a et 1.10b, présentant la même information mais sous deux formes différentes. La figure 1.10a représente l'évolution du coefficient de MORAN en fonction de la date, là où la figure 1.10b représente son évolution en fonction du rang de voisinage. Les dix courbes de la figure 1.10a correspondent à chaque rang de voisinage, là où les trente-trois courbes de la figure 1.10b correspondent à chaque rang. Comme précisé dans la légende nous nous sommes basés sur un voisinage par contiguïté nodale (voisinage de MOORE) de rang variable. Les rang supérieurs à un intègrent les voisins de rang inférieur, ainsi un voisinage de rang 3 prend en compte : « mes voisins, les voisins de mes voisins, et les voisins des voisins de mes voisins », on remarque aisément que l'augmentation du nombre de voisins est exponentielle, à titre d'exemple la matrice de voisinage de rang un est composée de 218 230 paires de voisins, avec une moyenne de 6 voisins par individus, le nombre de paires de voisins passe à 4 487 562 pour un voisinage de rang 5 (moyenne de 123), et atteint la somme de 18 319 734 paires de voisins et un nombre de voisins moyen de 501 pour un voisinage de rang 10. La figure 1.10b nous montre que le I de MORAN calculé par nos soins diminue en fonction du rang. Cette diminution continue n'est pas systématique, dans certains cas De la même manière la figure 1.10a nous permet d'observer l'évolution du coefficient de MORAN au cours du temps. De la même manière que l'écart-type ou le coefficient de GINI (cf. figures 1.4b et 1.6b) le coefficient de MORAN suit une tendance à la hausse. La valeur minimale correspond au dénombrement de 1793 (0,45 au rang 1) et le maximum correspond au recensement de 1999 (0,71). Sachant que l'indice de MORAN peut-être interprété comme la part de la variance expliquée par le voisinage [62], on passe d'une situation où 45 % de la variance est expliquée par les relations de voisinage, à une situation où 71 % de la variance est expliquée par les relations de voisinage. Compte tenu du grand nombre d'individus traités, les valeurs de l'indice de MORAN, obtenues sont importantes et signalent l'importance de la structure spatiale. Les valeurs positives de l'indice de MORAN indiquent que les individus proches ont tendance à se ressembler [70], et si l'on observe la figure 1.10a on constate que cette ressemblance ne fait qu'augmenter. Dans son ensemble l'étude de l'évolution de l'indice de MORAN corrobore donc les observations précédemment réalisées à partir de l'écart-type et du coefficient de GINI. Ainsi l'augmentation de l'écart-type est caractéristique d'une concentration de la population, laquelle est confirmée par l'augmentation du coefficient de GINI, or par définition une concentration de la population implique que des communes voisines partagent des population (dans notre cas une densité) comparables, et donc qu'elles se « ressemblent » ce qui implique une autocorrélation spatiale mesurée par l'indice de MORAN. L'augmentation de l'indice de MORAN est donc parfaitement comparable à l'augmentation de l'écart-type et du coefficient de GINI, cette dernière illustre le même phénomène, que l'on suppose être la *transition urbaine*.

L'utilisation des indicateurs locaux d'association spatiale (*local indicators of spatial association*, Lisa), méthode basée sur la désagrégation locale de l'indice de MORAN à pour objectif d'approfondir les observations réalisées à partir de l'écart-type, du coefficient de GINI et de l'indice de MORAN en introduisant la dimension spatiale dans cette étude. Les trois méthodes précédemment citées ne nous ont donné qu'une « image » du phénomène étudié à l'échelle métropolitaine, or ce que l'on observe est clairement une transformation profonde de la répartition de la population française, la lecture de divers ouvrages permet de supposer que le phénomène observé est la *transition urbaine* [3, 32], ce qui est corroboré par les observations réalisées à partir de la cartographie de la densité centrée. Mais nous n'avons employé aucune méthode statistique prenant en compte l'espace, l'analyse de la cartographie thématique étant avant une approche exploratoire. L'objectif est ici de valider l'hypothèse précédemment formulée d'une concentration de la population induite par la *transition urbaine*, en identifiant les zones de concentration et de déconcentration de la population de manière plus précise que ne le permet la cartographie thématique univariée.

Les *indicateurs locaux d'association spatiale* sont une méthode statistique développée par LUC ANSELIN en 1995 [2, 62, 64]. On peut résumer l'idée d'ANSELIN assez simplement, elle consiste en la désagrégation locale d'un indice global, en l'occurrence le I de MORAN. Ce dernier est exprimé de la manière suivante [62, 70] :

$$I = \frac{n}{m} \cdot \frac{\sum_i \sum_j w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2} \quad (1.7)$$

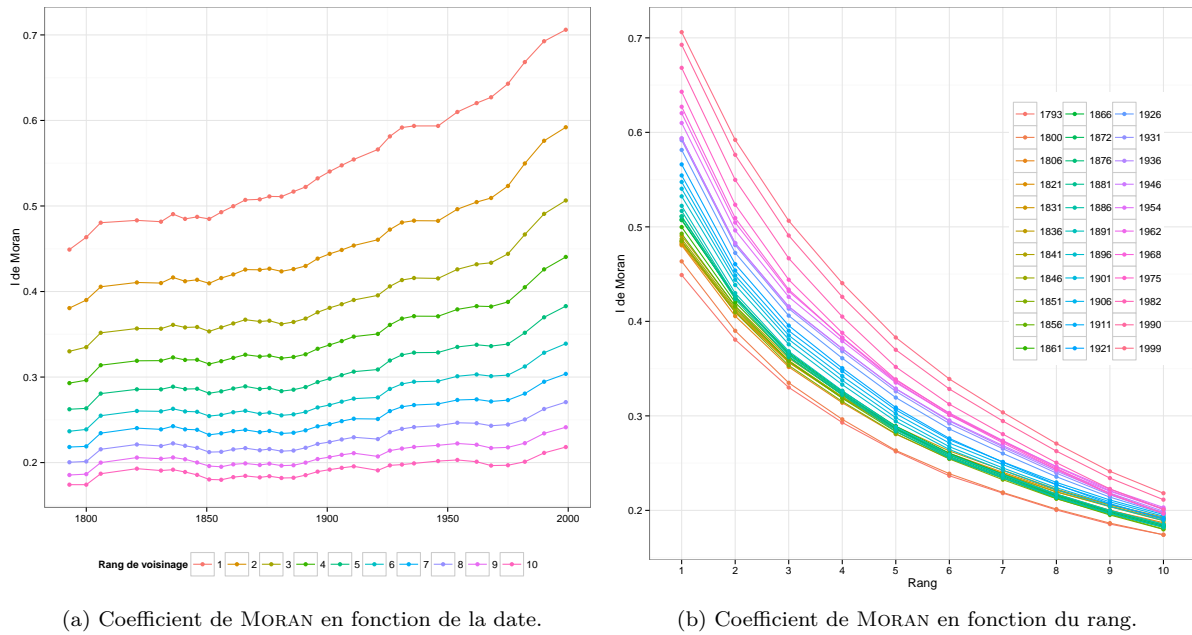


FIGURE 1.10 – Évolution du coefficient de MORAN pour la variable *écart à la densité moyenne* en fonction de l'année du recensement et du rang de voisinage de MOORE (contiguïté nodale) [1, 64].

Avec n le nombre d'individus dans l'échantillon, m le nombre de paires de voisins, w_{ij} la matrice de pondération (constante dans notre cas) i un individu et j un voisin de l'individu i . On remarque la très forte similarité de cette formule avec celle du coefficient de corrélation linéaire formulé de la manière suivante [74] :

$$r = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \quad (1.8)$$

Avec σ_{xy} la covariance qui est égale à [74] :

$$\sigma_{xy} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{y})(x_j - \bar{y}) \quad (1.9)$$

Et $\sigma_x \sigma_y$ le produit des écarts-types. Ainsi la structure de ces deux coefficients est quasiment identique, là où pour le coefficient de corrélation linéaire r on mesure la linéarité de la relation entre deux variables, le I de MORAN estime quant à lui la linéarité de la relation qui unit une variable à elle-même compte tenu des relations de voisinage. Le I de MORAN peut d'ailleurs être exprimé sous la forme [62] :

$$I = \frac{cov}{\sigma_x^2} = \frac{cov}{\sigma_x \sigma_x} \quad (1.10)$$

Avec cov la covariance et σ_x^2 la variance, soit le carré de l'écart-type. ANSELIN [2] à donc imaginé un indicateurs I_i dont la moyenne donnerait le I de MORAN [62]. Cette statistique, nommée I de MORAN local est exprimé par la formule suivante dans le cas où la matrice de voisinage est standardisée ($n = m$, *i.e.* chaque individu contribue de la même manière au I de MORAN Global) [63, 62] :

$$I_i = \frac{\sum_j w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_k - \bar{x})^2} \quad (1.11)$$

Les intérêts de recourir à un modèle local sont multiples, le tableau 1.4 est adapté d'une figure proposée par FOTHERINGHAM, BRUNSDON et CHARLTON [38] dans leur ouvrage *Geographically Weighted Regression*. Ce tableau a à l'origine pour but de montrer les distinctions entre le classique modèle de la régression linéaire et la régression géographiquement pondérée (désagrégation locale du précédent modèle), mais avec quelques adaptations ce dernier nous montre également les apports plus généraux de la désagrégation locale. Le recours à une méthode statistique désagrégée permet de cartographier les résultats et d'en tirer des conclusion sur la structure spatiale de phénomènes étudiés. Pour ce qui est des *lisa* Différents indicateurs sont cartographiables, le premier est l'indice de MORAN local (cf. équation 1.11), dont la moyenne correspond à l'indice global [62]. La cartographie de cet indice permet d'identifier les individus participant le plus à l'indice global, il est convient de préciser que la valeur des I locaux

Modèle global	Modèle local
Statistiques globales	Désagrégation locale de statistiques globales
Coefficient unique	Coefficients propres à chaque individu
Difficilement cartographiable	Facilement cartographiable
Non adapté aux SIG	Adapté aux SIG

TABLE 1.4 – Modèles locaux *vs* modèles globaux, traduit et adapté de FOTHERINGHAM, BRUNSDON et CHARLTON [38].

n'est pas bornée. Une autre cartographie possible est celle des clusters, il s'agit d'une carte combinant les informations sur la significativité *lisa* (estimé par le biais d'une méthode de Monte-Carlo³³ [2, 63]), l'autocorrélation spatiale et la valeur de la variable [63]. Cette dernière cartographie la ressemblance d'une commune à ses voisines en les regroupant en cinq classes basées sur les critères précédemment mentionnés [63] :

- Communes non significatives.
- Communes avec une densité (centrée) forte dans un voisinage de communes avec une densité forte.
- Communes avec une densité faible dans un voisinage de communes avec une densité faible.
- Communes avec une densité fortes dans un voisinage de communes avec une densité faible.
- Communes avec une densité faible dans un voisinage de communes avec une densité forte.

Cette cartographie permet donc de mettre en évidence « *des sous-espaces homogènes* » [63, P. 6], correspondant dans notre cas à des zones de forte ou de faible densité. Ces clusters sont également identifiables sur les diagrammes de MORAN (cf. 1.12) représentant en abscisse la valeur de la variable étudiée pour un individu et en ordonnée la valeur moyenne de la variable pour les voisins de l'individu. Lorsqu'il y a autocorrélation spatiale le nuage de point obtenu est de forme allongée et suit une progression partant du quart inférieur gauche à quart supérieur droit. Les individus situés à proximité de cette diagonale sont ceux dont le voisinage présente des valeurs semblables à la leur. Les quart supérieur gauche, et inférieur droit sont quant à eux occupés par les *outliers*, individus qui ne correspondent pas à leur voisinage. Ainsi chaque quart correspond à une des catégories précédemment mentionnée, le quart 1 (dans le sens des aiguilles d'une montre en partant du quart supérieur gauche) contient les individus avec un écart à la densité moyenne négatif avec un voisinage composé d'individus avec un écart à la densité moyenne positif, le quart 2 à des individus avec une valeur positive dans un voisinage de valeurs positives, le quart 3 à des individus avec une valeur positive dans un voisinage de valeurs négative et le quart 4 à des individus avec une valeur positive négative dans un voisinage de valeurs négatives, les individus non significatifs tendent quant à eux à ce concentrer à proximité du centre.

De ces trois éléments la carte de clusters, qui synthétise l'ensemble des informations, est la plus intéressante à analyser. Comme pour toutes nos précédentes productions, la structure de carte est profondément modifiée au cours du temps. Comme expliqué précédemment quarts clusters sont présents, auxquels s'ajoutent deux autres classes, une pour les communes non significatives et une classe pour les communes sans données. Les cartes 1.11 nous permettent de mieux percevoir les structures décrites précédemment, en 1793 les zones les plus denses sont localisés en région parisienne, dans la vallée du Rhin et le piémont vosgien, ainsi qu'en région lyonnaise et bordelaise. Le nord de la France présente également une structure intéressante. Villes moins visibles

En un siècle la structure présente le long de la côte normande, précédemment décrite, a grandement diminué, le « *ruban* » quasi continu qui liait les côtes d'armor à la côte d'opale est désormais rompu au niveau de la Basse et de la Haute Normandie. Les seuls clusters positifs qui demeurent sont quelques agglomérations, de taille variable, telles que le Havre, Rouen, Cherbourg ou encore Caen. Si le centre de la précédente structure c'est « *affaïssé* », les deux structures périphérique ce sont renforcées, le cluster breton, qui n'intégrait que les communes côtières du Nord de la Bretagne continue désormais jusqu'à la limite est du Morbihan, la structure s'est donc étendue, là où le cluster présent à la frontière Belge n'a pas réellement connu d'extension spatiale, mais au contraire un renforcement de la structure spatiale. Le cluster de valeurs positives centré sur Paris à également vu sa taille augmenter. Concernant les clusters composés de communes dont la densité est inférieure à la densité moyenne, on remarque la diminution du nombre de communes incluses dans le cluster situé au sud de la région parisienne.

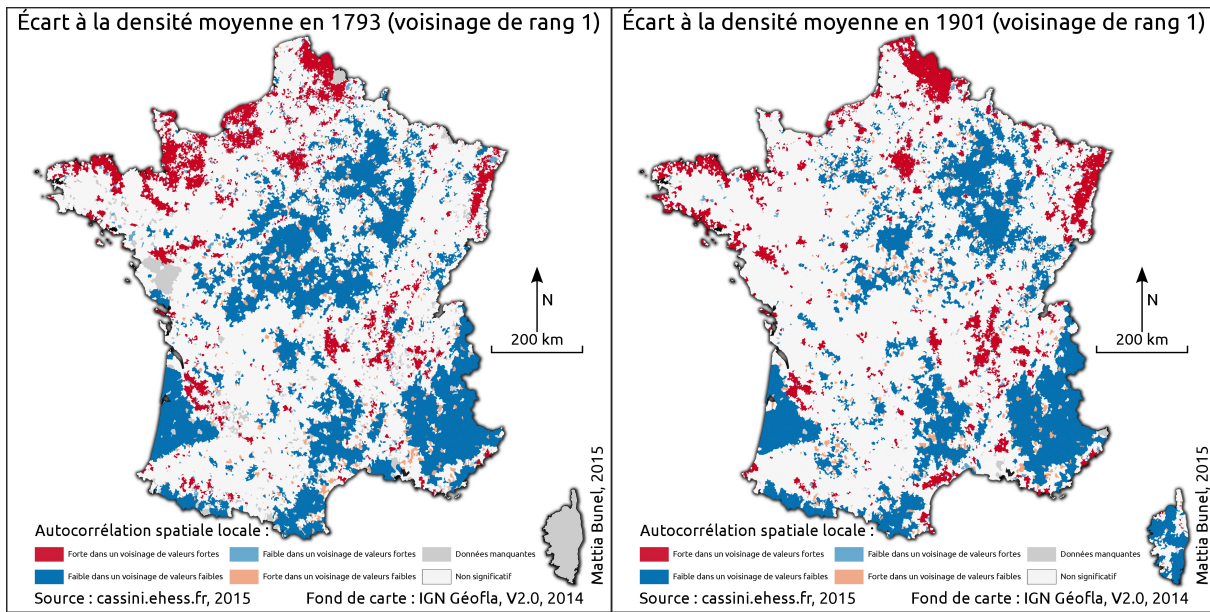
Comme précédemment, la période 1921-1946 est caractérisée par une augmentation de la taille des clusters centrés sur les grandes agglomérations, cette évolution concerne avant tout Paris, mais également des agglomérations de taille plus réduite, même si le contraste est moins important qu'entre 1901 et 1921. On remarque également que certaines communes, qui appartenaient à la classe une (fortes densités dans un voisinages de fortes densités) en 1921 quittent cette dernière pour rejoindre la classe des communes non significatives.

33. Pour toutes les réalisations à venir la significativité à été estimée à partir de 100 permutations, le seuil de significativité retenu est de 95 %.

Les changements observés entre 1946 et 1975, sont très importants. Si les évolutions précédentes laissaient déjà percevoir une hausse constante de l'importance des villes, c'est entre 1946 et 1975 que ce changement est le plus visible. Les clusters positifs sont toujours centrés sur les grandes agglomérations, mais leur taille a considérablement augmenté. Inversement la diagonale du vide semble de plus en plus présente. Là où elle était préalablement constituée de deux clusters non reliés, elle apparaît désormais quasi-continue, reliant les Ardennes aux Pyrénées.

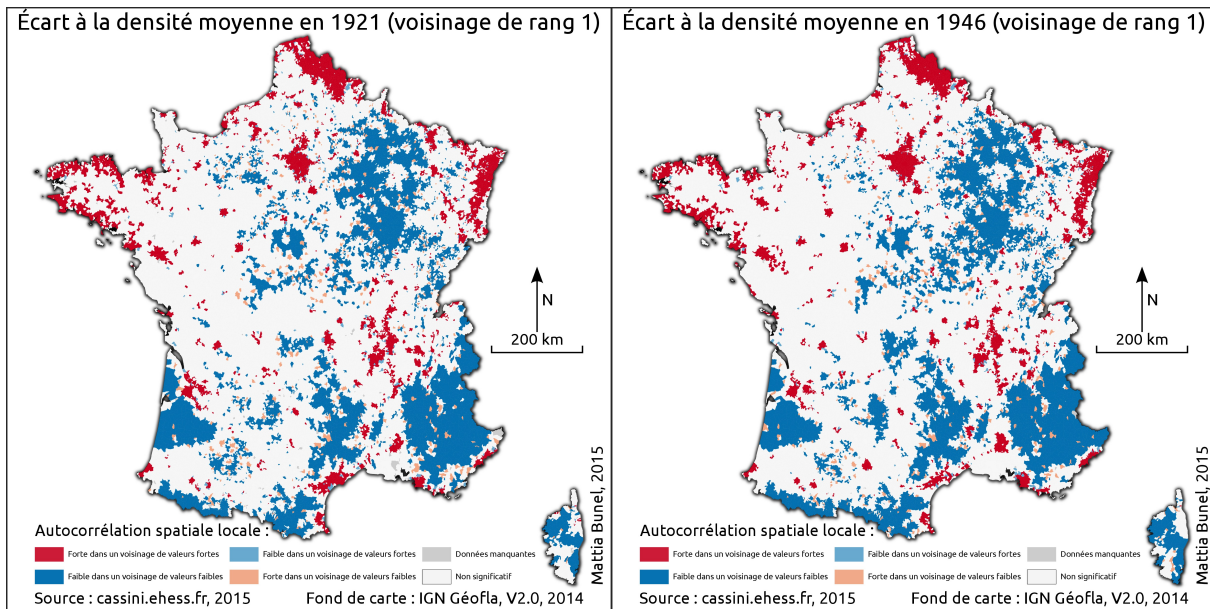
En 1999 on constate un renforcement de la structure déjà décrite en 1975. Les clusters restent semblables dans leur implantation, mais leur étendue augmente de façon conséquente. La région Lyonnaise et le sud de la vallée du Rhône en sont des exemples élogieux. La diagonale du vide, dont l'importance est ici maximale. Comme pour la série de cartes 1.9 c'est en 1999 que cette structure est la plus visible sur les cartes des clusters. Cette dernière est désormais parfaitement continue.

La figure 1.13 présente l'évolution de l'indice de MORAN local. À la vue des résultats issus de la cartographie des clusters (cf. figure 1.11) nous nous sommes demandés quelles étaient les communes qui participaient le plus à l'augmentation du I de MORAN global. La cartographie des I de MORAN locaux permet de répondre à cette question. Si l'on regarde tour à tour les six cartes composant la planche étudiée on peut juger de l'important changement dans la distribution des I locaux. En 1793 les fortes valeurs des I locaux sont correspondent majoritairement à des régions dont la densité est inférieure à la moyenne. Les plus visibles sont notamment l'Aquitaine, le massif alpin et la Sologne. Les zones les plus denses sont également visibles, notamment Paris, le département du Nord ou la vallée du Rhin mais elles ne constituent pas la majorité des communes participant à la valeur du I global. À partir de cette date le rôle des espaces de faible densité s'amenuise au profit des agglomérations. Seul le massif alpin conservera une part non négligeable dans la valeur du I de MORAN local, et ce même en 1999, date à laquelle les grandes valeurs du I de MORAN local sont majoritairement dues aux grandes agglomérations.



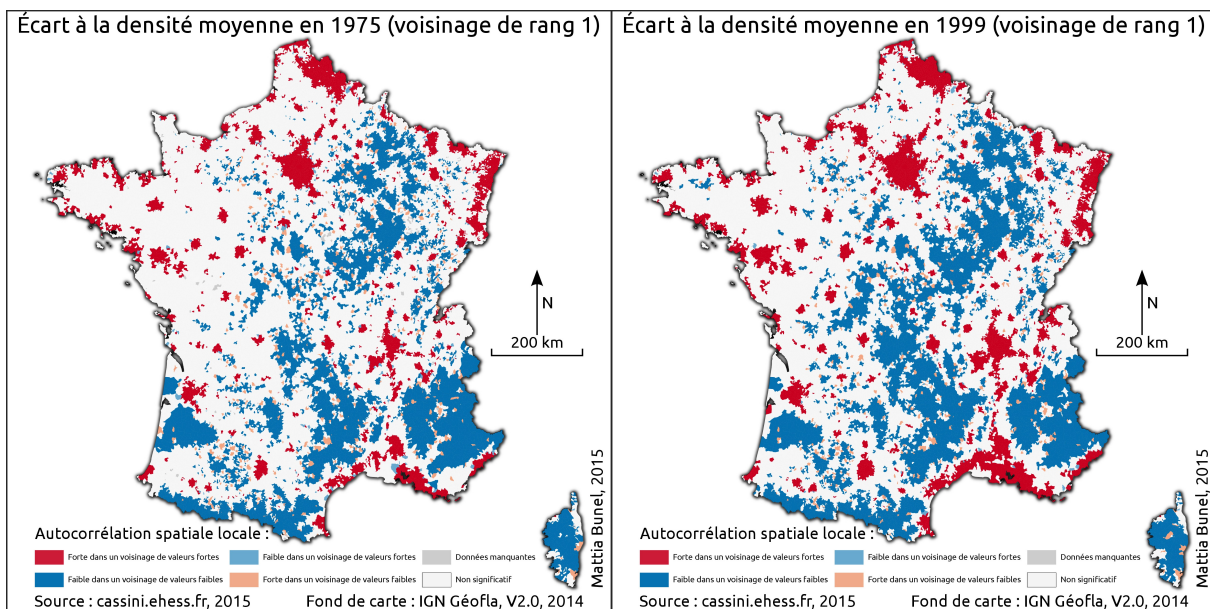
(a) Écart à la densité moyenne en 1793, $I = 0,45$.

(b) Écart à la densité moyenne en 1901, $T = 0,54$.



(c) Écart à la densité moyenne en 1921, $I = 0,56$.

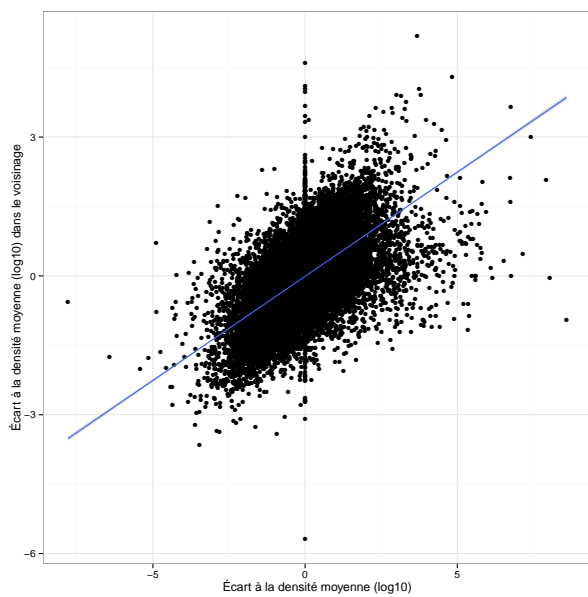
(d) Écart à la densité moyenne en 1946, $I = 0,59$.



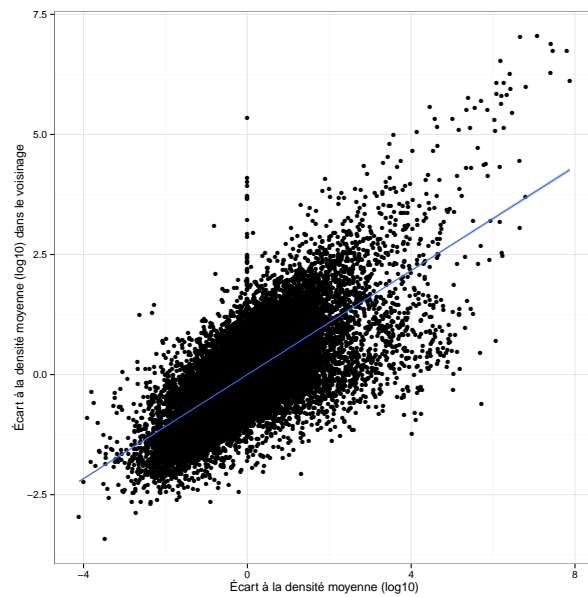
(e) Écart à la densité moyenne en 1975, $I = 0,63$.

(f) Écart à la densité moyenne en 1999, $I = 0,71$.

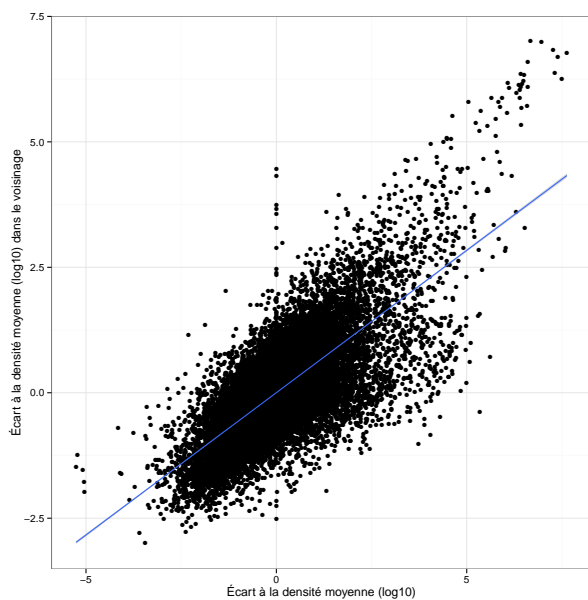
FIGURE 1.11 – Indicateurs locaux d'association spatiale pour la variable écart à la densité moyenne.



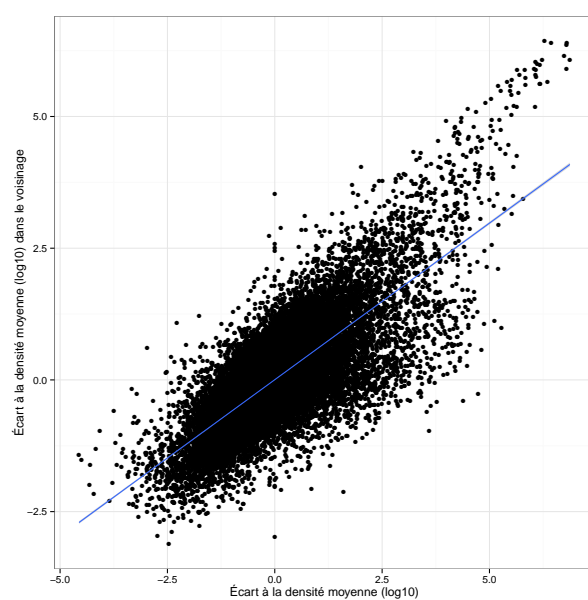
(a) Écart à la densité moyenne en 1793.



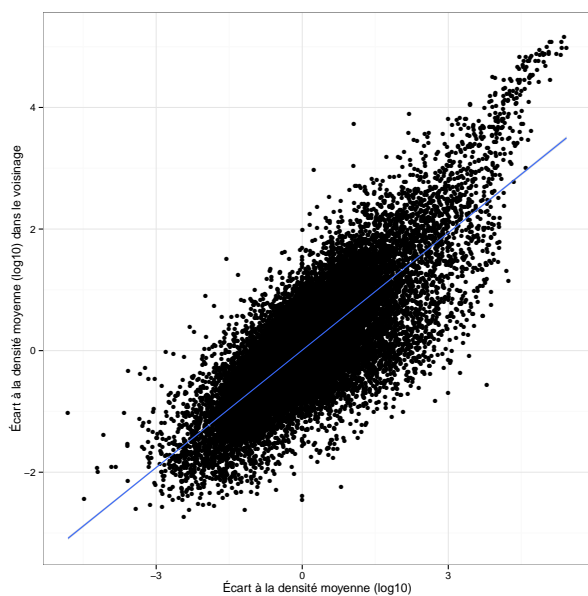
(b) Écart à la densité moyenne en 1901.



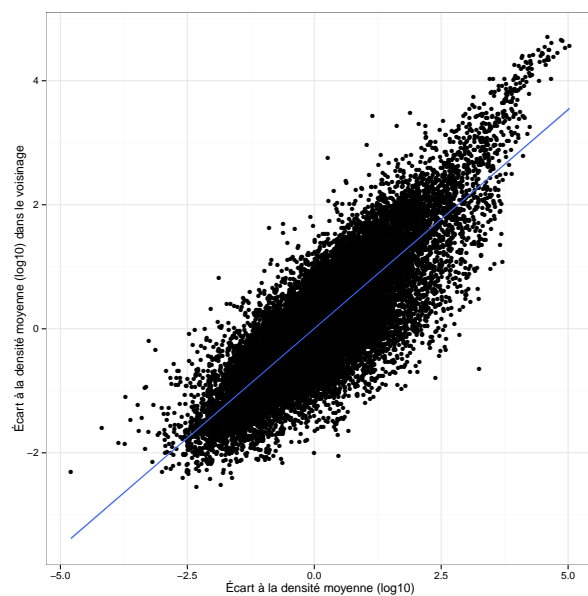
(c) Écart à la densité moyenne en 1921.



(d) Écart à la densité moyenne en 1946.

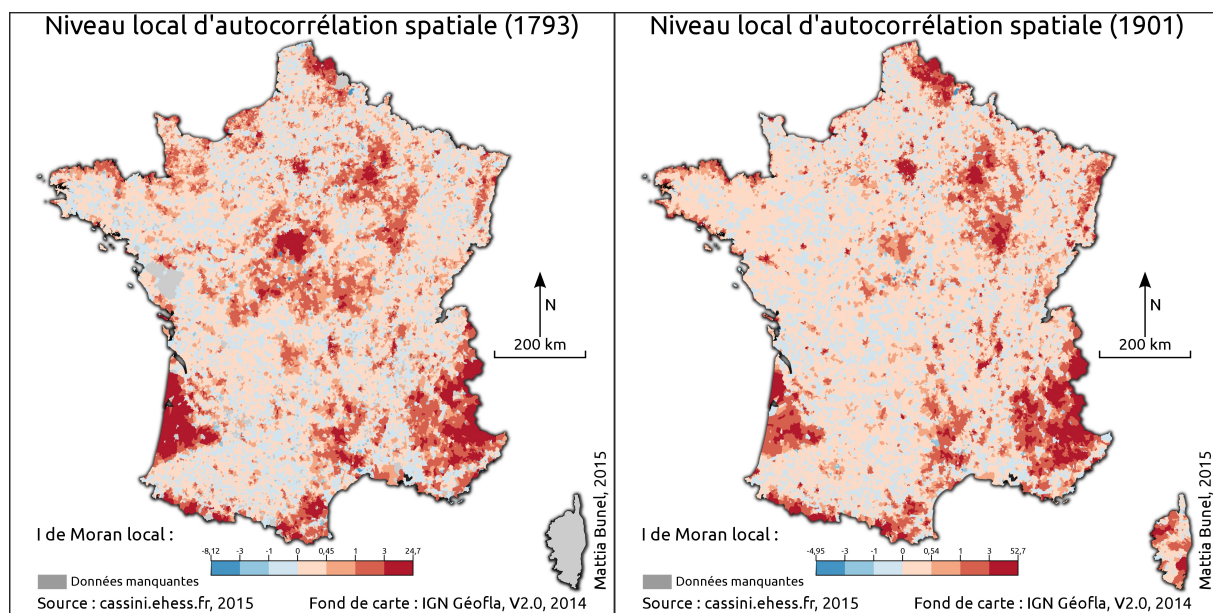


(e) Écart à la densité moyenne en 1975.



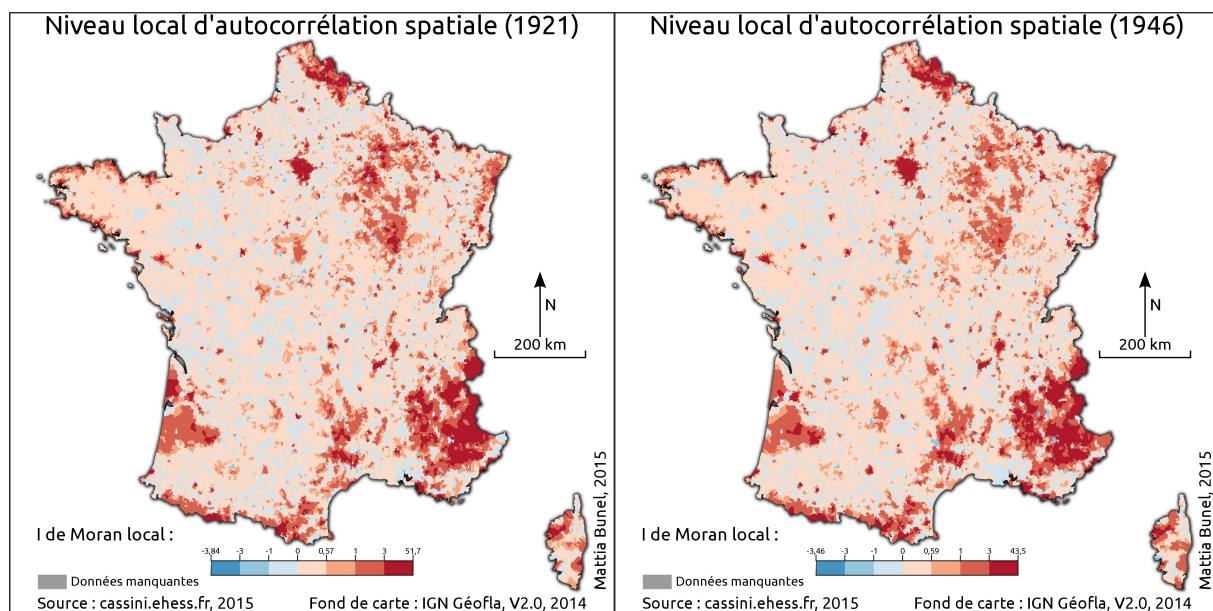
(f) Écart à la densité moyenne en 1999.

FIGURE 1.12 – Nuages de points de MORAN.



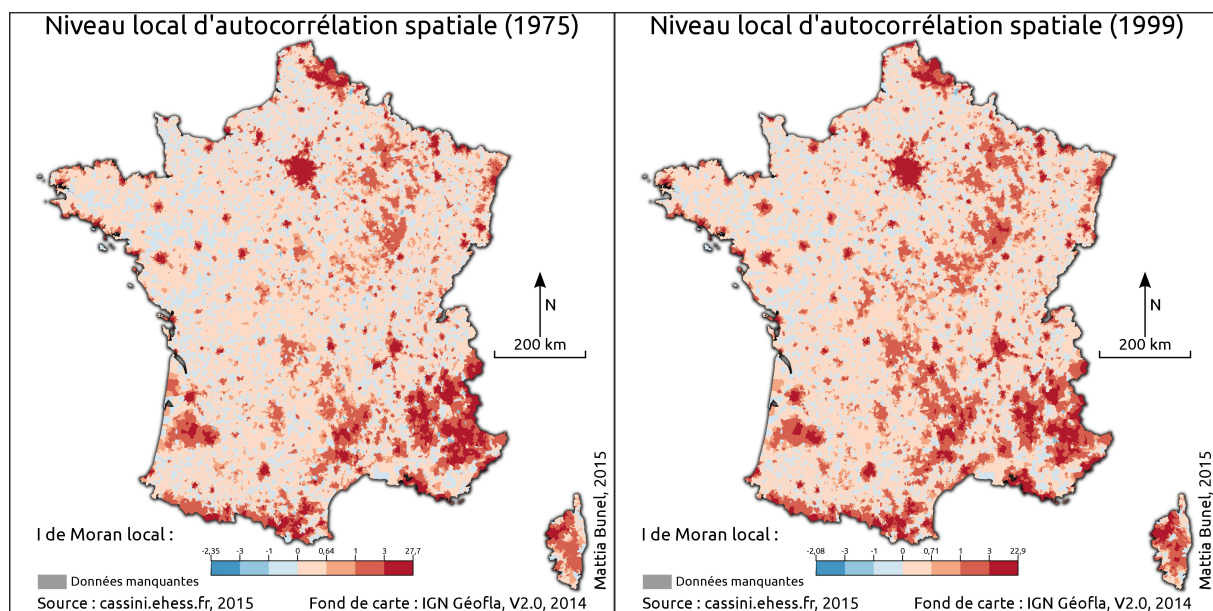
(a) Écart à la densité moyenne en 1793.

(b) Écart à la densité moyenne en 1901.



(c) Écart à la densité moyenne en 1921.

(d) Écart à la densité moyenne en 1946.



(e) Écart à la densité moyenne en 1975.

(f) Écart à la densité moyenne en 1999.

FIGURE 1.13 – Niveau local d'autocorrélation spatiale pour la variable *écart à la densité moyenne*.

L'utilisation des *indicateur locaux d'association spatiale* à permis de répondre aux questions posées par l'étude de l'évolution de divers indicateurs globaux (écart-type, coefficient de GINI) et par le recours à la cartographie thématique. Si nous savions au terme des précédentes recherche que nous étions en face d'une tendance à la concentration de la population, et plus généralement face à une profonde mutation de la distribution de la population française, nous n'avions que peu d'informations sur l'inscription spatiale de ce phénomène. Le recours à la cartographie thématique nous a permis d'apporter des réponses à ces questions, mais étant sujette à l'interprétation de l'auteur et ne permettant aucune quantification, cette méthode ne pouvait être une finalité. L'utilisation des *lisa* avait pour but de répondre à ces deux points, tout en permettant, une fois encore, de confirmer nos précédents résultats par le recoupement des méthodes. L'analyse de l'évolution du I de MORAN à confirmé les résultats mesurés (écart-type et coefficient de GINI) et observés (cartographie thématique) précédemment. L'augmentation du I de MORAN global indiquant un renforcement de la structure spatiale validant l'hypothèse de concentration. La cartographie des clusters à quant à elle permis de mettre en valeur les régions de fortes densités et inversement, et c'est en analysant la série temporelle émanant du cumul des cartes que nous pouvons reconstituer l'histoire de la répartition spatiale de la population. Les cartes des clusters sont éloquentes, on observe clairement un phénomène de concentration de la population dans les villes, ces dernières ne cessent de grandir et de prendre en importance. Mais cette évolution n'est pas constante, les changements les plus significatifs ayant lieux après 1945. De la même manière la répartition des zones les moins denses est modifiée durant cette période. La diagonale du vide est même de plus en plus présente, cette dernière atteint son apogée en 1999, date à laquelle elle présente la structure la plus compacte. Le recours aux indicateurs locaux d'association spatiale pose également d'autres questions, notamment induites par le traitement de la dimension historique. Lors de sa construction chaque carte est conçue indépendamment des autres, et c'est seulement lors du commentaire final que l'on traite l'ensemble comme une série historique. Il nous est donc impossible de savoir comment naissent les clusters. Au vu de la figure 1.3b nous savons que la densité de la population ne cesse d'augmenter, par conséquent les communes intégrant les clusters de fortes densités sont nécessairement des communes dont la densité a augmenté, et ce plus rapidement que la moyenne (pour que l'écart à la densité moyenne augmente), toutefois cette affirmation n'est pas valable pour les communes appartenant au cluster de faible densité, s'agit-il de communes dont la densité diminue ou de communes dont la densité reste stable, là où la densité moyenne augmente. Rapporté à la diagonale du vide, cette dernière est-elle plus visible car elle continue à se dépeupler, ou est-ce un effet de contraste? Nous allons essayer de répondre à ces questions dans la partie suivante au cours de laquelle nous réaliserons une typologie des communes permettant de synthétiser leur parcours dans la série temporelle.

1.3.3 Typologie des communes

Si l'objectif de cette étape a été précédemment décrit il reste à préciser par quelle méthode nous comptons réaliser cette typologie. Nous avons fait appel à une classification basée sur l'algorithme des *k-means*, les variables classées étant l'écart à la densité moyenne pour chaque date. Toutefois nous n'avons pas utilisé les trente-trois variables à notre disposition et ce pour une raison pratique. Comme précédemment expliqué nos données ne sont pas complètes, certaines communes, voir certains départements n'ont pas été dénombrés à une date t et ce pour de nombreuses raisons. Lors de nos précédentes réalisations ce détail ne posait guère de problèmes, chaque date étant traitée indépendamment. Mais dans le cas la méthode des *k-means* cela diffère. En effet l'algorithme ne peut fonctionner dans le cas où un individu n'a pas de données pour une des variables. Ainsi réaliser cette typologie sur l'ensemble des variables nécessiterait de faire fi de tous les individus pour les quelles nous n'avons pas les données des trente-trois recensements. Cette condition est bien trop pesante, et implique de ce passer de 20 % des individus, ce qui n'est pas envisageable. Nous avons donc cherché à réaliser cette typologie suivant le meilleur rapport couverture spatiale / couverture temporelle, nous avons donc supprimé les dénombrements présentant le plus grand nombre de *communes absentes*. Étant donnée le faible nombre de communes sans données des dénombrements de 1800 et 1806, notre typologie peut toutefois porter sur une grande amplitude temporelle³⁴. Une fois la grande majorité des recensements sélectionnés nous avons fait face à un second problème, cette manière de procéder impliquait de retirer tous les dénombrements entre 1821 et 1876. Il s'est avéré que cette amplitude était trop importante, la décision a donc été prise de « sacrifier » une partie des individus en sélectionnant un des dénombrements de cette période. Notre choix c'est porté sur le dénombrement de 1846, présentant le double avantage d'être au centre de cette période et de contenir un minimum d'individus absents. Ainsi le nord de la Lorraine est la seule région pour laquelle il manque une grande partie des communes, il s'agit clairement d'un choix pragmatique, il nous a semblé plus pertinent de privilégier un dénombrement à 600 communes.

34. Cent-quatre-vingt-dix-neufs ans, au lieu des deux-cent-six ans permis par la base de données.

Les résultats de cette typologie sont résumés par trois documents. Le premier est la carte des classes. Comme la légende le laisse entrevoir nous avons sélectionné une partition en cinq classes, le rendu cartographique permet de juger la qualité du partitionnement effectué. Cette classification est très intéressante pour notre travail car elle synthétise en une carte l'ensemble des observations réalisées. Les grandes agglomérations, et leurs banlieues appartenaient aussi lisiblement que les zones dont la densité est inférieure à la moyenne. Pour mieux comprendre ces classes il est intéressant d'étudier les statistiques de classes. Nous avons décidé de remplacer le classique tableau récapitulatif des classes par les graphiques 1.15a et 1.15b. Ces derniers représentent l'écart à la densité moyenne moyen (ou médian) pour chaque recensement et chaque classe. La classe centrale (classe 3) correspond à une classe moyenne, dont la densité moyenne reste toujours supérieure à zéro et ne varie quasiment pas. Les classes 1 et 2 sont quant à elles des classes composées de communes dont la densité est supérieure à la moyenne et augmente durant toute la période. La première classe regroupe les communes avec les plus fortes densités, et la croissance la plus importante. Notons que cette croissance se stoppe et diminue légèrement à partir de 1975. Si l'on regarde la carte 1.14 issue de cette même typologie on remarque que cette classe est uniquement composée d'agglomérations, les tailles de ces dernières sont variables, puisque cette classe intègre aussi bien Paris, Lyon et Marseille, qu'Angoulême, Dunkerque ou Besançon. Ces agglomérations ont donc vu leur densité croître, et ce de façon importante jusqu'en 1975, date à partir de laquelle on observe un *desserment* [3] et par conséquent une diminution de la densité. Cette observation correspond par conséquent au modèle de la *transition urbaine*, la phase de *métropolisation* étant désormais stoppée au profit d'une phase de *péri-urbanisation* [3] observable par la légère diminution de la densité sur les trois derniers recensements. Si ce n'est visible que sur la carte, on constate que la seconde classe est liée à la première. Il s'agit de communes pour lesquelles on identifie deux cas de figure différents. Le premier est celui des communes situées en périphérie des grandes agglomérations, il s'agit visiblement de la majorité d'entre-elles. Chaque commune appartenant à la première classe est en effet bordée d'une couronne de communes appartenant à cette seconde classe. Ces dernières sont d'une densité moindre et l'augmentation de cette dernière est plus lente. Le second cas de figure correspond à ces ensembles de communes contiguës, formant des filaments s'étendant le long des fleuves ou des littoraux. C'est notamment le cas des vallées du Rhône et de la Garonne, ou encore du littoral breton.

Inversement les classes 4 et 5 sont constituées des communes dont la densité est toujours inférieure à la moyenne et tend à diminuer. Mais au contraire de ce que l'on observe entre les classes 1 et 2, l'écart ne se creuse que peu entre les classes 4 et 5, ces dernières ont une évolution semblable, même si la diminution de la densité des communes de la classe 5 semble accélérer à partir des années 1950. Si l'on se rapporte à la carte on remarque qu'une grande partie des communes est incluse dans ces deux classes, notamment les communes qui appartenaient au cluster comprenant les communes avec une densité inférieure à la moyenne situées dans un voisinage semblable. Le graphique 1.15a répond donc aux questions que nous nous étions préalablement posées, les communes appartenant à ce cluster, et plus généralement les communes dont la densité était (dès le commencement de la série temporelle) inférieure à la moyenne voient leur densité décroître, et par conséquent perdent de la population. Ainsi les clusters identifiés par les *lisa* n'apparaissent pas par effet de contraste, il s'agit de zones pour lesquelles la densité diminue au cours de la période étudiée. Cette carte permet également de mettre en évidence les structures internes.

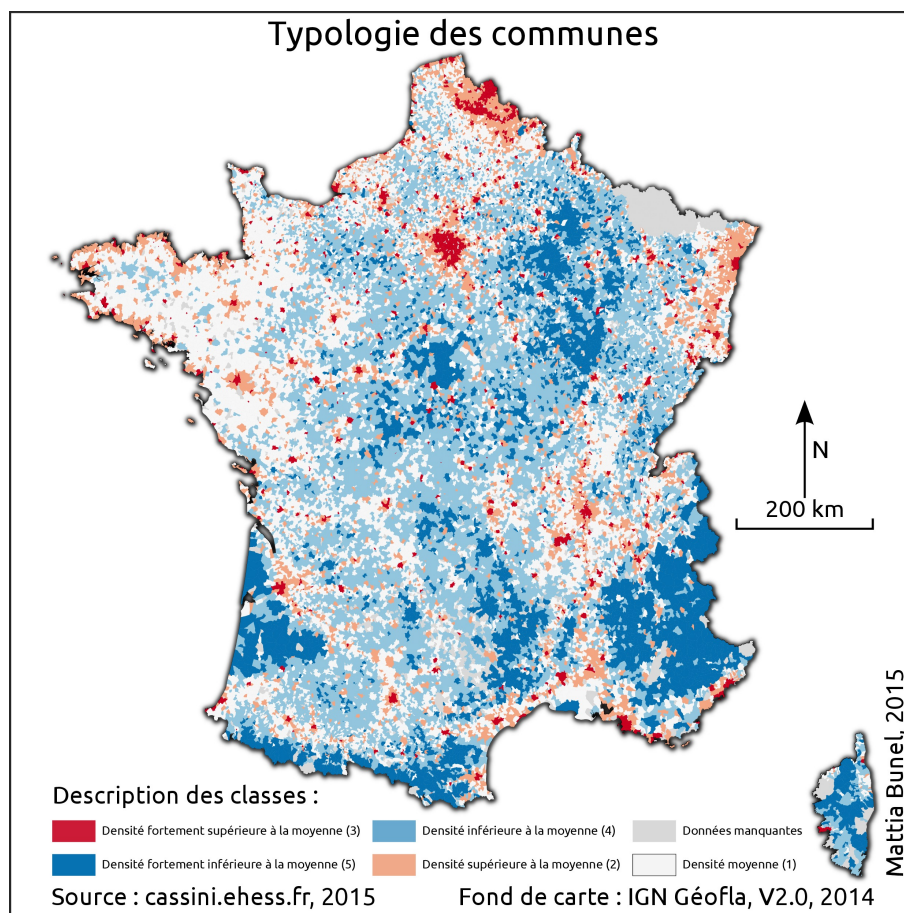


FIGURE 1.14 – Typologie des communes en cinq classes réalisée à partir de la méthode des *k-means* sur 25 recensements.

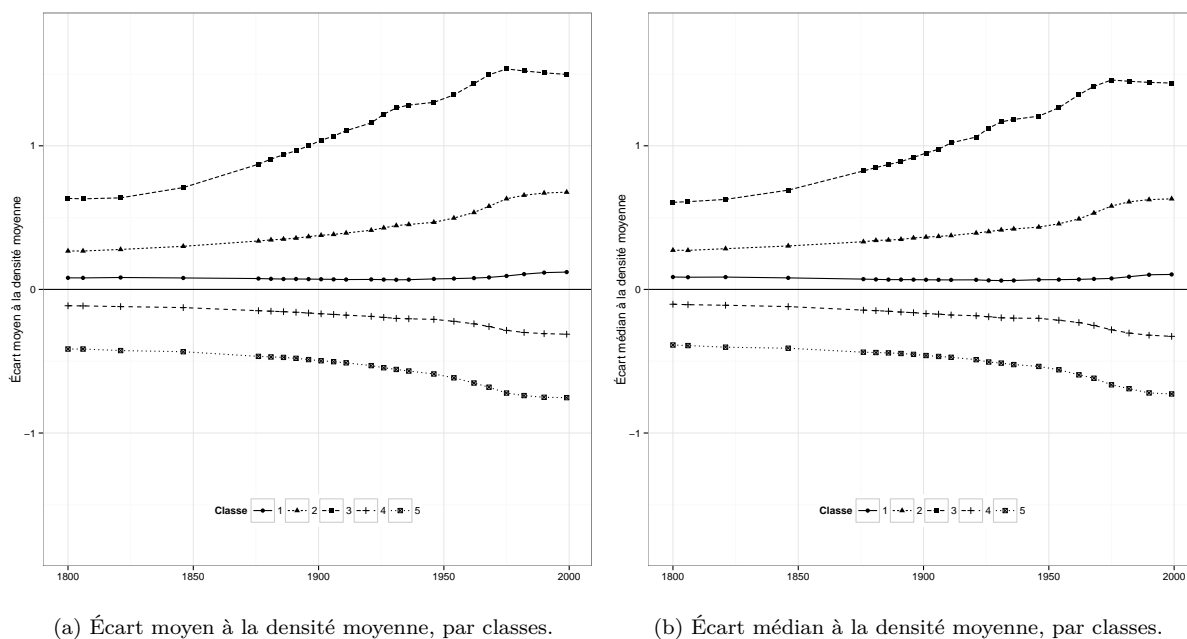


FIGURE 1.15 – Moyenne et médiane de la variable *écart à la densité moyenne* par classe et par date. Complément à la carte 1.14.

Conclusion

L'objectif initial de ce premier chapitre était à la fois de nous questionner sur l'existence d'une tendance à la concentration de la population française durant les XIX^e et XX^e siècles, mais également de revenir sur les données utilisées pour construire notre base de données.

Le premier tiers du chapitre est consacré à la présentation des données et plus généralement de l'histoire de la statistique française abordée dans ces grandes lignes. Ce volume peut paraître important, mais la décision de traiter l'histoire du recensement de cette manière émane d'une longue réflexion. Par le biais de cette approche nous avons souhaité interpeller le lecteur sur deux points principaux. Le premier est évidemment l'amplitude, sans commune mesure, de la base de données traitée. Si l'on se rapporte au nombre de modalités après agrégation (1 205 721) la base de données traitée n'est pas colossale, c'est au contraire par son amplitude temporelle (206 ans) que cette dernière s'illustre. Or en deux siècles la pratique du recensement a profondément été modifiée, chaque nouvelle opération apporte son lot de nouveautés, bulletin individuel standardisé en 1876, mécanisation en 1901, *etc.* S'affranchir de cette première partie historique équivaut à considérer l'ensemble des recensements comme des opérations semblables alors que de nombreuses différences subsistent. Le deuxième point porte sur la qualité des données. Nous avons montré que cette dernière est variable et peut-être scindée en deux composantes. La première, quantifiable, est celle des communes sans données à une date t . Les raisons de ces absences sont nombreuses, et parfois assez peu documentées. Il peut s'agir de communes non recensées pour des raisons politiques (le nord de la Vendée, sujet à des insurrections royalistes en 1793), de destructions ou de pertes d'archives, ou d'oublis. La majorité de ces absences sont concentrées dans la première moitié du XIX^e siècle (92 % des absences sont situées entre 1826 et 1872), et disparaissent (à quelques rares exceptions) à partir de 1876. La seconde composante est celle de la qualité intrinsèque du recensement. Les opérations ont été réalisées avec une application plus ou moins grande, certains recensements ne sont que des mises à jours de précédentes opérations (1811, 1826) [48, 31], d'autres sont sujets à des sur-enregistrements frauduleux (1921–1936) [29], *etc.* Les sources d'erreurs sont donc multiples et généralement non quantifiables ou non quantifiées³⁵. Nous sommes donc réduit à dresser une synthèse de la qualité des recensements à partir de simples observations, rarement argumentées, présentées par les auteurs. La seule conclusion que l'on peut véritablement tirer de ces observations est que la qualité des recensements s'accroît sensiblement avec le temps. Mais une nouvelle question se pose, *quid* de la qualité de notre variable, la population ? De toute évidence il s'agit de la seule variable que l'on peut étudier sur une aussi longue période. De nombreuses informations n'apparaissent que plus tardivement (distinction urbain-rural) ou sont sujettes à de très nombreux changements de définitions, rendant l'étude comparée extrêmement délicate (activités professionnelles). Si l'on suit un raisonnement logique, il est vraisemblable de supposer que la population est, dans tous les cas de figure, la variable la plus fiable. En effet il s'agit de la variable faisant intervenir le moins de composantes, ce qui tend à minimiser le risque d'erreurs. De plus toutes les autres données issues d'un recensement dépendent de la population, il est donc absurde de supposer qu'une variable autre que la population peut-être d'une meilleure qualité. Quoi qu'il en soit la qualité des données est une question trouble. Malgré les nombreuses références bibliographiques consultées, aucune ne nous apporte une réponse claire sur cette question. Nous avons donc réalisé notre propre synthèse sous la forme du tableau 1.2 (P. 20), en agrégeant les divers informations en provenance de la bibliographie. Il est vrai que non informations sur la question de la qualité restent limitées, seule l'étude de l'évolution du peuplement à l'échelle métropolitaine (cf. figure 1.3a page 1.3a) confrontant nos données à celles d'autres sources a permis une véritable mesure de la qualité de l'estimation de la population. Si nos données donnent des valeurs plus faibles que les estimations de DUPÂQUIER et al. [30, 32, 31], on constate que l'écart entre nos données est celles corrigées est quasi nul à partir de 1876 (date à partir de laquelle le nombre des communes lacunaires est quasi nul). Ce qui montre l'évolution temporelle de la qualité de nos données. C'est pour toutes ces raisons qu'il nous a semblé nécessaire de revenir aussi longuement sur l'histoire de nos données.

Dans un second temps nous nous sommes lancés dans l'étude de l'affirmation de LE BRAS et LABBE [50] présentée en introduction. Il nous est apparu logique de l'aborder en la scindant en deux. Dans un premier temps (premier chapitre) nous avons cherché à savoir si la repartition de la population était profondément modifiée durant cette période de deux-cent-six ans, et plus particulièrement si la population tendait à ce concentrer. L'absence d'observation d'une tendance à la concentration de la population aurait clairement fragilisé, voir invalidé l'affirmation de LE BRAS et LABBE [50]. Nos différentes approches ont clairement mis en évidence la tendance à la concentration de la population à l'échelle de la métropole. Une première analyse a été réalisée à partir de l'étude des indicateurs de dispersion, et plus particulièrement de *l'écart-type*, étude qui a été validée et approfondie par l'analyse du coefficient de GINI. Nous avons identifié plusieurs phases dans l'évolution de la concentration de la population (*stabilité, métropolisation,*

35. À l'exception des dénombrements de 1851 et 1856 pour lesquels DUPÂQUIER et al. [29] nous fournissent une quantification de la population non recensée. Toutefois rien ne nous indique qu'il s'agit du seul oubli lors de cette opération.

péri-urbanisation) et les avons décrites. Mais la validation de l'hypothèse de départ nous a amené à nous poser d'autres questions. Notamment sur la nature des régions concernées par cette mutation. Quelles sont les zones où la population se concentre et au contraire quelles sont les zones les moins denses, s'agit-il de régions qui stagnent dans un contexte où la population (et donc la densité moyenne) croît, ou s'agit-il de zones qui perdent en densité. Pour répondre à ces multiples questions il a été nécessaire d'aborder le sujet de manière plus précise, en intégrant la question spatiale dans nos recherches. La première étape a été de réaliser une cartographie thématique de *l'écart à la densité moyenne*. Cette série de cartes nous a permis d'observer les importantes transformations de la distribution de la population au cours de ces deux cent-six ans. Au vu de ces cartes le territoire paraît de plus en plus structuré, et les écarts entre villes et campagnes de plus en plus marqués. On y voit clairement que les villes prennent en importance, aussi bien spatiale que démographique (cf. figure 1.8) et ce au détriment des zones rurales, contrastant de plus en plus vigoureusement avec l'urbain. Le recours aux *indicateur locaux d'association spatiale* a permis d'observer plus en profondeur ce phénomène. L'augmentation du I de MORAN confirme les observations précédentes (population de plus en plus concentrée) et indique un renforcement de la structure spatiale. Les cartes des clusters valident également les précédentes observations, en nous montrant les oppositions entre fortes et faibles densités, entre urbain et rural. La structure observable sur les cartes est de plus en plus marquée, conformément au résultat attendu au vu de l'évolution du I de MORAN et fait apparaître certaines structures, dont la *diagonale du vide*, plus visible que jamais en 1999. La dernière étape a été de synthétiser tous ces résultats en une seule production cartographique. Les cartes des *k-means* montrent une typologie en cinq classes des communes. Les communes sont regroupées en fonction de l'évolution de leur *écart à la densité moyenne*. Les communes denses et fortement denses s'opposent aux communes peu denses et très peu denses. L'amplitude entre la densité moyenne de ces groupes augmente et illustre la concentration de la population. La représentation cartographique de ces classes nous permet d'identifier des structures déjà visualisées, la *diagonale du vide*, le massif alpin, les agglomérations et leur banlieues etc.

Mis en perspective avec la bibliographie nous avons pu produire un commentaire historique des événements observés. Le phénomène observé n'est autre que la *transition urbaine*, c'est à dire le passage d'une société à dominante rurale à une société à dominante urbaine [3]. Cette transition s'est opérée par le biais de migrations internes, principalement *l'exode rural*, dans un premier temps limité aux excédents démographiques des campagnes [31], puis massif et conduisant au dépeuplement des campagnes et à la baisse de leur densité. Mais l'urbanisation est également expliquée par « l'absorption » quasi-intégrale de la croissance démographique par les villes [31, P. 131]. L'analyse de nos productions est relativement simple, toutes nous montrent une façade d'un même phénomène, l'importance de la croissance (spatiale et démographique) des grandes agglomérations et plus généralement la *transition urbaine*, suivie d'une dynamique de *péri-urbanisation*. Mais les causes historiques de ces phénomènes sont nombreuses et s'inscrivent dans un système complexe où *transition démographique*, *industrialisation* et *transition urbaine* interagissent et conduisent aux résultats observés. Ce chapitre nous a donc permis de valider notre hypothèse de départ, et plus particulièrement de la mettre en perspective de grandes mutations historiques, et plus particulièrement de la profonde modification de la distribution de la population française entre le XIX^e et le XX^e siècle.

Nous pouvons désormais entamer la seconde étape de ce travail de recherche. Notre objectif pour le chapitre à venir est de nous questionner sur les liens entre la répartition de la population et l'environnement, ainsi que sur leur évolution. Au vu de l'affirmation de LE BRAS et LABBE [50], il existera un lien entre les spécificités environnementales d'une région et l'évolution de son peuplement. Les zones les plus « rudes » étant, selon cette affirmation, plus touchées par le dépeuplement. Le second chapitre sera l'occasion d'introduire une nouvelle composante dans notre travail, l'environnement.

Chapitre 2

L'influence de l'environnement sur la distribution de la population et ces variations

Introduction

Dans ce second chapitre notre objectif est d'introduire une nouvelle composante dans notre travail, l'environnement. L'hypothèse de LE BRAS et LABBE [50] laisse entendre qu'il existe un lien entre l'évolution de la distribution de la population et différentes caractéristiques physiques des communes. Dans ce chapitre nous testerons et explorerons cette hypothèse. Dans un premier temps nous reviendrons sur la théorie de la distribution de la population et nous présenterons le rôle et l'importance des différentes composantes. Puis nous étudierons le phénomène à travers une approche cartographique, suivie d'une approche statistique.

2.1 Théorie de la distribution de la population

La distribution de la population est la résultante d'un système complexe impliquant de nombreuses composantes. Cette dernière s'explique par une combinaison complexe de plusieurs facteurs [6]. Ces derniers sont généralement présentés en deux catégories, les *facteurs environnementaux* et les *facteurs historiques* [6, 65]. Mais DURAND-DASTÈS [34] propose également une typologie en trois composantes, environnement, temps et espace. Cette seconde approche distingue l'emplacement d'une commune et les différentes interactions spatiales de ces caractéristiques environnementales. nous avons choisi de présenter dans cette partie la théorie de la distribution de la population en nous basant sur cette première approche.

2.1.1 Distribution de la population et environnement

Quel que soit l'ouvrage traitant de la géographie de la population choisi tous insistent sur le rôle que joue l'environnement sur la distribution de la population. L'étude de la composante environnementale, et de son influence sur la distribution de la population est une préoccupation ancienne en géographie et est source de nombreux débats. Au début du XX^e siècle, le géographe allemand RATZEL considérait, selon une *approche déterministe* que la répartition de la population était essentiellement due à l'environnement [6], par extrapolation des environnements comparables présenteraient une population comparable. Pour l'école vidalienne cette conception est erronée, elle considère au contraire que le lien entre population et environnement est indirect, c'est ainsi que LUCIEN FEBVRE indique que « *La nature propose, l'homme dispose* » [6]. La position de l'école vidalienne est qualifiée de *possibiliste*. Aujourd'hui la conception *déterministe* du peuplement est tombée en désuétude, et l'on considère que la distribution de la population observée à un temps t dépend d'une interaction complexe de plusieurs composantes (dont l'environnement n'est qu'une partie) variant dans le temps et l'espace [3, 6, 65]. Parmi ces composantes l'environnement de joue qu'un rôle mineur, l'élément principal étant l'histoire et les composantes qui en découlent (ancienneté du peuplement, vitesse d'accroissement de la population, *etc.*) [6, 60]. Le terme d'environnement, utilisé jusqu'ici est volontairement assez vaste, par cette expression nous souhaitons parler de l'ensemble des composantes rattachées à l'environnement pouvant avoir directement ou indirectement un effet sur le peuplement humain. L'ensemble de ces composantes est vastes, mais il est possible de les regrouper en

trois (quatre selon BAUELLE [6]) classes que sont les contraintes climatiques, l'altitude et les reliefs, l'hydrographie (et la pédologie).

À l'échelle de la planète la principale contrainte est de nature climatique. Il existe d'importants contrastes de population entre les différents milieux climatiques. La majorité de la population est concentrée dans les milieux tempérés et subtropicaux, alors que les milieux arides et polaires ne totalisent que 2 % de la population mondiale [6, 76]. On observe que les milieux les moins denses correspondent généralement aux climats les plus contraignants [6, 65]. Mais cette répartition en fonction du climat est plus complexe qu'un gradient équateur-pôles. On constate qu'il existe également une différence de densité entre les côtes occidentales et orientales, variant avec la latitude. Au niveau des latitudes tempérées, la façade occidentale est plus dense que la façade orientale. En effet les hivers sur les façades orientales sont beaucoup plus rigoureux que sur les façades occidentales. Il en résulte une différence de densité très importante, puisque « le Royaume-Uni compte 245 hab./km², là où le Labrador, situé à la même latitude, n'atteint même pas 1 hab./km² » [6]. On retrouve ces mêmes contrastes aux latitudes intertropicales, mais cette fois-ci c'est la côte orientale qui se trouve être la plus dense.

Toutefois, on peut se demander de quelle manière certains climats sont plus contraignants que d'autres. Même au plus hautes latitudes il existe des civilisations et des cultures qui ont su survivre et se développer [65]. L'homme est physiquement capable de survivre aux hautes latitudes, mais les différentes civilisations qui se sont développées dans ces régions n'ont pu baser leur alimentation sur l'agriculture [6]. En effet les grands froids caractéristiques des climats polaires, ont pour effet de rendre l'agriculture très difficile, voire impossible [6, 65] et c'est là l'une des premières limitations liées au climat « [...] la principale chute de densité intervient là où l'agriculture devient impossible. » [6]. Aussi, l'influence du froid est indirecte, c'est avant tout l'impossibilité de cultiver qui explique les faibles densités des hautes latitudes et non le froid en lui-même [6]. Il existe cependant certaines exceptions à ce modèle, notamment certaines concentrations de populations isolées dans les hautes latitudes, situées là où l'intérêt économique ou stratégique du site nécessite ou incite l'implantation, malgré l'inhospitalité du milieu (par exemple des ressources minières), on pourra citer les villes de Moumansk et Norilsk en Sibérie (respectivement les villes de plus de 250 000 et de 100 000 les plus septentrionales au monde) [6, 7], ou encore la mine de diamants de Diavik située à l'extrême nord du Canada. Bien qu'il soit la principale contrainte liée aux hautes latitudes, le froid n'est pas la seule problématique à laquelle sont sujettes les populations vivant aux plus hautes latitudes. La durée du jour, ainsi que sa forte variation sont deux contraintes agissant de façon néfaste sur l'être humain [6]. Selon BAUELLE [6] « la longueur de la nuit polaire explique certaines carences en vitamines à l'origine d'une baisse de la fécondité et par conséquent d'une plus faible croissance démographique. »

Si les hautes latitudes sont synonymes de faibles densités, les zones les plus chaudes peuvent également être faiblement peuplées. Comme pour le froid, ce n'est pas la résistance humaine qui pose véritablement problème mais les conséquences des fortes chaleurs. Le manque d'eau est une véritable problématique qui, comme les grands froids rend l'agriculture sporadique, voire impossible [6, 65] et par conséquent les milieux désertiques sont les moins denses après les milieux polaires. Cependant la distribution de la population dans les zones très chaudes et arides est sensiblement différente de celle des milieux polaires [6]. Aux hautes latitudes les côtes sont les endroits les plus peuplés, contrastant fortement avec l'arrière pays. Un exemple notable est celui de l'Islande, où 50 % de la population est concentrée dans la capitale portuaire, Reykjavik. Toutefois, la répartition de la population dans les zones arides est plus sporadique, plus discontinue. La population est regroupée autour des oasis, mais également près des sources généralement situées sur les piémonts [6]. Certains pays essaient de lutter contre le manque d'eau en développant divers programmes, basés à la fois sur des barrages (Égypte), sur l'exploitation de nappes d'eau fossiles (Libye), ou sur la désalinisation de l'eau de mer [6]. Toutefois ces programmes sont très coûteux, et le prix de production d'eau douce est répercuté au produit fini, c'est à dire les produits agricoles cultivés sur des terres gagnées sur les déserts [6, 65].

Nous avons cités les deux climats limitant les plus importants, mais nous pourrions en citer un troisième, le climat tropical humide à l'origine des « déserts verts » tels que les forêts équatoriales [6, 65]. Ici la principale problématique est la salubrité de l'environnement. Ce type de climat favorise la multiplication de maladies parasitaires dont les moustiques, et les mouches sont vecteurs [6, 65] ce qui est la principale raison de la faible densité de population de ces régions. Mais il n'est pas nécessaire de revenir plus longuement sur les déserts humains arides tropicaux et polaires. Comme nous l'avons expliqué la faible densité de population de ces régions est expliquée par leur climat, extrêmement contraignant. Or de tels climats n'existent pas en France métropolitaine, pas plus que de grands déserts humains tels que nous pourrions en trouver en Guyane ou en Algérie. Si notre étude portait sur de tels espaces le recours à des variables climatiques aurait été justifié, mais ce choix ne semble pas adapté à notre étude. Comme nos précédents exemples le laissent apparaître la composante climatique agit à petite échelle, à trop petite échelle pour être prise en considération dans notre travail. Il existe certes une variation notable du climat à l'échelle de la métropole française, mais celle-ci est trop faible pour jouer sur la distribution de la population. À titre d'exemple nous avons, lors de nos recherches, testé la corrélation entre différentes composantes

climatiques et la densité de la population en 1999. Quel que soit la variable explicative retenue (température, pluviométrie, ensoleillement...) les coefficients de détermination obtenus sont systématiquement inférieurs à 1 %. Par conséquent nous ne traiterons pas de la composante climatique durant ce travail.

La seconde composante environnementale dont nous parlerons est l'altitude. Notez que lorsque nous parlons d'altitude nous souhaitons parler plus généralement des massifs montagneux, caractérisés par leur altitude, mais également par leur relief. De manière analogue aux travaux de SMALL et COHEN [76] nous traiterons que de l'altitude lors de la réalisation de nos différents modèles, les deux composantes étant intrinsèquement liées. L'altitude limite le peuplement par le manque d'oxygène, là où les reliefs limitent le peuplement pour des raisons d'accessibilité [6, 65]. On constate que la densité moyenne baisse avec l'altitude, mais cette décroissance n'est pas linéaire, elle est plutôt marquée par un effet de seuil. La majorité de la population mondiale est située dans les vallées et dans les zones faiblement marquées par les reliefs. Aux environs de 6 000 m d'altitude, l'homme est sujet au « *mal des montagnes* » induit par la baisse du taux d'oxygène : l'individu est alors dans un état d'hypoxie [65]. Toutefois les records d'altitude d'établissements humains ne dépassent pas les 5 000 m d'altitude. L'altitude peut également avoir un effet négatif sur l'agriculture, en rendant impossible la culture de certains végétaux. Mais l'altitude peut également avoir un effet inverse sur les densités. Il existe certaines zones (Éthiopie, Bolivie) où les densités augmentent avec l'altitude. La hausse de l'altitude protège les populations de certaines maladies tropicales dont nous parlions précédemment, mais également des plus fortes températures, l'altitude a un effet qui dépend principalement du milieu [6]. DOLFUS [27] a créé le terme de *montagnes relatives* pour qualifier ce fait. La principale conséquence de l'existence des reliefs est l'enclavement. Les reliefs rendent la circulation plus difficile, plus lente, voire dangereuse [6]. Les conséquences de cet enclavement peuvent être paradoxales. Il peut conduire à de faibles densités, mais également protéger une population d'invasions. De nombreuses populations (les Cathares par exemple) se sont réfugiés dans les montagnes pour fuir les invasions et les persécutions. Une terre plus riche, et mieux desservie et mécaniquement plus accessible et plus convoitée, on parle du *paradoxe de Bagdad* [6]. Les reliefs peuvent donc expliquer certaines fortes densités, notamment celle des *montagnes refuges*, dont de nombreux exemples existent de part et d'autre de la planète (Khabilie, Albanie, pays Dogon) [6, 27, 65]. Le processus qui conduit à la création de montagnes refuges et avant tout historique, l'homme choisi de sacrifier la productivité et l'accessibilité au profit de la sécurité [6, 65]. De plus les terres sont plus difficiles à cultiver, les sols sont plus pauvres donc moins productifs et l'érosion les appauvrit encore plus, les fortes pentes rendent la culture plus délicate, et l'irrigation quasi impossible sans structures lourdes [6]. Ces différents facteurs rendent les hauts reliefs difficiles à vivre. Mais les hauts reliefs sont également à l'origine des fortes densités des piémonts. Parfaitement situés au pied de ces obstacles, ils en deviennent des points de passage stratégiques [6, 27, 65]. Contrairement au climat, la composante altimétrique est parfaitement adaptée à notre échelle d'étude. Plusieurs massifs montagneux sont présents en France métropolitaine, et les cartes réalisées dans le précédent chapitre indiquent clairement que les communes situées aux plus hautes altitudes sont les moins peuplées. Nous retenons donc cette variable pour notre travail et présenterons ultérieurement les modalités de sa construction.

L'hydrographie est également un facteur qui joue un rôle important dans la répartition de la population. Une grande partie de la population mondiale est généralement concentrée dans les vallées fluviales et le long des côtes [76] :

« *Il est facile et assez spectaculaire de constater le haut degré de corrélation existant entre les vallées fluviales et la répartition des populations.* » [67].

L'hydrographie peut être décomposée en deux points. Les fleuves sont la première composante de l'hydrographie. Ils présentent de nombreux avantages expliquant la concentration de population le long de leurs rives. La quantité d'eau apportée par les fleuves permet l'irrigation. Or irriguer permet d'augmenter la production agricole et, de plus, les sols alluviaux sont généralement très fertiles [6]. Comme nous avons pu le voir précédemment, il existe un lien entre densités et productivité agricole, l'irrigation, en augmentant la productivité agricole rend les fleuves attractifs pour les sociétés humaines. En outre, si leur débit est suffisant, ils peuvent également faire office de voie de communication, élément déterminant dans l'implantation humaine. LE BRAS [49] parle de « *filaments urbains* », pour décrire les couloirs de forte densité s'étendant le long des vallées fluviales et reliant le centre de la France (Paris), aux littoraux. La France est caractéristique de cette distribution, toutefois cette réflexion a été étendue à l'Europe dans son ensemble. Comme dans le cas des montagnes le rapport entre fleuves et population varie. Dans les milieux tropicaux les fleuves (si leur débit n'est pas trop important) peuvent être vecteurs de différentes maladies comme la *bilharziose* [6]. Mais certains fleuves sont également sujets à de violentes crues, rendant toute implantation délicate. La seconde composante de l'hydrographie est l'ensemble des côtes. Le littoral est également attractif pour les populations, à un point tel que la DATAR considère que dans le cas de la France « *la géographie des bassins et des littoraux domine toutes les autres formes de peuplement* » [22]. Les littoraux présentent des avantages comparables à ceux des fleuves, bien qu'ils ne permettent pas l'irrigation, la

mer constitue une ressource non négligeable, mais également une voie de communication très importante. Toutefois, des exceptions existent, c'est notamment le cas des déserts côtiers (Libye, Égypte). À l'échelle de la France métropolitaine on retrouve également des zones littorales peu denses (Aquitaine, Camargue), mais elles n'entrent pas dans la catégorie des desserts côtiers. Comme pour l'altitude, le recours à cette nouvelle composante est parfaitement adapté à notre échelle et nous travaillerons également à partir de l'hydrographie lors de la réalisation des modèles ultérieurs. Toutefois nous avons choisi de ne travailler qu'à partir des littoraux et non des fleuves. Les raisons de ce choix sont multiples mais la principale étant que nos différents essais utilisant la distance au fleuves comme variable explicative n'ont donné que des résultats assez faibles. Nous avons donc jugé à la vue du travail nécessaire pour exploiter cette troisième variable qu'il serait préférable de se concentrer sur les deux premières, plutôt que de traiter superficiellement trois composantes.

Si les trois facteurs précédemment cités font l'objet d'un consensus, BAUELLE [6] relève un dernier facteur. Selon lui la qualité des sols est un facteur important pour expliquer les variations locales du peuplement. De la qualité des sols découle leur capacité productive. Or comme nous l'avons précédemment vu, il existe un rapport entre productivité agricole et peuplement. Toutefois il faut relativiser ce constat, la qualité des sols influe aussi bien sur la capacité productive que sur la demande de main-d'œuvre. Aujourd'hui les zones les plus productives (Beauce, Middle Ouest) sont des zones peu peuplées. L'hypermécanisation ayant permis d'augmenter la production tout en réduisant la main d'œuvre [6]. Si cette variable aurait pu être intéressante à traiter, il s'est avéré impossible de réaliser ce traitement du fait de la disponibilité des données. Malgré nos recherches, nous n'avons trouvé aucune base de données nous permettant d'intégrer cette dimension à notre étude.

Le rapport entre peuplement et contraintes physiques fait l'objet d'un consensus. Tous les auteurs relèvent la disparité des densités à l'échelle mondiale. Selon PINCHEMEL et PINCHEMEL [65], « *Trois relations entre la trame du peuplement et les données physiques s'imposent : le peuplement de la terre est lié aux littoraux, aux basses terres, aux vallées.* », cette observation est confirmée par les travaux de SMALL et COHEN [76]. Chacun des paramètres précédemment cités n'agit pas de façon unique : tous se combinent pour donner naissance aux différents milieux existants [6]. Cet enchaînement de facteurs fonctionne comme un système en constante interaction avec les sociétés humaines, et par conséquent la répartition des populations. Mais même en faisant appel à ces différents facteurs et en les combinant, il existe des variations au sein de milieux identiques inexplicables par de simples facteurs physiques. L'action de l'environnement sur le peuplement est avant tout indirecte [6]. La perception des contraintes par l'homme joue un rôle déterminant. Le concept précédemment cité de « *montagnes relatives* » renvoie à cette idée. Le point fondamental est plus les rapport entre homme et environnement que les contraintes environnementales elles-mêmes. Comme le signale PUMAIN [68, P. 430], « *les géographes ont très vite récusé une intervention déterminante du milieu naturel dans l'explication des inégalités du peuplement.* » C'est avant tout l'héritage des générations précédentes sous la forme de différentes infrastructures, qui détermine le peuplement. C'est cet héritage que François Durand-Dastès, qualifie d'*espace reçu*. Pour conclure on aurait tort de mésestimer l'importance des héritages, et plus généralement de l'histoire, PUMAIN [68] insiste sur l'importance des « *processus cumulatifs dans l'émergence des grands foyers de peuplement [...]* ». Selon Pierre Gourou, « *en dehors des limitations désertiques, la carte générale de la répartition de la population dans le monde s'explique avant tout par la civilisation et l'histoire.* » [6, 77].

2.1.2 L'importance déterminante de la composante historique

Bien que déterminante la composante environnementale n'est pas la principale composante influant la distribution des populations. La composante historique joue un rôle capital et les cartes précédemment réalisées en témoignent.

Selon THUMERELLE [77], « *La répartition de la population, est d'abord le résultat d'un processus historique.* » Le simple recours aux contraintes environnementales précédemment citées ne suffit pas à expliquer entièrement la distribution de la population, et les cartes précédemment réalisées en témoignent (Cf. Chapitre 1). Le modèle de distribution de la population en fonction des composantes environnementales est imparfait, puisque sujet à de nombreux contres exemples [6]. On peut citer deux exemples assez éloquentes, le premier étant le cas des *montagnes refuges*. L'altitude et les reliefs constituent une barrière naturelle diminuant les moyens de subsistance (peu de terres arables et diminution du nombre de cultures possibles [65]) et entravant la circulation [6, 65]. Il en résulte généralement une diminution de la population avec l'augmentation de l'altitude [65, 76]. Mais dans certaines situations ces inconvénients se transforment en avantages. Dans le cas des montagnes refuges l'enclavement devient une garantie de sécurité et conduit à une distribution de la population allant à l'encontre du modèle précédemment décrit (diminution de la densité lors de l'augmentation de l'altitude) [27]. À l'échelle de la planète de nombreux exemples existent (c'est notamment le cas de l'Albanie, de la Khabilie, ou encore du pays Dogon). De manière analogue l'existence de *déserts côtiers* va à l'encontre du modèle général, impliquant la présence

d'importantes densités sur les littoraux, lesquelles diminuent avec l'éloignement du littoral [65, 76]. Sans aller jusqu'à parler de *désert côtier*, les faibles densités de population en Aquitaine, sont un des exemples de faibles densités littorales.

Ces contre-exemples ne doivent pas leur existence uniquement à de simples composantes environnementales, d'autres facteurs doivent rentrer en considération. C'est en étudiant les peuplements successifs que l'on prend conscience de l'importance du rôle de l'histoire sur la distribution actuelle des populations. Comme pour les contraintes environnementales, l'action du temps sur la survie des populations est complexe, et c'est l'interaction de plusieurs phénomènes sur une certaine durée qui a conduit à la situation actuelle. L'image des sédiments, chère à VIDAL DE LA BLACHE illustre bien l'idée que la distribution de la population observable à une date t est le fruit d'une longue évolution, et comme les limons, la distribution de la population est la résultante de plusieurs années de sédimentation, constante ou non. De la même manière qu'une carte *chorochromatique* [13] ne représente que l'apparence externe et plane de volumes enfouis, les cartes *choroplètes* précédemment présentés ne figurent que la résultante d'une longue évolution de la distribution de la population.

Par conséquent la distribution de la population observable aujourd'hui est le fruit de siècles de développement humain. Un des premiers facteurs notables est l'ancienneté du peuplement, les premiers foyers de populations font aujourd'hui encore partie des zones les plus denses. Selon Fernand Braudel, « *Où l'homme se trouvait en 1500, il se trouve encore* » [6]. L'ancienneté du peuplement conduit (à conditions climatiques égales) à un peuplement plus dense. BAUDELLÉ [6] constate que généralement l'ancienneté du peuplement conduit à de plus fortes densités, qui elles-mêmes conduisent à un système agricole plus intensif, par conséquent plus productif. Toutefois ce constat n'est pas invariable. L'histoire n'est pas linéaire, des anciens foyers de peuplement ont pu, pour diverses raisons (guerres, invasions, fluctuations politiques) devenir des zones peu peuplées aujourd'hui, c'est par exemple le cas de la péninsule du Yucatan [6]. À notre échelle (spatiale et temporelle) nous ne pouvons pas prendre en considération cet élément. Comme nous l'avons déjà indiqué, en 1793, *i.e.* au début de notre série statistique, l'intégralité de la France métropolitaine est peuplée et il faudrait étudier une période énormément plus longue pour mettre en exergue ce phénomène. Ce dernier est particulièrement important lorsque l'on cherche à comparer la distribution de la population de différents pays, régions ou continents et justifie, à titre d'exemple, les principales différences de population entre pays colons et colonisés (à climat comparable) [6].

Mais les migrations, qu'elles soient volontaires, ou forcées, jouent un rôle encore plus important dans la distribution de la population et son évolution. Nous l'avons vu dans le premier chapitre, la concentration de la population observée en France métropolitaine durant notre période d'étude, et plus généralement à transformation de la distribution de la population sont expliquées par la *transition urbaine*. Phénomène dû en partie à un *exode rural* massif, qui est un important phénomène migratoire [3]. L'effet des migrations est généralement plus important que l'ancienneté du peuplement [60]. Le meilleur exemple est le continent américain, massivement peuplé à partir du XV^e siècle. L'émigration vers le continent a eu d'importantes répercussions. Non seulement le continent a été massivement peuplé, mais certains pays ont subi une émigration massive (Irlande, Royaume-Uni), dont les répercussions ont été très importantes. On estime qu'entre 1846 et 1932 l'Irlande a perdu le tiers de sa population [6, 65]. La vitesse de l'accroissement est un autre facteur essentiel expliquant la répartition du peuplement aujourd'hui. NOÏN [60] estime qu'il s'agit du facteur le « *plus important de tous pour comprendre la formation des densités* ». En effet à population de départ identique, de petits écarts de la vitesse d'accroissement peuvent conduire à de grandes différences de population. La vitesse d'accroissement varie suivant les époques et les pays. Certains événements historiques ont conduit à une hausse de la vitesse d'accroissement de la population, et donc à un fort accroissement de la population. Un des exemples les plus connus est celui de la révolution néolithique. En adoptant un mode de vie basé sur l'agriculture, permettant une production de nourriture plus régulière et importante que la chasse et la pêche, l'espèce humaine a connu un bond démographique important [6]. Ce même phénomène a eut lieu durant notre période d'étude. La révolution industrielle ayant également été l'occasion d'un important bond démographique [6]. Une fois encore, en adoptant un paradigme différent, permettant une production alimentaire plus importante, les êtres humains ont vu leur nombre augmenter de façon exponentielle. La révolution néolithique a fait passer la population humaine de 5 à 50 millions. Quant à la révolution industrielle, on estime qu'elle a permis à la population humaine d'être multipliée par 5. La croissance de la population est sujette à une progression par bonds [6, 65, 77].

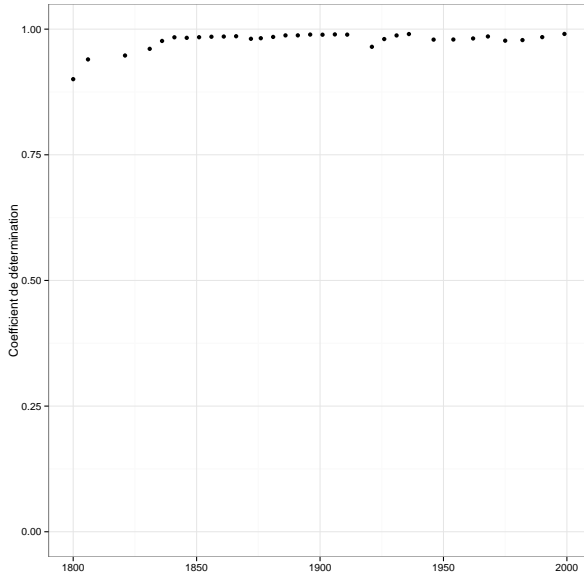
Tous ces éléments indiquent que le peuplement et par conséquent sa distribution sont sujets à une forte inertie. À l'échelle de la planète les grands foyers de population actuels sont les descendants des grands foyers de population du néolithique [6], et les grandes migrations d'hier ont encore des effets à l'heure actuelle. Ainsi il est très probable que *le peuplement d'aujourd'hui* soit avant tout expliqué par *le peuplement d'hier*, et par conséquent que la série statistique étudiée soit fortement autocorrélée temporellement comme elle l'est spatialement.

Il est possible, au même titre que l'autocorrélation spatiale, de quantifier l'importance de la compo-

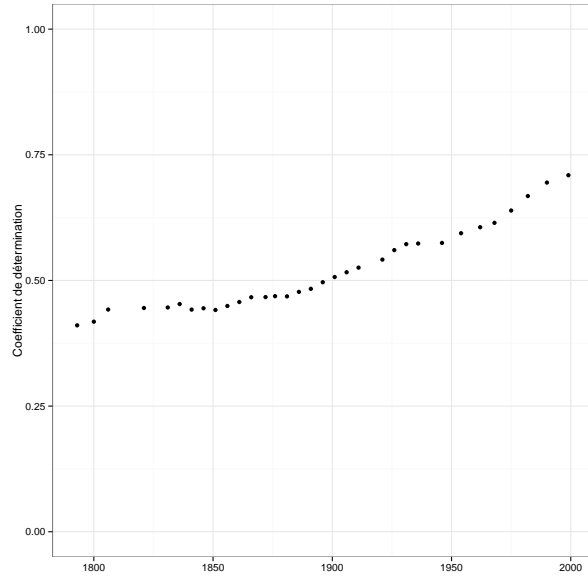
sante temporelle en mesurant l'autocorrélation temporelle de la variable étudiée. Comme nous l'avons déjà expliqué dans le premier chapitre la mesure de l'autocorrélation spatiale est préférentiellement réalisée à l'aide de l'indice de MORAN, qui est une adaptation du coefficient de corrélation à la problématique de l'autocorrélation spatiale. De manière similaire on peut mesurer l'autocorrélation temporelle en calculant la corrélation entre la variable à une date t et la valeur de la variable à une date $t - 1$. La figure 2.1a représente l'évolution du coefficient de détermination du modèle de régression linéaire où la variable indépendante correspond à *l'écart à la densité moyenne* au recensement $t - 1$ et la variable dépendante à *l'écart à la densité moyenne* au recensement t . Par conséquent seuls 32 points sont rapportés sur le graphique 2.1a. On remarque immédiatement les très fortes valeurs du coefficient de détermination. Ce dernier varie entre 0,90 (1800) et 0,99 (1999), avec une valeur moyenne de 0,97. Ces valeurs sont absolument *colossales*, et dans la situation où l'autocorrélation temporelle est la plus faible (1800), 90 % de la variance à la date t est expliquée par les valeurs à la date $t - 1$. De plus il convient de rappeler que nous travaillons sur plus de 30 000, ce qui rend la valeur du coefficient de détermination d'autant plus importante. Une telle valeur de l'autocorrélation temporelle confirme l'importance de la composante temporelle, *l'écart à la densité moyenne* d'une commune dépend en quasi-totalité de son précédent *l'écart à la densité moyenne*.

À ce stade nous avons cherché à savoir si la part de l'information non prise en compte par ces modèles pouvait être expliquée par l'autocorrélation spatiale, précédemment calculée. Nous avons donc réalisé un second modèle, de façon similaire au précédent, mais où la variable indépendante n'est plus l'écart à la densité moyenne à la date $t - 1$, mais la valeur moyenne des voisins à la date t suivant une matrice de voisinage de rang 1 (il s'agit comme précédemment d'un voisinage de MOORE). Il en résulte la figure 2.1b, logiquement très semblable à la figure 1.10a. Bien très importante compte tenu du nombre d'individus, l'autocorrélation spatiale explique une part de la variance nettement moins importante, et est plus sujette aux fluctuations que l'autocorrélation temporelle. La troisième figure illustre l'évolution du coefficient de détermination entre l'écart à la densité moyenne à la date t est la moyenne de l'écart à la densité moyenne des voisins à la date $t - 1$. Les valeurs du coefficient de détermination sont semblables à celles obtenues sur la figure 2.1b, mais sont légèrement inférieures. On note également que l'évolution du coefficient de détermination est identique à celle la figure 2.1b.

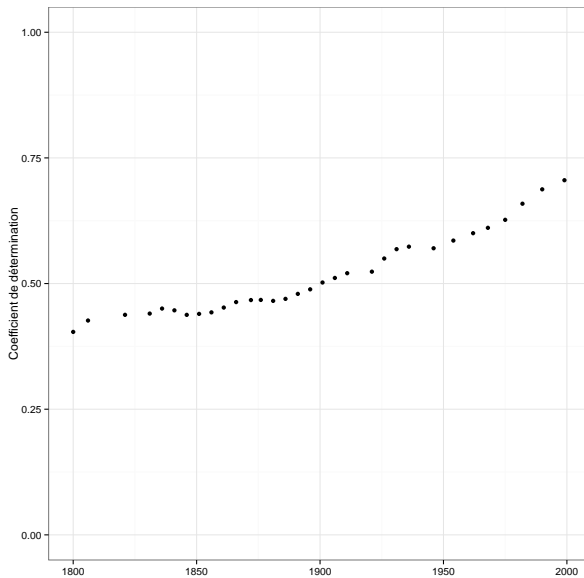
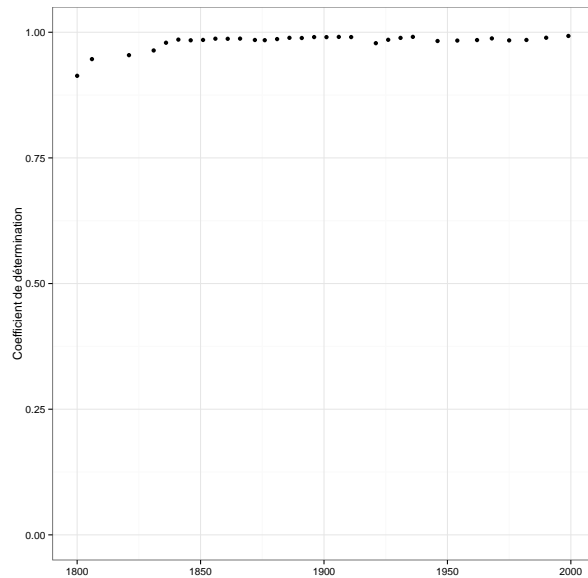
Pour chercher à quantifier l'importance des différentes autocorrélations nous avons réalisé un nouveau modèle (cf. figure 2.1d) combinant les modèles précédemment présentés. Ici la variable dépendante reste l'écart à la densité moyenne à la date t , et les variables indépendantes sont au nombre de trois, *l'écart à la densité moyenne* de la commune à la date $t - 1$, la moyenne des *écarts à la densité moyenne* pour les voisins et la moyenne des *écarts à la densité moyenne* pour les voisins à la date $t - 1$. La comparaison des figures 2.1a et 2.1d permet de remarquer une légère augmentation des coefficients de détermination. La valeur minimale passe à 0,91 (1800) et la valeur moyenne passe à 0,98. Les plus fortes augmentations concernent les valeurs les plus basses notamment le dénombrement de 1921 qui voit son coefficient de détermination passer de 0,96 à 0,98. L'augmentation apportée par l'ajout de la valeur des voisins au précédent modèle ne fait qu'augmenter très légèrement l'explication de la variance, la mesure de l'autocorrélation temporelle étant déjà extrêmement efficace. Mais il est extrêmement intéressant de constater les très fortes valeurs du coefficient de détermination, indiquant que la quasi totalité de la variance est expliquée par les composantes historiques et spatiales.



(a) Autocorrélation temporelle



(b) Autocorrélation spatiale

(c) Autocorrélation spatiale (voisins au temps $t - 1$)

(d) Combinaison des modèles précédents

FIGURE 2.1 – Évolution du coefficient de détermination en fonction de l'année du recensement pour quatre modèles d'autocorrélation. Les quatre figures ont été réalisées à l'aide d'une régression linéaire. La figure 2.1a à été réalisée en réalisant un modèle de régression linéaire entre l'écart à la densité moyenne à la date t avec l'écart à la densité à la date $t - 1$, la figure 2.1b en réalisant un modèle similaire entre la modalité d'une commune et la moyenne des modalités de ces voisins (voisinage de MOORE de rang 1), la figure 2.1c est identique à la précédente, si ce n'est qu'il s'agit de la valeur moyenne des voisins à la date $t - 1$, la figure 2.1d quant à elle est une combinaison des trois modèles précédents.

Sachant suite à des recherches bibliographiques l'importance de la composante temporelle en ce qui concerne la distribution de la population, nous nous sommes attelés à essayer de quantifier l'influence de la composante temporelle. Grâce à l'importante base de données dont nous disposons, il nous a été possible de calculer l'autocorrélation temporelle entre chaque recensement sur une période de 206 ans. La conclusion de ces mesures est sans appel. La distribution de la population à un recensement t dépend en moyenne à 97 % de la distribution de la population au précédent recensement $t - 1$. Ce constat peut paraître évident, et on pourra notamment avancer que la valeur des coefficients de détermination est avant tout explicable par la matrice de voisinage choisie. Il est vrai qu'avec une période censitaire médiane de 5 ans, l'écart entre chaque recensement est suffisamment faible pour que les changements dans la distribution de la population soient limités, et donc limitent la diminution du coefficient de détermination. En plus de nous donner une indication sur l'importance de la composante historique, le calcul de l'autocorrélation temporelle nous apporte également des éléments de réponse concernant la part de la variance expliquée par d'autres modèles, ne prenant pas en compte la composante temporelle.

2.1.3 Sources, données et traitements

La prise en compte de plusieurs composantes environnementales dans notre travail a nécessité la construction de nouvelles variables *ad hoc*, et ce à partir de nouvelles données. Comme expliqué précédemment trois (ou quatre) composantes environnementales sont généralement présentées comme déterminantes pour ce qui est de la distribution des populations. La première est bien évidemment la composante *climatique*. De nombreuses variables peuvent être utilisées pour illustrer cette composante, mais nous avons choisi de ne pas l'inclure à notre travail. Ce choix n'est pas motivé par une absence de données (bien au contraire les données portant sur le climat sont faciles à obtenir), mais par une question d'échelle. La composante climatique peut expliquer en partie la distribution de la population, mais seulement à petite échelle, lorsque l'étendue du territoire est suffisante pour voir apparaître de grandes variations climatiques [6]. De même la France est un pays située dans une zone tempérée [6] et aucune partie du territoire n'est caractérisée par un climat si extrême qu'il s'avère être un facteur « *limitant* » [6]. Si nous avions travaillé sur un autre pays, tel que l'Algérie, la composante climatique se serait sans doute avérée nettement plus importante. Mais dans le cas de la France la composante climatique joue à trop petite échelle pour être déterminante.

L'altitude est au contraire une composante importante. Qui joue aussi bien un rôle à l'échelle continentale [76] que locale. Les différentes cartes présentées dans le premier chapitre laissent bien apparaître les massifs montagneux comme étant des zones de densité inférieure à la moyenne. Pour intégrer la composante altimétrique dans notre travail nous avons commencé par obtenir les données d'altitude à partir du site de l'IGN. Les seules données d'altitude disponibles à l'échelle de la France métropolitaine gratuitement et librement sont les données issues du MNT BD ALTI, dont la précision maximale est un pas de 75 m. À partir de ce fichier nous avons attribué une altitude moyenne à chaque commune à l'aide d'une *Béta-interpolation* réalisée avec le logiciel SAGA GIS. C'est cette variable, que nous désignerons désormais sous le nom d'altitude, qui sera utilisée dans les traitements ultérieurs.

La seconde composante retenue est celle de l'hydrographie. Les données utilisées ont été produites par l'*office national de l'eau et des milieux aquatiques* (ONEMA) et ont été obtenues à partir du site *data.gouv.fr*. La base de données utilisée porte le nom de BD CARTHAGE (Base de Données sur la CARTographie THématique des AGences de l'eau et du ministère chargé de l'environnement) et la version utilisée a été mise à jour en septembre 2014. Contrairement à la base BD ALTI cette dernière est composée de multiples couches SIG. Nous n'avons utilisé que la couche *hydrographie surfacique*, dont nous avons sélectionné les polygones correspondant à la pleine mer. Une fois cette tâche réalisée nous avons transformé ces polygones en points, dans le but d'obtenir le tracé du littoral sous la forme d'un nuage de points. L'extension *mmgis* a ensuite été utilisée pour calculer la distance séparant chaque commune (ou plus précisément son centroïde) au point le plus proche (une des points parmi les 170 129 points créés à partir de la couche *hydrographie surfacique*) et donnant ainsi la distance entre une commune et le littoral le plus proche. Cette valeur est exprimée en mètres (et non en kilomètres) et varie entre 76 mètres et 405 kilomètres. C'est ici qu'on perçoit les limites de cette méthode de calcul basée sur la distance au centroïde. Aucune commune, même les communes littorales ou les îles ne possèdent une distance au littoral nulle, aucun centroïde n'étant confondu avec un des points de la couche *distance au littoral*. Pour conclure cette courte présentation de ces deux nouvelles variables nous avons réalisé le tableau 2.1, synthétisant les différentes informations sur ces nouvelles variables, notamment les sources ou leur distribution statistique.

Sources	Altitude	Distance au littoral
Producteur	IGN	ONEMA
Nom	BD ALTI 75 m	BD CARTHAGE
Site de téléchargement	<i>ign.fr</i>	<i>Data.gouv.fr</i>
Date de mise à jour	Octobre 2013	Septembre 2014
Licence	Licence ouverte 1.0	Licence ouverte 1.0
Traitements	Altitude	Distance au littoral
Modification données	<i>Na.</i>	Extraction des contours
Calcul de la variable	<i>Béta-interpolation</i>	Calcul de la distance
Distribution	Altitude	Distance au littoral
Unité	<i>mètres</i>	<i>mètres</i>
Moyenne	277,8	177 781,0
Minimum	0,1	75,9
Premier quartile	103,7	70 820,1
Médiane	185,5	152 619,5
Troisième quartile	333,7	270 800,4
Maximum	2 700,0	405 088,9
Écart-type	289,9	127 498,3

TABLE 2.1 – Variables environnementales : Sources, traitements et distribution.

Cette première partie nous a permis de présenter les différents éléments expliquant la distribution de la population en fonction de l'environnement. Deux points importants sont à noter. Le premier est que la composante la plus importante reste le temps, la distribution de la population à un temps t dépendant avant tout de la distribution de la population au temps $t - 1$, comme notre estimation de l'autocorrélation temporelle l'a montré. Le second point est que le rôle de l'environnement n'est pas direct. Celui-ci a avant tout un rôle limitant, mais mêmes les régions les plus contraignantes sont habitées. Cette observation est d'autant plus valable lorsqu'elle est appliquée à la France métropolitaine, cette dernière ne possédant pas de régions au climat si rude qu'il en devient très contraignant.

2.2 Approche cartographique

Dans cette seconde partie nous allons étudier l'influence de l'environnement sur la distribution de la population et sur son évolution. Comme expliqué dans l'introduction nous commencerons par adopter une démarche cartographique qui sera ensuite complétée par des modèles statistiques dans la partie suivante.

2.2.1 Évolution de la superficie

Pour répondre à ces questions nous avons réalisé les figures 2.2a et 2.2b. Il s'agit d'une production orientale et qui nécessite par conséquent quelques explications. Le phénomène de *concentration de la population* peut-être exprimé de deux manières différentes, on peut le présenter comme étant l'augmentation de la population contenue dans une surface fixe, *i.e.* que la concentration de la population, et donc la densité d'une commune quelconque (mais également de tout autre espace) augmente lorsque sa population augmente (dans le cas où sa surface ne varie pas). Les figures 2.2a et 2.2b sont basées sur la relation inverse, la concentration de la population peut-être considérée comme la diminution de la surface occupée par une population fixe. Les figures 2.2a et 2.2b illustrent donc la part de la superficie totale occupée par une même part de la population. A bien des égards les figures 2.2a et 2.2b ressemblent fortement à la figure 1.6c, illustrant l'évolution de la part de l'effectif contenant une part fixe de la population. Mais la véritable innovation de nos figures intervient avec l'introduction de la question environnementale dans leur construction. Chaque « bande » visible sur les figures 2.2a et 2.2b correspondent à un quart de la population totale, ordonnées selon l'altitude ou la distance à la mer. Ainsi, pour la figure 2.2a, la « bande » la plus claire représente l'évolution de la superficie occupée par le quart de la population situé aux plus basses altitudes, et la « bande » la plus sombre le quart de la population situé aux plus hautes altitudes. De la même manière pour la figure 2.2b la « bande » la plus claire correspond à la superficie occupée par le quart de la population le plus proche des littoraux. D'un point de vue purement technique la réalisation de ces figures est assez simple, et est inspirée du calcul des fréquences cumulées. On commence par calculer la part de la population et de la Superficie totale représentée par une commune (on travaille par conséquent à partir de la population, et non de la variable de travail utilisée tout au long de la première partie) et on réitère cette opération pour chaque dénombrement (uniquement pour la population, le fond

de carte étant invariable). Là où intervient la principale différence avec le calcul des fréquences cumulées et que l'on fait le cumul des fréquences en fonction de la variable environnementale choisie et non en fonction du rang. En conclusion on obtient la part de la superficie occupée par les n % communes les plus basses ou les plus proches du littoral. Notre choix s'est porté sur des quartiles. Notons également que les fluctuations visibles, majoritairement avant 1850, sur les figures 2.2a et 2.2b sont imputables aux communes sans données (cf. tableau 1.2). Toutefois ces fluctuations ne nous empêchent pas de percevoir l'évolution globale du phénomène étudié.

Altitude

Des deux figures réalisées la figure 2.2a est sans aucun doute la plus éloquente. L'évolution du phénomène étudié est clairement visible, entre 1793 et 1999 on remarque une diminution de la superficie occupée par les 50 % de la population situés aux plus faibles altitudes, une augmentation de la surface occupée par le quart de la population situé aux plus hautes altitudes et une stagnation de la troisième classe. Deux classes occupent une superficie moindre (de 0 à 50 %) une classe stagne (de 50 % à 75 %) et une classe occupe une plus grande superficie (de 75 à 100 % de la population de cumulée). La variation peut-être très importante, comme l'indiquent les valeurs du tableau 2.2. La superficie occupée par le premier quartile diminue de moitié et celle du dernier quartile augmente de 14 points. En 1999 le dernier quartile occupe presque la moitié de la superficie totale, là où le premier quartile en occupe 9 %. Étant donné que nous travaillons à partir d'un pourcentage de la population totale cela ne signifie pas que les plus hautes communes perdent nécessairement de la population, mais que leur importance relative diminue, plus de communes étant nécessaires pour représenter un quart de la population. Inversement la part de la superficie totale occupée par les deux premières classes diminue fortement, la même part de la population étant concentrée dans une superficie de plus en plus petite. Si l'on se réfère au tableau 2.2 on peut observer plus en détail l'évolution de la part de la superficie totale occupée par un quartile de la population. Les deux premiers quartiles (Q1 et Q2) subissent une baisse quasi-continue, mais là où le second quartile voit la part de surface qu'il occupe diminuer jusqu'en 1990, celle du premier quartile atteint son minimum en 1906, puis elle reste stable. Ce qui implique qu'aucune commune ne quitte cette classe entre 1906 et 1911. Le tableau 2.2 confirme également la stagnation du quartile Q3, la part de la superficie occupée par ce dernier reste comprise en 25 et 29 % tout au long de la période. Quant à la classe Q4, il s'agit de la seule classe dont la part de la superficie occupée augmente globalement, pour atteindre son maximum en 1999. Différentes variations sont perceptibles, notamment en ce qui concerne la classe Q4. Jusqu'à la moitié du XIX^e siècle la part de la surface occupée fluctue aussi bien à la hausse qu'à la baisse, et ce n'est qu'à partir du dénombrement de 1856 qu'apparaît une véritable tendance à la hausse qui dure jusqu'en 1999 (à l'exception des dénombrements de 1861 et de 1876). À l'inverse la tendance à la diminution de la surface occupée par la classe Q1 n'est pas soumise à des fluctuations. Celle-ci est initiée dès le dénombrement de 1800 et continue jusqu'en 1906. Seul le dénombrement de 1836 contraste avec cette observation, puisqu'il est caractérisé par une forte augmentation (10,35 %) de la superficie occupée immédiatement suivie d'une importante diminution (-9,59%) de cette même superficie. Cette importante augmentation de la superficie occupée par la première classe (Q1) se retrouve également à la même date pour la seconde classe (Q2) qui subit une augmentation de 7,56 % entre 1831 et 1836, suivie d'une diminution de 3,63 % entre 1836 et 1841. Ce « bond » est évidemment imputable aux communes *absentes* et *lacunaires*. Le tableau 1.2 nous indique qu'avec ses 1 191 communes sans données le dénombrement de 1836 n'est pas le plus touché par l'absence de données, mais la carte 2.3a et plus généralement la planche 2.3 permet de constater que la grande majorité des communes sans données (*i.e.* les départements du Gers, de la Charente et de la Loire-Atlantique) sont incluses dans les deux premières classes, comme nous le montre la carte 2.3b. Et il s'agit du seul dénombrement pour lequel une part importante des communes absentes appartiennent aux classes Q1 et Q2, dans les autres cas (1831, 1851, *etc.*) les communes absentes sont principalement localisés dans des départements appartenant aux classes Q3 et Q4 (cf. carte 2.3) (Savoie, Haute-Savoie, Vosges, *etc.*) ce qui se traduit également par une importante variation de la part de la superficie occupée (c'est notamment le cas en 1831 avec une augmentation de 4,01 % de la surface occupée et une diminution de 6,45 % en 1851. De la même manière la classe Q4 voit la part de la surface qu'elle occupe augmenter de 3,66 % en 1831 et de 8,34 % en 1851, ce qui constitue une des plus grandes augmentations). Lors du calcul de la part de la superficie occupée par les différents quartiles, ces communes de population inconnue (et donc de modalité nulle) n'ont pas participé à l'augmentation de la valeur de la population cumulée, alors qu'elles ont participé à l'augmentation de la part de la superficie occupée, ce qui se traduit par une augmentation de la part de la superficie occupée de certaines classes (*i.e.* celles contenant les communes sans données) aux prix d'une diminution de la part de la superficie occupée par les autres classes (le total étant toujours égal à 100 %).

Après avoir décrit cette figure il est désormais nécessaire de ce questionner sur l'interprétation de ces résultats. Pour ce faire nous allons débiter par le dénombrement de 1793. Le tableau 2.2 et la

figure 2.2a nous montrent qu'il s'agit de la situation où les quatre classes présentent les valeurs les plus proches. Les deux premières classes présentent une part de la superficie inférieure à 25 % (*i.e.* à la valeur correspondant à une répartition égalitaire entre les quatre classes). Inversement les classes Q3 et Q4 occupent une superficie supérieure à 25 %. Étant donnée que la population de chaque classe est identique, on peut interpréter la part de la superficie occupée comme une mesure de la densité de population, plus la superficie occupée est faible, plus la densité est importante. Ainsi en 1793, il existe un gradient de densité entre la classe Q1 (la plus dense) et la classe Q4. Cette structure est valable pour chacune des trente-trois dates étudiés. Le fait que la densité moyenne tende à diminuer avec l'altitude n'est pas surprenant, nous avons déjà parlé de ce point précédemment et de nombreux travaux ont mis en évidence ce point [6, 8, 65, 76]. Plus que la structure à un temps t , nous sommes avant tous intéressés par l'évolution de cette structure. Si, comme dit plus haut, sa hiérarchie ne varie pas, la densité relative des classes change fortement. À partir de 1836, la classe Q1 subit une diminution continue de la superficie qu'elle occupe, ce qui traduit une augmentation de la densité de population et par conséquent une concentration de la population aux plus basses altitudes. Inversement, la superficie occupée par la dernière classe ne cesse d'augmenter, ce qui implique une déconcentration de la population située aux plus hautes altitudes. Il est nécessaire de mettre ces observations en perspective à l'aide d'un commentaire historique et géographique. Au vu des ces tableaux et de ces graphiques on peut supposer que les variations observées correspondent à de grandes mutations de la distribution de la population et notamment à la tendance au dépeuplement des montagnes observable à partir de la moitié du XIX^e siècle [8, 31, 32]. Mais nous attendrons la prochaine sous-partie pour approfondir ce commentaire, notamment en procédant à la cartographie des réalisations précédemment présentées.

Distance à la mer

La figure 2.2b ne présente pas une structure aussi marquée que celle de la figure 2.2a. Ici, aucune classe ne semble gagner en importance, toutes restent proches des mêmes valeurs. Seule la classe Q4 voit sa superficie augmenter, au prix de la diminution de celle de la classe Q2. La tableau 2.3 confirme cette observation. Si l'on se réfère à la première ligne du tableau (1793) on remarque que nous sommes beaucoup plus proches du modèle de distribution homogène (Q1 25 %, Q2 25%, Q3 25% et Q4 25%) que pour le tableau précédent. Mais comme précédemment la comparaison des premières et des dernières lignes indique également un éloignement, certes moins marqué, du modèle de distribution homogène. Entre 1793 et 1999 la surface occupée par le premier quart de la population diminue de 2 points, celle du second quart augmente de 2 points, celle du troisième quart diminue de 5 points et celle du dernier quart augmente de 5 points. Non seulement l'intensité des variations est sans commune mesure avec celles observées sur le tableau 2.2a, mais l'évolution des classes fluctue beaucoup plus. Là où précédemment nous observions une franche tendance à la diminution de la superficie pour la classe Q1 (à partir de 1836), pour la distance au littoral cette dernière fluctue beaucoup plus. Les phases d'augmentations côtoient les phases de diminution sans réelle logique. Si comme nous l'avons dit précédemment la superficie occupée par la classe Q1 diminue de deux points entre 1793 et 1999, elle ne diminue que d'un point entre 1793 et 1982 et d'un autre point entre 1982 et 1999. On peut faire les mêmes observations pour les autres classes, des périodes d'augmentation de la superficie côtoient des périodes de diminutions sans réelle logique. Seule la classe Q4 échappe à cette règle, puisqu'elle subit une augmentation continue à partir de 1876, même si cette dernière reste faible comparée à celle que nous avons observé précédemment pour la classe Q4.

Nous ne pouvons pas donner de seuil à partir du quel l'augmentation (ou la diminution) de la surface occupée par un quartile de la population serait suffisante pour que l'on puisse estimer qu'il y a une réelle modification de la répartition de la population en fonction d'une composante environnementale. Les grandes variations observées précédemment en fonction de l'altitude laissent supposer que cette modification est bien réelle, mais cette observation n'est pas applicable au cas de la distance au littoral. Ce second cas est plus délicat, au vu des résultats obtenus il semble clair que si il existe un lien entre la distribution de la population et la distance au littoral ce dernier est nettement moins *fort* qu'il ne l'est entre la distribution de la population et l'altitude. De plus nous n'observons pas comme précédemment une structure en gradient, impliquant une diminution de la densité moyenne en fonction de l'altitude. Ici la classe Q3 occupe systématiquement une surface plus petite que la classe Q2, voir de la classe Q1 (à partir de 1866), la classe Q3 est donc globalement plus dense que ces dernières (en tout cas à partir de 1866). À ce stade nous nous sommes demandés si le problème ne venait pas du découpage en quartiles, qui convenait parfaitement au cas précédent, mais qui pourrai s'avérer trop grossier pour cette seconde étude. Nous avons donc réalisé un second découpage, mais cette fois par déciles. Celui-ci se présente uniquement sous la forme d'un tableau (cf. figure 2.4, nous avons choisis de ne pas faire figurer la colonne de variation par rapport à la date précédente pour des raisons de lisibilité ainsi que de mise en page), une figure similaire aux précédentes s'étant avérée peu lisible avec un si grand nombre de classes.

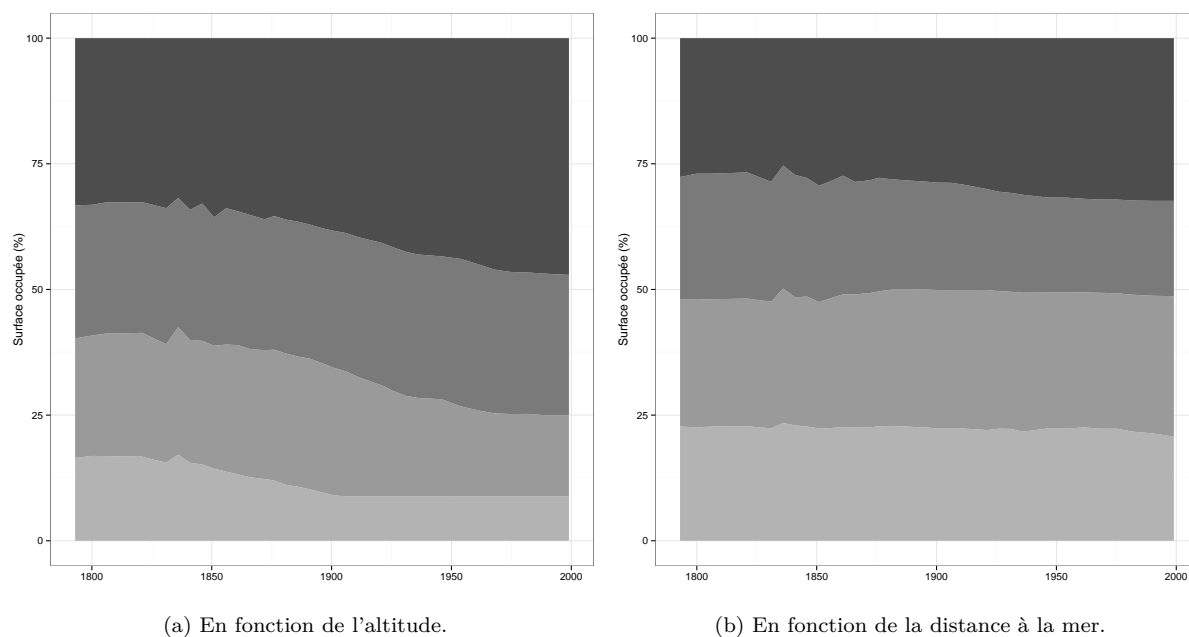


FIGURE 2.2 – Part de la superficie totale occupée pour chaque quartile de la population en fonction de l'altitude et de la distance à la mer.

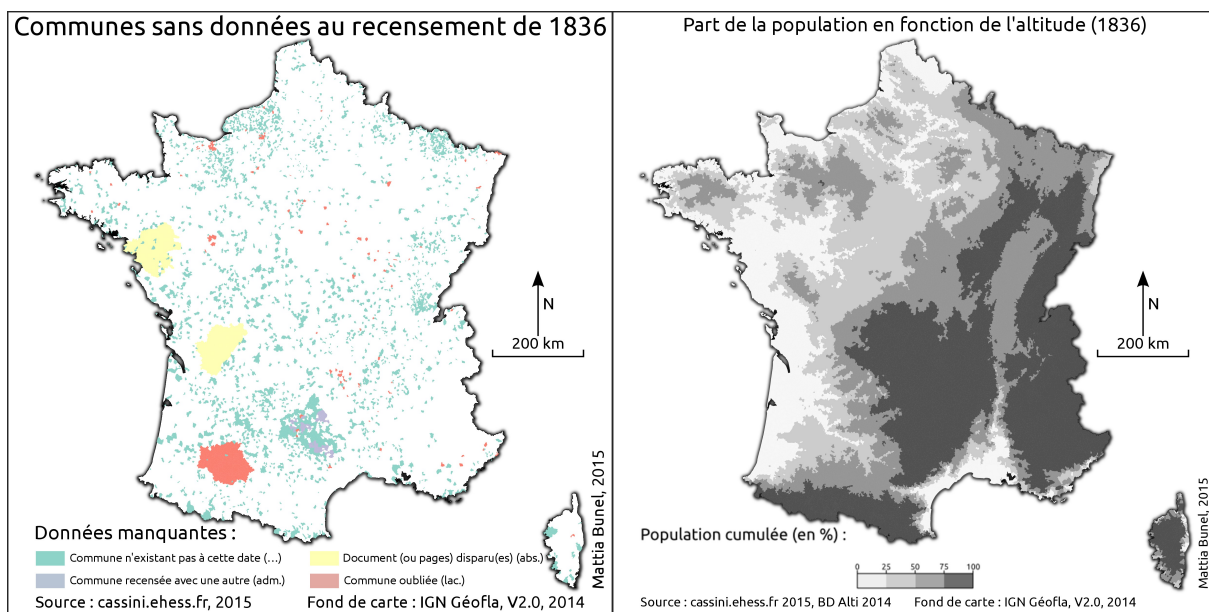
Ce tableau nous apporte plus des détails, et nous permet de remarquer que seules certaines classes (la moitié) voient la surface qu'elles occupent varier de manière significative. C'est le cas des classes Q1, Q2, Q4, Q6, Q7, Q8 et Q10. Pour ces classes la variation la plus faible est de l'ordre de trois points, alors qu'elle est généralement de l'ordre d'un point pour les 4 autres classes. Parmi ces six classes la première est la plus intéressantes, notamment parce qu'elle présente la plus grande diminution de la superficie occupée (avec la classe Q6) de la série. Cette dernière représente les 10 % de la population totale les plus proches des mers et océans. En 1793, cette part de la population occupe 8,78 % du territoire, et cette part diminue de façon presque continue, pour atteindre la valeur de 3,55 % en 1999 (le minimum est de 3,50 % en 1975)¹. Ainsi la même part de la population (population qui on le rappelle augmente tout au long de la période, à quelques exceptions, cf. figure 1.3a) occupe une superficie de plus en plus faible et il en résulte une densification et une par conséquent une concentration de la population vivant sur la bande littorale. GÉRARD-FRANÇOIS DUMONT a proposé le terme de « litturbanisation » pour qualifier le « développement du peuplement des espaces littoraux et sublittoraux » [82]. Ce nouveau découpage nous permet de mettre indirectement en évidence le « développement du peuplement des espaces littoraux [...] », mais non celui des espaces sublittoraux. Les classes Q2, Q3, Q4 subissent quant à elles une augmentation de leur superficie peu marquée sauf pour la classe Q4 (plus marquée). Les 30 % de la population situées après les 10 % les plus proches du littoral, occupent une superficie de plus en plus importante, ce qui va à l'encontre des résultats attendus par le modèle de DUMONT. Mais cette différence est probablement due à la méthode employée, celle-ci n'est pas conçue pour mettre en évidence un phénomène tel que la litturbanisation. Là où ZANINETTI [82] étudie l'évolution de la population en fonction d'un découpage géographique (dans ce cas il s'agit d'un cumul entre un découpage par contiguïté, pour les communes littorales, et par distance, pour les communes sublittorales) notre méthode nous permet de nous interroger sur l'évolution de la surface occupée par une part de la population fixe, la mise en évidence de la densification des communes littorales est uniquement permise par la concordance entre l'échelle choisie (par déciles) et le phénomène qui concerne la même part de la population.

La seconde classe connaissant la plus grande diminution de sa superficie et la classe Q6. L'explication de cette importante diminution est simplement que cette classe contient la commune de Paris. Or comme nous l'avons montré dans le précédent chapitre, les grandes agglomérations, et notamment Paris on subit une forte augmentation de leur population au cours de la période étudiée, notamment à cause de l'industrialisation, de l'exode rural et plus généralement du phénomène de *métropolisation* [3, 31, 32].

1. Notons également que le chiffre de 3,55 % de la superficie, occupée par 10 % de la population en 1999 corrobore les observations de ZANINETTI [82], lequel indique que les 871 communes littorales concentrent en 1999 10 % de la population sur 4 % du territoire.

Date	Q1	Var. (%)	Q2	Var. (%)	Q3	Var. (%)	Q4	Var. (%)
1793	16,48		23,77		26,49		33,26	
1800	16,89	2,47	23,95	0,76	26,00	-1,83	33,15	-0,31
1806	16,83	-0,34	24,39	1,84	26,10	0,39	32,67	-1,47
1821	16,72	-0,64	24,64	1,02	26,01	-0,37	32,63	-0,13
1831	15,52	-7,22	23,61	-4,18	27,05	4,01	33,82	3,66
1836	17,12	10,34	25,39	7,56	25,70	-4,98	31,78	-6,04
1841	15,48	-9,59	24,47	-3,63	25,89	0,74	34,15	7,47
1846	15,21	-1,77	24,57	0,43	27,32	5,51	32,90	-3,68
1851	14,36	-5,59	24,44	-0,54	25,56	-6,45	35,64	8,34
1856	13,76	-4,13	25,26	3,36	27,13	6,15	33,84	-5,05
1861	13,21	-3,98	25,71	1,78	26,61	-1,93	34,46	1,84
1866	12,64	-4,33	25,50	-0,84	26,68	0,28	35,17	2,07
1872	12,28	-2,84	25,65	0,59	26,03	-2,46	36,04	2,46
1876	12,02	-2,17	26,00	1,35	26,59	2,16	35,40	-1,78
1881	11,13	-7,40	26,10	0,49	26,65	0,23	36,10	1,98
1886	10,76	-3,27	25,89	-0,89	26,83	0,70	36,51	1,14
1891	10,23	-4,94	25,99	0,40	26,68	-0,56	37,09	1,59
1896	9,64	-5,82	25,69	-1,15	26,85	0,62	37,82	1,96
1901	9,02	-6,34	25,37	-1,26	27,26	1,53	38,34	1,39
1906	8,83	-2,12	24,87	-1,98	27,53	0,99	38,77	1,11
1911	8,83	0,00	23,76	-4,46	27,82	1,07	39,58	2,10
1921	8,83	0,00	22,07	-7,09	28,36	1,88	40,74	2,93
1926	8,83	0,00	20,90	-5,29	28,59	0,87	41,66	2,26
1931	8,83	0,00	20,01	-4,29	28,65	0,18	42,51	2,03
1936	8,83	0,00	19,59	-2,09	28,54	-0,38	43,03	1,24
1946	8,83	0,00	19,32	-1,39	28,43	-0,37	43,41	0,88
1954	8,83	0,00	17,90	-7,32	29,33	3,17	43,93	1,19
1962	8,83	0,00	17,02	-4,91	29,03	-1,04	45,11	2,70
1968	8,83	0,00	16,52	-2,97	28,60	-1,47	46,05	2,07
1975	8,83	0,00	16,38	-0,84	28,26	-1,20	46,53	1,04
1982	8,83	0,00	16,41	0,22	28,12	-0,48	46,63	0,21
1990	8,83	0,00	16,11	-1,88	28,18	0,20	46,88	0,54
1999	8,83	0,00	16,11	0,00	27,96	-0,78	47,10	0,47

TABLE 2.2 – Part de la superficie occupée (en %) (et sa variation) par décile de la population française en fonction de l'altitude. La colonne *Q1* correspond au quart de la population situé aux plus faibles altitudes et la colonne *Q4* au quart de la population situé aux plus hautes altitudes.



(a) Communes sans données.

(b) Population cumulée en fonction de l'altitude.

FIGURE 2.3 – Population cumulée en fonction de l'altitude en 1836.

Date	Q1	Var. (%)	Q2	Var. (%)	Q3	Var. (%)	Q4	Var. (%)
1793	22,73		25,25		24,40		27,62	
1800	22,55	-0,82	25,44	0,77	25,05	2,61	26,95	-2,47
1806	22,73	0,81	25,32	-0,50	24,98	-0,27	26,96	0,04
1821	22,78	0,21	25,41	0,35	25,10	0,47	26,71	-0,95
1831	22,38	-1,81	25,20	-0,83	23,84	-5,28	28,58	6,55
1836	23,44	4,54	26,76	5,82	24,46	2,51	25,35	-12,76
1841	22,96	-2,11	25,49	-4,95	24,33	-0,53	27,22	6,89
1846	22,73	-0,98	25,86	1,40	23,58	-3,16	27,83	2,18
1851	22,36	-1,67	25,14	-2,83	23,13	-1,95	29,36	5,24
1856	22,43	0,31	25,86	2,76	23,31	0,75	28,41	-3,37
1861	22,62	0,82	26,42	2,11	23,61	1,29	27,36	-3,83
1866	22,52	-0,41	26,51	0,35	22,38	-5,50	28,59	4,31
1872	22,55	0,10	26,72	0,80	22,45	0,30	28,28	-1,07
1876	22,78	1,04	26,83	0,41	22,60	0,70	27,78	-1,82
1881	22,85	0,28	27,12	1,06	21,97	-2,88	28,06	1,00
1886	22,78	-0,32	27,23	0,41	21,74	-1,05	28,25	0,67
1891	22,63	-0,66	27,41	0,63	21,57	-0,82	28,40	0,54
1896	22,55	-0,34	27,43	0,07	21,46	-0,50	28,56	0,57
1901	22,34	-0,95	27,55	0,44	21,37	-0,42	28,75	0,63
1906	22,34	0,00	27,54	-0,02	21,34	-0,14	28,78	0,12
1911	22,34	0,00	27,55	0,02	21,00	-1,63	29,12	1,16
1921	22,03	-1,39	27,86	1,14	20,08	-4,55	30,02	3,00
1926	22,30	1,21	27,37	-1,81	19,78	-1,52	30,55	1,72
1931	22,26	-0,20	27,29	-0,31	19,67	-0,55	30,79	0,77
1936	21,71	-2,51	27,65	1,31	19,49	-0,96	31,16	1,19
1946	22,34	2,80	27,07	-2,15	18,91	-3,06	31,69	1,69
1954	22,34	0,00	27,07	0,00	18,87	-0,18	31,72	0,11
1962	22,55	0,92	26,87	-0,72	18,61	-1,41	31,97	0,78
1968	22,30	-1,10	27,04	0,61	18,55	-0,29	32,11	0,42
1975	22,30	0,00	26,95	-0,31	18,64	0,44	32,11	0,00
1982	21,71	-2,72	27,24	1,03	18,75	0,62	32,30	0,60
1990	21,38	-1,56	27,37	0,50	18,88	0,65	32,37	0,22
1999	20,72	-3,18	27,94	2,01	18,97	0,51	32,37	0,00

TABLE 2.3 – Part de la superficie occupée (en %) (et sa variation) par quart de la population française en fonction de sa distance au littoral. La colonne *Q1* correspond au quart de la population le plus proche des littoraux et la colonne *Q4* au quart de la population le plus éloigné des littoraux.

Date	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10
1793	8,78	9,79	8,62	10,48	10,31	7,93	11,52	10,53	11,92	10,11
1800	8,44	9,76	8,81	10,67	10,32	8,50	11,25	11,05	11,69	9,53
1806	8,55	9,87	8,83	10,58	10,22	8,29	11,36	11,05	11,85	9,41
1821	8,47	10,00	8,81	10,58	10,32	8,41	11,32	10,91	11,95	9,22
1831	8,09	9,89	8,79	10,70	10,12	7,51	11,33	10,88	13,24	9,47
1836	8,32	10,63	9,41	11,42	10,42	8,15	10,85	10,65	11,60	8,55
1841	8,16	10,36	8,98	10,69	10,26	7,20	11,92	11,25	12,43	8,74
1846	7,86	10,32	9,10	10,85	10,47	6,92	11,87	10,08	12,96	9,58
1851	7,62	10,16	8,84	10,61	10,28	6,30	11,89	10,44	12,90	10,96
1856	7,35	10,38	9,00	11,02	10,53	5,77	12,36	10,65	12,81	10,12
1861	7,30	10,59	9,13	11,24	10,78	5,48	12,71	10,77	12,41	9,60
1866	7,16	10,47	9,09	11,45	10,87	4,96	12,52	10,25	12,02	11,21
1872	7,16	10,53	8,95	11,59	11,05	4,90	12,76	10,14	12,11	10,82
1876	7,19	10,79	9,17	11,86	10,61	4,85	12,92	10,06	12,60	9,95
1881	7,06	10,91	9,26	12,16	10,58	4,07	13,05	10,12	12,70	10,09
1886	6,75	11,02	9,31	12,32	10,60	3,78	13,17	10,12	12,67	10,26
1891	6,48	11,12	9,12	12,58	10,73	3,49	13,31	10,05	12,85	10,27
1896	6,24	11,36	8,96	12,73	10,68	3,20	13,38	10,18	13,01	10,25
1901	5,96	11,57	8,81	12,84	10,69	2,78	13,62	10,15	13,24	10,33
1906	5,71	11,83	8,75	12,91	10,68	2,58	13,71	10,26	13,33	10,25
1911	5,54	11,99	8,75	13,07	10,53	2,37	13,52	10,38	13,52	10,33
1921	5,22	12,18	8,22	13,79	10,49	1,70	12,94	11,35	13,18	10,93
1926	5,10	12,44	8,36	13,49	10,28	1,88	12,39	11,94	12,93	11,19
1931	4,59	12,70	8,51	13,75	9,99	1,93	11,89	12,52	12,90	11,21
1936	3,86	12,78	8,84	13,69	10,19	1,84	11,97	12,72	12,90	11,21
1946	5,29	12,24	8,65	13,48	9,73	1,73	11,42	13,24	12,82	11,39
1954	4,77	12,77	8,67	13,87	9,33	1,73	11,29	13,40	12,97	11,21
1962	4,37	13,17	9,00	14,49	8,39	2,09	10,71	13,59	12,93	11,26
1968	3,82	13,42	8,96	14,72	8,43	2,32	10,24	13,52	13,20	11,39
1975	3,50	13,05	9,24	14,99	8,47	2,90	9,87	13,28	13,24	11,46
1982	3,52	12,62	9,34	14,40	9,07	3,00	9,47	13,67	13,26	11,66
1990	3,61	12,06	9,59	14,17	9,33	3,08	9,08	13,63	13,04	12,42
1999	3,55	11,59	9,74	14,03	9,74	3,02	8,80	13,84	12,96	12,72

TABLE 2.4 – Part de la superficie occupée (en %) (et sa variation) par quart de la population française en fonction de sa distance au littoral. La colonne *Q1* correspond au dixième de la population le plus proche des littoraux et la colonne *Q10* au dixième de la population le plus éloigné des littoraux.

Cette sous-partie avait pour objectif de présenter les résultats de la méthode que nous avons développée pour étudier la distribution de la population en fonction d'une composante environnementale (dans notre cas l'altitude et la distance au littoral). Deux cas de figure se sont présentés, tout d'abord nous avons commencé par étudier la distribution de la population en fonction de l'altitude. Les quatre quartiles présentaient tous une évolution facilement décelable, les deux premiers voyaient leur superficie décroître (de manière plus marquée pour le premier quartile), le troisième quartile conservait une surface proche et le dernier voyait sa surface augmenter de manière significative. Ces transformations nous laissent supposer qu'il y a eut une transformation de la distribution de la population en fonction de l'altitude, cette dernière c'est concentrée dans les zones les plus basses, aux prix de la perte d'importance (relative) des communes situées aux plus hautes altitudes. Cette observation est corroborée par certaines de nos lectures bibliographiques, notamment par JACQUELINE BEAUJEU-GARNIER qui dès 1958 parlait de « dépeuplement [...] des montagnes » [8]. De même une structure en gradient était perceptible tout au long de la période étudiée, et ce malgré les différentes variations. La superficie occupée par les différents quartile augmente avec l'altitude, ce qui implique une diminution de la densité avec l'altitude, phénomène également confirmé par nos lectures [6, 76]. Pour ce qui est de l'évolution de la distribution de la population en fonction de la distance au littoral le constat est fort différent. Ici la division en quartiles s'est avérée insuffisante pour voir apparaître une quelconque structure et il a été nécessaire de recourir à une subdivision plus fine, les déciles, pour plus de précision. La structure la plus intéressante était celle de la première classe Q1, qui nous a permis de mettre en évidence la densification du littoral, composante du phénomène de *litturbanisation* [82]. Mais la méthode utilisée présente des inconvénients et nous n'avons pas pu mettre en évidence la densification des communes probablement à cause d'une échelle encore trop grossière. De plus nous n'avons pas observé l'existence d'un gradient comparable à celui de la variable altitude, là où cette dernière met en évidence une structure concernant l'ensemble du territoire, la distance au littoral ne semble impacter que les communes littorales, ou très proches du littoral.

Au vu de nos résultats il nous est apparu intéressant de compléter cette observation en réalisant une cartographie des communes selon leur classe d'appartenance, ce complément peut nous apporter de nouvelles informations et ce au prix d'une mise en œuvre simple.

2.2.2 Cartes

Comme indiqué dans la conclusion de la précédente sous-partie, nous allons compléter le précédent commentaire en cartographiant les résultats précédemment commentés. De la même manière que pour les cartes déjà présentées nous ne commenterons qu'un échantillon de l'ensemble des cartes réalisées, comme précédemment l'intégrale des cartes figure dans le volume annexe.

Altitude

La planche 2.5 contient six cartes représentatives de l'évolution de la distribution de la population en fonction de l'altitude précédemment décrite. Les dates retenues sont le début de la série statistique (1793), sa fin (1999) ainsi que les recensements de 1851, 1901, 1946 et 1975. Conformément aux résultats présentés dans la précédente partie, nous voyons une diminution du nombre de communes appartenant aux classes Q1 et Q2, et par conséquent une diminution de l'aire occupée par ces classes. À l'inverse la classe Q4 englobe de plus en plus de communes, et son aire ne cesse d'augmenter, jusqu'à atteindre son étendue maximum en 1999.

C'est en 1793 que la classe Q1 occupe une surface maximale, la carte 2.5a, nous permet d'identifier les communes incluses dans cette dernière. On y retrouve logiquement la majorité des communes littorales, à l'exception de nombreuses communes de la côte d'azur et du littoral corse, ainsi que certaines communes du littoral normand. De même certaines communes sublittorales (notamment au niveau de la côte atlantique) appartiennent à cette classe, à ces dernières s'ajoutent une partie des communes appartenant aux grandes vallées fluviales, on distingue clairement le sud de la vallée du Rhône, ainsi qu'une importante partie des vallées de la Seine et de la Garonne. C'est dans la seconde classe (Q2) que l'on retrouve la grande majorité des communes littorales non incluses dans la classe précédente, notamment la majorité des communes du littoral corse et de la côte d'azur. À ces dernières s'ajoutent une partie des communes de la Bretagne intérieure (*Arcoat*), ainsi que la région de la Beauce. La troisième classe contient des communes dont l'altitude moyenne est (environ) supérieure à deux cents mètres, tout en restant inférieure à trois cent mètres. On retrouve ainsi la majeure part de la vallée du Rhône et de la Saône, ainsi que les régions de l'Alsace et de la Lorraine, à l'exception de la partie vosgienne. La dernière classe (Q4) est évidemment composée des communes situées aux plus hautes altitudes, dans notre cas il s'agit des communes dont l'altitude moyenne est supérieure à trois cents mètres. L'éventail est donc important car cette classe concentre les 10 538 communes les plus hautes, dont l'altitude moyenne varie de 300 à 2 700 mètres. Un rapide coup d'œil sur la carte 2.5a suffit pour remarquer que la majorité des

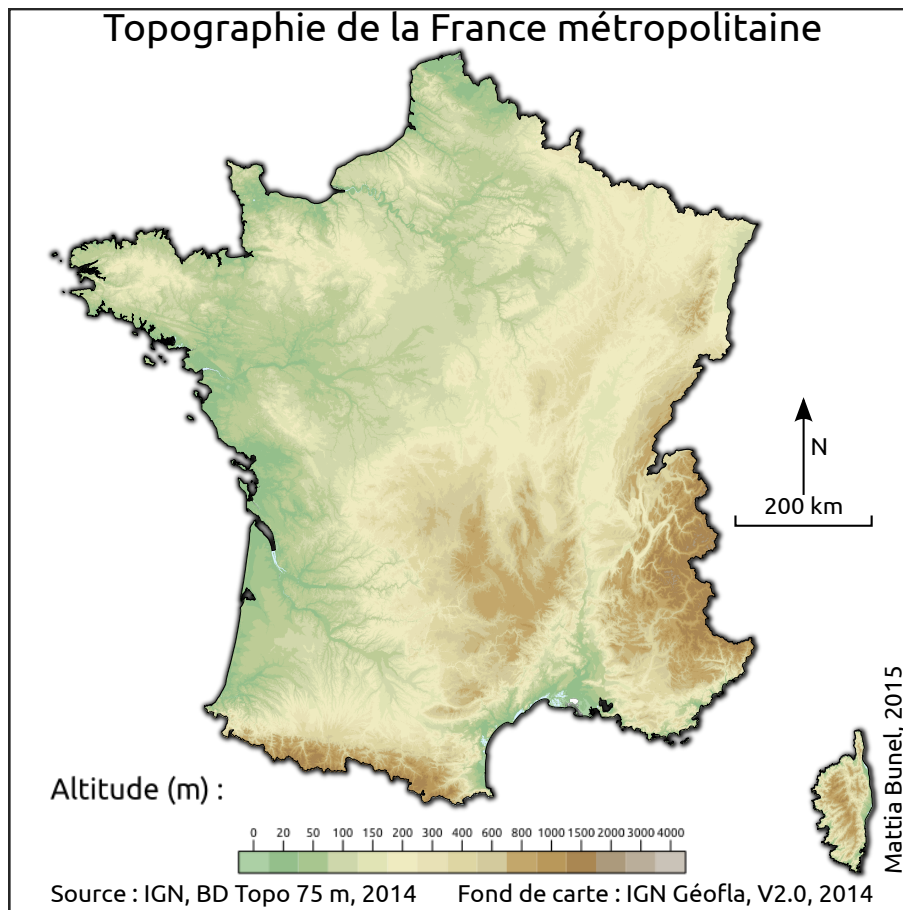


FIGURE 2.4 – Topographie de la France métropolitaine.

ces communes sont localisées dans les massifs montagneux et leurs piémonts. On retrouve notamment le massif pyrénéen, le massif central, ou encore une importante structure continue composée des Alpes, du Jura et des Vosges. Il est évident que le présent commentaire est fortement similaire à un simple commentaire de la France physique, mais cet état de fait est simplement dû au fait que nous n'avons commenté qu'une seule et unique carte. Comme nous l'avons déjà signalé cette méthode prend tout son sens lorsque qu'elle est utilisée à des fins comparatives. Les structures observées sont parfaitement logiques et se retrouvent notamment sur la carte 2.4, mais c'est en comparant l'évolution géographique de ces aires que nous pourrions approfondir le sujet.

Entre 1793 et 1851 les changements sont minimes. Si l'on se réfère au tableau 2.2 on remarque que la première classe a vu sa superficie diminuer de deux points, celle de la seconde classe à légèrement augmentée, celle de la troisième classe à faiblement diminué, quant à celle de la dernière classe elle n'a quasiment pas évolué. Et cette absence d'évolution se retrouve également sur la carte 2.5b, les structures précédemment décrites restent valables, et les changements de classe ne concernent que certaines communes localisées. C'est en comparant les cartes 2.5c et 2.5d que l'on commence à remarquer de nombreux changements. Les changements les plus éloquentes sont ceux concernant la première classe (Q1), dont la part de la superficie atteint une de ces plus faibles valeurs. Désormais une grande partie des communes précédemment classées dans la première classe (Q1) appartiennent à la seconde classe, c'est notamment le cas d'une grande partie des communes *sublittorales* bretonnes, mais également des communes littorales normandes. L'intégralité des communes appartenant précédemment à la première classe sont désormais incluses au sein de la seconde classe, qui tend à se « rapprocher » du littoral, et dont la part de la superficie à légèrement augmenté. Les mutations des classes Q3 et Q4 sont moins visibles, mais deviennent véritablement importantes entre 1901 et 1946.

En 1946, la superficie de la première classe plafonne à son minimum (et ce depuis 1906). Comme précédemment cette dernière ne contient pratiquement que des communes littorales et sublittorales, mais également à certaines communes appartenant à des vallées fluviales. On note également que, comme en 1793, le département du nord appartient également à la première classe (Q1). La surface de la seconde classe a également diminué, les communes appartenant précédemment à cette classe se retrouvent incluses dans la classe Q3, c'est notamment le cas des communes de la Beauce, ou encore d'une partie des communes de la vallée du Rhin. Mais la classe qui a subi la plus grande évolution de sa superficie est sans conteste la

classe Q4 qui occupe désormais 43 % de la superficie totale. De façon logique cette dernière s'étend sur les communes appartenant auparavant à la classe Q3. On est désormais en présence d'une structure continue reliant tous les grands massifs entre eux, à l'exception des Pyrénées. Seules les vallées du Rhône et de la Saône limitent l'ampleur du contact direct entre le massif alpin et les Cévennes. Mais si l'on compare l'évolution de cette vallée depuis la première carte on note que la part des communes appartenant à la classe Q3 se réduit grandement, de plus en plus de communes appartenant à la classe Q4.

Comme pour la carte de 1946, la principale évolution visible sur les cartes de 1975 et de 1999 et l'importante augmentation de l'aire occupée par la classe Q4 (qui atteint son maximum en 1999). Le quart de la population est désormais concentré dans la moitié la plus élevée du territoire national. Les classes Q2 et surtout Q1 conservent les structures précédemment décrites. Même si le tableau 2.2 indique une diminution de la superficie occupée par la classe Q2 et ce au profit de la classe Q3. L'étude comparée de ces six cartes nous donne plus d'informations sur la stagnation de la part de la superficie occupée par la classe Q3, notamment visible à partir du tableau 2.2. On voit clairement que cette classe est encadrée par une classe qui tend à perdre en superficie (Q2) et une autre qui tend à en gagner (Q4). Ainsi la classe Q3 gagne en superficie en englobant les communes appartenant auparavant à la classe Q2, mais en perd au profit de la classe Q4, en résulte une stagnation de la superficie de la classe, mais une évolution de son inscription spatiale.

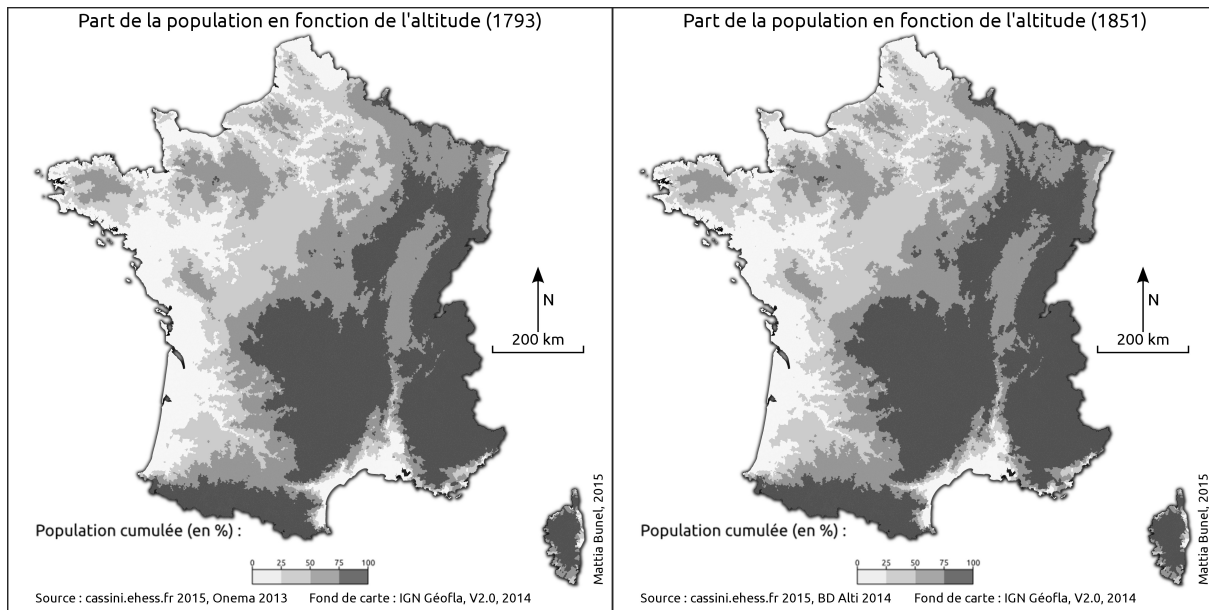
Distance à la mer

La planche 2.6 est également composée de six cartes. Les dates retenues sont le début de la série statistique (1793), sa fin (1999) ainsi que les recensements de 1851, 1901, 1946 et 1975. Au vu des résultats précédemment obtenus, nous avons pris la décision de commenter les cartes des déciles, et non celles des quartiles (toutefois présentes dans les annexes).

Les cartes de la planche 2.6 présentent une structure bien plus géométrique que les cartes précédemment commentées, quelque soit la date, les classes s'organisent par « bandes » suivant un gradient littoral intérieur des terres induit par la nature même du calcul. On remarque qu'il existe deux catégories de classes, celles qui sont continues et qui ne forment qu'un *bloc* et celles discontinues, composées de plusieurs zones (généralement deux, voir trois en comptant la Corse) non reliées entre elles. La première catégorie correspond aux communes les plus éloignées du rivage, la seconde correspond aux communes les plus proches du littoral, qui se retrouvent subdivisées entre leur partie méditerranéenne et leur partie atlantique. Le nombre de classes appartenant à chaque catégorie ne varie pas, il reste de 6 pour la première et de 4 pour la seconde. Comme nous l'avons indiqué lors du commentaire du tableau 2.4 les principales transformations s'opèrent au niveau de la classe Q1. Cette dernière correspond à la bande littorale, et regroupe 10 % de la population totale quelque soit la date. Sa superficie, maximale en 1793, ne cesse de réduire jusqu'en 1975. Et cette réduction est parfaitement visible sur les cartes 2.6, contrairement à sa légère augmentation de superficie entre 1975 et 1999.

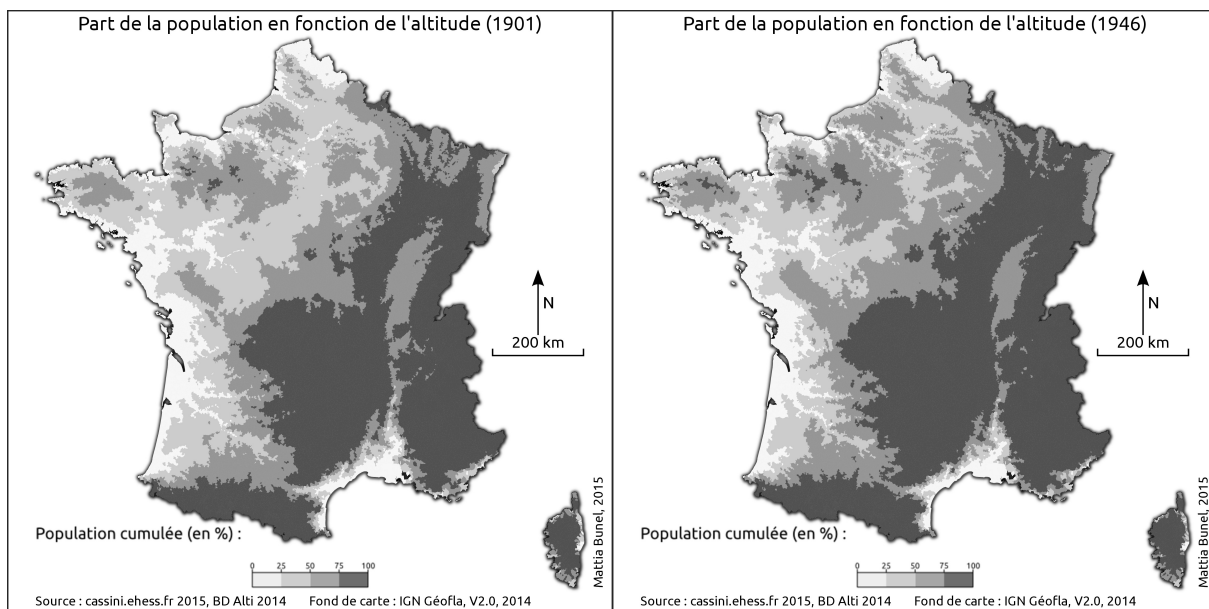
En 1793, la classe Q1 correspond à une bande littorale large de plusieurs dizaines de kilomètres et intégrant des communes assez éloignées du littoral, notamment la commune d'Aix-en-Provence. Mais cette profondeur diminue assez rapidement et dès 1851 on remarque un rétrécissement de cette bande littorale. Ce dernier continue et à partir de 1901, un certain nombre de communes contiguës au littoral (notamment la commune d'Arles) se retrouvent relégués dans la seconde classe. Le cas de la Corse est intéressant, car il rend, par le faible nombre de classes, le phénomène particulièrement lisible. La superficie occupée par le premier décile se réduit et la seconde classe « absorbe » en son sein les communes précédemment rattachées à la première classe. Et cet état de fait se retrouve entre de nombreuses classes, c'est notamment le cas de la classe Q6 qui est peu à peu absorbée par la classe Q7.

Contrairement à au cas précédent où toutes les classes présentaient un intérêt, l'étude de l'inscription spatiale de la population en fonction de la distance au littoral ne présente d'intérêt que pour les premières classes, et plus particulièrement la classe Q1. Bien que cette dernière ne soit pas la seule à voir sa superficie diminuer de manière importante, il s'agit de la seule classe pour laquelle nous pouvons clairement mettre cette évolution en rapport avec la distance au littoral. Si l'on se penche sur le cas de la classe Q6, on peut clairement expliquer la diminution de sa superficie par la forte densification de Paris, elle-même induite par une transformation globale de la distribution de la population (*métropolisation*) qui s'explique par des causes sociales et technologiques (*industrialisation* et *exode rural*) et non environnementales [79, 31, 32, 80]. Cette réduction de superficie est donc un *artefact*, et un potentiel risque d'interprétation erronée. Contrairement à la classe Q1, dont la réduction est l'indication d'une concentration de la population sur la frange littorale. Ainsi contrairement aux précédentes cartes qui mettaient en évidence une modification globale de la répartition de la population, ces nouvelles cartes nous apportent moins d'éléments et laissent supposer que ce critère est moins porteur de sens.



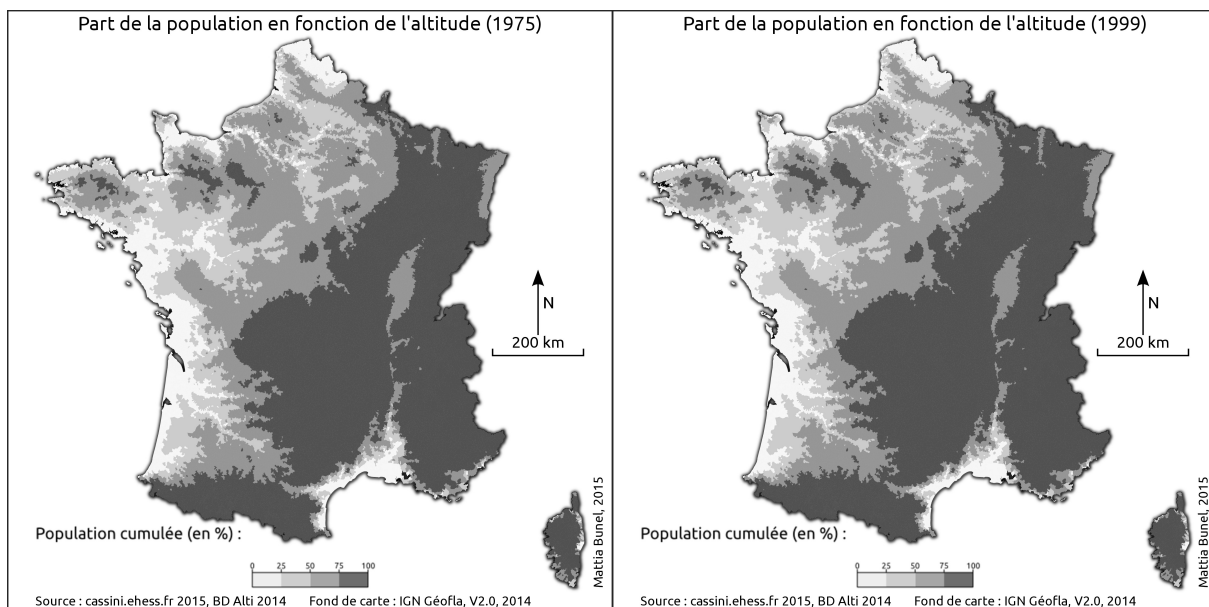
(a) Écart à la densité moyenne en 1793.

(b) Écart à la densité moyenne en 1901.



(c) Écart à la densité moyenne en 1936.

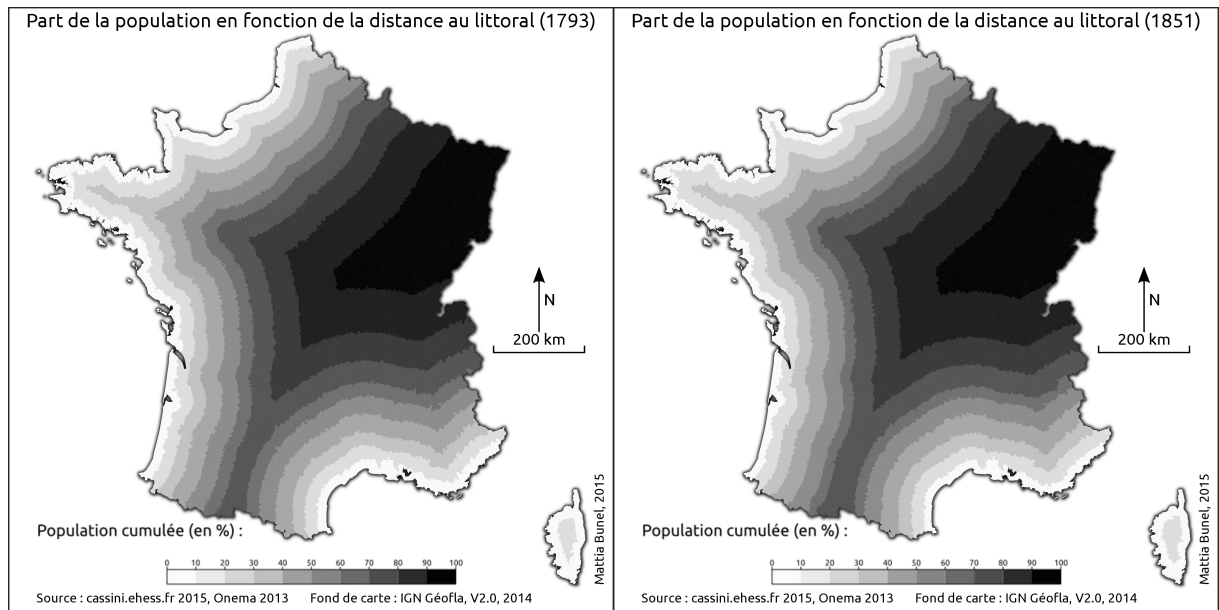
(d) Écart à la densité moyenne en 1946.



(e) Écart à la densité moyenne en 1975.

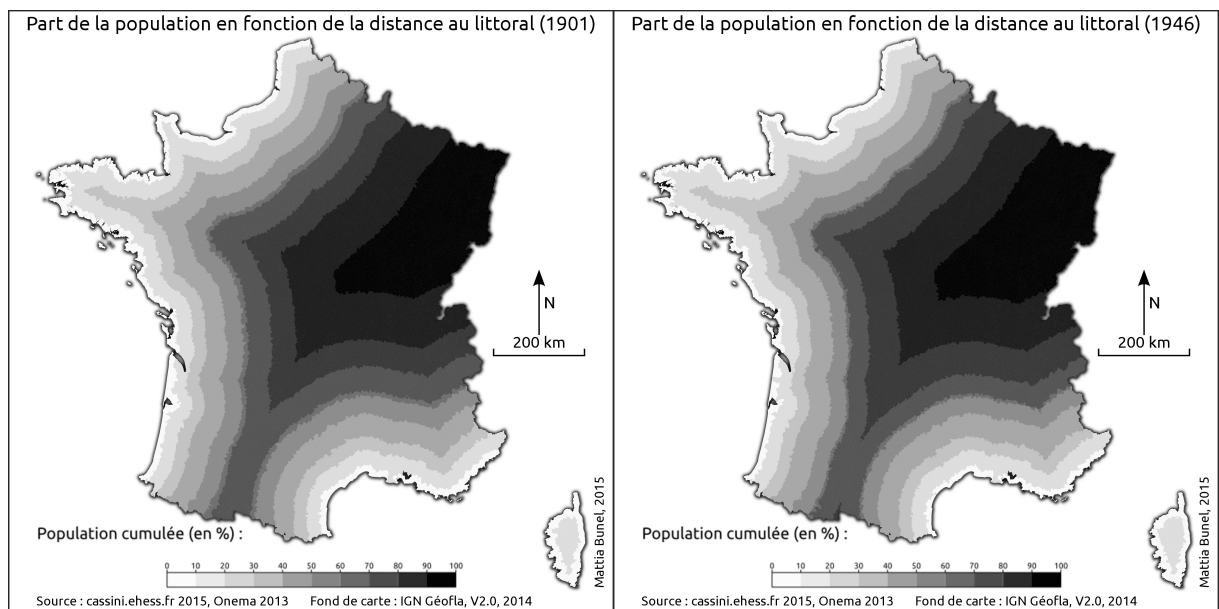
(f) Écart à la densité moyenne en 1999.

FIGURE 2.5 – Population cumulée en fonction de l'altitude.



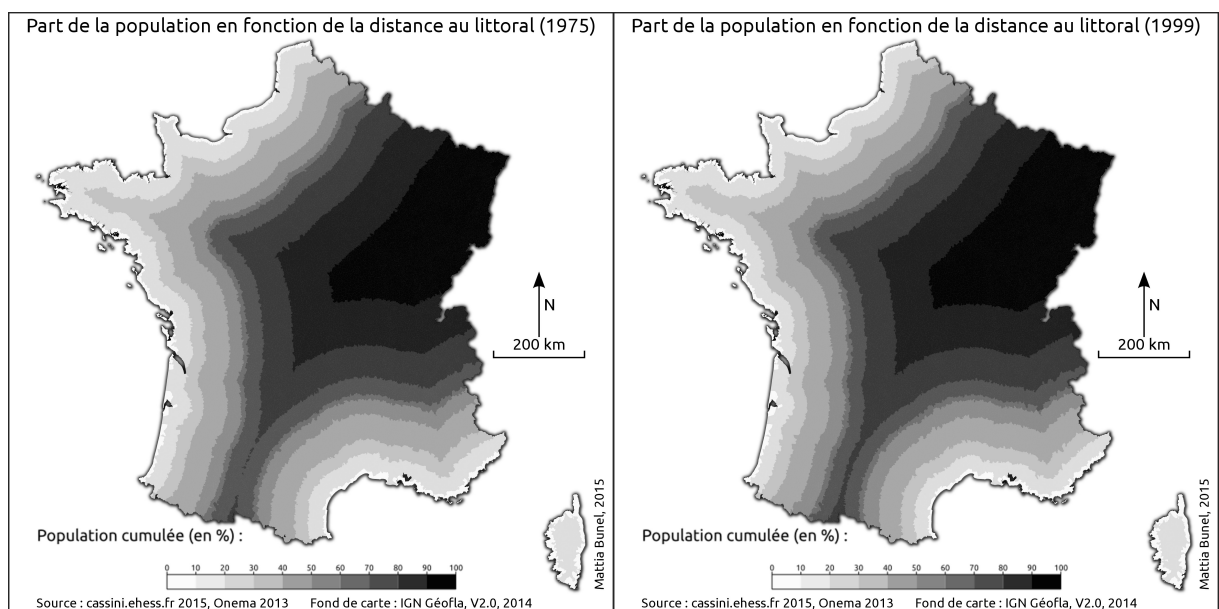
(a) Écart à la densité moyenne en 1793.

(b) Écart à la densité moyenne en 1901.



(c) Écart à la densité moyenne en 1936.

(d) Écart à la densité moyenne en 1946.



(e) Écart à la densité moyenne en 1975.

(f) Écart à la densité moyenne en 1999.

FIGURE 2.6 – Population cumulée en fonction de la distance à la mer.

L'objectif de cette partie était de nous questionner sur l'existence d'un lien entre la transformation de la distribution de la population mise en évidence dans le premier chapitre et différentes composantes environnementales influant la distribution de la population, l'altitude et la distance au littoral [6, 65, 76]. Nous savions au vu des résultats présentés dans la précédente partie, l'importance de l'autocorrélation temporelle dans ces transformations et par conséquent le rôle limité des différentes composantes environnementales. Avant de recourir à de puissantes méthodes quantitatives nous avons voulu explorer la question à l'aide d'une méthode plus simple. En étudiant la part de la superficie occupée par des quantiles de la population totale ordonnés en fonction d'un des deux critères étudiés nous avons pu mettre en évidence différents phénomènes liés à la concentration de la population sur la période étudiée.

Dans un premier temps, nous avons constaté que la population se concentrait dans les communes les plus basses et ce au prix d'une perte d'importance relative des communes situées aux plus hautes altitudes. Là où les 50 % de la population situés aux plus faibles altitudes occupent 40 % du territoire total en 1793, cette même proportion n'occupe que 25 % du territoire en 1999. Rapporté à l'accroissement quasi-constant de la population française sur cette période, il en découle une importante augmentation de la densité des communes situées aux plus faibles altitudes. Inversement, le nombre de communes incluses dans la dernière classe ne cesse d'augmenter, puisque cette dernière occupe quasiment 50 % du territoire total en 1999 pour 25 % de sa population. L'évolution observée est assez claire, les communes situées aux plus faibles altitudes prennent de plus en plus d'importance, et à l'inverse les communes situées aux plus hautes altitudes perdent en importance. Au vu de ces résultats il nous paraît indispensable d'étudier plus en détail l'évolution de la distribution de la population en fonction de l'altitude dans l'optique de chercher à savoir si l'évolution de la distribution de la population en fonction de la distance au littoral est bien différente. Comme nous l'avons déjà indiqué une seule des dix classes étudiées subit une réelle transformation. La diminution de la superficie de la première classe illustre une tendance à la concentration de la population sur le littoral. Mais il est intéressant de remarquer que les mêmes cartes réalisées pour la variable *altitude* permettaient de réaliser les mêmes conclusions. En effet les classes Q1 de la première et de la seconde série concernent (en partie) les mêmes communes, l'altitude moyenne étant corrélée à la distance à la mer. Ainsi la diminution de la superficie de la classe Q1 (altitude) indique également une concentration de la population sur les littoraux. Les résultats des deux séries de cartes se recoupent. Toutefois la première série est plus riche en informations. Là où l'étude de la distribution de la population en fonction de la distance au littoral nous a uniquement permis de mettre en évidence une tendance à la concentration de la population sur le littoral, la même étude basée sur l'altitude nous a permis de confirmer cette observation, tout en nous indiquant une l'existence d'une importante diminution de la superficie occupée par la quatrième classe et par conséquent de la perte d'importance des communes montagnardes, vraisemblablement causée par le dépeuplement des montagnes au cours du XIX^e siècle [8, 31, 32].

Les conclusions de cette partie ne font que nous donner des pistes et non pas des éléments de démonstration. Ces derniers seront apportés dans la partie suivante, portant sur l'application d'une méthodologie quantitative aux questions soulevés par notre travail. Nous chercherons notamment à explorer plus en profondeur le lien entre environnement et évolution de la distribution de la population à l'aide de différentes méthodes quantitatives.

2.3 Approche quantitative

Dans cette troisième partie nous approfondirons les résultats obtenus précédemment à l'aide d'un ensemble de méthodes quantitatives. Notre première approche constituera en la classique application d'un modèle de régression linéaire et sera suivie par une analyse de la variance et une nouvelle classification, faisant écho à celle réalisée précédemment.

2.3.1 Régression linéaire

Nous pouvons estimer et quantifier le lien entre variables environnementales et distribution de la population à l'aide d'une méthode statistique régulièrement utilisée, la *régression linéaire*. Cette méthode a précédemment été utilisée pour quantifier l'autocorrélation temporelle et spatiale. Elle consiste en la réalisation d'un modèle de la forme :

$$Y = \beta_0 + \beta_1 X + \epsilon \quad (2.1)$$

Où Y est la variable dépendante, X la variable explicative, β_0 et β_1 les paramètres du modèle et ϵ les résidus du modèle [21]. Le modèle ici présenté correspond à une régression simple (une seule variable explicative), mais peut facilement être généralisé avec l'ajout de nouvelles variables explicatives, également dites indépendantes. L'objectif de ces modèles est multiples, nous cherchons à la fois à vérifier l'existence d'un lien entre distribution de la population et certaines variables environnementales, puis à quantifier

ce lien et enfin à confronter les valeurs du modèle aux valeurs réelles pour mettre en exergue les points de divergence.

Le premier modèle réalisé a pour variable dépendante *l'écart à la densité moyenne* à la date t et pour variable explicative l'altitude communale moyenne. Une fois ce modèle type réalisé, nous l'avons reproduit pour chaque recensement et en avons tiré les figures 2.7a et 2.7b. Nous avons pris la décision de ne pas afficher ici les résultats numériques des différents modèles réalisés et ce pour des questions de lisibilité. Mais Quelque soit le recensement sélectionné, la variable explicative (altitude communale moyenne) est systématiquement significative au seuil de 0,001 %. Toutefois les coefficients de détermination obtenus restent faibles, la part de la variance expliquée reste inférieure à 15 %. Mais si l'on se réfère à la figure 2.7a nous constatons que la valeur du coefficient de détermination tend à augmenter au cours du temps, par conséquent le modèle se révèle de plus en plus pertinent. De manière similaire, la valeur du *coefficient de la variable altitude* diminue avec l'augmentation du coefficient de détermination (cf. figure 2.7b). D'un point de vue géographique cela signifie que l'écart à la densité décroît de manière de plus en plus importante à mesure que l'altitude augmente. Ces modèles mettent ainsi en évidence l'augmentation de la répulsivité des montagnes.

L'analyse de ces modèles est complétée par les cartes 2.8, 2.9 et 2.10. Ces dernières permettent la juxtaposition des cartes des valeurs estimées par le modèle, des résidus et des valeurs réelles et ce pour six dates que nous avons jugées pertinentes. Quelle que soit la planche, la première carte est toujours celle des valeurs estimées par le modèle. Nous remarquons qu'elle sont quasiment toutes identiques, ce qui s'explique par l'invariabilité de la variable explicative (altitude). L'écart à la densité moyenne estimée par le modèle laisse apparaître une structure en gradient. L'écart à la densité moyenne devient de plus en plus important (dans le négatif) avec l'augmentation de l'altitude. Ainsi les zones les moins denses correspondent aux principaux massifs montagneux (Alpes, Massif Central. . .) et les zones les plus denses aux plus faibles altitudes (principalement les littoraux). Si nous confrontons ces cartes à celles de *l'écart à la densité moyenne* aux mêmes dates nous remarquons que les deux cartes ne correspondent absolument pas. Certes ce gradient existe, mais il n'occupe qu'une part minime dans la structure spatiale de l'écart à la densité moyenne, part qui est par ailleurs quantifiée par le coefficient de détermination. Les cartes des résidus permettent d'observer la distribution spatiale des imprécisions du modèle. Malgré la grande étendue temporelle illustrée par ces six modèles, les résidus concernent systématiquement deux types de communes, dont le nombre et la localisation varie dans le temps. On peut distinguer deux catégories de résidus, les premiers sont les résidus négatifs, il s'agit de communes pour lesquelles l'écart à la densité moyenne estimée par le modèle est supérieure à celle estimée par le modèle. L'exemple de l'Aquitaine est parfaitement représentatif de cette catégorie, quel que soit le recensement étudié, les communes de cette région apparaissent comme des résidus négatifs. La raison est assez simple, nous avons vu sur les cartes de *l'écart à la densité moyenne* que cette région a toujours été caractérisée par une densité inférieure à la densité nationale, ce qui conduit par ailleurs la majorité des communes à appartenir à la cinquième classe de la typologie réalisée à la fin du premier chapitre. Or, comme nous l'avons précédemment indiqué, le modèle estime que l'écart à la densité moyenne diminue avec l'augmentation de l'altitude, réciproquement, les communes littorales, par corollaire de faible altitude auraient une densité élevée et par conséquent à écart à la densité moyenne positif. Cette contradiction résulte par l'apparition de résidus négatifs, qui s'inscrivent dans la durée. Cette situation est également observable pour les communes de la diagonale du vide et plus généralement pour l'ensemble des communes rurales (à partir de 1946). Ici la structure n'est pas permanente, mais au contraire évolue. Comme précédemment il s'agit de communes dont l'écart à la densité moyenne est inférieur à ce que laisserait supposer leur altitude. Au vu des indications précédemment énoncées, notamment dans le premier chapitre l'accroissement du nombre de communes rurales appartenant à la catégorie des résidus négatifs est explicable par le phénomène *d'exode rural* et correspond également à nos précédentes observations, concluant à une tendance à la concentration de la population. De manière similaire on trouve également des résidus positifs, *i.e.* des communes dont *l'écart à la densité moyenne* est supérieur à ce que ne le laisserait supposer leur altitude. Il est très facile, au vu des cartes des résidus, d'identifier les communes concernées. Il s'agit bien entendu des principales agglomérations. De manière analogue à ce qui avait été annoncé précédemment, on peut voir les phénomènes de *métropolisation* et de concentration de la population à travers la carte des résidus. Ces cartes nous informent sur les structures non expliquées par l'altitude. Et l'on trouve notamment les grandes agglomérations dont la densité ne dépend pas (ou peu) de la composante altimétrique. Si l'on se réfère à la précédente partie et notamment aux écrits de BAUELLE [6], mais également au premier chapitre il est logique que les agglomérations apparaissent comme des résidus positifs, leur densité étant avant tout expliquée par la composante historique, et plus particulièrement par de nombreuses transitions historiques, telles que *l'industrialisation* et la *métropolisation*, en ce qui concerne le XIX^e siècle. D'une certaine façon, il est tout à fait pertinent de considérer que les cartes 2.8c, 2.8d, 2.9c, 2.9d, 2.10c, et 2.10d ne sont que la représentation des écarts à la densité moyenne après la « *neutralisation* » de la composante altimétrique. Ainsi ces cartes correspondent aux structures que nous pourrions observer dans le cas où le

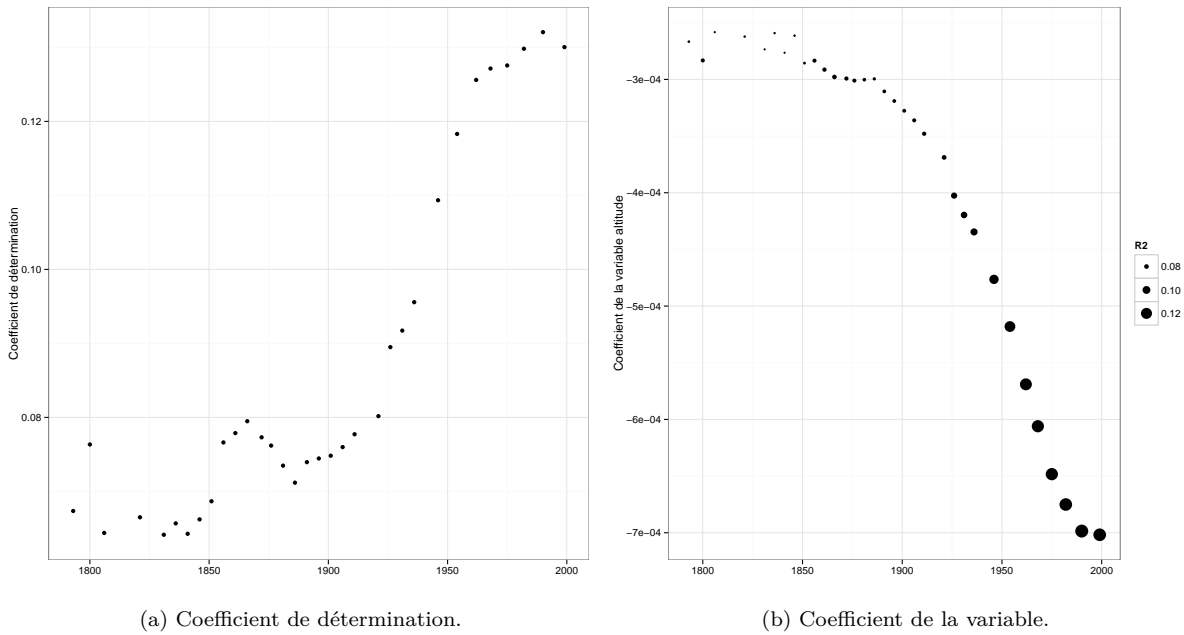
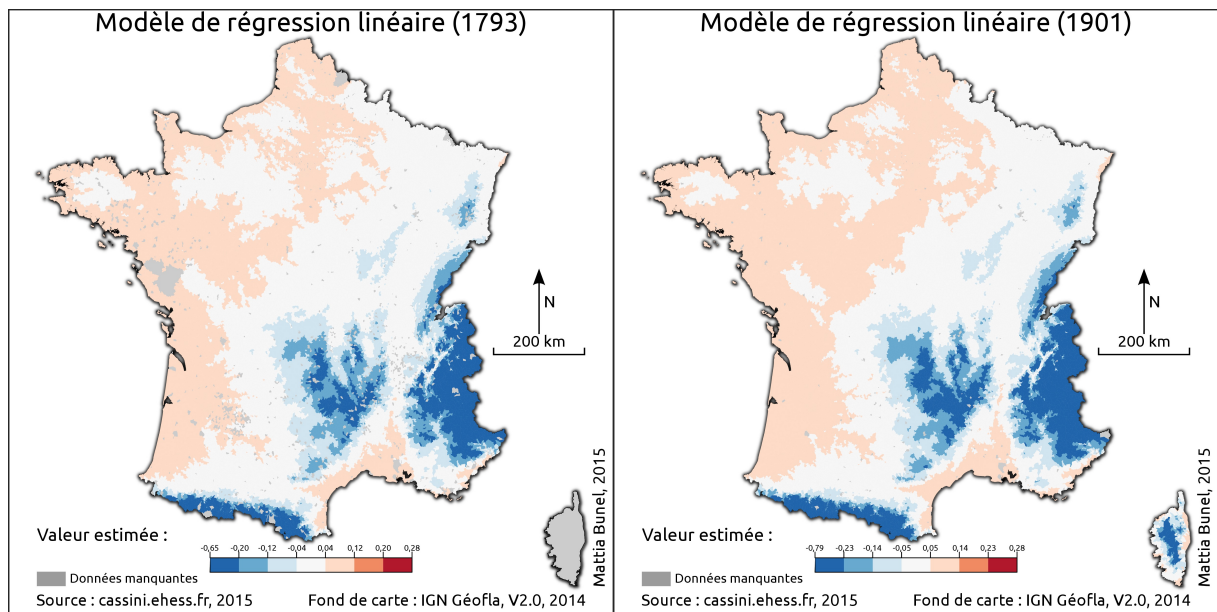
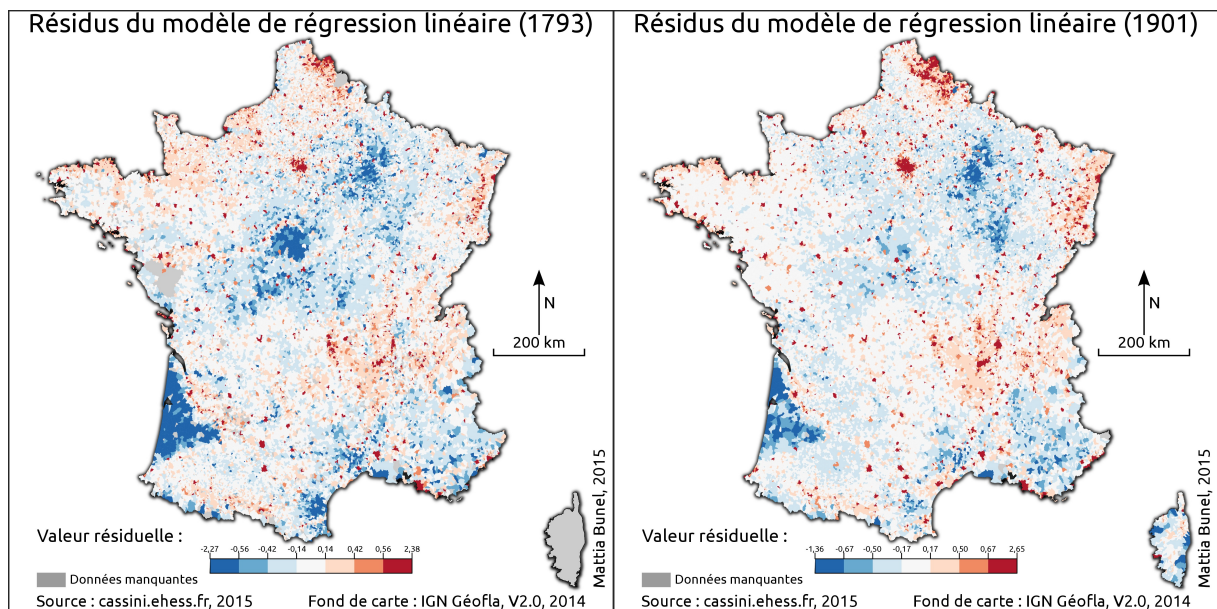


FIGURE 2.7 – Modèles de régressions, écart à la densité moyenne en fonction de l'altitude.

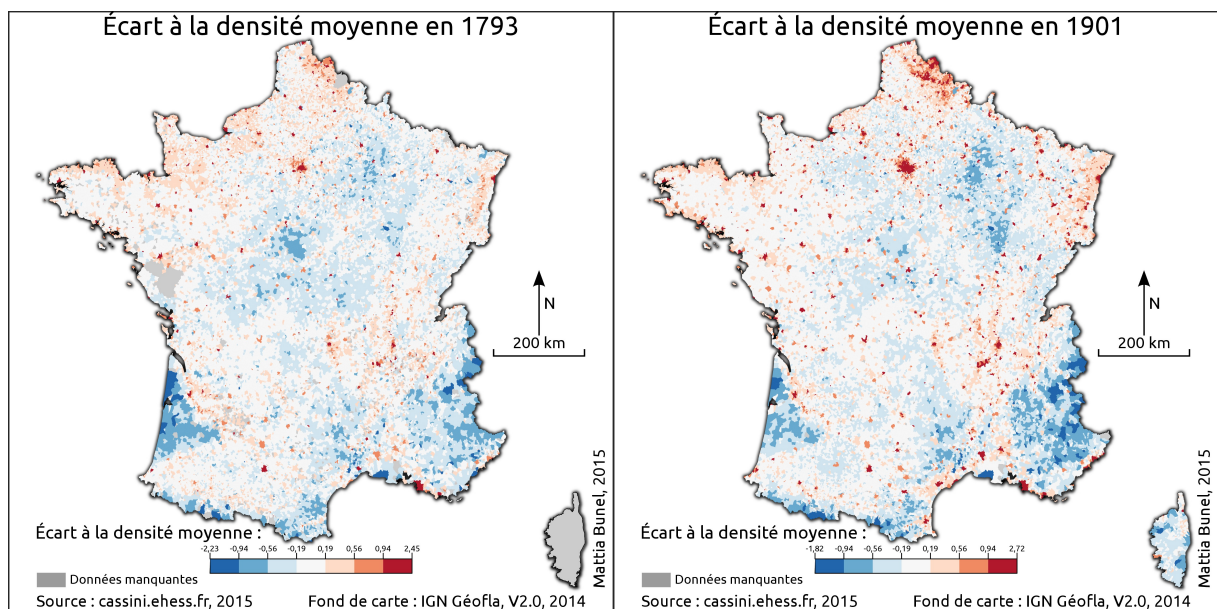
territoire serait d'altitude invariable. Et malgré la « *suppression* » d'une composante environnementale importante dans la structuration de la distribution de la population [6, 65, 76], nous retrouvons en grande partie (sur les cartes des résidus) les structures précédemment décrites, et leur évolution. Ainsi ces cartes démontrent l'importance de la composante historique, qui demeure la seule capable d'expliquer les fortes (et une partie des faibles) densités.

La seconde série de modèles réalisés a pour variable explicative la distance au littoral et pour variable à expliquer *l'écart à la densité moyenne* à la date t . Les productions graphiques réalisées pour ces modèles sont identiques à celles présentées précédemment. Le premier constat que nous pouvons faire en comparant le graphique 2.11 et son équivalent pour les précédents modèles, le graphique 2.7 est que les coefficients de détermination obtenus pour les modèles de la seconde série, sont globalement plus faibles. Si l'on omet le dénombrement de 1793, dont la valeur du coefficient de détermination est particulièrement importante, et donc probablement biaisée, c'est lors du dénombrement de 1946 que l'on atteint la valeur maximale du coefficient de détermination, qui est alors de 0,01, *i.e.* que seul 1 % de la variance totale est expliqué par le modèle, là où précédemment les meilleurs modèles avaient un coefficient de détermination dix fois supérieur. Ces nouveaux modèles ont un *coefficient de détermination* si faible qu'ils s'avèrent complètement inefficaces. Mais nous avons tout de même pris la décision de présenter six des trente-trois cartes de résidus. Faible coefficient de détermination oblige, les cartes des résidus sont quasiment identiques aux cartes de *l'écart à la densité moyenne* précédemment réalisées. En effet un coefficient de détermination de 1 %, indique que les 99 % restants de l'information initiale se retrouvent dans les résidus, par conséquent fortement similaires aux données initiales. Les conclusions précédemment réalisées laissent supposer l'obtention de résultats comparables à ceux que nous avons obtenus. Ces derniers nous montrent que l'écart à la densité moyenne pour les communes de la France métropolitaine à une date t , n'est pas explicable par la distance au littoral. Mais cette observation semble incompatible avec les théories présentées dans différents ouvrages, et décrites au début de ce chapitre. Nos résultats semblent contredire le fait que « *la géographie des bassins et des littoraux domine toutes les autres formes de peuplement* » [22]. La réponse à ce paradoxe a vraisemblablement été présentée dans la précédente partie, rappelons que nous avons montré qu'un découpage en quartiles, pertinent pour étudier l'évolution du peuplement en fonction de l'altitude, était totalement inadapté à la même étude basée sur la distance au littoral. C'est en recourant à l'échelle du décile que nous avons commencé à obtenir des résultats intéressants. C'est derniers nous ont montré que seules quelques classes subissaient une diminution notable de leur surface, notamment les classes regroupant les communes *littorales* et *sublittorales*. Pour ce qui est d'une grande partie des autres communes, la distance au littoral semblait n'avoir aucune importance. Or les modèles présentés portent sur l'ensemble des communes de la France métropolitaine, et intègrent par conséquent un grand nombre de communes susceptibles de biaiser le modèle. Ce qui est probablement le cas des grandes agglomérations non littorales. On peut par conséquent en déduire que cette approche n'est pas des plus optimales. Pour essayer d'améliorer ce modèle on pourrait traiter la distance au littoral comme une variable qualitative,

(a) Valeur estimée en 1793, $R^2 = 0,07$.(b) Valeur estimée en 1901, $R^2 = 0,08$.

(c) Résidus en 1793.

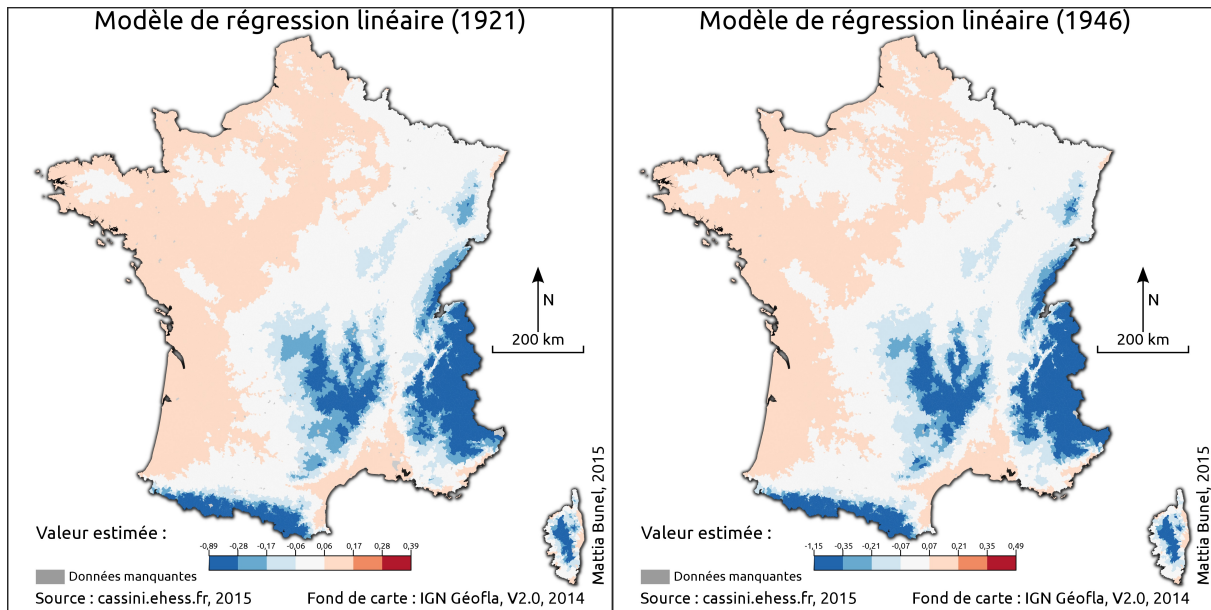
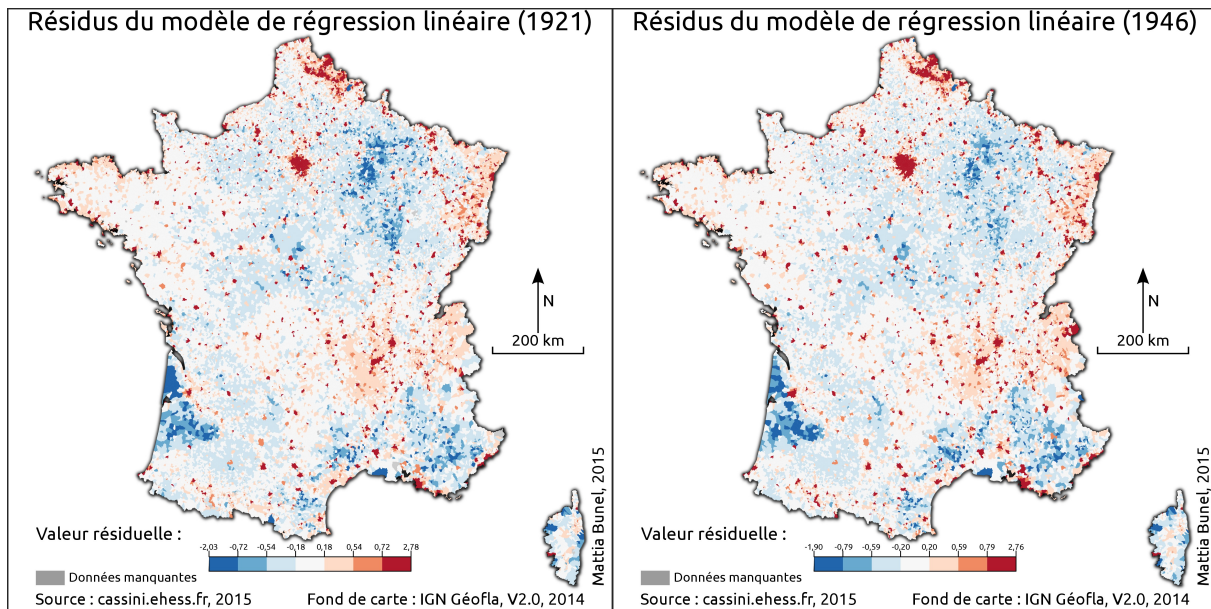
(d) Résidus en 1901.



(e) Écart à la densité moyenne en 1793.

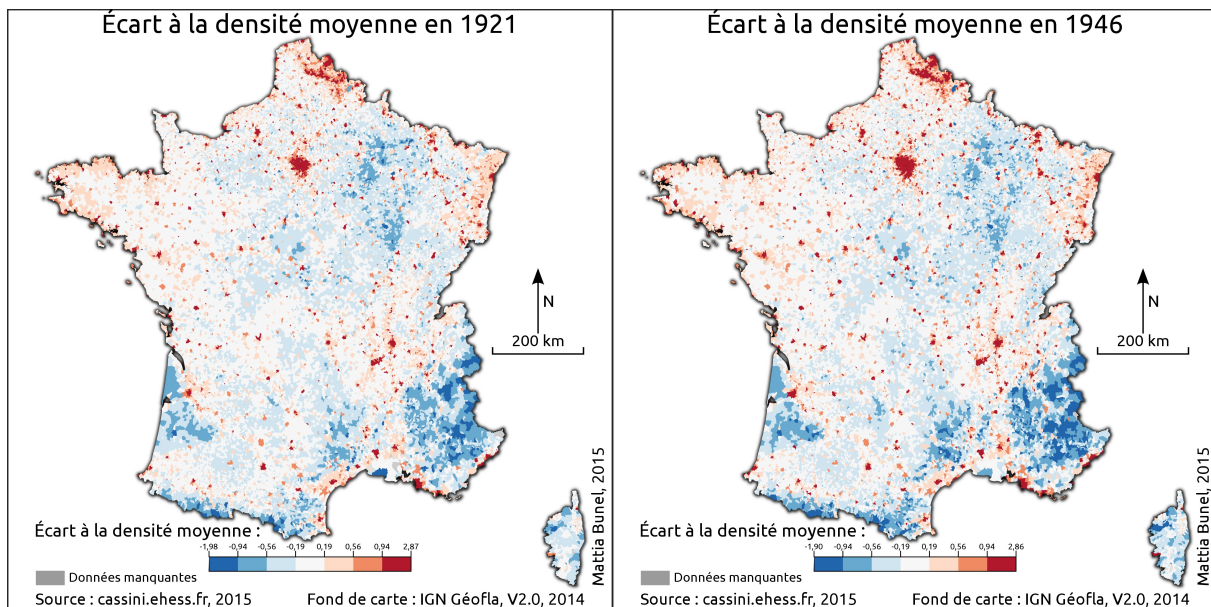
(f) Écart à la densité moyenne en 1901.

FIGURE 2.8 – Modèles de régressions, écart à la densité moyenne en fonction de l'altitude (1).

(a) Valeur estimée en 1921, $R^2 = 0,08$.(b) Valeur estimée en 1946, $R^2 = 0,11$.

(c) Résidus en 1921.

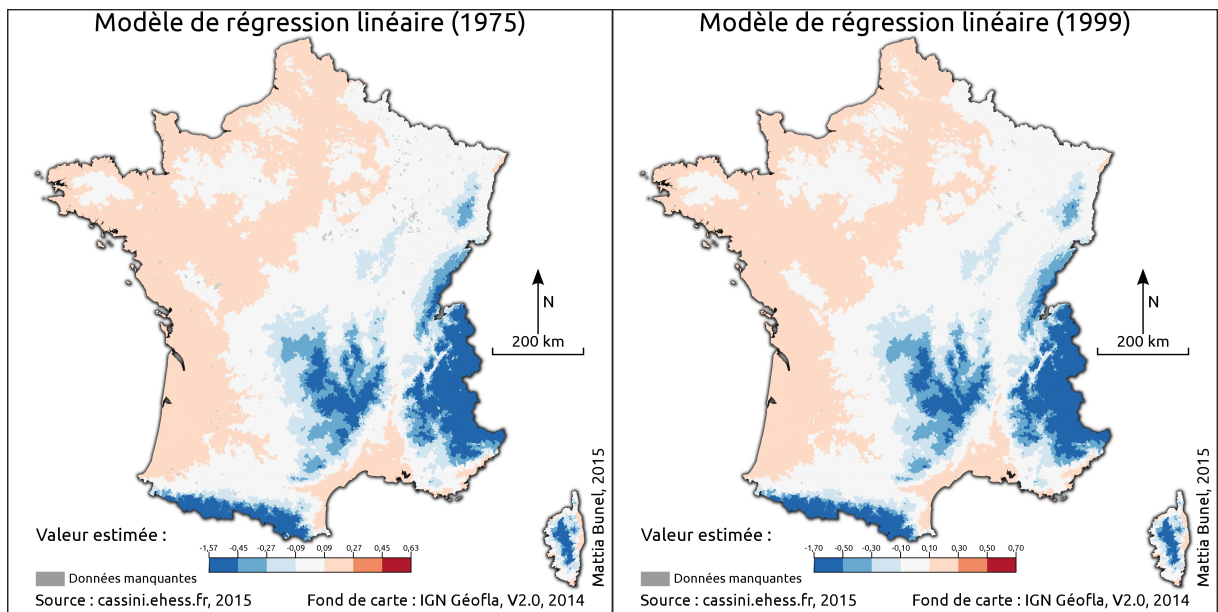
(d) Résidus en 1946.



(e) Écart à la densité moyenne en 1921.

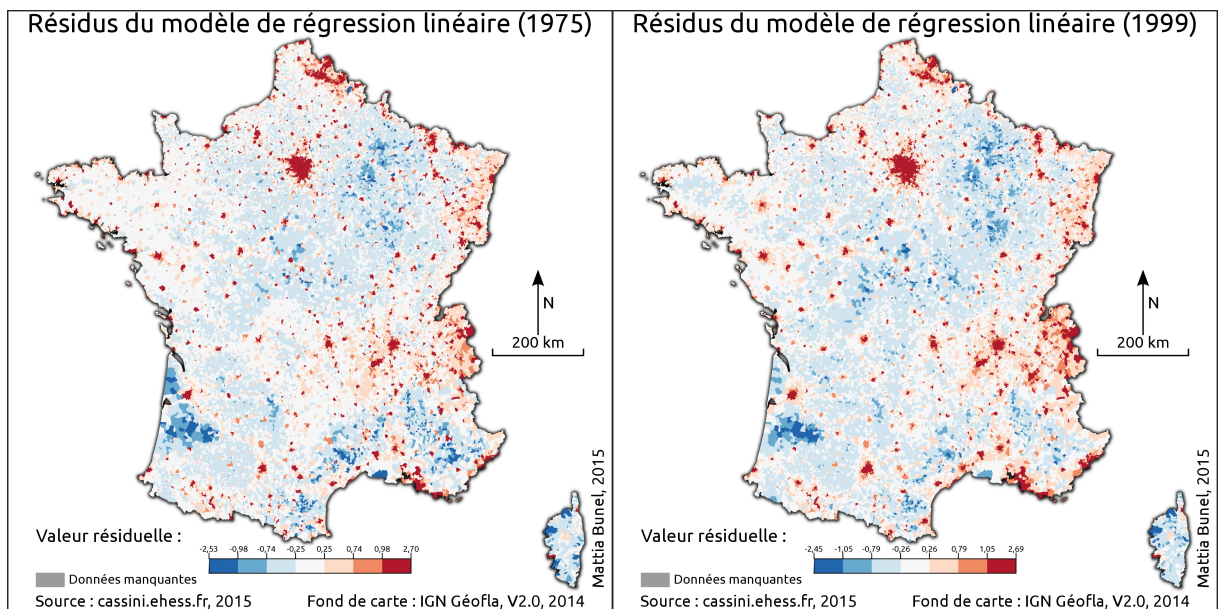
(f) Écart à la densité moyenne en . 1946.

FIGURE 2.9 – Modèles de régressions, écart à la densité moyenne en fonction de l'altitude (2).



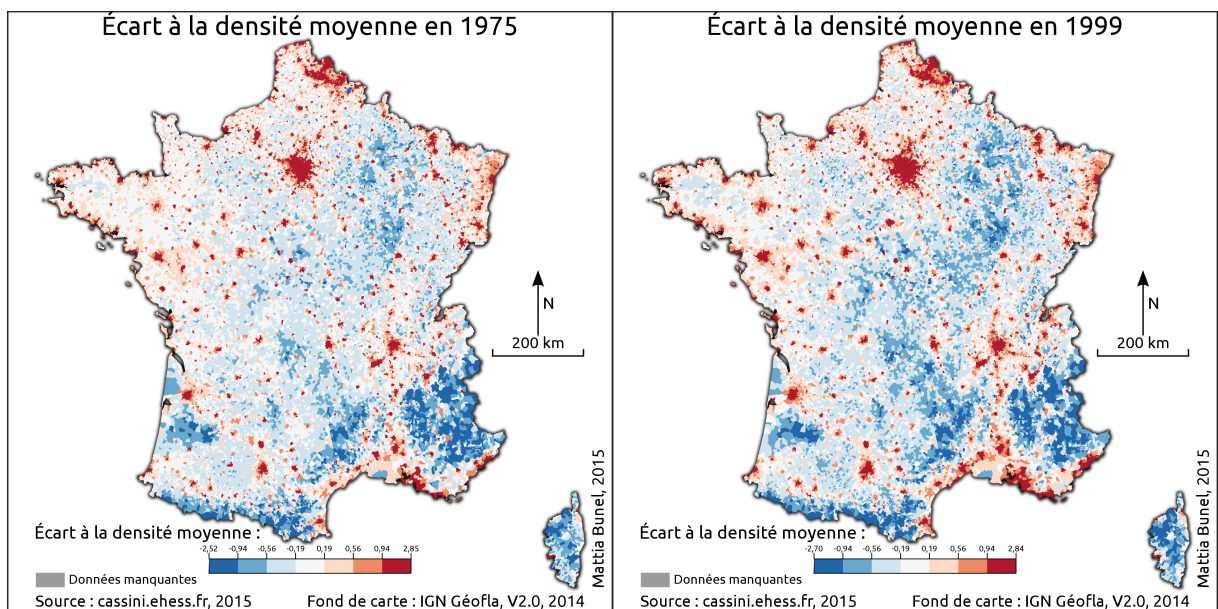
(a) Valeur estimée en 1975, $R^2 = 0,13$

(b) Valeur estimée en 1999, $R^2 = 0,13$.



(c) Résidus en 1975.

(d) Résidus en 1999.



(e) Écart à la densité moyenne en 1975.

(f) Écart à la densité moyenne en 1999.

FIGURE 2.10 – Modèles de régressions, écart à la densité moyenne en fonction de l'altitude (3).

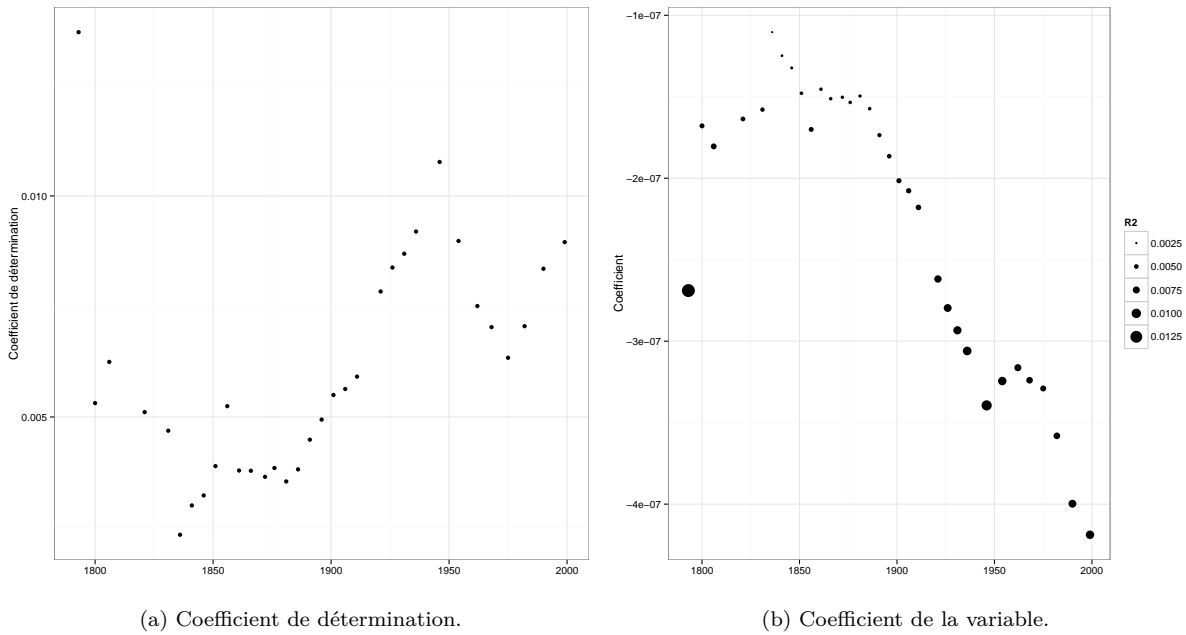


FIGURE 2.11 – Modèles de régressions, *écart à la densité moyenne* en fonction de la distance au littoral.

voire binaire et supposer que seules les communes littorales, ou situées à une certaine distance du littoral se voient attribuer une valeur, et toutes les communes ne validant pas cette condition se voient attribuer une seconde valeur. Le principal problème lié à cette méthode est la définition du seuil, il serait dès lors nécessaire de mettre en place une procédure permettant de sélectionner le seuil le plus pertinent, une méthode similaire à la *cross-validation* [38] ou aux coefficient AIC et BIC [21] aurait pu être envisagée. Mais nous n'avons pas choisi cette option, et avons préféré garder cette variable et l'appliquer à de nouveaux modèles.

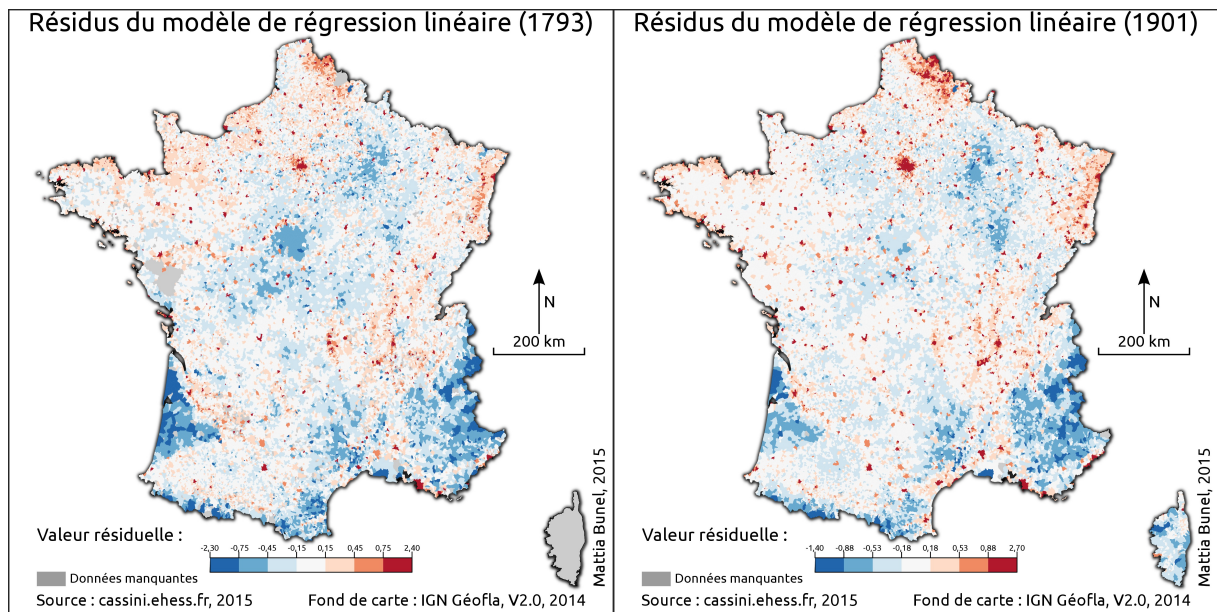
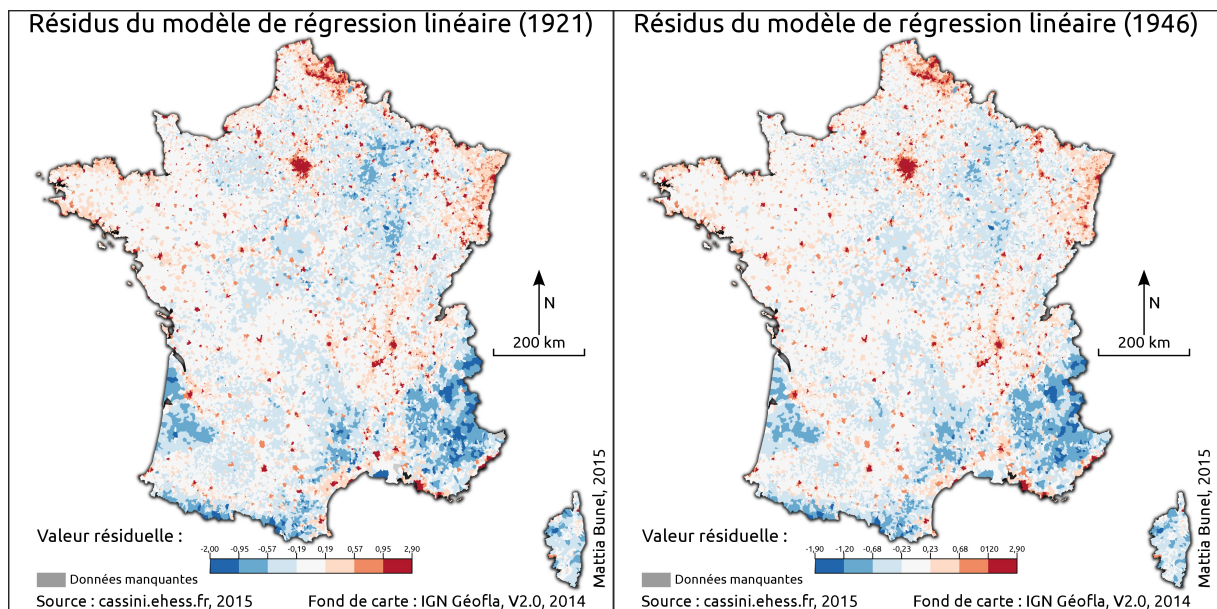
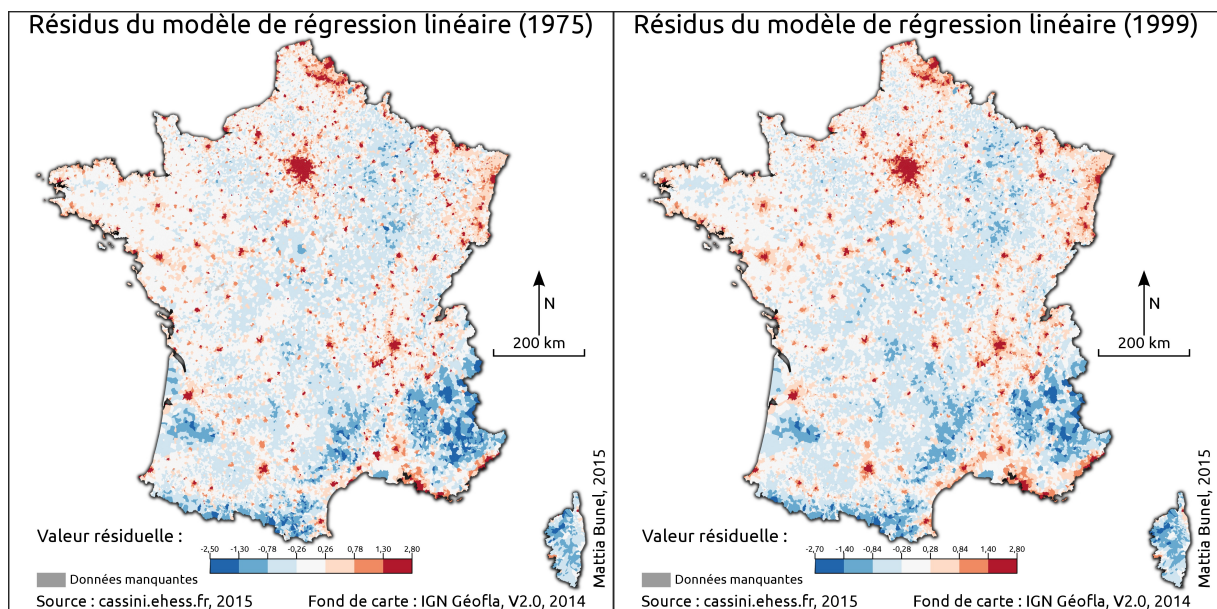
(a) Résidus en 1793, $R^2 = 0,014$.(b) Résidus en 1901, $R^2 = 0,005$.(c) Résidus en 1921, $R^2 = 0,006$.(d) Résidus en 1946, $R^2 = 0,010$.(e) Résidus en 1975, $R^2 = 0,006$.(f) Résidus en 1999, $R^2 = 0,009$.

FIGURE 2.12 – Modèles de régressions, écart à la densité moyenne en fonction de la distance au littoral.

La réalisation de ces différents modèles de régression nous a permis de montrer et de quantifier les liens existant entre distribution de la population. Quel que soit le modèle construit il y a corrélation entre l'écart à la *densité moyenne* et l'altitude ou la distance au littoral. Si dans le cas des modèles expliquant *l'écart à la densité moyenne* en fonction de la distance au littoral les *coefficients de détermination* obtenus sont très faibles, il n'en va pas de même pour les modèles expliquant *l'écart à la densité moyenne* en fonction de l'altitude. Certes les *coefficients de détermination* obtenus pour ces seconds modèles restent dans l'absolu assez faibles, avec une part de la variance expliquée comprise aux alentours de 10 %, mais rappelons que l'action des composantes environnementales est avant tout indirecte et que par conséquent de tels coefficients de déterminations restent honorables. Il est très intéressant d'analyser les différentes cartes des résidus précédemment présentées. Comme nous l'avons déjà indiqué, la structure des résidus présente d'importantes similitudes avec celle des grandes agglomérations mais également des grands vides, tels que la *diagonale du vide* ou l'Aquitaine. La présence de ces structures dans les résidus du modèle nous indique qu'elles ne peuvent être expliquées par leurs seules caractéristiques physiques, ce qui corrobore nos précédentes explications. Différents phénomènes démographiques et migratoires ont conduit à la concentration de la population dans les agglomérations (métropolisation) et le dépeuplement des régions rurales au profit de ces mêmes agglomérations (exode rural) et ces phénomènes n'ont probablement aucun lien avec les caractéristiques physiques des endroits. On peut toutefois se demander si il n'existerait pas un lien, non pas entre la distribution de la population et l'environnement, mais entre l'évolution de la distribution de la population et l'environnement. À titre d'exemple on peut se demander si les communes de montagne ont vu leur écart à la densité moyenne se réduire de manière plus importante que les autres communes rurales et c'est cette hypothèse que nous allons dès à présent tester .

2.3.2 Analyses de la variance

Comme nous l'avons vu précédemment il existe un lien entre distribution de la population et environnement. Ce dernier a pu être quantifié à l'aide des modèles précédemment présentés. Les coefficients de détermination obtenus restent assez faibles, mais la réalisation de ces modèles n'avait pour objectif que de quantifier l'importance de l'environnement dans la distribution de la population et non d'utiliser ces modèles dans un but prédictif.

Si l'on se réfère à l'affirmation de LE BRAS et LABBE [50] précédemment discutée, l'évolution de la distribution de la population touche différemment chaque commune. Ce point a déjà été abordé dans le précédent chapitre, et l'on sait, grâce à la typologie précédemment réalisée que l'on peut scinder les « *trajets* » des communes en cinq catégories bien différentes. Mais nous pouvons compléter cette analyse en nous posant de nouvelles questions. Lorsque BEAUJEU-GARNIER [8] parle de « *dépeuplement des montagnes* » un lien direct est fait entre évolution de la distribution de la population et environnement. Des communes partageant le même environnement suivent une évolution de leur population similaire. Par conséquent nous nous sommes demandés si les communes appartenant aux mêmes classes dans la typologie précédemment réalisée partageaient également une altitude et une distance au littoral similaires. L'existence d'une certaine cohésion des caractéristiques environnementales au sein d'une même classe permettrait de démontrer que l'évolution de l'écart à la densité de certaines communes est lié à leurs spécificités environnementales.

Pour répondre à cette question nous avons repris les résultats de la typologie du précédent chapitre. Chaque commune se voit attribuer une variable qualitative nominale correspondant à sa classe d'appartenance (définie à l'aide de la méthode des *k-means*, cf. chapitre 1 partie 3 sous partie 3). Comme nous reprenons les résultats de cette typologie nous héritons également des problèmes qui y sont liés, à savoir l'absence de certaines communes. Par conséquent cette analyse ne portera pas sur la totalité des communes (se référer à la partie précédemment citée pour plus de détails).

Notre méthodologie est basée sur l'emploi d'une analyse de la variance (Anova), méthode permettant de modéliser la relation entre une ou plusieurs variables qualitatives (dans notre cas la classe d'appartenance issue de la typologie précédemment réalisée que nous noterons A) et une variable à expliquer quantitative, notée Y [21, 74]. Nous n'allons faire appel qu'à une analyse à un facteur, où l'on compare les moyennes de Y pour chaque modalité I de A [21]. On construit donc un modèle de la forme suivante :

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij} \quad (2.2)$$

Avec y_{ij} la j^e valeur observée pour la sous-population i , ϵ_{ij} les résidus du modèle, α_i l'effet propre à la variable i et μ l'effet moyen [21].

Dans notre cas nous avons réalisé deux modèles, le premier a pour variable dépendante l'altitude communale moyenne, et le second la distance au littoral du centroïde de la commune. Avant de procéder à ces deux analyses de variance nous avons construit les *boxplots* 2.13a et 2.13b. Ces derniers décrivent respectivement la distribution de l'altitude et de la distance au littoral pour chaque classe issue de la typologie.

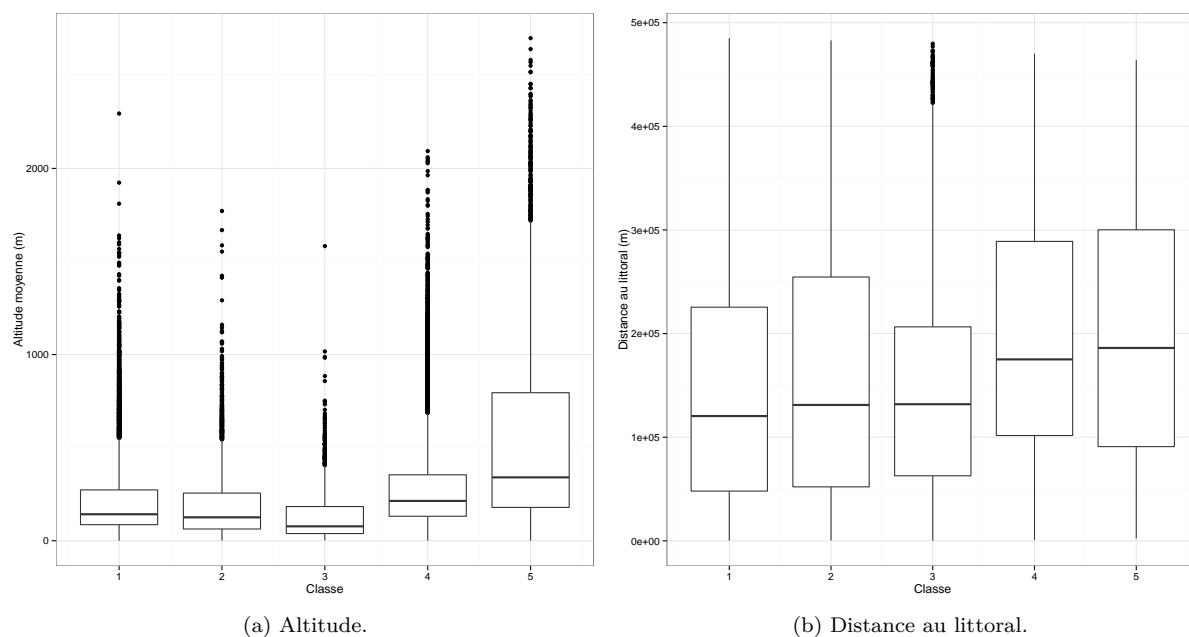


FIGURE 2.13 – Boîtes de Tuckey selon les classes issues de la typologie précédemment réalisée (cf. figure 1.14).

Au vu de ces graphiques il nous est difficile de conclure sur l'existence, ou non d'un effet induit par la classe d'appartenance. Certes la distribution des variables *altitude* et *distance au littoral* varie suivant la classe, mais cette variation n'est pas très importante. De plus les classes se recoupent, à titre d'exemple la classe n° 5 contient aussi bien les plus hautes communes de France que des littorales (cf. carte 1.14). Si il s'agit de la seule classe à intégrer des communes de forte altitude, toutes les classes contiennent des communes de faible altitude, ce qui est également valable pour la distance au littoral en fonction des différentes classes, toutes contiennent aussi bien des communes proches, que des communes éloignées du littoral. Ces observations vont à l'encontre des hypothèses statistiques que nous souhaitons tester, à savoir que l'altitude et la distance au littoral dépendent de la classe. Pour apporter une réponse définitive à cette questions nous allons effectuer deux analyses de la variance dont les résultats sont inscrits dans les tableaux 2.5 et 2.6. Quel que soit le modèle, le test de significativité de FISHER démontre l'existence d'une certaine cohérence au sein des classes. Dans les deux cas, la valeur du F de FISHER est telle que la *p-value* associée correspond à une probabilité de rejet erroné de l'hypothèse H_0 inférieure à $< 2, 2 \cdot 10^{-16}$. On peut donc rejeter l'hypothèse nulle et conclure à la significativité des facteurs *Altitude* et *Distance au littoral* au niveau $< 0, 001\%$ [21, 74]. Toutefois si l'on calcule la part de la variance expliquée (*i.e.* le rapport entre la variation *intra-classes* et la variation *inter-classes* [28]) on constate que celle-ci est relativement faible. Elle est de 13,4 % pour le premier modèle et de 2,4 % pour le second. Ainsi si il existe une relation réellement significative entre la précédente classification et les deux variables environnementales utilisées cette dernière n'est pas assez importante pour que l'on puisse considérer que les classes présentent une certaine homogénéité du point de vue environnemental. Et cet état de fait est parfaitement logique. Si l'on se réfère à la figure 2.13a, ou mieux encore à la carte 1.14, on remarque que des communes situées dans des milieux très différents appartiennent à la même classe. L'exemple le plus parlant est celui de la classe n° 5. Pour rappel cette dernière regroupe les communes ayant vu leur écart à la densité moyenne réduire au cours de la période étudiée. Or cette classe contient aussi bien des communes de l'Aquitaine que des communes Alpines, diamétralement opposées par leur altitude moyenne, mais également les communes pyrénéennes et des communes champenoises, opposées (mais dans une moindre mesure) par leur distance au littoral. Ainsi dans une même classe figurent des communes proches du littoral et de faible altitude, des communes montagneuses et proches du littoral, et des communes d'altitude moyenne et éloignées du rivage. Il est par conséquent évident que la typologie précédemment réalisée ne peut nous permettre de montrer qu'il existe un lien entre l'environnement et l'évolution de la distribution de la population. Les méthodes utilisées jusqu'ici ne suffisent pas et nous sommes dans l'obligation de compléter nos recherches pour apporter la meilleure réponse possible à notre problématique.

	Ddl	Somme des carrés	Moyenne des carrés	F	Pr
Classe	4	386 332 796	96 583 199	1 335,8	$< 2,2 \cdot 10^{-16}$
Résidus	34441	2 490 294 665	72 306		

TABLE 2.5 – Analyse de la variance. $R^2 = 13,43\%$

	Ddl	Somme des carrés	Moyenne des carrés	F	Pr
Distance	4	$1,3500 \cdot 10^{13}$	$3,3750 \cdot 10^{12}$	215,25	$< 2,2 \cdot 10^{-16}$
Résidus	34441	$5,4001 \cdot 10^{14}$	$1,5679 \cdot 10^{10}$		

TABLE 2.6 – Analyse de la variance. $R^2 = 2,44\%$

2.3.3 Typologie

Au vu des résultats précédents nous sommes encore dans l'incapacité de répondre à notre problématique. Si nous avons mis en évidence l'évolution de la distribution de la population dans le premier chapitre, et que nous avons quantifié les effets des composantes environnementales dans le second, nous n'avons pas encore proposé une synthèse de ces résultats. À la fin du précédent chapitre nous avons réalisé une typologie basée sur la méthode des *k-means*. Celle-ci présentait un résultat intéressant, nous permettant de condenser l'évolution de la distribution de la population en une figure, les parcours des communes étant efficacement classifiables. Cette typologie constituait en quelque sorte l'aboutissement du précédent chapitre car elle se basait sur tous nos précédents résultats pour faire émerger une nouvelle information. Mais notre travail ne portait pas uniquement sur l'évolution de la distribution de la population, mais sur les liens entre l'environnement et cette évolution. Les modèles linéaires nous ont montrés que la part de la distribution de la population expliquée par l'altitude et la distance au littoral était faible, et que la composante historique était capitale. Ces modèles nous ont montré que cette approche n'était pas des plus pertinentes et qu'il était peu, voir inefficace d'expliquer *l'écart à la densité* par l'altitude ou la distance au littoral. Nous nous sommes donc demandés si les classes précédemment réalisées possédaient une certaine cohésion au niveaux des deux composantes environnementales précédemment citées, et si des communes à l'évolution de la densité comparable présentaient des caractéristiques physiques similaires. Malheureusement le résultat de cette approche n'est pas concluant, des communes aux caractéristiques physiques très différentes étant situées dans les mêmes classes. Toutefois nous pensons qu'en raffinant les résultats de cette typologie nous pourrions obtenir une nouvelle typologie permettant de décrire l'évolution de l'écart à la densité des communes en fonction de leurs caractéristiques physiques. Ainsi nous pourrions conclure ce travail avec une classification basée sur nos précédents travaux, et synthétisant au mieux l'information.

Nous avons choisi de ne pas effectuer une nouvelle classification, mais au contraire de raffiner les résultats précédemment obtenus à la fin du premier chapitre. Ce choix à principalement été motivé par la volonté de permettre une comparaison des résultats entre les deux typologies, comme dit précédemment notre objectif est de « *raffiner* » les résultats et non d'en produire de nouveaux. La démarche suivie est extrêmement simple, elle consiste en une nouvelle classification à l'aide de la méthode des *k-means* des communes appartenant à la même classe. Les variables utilisées sont la distance au littoral et l'altitude communale moyenne. Du fait de la différence des ordres de grandeur ces dernières ont été préalablement centrées et réduites, afin d'éviter un biais important en faveur de la distance au littoral.

La première étape réalisée a consisté en la sélection du nombre de sous-classes optimales pour chaque classe. La procédure suivie est relativement simple, pour chaque classe nous avons opéré neufs découpages de 2 à 10 classes et avons sélectionné l'un d'eux à partir de la part de la variance expliquée par la classification. Nous n'avons pas sélectionné la partition présentant la plus grande part de la variance expliquée, mais nous avons au contraire cherché à sélectionner le découpage présentant le meilleur rapport entre part de la variance expliquée et le nombre de classes. Ainsi l'ajout d'une classe conduit généralement à l'augmentation de la part de la variance expliquée. Cette augmentation est variable, mais elle tend à diminuer avec le nombre de classes. Généralement on constate l'existence d'un seuil, à partir duquel l'augmentation du nombre de classes n'augmente presque plus la part de la variance expliquée, ce seuil correspond à un bon repère pour choisir le nombre de classes [21]. Dans notre cas, à partir de ce seuil l'augmentation de la part de la variance expliquée devenait généralement inférieure à deux points. Le tableau 2.7 représente le nombre de sous-classes sélectionnées pour chaque classes. On remarque que ce nombre est identique pour chaque classe. Il ne s'agit pas d'un choix délibéré, mais bien, comme expliqué précédemment du seuil correspondant à une diminution notable de la vitesse d'augmentation de la part de la variance expliquée. On peut remarquer que la valeur de cet indicateur est semblable pour chaque découpage, cette dernière est aux alentours de 85 %, ce qui en fait des découpages efficaces.

Mais la principale analyse de ces résultats ce doit d'être effectuée à partir des cartes qui en sont

issues. Malheureusement le grand nombre de classes ne nous permet pas d'en faire figurer l'intégralité sur une seule carte, sans sérieux problèmes de lisibilité. Nous avons donc représenté la classification interne à chaque classe sur six cartes différentes (cf. figure 2.14). Il est également important de noter que les couleurs des classes ne sont pas faites pour être comparées entre les différentes cartes, certes la même couleur correspond toujours au même numéro de sous-classe, mais les caractéristiques environnementales des communes appartenant à ces classes ne sont pas comparables. Si l'on compare les différentes cartes présentées, on remarque que certaines communes semblent appartenir à plusieurs classes (et donc plusieurs cartes), notamment dans les zones denses. Nous avons vérifié plusieurs fois cette anomalie et cette dernière est expliquée par la procédure d'export des cartes réalisées. Nous transformons ces dernières en pdf à partir du svg, puis en jpg à partir du pdf. Lors de ces opérations les géométries peuvent être simplifiées, notamment sur les grands aplats de couleurs, ce qui conduit à cette impression de superposition.

Les cartes de la planche 2.14 illustrent les découpages en sous-classes réalisés au sein de la première typologie. La carte de cette dernière est par ailleurs également présentée en juxtaposition, pour rappel. La classe correspond aux communes dont l'écart à la densité moyenne reste stable, et légèrement supérieur à la moyenne. Nous avons préalablement qualifié cette classe de « *classe moyenne* ». Ce nouveau découpage en six classes permet d'affiner les catégories en fonction des caractéristiques environnementales. La carte 2.14b, illustrant l'appartenance des communes à l'une de ces six classes permet de comprendre l'organisation des communes en fonction de leur composantes environnementales. Une première classe, extrêmement visible regroupe l'ensemble des communes *littorales* et *sublittorales* appartenant à la classe moyenne. Il s'agit de communes dont l'altitude moyenne, ainsi que la distance à la mer sont faibles. Une seconde classe, épouse les contours de la classe précédente. Cette dernière regroupe les communes avec une altitude moyenne et une distance au littoral plus élevée que celle de la classe précédente. Les communes appartenant à ces classes sont généralement situées à une centaine de kilomètres du littoral. La troisième catégorie est celle des communes de moyenne altitude, que l'on retrouve notamment dans le haut de la vallée du Rhône et de la Saône, puis viennent les communes très éloignées du rivage, telles que certaines communes d'Alsace, ainsi que les communes de haute et moyenne montagne, à l'altitude importante. Un regard sur la figure 2.15a nous permet de remarquer que l'évolution de l'écart moyen à la densité moyenne pour les communes de ces différentes sous classes était fortement similaire d'une classe à l'autre, et ce quel que soit la date. Cette observation nous conduit à nous poser une nouvelle question, cette typologie en sous-classes est-elle pertinente pour cet ensemble de communes ?

Nous apportons la réponse à cette question à travers la figure 2.15b. En effet nous sommes demandés si ce second découpage ne nous conduisait pas à modifier la représentativité des moyennes figurées sur le document 2.15a. Rappelons en effet que ce second découpage ne prend pas en compte l'évolution de l'écart à la densité moyenne, mais uniquement les caractéristiques physiques des communes. Ainsi l'algorithme peut conduire à la création de sous-classes à la dispersion plus importante que la classe dont sont issues les communes, toutefois le fait de procéder en deux découpages successifs à également pour effet de limiter ce phénomène. La méthode d'évaluation que nous proposons à l'aide de la figure 2.15b consiste à comparer l'évolution de la dispersion statistique causée par le second découpage. Nous avons calculé le *coefficient de variation* pour chaque classe de la première typologie, puis nous avons fait de même pour la seconde typologie. La valeur représentée sur la figure 2.15b est la différence entre ces deux valeurs. Une valeur négative indique que le *coefficient de variation* de la seconde classification est plus faible que celui de la première classification à la date t et une valeur positive que le *coefficient de détermination* a augmenté suite au second découpage. Par conséquent une valeur négative indique que les communes appartenant à la seconde classe sont plus proches de *l'écart moyen à la densité moyenne* à la date t que ne l'était l'ensemble des communes appartenant à la classe dont elles sont issues, et inversement.

Les communes préalablement classées dans la première classe par la première typologie ont pour code (11, 12, 13, 14, 15, et 16), correspondant à la première sous-classe de la classe 1, la seconde sous-classe de la classe 1 et ainsi de suite. Les communes issues de la première classe (première typologie), sont celles pour lesquelles *l'écart au coefficient de variation* de la classe de référence est le plus important (que ce soit dans le positif où le négatif), par conséquent cette seconde classification permet une réduction de la dispersion statistique pour certaines sous-classes et cause une augmentation de cette même dispersion pour d'autres sous-classes. La raison de ces importantes différences est à chercher dans la description de la première classe réalisée durant la première typologie, nous avons dès lors qualifié cette dernière de « *classe moyenne* ». Cette dernière regroupe des communes dont le comportement correspond à la variation globale, et comme nous le montrent les cartes 2.14a et 2.14b, un grand nombre de communes appartient à cette classe. Or le second découpage a affiné la première classification en y ajoutant la composante environnementale. Mais malgré ce second découpage, certaines classes continuent à occuper une vaste portion du territoire, notamment les classes 4 et 6. Si l'on se réfère à la figure 2.15b on note également que ces deux classes correspondent également à celles possédant le plus fort *écart au coefficient de variation de la classe originelle*. Cette observation nous permet de supposer que cette forte augmentation du *coefficient de variation* est imputable à la grande étendue de ces classes, qui certes

possèdent une cohérence en terme de géographie physique, mais au prix d'une augmentation du *coefficient de variation*.

Pour la seconde classe le nombre de communes concernées est bien moindre. Lors de la précédente typologie nous avons identifié cette classe comme étant celle regroupant des communes urbanisées de moyenne importance, principalement situées en banlieues des grandes agglomérations, à la densité plus importante, regroupées dans la troisième classe. L'évolution de la moyenne de *l'écart à la densité moyenne* présente plus de disparités entre les classes que précédemment. Certes 5 des 6 classes suivent une évolution groupée, mais la première, correspondant à certaines communes de la vallée du Rhône et du massif central (altitude moyenne à forte et distance au littoral moyenne) ce détache sensiblement de l'ensemble à partir du recensement de 1954. L'évolution des deux groupes est similaire, et l'on observe dans les deux cas une diminution de la vitesse de croissance à partir de 1975. Pour l'ensemble de ces sous-classes le recours à une seconde typologie s'avère efficace, puisqu'il conduit à une diminution globale des *coefficients de variation*, à l'exception de celui de la première classe qui subit une légère augmentation. Dans l'ensemble, le second découpage de cette classe est une réussite, puisque les classes obtenues présentent une certaine cohérence pour ce qui est de l'évolution leur *écart à la densité moyenne* et ce malgré le découpage basé sur des critères environnementaux.

La troisième classe correspond aux communes urbaines, dont *l'écart à la densité moyenne* est fortement supérieur à la moyenne, et ne cesse d'augmenter (cf. première typologie). Comme précédemment une classe (classe 6) ce distingue fortement des cinq autres. L'évolution de *l'écart à la densité moyenne* est également comparable (dans ces grandes lignes) à celui des cinq autres classes, mais sa vitesse d'évolution est bien supérieure. Si l'on se réfère à la carte 2.14d, on remarque que cette sous-classe concentre un grand nombre de communes, dont la ville de Paris, ainsi que sa proche banlieue. La dispersion statistique est légèrement plus importante qu'avant la première classification, mais il est intéressant de voir la pertinence des *clusters* obtenus après cette seconde typologie. Le recours à cette nouvelle classification permet dans ce cas d'extraire un comportement particulier, ce qui appuie l'hypothèse d'un lien entre évolution de *l'écart à la densité moyenne* et *variables environnementales*.

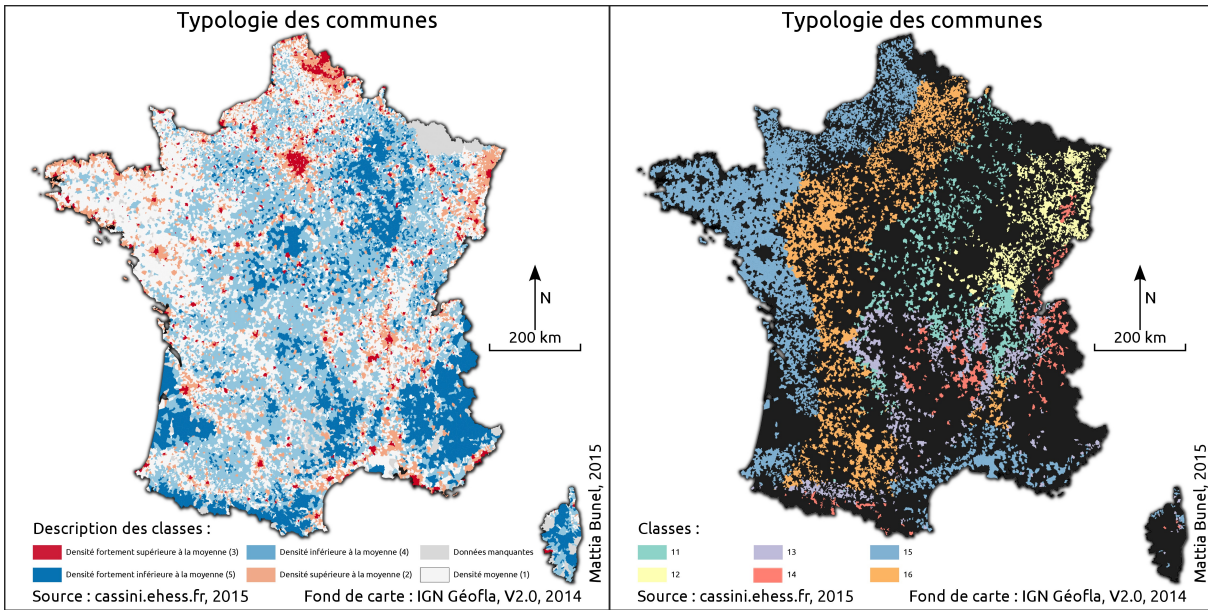
La quatrième classe issue de la première typologie, regroupe les communes dont *l'écart à la densité moyenne* est systématiquement inférieur à la moyenne. Le profil moyen de cette classe, présenté à la fin du premier chapitre, montrait une tendance à l'augmentation (dans le négatif) de *l'écart à la densité moyenne*, au cours du temps, ce dernier s'accélérait à partir du dénombrement de 1946. Si l'on se réfère à la figure 2.15a, on remarque (comme pour toutes les autres classes) que les sous-classes suivent une évolution de leur *écart à la densité moyenne* comparable à celle de la classe dont elles sont issues (ce qui est une résultante logique de leur méthode de construction). Mais, et c'est ce qui nous intéresse ici, on observe des variations entre ces classes, comme par exemple une modification de la hiérarchie, et une tendance à la dispersion au cours du temps de *l'écart à la densité moyenne* des sous-classes par rapport à *l'écart à la densité moyenne* de la classe issue de la précédente typologie. En ce rapportant à la carte 2.14e il nous est possible de décrire les caractéristiques physiques des communes composant les différentes sous-classes. La première sous-classe comprend principalement des communes situées dans la partie nord de la *diagonale du vide*, ces dernières ont une altitude moyenne et sont assez éloignées du littoral. La seconde sous-classe est composée de communes à l'altitude plus élevée que celle de la classe précédente. Il s'agit de communes de moyenne montagne, comme une partie du piémont pyrénéen, ou la partie nord-ouest du massif central. La quatrième sous-classe est composée de communes situées en moyenne à une centaine de kilomètres du littoral. Cette sous-classe est parallèle à la sous-classe 5, principalement composée de communes *sublittorales* et littorales. Cette structure en deux bandes parallèles est très semblable à celle visible sur la carte 2.14b. La quatrième sous-classe est composée de communes situées au niveaux des régions de l'Alsace, de la Lorraine ou de la Bourgogne. Il s'agit de communes avec une altitude légèrement plus élevée que la moyenne, et situées à une grande distance du littoral. La sixième et dernière sous-classe est composée des plus hautes communes, on retrouve notamment quelques communes alpines ou pyrénéennes, mais elles appartiennent dans leur grande majorité à la classe 5 (première typologie). Le graphique 2.15a nous permet de remarquer que les sous-classes 2 et 6, correspondant à des communes montagneuses sont celles dont *l'écart à la densité moyenne* est le plus important sur les dernières années de notre série statistique, à l'inverse la sous-classe 5 correspondant à des communes *littorales et sublittorales* est celle qui voit la valeur moyenne de *l'écart à la densité moyenne*, diminuer le moins, conduisant même à une profonde modification de la hiérarchie, puisque cette classe, dernière au début du XX^e siècle devient la première à partir de 1946.

La cinquième et dernière classe regroupe les communes dont *l'écart à la densité moyenne*, nettement inférieur à la moyenne, suit une tendance à la diminution. Ce nouveau découpage permet de réaliser une distinction entre les communes. Certaines sous-classes sont clairement rattachées à un environnement particulier, comme par exemple la sous-classe 3, constituée de communes de haute montagne, ou la sous-classe 4, regroupant des communes de faible altitude et proches du littoral. Comme pour la classe 3, ce nouveau découpage fait apparaître d'importantes variations entre les sous classes. Encore une fois,

Classe originale	Découpée en...	Part de la variance expliquée (%)
1		86,4
2		85,5
3	6	85,8
4		85,3
5		87,8

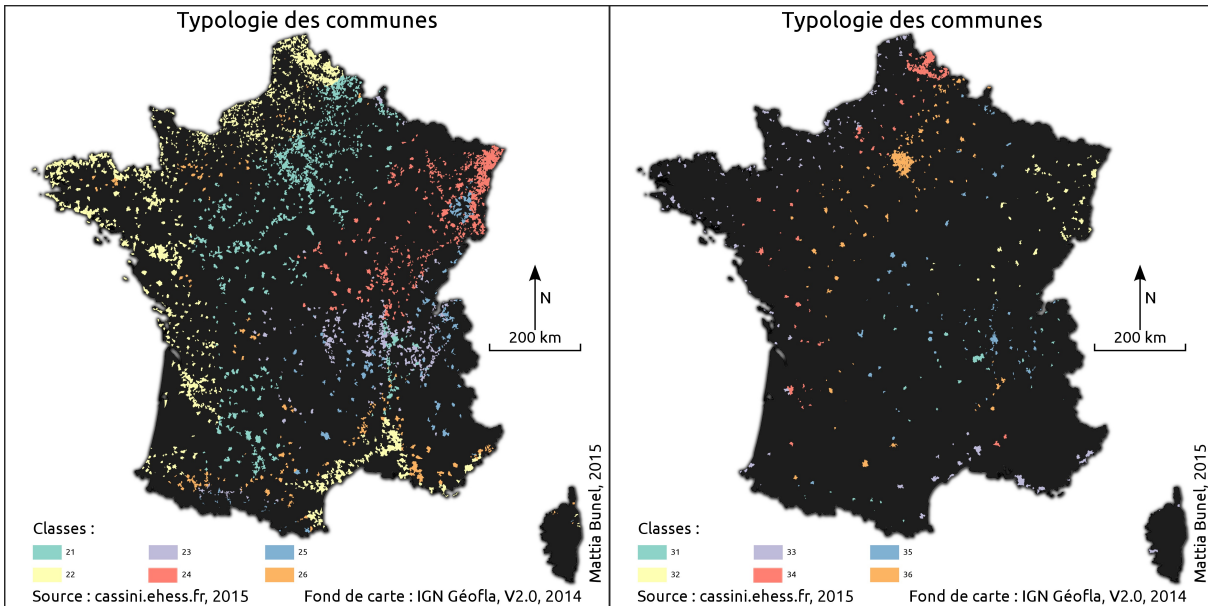
TABLE 2.7 – Part de la variance expliquée par la seconde typologie.

toutes les sous-classes voient leur *écart à la densité moyenne* évoluer de manière similaire, la première typologie ayant été réalisée selon ce critère. Deux classes se distinguent par leurs valeurs extrêmes, il s'agit des sous-classes 53 et 55, correspondant aux hautes et moyennes montagnes. Ces dernières possèdent un *écart moyen à la densité moyenne* semblable, et convergent l'une vers l'autre. À l'origine la sous-classe 6, correspondant aux zones de moyenne montagne, était caractérisée par un *écart moyen à la densité moyenne* plus faible que celui de la sous-classe 4 (communes relativement proches du littoral). Mais cette situation s'inverse à partir de 1876, dénombrement à partir duquel les deux courbes se chevauchent. Si l'on observe l'évolution de *l'écart moyen à la densité moyenne* pour la sous-classe 4, on remarque que celui-ci diminue assez lentement, et finit même par augmenter à partir de 1975, ce qui conduit cette sous-classe à être celle dont *l'écart moyen à la densité moyenne* est le plus important (tout en restant quasiment égal à celui de première sous-classe) des six sous-classes étudiées. Dans l'absolu, *l'écart à la densité moyenne* à vu sa valeur augmenter (dans le négatif) au cours du temps, mais si l'on se réfère à la hiérarchie, on constate que les communes de cette sous-classe sont celles qui ont perdu le moins d'importance, parmi les communes de la classe 5 (première typologie). À l'inverse des communes de montagnes, qui voient leur importance relative diminuer, ce phénomène est particulièrement intéressant si l'on se réfère à la sous-classe 2, correspondant (si l'on se réfère à la carte 2.14b) à des communes de montagne, à l'altitude moyenne plus faible, que celle des communes des sous-classes 3 et 6. Cette sous-classe, caractérisée par l'écart à la densité moyenne le plus faible entre 1826 et 1926, voit la valeur de celui-ci diminuer grandement, la faisant passer du premier rang au quatrième. Si l'on compare l'évolution de ces sous-classes à la carte 2.14f on peut voir apparaître une structure logique liant caractéristiques physiques de communes et évolution de *l'écart moyen à la densité moyenne*. Les trois sous-classes avec *l'écart à la densité moyenne* le plus faible (à partir de 1968), correspondent, comme indiqué précédemment à des communes montagnaises. Le découpage en sous-classes de la classe 5 s'avère pertinent, puisque les classes obtenues présentent des profils physiques différents. Ainsi, là où communes littorales et de haute-montagne ne faisaient auparavant qu'un tout, expliquant probablement en partie les faibles coefficients de détermination obtenus dans les modèles précédents, nous sommes désormais en mesure de distinguer ces classes et leur comportement.



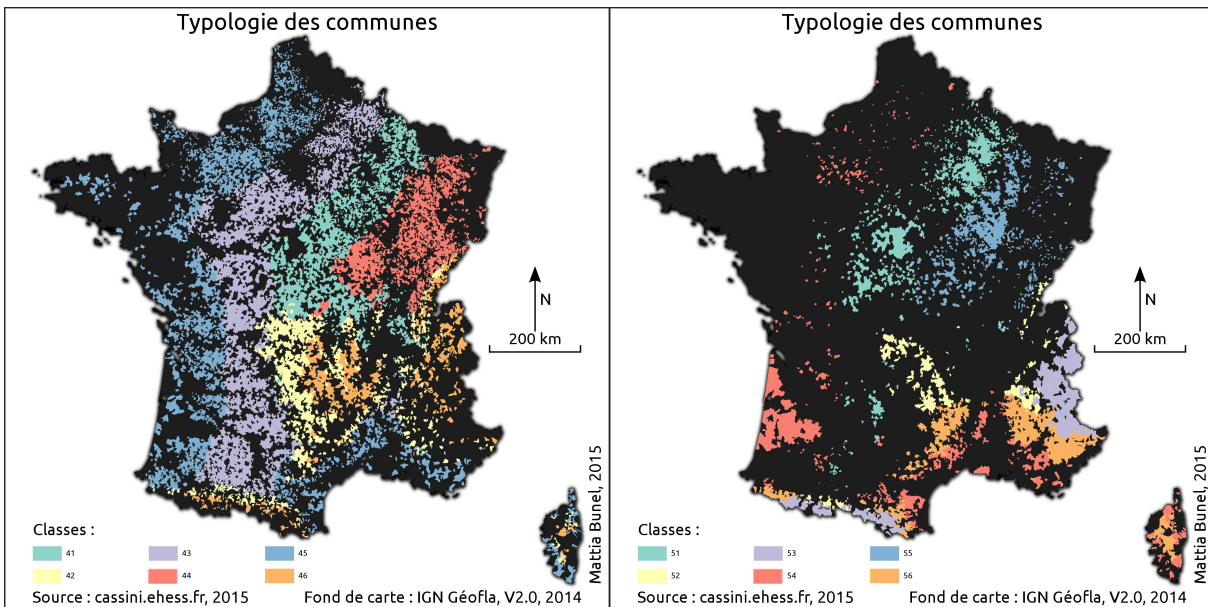
(a) Typologie initiale.

(b) Subdivision de la classe 1.



(c) Subdivision de la classe 2.

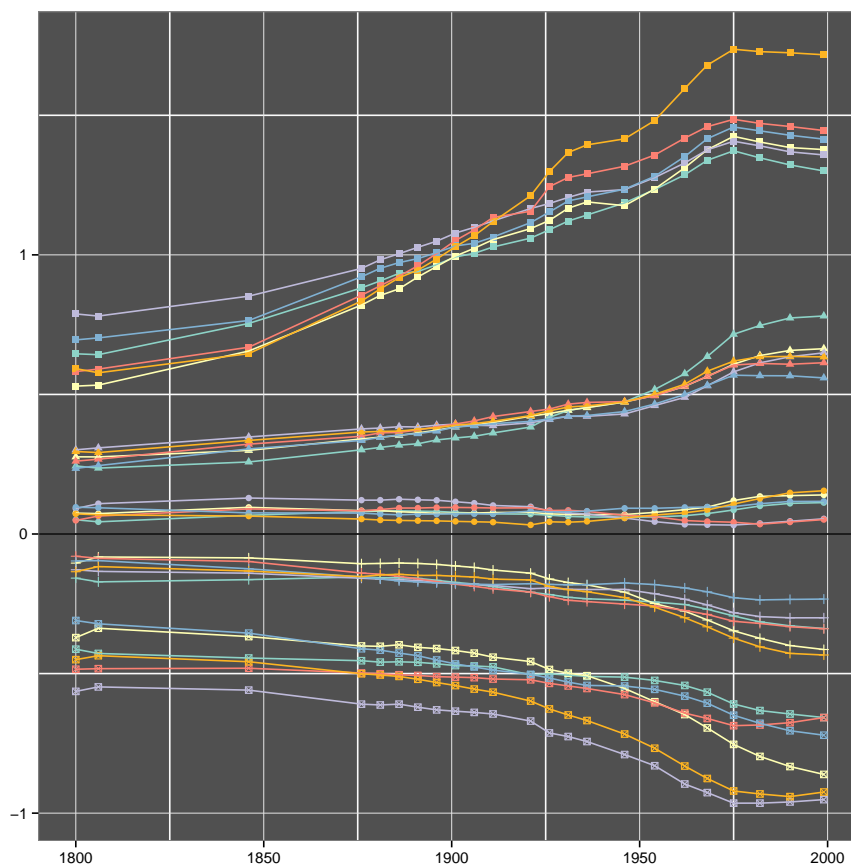
(d) Subdivision de la classe 3.



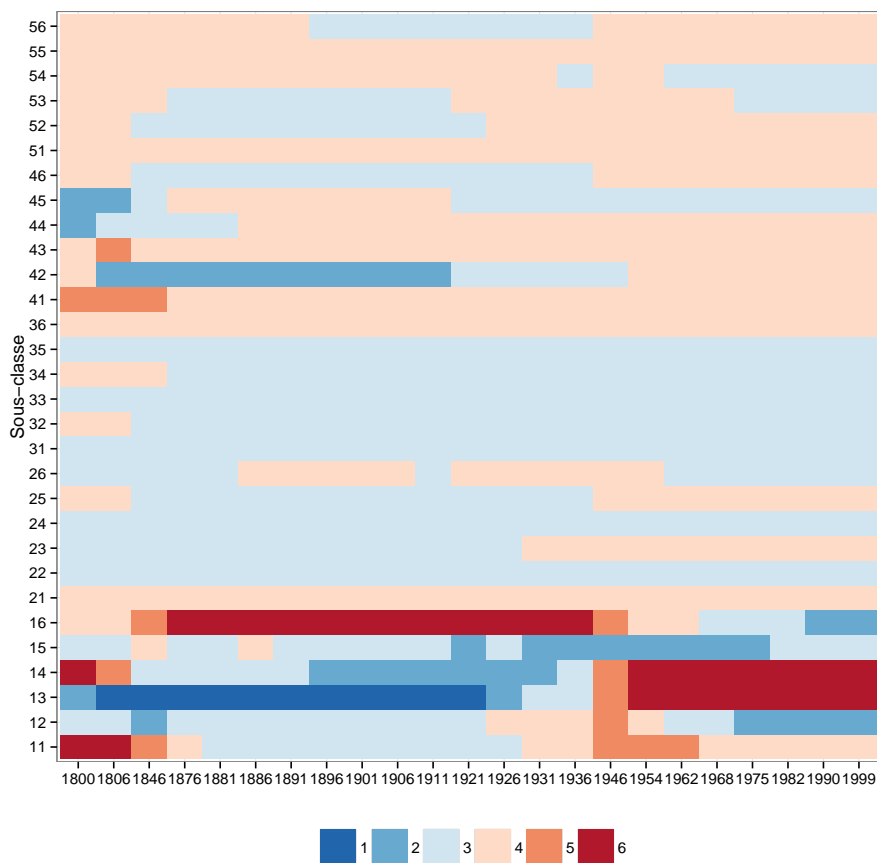
(e) Subdivision de la classe 4.

(f) Subdivision de la classe 5.

FIGURE 2.14 – Seconde typologie : cartes.



(a) Écart moyen à la densité moyenne par classe (variation de forme) et sous-classes (variation de couleur, les couleurs correspondent à celles figurées sur les cartes 2.14)



(b) Évolution du *coefficient de variation* suite à la seconde classification, en fonction de la sous-classe et de la date. 1, (-1; -0,5); 2, (-0,5; -0,25); 3, (-0,25; 0); 4, (0; 0,25); 5, (0,25; 0,5); 6, (0,5; 5), avec la borne supérieure exclue.

FIGURE 2.15 – Seconde typologie : autres documents.

Le recours à cette seconde typologie est dans l'ensemble extrêmement intéressant, puisque cette réali- sation nous a permis d'obtenir de nouvelles informations et de nouveaux éléments de réponse complétant nos précédents travaux. Pour rappel cette seconde typologie était basée sur la réutilisation des classes issues de la première typologie réalisée à la fin du premier chapitre. Cette dernière avait classé les communes suivant *l'écart à la densité moyenne* ainsi que son évolution. Par construction ces classes possédaient une certaine cohérence, mais nous sommes partis de l'hypothèse que la composante environnementale pouvait permettre de raffiner les résultats en classant les communes d'une même classe en plusieurs sous-classes cohérentes d'un point de vue environnemental. Il était en effet délicat d'étudier des classes composées de régions aussi différentes que les Alpes et l'Aquitaine (dont la majorité des communes appartiennent à la même classe, la cinquième, à l'issue de la première typologie). L'objectif de cette seconde classification était simple, il s'agissait de raffiner les résultats précédemment obtenus en fonction des deux variables environnementales étudiées, dans le but d'obtenir des sous-classes, présentant dans l'idéal des profils variés, signe d'une pertinence certaine de notre démarche. Et cette seconde classification a rempli son rôle, en nous transmettant trente classes dont certaines (11, 53, 56) s'éloignent grandement du modèle initial. Toutefois une question se posait, ces sous-classes assez différentes du modèle initial ne sont-elles pas artificiellement créées par une ou plusieurs valeurs aberrantes ? Nous avons proposé une estimation de ce phénomène à l'aide de la figure 2.15b. L'idée sous-jacente était que ce phénomène pouvait être quantifié à partir du *coefficient de variation*, une augmentation de ce dernier indiquant une plus forte dispersion et donc potentiellement une représentativité moins importante. Le *coefficient de variation* n'étant pas suffisant dans notre cas nous avons travaillé à partir de la différence entre le coefficient de variation de *l'écart à la densité moyenne* à la date t pour chaque sous-classe et du coefficient de variation de *l'écart à la densité moyenne* à la date t pour la classe de laquelle sont dérivées les sous classes (1 pour 11, 12, 13, 14, 15, 16 ; 2 pour 21, 22, 23, 24, 25, 26 etc.) Le résultat de cette opération est mitigé, et varie fortement en fonction des classes et de la date, mais dans l'ensemble il apparait que seules les sous-classes de la première classe subissent une augmentation du *coefficient de variation* réellement importante. Dans la grande majorité des cas le *coefficient de variation* augmente ou diminue très légèrement, ce qui laisse supposer que cette nouvelle classification reste pertinente, malgré l'augmentation importante du coefficient de variation pour certaines classes. Globalement on observe une tendance à la dispersion de la valeur de *l'écart à la densité moyenne* pour les sous classes sur le dernier quart de la série temporelle. Là où, jusqu'en 1946 les séries étaient sinon confondues tout du moins parallèles, on observe désormais des écarts entre sous-classes de plus en plus importants et sans doute causés par la concentration de la population préalablement mise en évidence.

Conclusion

L'objectif de cette seconde partie était de nous questionner sur le lien unissant distribution de la population et composantes environnementales. Comme nous avons pu le voir dans nos diverses recherches bibliographiques, mais également dans les cartes de la distribution de la population présentés dans le premier chapitre, la distribution de la population est contrainte par l'environnement, les milieux les plus *rudes*, étant généralement les moins denses. Une de nos premières réflexions a été de nous interroger sur la distribution de la population en fonction de deux variables environnementales que sont l'altitude moyenne des communes et leur distance au littoral. À l'aide de ces variables nous avons construit différentes cartes illustrant la part de la surface du territoire occupée par une part déterminée de la population classée suivant une de ces deux variables environnementales. Les résultats de cette approche sont intéressants car ils nous ont permis de mettre en valeur l'augmentation de la superficie occupée par le quart de la population situé aux plus hautes altitudes, et une concentration de la population aux plus basses altitudes. Ce résultats va dans le sens du dépeuplement subit par les montagnes au cours du XX^e siècle et notamment décrit par BEAUJEU-GARNIER [8]. L'application de la même méthode à la distance au littoral s'est révélée moins efficace, et nous avons été contraint de passer des quartiles aux déciles pour mettre en évidence des transformations. En consultant quelques références portants sur les phénomènes de *litturbanisation* nous en sommes venus à la conclusion que la concentration de la population sur les communes littorales était un phénomène agissant à une trop grande échelle pour être mis en évidence par les quartiles de la population, seuls les déciles s'avérant efficaces, car coïncidant parfaitement avec l'échelle du phénomène. Mais cette approche ne proposait qu'un résultat graphique, et nous avons pris la décision d'appliquer différents modèles statistiques à notre problématique. La première modélisation effectuée consistait en l'application de modèles de régression linéaire aux variables étudiées. Nous avons réalisé 66 modèles répartis en deux catégories, les premiers expliquant l'écart à la densité moyenne à la date t en fonction de l'altitude et les seconds, l'écart à la densité moyenne à la date t en fonction de la distance au littoral. Quelle que soit la date retenue les coefficients de détermination des modèles ayant l'altitude comme variable explicative se sont révélés nettement plus importants que ceux des modèles

ayant pour variable explicative la distance au littoral. Mais concevoir de tels modèles revient à supposer l'existence d'un lien direct et continu entre les deux variables étudiées. Or comme nous avons pu le voir lors de nos différentes lectures l'action de l'environnement sur la distribution de la population est avant tout la résultante d'une interaction indirecte. Nous nous sommes par conséquent demandés si les classes issues de la typologie précédemment réalisée ne présentaient pas une certaine cohérence du point de vue des caractéristiques physiques des communes les composant. Nous avons donc réalisé deux analyses de la variance, certes significatives, mais possédant un coefficient de détermination assez faible, explicable par la cohabitation dans une même classe de communes sensiblement différentes d'un point de vue environnemental. Pour remédier à ce problème nous avons décidé de décomposer les classes issues de la première typologie en 6 sous-classes, créées à partir des deux variables environnementales étudiées et d'observer l'évolution de l'écart à la densité moyenne de chacune de ces sous-classes. Cette opération s'est révélée concluante, et d'autant plus intéressante si les classes originelles possédaient des comportements « *extrêmes* » (notamment les classe 3 et 5). Les sous-classes ainsi obtenues présentaient une cohérence au niveau de leurs caractéristiques environnementales, et voyait leur écart à la densité moyenne évoluer de manière différente. L'exemple des communes de la classe 5 est particulièrement éloquent. Nous savions, au terme de la précédente typologie que les communes incluses dans cette classe voyaient leur écart à la densité moyenne (fortement négatif) devenir de plus en plus important. Mais cette classe était composée de nombreuses communes aux caractéristiques très variables. Le recours à cette seconde classification nous a permis d'isoler les communes montagneuses des communes littorales, nous avons ainsi constaté que les communes de haute et de moyenne montagne étaient celles qui voyaient leur *écart à la densité moyenne* diminuer le plus fortement, ce qui implique que ces dernières voient leur importance relative diminuer plus fortement que celles des autres communes. Et cette observation est parfaitement compatible avec le *dépeuplement des montagnes* décrit par BEAUJEU-GARNIER [8], mais également avec l'hypothèse de LE BRAS et LABBE [50] selon laquelle les milieux les plus rudes sont les plus rapidement dépeuplés.

Conclusion

Nous voici à présent au terme de ce travail de recherche portant sur un sujet aussi vaste que l'étude de la distribution de la population, sa variation et ces liens avec l'environnement, et ce sur une période de deux cent six ans. Il serait orgueilleux de prétendre que nous avons fait plus qu'effleurer un sujet si vaste et complexe que son étude a occupé de nombreux et éminents géographes durant des décennies et qui pourtant était encore récemment qualifié d'« énigme » par [68]. Il est par conséquent évident que notre travail, de part sa nature même, ne constitue qu'une nouvelle exploration d'un sujet massivement étudié. Toutefois ce dernier possède également quelques qualités qui le rendent original. Le point le plus important est bien entendu la base de données traitée. Certes, cette dernière n'est pas impressionnante de part sa taille (\simeq 1,4 millions de modalités), son échelle (communale) ou son nombre de variables (33), mais avant tout par son amplitude temporelle. Certains travaux, notamment ceux du projet ARCHAEOMEDES [3] on traité des données similaires sur une période (presque) identique et à la même échelle. Mais là où ce dernier ne portait que sur deux départements nous avons effectué notre étude à l'échelle de la métropole française. Bien entendu la base de données n'est pas parfaite, et son utilisation nous a confronté à plusieurs problèmes, aussi bien liés à la qualité de la donnée issue des opérations de recensement, qu'à l'absence de nombreuses communes. Ces problèmes étaient inhérents à l'utilisation de dénombremments aussi anciens et nous avons été contraints de nous en accommoder. Et c'est probablement le point le plus problématique de ce travail. Nous avons hérité d'une base de données très peu documentée et il a été nécessaire d'étudier et de recouper chaque document traitant du recensement de la population dans l'histoire pour reconstituer une histoire des données et par conséquent estimer leur qualité. Ce travail, long et délicat, a occupé la majorité du temps nécessaire à la construction ainsi qu'à la présentation de la base de données.

Pour répondre à notre hypothèse de départ nous avons pris la décision de scinder notre travail en deux parties, dans l'optique de simplifier au maximum notre recherche. L'objectif de notre première partie était de nous questionner sur l'évolution de la distribution de la population et plus particulièrement sur l'existence d'une tendance à la concentration de la population. Notre travail sur ce premier chapitre a été à la fois intéressant et efficace. Le recours aux différentes présentées a permis d'aborder notre problématique suivant plusieurs axes, se complétant mutuellement. Mais ce premier chapitre ne se résume pas aux méthodes statistiques, aussi diverses soit elles, employées. C'est avant tout grâce à la recherche bibliographique que nous avons pu lier nos observations à un phénomène plus global, la *transition urbaine*.

Le second chapitre nous a longtemps semblé moins aboutit. Nos résultats sont sensiblement moins impressionnants et aboutis que précédemment. Dans un premier temps nous avons supposé que la cause de ces résultats relativement mauvais était simplement explicable par l'utilisation de modèles non adaptés au(x) phénomène(s) étudié(s). Cette analyse, certes non dénuée de fondements, nous semble désormais clairement insuffisante. Il est avant tout nécessaire de nous demander si nous pouvons valider l'hypothèse de l'existence d'un lien entre l'évolution de la distribution de la population et les spécificités environnementales des communes. Certains des éléments apportés lors de ce travail étayent cette hypothèse, mais à ce stade aucun ne suffit à la valider totalement. Mais alors pourquoi sommes nous dans l'incapacité de montrer l'existence d'un tel lien, là où montrer la tendance à la concentration de la population c'était avéré un être un exercice simple ? Sans doute car l'effet de l'environnement est bien plus faible que celui de la *transition urbaine*. Les distinctions environnementales se retrouvent minimisées par une mutation nationale, marquant profondément la distribution de la population.

D'une certaine manière les résultats de ce travail rejoignent ceux de notre précédent travail de recherche. La composante environnementale, s'avère peu efficace pour expliquer la distribution de la population, contrairement à la composante historique. Ce constat et extrêmement stimulant intellectuellement, et nous ouvre de nouvelles voies de réflexion. La composante historique s'avère, au même titre que la composante spatiale, intéressante à étudier et riche en enseignements. Et l'étude conjointe de ces deux aspects constitue une excellente base de réflexion pour un nouveau travail

Annexe A

Documents relatifs à la base de données

A.1 Logiciels utilisés

Dans cette annexe figure la liste des logiciels les plus utilisés lors de la réalisation de ce travail. Ces derniers sont classés par ordre alphabétique et son associés au(x) numéro(s) de version(s) utilisé(s), ainsi qu'au nom de l'équipe de développement (ou du développeur). Dans le cas où nous avons fait usage de *modules*¹ supplémentaires, nous les avons adjoints à la liste. Ces derniers sont classés par fréquence d'utilisation.

La constitution de cette annexe répond à une triple attente. Dans un premier lieu elle permet au lecteur curieux d'obtenir plus d'informations sur la méthodologie employée. Nous avons volontairement choisi de ne travailler qu'à partir de logiciels libres, qui sont, par définition, faciles à obtenir. De plus la disponibilité du code source (faisant partie des caractéristiques des logiciels libres) permet d'étudier le fonctionnement de l'algorithme et par conséquent d'expliquer certains résultats aberrants. La disponibilité du code source de logiciels permet donc une totale transparence méthodologique². Dans un deuxième temps cette annexe peut également servir dans le cas où le fonctionnement de certaines fonctions se retrouve modifié. Il sera donc possible de retrouver la version précédemment utilisée à l'aide de ce document³. Cette partie est également l'occasion de remercier les différents développeurs qui ont rendu par leur travail, le mien possible.

- Grass Gis 7.0, Grass Development Team
- Inkscape 0.48.4, Inkscape Development Team
- OpenGeoDa 1.4, [Luc Anselin et al.](#)
- Python 3.4, Python Software Foundation
 - BeautifulSoup 4.3.2, [Leonard Richardson](#)
 - requests 2.2.1, [Kenneth Reitz](#)
- Qgis 2.10 *Pisa*, 2.8 *Wien*, 2.6 *Brington*, et 2.4 *Chugiak*, Qgis Development Team
 - Table Manager 0.6, [Borys Jurgiel](#)
 - QSpatiaLite 7.0.4, Romain Riviere
 - Mmqgis 2015.2.11, Michael Minn
 - fTools 0.6.2, [Carson Farmer](#)
- (Gnu) R 3.1.1 *Sock it to Me*, The R Fondation for Statistical Computing
 - ggplot 1.0.1, [Hadley Wickham](#)
 - sp 1.0-17, [Roger Bivand](#)
 - spdep 0.5-88, [Roger Bivand](#)
 - maptools 0.8-34, [Roger Bivand](#)

1. Cette dénomination à été choisie pour unifier en un seul terme des éléments à la fonction semblable mais à la dénomination différente. On parlera de *Bibliothèques* (ou de *library*) pour Python, de *plug-in* pour Qgis, de *package* pour Gnu R. *Spatialite* est quant à elle la *cartouche spatiale* de SQLite.

2. Il est à noter que lors de ce travail nous avons régulièrement consulté de code source de certaines fonctions du logiciel R. En effet dans certains cas il était nécessaire de comprendre comment la fonction traitait les valeurs nulles ou manquantes.

3. Une mise à jour récente de la bibliothèque BeautifulSoup a changé le *parseur* utilisé par la bibliothèque. Par conséquent, cette dernière qui était très efficace pour traiter des fichiers html non conformes aux recommandations du W3C, est subitement devenue incapable de traiter ces mêmes fichiers [75]. Par conséquent de nombreux développeurs Python ont pris la décision de revenir à une version antérieure de cette bibliothèque. Dans notre cas les fichiers html traités étaient suffisamment *propres* pour que cette action ne soit pas nécessaire.

- stats 3.1.1, R Core Team and contributors worldwide
- ineq 0.2-33, [Achim Zeileis](#)
- lmtest 0.9-33, [Torsten Hothorn](#)
- Saga Gis 2.1, Saga Development Team
- SQLite 3.8.2, [D. Richard Hipp](#)
- Spatialite 4.0.1, Alessandro Furieri

A.2 Un exemple de notice communale

NOTICE COMMUNALE		Ldh/EHESS/Cassini																																									
Eaux-Bonnes		EXPORTER																																									
superficie	3 852 ha																																										
altitude	520 m / 2 619 m																																										
coordonnées	Lambert II étendu x : 376 998,00 longitude : 0°23'31" W y : 1 778 326,00 latitude : 42°58'23" N																																										
code insee	64 2 16 204																																										
statut(s)	commune																																										
Administration actuelle (recensement 1999)																																											
région	Aquitaine																																										
département	Pyrénées-Atlantiques																																										
arrondissement	Oloron-Sainte-Marie																																										
canton	Laruns																																										
Administration ancienne (1789-1999)																																											
souveraineté	1789, royaume de France 1790, Basses Pyrénées																																										
département	1793, Basses Pyrénées 1801, Basses-Pyrénées [1969, Pyrénées-Atlantiques]																																										
district	1793, Oloron																																										
arrondissement	1801, Oloron [Oloron ... 1858, Oloron-Sainte-Marie]																																										
canton	1793, Bielle 1801, Laruns																																										
municipalité	1793, Aas																																										
Le nom																																											
ancien(s)	information non disponible actuellement																																										
nom(s) rév.	information non disponible actuellement																																										
an II	1793, Aas																																										
Bull. des Lois	1801, Aast Aas 1861, Eaux-Bonnes																																										
		Le territoire communal																																									
		<i>f. / c. / t. *</i> Absorbe en 1861, Assoste																																									
		<i>chef-lieu</i> -																																									
		<i>modif. limites</i> -																																									
		* fusion / cession / transfert de territoires communaux																																									
		Le nombre d'habitants (voir diagramme d'évolution)																																									
		Le recensement de 1826, qui ne serait qu'une réactualisation de celui de 1821, n'a pas été retenu.																																									
		Le recensement de 1871 a été, pour cause de guerre, repoussé à l'année 1872.																																									
		Le recensement de 1941, réalisé selon des instructions différentes, ne peut être qualifié de recensement général, et n'a donné lieu à aucune publication officielle.																																									
		Les résultats provisoires du recensement par sondage annuel réalisé en 2004, 2005 et 2006 selon les communes sont tous, par convention, affichés à 2006 (voir dossier Sources et site Insee).																																									
		Pour l'Alsace-Lorraine, les recensements de la période 1870-1919 ont eu lieu aux années 0 et 5 sauf 1871, et 1915 qui n'a pas été réalisé.																																									
		Pour Nice et la Savoie, les recensements de la période 1814-1860 ont eu lieu en 1815, 1822, 1838, 1848 et 1858.																																									
		<table border="1"> <thead> <tr> <th>autre</th> <th>1841</th> <th>1876</th> <th>1911</th> <th>1962</th> </tr> </thead> <tbody> <tr> <td>1793</td> <td>206</td> <td>1846 477</td> <td>1881 828</td> <td>1921 485</td> </tr> <tr> <td>1800</td> <td>152</td> <td>1851 366</td> <td>1886 874</td> <td>1926 455</td> </tr> <tr> <td>1806</td> <td>206</td> <td>1856 404</td> <td>1891 812</td> <td>1931 453</td> </tr> <tr> <td>1821</td> <td>259</td> <td>1861 762</td> <td>1896 775</td> <td>1936 462</td> </tr> <tr> <td>1831</td> <td>257</td> <td>1866 917</td> <td>1901 768</td> <td>1946 510</td> </tr> <tr> <td>1836</td> <td>269</td> <td>1872 836</td> <td>1906 684</td> <td>1954 558</td> </tr> <tr> <td></td> <td></td> <td></td> <td>2006</td> <td></td> </tr> </tbody> </table>		autre	1841	1876	1911	1962	1793	206	1846 477	1881 828	1921 485	1800	152	1851 366	1886 874	1926 455	1806	206	1856 404	1891 812	1931 453	1821	259	1861 762	1896 775	1936 462	1831	257	1866 917	1901 768	1946 510	1836	269	1872 836	1906 684	1954 558				2006	
autre	1841	1876	1911	1962																																							
1793	206	1846 477	1881 828	1921 485																																							
1800	152	1851 366	1886 874	1926 455																																							
1806	206	1856 404	1891 812	1931 453																																							
1821	259	1861 762	1896 775	1936 462																																							
1831	257	1866 917	1901 768	1946 510																																							
1836	269	1872 836	1906 684	1954 558																																							
			2006																																								
		légende - 0 hab. : commune inhabitée - ... : commune n'existant pas à cette date - adm. : commune recensée avec une autre - lac. : commune oubliée sur la publication du recensement - ill. : information illisible - vide : information non disponible actuellement - abs. : document (ou pages) disparu(es)																																									
		La paroisse sous l'ancien régime																																									
		<i>intendance ou généralité</i> information non disponible actuellement																																									
		<i>élection ou diocèse civil</i> information non disponible actuellement																																									
		<i>subdélégation</i> information non disponible actuellement																																									
		<i>parlement</i> information non disponible actuellement																																									
		<i>coutume</i> information non disponible actuellement																																									
		<i>bailliage ou sénéchaussée</i> information non disponible actuellement																																									
		<i>diocèse religieux</i> information non disponible actuellement																																									
		<i>vocable ou paroisse</i> information non disponible actuellement																																									

FIGURE A.1 – Un exemple de notice : la commune d'Eaux-Bonnes (Pyrénées-Atlantiques)

A.3 Évolution de la définition de la population

Date	Définition de la population	Source
1793	<i>Population de droit</i>	[55]
1800	Population de droit	[44]
1806	Population de droit	[44]
1821	Population de droit	[44]
1831	Population de droit	[44]
1836	Population de droit	[31, 44]
1841	Population de fait	[31, 44]
1846	Population de fait	[31, 44]
1851	Population de fait	[44]
1856	Population de fait	[44]
1861	Population de fait	[44]
1866	Population de fait	[44]
1872	Population de fait	[44]
1876	Population de fait	[44]
1881	Définitions multiples	[44]
1886	Définitions multiples	[44]
1891	Définitions multiples	[44]
1896	Définitions multiples	[44]
1901	Définitions multiples	[44]
1906	Définitions multiples	[44]
1911	Définitions multiples	[44]
1921	Définitions multiples	[44]
1926	Définitions multiples	[44]
1931	Définitions multiples	[44]
1936	Définitions multiples	[44]
1946	Définition multiples	[44]
1954	Population de fait	[44]
1962	Définitions multiples	[44]
1968	Définitions multiples	[44]
1975	Définitions multiples	[44]
1982	Définitions multiples	[44]
1990	Définitions multiples	[44]
1999	Définitions multiples	[44]

TABLE A.1 – Définition de la population utilisée lors du recensement correspondant. Les hypothèses figurent en *italique*.

Date	Population retenue	Source
1793	<i>Population de droit</i>	
1800	<i>Population de droit</i>	
1806	<i>Population de droit</i>	
1821	<i>Population de droit</i>	
1831	<i>Population de droit</i>	
1836	<i>Population de droit</i>	
1841	<i>Population de fait</i>	
1846	<i>Population de fait</i>	
1851	<i>Population de fait</i>	
1856	<i>Population de fait</i>	
1861	<i>Population de fait</i>	
1866	<i>Population de fait</i>	
1872	<i>Population de fait</i>	
1876	<i>Population de fait</i>	
1881	Population légale	[48]
1886	Population légale	[48]
1891	Population légale	[48]
1896	Population légale	[48]
1901	Population légale	[48]
1906	Population légale	[48]
1911	Population légale	[48]
1921	Population légale	[48]
1926	Population légale	[48]
1931	Population légale	[48]
1936	Population légale	[48]
1946	Population légale	[48]
1954	<i>Population de fait</i>	
1962	Population sans doubles comptes	[48]
1968	Population sans doubles comptes	[48]
1975	Population sans doubles comptes	[48]
1982	Population sans doubles comptes	[48]
1990	Population sans doubles comptes	[48]
1999	Population sans doubles comptes	[48]

TABLE A.2 – Population retenue lors de la construction de la base de données *cassini.chess*. Les hypothèses non étayées par des sources sont figurées en *italique*.

Annexe B

Lois relatives au recensement

B.1 Extrait de la loi relative à la démocratie de proximité

L'extrait suivant est l'article 156 de la loi n° 2002-276 du 27 février 2002 relative à la démocratie de proximité. Il fait partie du titre V (Des opérations de recensement) du chapitre III (Procédure d'enquête publique) (avec les articles 157 et 158). Il s'agit d'un document notable, qui fixe la structure légale de la méthodologie du recensement par échantillon tournant [52] dont la première application date de janvier 2004 [56, 46].

I. – Le recensement de la population est effectué sous la responsabilité et le contrôle de l'Etat.

II. – Le recensement a pour objet : 1° Le dénombrement de la population de la France ; 2° La description des caractéristiques démographiques et sociales de la population ; 3° Le dénombrement et la description des caractéristiques des logements. Les données recueillies sont régies par les dispositions de la loi n° 51-711 du 7 juin 1951 sur l'obligation, la coordination et le secret en matière de statistiques et de la loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés.

III. – La collecte des informations est organisée et contrôlée par l'Institut national de la statistique et des études économiques. Les enquêtes de recensement sont préparées et réalisées par les communes ou les établissements publics de coopération intercommunale, qui reçoivent à ce titre une dotation forfaitaire de l'Etat.

IV. – L'article L. 2122-21 du code général des collectivités territoriales est complété par un 10° ainsi rédigé : « 10° De procéder aux enquêtes de recensement. »

V. – Lorsque l'établissement public de coopération intercommunale a reçu des communes qui le constituent compétence pour préparer et réaliser les enquêtes de recensement, l'organe délibérant de l'établissement peut, par délibération, charger le président de l'établissement de procéder à ces enquêtes. Dans le cas où une commune ou un établissement public de coopération intercommunale refuserait ou négligerait d'accomplir cette mission, le représentant de l'Etat dans le département peut, après l'en avoir requis, y pourvoir d'office. Les enquêtes de recensement sont effectuées par des agents recenseurs, agents de la commune ou de l'établissement public de coopération intercommunale affectés à cette tâche ou recrutés par eux à cette fin. Lorsque l'activité exercée par un agent recenseur présente un caractère accessoire, elle est exclue de l'interdiction prévue à l'article L. 324-1 du code du travail. L'inéligibilité prévue au douzième alinéa de l'article L. 231 du code électoral s'applique à tous les agents recenseurs, quel que soit le nombre d'habitants de la commune.

VI. – Les dates des enquêtes de recensement peuvent être différentes selon les communes. Pour les communes dont la population est inférieure à 10 000 habitants, les enquêtes sont exhaustives et ont lieu chaque année par roulement au cours d'une période de cinq ans. Pour les autres communes, une enquête par sondage est effectuée chaque année ; la totalité du territoire de ces communes est prise en compte au terme de la même période de cinq ans. Chaque année, un décret établit la liste des communes concernées par les enquêtes de recensement au titre de l'année suivante.

VII. – Pour établir les chiffres de la population, l'Institut national de la statistique et des études économiques utilise les informations collectées dans chaque commune au moyen d'enquêtes de recensement ex-

haustives ou par sondage, les données démographiques non nominatives issues des fichiers administratifs, notamment sociaux et fiscaux, que l'institut est habilité à collecter à des fins exclusivement statistiques, ainsi que les résultats de toutes autres enquêtes statistiques réalisées en application de l'article 2 de la loi n° 51-711 du 7 juin 1951 précitée. A cette fin, les autorités gestionnaires des fichiers des organismes servant les prestations de base des régimes obligatoires d'assurance maladie transmettent à l'Institut national de la statistique et des études économiques les informations non nominatives qu'il appartient à l'institut d'agrèger cinq ans après leur réception, à un niveau géographique de nature à éviter toute identification de personnes.

VIII. – Un décret authentifie chaque année les chiffres des populations de métropole, des départements d'outre-mer et de Saint-Pierre-et-Miquelon, des circonscriptions administratives et des collectivités territoriales.

IX. – Les informations relatives à la localisation des immeubles, nécessaires à la préparation et à la réalisation des enquêtes de recensement, sont librement échangées entre l'Institut national de la statistique et des études économiques, les communes et les établissements publics de coopération intercommunale intéressés.

X. – Le premier décret authentifiant les chiffres de population en application du VIII sera publié à la fin de la première période de cinq ans mentionnée au VI.

B.2 Ordonnance du 16 janvier 1822

Ce document étant régulièrement cité dans le corps du document nous avons pris la décision de joindre ci-dessous un verbatim¹ de cette ordonnance. Le texte ci-dessous provient du tome 23 de la Collection complète des lois, décrets d'intérêt général, traités internationaux, arrêtés, circulaires, instructions, etc. publié en 1828 [35, p. 559-560].

16 = 23 JANVIER 1822. – *Ordonnance du Roi portant approbation du tableau de la population du royaume par département.* (VII, Bulletin DI, n° 12, 038.)

Voy. ordonnance du 15 mars 1827.

Louis, etc., sur le rapport de notre ministre secrétaire-d'Etat de l'intérieur ; vu les états de recensement de la population des départements de notre royaume, arrêtés au 1er janvier 1822 par les préfets ; considérant que ces états doivent servir de base à diverses opérations administratives qui se règlent sur la population ; notre Conseil entendu, nous avons ordonné et ordonnons ce qui suit :

Art. 1er. Le tableau ci-annexé [cf. tableau B.1 p. 97] de la population de notre royaume par département est approuvé, et sera considéré comme seul authentique pendant cinq ans, à compter du 1er janvier 1822.

2. Il sera statué ultérieurement sur l'état des villes et bourgs de notre royaume qui renferment une population agglomérée de quinze cents âmes et au-dessus.

3. Nos ministres secrétaires d'Etat sont chargés, chacun en ce qui le concerne, de l'exécution de la présente ordonnance, qui sera insérée au Bulletin des Lois.

1. Quelques coquilles typographiques, ainsi que certaines imprécisions de la *reconnaissance optique des caractères* (OCR/ROC) ont toutefois été corrigées.

DÉPARTEMENTS	POPULATION	DÉPARTEMENTS	POPULATION
Ain	328 838	Lot	275 296
Aisne	459 666	Lot-et-Garonne	330 121
Allier	280 025	Lozère	133 934
Alpes (Basses)	149 310	Maine-et-Loire	442 859
Alpes (Hautes)	121 418	Manche	594 196
Ardèche	304 339	Marne	309 444
Ardennes	266 985	Marne (Haute)	309 444
Arriège	234 878	Maienne	343 819
Aube	230 688	Meurthe	379 985
Aude	253 194	Meuse	292 385
Aveyron	339 422	Morbihan	416 224
Bouches du Rhône	313 614	Moselle	376 928
Calvados	492 613	Nièvre	257 990
Cantal	252 100	Nord	905 764
Charente	374 541	Oise	375 817
Charente-Inférieure	409 477	Orne	422 884
Cher	239 561	Pas-de-Calais	626 584
Corrèze	273 418	Puy-de-Dôme	533 410
Corse (Ile de)	180 348	Pyénées (Basses)	399 474
Côte-d'Or	358 148	Pyénées (Hautes)	212 077
Côte-du-Nord	552 424	Pyénées-Orientales	143 054
Creuse	248 785	Rhin (Bas)	502 638
Dordogne	453 136	Rhin (Haut)	370 062
Doubs	242 663	Rhône	391 580
Drome	273 511	Saône (Haute)	308 171
Eure	416 178	Saône-et-Loire	498 057
Eure-et-Loir	264 448	Sarthe	428 432
Finistère	483 095	Seine	821 706
Gard	334 164	Seine-Inférieure	655 804
Garonne (Haute)	391 118	Seine-et-Marne	303 150
Gers	301 336	Seine-et-Oise	424 490
Gironde	522 041	Sèvre (Deux)	279 845
Hérault	324 126	Somme	508 910
Ille-et-Vilaine	533 207	Tarn	313 713
Indre	230 373	Tarn-et-Garonne	238 143
Indre-et-Loire	282 372	Var	305 096
Isère	505 585	Vaucluse	224 431
Jura	301 768	Vendée	316 587
Landes	256 311	Vienne	260 697
Loir-et-Cher	227 527	Vienne (Haute)	272 330
Loire	343 524	Vosges	357 727
Loire-Inférieure	433 815		
Loiret	291 394	TOTAL GÉNÉRAL	30 465 291

TABLE B.1 – Relevé général de la population des départements du royaume en 1821

Bibliographie

- [1] Luc ANSELIN. *GeoDa 0.9 User's Guide*. Center for Spatially Integrated Social Science, 2003. URL : <http://geodacenter.org/downloads/pdfs/geoda093.pdf>.
- [2] Luc ANSELIN. « Local Indicators of Spatial Association-LISA ». In : *Geographical analysis* 27.2 (avr. 1995), p. 93–115. ISSN : 0016-7363. DOI : 10.1111/j.1538-4632.1995.tb00338.x. URL : <http://dx.doi.org/10.1111/j.1538-4632.1995.tb00338.x>.
- [3] ARCHAEOEMEDES. *Des oppida aux mérotopoles*. Paris : Economica, 1997.
- [4] Michel ARMATTE. « Lucien March (1859-1933) : une statistique mathématique sans probabilité ? » In : *Courrier des statistiques* 123 (2008). URL : http://www.insee.fr/fr/ffc/docs_ffc/cs123b.pdf.
- [5] Fabrice BARDET. « Du recensement au sondage de la population. L'exception démocratique française ». In : *Politix* 3.79 (2007). ISSN : 0295-2319. DOI : 10.3917/pox.079.0195. URL : <http://www.cairn.info/revue-politix-2007-3-page-195.htm>.
- [6] Guy BAUELLE. *Géographie du peuplement*. Paris : Armand Colin, 2000.
- [7] Céline BAYOU et Éric LE BOURHIS. « Norlisk et Mourmansk ». In : *Le courrier des pays de l'est* 1066 (2007).
- [8] Jacqueline BEAUJEU-GARNIER. *Géographie de la population*. T. 1. Paris : Librairie de Médicis, 1956. 429 p.
- [9] Jacques BERTILLION. *Des recensements de la population, de la nuptialité, de la natalité et de la mortalité à Paris pendant le XIX^e siècle et les époques antérieures*. In : *Annuaire statistique de la ville de Paris 1905*. Paris, 1907.
- [10] Jean-Noël BIRABEN. « L'histoire du peuplement humain des origines à nos jours. Histoire du peuplement et prévisions ». In : *Démographie : analyse et synthèse*. Sous la dir. de Graziella CASELLI, Jacques VALLIN et Guillaume WUNSCH. T. 5. INED, 2004. Chap. 66.
- [11] Michel BUSSI. « Quand l'INSEE projette d'enterrer le recensement de population... et la Géographie avec ! » In : *Cybergeo* (1999). URL : <http://cybergeo.revues.org.lama.univ-amu.fr/5531>.
- [12] Jean CASEVITZ. « Le Recensement du 10 mars 1946 ». In : *L'information géographique* 11.3 (1947).
- [13] Colette CAUVIN. *Cartographie thématique. Des transformations incontournables*. T. 2. Paris : Hermes, 2007.
- [14] Joël CHARRE. *Statistiques et territoire*. Montpellier : RECLUS, 1995.
- [15] Jean-Claude CHESNAIS. « La transition démographique : étapes, formes, implications économiques. Etude de séries temporelles (1720-1984) relatives à 67 pays ». In : *Population (French Edition)* 41.6 (nov. 1986), p. 1059. ISSN : 0032-4663. DOI : 10.2307/1532931. URL : <http://dx.doi.org/10.2307/1532931>.
- [16] François CLANCHÉ. « Bilan et enseignements de la rénovation du recensement français ». In : *La statistique et les 150 ans de l'unité italienne*. Colloque de la Société italienne de statistique. Bologne, 2011. URL : <http://unstats.un.org/unsd/censuskb20/Attachment484.aspx>.
- [17] François CLANCHÉ. « Focus? Le nouveau recensement en continu : une méthode originale, des résultats nombreux et faciles d'accès ». In : *Informations sociales* 3.183 (2014). ISSN : 0046-9459.
- [18] Nicole COEFFIC. « L'enquête post-censitaire de 1990. Une mesure de l'exhaustivité du recensement ». In : *Population* 48.6 (nov. 1993), p. 1655. ISSN : 0032-4663. DOI : 10.2307/1534158. URL : <http://dx.doi.org/10.2307/1534158>.
- [19] CONTRIBUTEURS DE WIKIPÉDIA. *1815 en France*. Sous la dir. de l'encyclopédie libre WIKIPÉDIA. 2015. URL : https://fr.wikipedia.org/wiki/1815_en_France.

- [20] CONTRIBUTEURS DE WIKIPÉDIA. *Liste des ministres français de l'Intérieur*. Sous la dir. de l'encyclopédie libre WIKIPÉDIA. 2015. URL : http://fr.wikipedia.org/wiki/Liste_des_ministres_fran%C3%A7ais_de_l'Int%C3%A9rieur.
- [21] Pierre-André CORNILLON, Arnaud GUYADER, Husson FRANÇOIS, Nicolas JÉGOU, Julie JOSSE, Maëla KOLAREG, Éric MATZNER-LØBER et Laurent ROUVIÈRE. *Statistiques avec R*. Rennes : Presses universitaires de Rennes, 2012.
- [22] DATAR. *Aménagement du territoire et recensement de la population en 1999*. Paris : DATAR, 2000.
- [23] Jean-Marc DEBARD et François LASSUS. *Le Dénombrement du comté de Bourgogne de 1688 : essai sur la situation démographique en Franche-Comté au lendemain de son annexion à la France*. Sous la dir. de Jean-Marc DEBARD. Besançon : Académie des sciences, belles-lettres et arts, 1979. URL : http://ccfr.bnf.fr/portailccfr/jsp/ccfr/sitemap/bmr_sitemap_view.jsp?record=bmr:UNIMARC:1243283.
- [24] Jean-Michel DECROLY. « La répartition des hommes à la surface de la planète. Histoire du peuplement et prévisions ». In : *Démographie : analyse et synthèse*. Sous la dir. de Graziella CASELLI, Jacques VALLIN et Guillaume WUNSCH. T. 5. INED, 2004. Chap. 70.
- [25] DÉPARTEMENT DES AFFAIRES ÉCONOMIQUES ET SOCIALES. *Principes et recommandations concernant les recensements de la population et des logements*. New York : ONU, 2009. ISBN : 978-92-1-261214-0.
- [26] Alain DESROSIÈRES, Jacques MAIRESSE et Michel VOLLE. « Les temps forts de l'histoire de la statistique française ». In : *Economie et statistique* 83 (1976). DOI : 10.3406/estat.1976.2401. URL : http://www.persee.fr/web/revues/home/prescript/article/estat_0336-1454_1976_num_83_1_2401.
- [27] Olivier DOLFUS. « Le système Monde ». In : *Géographie universelle*. Sous la dir. de Roger BRUNET. T. 1.2. Belin-Reclus, 1990.
- [28] Pierre DUMOLARD, Nathalie DUBUS et Laure CHARLEUX. *Les statistiques en géographie*. Paris : Belin, 2003.
- [29] Jacques DUPÂQUIER, Jean-Noël BIRABEN, Robert ÉTIENNE, Charles PIETRI, Luce PIETRI, Henri BAUTIER, Henri DUBOIS, Arlette HIGOUNET-NADAL et Christiane KLAPISCH-ZUBZER. *Histoire de la population française. Des origines à la Renaissance*. Sous la dir. de Jacques DUPÂQUIER. T. 1. Paris : PUF, 1988.
- [30] Jacques DUPÂQUIER, Guy CABOURDIN, Bernard LEPETIT, Jean-Pierre POUSSOU, Jean-Noël BIRABEN, Didier BLANCHET, Alain BLUM, Hector GUTIERREZ, Jean-Pierre GUTTON, François LEBRUN, Alain BIDEAU, Jean-Pierre BARDET, Jacques HOUDAILLE, Antoinette FAUXVE-CHAMOUX, Jean-Yves GRENIER, André BURGUIÈRE et Jean-Claude PERROT. *Histoire de la population française. De la Renaissance à 1789*. Sous la dir. de Jacques DUPÂQUIER. T. 2. Paris : PUF, 1988.
- [31] Jacques DUPÂQUIER, Alain DROUARD, Maurice GARDEN, Hervé LE BRAS, Alfred SAUVY, Jean-Claude CHESNAIS, Guy DESPLANQUES, Jacques VALLAIN, Noël BIRABEN, Denise PUMAIN, Daniel COURGEAU, Philippe BOURCIER DE CARBON et Martine SEGALIN. *Histoire de la population française. De 1914 à nos jours*. Sous la dir. de Jacques DUPÂQUIER. T. 4. Paris : PUF, 1988.
- [32] Jacques DUPÂQUIER, René LE MÉE, Joseph GOY, Maurice GARDEN, Hervé LE BRAS, Bernard LEPETIT, Jean-Pierre POSSOU, Daniel COURGEAU, Jean-Pierre BARDET, Alain BIDEAU, Jean-Noël BIRABEN, Jacques LÉONARD, Bernard LÉCUYER, Patrice BOURDELAIS, Agnès FINE, Martine SEGAKEN, Yves CHARBIT et André BEJIN. *Histoire de la population française. De 1789 à 1914*. Sous la dir. de Jacques DUPÂQUIER. T. 3. Paris : PUF, 1988.
- [33] Jacques DURAND. « L'attraction des nombres ronds et ses conséquences économiques ». In : *Revue française de sociologie* (1961). URL : http://www.persee.fr/web/revues/home/prescript/article/rfsoc_0035-2969_1961_num_2_3_5938.
- [34] François DURAND-DASTÈS. « La mémoire de Gaïa ». In : *Géopoint 90*. Sous la dir. de Groupe DUPONT. 1990.
- [35] Jean-Baptiste DUVERGIER et CONSEIL D'ÉTAT. *Collection complète des lois, décrets d'intérêt général, traités internationaux, arrêtés, circulaires, instructions, etc.* vol. 23. Paris : Sirey, 1828. URL : <https://books.google.fr/books?id=W2ZFQAAMAAJ>.
- [36] Orlando FIGES. *The Crimean War : A History*. New York : Picador, 2010.
- [37] Vincent FONSAGRIVE. « À propos du recensement du 10 mars 1946 ». In : *Population* 1.2 (avr. 1946), p. 203–212. ISSN : 0032-4663. DOI : 10.2307/1523988. URL : <http://dx.doi.org/10.2307/1523988>.

- [38] A. Stewart FOTHERINGHAM, Christon BRUNSDON et Martin CHARLTON. *Geographically Weighted Regression : The Analysis of Spatially Varying Relationships*. London : Wiley, 2002.
- [39] Alain GODINOT. « La rénovation du recensement de la population ». In : *Courrier des statistiques* (2003). URL : <http://www.epsilon.insee.fr/jspui/bitstream/1/8438/1/cs105-106b.pdf>.
- [40] Louis HENRY. « Le controle des recensements ». In : *Population* 4.2 (avr. 1949), p. 231. ISSN : 0032-4663. DOI : [10.2307/1524044](https://doi.org/10.2307/1524044). URL : <http://dx.doi.org/10.2307/1524044>.
- [41] INSEE. *Cinquante ans d'Insee : 1946-1996*. Paris : INSEE, 1996.
- [42] INSEE. *Code Officiel Géographique*. 2015. URL : <http://www.insee.fr/fr/methodes/nomenclatures/cog/>.
- [43] INSEE. *Le recensement à l'étranger*. 2005. URL : <http://www.insee.fr/fr/ppp/sommaire/imeths01zx.pdf>.
- [44] INSEE. *Le recensement de la population dans l'Histoire*. 2005. URL : <http://www.insee.fr/fr/ppp/sommaire/imeths01c.pdf>.
- [45] INSEE. *Population sans doubles comptes (recensement de la population 1999)*. 2015. URL : <http://www.insee.fr/fr/methodes/default.asp?page=definitions/pop-sans-double-comptes-rp99.htm>.
- [46] INSEE. *Recensement de la population : Définition*. 2015. URL : <http://www.insee.fr/fr/methodes/default.asp?page=definitions/recensement-de-la-population.htm>.
- [47] Claire KUBRAK. *Concentration et spécialisation des activités économiques : des outils pour analyser les tissus productifs locaux*. Paris : INSEE, 2013. URL : http://www.insee.fr/fr/publications-et-services/docs_doc_travail/H2013-01.pdf.
- [48] LDH. *Des villages de Cassini aux communes d'aujourd'hui : Sources*. Ehess. 2004. URL : http://cassini.ehess.fr/cassini/fr/html/5_donnees.htm.
- [49] Hervé LE BRAS. *Le Peuplement De l'Europe*. Paris : La documentation Française, 1997.
- [50] Hervé LE BRAS et Morgane LABBE. *La planète au village*. Paris : Éditions de l'Aube, 1993. URL : <http://hdl.handle.net/2042/28523>.
- [51] Monique LE GUEN. « La boîte à moustaches de TUKEY, un outil pour initier à la Statistique ». In : *Statistiquement Votre - SFDS* 4 (2001), p. 1-3. URL : <https://halshs.archives-ouvertes.fr/halshs-00287697>.
- [52] Olivier LEFEBVRE. « Le recensement de la population en France ». In : 57th Session of the International Statistical Institute (ISI) : our past, present and future. 2009. URL : http://unstats.un.org/unsd/censuskb20/Attachments/2009FRA_ISI-GUIDb7b8b1cb4c914a999e0f85bffab8843a.pdf.
- [53] Henri LERIDON. « La seconde révolution contraceptive : la régulation des naissances en France de 1950 à 1985 ». In : *Population (French Edition)* 42.2 (mar. 1987), p. 359. ISSN : 0032-4663. DOI : [10.2307/1533089](https://doi.org/10.2307/1533089). URL : <http://dx.doi.org/10.2307/1533089>.
- [54] Colin MCEVEDY et Richard JONES. *Atlas of world population history*. Harmondsworth : Penguin Books, 1978.
- [55] Paul MEURIOT. « Le recensement de l'an II ». In : *Journal de la société statistique de Paris* (1918). URL : http://www.numdam.org/item?id=JSFS_1918__59__34_0.
- [56] MINISTÈRE DE L'INTÉRIEUR. *Loi n° 2002-276 du 27 février 2002 relative à la démocratie de proximité*. 2002. URL : <http://legifrance.gouv.fr/eli/loi/2002/2/27/2002-276/jo/texte>.
- [57] Claude MOTTE, Isabelle SÉGUY et Christine THÉRÉ. *Communes d'hier, communes d'aujourd'hui : les communes de la France métropolitaine, 1801-2001 : dictionnaire d'histoire administrative*. Paris : INED, 2003, p. 406. URL : https://books.google.fr/books?id=fXnCxBqCL6cC&dq=des+villages+de+cassini+aux+communes+d'+aujourd'hui&lr=&hl=fr&source=gbs_navlinks_s.
- [58] Claude MOTTE et Marie-Christine VOULOIR. « Le site cassini.ehess.fr un instrument d'observation pour une analyse du peuplement ». In : *Conférence catographique internationale* 191 (2007). URL : <http://www.lecfc.fr/new/articles/191-article-7.pdf>.
- [59] Jacques NECKER. *De l'Administration des finances de la France*. T. 1. Paris, 1784. URL : <http://gallica.bnf.fr/ark:/12148/btv1b8613396p>.
- [60] Daniel NOIN. *Atlas de la population mondiale*. Montpellier : Reclus, 1991.
- [61] Daniel NOIN. *Géographie de la population*. Paris : Armand Colin, 2005.

- [62] Sébastien OLIVEAU. « Autocorrélation spatiale : leçons du changement d'échelle ». In : *Espace géographique* 1 (avr. 2010), p. 51–64. URL : <https://halshs.archives-ouvertes.fr/halshs-00469386>.
- [63] Sébastien OLIVEAU. « Les indicateurs locaux d'association spatiale (LISA) comme méthode de régionalisation : Une application en Inde ». In : *Théoquant* (2005).
- [64] Sébastien OLIVEAU. « Modernisation villageoise et distance à la ville en Inde du Sud ». Thèse de doctorat. Paris : Université Paris I Pantheon - Sorbonne, nov. 2004.
- [65] Philippe PINCHEMEL et Geneviève PINCHEMEL. *La face de la Terre*. Paris : Armand Colin, 1988.
- [66] Mario POLÈSE. « Urbanisation et développement économique ». In : *Encyclopédie de Géographie*. Sous la dir. d'Antoine BAILLY, Robert FERRAS et Denise PUMAIN. Economica, 1995. Chap. 37.
- [67] Christian PRADEAU. « L'information géographique ». In : *Réseaux hydrographiques et organisation de l'espace* 63 (1999).
- [68] Denise PUMAIN. « Le Peuplement ». In : *Encyclopédie de Géographie*. Sous la dir. d'Antoine BAILLY, Robert FERRAS et Denise PUMAIN. Economica, 1995. Chap. 23.
- [69] Denise PUMAIN. *Peuplement*. Sous la dir. d'HYPERGÉO. 2014. URL : <http://www.hypergeo.eu/spip.php?article442>.
- [70] Denise PUMAIN et Thérèse SAINT-JULIEN. *L'analyse spatiale. Localisations dans l'espace*. T. 1. Paris : Armand Colin, 1997. 167 p.
- [71] Ricco RAKOTOMALALA. *Tests de normalité : Techniques empiriques et tests statistiques*. Lyon, 2001. URL : http://eric.univ-lyon2.fr/~ricco/cours/cours/Test_Normalite.pdf.
- [72] Marcel RONCAYOLO. *Histoire de la France urbaine. la ville aujourd'hui*. Sous la dir. de Georges DUBY. T. 5. Paris : Seuil, 1985.
- [73] Thérèse SAINT-JULIEN. *Diffusion spatiale*. Sous la dir. d'HYPERGÉO. 2014. URL : <http://www.hypergeo.eu/spip.php?article11>.
- [74] Gilbert SAPORTA. *Probabilités, analyse des données et statistique*. Paris : Technip, 2011.
- [75] Sébastien SAUVAGE. *BeautifulSoup et html invalide*. 2013. URL : <http://sebsauvage.net/links/?PwAG5g>.
- [76] Christopher SMALL et Joel E. COHEN. « Continental physiography, climate, and the global distribution of human population ». In : *Current Anthropology* (2004), p. 269–277.
- [77] Pierre-Jean THUMERELLE. *Les populations du monde*. Paris : Nathan, 1996. 384 p.
- [78] John TUKEY et John WILDER. *Exploratory Data Analysis*. London : Addison-Wesley, 1977.
- [79] Jacques VALLIN. « La transition démographique européenne : 1740-1940. Histoire du peuplement et prévisions ». In : *Démographie : analyse et synthèse*. Sous la dir. de Graziella CASELLI, Jacques VALLIN et Guillaume WUNSCH. T. 5. INED, 2004. Chap. 68.
- [80] Jacques VALLIN et Graziella CASELLI. « Quand l'Angleterre rattrapait la France ». In : *Population et sociétés* 346 (1999).
- [81] Dirk VAN DE KAA. « Europe's second demographic transition ». In : *Population bulletin* 42.1 (1987), p. 1–59.
- [82] Jean-Marc ZANINETTI. « L'urbanisation du littoral en France ». In : *Population & Avenir* 677.2 (2006), p. 4. ISSN : 1968-3952. DOI : [10.3917/popav.677.0004](https://doi.org/10.3917/popav.677.0004). URL : <http://dx.doi.org/10.3917/popav.677.0004>.

Table des figures

1.1	Timbres à l'effigie du recensement de la population.	18
	(a) 1982 (n° 2202)	
	(b) 1999 (n° 3223)	
1.2	Communes sans données en fonction du recensement.	21
	(a) <i>Écart à la densité moyenne</i> en 1793.	
	(b) <i>Écart à la densité moyenne</i> en 1901.	
	(c) <i>Écart à la densité moyenne</i> en 1921.	
	(d) <i>Écart à la densité moyenne</i> en 1946.	
	(e) <i>Écart à la densité moyenne</i> en 1975.	
	(f) <i>Écart à la densité moyenne</i> en 1999.	
1.3	Évolution de la population et de la densité.	26
	(a) Évolution de la population française.	
	(b) Évolution de la densité de population.	
1.4	Évolution de <i>l'écart-type</i> et pente locale associée.	29
	(a) Distribution de la variable <i>écart à la densité moyenne</i> pour chaque recensement.	
	(b) Évolution de <i>l'écart-type</i> .	
	(c) Pente locale pour la figure 1.4b.	
1.5	<i>Boxplot</i> de la variable <i>écart à la densité moyenne</i>	31
	(a) <i>Boxplot</i> médiane quartiles.	
	(b) <i>Boxplot</i> moyenne écart-type.	
1.6	Concentration de la population française entre 1793 et 1999.	35
	(a) Mesure de la concentration de la population française pour chaque dénombrement	
	(b) Évolution du coefficient de GINI	
	(c) Concentration de la population française.	
1.7	Courbe logistique	36
1.8	Taux d'urbanisation en France entre 1700 et 2007.	36
1.9	Cartographie de la variable <i>écart à la densité moyenne</i>	39
	(a) <i>Écart à la densité moyenne</i> en 1793.	
	(b) <i>Écart à la densité moyenne</i> en 1901.	
	(c) <i>Écart à la densité moyenne</i> en 1921.	
	(d) <i>Écart à la densité moyenne</i> en 1946.	
	(e) <i>Écart à la densité moyenne</i> en 1975.	
	(f) <i>Écart à la densité moyenne</i> en 1999.	
1.10	Évolution du coefficient de MORAN.	41
	(a) Coefficient de MORAN en fonction de la date.	
	(b) Coefficient de MORAN en fonction du rang.	
1.11	<i>Indicateurs locaux d'association spatiale</i>	44
	(a) <i>Écart à la densité moyenne</i> en 1793, $I = 0,45$.	
	(b) <i>Écart à la densité moyenne</i> en 1901, $T = 0,54$.	
	(c) <i>Écart à la densité moyenne</i> en 1921, $I = 0,56$.	
	(d) <i>Écart à la densité moyenne</i> en 1946, $I = 0,59$.	
	(e) <i>Écart à la densité moyenne</i> en 1975, $I = 0,63$.	
	(f) <i>Écart à la densité moyenne</i> en 1999, $I = 0,71$.	
1.12	Nuages de points de MORAN.	45
	(a) <i>Écart à la densité moyenne</i> en 1793.	
	(b) <i>Écart à la densité moyenne</i> en 1901.	
	(c) <i>Écart à la densité moyenne</i> en 1921.	
	(d) <i>Écart à la densité moyenne</i> en 1946.	
	(e) <i>Écart à la densité moyenne</i> en 1975.	
	(f) <i>Écart à la densité moyenne</i> en 1999.	
1.13	Niveau local d'autocorrélation spatiale	46

(a)	Écart à la densité moyenne en 1793.	
(b)	Écart à la densité moyenne en 1901.	
(c)	Écart à la densité moyenne en 1921.	
(d)	Écart à la densité moyenne en 1946.	
(e)	Écart à la densité moyenne en 1975.	
(f)	Écart à la densité moyenne en 1999.	
1.14	Typologie des communes.	49
1.15	Moyenne et médiane par classe et par date.	49
(a)	Écart moyen à la densité moyenne, par classes.	
(b)	Écart médian à la densité moyenne, par classes.	
2.1	Évolution du coefficient de détermination en fonction de l'année.	58
(a)	Autocorrélation temporelle	
(b)	Autocorrélation spatiale	
(c)	Autocorrélation spatiale (voisins au temps $t - 1$)	
(d)	Combinaison des modèles précédents	
2.2	Part de la superficie occupée par classe.	63
(a)	En fonction de l'altitude.	
(b)	En fonction de la distance à la mer.	
2.3	Population cumulée en fonction de l'altitude en 1836.	64
(a)	Communes sans données.	
(b)	Population cumulée en fonction de l'altitude.	
2.4	Topographie de la France métropolitaine.	68
2.5	Population cumulée en fonction de l'altitude.	70
(a)	Écart à la densité moyenne en 1793.	
(b)	Écart à la densité moyenne en 1901.	
(c)	Écart à la densité moyenne en 1936.	
(d)	Écart à la densité moyenne en 1946.	
(e)	Écart à la densité moyenne en 1975.	
(f)	Écart à la densité moyenne en 1999.	
2.6	Population cumulée en fonction de la distance à la mer.	71
(a)	Écart à la densité moyenne en 1793.	
(b)	Écart à la densité moyenne en 1901.	
(c)	Écart à la densité moyenne en 1936.	
(d)	Écart à la densité moyenne en 1946.	
(e)	Écart à la densité moyenne en 1975.	
(f)	Écart à la densité moyenne en 1999.	
2.7	Modèles de régressions, écart à la densité moyenne en fonction de l'altitude.	74
(a)	Coefficient de détermination.	
(b)	Coefficient de la variable.	
2.8	Modèles de régressions, écart à la densité moyenne en fonction de l'altitude (1).	75
(a)	Valeur estimée en 1793, $R^2 = 0,07$.	
(b)	Valeur estimée en 1901, $R^2 = 0,08$.	
(c)	Résidus en 1793.	
(d)	Résidus en 1901.	
(e)	Écart à la densité moyenne en 1793.	
(f)	Écart à la densité moyenne en 1901.	
2.9	Modèles de régressions, écart à la densité moyenne en fonction de l'altitude (2).	76
(a)	Valeur estimée en 1921, $R^2 = 0,08$.	
(b)	Valeur estimée en 1946, $R^2 = 0,11$.	
(c)	Résidus en 1921.	
(d)	Résidus en 1946.	
(e)	Écart à la densité moyenne en 1921.	
(f)	Écart à la densité moyenne en 1946.	
2.10	Modèles de régressions, écart à la densité moyenne en fonction de l'altitude (3).	77
(a)	Valeur estimée en 1975, $R^2 = 0,13$	
(b)	Valeur estimée en 1999, $R^2 = 0,13$.	
(c)	Résidus en 1975.	
(d)	Résidus en 1999.	
(e)	Écart à la densité moyenne en 1975.	
(f)	Écart à la densité moyenne en 1999.	

2.11	Modèles de régressions, <i>écart à la densité moyenne</i> en fonction de la distance au littoral. .	78
	(a) Coefficient de détermination.	
	(b) Coefficient de la variable.	
2.12	Modèles de régressions, <i>écart à la densité moyenne</i> en fonction de la distance au littoral. .	79
	(a) Résidus en 1793, $R^2 = 0,014$.	
	(b) Résidus en 1901, $R^2 = 0,005$.	
	(c) Résidus en 1921, $R^2 = 0,006$.	
	(d) Résidus en 1946, $R^2 = 0,010$.	
	(e) Résidus en 1975, $R^2 = 0,006$.	
	(f) Résidus en 1999, $R^2 = 0,009$.	
2.13	<i>Boîtes de Tuckey</i> selon les classes issues de la typologie.	81
	(a) Altitude.	
	(b) Distance au littoral.	
2.14	Seconde typologie : cartes.	86
	(a) Typologie initiale.	
	(b) Subdivision de la classe 1.	
	(c) Subdivision de la classe 2.	
	(d) Subdivision de la classe 3.	
	(e) Subdivision de la classe 4.	
	(f) Subdivision de la classe 5.	
2.15	Seconde typologie : autres documents.	87
	(a) Écart moyen à la densité moyenne par classe (variation de forme) et sous-classes (variation de couleur, les couleurs correspondent à celles figurées sur les cartes 2.14)	
	(b) Évolution du <i>coefficient de variation</i> suite à la seconde classification, en fonction de la sous-classe et de la date. 1, (-1 ; -0,5) ; 2, (-0,5 ; -0,25) ; 3, (-0,25 ; 0) ; 4, (0 ; 0,25) ; 5, (0,25 ; 0,5) ; 6, (0,5 ; 5), avec la borne supérieure exclue.	
A.1	Un exemple de notice : la commune d'Eaux-Bonnes (Pyrénées-Atlantiques)	92

Liste des tableaux

1.1	Quelques exemples d'organisation du recensement.	19
1.2	Essai de synthèse de la qualité des recensement.	20
1.3	Tests de normalité.	30
1.4	Modèles locaux <i>vs</i> modèles globaux.	42
2.1	Variables environnementales : Sources, traitements et distribution.	60
2.2	Part de la superficie occupée (en %) (et sa variation).	64
2.3	Part de la superficie occupée (en %) (et sa variation).	65
2.4	Part de la superficie occupée (en %) (et sa variation).	66
2.5	Analyse de la variance (1).	82
2.6	Analyse de la variance (2).	82
2.7	Part de la variance expliquée par la seconde typologie.	85
A.1	Définition de la population utilisée lors du recensement correspondant.	93
A.2	Population retenue lors de la construction de la base de données <i>cassini.ehess</i>	94
B.1	<i>Relevé général de la population des départements du royaume en 1821</i>	97