



HAL
open science

Les effets de contexte en géographie. Des fondements théoriques à la modélisation statistique

Thierry Feuillet

► **To cite this version:**

Thierry Feuillet. Les effets de contexte en géographie. Des fondements théoriques à la modélisation statistique. Géographie. Université Lyon 3 Jean Moulin, 2021. tel-03357283

HAL Id: tel-03357283

<https://hal.science/tel-03357283>

Submitted on 28 Sep 2021

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.



Habilitation à diriger des recherches

UNIVERSITÉ DE LYON 3

Présentée par

Thierry FEUILLET

Volume 1 - Essai inédit

Les effets de contexte en géographie

Des fondements théoriques à la modélisation statistique

Dossier présenté le 24 septembre 2021

Jury :

<i>Rapporteurs :</i>	Lena SANDERS	- DR CNRS - UMR Géographie-cités
	Jean-Michel OPPERT	- PU-PH - Sorbonne Université (Paris)
	Hugo PILKINGTON	- PU - Université Paris 8 (Saint-Denis)
<i>Garant :</i>	Étienne COSSART	- PU - Université Jean Moulin Lyon 3
<i>Examineurs :</i>	Virginie CHASLES	- PU - Université Jean Monnet Saint-Etienne
	Arnaud BANOS	- DR CNRS - UMR IDEES, IRED

Remerciements

Certaines rencontres modifient significativement le cheminement des choses, et je dois ainsi à quelques personnes en particulier d'en être à cette étape aujourd'hui.

Il s'agit d'abord de Denis Mercier et de Pascal Bertran, qui m'ont permis de finaliser ma thèse. Leur confiance et leurs encouragements m'ont incité à poursuivre dans la recherche à un moment où c'était plutôt compromis, et ont finalement permis d'initier la dynamique qui m'a mené à l'HDR. Juste après la thèse, les missions en Islande entre 2010 et 2012, pilotées par Denis Mercier, ont constitué un magnifique tremplin dans mes activités de recherche et ont réellement marqué le point de départ de mes réflexions sur les effets contextuels. Un grand merci, Denis, pour cette aventure humaine et scientifique.

L'autre moment essentiel dans mon parcours a été ma rencontre en 2013 avec Héléne Charreire et Jean-Michel Oppert, dont la générosité, l'humanité et la modestie m'ont réconcilié avec le monde académique. Ils m'ont fait découvrir la géographie de la santé (dans un premier temps à travers le projet ACTI-Cités, puis Spotlight), domaine si propice aux analyses spatiales et contextuelles, dans lequel je continue de m'épanouir scientifiquement. Cela a été une réelle chance de les rencontrer, et c'est toujours un vrai plaisir de travailler en leur compagnie. Jean-Michel Oppert m'a beaucoup inspiré par sa capacité à hiérarchiser de façon extrêmement pertinente les questionnements et raisonnements scientifiques, et par sa rigueur incroyable dans l'écriture des articles, aucun mot n'étant jamais laissé au hasard. J'ai apprécié toutes ces heures passées à peaufiner l'écriture de ces papiers, et j'ai même fini par admettre les bienfaits des coupes sévères... Héléne, c'est la générosité, l'altruisme, la fiabilité, l'efficacité. Merci beaucoup à tous les deux pour votre confiance et votre fidélité.

Parfois aussi, ce sont des bouts de discussion informelle avec certains collègues qui nous aident à y voir plus clair. Celles ou ceux avec qui le partage de longueur d'onde scientifique est absolument évident. À ce titre, je pense en particulier à Paul Salze et à nos discussions sur l'hétérogénéité spatiale, à Besançon ou à Strasbourg, qui ont fait murir les idées développées dans cette HDR. Je dois aussi beaucoup à Hadrien Commenges pour m'avoir aidé à être autonome sur R, ce qui m'a ouvert de multiples horizons. Depuis mon recrutement en 2015 à Paris 8, j'ai également bénéficié de précieux échanges avec les collègues du quotidien – Jean-François Valette et Johan Milian en particulier – et qui m'ont enrichi, dans toutes les dimensions de notre métier d'enseignant-chercheur.

Je remercie évidemment sincèrement Étienne Cossart, dont la finesse et l'intelligence des conseils m'ont largement aidé à clarifier mes idées et mes propos. Il était le garant parfait pour ce travail : suivant mes travaux depuis 2010, partageant mon affinité pour l'analyse spatiale quantitative, et étant spécialiste de géographie

physique tout en restant largement ouvert aux autres branches de la géographie. Merci pour toutes les discussions que nous avons eues, qui m'ont fait progresser.

Je suis très reconnaissant envers les membres du jury d'avoir accepté de juger ce dossier d'HDR. Il s'agit d'abord de Lena Sanders, dont les publications m'ont beaucoup inspiré (comme beaucoup d'autres géographes quantitativistes!), et avec qui je suis fier de discuter de ce travail. Arnaud Banos est naturellement un autre nom incontournable de l'analyse spatiale quantitative. C'est toujours un réel plaisir de lire ses textes si limpides. Merci également à Virginie Chasles – que je connais malheureusement encore peu, mais dont j'apprécie beaucoup la pertinence et la clarté des propos sur les questions d'inégalités spatiales de santé – et à Hugo Pilkington, que je côtoie au quotidien depuis 2015 à Paris 8 et avec qui j'échange régulièrement de façon fructueuse en géographie de la santé et en statistique. Enfin, merci encore à Jean-Michel Oppert, dont la présence dans ce jury me semblait évidente et obligatoire.

Merci enfin à ma compagne Julie pour m'avoir aidé à débloquer bien des raisonnements... À ma fille Zélie, ce ne sont pas des remerciements que je veux adresser, mais des excuses... pour toutes les fois où le travail est passé avant elle. J'en ai réellement pris conscience quand, à 6 ans, elle m'a dit : "Laisse-moi tranquille, je travaille", en faisant semblant de tapoter sur un faux ordinateur qu'elle avait préalablement dessiné, avec tous les détails. Promis, demain, j'arrête...

Sommaire

Introduction	1
1 Les effets de contexte dans leur cadre théorique et conceptuel	7
1.1 Définition d'un contexte spatial	8
1.2 Structuration de l'espace par systèmes de contextes spatiaux	9
1.3 Le rôle de l'hétérogénéité et de la dépendance spatiales dans la structuration des contextes spatiaux	16
1.4 Apports de la considération des systèmes de contextes spatiaux dans les analyses géographiques	25
2 Les études contextuelles, leur modélisation statistique et leurs limites	33
2.1 Le cheminement du déterminisme au contextualisme : historique de l'étude des liens entre l'humain et son environnement	33
2.2 La modélisation statistique des effets contextuels en analyse transversale : modèles contextuels, multiniveaux et spatiaux	41
2.3 Typologie des erreurs associées à la modélisation statistique des effets contextuels	52
3 De la non-stationnarité spatiale à une géographie réaliste des contextes	61
3.1 Justification et présentation du protocole méthodologique combinant modèles GWR et multiniveau	62
3.2 Application du modèle hybride multiniveau-GWR à la géographie de l'obésité en région parisienne	69
3.3 Implications des résultats en géographie et en santé publique	82
Conclusion	89

Introduction

Plus on pédale moins fort, moins on avance plus vite

Coluche

Si mon profil peut aujourd'hui être considéré comme celui d'un spécialiste d'analyse spatiale quantitative¹, ma formation initiale ne m'avait pourtant pas pré-disposé à cela. En effet, je n'ai disposé que de très peu – si ce n'est d'aucun – enseignement en géographie théorique ou en analyse spatiale durant tout mon cursus universitaire, et quasiment aucun non plus en statistiques. Mon premier contact avec le quantitatif a été très tardif, en fin de thèse (2009). Une des raisons en est que mon premier directeur de thèse ne voulait pas en entendre parler, alors que le sujet s'y prêtait pourtant assez bien. Je me souviens d'un sentiment de frustration, car je ressentais déjà des affinités avec les approches quantitatives, qui me semblaient pouvoir permettre d'inscrire les hypothèses de travail et les interprétations d'analyses dans un cadre rigoureux. J'avais souffert en début de thèse de ne pas être capable d'imaginer une stratégie d'échantillonnage spatial pertinente, de ne pas vraiment savoir que faire de mes observations, de ne pas savoir comment valider ou invalider mes hypothèses, de ne pas aborder correctement la question de la représentativité spatiale de mes résultats et celle des croisements d'échelles, autant de questions que les approches quantitatives m'ont ensuite permis de résoudre. Le premier grand tournant de ma carrière naissante a été le moment où, alors qu'il relisait mes premiers résultats, mon second directeur de thèse (Pascal Bertran, qui m'a suivi pour la quatrième et dernière année en co-direction avec Denis Mercier) m'a dit : « C'est bien tout ça, mais ça manque un peu de quantitatif ». Je ne suis jamais passé aussi vite de la frustration au soulagement, et bientôt à l'épanouissement. Néanmoins, mon retard était difficile à rattraper dans la contrainte du calendrier doctoral. C'est donc juste après la thèse que j'ai commencé réellement, en autodidacte, à m'investir dans la géographie quantitative.

La découverte des statistiques spatiales s'est faite dans la foulée. Le retour d'un relecteur en 2011 sur un article traitant de l'analyse spatiale des névés pyrénéens (oui, cela a un intérêt pour certains) m'a offert l'opportunité de découvrir une méthode en particulier, la GWR, dont l'objet central (l'hétérogénéité spatiale) constituera en quelque sorte le fil conducteur de mes préoccupations scientifiques futures. Critiquant l'absence de considération de la non-stationnarité spatiale dans mes analyses, ce relecteur m'a signalé cette méthode comme un recours opportun. Je me souviens de la fascination que m'ont procuré les premières cartes des coefficients de régression que j'ai dessinées. D'un coup, toute la complexité des processus

1. Analyse spatiale entendue dans sa définition large, regroupant l'ensemble des modèles et des concepts permettant de mettre en œuvre des techniques et des méthodes visant à caractériser les configurations et les propriétés de l'espace [Pumain et Saint-Julien, 2010].

géographiques et de leur empreinte spatiale sautait aux yeux. Il ne s'agissait plus seulement de constater la variabilité spatiale d'une variable, comme le montrerait n'importe quelle carte habituelle, mais celle de la corrélation entre variables. La corrélation peut très bien être la même partout, si les variables en jeu covarient de façon identique dans l'espace. Par exemple, la relation entre la topographie et la persistance des névés est à peu près la même partout. Mais la variation de la covariation : voilà la réelle hétérogénéité spatiale ! La relation entre l'altitude et la présence de névés varient selon la localisation ! J'ai appliqué pour la première fois cette méthode pour décrire la non-stationnarité des facteurs de localisation des glissements paraglaciaires en Islande du nord [Feuillet *et al.*, 2014], dans le cadre de ma première expérience post-doctorale dans ce milieu subarctique, supervisée par Denis Mercier, et en collaboration avec Etienne Cossart et Armelle Decaulne (figure 1). Passée la fascination procurée par ces cartes de corrélations, les interrogations sur les causes de ces structures spatiales sont apparues. Pourquoi la relation étudiée est-elle positive ici ? Pourquoi est-elle négative précisément dans ce secteur ? Parce qu'elle dépend du contexte local, soit, mais pourrait-on appréhender cet effet du contexte sur les relations statistiques de façon protocolaire et systématique, dans une volonté d'établir un cadre explicatif suffisamment englobant pour s'appliquer à toute thématique, à tout objet d'étude ?

En me plongeant dans la littérature liée à la GWR, j'ai rapidement pu constater que les articles du trio fondateur (par exemple Brunsdon *et al.* [1996]; Fotheringham *et al.* [2002]), de même que la grande majorité des études empiriques mobilisant la méthode, restaient très lapidaires sur les causes de la non-stationnarité spatiale, n'insistant en général que sur sa description et ses implications sur l'objet étudié. Une explication récurrente dans les articles de Stewart Fotheringham est de considérer simplement que la non-stationnarité résulte du caractère intrinsèquement hétérogène de l'espace, faisant écho en cela à la fameuse part explicative de l'espace : "For whatever reason, some relationships are intrinsically different across space. Perhaps, for example, there are spatial variations in people's tastes or attitudes or there are different administrative, political, or other contextual issues that produce differing responses to the same stimuli across space" [Fotheringham *et al.*, 1997, p. 60]. C'est, d'une certaine façon, également ce message qu'envoie Michael Goodchild dans ses articles passionnants – presque philosophiques – sur l'hétérogénéité spatiale. Cette explication semble néanmoins être une façon de botter en touche, car elle évite de s'interroger sur les causes réelles de l'hétérogénéité de l'espace. L'espace est un médiateur, il n'est pas un facteur en soi. Les mécanismes de cette médiation nourrissent d'ailleurs tout l'intérêt de la géographie en tant que discipline. La lecture de François Durand-Dastès, plus tardive, m'a rassuré sur la véracité de ce constat.

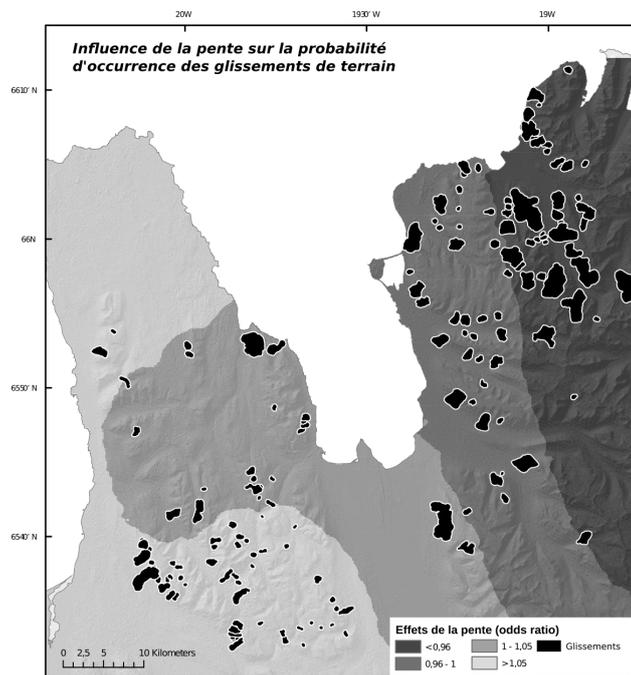


FIGURE 1 – Influence de la pente sur la probabilité d’occurrence des glissements de terrain paraglaciers en Islande du nord, révélée par un modèle GWR. Cet effet est non-stationnaire : il est négatif à l’est du fjord, mais positif à l’ouest. Dès lors, la question de la délimitation des contextes spatiaux et de la caractérisation de leur spécificité devient un enjeu majeur dans la compréhension des logiques de localisation de ces modelés. Figure extraite de [Feuillet *et al.* \[2014\]](#).

C’est à ce moment de mes interrogations que j’ai, par un surprenant concours de circonstances, atterri en tant que post-doctorant (en 2013) dans le domaine de la géographie de la santé. Ce contrat a été réalisé dans le cadre du projet ACTI-Cités – piloté par Jean-Michel Oppert, en collaboration avec Hélène Charreire – et visait à étudier les déterminants environnementaux de la mobilité active en milieu urbain à l’aide de la cohorte NutriNet-Santé. On pourrait croire à première vue que le passage entre la géomorphologie des milieux froids à l’épidémiologie nutritionnelle est verglacé, mais cette ouverture thématique a au contraire constitué une merveilleuse opportunité d’approfondir ces questionnements. J’y ai découvert toute la littérature sur les effets de quartiers, qui a compensé toute celle qui n’existait pas sur les effets de site en géomorphologie. La lecture des articles de [Macintyre *et al.* \[2002\]](#) et de [Cummins *et al.* \[2007\]](#) a constitué en particulier une sorte de révélation, faisant naître des connexions entre des idées qui restaient jusqu’alors floues et éparpillées. Ces auteurs confirmaient ma première impression selon laquelle chacun reconnaît que les contextes spatiaux agissent explicitement sur de multiples éléments, mais qu’ils sont trop souvent considérés comme des boîtes noires, si bien que les mécanismes de production de l’hétérogénéité locale restent mystérieux.

Contrairement à ce que je pensais dans un premier temps, les effets de contexte sont plus complexes en géographie humaine qu'en géographie des milieux peu anthropisés. Tout simplement car dans ces derniers, les agents qui créent et composent des contextes locaux (des roches, des pentes, des conditions climatiques, etc.) se comportent de façon à peu près prévisibles, et que par conséquent leurs interactions, bien que parfois complexes évidemment, restent intelligibles. Les choses se compliquent quand ces contextes relèvent en partie des comportements des humains qui les composent, car ces comportements sont infiniment plus variables et chaotiques, autant dans leur état que dans leurs liens avec l'espace. Ce qu'a révélé à mes yeux l'article de *Cummins et al.* [2007] et qui constitue, il me semble, la richesse de la géographie de la santé, c'est que les propriétés des lieux sont liées de manière systémique aux comportements humains, signifiant qu'ils se co-construisent sans qu'il ne soit possible d'établir des causalités unidirectionnelles dans un sens ou dans l'autre. Ainsi, les lieux englobent des caractéristiques environnementales, des humains, et leurs relations, le tout sans hiérarchie ni antériorité. Ce sont ces lieux, et leur dynamique spécifique, qui expliquent la non-stationnarité spatiale. Reste donc à les délimiter, les définir, et caractériser leurs effets.

L'épidémiologie et la géographie de la santé m'ont apporté un autre élément considérable, du point de vue de la quantification des effets de contexte cette fois : la modélisation multiniveau. Cette méthode, très peu employée en géographie physique, est au contraire incontournable dans la modélisation statistique des effets de contexte en sciences sociales (probablement en grande partie parce que la méthode repose sur une vision discrète de l'espace, alors que la géographie physique a pour habitude de le considérer dans sa dimension continue). En m'appropriant cette méthode, notamment à travers la lecture des articles si didactiques de Kelvin Jones en géographie et d'Ana Diez-Roux en épidémiologie, l'idée d'une hybridation potentielle avec la GWR a commencé à germer. La GWR est souvent utilisée comme une méthode exploratoire, descriptive, alors que le multiniveau établit statistiquement les effets de contexte, dans un cadre inférentiel plus solide que la GWR. Néanmoins, le multiniveau utilise presque systématiquement des contextes aux limites basées non pas sur les données impliquées dans les analyses, mais sur des maillages pré-établis, le plus souvent des mailles administratives. Si les contextes locaux produisent la non-stationnarité révélée par les GWR, alors pourquoi ne pas plutôt utiliser les GWR pour délimiter les contextes des modèles multiniveaux ? Cette idée m'apparaît d'autant plus séduisante qu'elle semble pouvoir constituer un pont entre l'analyse de processus continus dans l'espace (pour lesquels des méthodes comme la GWR sont particulièrement bien adaptées) – typiquement préférée en géographie physique donc – et l'analyse de processus discrets au sein de contextes spatiaux (préférée en géographie humaine, et que la modélisation multiniveau autorise). Combiner les deux approches reviendrait donc en quelque sorte à relier une géographie des surfaces à une géographie des lieux.

L'objectif de cet essai est d'établir les fondements de cette hybridation méthodologique, puis de l'appliquer à la géographie de l'obésité en Île-de-France (la géographie de la santé se prêtant tellement bien à ces questionnements, comme nous l'expliquerons au cours de ce travail). Mais avant d'aborder ces aspects méthodologiques et empiriques en troisième partie, les deux premières seront consacrées au volet théorique. Comme le dit si bien M. Goodchild, "if we let the data speak for themselves, we must be aware at the same time that they are not necessarily speaking for geography" [Goodchild, 2016]. J'entends cette phrase comme un avertissement à ne pas déconnecter les données spatiales des théories géographiques. C'est la rigueur du cadre théorique qui détermine la pertinence du choix des méthodes d'analyse et de leur dimension heuristique. Au-delà de la nécessité de définir les concepts avant toute chose – d'autant plus quand ils circulent à travers plusieurs disciplines comme c'est le cas en analyse contextuelle – il m'a donc semblé indispensable de contribuer à proposer les bases d'une théorie géographique des effets de contexte, qui serait valable pour toute thématique. L'augmentation récente des capacités computationnelles, de l'accès aux données, mais également de la disponibilité des outils, permet presque aujourd'hui à n'importe qui de produire des analyses quantitatives impliquant des données spatialisées. En parallèle, la compétition entre les éditeurs scientifiques – et entre les scientifiques eux-mêmes – renforce la course aux publications², probablement au détriment de la qualité scientifique d'un certain nombre d'entre eux. Or, il est plus facile et plus rapide de publier des résultats d'analyses quantitatives que les fondements théoriques des processus que l'on étudie. Il en résulte que parmi les centaines et les centaines d'articles publiés chaque année sur la modélisation statistique de données spatiales et contextuelles, peu finalement confrontent leurs résultats à une quelconque théorie géographique qui permettrait de les mettre en perspective en les positionnant dans un espace conceptuel. Il n'est pas suffisant de ne s'en remettre qu'au caractère intrinsèquement hétérogène de l'espace !

Voilà pourquoi la première partie de cet essai est consacrée à une réflexion théorique sur la notion de contexte spatial en géographie, des mécanismes de leur genèse aux principes de leurs effets. En s'appuyant sur une littérature variée d'un point de vue disciplinaire, l'objectif est principalement de montrer comment cette notion de contexte spatial facilite les convergences sémantiques et théoriques, y compris pour les études n'ayant pas mobilisé explicitement ce concept. La volonté est de montrer dans quelle mesure les contextes et leurs effets représentent un dénominateur commun à tous ces travaux et une clé de lecture particulièrement pertinente de l'espace et de ses structures. Bien qu'inspiré des travaux sur les effets de quartiers, ce cadre théorique est résolument pensé de façon à pouvoir s'appliquer à n'importe quelle thématique géographique, de la géomorphologie à la géographie sociale. Cette volonté traduit un positionnement scientifique plus large – probable-

2. Entre 1996 et 2018, le nombre d'articles scientifiques publiés chaque année dans le monde est passé de 972 000 à 2 556 000, et les articles chinois ont augmenté à eux seuls de près de 1500 % sur la même période [NSF, 2020].

ment issu de mes propres trajectoires scientifiques bien sûr – qui consiste à vouloir dépasser les cloisonnements internes à la discipline géographique, en particulier cette distinction entre géographie physique et géographie humaine malheureusement toujours aussi présente aujourd’hui qu’il y a un siècle. J’adhère ainsi à la volonté de [Bunge \[1962\]](#), qui espérait qu’un jour la géographie ne serait plus divisée qu’entre points, lignes et surfaces, et pas entre thématiques. Ce qui soude l’ensemble des géographes, c’est la volonté de comprendre les structures spatiales, de caractériser et de comparer des lieux, quel que soit l’objet étudié. Il n’y a donc pas de raison de penser qu’un cadre théorique décrivant les mécanismes de production et de fonctionnement des contextes spatiaux ne puisse pas être valable pour toute thématique.

La deuxième partie est en quelque sorte une revue de la littérature, entre épistémologie et état de l’art des études contextuelles, quantitatives mais pas seulement, en géographie et dans les disciplines connexes. Cette partie est donc une bonne transition entre la première – théorique – et la dernière – empirique. Nous concluons enfin cet essai sur les perspectives de recherche que ces réflexions auront entraînées, comme autant de projets pour l’avenir qui contribueront à perfectionner l’analyse des effets de contexte en géographie.

Les effets de contexte dans leur cadre théorique et conceptuel

Un enjeu majeur de la géographie consiste à questionner les logiques de localisation des phénomènes spatialisés à travers la caractérisation et l'analyse des structures spatiales sous-jacentes. Ces questions de localisation consistent en trois interrogations majeures :

- (i) Où se produisent ces phénomènes ? Cela implique une analyse des occurrences ou des densités d'objet géographiques (par exemple : où observe-t-on des glissements de terrain ?).
- (ii) Leurs caractéristiques sont-elles spatialement variables ? Cela implique une analyse des quantités (ou des qualités). Par exemple : où les glissements de terrain sont-ils les plus massifs ou les plus meurtriers ?
- (iii) Les facteurs de localisation ou de variation des quantités sont-ils les mêmes partout ? Cela implique cette fois une analyse des co-occurrences ou des covariations. Par exemple : le commandement du versant ou la nature de la roche expliquent-ils systématiquement l'occurrence ou la masse des glissements ?

Une des complexités de la géographie – qui en fait aussi sa richesse – provient du caractère fondamentalement hétérogène de l'espace. Cette hétérogénéité spatiale, ubiquiste, a pour conséquence de faire varier ces occurrences, ces quantités et ces covariations au gré des contextes spatiaux. Ainsi, une façon possible d'aborder ces logiques de localisation consiste à considérer la surface terrestre comme une juxtaposition de ces contextes spatiaux, et à supposer que c'est par la localisation et la compréhension du fonctionnement de ces contextes que les logiques de distribution spatiale de toute observation d'échelle inférieure peuvent être appréhendées. Cette première partie vise à présenter en détail cette vision de l'espace, vision ayant servi de cadre aux raisonnements et analyses subséquentes de ce travail. Il est en effet d'autant plus souhaitable d'inscrire ces raisonnements dans un cadre théorique et conceptuel bien circonscrit, que ce champ transdisciplinaire (les effets de contexte) véhicule un nombre important de concepts nomades et implique différentes théories géographiques.

1.1 Définition d'un contexte spatial

Il convient de débiter par une définition de la notion de contexte spatial et une justification du choix d'en faire le concept central du présent travail. Les contextes spatiaux peuvent être définis comme des portions d'espace dont la proximité spatiale des éléments les constituant peut engendrer des dynamiques internes spécifiques – les effets de contexte – qui tendent en retour à renforcer leur existence et leur singularité. Un effet de contexte est un effet d'appartenance : il existe en effet dès lors que le niveau des attributs des objets spatiaux composant le contexte est affecté par le fait même d'appartenir à ce contexte. Cette définition est équivalente à celle du contexte territorial proposée par [Grasland \[2001\]](#)¹. L'auteur réserve la notion de contexte spatial à des zones non bornées, et qui impliquent donc des effets de localisation (distances) et non d'appartenance. Cette nuance entre une vision discrète (contexte territorial) et une vision continue (contexte spatial) de l'espace – et le passage de l'une à l'autre – constituent d'ailleurs un élément essentiel de ce travail et sera largement discutée par la suite, en particulier dans la partie 3. La notion de contexte spatial a été préférée à celle de contexte territorial dans le cadre de ce volume, d'une part pour son acception plus englobante (le concept de territoire implique un espace politiquement construit et délimité par des frontières, adapté à la géographie humaine mais moins aux objets de la géographie physique) et d'autre part car la notion même de contexte sous-entend déjà l'existence d'une zone délimitable.

La notion de lieu est proche, mais est également plus réductrice, raison pour laquelle elle n'a pas été retenue. Elle est plus réductrice à double titre :

- (i) D'un point de vue scalaire d'abord, car la notion de lieu se restreint à une acception plutôt locale, comme une place, une rue ou un quartier, quand un contexte spatial peut tout autant désigner une région entière.
- (ii) D'un point de vue disciplinaire ensuite : le lieu est entendu comme une construction sociale [[Clerc, 2004](#)], une aire géographique produite, identifiée et contextualisée par l'Homme, si bien qu'il tend à exclure de son champ les espaces peu ou pas anthropisés, ce qui est contraire à l'objectif de généralité de ce travail.

Néanmoins, le lieu est bien considéré comme un type de contexte spatial :

"Place is a context. In social science place is used as a means to define context, the geographic area within which humans live their lives and, more precisely, the relationships between people and their environment"
[[Roche, 2016](#), p. 3]

Les effets de lieu sont donc un type d'effets de contexte, au même titre que les effets de quartiers². A l'opposé, la notion de région, entendue comme produit

1. <http://grasland.script.univ-paris-diderot.fr/agreg/module7/index.htm>

2. Sur cette question terminologique dans le domaine des effets de quartier, [Petrović et al. \[2020\]](#)

du processus de régionalisation, c'est-à-dire de découpage de l'espace en régions homogènes [Pumain et Saint-Julien, 2010], est associée à une échelle macro, qui de fait représente mal les contextes locaux.

Les contextes spatiaux

Les contextes spatiaux peuvent être définis comme des portions d'espace dont la proximité spatiale des éléments les constituant permet d'engendrer des dynamiques internes spécifiques – les effets de contexte – qui tendent en retour à renforcer leur existence et leur singularité. Il y a effet de contexte dès lors que le niveau des attributs des objets spatiaux composant le contexte est affecté par le fait même d'y appartenir [Grasland, 2001].

L'autre concept candidat est celui de système spatial. Il peut être défini comme « un ensemble d'entités spatiales (les éléments) entre lesquelles existe un ensemble de relations » [Sanders, 1999, p. 26]. Bien entendu, la notion de système est centrale dans cette approche de l'espace. Les contextes spatiaux en sont, puisque ce qui les caractérisent – les effets de contexte – sont principalement issus des interactions entre les éléments (humains ou naturels) qui les constituent (*cf.* §1.3.2). Néanmoins, la notion de système désignent aussi bien le réseau de contextes en interaction. Les contextes sont alors vus comme des sous-systèmes, rejoignant ainsi l'interprétation très englobante du système spatial selon Pinchemel et Pinchemel [1997]. Dans ce travail, pour éviter toute ambiguïté terminologique, la notion de système sera réservée à la qualification d'un ensemble de contextes spatiaux en interaction dans un environnement (systèmes de contextes spatiaux).

1.2 Structuration de l'espace par systèmes de contextes spatiaux

Cette vision modélisée de l'espace, ainsi que tous les concepts qu'elle associe, constitue une synthèse revisitée et contextualisée de lectures issues de champs variés, mais principalement inspirées des travaux de géographes, sociologues et épidémiologistes sur l'hétérogénéité spatiale, les systèmes complexes et les effets de quartiers/lieux. Il s'inspire en particulier des réflexions de Durand-Dastès [1992, 1984] sur les mécanismes de production des localisations (facteurs de stabilité, interaction spatiale et processus diachroniques), mais en distinguant explicitement les forces productrices de contextes (effets verticaux) des forces qui les maintiennent

arrivent d'ailleurs à la même conclusion à propos de la pertinence d'un terme plus générique : "the conceptualizations of neighbourhood in neighbourhood effects research should more closely match the underlying mechanisms. This implies that the term 'spatial context effects' more closely matches what we try to understand than the term 'neighbourhood effects'".

ou les renforcent (effets horizontaux), dans une logique systémique. Il fait également écho dans sa structure aux systèmes de villes de Christaller, repris par de nombreux thématiciens, dont Denise Pumain en particulier. La différence est ici qu'il s'applique à des objets géographiques et des échelles variés, pas seulement les villes. Ce cadre théorique n'a donc ni la prétention d'être inédit, ni celle d'être exclusif évidemment, mais présente néanmoins l'intérêt de relier de façon logique les différentes composantes d'un système de mécanismes complexes, d'intégrer des lois et des concepts fondamentaux de la géographie théorique et de l'analyse spatiale (hétérogénéité, dépendance, interaction, échelle, etc.), et enfin d'être robuste aux changements d'échelles par sa composante fractale.

Ce modèle de systèmes de contextes spatiaux est illustré de façon synthétique dans la figure 1.1.

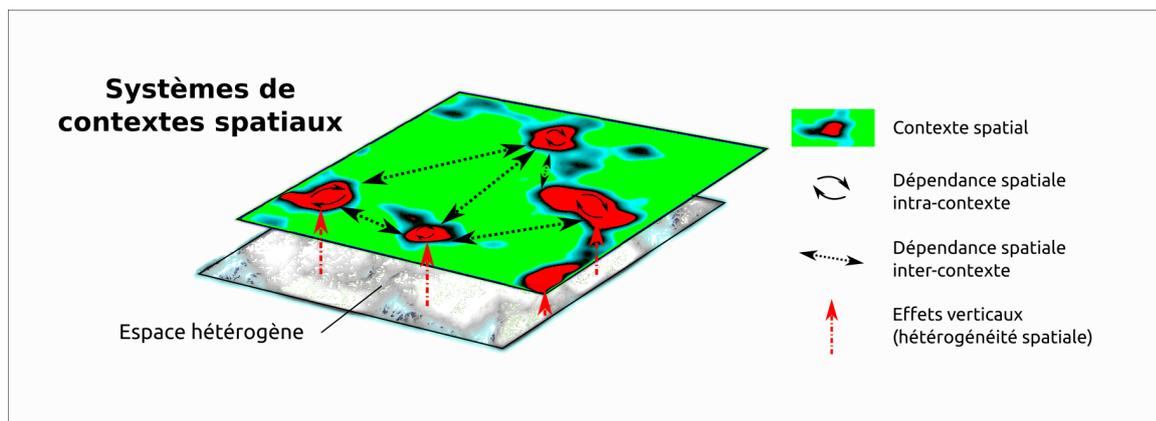


FIGURE 1.1 – Représentation schématique des systèmes de contextes spatiaux dans un espace hétérogène.

Le schéma représente un environnement (portion d'espace dont la dimension dépend de l'objet étudié), dans lequel sont répartis des contextes spatiaux connectés entre eux par des interactions spatiales. La localisation initiale de ces contextes est principalement déterminée par des processus de long terme, qui structurent en profondeur l'hétérogénéité de l'espace, tandis que les contextes eux-mêmes sont animés d'effets propres. Cette structuration de l'espace par contextes se répète à travers les échelles, les contextes pouvant par exemple désigner des rues, des quartiers, des villes ou des régions ; des coteaux, des versants, des bassins-versants ou des massifs. Le choix de l'échelle d'analyse dépend de la nature du phénomène étudié. L'ensemble constitue un système maintenu cohérent par l'ensemble de ces interactions.

Une analogie pourrait être faite avec le système atmosphérique. L'atmosphère constitue l'environnement et les centres de pression (anticyclones et dépressions)

les contextes spatiaux. La localisation de certains de ces centres de pression sont déterminés par les caractéristiques de l'espace sous-jacent, comme c'est le cas pour l'anticyclone de Sibérie par exemple, ou par les interactions entre cellules (centres de pression dynamiques). Ces centres de pression sont des entités qui fonctionnent de façon subautonome, avec une dynamique propre (tourbillons, fronts, subsidence, ascendance, etc.), mais qui sont néanmoins connectés par des échanges de flux (dont l'intensité est fonction de la proximité et de la puissance des centres, exactement comme selon les principes du modèle gravitaire) maintenant leur équilibre, et donc la stabilité de l'ensemble du système. Cette image peut être transposée à un système de quartiers, un système de villes ou un système de régions.

Entrons maintenant plus en détail dans les mécanismes de production et de maintien de ces contextes spatiaux. Ces mécanismes sont les trois principaux fondements de la géographie théorique et de l'analyse spatiale : l'hétérogénéité spatiale, la dépendance spatiale et les lois d'échelles. La géographie théorique et quantitative tente depuis les années 1960 d'établir des lois géographiques. Rappelons qu'une loi, dans son acception scientifique, est un postulat qui peut être vérifié par l'observation empirique des faits. Une loi n'est pas nécessairement toujours vraie – même en physique – mais elle l'est suffisamment souvent pour pouvoir être considérée comme une norme, ou tout du moins comme une structure persistante, régulière. Elle n'est pas non plus nécessairement causale. Comme le rappelait Miller [2004], la loi de la gravité de Newton n'est pas une explication. Le principe même de l'existence de lois en sciences sociales est toujours discuté. Dans une conférence organisée à Paris en décembre 2016 par Claude Grasland sur le thème des lois en géographie, Michael Goodchild rapportait la réponse du physicien Ernest Rutherford à cette question : “the only result that can possibly be obtained in the social sciences is : some do, and some don't”. Néanmoins, les géographes s'y exercent donc, et un certain consensus existe aujourd'hui sur l'existence de deux principales lois en géographie :

- (i) La dépendance spatiale constitue pour la plupart des géographes théoriciens et quantitativistes, et plus largement pour tous les chercheurs se réclamant des sciences de l'information géographique, la première loi de la géographie. Elle est attribuée à Waldo Tobler, qui écrivait en 1970 en relation avec ses études de simulation de la croissance urbaine : “I invoke the first law of geography : everything is related to everything else, but near things are more related than distant things” [Tobler, 1970]. La dépendance spatiale traduit donc le lien persistant existant entre la ressemblance et la distance : deux observations ou deux lieux proches ont une probabilité de se ressembler qui est plus forte que s'ils étaient éloignés. C'est cette loi qui régit le modèle gravitaire et l'intensité des interactions spatiales, entre autres. Sa traduction statistique est l'auto-corrélation spatiale, propriété selon laquelle les corrélations sont plus fortes à mesure que la distance entre les observations diminue. Elle implique dans les modèles statistiques une sur-estimation du nombre d'observations, puisque théoriquement supposées indépendantes.

- (ii) L'hétérogénéité spatiale est considérée comme la deuxième loi de la géographie [Goodchild, 2004]. Dans son sens très général, l'hétérogénéité spatiale se réfère à la non-uniformité (ou la diversité) des phénomènes géographiques, et peut être résumée par le fait qu'aucun lieu ne peut représenter à lui seul la surface de la Terre [Goodchild, 2004; Anselin, 1989]. Comme l'écrit Goodchild [2009], sans hétérogénéité spatiale, "European explorers would not have found new worlds, but new samples of old ones". Wu [2004, p. 125] la définit également de façon limpide ainsi : "Spatial heterogeneity is ubiquitous across all scales and forms the fundamental basis of the structure and functioning of landscapes, be they natural or cultural". Ainsi, l'hétérogénéité spatiale est considérée comme ubiquiste et applicable à des domaines extrêmement variés (topographie, sol, climat, paysages, distribution des espèces, urbanisation, industrialisation, etc.). Sa traduction statistique est la non-stationnarité spatiale, c'est-à-dire l'instabilité des paramètres statistiques de variables géographiques aléatoires, que ces paramètres soient des moyennes, des variances ou des covariances. Ainsi, toutes distributions ou caractéristiques d'observations spatialisées varient d'un lieu à l'autre. Elle implique dans les modèles statistiques des problèmes de représentativité spatiale des observations.

Le troisième grand principe de la géographie, s'il en est un, pourrait être représenté par le principe fractal, tel que l'a justement proposé Goodchild dans ce même séminaire de 2016. Ce concept désigne la répétition des mêmes structures à travers les échelles, et implique de devoir systématiquement préciser la résolution spatiale des observations que l'on mène. Il est manifeste qu'en géographie, les systèmes spatiaux tendent souvent à s'emboîter, et que pris au sens large, ce principe fractal constitue finalement une loi scalaire qui caractérise ces emboîtements.

**Les deux principales lois de la géographie :
la dépendance et l'hétérogénéité spatiales**

La dépendance spatiale (déclinaison statique de la loi de Tobler) stipule que tout interagit avec tout, mais que les choses proches se ressemblent davantage que les choses éloignées. Par exemple, un quartier pauvre a plus de chances d'être entouré d'autres quartiers pauvres que de quartiers riches.

L'hétérogénéité spatiale stipule quant à elle que tout caractère physique et social a tendance à varier spatialement (la température, les pentes, les densités de population, les richesses, etc.). Ainsi, aucun lieu ne peut être parfaitement représentatif de tous les autres.

C'est d'ailleurs bien une loi scalaire de ce type qui lie hétérogénéité et dépendance spatiales. Certains chercheurs, comme Anselin ou Goodchild, distinguent les deux en considérant que l'hétérogénéité spatiale est un effet de premier ordre (dans la mesure où elle ne concerne qu'un lieu à la fois), tandis que la dépendance spatiale

implique nécessairement une interaction entre deux lieux au moins, et serait donc un effet de second ordre. On peut néanmoins voir les choses différemment, et supposer que les liens subtils qui unissent dépendance et hétérogénéité sont d'ordre scalaire : l'hétérogénéité spatiale ne serait autre que la variance locale de la dépendance spatiale. En d'autres termes, la dépendance spatiale concernerait une échelle fine et impliquerait nécessairement, à une échelle plus large, une hétérogénéité spatiale. Comme l'indique [Getis \[1994\]](#), la dépendance spatiale est un cas spécial d'homogénéité spatiale. On pourrait illustrer cette position par un exemple de géographie sociale. Imaginons une ville fortement ségréguée d'un point de vue spatial : les quartiers sont homogènes (soit riches, soit défavorisés), et sont donc caractérisés par une autocorrélation spatiale forte des valeurs mesurées de la richesse et de probablement bien d'autres variables. Au sein de ces quartiers, pris séparément, les observations seraient spatialement dépendantes, les variances seraient faibles, de même que l'indice de Gini. Mais si l'on monte en échelle, et qu'on réalise les calculs au niveau de la ville entière, alors une hétérogénéité spatiale apparaîtrait, traduisant une mixité sociale. Cet exemple théorique est illustré de manière très simple dans la figure 1.2.

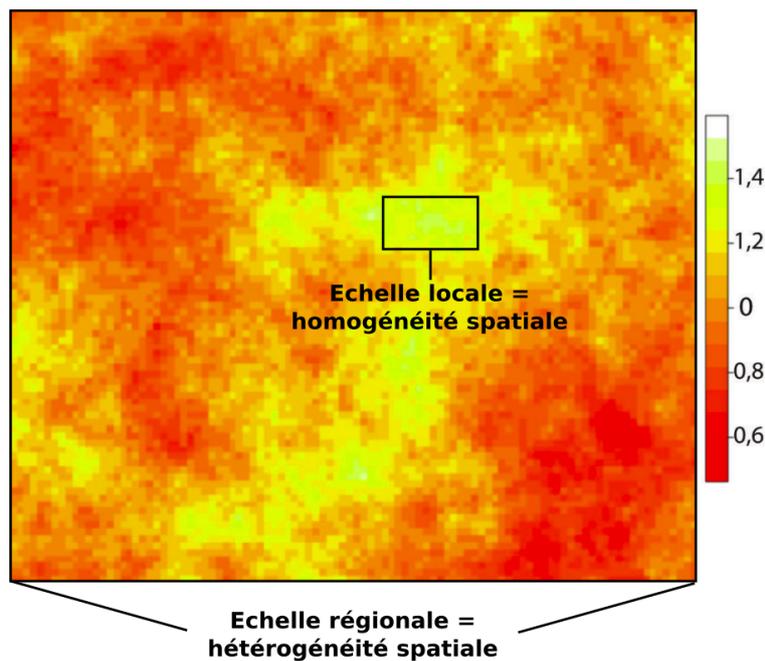


FIGURE 1.2 – Illustration théorique du rôle essentiel de l'échelle dans la distinction entre homogénéité et hétérogénéité spatiales, ici dans le cadre d'un champ de valeurs simulées. L'hétérogénéité, d'ordre scalaire supérieur, incorpore l'homogénéité. Sa caractérisation est donc dépendante des bornes de la zone d'étude.

Vu sous cet angle, l'hétérogénéité incorpore la dépendance, la dépendance impliquant l'hétérogénéité à une échelle supérieure. [Dutilleul et Legendre \[1993\]](#) arrivent à la même conclusion en écologie spatiale :

"scale is a crucial aspect of spatial heterogeneity. Shifts of scale, for an ecological entity, may lead from perceived homogeneity to heterogeneity, and vice versa; the information contained at one level of resolution may look like noise at another. In some cases, there may be a duality in nature between small-scale heterogeneity and large-scale homogeneity"

Jiang [2015] considère à ce titre l'hétérogénéité spatiale comme une loi scalaire. L'auteur va même plus loin en estimant que cette loi, bien ordonnée en profondeur, produit l'ensemble des « désordres » (c'est-à-dire des disparités spatiales) que l'on observe en surface. Ces « désordres » sont néanmoins bien régis par des lois, comme la loi de Zipf pour la taille des villes, ou la loi de Pareto pour bon nombre de distribution d'objets géographiques. Suivant ce raisonnement, Jiang inverse la logique en considérant l'hétérogénéité spatiale comme une loi de premier ordre (puisque'elle produit à elle seule l'ensemble du désordre), et la dépendance spatiale comme un loi de second ordre (puisque'elle implique l'interaction entre deux lieux au moins).

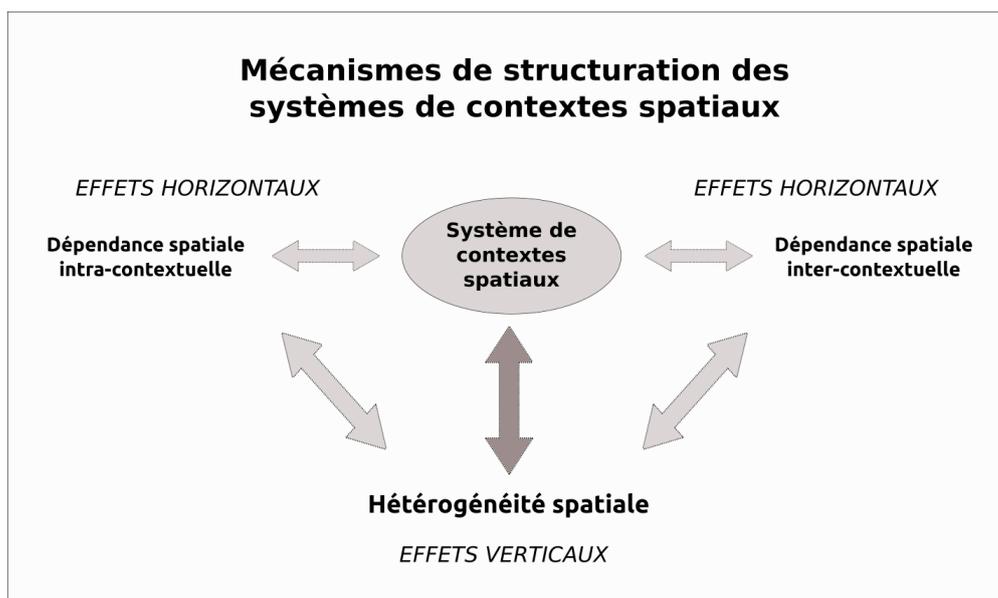


FIGURE 1.3 – Représentation sagittale des mécanismes de structuration des systèmes de contextes spatiaux, combinant effets verticaux (hétérogénéité spatiale) et effets horizontaux (dépendance spatiale).

Pour en revenir au sujet de ce travail, il est un fait que ces deux principes fondamentaux de la géographie – hétérogénéité et dépendance – sont intrinsèquement associés à la localisation des contextes spatiaux. Leur rôle est représenté schématiquement dans la figure 1.3, qui est la version sagittale, ou mécaniste, en quelque sorte, de la figure précédente. On y retrouve la structure en trois branches explicatives de Durand-Dastès [2001]. Dans ce schéma (reproduite de façon revisitée dans la figure 1.4), l'auteur proposait une démarche théorique visant à expliquer ce

qui se trouve dans un lieu donné j à un temps t par trois mécanismes : les relations écologiques, les interactions spatiales et les processus diachroniques. Bien que proche en bien des aspects, les mécanismes explicatifs de la structuration des systèmes de contextes spatiaux présentés dans la partie suivante apportent quelques nuances terminologiques, en plaçant notamment les deux grandes lois de la géographie au cœur du schéma. Cela permet aussi de mieux discriminer les relations écologiques des processus diachroniques, dans la mesure où les relations écologiques peuvent impliquer des liens aux héritages difficilement discernables des processus diachroniques.

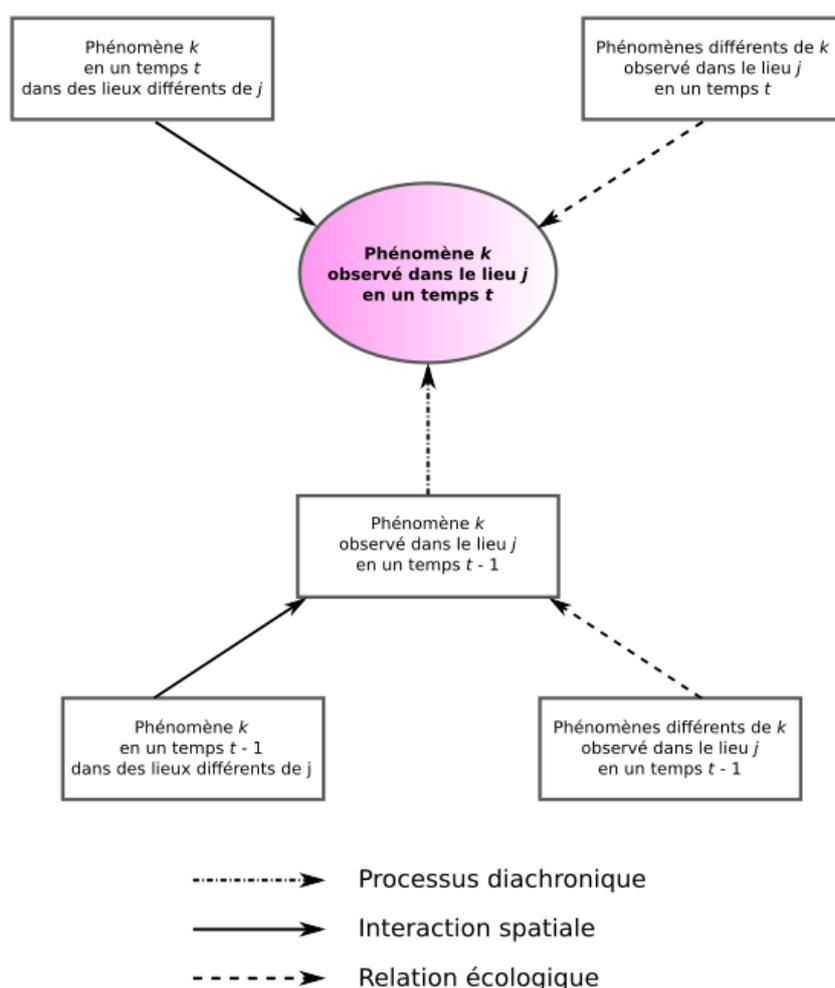


FIGURE 1.4 – Modèle d’explication des localisations selon Durand-Dastès [2001] (revisité). Ce modèle est également fondé sur trois branches explicatives dans deux logiques synchroniques : les processus diachroniques, les interactions spatiales et les relations écologiques.

1.3 Le rôle de l'hétérogénéité et de la dépendance spatiales dans la structuration des contextes spatiaux

Spatial dependency and spatial heterogeneity create local context that shapes physical and human processes as they evolve on the surface of the Earth
Flake [1998], cité par Miller et Goodchild [2015]

1.3.1 Des effets verticaux : l'hétérogénéité spatiale et les processus de long terme

L'hétérogénéité spatiale participe à la pré-figuration des contextes spatiaux. Cet effet vertical fait étroitement écho aux effets écologiques de Pumain et Saint-Julien [2010, p. 26], décrits comme les effets de la superposition d'objets ou de configurations au sein d'un même lieu, ou en d'autres termes les "rapports au milieu" [Durand-Dastès, 2001]. Cela concerne également les processus diachroniques de Durand-Dastès [1992, 1984] et son concept d'espace reçu, ou d'espace-mémoire [Durand-Dastès, 1991]. L'auteur explique que chaque génération de sociétés reçoit un espace déjà structuré par son histoire, fondamentalement hétérogène, empreint donc d'héritages qui contribuent à déterminer la localisation des phénomènes actuels. P. Pinchemel décrivait ces héritages de façon équivalente : "Les structures géographiques sont inscrites à la surface de la terre par les sociétés, comme une écriture ; les signes en sont les points, les lignes et les surfaces dont se servent les hommes depuis l'aube de l'humanité pour délimiter, utiliser, organiser cette surface terrestre" [Pinchemel, 1981]. Pour élargir ces propos, on peut considérer que cette hétérogénéité a été nourrie par des processus de long terme, autant physiques que sociaux. Les processus physiques sont ceux qui structurent les grands éléments du paysage, tels que les vallées, les versants, les plateaux, et qui favorisent, par leurs caractéristiques propres (pente, lithologie, nature et épaisseur du sol, humidité, etc.), tel ou tel usage du sol. Les processus sociaux de long terme concernent la façon dont l'espace a été aménagé, occupé et administré par les générations passées, et dont on en ressent encore aujourd'hui l'inertie à travers ce qu'on pourrait appeler des effets d'hystérésis³. L'effet de cette hétérogénéité spatiale sur la localisation des contextes spatiaux est vertical, dans la mesure où la géographie de ces contextes – et des réseaux qu'ils forment – se superposent (en partie) aux formes structurelles de l'espace hétérogène. C'est donc en quelque sorte une force créatrice d'organisations spatiales subséquentes, et qui donc précède les jeux de relations interactives entre ces organisations, de même que leur fonctionnement. Pour qu'il y ait des effets de contexte, il faut d'abord qu'il y ait des contextes. L'hétérogénéité a donc un rôle

3. L'hystérésis désigne de façon générale le retard de l'effet sur la cause. Ce concept est utilisé en sciences dures comme en sciences humaines, notamment en économie où il décrit une situation dans laquelle le taux de chômage reste élevé alors que la cause de la hausse a disparu.

précurseur fondamental dans ce schéma spatial, c'est pourquoi la flèche qui y est associée, dans la figure 1.3, apparaît surlignée. Il convient de préciser que ce n'est pas pour autant que l'espace doit être considéré comme un support passif. Ses caractéristiques structurelles sont issues de processus spécifiques, parfois encore actifs, si bien que ce n'est pas l'espace qui doit être considéré comme une cause, mais bien ce qui l'a structuré :

"space should be used as a predictor variable only as a last resource as it does not provide any insights about the causality relationships" [Dale et Fortin, 2014, p. 31]

Goodchild va dans le même sens en écrivant : "Spatial properties such as latitude, distance, or direction could never be said to truly explain anything" [Goodchild, 2008, p. 13]. Ce qui détermine les comportements de mobilité, ce n'est pas la distance, mais le coût de cette distance.

1.3.2 Des effets horizontaux : dépendance spatiale et effets de voisinage

Des processus horizontaux (diffusion, influence, échanges, etc.) se relaient et se combinent alors pour structurer cette organisation de l'espace en contextes, qui se traduit par des configurations spatiales préférentielles des aménagements, des richesses, des ressources, des densités de toutes sortes. Les notions de distance et de proximité deviennent alors centrales, car elles vont déterminer l'intensité de ces processus horizontaux. Ces processus vont engendrer une systémogénèse. Un système est défini comme un ensemble d'éléments en interaction. C'est précisément ce qui caractérise ces contextes spatiaux : la convergence géographique de ses éléments constitutifs va engendrer des interactions. Ces interactions représentent les effets de contexte *stricto sensu*. C'est également ce qui caractérise les relations entre des contextes relativement proches, liés par des interactions spatiales dont l'intensité est déterminée à la fois par la proximité et les caractéristiques des contextes. Il s'agit d'un autre type d'effet de contexte, entendu au sens large. L'ensemble de ces interactions, inter- et intra-contextuelles, constituent une force de stabilisation ou de renforcement de l'ensemble du système. Ce fonctionnement systémique de l'espace est bien connu dans le domaine des simulations de systèmes géographiques complexes [Mathian et Sanders, 2014]. Un exemple théorique pourrait en être le modèle de Schelling appliqué à un espace réticulaire présentant des structures spatiales sous-jacentes (cliques), décrit par Banos [2010] : ses simulations montrent que les caractéristiques des cliques déterminent les structurations spatiales ségréгатives de l'ensemble de l'environnement. Ces effets de cliques ne sont autre que des effets de contextes locaux appliqués à un espace réticulaire et fonctionnant au sein d'un système complexe. Un exemple équivalent en géomorphologie pourrait être illustré par les travaux de Cossart et Fressard [2017] qui, toujours dans un espace réticulaire et à l'aide de la théorie des graphes, ont montré que l'efficacité de l'ensemble du transfert sédimentaire dans un bassin-versant pouvait dépendre des

caractéristiques structurelles locales d'un très petit nombre de nœuds.

En sciences sociales, les contextes ont longtemps été vus – et continuent de l'être dans de nombreuses publications aujourd'hui – comme des boîtes noires dont on constate l'influence quasi mystique sur des objets, mais dont on ne comprend pas précisément les mécanismes. Ce flou a été parfaitement décrit par *Macintyre et al.* [2002] dans un article référence sur les liens entre lieux et santé :

"'Context' is thus often treated as a residual category, containing those factors influencing human health behaviours or health which remain once every imaginable individual characteristic is taken into account. It is indeed a black box, an unspecified "miasma" which somehow, but we do not know how, influences some aspects of health, health-related behaviour or health risks in some population groups" [*Macintyre et al.*, 2002, p. 129]

Néanmoins, les mécanismes théoriques œuvrant à travers les effets de contexte ont tout de même été l'objet d'articles théoriques, principalement dans le domaine des effets de quartier sur les comportements individuels, notamment par les sociologues. Différents auteurs ont proposé différentes typologies de mécanismes, se recoupant souvent partiellement, chacune étant orientée selon les angles privilégiés dans la discipline ou la spécialité de son auteur. En revanche, de façon étonnante, les géographes ne semblent pas s'être tellement adonnés à ce type d'exercice, comme le relèvent d'ailleurs Macintyre et ses co-auteurs (issus de la santé publique et de l'épidémiologie) : "Although some of the work on 'place effects' on health has been undertaken by geographers (not all of whom have been convinced that there are any 'place effects' [...]), much of it has been undertaken by researchers from outside geography and has not been informed by any specifically geographical perspective" [*Macintyre et al.*, 2002, p. 126]. Depuis 2002, de nombreux géographes de la santé se sont investis dans ce domaine, mais il faut bien en effet reconnaître qu'il est difficile de trouver actuellement, dans la littérature scientifique, des essais de typologie des effets de contextes spatiaux conçus par des géographes, mobilisant le bagage théorique propre à leur discipline. Pour corriger en partie cette lacune malheureuse, nous passerons en revue deux de ces typologies, les plus significatives, pour tenter ensuite d'en proposer une nouvelle empruntant une perspective explicitement géographique, suffisamment englobante pour concerner tous les types d'effets de contexte spatial, et pas seulement les effets de quartier, toujours dans le même objectif de généralité.

Celle de *Galster* [2012], tout d'abord, définit quatre types d'effets, regroupant un total de 15 mécanismes⁴ :

4. Si l'énonciation de ces effets peut paraître par trop atomisée de prime abord, il convient de garder en tête que la plupart d'entre eux se combinent en réalité, comme l'illustrera la typologie finale qui sera proposée en fin de section.

- les effets d'interaction sociale, qui concernent tous les mécanismes impliquant des échanges entre les individus d'un même contexte : effets de contagion, socialisation collective, réseaux interpersonnels, compétition, etc.
- les effets environnementaux, qui se réfèrent aux effets directs des attributs des contextes locaux sur les individus et leur santé, sans médiation comportementale. Cela intègre par exemple la pollution de l'air ou de l'eau, l'exposition au bruit, à la délinquance, ou encore à l'esthétique du quartier.
- les effets géographiques, qui désignent les avantages et les limites que tirent les quartiers de leur position géographique dans un contexte territorial plus large (distance aux bassins d'emplois, accessibilité, etc.).
- les effets institutionnels enfin, qui impliquent les actions menées par des acteurs (publics ou privés) extérieurs aux quartiers en question, et qui déterminent les caractéristiques du quartier (cela inclut par exemple la stigmatisation, la qualité des ressources locales en termes d'écoles, de commerces, d'hôpitaux, etc.)⁵.

La typologie de [Jencks et Mayer \[1990\]](#) est assez proche de celle de Galster. Par ailleurs, l'une des typologies plus évoquées dans la littérature sur les effets de quartier est celle de [Manski \[1993\]](#). Elle a l'avantage d'être générique, mais l'inconvénient d'être pensée dans une logique statistique, en mobilisant une terminologie propre aux méthodes de régression qui peut porter à confusion. Cette typologie inclut également quatre effets, dont les deux premiers seulement sont des effets sociaux :

- les effets endogènes, qui désignent la propension d'un individu à se comporter de la même façon que les individus du même contexte. Ces effets sont conceptuellement proches des effets d'interaction sociale de Galster.
- les effets contextuels (ou exogènes), qui désignent la propension d'un individu à se comporter en fonction des caractéristiques des autres individus.
- les effets écologiques, traduisant le fait que les individus d'un même contexte se comportent de façon similaire car ils sont soumis au même environnement institutionnel.
- les effets corrélés enfin, selon lesquels les individus d'un même contexte se comportent de façon similaire car ils disposent de caractéristiques individuelles similaires.

Pour illustrer son propos, l'auteur prend l'exemple de la réussite scolaire d'un adolescent, le contexte étant donc représenté par son école. Il y a effet endogène quand sa réussite scolaire tend à varier avec la réussite scolaire moyenne des autres élèves de l'école. Il y a effet contextuel quand sa réussite tend à varier, par exemple, en fonction de la composition socio-économique des autres élèves. L'effet écologique est impliqué quand les élèves d'une même école disposent d'environnements familiaux similaires, et que ces contextes familiaux tendent à leur tour à affecter la

5. On pourrait considérer ces effets institutionnels comme un mécanisme d'initiation des effets environnementaux, plutôt que comme un effet à part entière.

réussite scolaire. Enfin, les effets corrélés désignent par exemple la situation dans laquelle la réussite scolaire dépend en partie du fait que les enfants d'une même classe ont tous à peu près le même âge.

En s'inspirant de ces deux typologies, partiellement redondantes, il est possible d'en dessiner une autre, suffisamment générique pour pouvoir nourrir l'interprétation, théorique au moins, des effets de contextes spatiaux, au-delà de la géographie sociale seulement. Le raisonnement étant spatialisé, cette typologie distingue tout d'abord deux sous-types d'effets : les effets endogènes, qui désignent des processus dont l'initiation est interne aux contextes, et les effets exogènes⁶, qui concernent le voisinage des contextes, c'est-à-dire leur position dans un environnement plus large, abritant d'autres contextes (l'initiation des processus est extérieure aux contextes, donc) :

- Les effets endogènes peuvent être classés en deux types : les effets d'interactions et les effets de normes locales partagées. Le premier résulte des interactions entre agents d'un même lieu, liés par leur proximité géographique, qui peuvent favoriser une convergence des comportements ou des états, via des processus de contagion ou de mimétisme par exemple. Cela s'observera à l'échelle locale surtout : rue, quartier, versant, etc. Si les agents en question sont des humains, il s'agira d'interactions sociales. Ce premier effet est donc proche de du premier défini par Galster, ou des deux effets sociaux de Manski. Le second effet endogène est celui de la norme locale partagée : le fait de partager un même contexte, avec ses co-occurrences de caractéristiques nécessairement singulières et ses spécificités issus de processus verticaux ou institutionnels, contribue à faire converger les états dans cet environnement, sans obligation d'interactions. Il correspond aux effets environnementaux de Galster, autant qu'aux effets écologiques de Manski. Ces deux effets endogènes peuvent être considérés comme relevant d'une dépendance spatiale intra-contextuelle, tel qu'elle est illustrée dans la figure 1.5. Comme signalé par Merlo *et al.* [2005], la combinaison de ces deux effets endogènes évoque, en sciences sociales, la production de la cohésion sociale de Durkheim [1893], selon laquelle le partage de valeurs et normes communes via l'appartenance à une même communauté génère une conscience collective. Cette conscience devient un fait contextuel à part entière, qui dépasse la somme des caractéristiques individuelles.
- Les effets exogènes correspondent en substance aux effets géographiques de Galster, et peuvent également être divisés en deux types (figure 1.5) : les effets de voisinage et les effets de position. Les effets de voisinage concernent les

6. Notons que ces termes d'endogénéité et d'exogénéité ne sont pas utilisés dans le même sens que dans les textes de Manski. Ils ont un sens ici géographique (effets internes ou externes au contexte), quand Manski les mobilise dans la terminologie des modèles de régression (variables endogènes, ou à expliquer, et variables exogènes, ou explicatives).

interactions spatiales entre un contexte et ses voisins directs, tandis que les effets de position se réfèrent à la place d'un contexte dans une hiérarchie plus large. Cela implique les distances à des concentrations fonctionnelles ou à des éléments structurants d'un environnement plus large, comme la distance aux bassins d'emplois, aux ressources, aux aménités, aux cours d'eau, aux aéroports, etc. Par exemple, le fait qu'un quartier soit situé en zone périurbaine ou en centre-ville a des implications sur les comportements individuels (en termes de mobilité notamment) – même si les quartiers présentent les mêmes caractéristiques intrinsèques – ce qui relève typiquement de ces effets de position. Effets de voisinage et effets de position sont conceptuellement assez proches, mais les effets de position n'impliquent pas nécessairement de voisinage direct. Ces deux effets exogènes peuvent être considérés comme relevant d'une dépendance spatiale inter-contextuelle (figure 1.5).

Ainsi, les effets de contexte peuvent être vus comme agissant à deux échelles : au sein même des contextes, et à une échelle supra-contextuelle, en lien avec la position du contexte dans son environnement (ou, en géographie humaine, dans son territoire). Par exemple, des comportements individuels ou des états de santé peuvent être expliqués par les caractéristiques locales du quartier de résidence (effets endogènes), mais également, de façon simultanée, par la position de ce quartier dans la hiérarchie du système urbain (effets exogènes). C'est ce que nous avons par exemple illustré dans un article récent sur l'obésité en France, dans lequel nous avons mis en relation le risque d'être en surpoids ou obèse avec l'environnement socio-économique du quartier de résidence (IRIS, caractérisant donc ici le contexte spatial *stricto sensu*), dans un échantillon de 70 000 personnes [Feuillet *et al.*, 2020b]. Cet environnement socio-économique a été quantifié à travers l'indice de défaveur sociale français [Rey *et al.*, 2009], variable latente calculée sur la base du taux de chômage, du pourcentage d'ouvriers, du pourcentage de personnes non diplômées et du revenu médian, le tout au niveau de l'IRIS. Le premier résultat a été de révéler une relation significative entre cet indice de défaveur sociale et le risque de surpoids, ce qui correspond à un effet endogène. Cet effet endogène peut par exemple être dû au fait que les habitants d'un même quartier sont soumis aux mêmes contraintes obésogènes (par exemple l'absence d'aménagements urbains favorables à la mobilité active). Mais de façon plus originale, nous avons également montré que cette relation variait en fonction de l'appartenance de chaque IRIS au contexte territorial. Ce contexte territorial a été défini à partir d'une combinaison de nomenclatures de l'INSEE, sur des critères à la fois fonctionnels et morphologiques, et résultant en une classification en 10 contextes (par exemple : banlieue de grande aire urbaine, ville-centre de petite aire urbaine, zone rurale, petite couronne parisienne, etc.). Ainsi, la relation entre le niveau de défaveur sociale du quartier et le risque de surpoids est apparu très fort en banlieue des grandes villes, mais nulle ou faible en milieu rural ou périurbain (figure 1.6). Ce second résultat illustre un effet exogène (effet de position), relatif à la situation géographique du contexte spatial (quartier de résidence) dans la hiérarchie urbaine

(déterminant elle-même tout une série de facteurs de comportements obésogènes, comme l'accessibilité aux réseaux de transport en commun et à l'emploi).

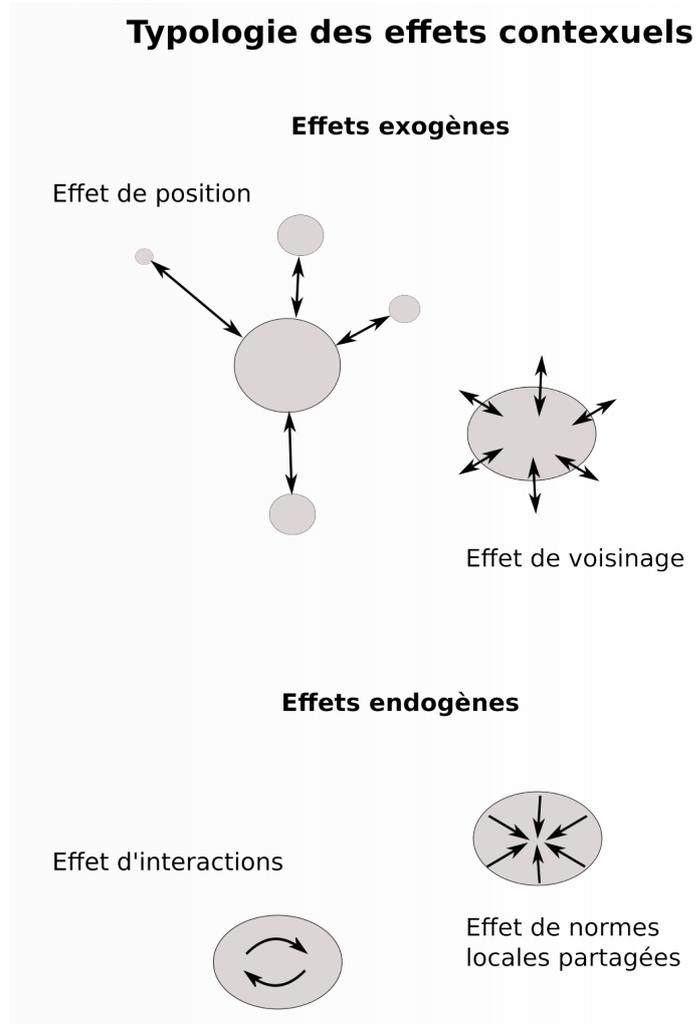


FIGURE 1.5 – Typologie des effets contextuels. Ces effets sont distingués selon qu'ils œuvrent au sein des contextes uniquement (effets endogènes) ou qu'ils sont initiés à l'extérieur des contextes (effets exogènes).

Les typologies, comme tout modèle, présentent l'intérêt de simplifier la réalité pour faciliter la représentation de sa complexité, mais aussi l'inconvénient, par le principe même de la classification, de séparer des éléments qui sont fondamentalement connectés. Les quatre effets présentés ci-dessus sont ne peuvent pas être strictement distingués les uns des autres par l'observation, car ils œuvrent au sein de systèmes, dans lesquels les problèmes d'antériorité et d'endogénéité sont insolubles par principe. Ces effets se nourrissent mutuellement, se renforcent ou s'affaiblissent, se chevauchent, sans qu'il soit réellement possible de déterminer

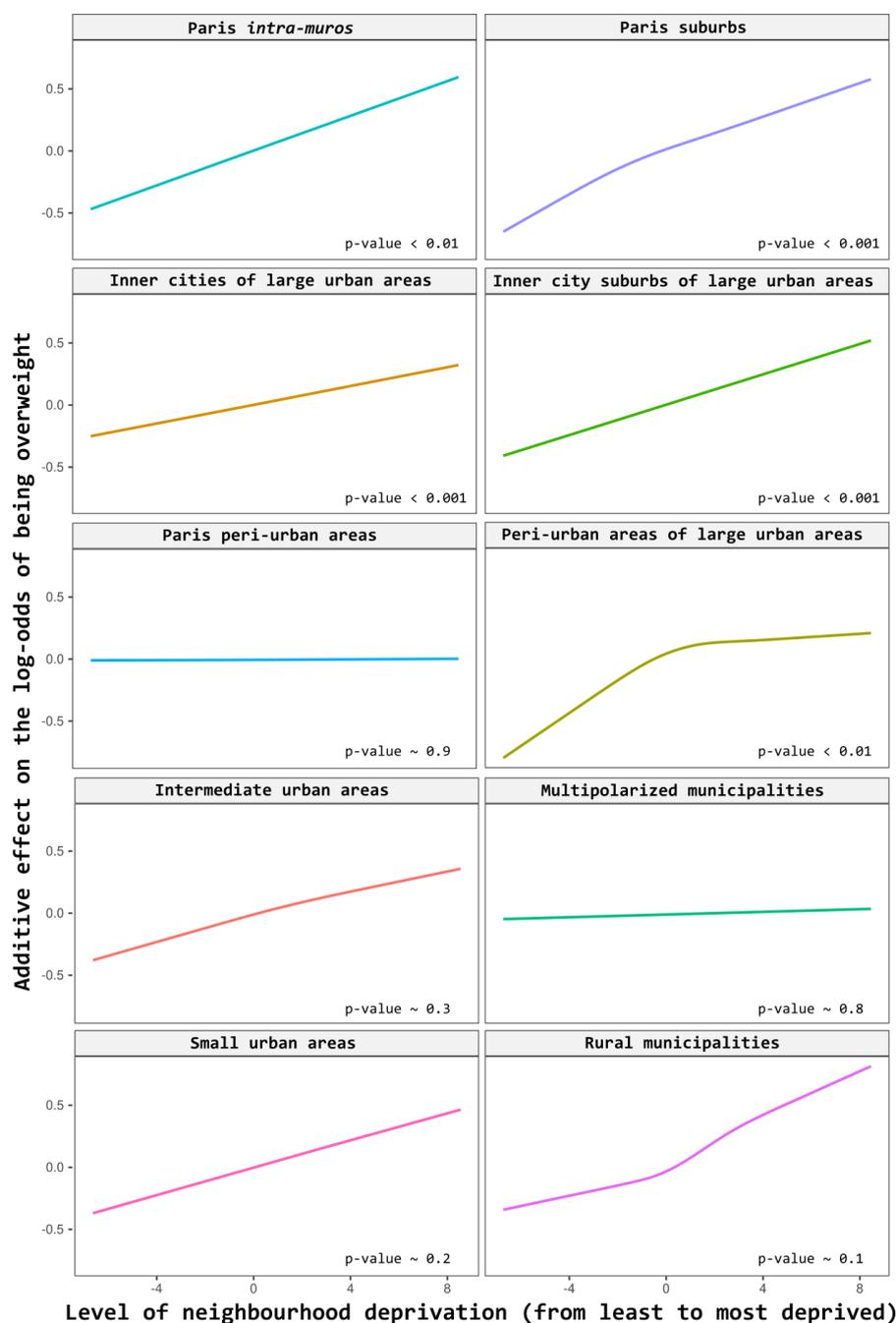


FIGURE 1.6 – Relations entre le risque d'être en situation de surpoids et le niveau de défaveur sociale du quartier de résidence, en fonction du contexte urbain, dans un échantillon de 70 000 personnes de la cohorte Nutrinet-Santé. Ces graphiques illustrent le caractère multiscalaire des effets de contexte : globalement, la relation entre le niveau de défaveur sociale du quartier de résidence et le surpoids est positive, reflétant un effet contextuel endogène (lui-même issu d'effets d'interactions et d'effets de normes locales partagées). Néanmoins, cette relation varie selon le contexte urbain, indiquant cette fois un effet exogène (effet de position). Figure extraite de [Feuillet et al. \[2020b\]](#).

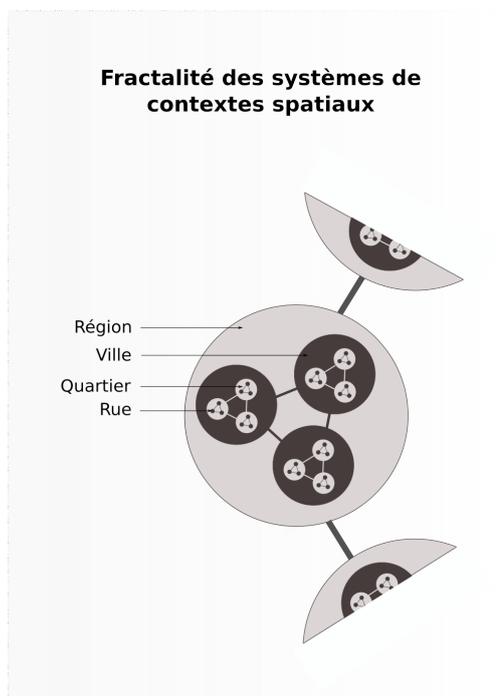


FIGURE 1.7 – Représentation schématique de la composante fractale des systèmes de contextes spatiaux. Les mécanismes systémiques œuvrent à toutes les échelles.

lequel d'entre eux a agi en premier. C'est exactement le paradoxe de la poule et de l'œuf, pour lequel l'approche systémique est la mieux à même de répondre.

Ainsi, processus verticaux (effets de l'hétérogénéité spatiale) et processus horizontaux (effets de la dépendance spatiale) se relaient ou se combinent pour structurer les systèmes de contextes spatiaux. Comme précédemment évoqué, cette considération de l'espace par systèmes de contextes permet d'impliquer naturellement les principaux fondements de la géographie, dont le principe fractal constitue une dimension, au même titre que l'hétérogénéité et la dépendance spatiales. Le principe fractal désigne, au sens large, la répétition des structures à travers les échelles [Mandelbrot, 1982]. En d'autres termes, comme le résume Goodchild [2003], "the closer you see the more you look". C'est précisément ce qui caractérise cette vision par systèmes de contextes : quelles que soient les échelles d'analyses, des contextes spatiaux connectés se dessinent (figure 1.7). Cette structure se répète à différentes résolutions spatiales pour former un emboîtement caractéristique des systèmes spatiaux. Ce schéma n'est bien entendu pas sans rappeler la théorie des lieux centraux ou les structures de Mandelbrot, dont il s'inspire.

1.4 Apports de la considération des systèmes de contextes spatiaux dans les analyses géographiques

Les effets de contexte sont naturellement et inextricablement liés à la géographie. Quels que soient les objets d'étude, les échelles et les approches, ces effets sont essentiels dans l'interprétation des structures spatiales des faits étudiés et dans la compréhension de leurs processus générateurs. Deux principales catégories de raisons justifient de considérer cette vision de l'espace sous forme de systèmes de contextes et de ne pas ignorer leurs effets dans l'analyse de phénomènes spatialisés : des raisons épistémologiques et des raisons théoriques.

1.4.1 Les effets de contexte en géographie : des conciliateurs épistémologiques

En tant que discipline, la géographie trouve sa raison d'être dans la description et la comparaison des lieux, ainsi que dans l'appréhension de leur organisation spatiale. La partie qui suit a pour but de souligner la manière dont la prise en compte des effets contextuels tend à fédérer les approches et les concepts en géographie, en d'autres termes à réconcilier les différentes sensibilités d'une science qui souffre parfois de sa diversité thématique.

La pertinence d'une approche nomothétique comparée à une approche idiographique a longtemps fait l'objet de débats en géographie. En tant que science de l'espace et des territoires, il était logique que des oppositions naissent entre des défenseurs de la spécificité des lieux et des partisans d'une approche globale plus lissée et comparative. Cette conception idiographique trouve son apogée en France dans la première moitié du 20^{ème} siècle, dans les traces de Vidal de la Blache et de ses élèves, et est marquée par une démarche scientifique consistant à multiplier les monographies régionales exhaustives sans nécessairement les connecter. Cette volonté d'encyclopédisme, cette quête du particularisme, est parfaitement illustrée par un examen rapide des thèses de géographie soutenues en France entre 1872 à 1972 et référencées par Briend et Joseph [1997] : sur les 238 thèses soutenues, 37 % impliquaient une vision territoriale très localisée (une ville ou un lieu par exemple), 55 % une échelle régionale et 8 % seulement une échelle globale (supra-nationale ou purement thématique, sans attache géographique précise). À titre d'illustration de cette logique idiographique visant à accumuler des connaissances géographiques locales, on peut citer la thèse de Clozier [1940] intitulée *La gare du Nord*, ou les très nombreux essais régionaux, comme celle de George [1935] : *La région du bas Rhône : étude de géographie régionale*, de Juillard [1953] : *La vie rurale dans la plaine de basse Alsace. Essai de géographie sociale*, ou encore celle de Tricart [1949] : *La partie orientale du bassin de Paris. Étude morphologique*.

Dès 1953, Fred K. Schaefer, géographe initiateur de la révolution quantitative,

écrivait :

"The present conditions of the field (of geography) indicate a stage of development well known from other sciences which finds most geographers still busy with classifications rather than laws" (Schaefer [1953, p. 229], cité par Golledge et Amedeo [1968]).

En d'autres termes, « le géographe se mue en connaisseur érudit des régions d'un territoire » [Varenne, 2017, p. 29], sans perspective généralisatrice. Cette approche fait écho aujourd'hui à ce qui est appelé la géographie régionale. C'est à partir des années 1960 que va apparaître un contre-courant porté par la géographie quantitative, en particulier grâce à Haggett [1965] et son ouvrage séminal *Locational analysis in human geography*, contre-courant dans lequel la recherche de régularités dans l'organisation spatiale des phénomènes, voire de lois théoriques, prend le pas sur la caractérisation empirique des singularités territoriales. En Europe, cette dynamique de recherche est illustrée par la tenue régulière des *European Colloquium of Theoretical and Quantitative Geography* (ECTQG), débutés en 1978 à Strasbourg. En France, un colloque de géographie quantitative est organisé à Besançon dès 1972 [Pumain et Robic, 2002] et préfigurera les rencontres de Théo Quant. Ce mouvement est initié par des géographes (des « passeurs », selon le terme employé par Sanders [2011]) comme Philippe Pinchemel, Antoine Bailly, François Durand-Dastès, Henri Reymond, Jean-Bernard Racine, Denise Pumain ou encore Roger Brunet. Ce dernier justifiait l'élan de cette nouvelle géographie théorique et quantitative dès 1976 par l'impasse que représentait l'accumulation de monographies et la lassitude à l'égard d'une géographie générale aboutissant à des typologies sans principes (Brunet [1976], cité par Cuyala [2015]).

La dichotomie idiographie/nomothétie résonne avec d'autres dualités scientifiques récurrentes. Ainsi, dans l'approche nomothétique même, on continue souvent d'opposer l'approche inductive, partant des observations factuelles, à l'approche déductive, partant d'hypothèses théoriques et reflétant en cela la conception nomothétique. De même, l'opposition global vs local est régulière. Dans une certaine mesure, la distinction tenace et parfois prétendue exclusive entre approches quantitative et qualitative relève de la même logique. L'approche qualitative est souvent associée à une analyse fine des mécanismes sur un échantillon réduit d'observations, quand l'analyse quantitative vise à synthétiser des informations plus grossières, mais sur un échantillon plus large. Tous ces débats consistent finalement à opposer différentes façons de caractériser l'espace et, en pratique, ces positions ne sont pas incompatibles : par exemple, l'identification de singularités locales ne compromet pas la généralisation, tandis que la recherche de lois générales n'exclut pas les spécificités. Ce ne sont pas des alternatives. De même, les allers-retours entre hypothèses théoriques et confrontation aux faits observés sont habituels et nécessaires. Comme le rappellent Gana et Broc [2018], « il est rare qu'une démarche inductive soit complètement athéorique, totalement extrinsèque à toute hypothèse a priori, et où les résultats empiriques s'en remettraient à la pure loi du hasard »

[Gana et Broc, 2018, p. 267].

Il est possible d'apprécier cette conception à travers une analogie avec l'inférence statistique. L'inférence statistique est le cadre qui permet de généraliser des observations, c'est-à-dire de tirer des conclusions générales (généralisation) d'une observation partielle du réel. En statistique inférentielle, on considère que cette observation est une réalisation d'un processus parent, englobant, dont il s'agit de comprendre les logiques de fonctionnement. Ces observations empiriques ne prennent donc tout leur sens que dans la mesure où elles représentent (plus ou moins bien selon la pertinence de l'échantillonnage) le processus d'ordre supérieur qui les a engendrées.

Ces dualités structurent en un sens la discipline géographique, mais la brident par ailleurs, comme tout cloisonnement. Plusieurs raisons permettent de penser qu'appréhender les analyses géographiques par la question des effets de contexte dans une logique systémique contribue à surmonter naturellement ces oppositions. Accepter l'ubiquité des effets de lieu, quels que soient les faits ou les processus étudiés, implique de reconnaître l'hétérogénéité des structures spatiales et la coexistence parfois contiguë de configurations spécifiques. Les effets de contexte prennent ainsi tout leur sens quand on les emploie au pluriel, et ce pluriel implique d'appréhender les variantes locales d'un phénomène comme des éléments intrinsèquement constitutifs des structures dégagées à un niveau supérieur [Sanders, 2001]. Surtout, cela implique d'avoir une vision d'ensemble de l'organisation spatiale de ces configurations, ou en d'autres termes de concevoir une géographie des spécificités locales. La spécificité d'un lieu est ainsi mise en perspective par sa juxtaposition avec d'autres lieux, d'autres spécificités, pour aboutir à la définition d'une structure d'organisation, qui correspond au système de contextes spatiaux définis dans la partie précédente. Ainsi, comprendre les effets de lieu permet à la fois de caractériser la spécificité des lieux, tout en appréhendant la logique de répartition des contextes spatiaux à une échelle supérieure. Le développement de ce type d'analyses basées sur les lieux (qualifiées de *place-based analyses* par les anglo-saxons) représentent, selon Goodchild [2004], "a middle position in the old debate in geography between nomothetic and idiographic science, rejecting the position that all places are unique in favor of general principles whose parameters vary from one place to another" [Goodchild, 2004, p. 711]. C'est dans ce sens que l'analyse des effets de lieu peut être considérée comme un réconciliateur épistémologique.

Illustrons cette position par un exemple (figures 1.8 et 1.9). Dans une étude récente [Feuillet *et al.*, 2020a], nous avons cherché à étudier la dynamique spatiale de la limite supérieure de l'arbre (*treeline*) dans les Pyrénées, depuis les 60 dernières années. Il se trouve qu'à l'échelle régionale (ici la partie orientale du massif), cette dynamique répond au réchauffement climatique (augmentation des températures estivales) par une remontée altitudinale, lente mais régulière, de $+0,73 \text{ m/an}^{-1}$. Il s'agit donc là de la tendance générale du processus étudié, celle à laquelle on

s'arrêterait dans le cadre d'une démarche purement nomothétique. Imaginons maintenant que dans une démarche d'analyses locales, nous ne sélectionnions qu'un petit échantillon de versants, et que ces échantillons révèlent à la fois de fortes remontées et des régressions altitudinales. Dans les deux cas, les facteurs locaux prédominent sur le facteur climatique. Les fortes remontées s'expliquent par une déprise agropastorale particulièrement significative et provoquant un rebond puissant de la *treeline*, tandis que les processus de régression sont dus à une suractivité géomorphologique qui freine la colonisation des arbres. En résumé, l'analyse régionale efface le rôle des facteurs locaux, tandis que les analyses locales échouent à enregistrer le signal climatique d'ordre régional. Ainsi, aucune des deux analyses n'est valable tant qu'elles s'excluent. Abordons maintenant la problématique sous l'angle des effets de contexte. On peut considérer que les versants sur lesquels la dynamique de la *treeline* se comporte de façon particulière, c'est-à-dire là où elle s'écarte le plus de la tendance régionale, sont soumis à des effets de contextes locaux : la combinaison des caractéristiques propres à chacun de ces versants (activités géomorphologiques, trajectoires agropastorales, etc.) affectent de façon spécifique la dynamique de la *treeline* en chaque lieu. Ces lieux, ou contextes, évoluent alors de façon subautonome et systémique, intégrant des boucles de rétroaction internes : la dynamique de l'arbre joue localement sur elle-même, en amplifiant le processus initial (ces effets sont qualifiés d'endogènes dans la terminologie de Manski, cf. § 1.3.2). Comprendre la dynamique de la *treeline* et ses ressorts requiert donc d'avoir une vue d'ensemble de ces effets de contexte, car effets régionaux et effets locaux se subsument. Cette subsomption doit être appréhendée dans un premier temps par la description de la géographie des effets de contexte – de leur agencement spatial et de leurs caractéristiques – qui permettra alors d'identifier de façon pertinente et représentative les versants qui devront subséquemment faire l'objet d'observations tendant vers l'exhaustivité.

Une géographie des lieux, intégrant par essence la question des effets contextuels avec lesquels ils sont intrinsèquement liés, serait en quelque sorte le pendant explicatif d'une géographie de l'espace, davantage descriptive. L'une cherche à décrire, l'autre à expliquer. Les effets contextuels sont des processus générateurs de structures spatiales, contrairement par exemple à l'hétérogénéité spatiale qui n'est qu'une description transversale de ces structures. La géographie des lieux et des effets de contexte ne cherche pas seulement à décrire les structures spatiales, mais également à les expliquer.

1.4.2 L'importance des effets de contexte dans les chaînes causales

Ne pas considérer les contextes et leurs effets peut mener à une interprétation biaisée des chaînes causales théoriques, à au moins deux niveaux (figure 1.10). Tout d'abord, les contextes spatiaux en question sont eux-mêmes parfois facteurs explicatifs du phénomène étudié (noté « variable Y » dans la figure 1.10). Par

1.4. Apports de la considération des systèmes de contextes spatiaux dans les analyses géographiques

29

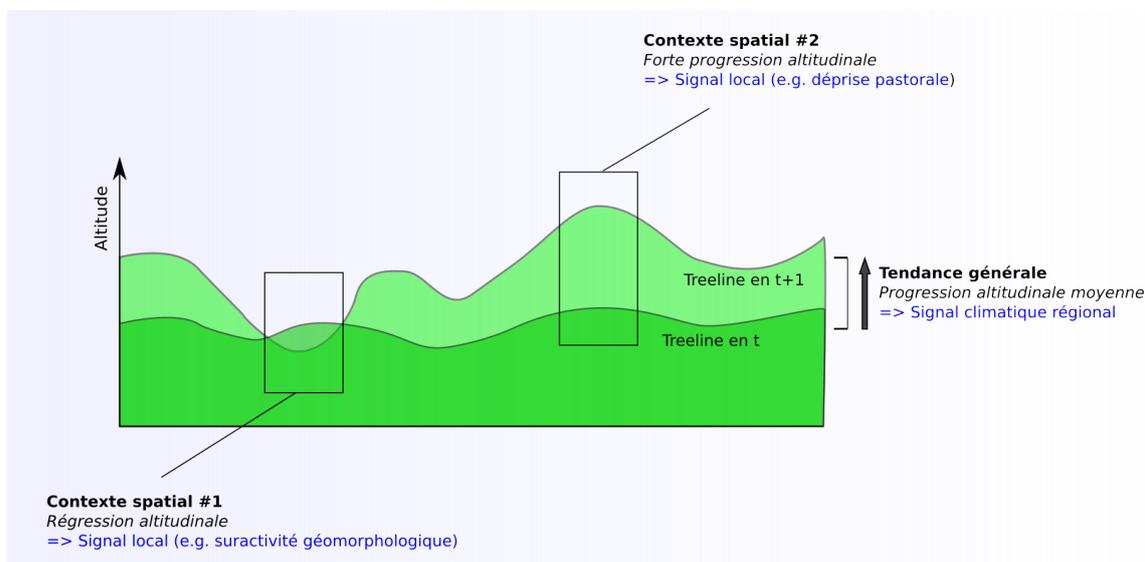


FIGURE 1.8 – Illustration théorique de l'importance des contextes spatiaux dans l'analyse de la dynamique de la *treeline*. A l'échelle régionale, la *treeline* a progressé en altitude, en réponse au réchauffement climatique. Dans le contexte spatial n°1, la *treeline* a régressé, en raison d'une suractivité géomorphologique locale. Dans le contexte spatial n°2, la *treeline* a très fortement progressé à cause d'un abandon agro-pastoral local marqué. Pour comprendre entièrement ce processus écologique, il convient donc de combiner les analyses locales et régionales, ou en d'autres termes d'adopter une géographie régionale des spécificités locales.



FIGURE 1.9 – Illustration photographique de deux contextes spatiaux d'évolution de la limite supérieure de la strate arborée en milieu montagnard. Dans le contexte de gauche (massif du Carlit), la forme très disséquée de l'écotone indique une dynamique de l'arbre contrainte par les conditions topo-géomorphologiques. On distingue par exemple des tabliers d'éboulis nus de toute végétation, défavorables donc à la croissance de l'arbre. Sur l'image de droite (massif du Puigmal), la limite est abrupte et régulière, passant de l'étage montagnard à l'étage alpin sans transition. Cette forme éconotonale est typique des milieux ayant subi une forte pression agro-pastorale dans les estives, et qui connaissent aujourd'hui une déprise favorisant la colonisation rapide de l'arbre.

exemple, la probabilité de contracter la COVID-19 dépend de facteurs individuels (respect des gestes barrières, comorbidité, âge, puissance du système immunitaire, etc.), mais également du niveau de prévalence de la maladie dans le quartier ou la ville où l'on vit, de la présence de clusters spatiaux. Ne pas considérer ces clusters revient à ignorer une partie des facteurs explicatifs du phénomène étudié. En analyse quantitative, cette omission du contexte mène à un biais inférentiel connu sous le nom d'erreur atomiste, ou erreur individualiste [Alker, 1969; Susser, 1973]⁷, qui consiste à attribuer à un groupe (ou, en analyse géographique, à un contexte spatial) des relations établies à un niveau individuel. Bien que moins connue et moins étudiée sur le plan théorique que l'erreur écologique, l'erreur atomiste est un biais majeur des études écologiques (*cf.* encart ci-dessous).

Erreur écologique et erreur atomiste

L'erreur écologique est un biais d'interprétation fréquent en analyse contextuelle, qui consiste à attribuer à des individus des relations statistiques qui ont été estimées à un niveau agrégé (le mot écologique faisant ici référence à ce niveau agrégé, ou contextuel). Elle a été formalisée par **Robinson [1950]**, un sociologue américain, qui a mis en relation le taux d'analphabétisme et le taux de personnes immigrées aux États-Unis. Au niveau agrégé (états), cette corrélation était négative, signifiant que plus le taux de personnes immigrées est élevé, plus le taux d'analphabétisme est faible. En revanche, au niveau individuel, la corrélation était positive, signifiant que les immigrés sont en moyenne plus analphabètes que les américains autochtones. L'erreur écologique aurait consisté à conclure le contraire, uniquement sur la base des estimations faites au niveau agrégé.

L'erreur atomiste (ou individualiste) est considérée comme l'inverse de l'erreur écologique, c'est-à-dire le fait d'inférer des conclusions à une échelle agrégée à partir de données acquises au niveau individuel uniquement. Ces deux erreurs - écologiques et atomistes - sont une manifestation d'un problème plus vaste, le MAUP (*Modifiable Areal Unit Problem*, **Openshaw [1977]**), qui stipule plus généralement que l'échelle d'agrégation des données a un impact sur les corrélations statistiques.

L'autre possibilité est que les contextes spatiaux soient impliqués dans les chaînes causales par un effet de modération de la relation étudiée. Selon la terminologie des relations causales proposée par **Bauman *et al.* [2002]**, un effet de modération se produit lorsque la relation entre un facteur causal X et une variable d'intérêt Y varie en fonction d'une troisième variable Z. Ce concept est donc proche de celui d'interaction statistique. En épidémiologie par exemple, un facteur de modération classique est le genre. Dans notre cas, ce facteur est le contexte,

7. Déjà décrit, sans le nommer, par **Dogan et Rokkan [1967]**.

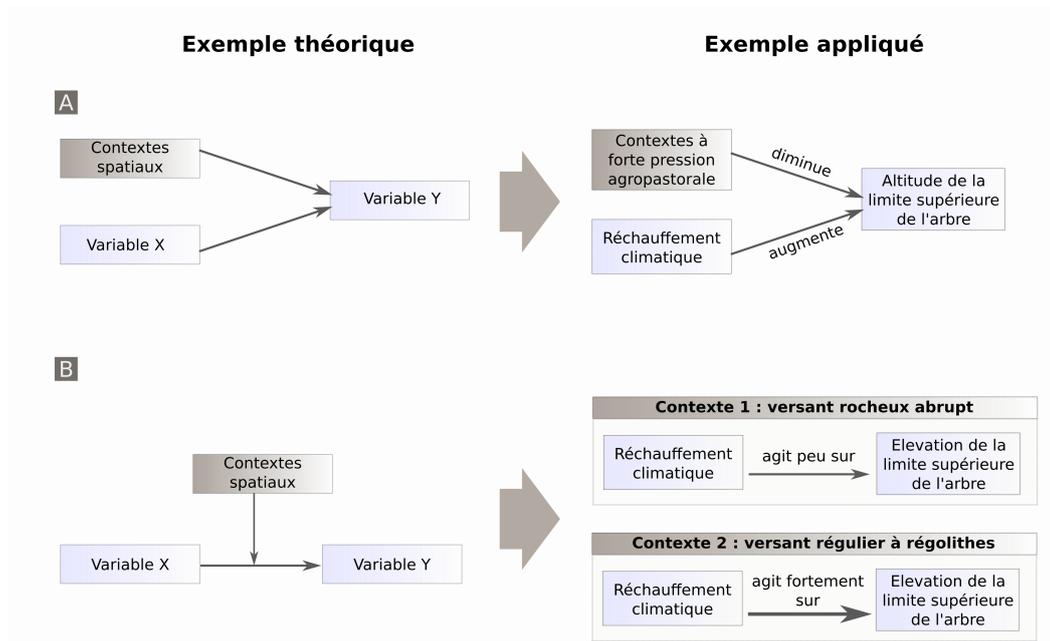


FIGURE 1.10 – Illustration théorique de l'importance de la considération des contextes spatiaux dans les chaînes causales. A. Les contextes spatiaux peuvent être un facteur causal direct du phénomène observé. Les ignorer mène à commettre l'erreur atomiste. B. Les contextes spatiaux peuvent agir comme facteurs de modération de la relation étudiée, engendrant sa non-stationnarité spatiale. Cette configuration représente la variante spatiale du paradoxe de Simpson, elle-même cas particulier d'erreur écologique.

signifiant que la relation entre X et Y varie selon les contextes spatiaux (figure 1.10). Pour illustrer ce propos, reprenons l'exemple de la dynamique de la *treeline* dans les Pyrénées. Au niveau régional, on étudie la relation entre l'évolution climatique (facteur X) et la variation altitudinale de la limite de l'arbre (variable Y). Cette relation va être positive dans la plupart des cas (plus le climat se réchauffe, plus l'arbre monte), mais peut néanmoins être nulle ou négative dans certains contextes locaux, en raison de leurs particularités intrinsèques (forte pression touristique locale, activités géomorphologiques intenses, etc.) qui affectent la relation principale en modifiant son sens ou son intensité. Dans ce cas, les contextes locaux sont bien des modérateurs de la relation étudiée. Il en est de même lorsque on étudie les effets d'une politique publique (comme une politique de promotion de la santé ou une politique d'aménagement du territoire par exemple) sur l'ensemble d'un territoire : cet effet n'aura probablement pas le même niveau d'efficacité partout, en fonction des contextes locaux.

Cette première partie visait à présenter les principes d'une vision théorique de l'espace organisé en systèmes de contextes spatiaux. Selon cette vision, la localisation initiale des contextes spatiaux est principalement déterminée par l'hétérogénéité

spatiale (effet vertical) issue de processus de long terme. Mais cela ne fait pas pour autant de la géographie la fille de l'histoire, selon les termes de [Pumain \[2003\]](#), car l'ensemble de la structuration tient son équilibre de la dépendance spatiale (effet horizontal) qui nourrit les interactions à la fois entre et au sein des contextes spatiaux, et renforce en retour l'hétérogénéité. Les effets propres à chacun de ces contextes, qualifiés selon les objets d'étude et les échelles d'effets de lieu, de quartier ou de site, contribuent donc à la production des structures spatiales à l'échelle supra-contextuelle et occupent en conséquence un rôle central dans les analyses géographiques.

Les études contextuelles, leur modélisation statistique et leurs limites

Après avoir défini dans la première partie le périmètre conceptuel et théorique des effets de contexte en géographie, cette deuxième partie va être consacrée aux études contextuelles proprement dites, d'un point de vue plus empirique donc. Une rapide synthèse de la façon dont les géographes se sont approprié l'étude des liens entre l'Humain et son environnement, et comment cette vision a progressivement évolué d'une position possibiliste à une acception plus systémique nous semblait un bon moyen de débiter cette partie. La deuxième sous-partie sera consacrée à une revue des principaux modèles statistiques régulièrement mobilisés dans cette littérature sur les effets de contexte, et qui reflète aussi une intéressante évolution, partant de méthodes relativement simplistes, d'un point de vue théorique comme statistique, jusqu'aux modèles multiniveaux, qui constituent toujours aujourd'hui une des méthodes les plus abouties pour mettre en valeur des effets contextuels à partir d'observations. Enfin, la dernière sous-partie a pour but de présenter une synthèse des erreurs théoriques et méthodologiques régulièrement rencontrées dans cette littérature, avec une attention particulière portée aux erreurs géographiques.

2.1 Le cheminement du déterminisme au contextualisme : historique de l'étude des liens entre l'humain et son environnement

2.1.1 Déterminisme, possibilisme et écologie humaine

L'étude des relations entre l'homme et son environnement, qu'il soit naturel, bâti ou social, est ancienne. Elle démarre des visions déterministes du 19^{ème} siècle, qui voyaient une relation causale, unidirectionnelle, liant l'environnement naturel (climat, relief, hydrologie, etc.) au comportement de l'homme et à l'organisation des sociétés. Cette pensée est portée par les géographes prussiens/allemands, notamment Carl Ritter (lui-même élève de Humboldt) puis Friedrich Ratzel. C'était déjà un progrès pour la géographie de l'époque, qui jusque-là n'impliquait pas l'homme comme élément essentiel des paysages, et cette période marque ainsi le

début de la géographie moderne. D'après [Johnston \[1991\]](#), cité par [Varenne \[2017\]](#), ce courant déterministe est une conséquence de la diffusion du darwinisme qui véhiculait, suite à la publication de l'*Origine des espèces* [[Darwin, 1859](#)], « cette hypothèse forte de l'influence [...] de l'environnement sur les formes de vie et donc aussi probablement sur les sociétés » [[Varenne, 2017](#), p. 27]. Pour autant, considérer le milieu naturel principalement comme une contrainte pour les sociétés a assez vite été discuté, notamment par les géographes français Elisée Reclus puis Vidal de la Blache [[Pelletier, 2016](#)], pour mener au début du 20^{ème} siècle au concept de possibilisme [[Febvre, 1922](#)], qui sera porté par l'école française de géographie [[Claval, 1998](#)]. La conception du lien entre milieu et société évolue vers une vision non plus causale, mais relationnelle : les milieux résultent également de ce que les hommes en font, et l'organisation des sociétés traduit donc cette relation réciproque entre milieu naturel et activités sociales et économiques. Cette école française régionalise alors les terrains d'étude, à la recherche de régions homogènes, dans lesquelles les interactions entre homme et environnement vont être décrites. C'est finalement déjà, à une certaine échelle, une façon de considérer l'importance des effets de contexte et leur fonctionnement systémique.

Néanmoins, comme le rappelle [Varenne \[2017\]](#), ce possibilisme français n'a parfois été considéré que comme une variante du déterminisme [[Haggett, 1965](#); [Harvey, 2011](#)], et que la réelle vision iconoclaste préconisant la réciprocité des liens humains-environnement – sans causalité unidirectionnelle, que ce soit dans un sens ou dans l'autre – n'apparaît qu'à travers la conception écologique de ces relations. En effet, parallèlement à l'affirmation de cette école française possibiliste se développe outre-atlantique le champ de l'écologie humaine, expression employée pour la première fois en 1907 par E.S. Richards, puis reprise par les sociologues de l'école de Chicago avec une étonnante modernité :

"Human ecology is fundamentally interested in the effect of position, in both time and space, upon human institutions and human behavior"
[[Park et Burgess, 1925](#), p. 64].

S'appuyant sur le vocabulaire de l'écologie (tel que la notion d'habitat ou de compétition), ce courant s'intéresse alors au jeu d'interactions instauré entre un individu (ou groupe d'individus) et ses conditions de vie, son « habitat » [[Rhein, 2003](#)], et va être intégré par l'école américaine de géographie dans la première moitié du 20^{ème} siècle. Déjà en 1923, le géographe Harlan Burrows considérait la géographie comme la science de l'écologie humaine, visant à étudier les interactions entre l'homme et l'environnement en minimisant le danger d'assigner aux facteurs environnementaux une influence déterministe qu'ils n'exercent pas [[Barrows, 1923](#)]. D'autres géographes ont suivi cette voie, comme Carl Sauer [[Sauer, 1927](#)], Richard Harstshorne [[Hartshorne, 1939](#)] ou Maximilien Sorre en France [[Sorre, 1948](#)]. L'écologie humaine constitue d'ailleurs encore aujourd'hui le cadre théorique et conceptuel revendiqué par beaucoup de géographes.

2.1.2 De l'écologie humaine aux effets de quartier : le poids des contextes locaux

Social scientists have long suffered an interest in neighborhood effects, which they view as a special case of context effects – the raison d'être of social science

[Oakes, 2004, p. 1930]

Cette écologie humaine va constituer le terreau fertile des premières études réellement contextuelles, dans le sens où les effets sur les activités humaines de contextes spécifiquement délimités vont être analysés de façon méthodique. Ces études sont intimement associées à cette sociologie urbaine du début du 20^{ème} siècle aux États-Unis sur les liens entre lieu de résidence (typiquement le quartier) et comportements individuels. Dès 1914, R.A. Woods, sociologue américain, écrivait :

"The neighborhood is the most satisfactory and illuminating form of the social extension of personality, of the interlacing and comprehensive complex of the interplay of personalities ; the social unit which can by its clear definition of outline, its inner organic completeness, its hair-trigger reactions, be fairly considered as functioning like a social mind" [Woods, 1914, p. 580]

Cette vision systémique – le quartier, le tout, tient son unité de l'interdépendance de ces éléments, les individus – a influencé la plupart des études contextuelles qui ont suivi, y compris celles sur les effets de quartier qui vont réellement prendre forme à partir des années 1950-60. Si l'on doit la première mention des effets de quartier à l'économiste Friedman [1955], ce sont les sociologues américains qui seront à l'origine de la première génération de ce type d'études. Dans cette littérature sociologique américaine, il s'agissait de comprendre les effets du contexte local à travers deux mécanismes : d'une part les effets de la concentration en un même lieu de groupes de personnes partageant des caractéristiques socio-économiques semblables, et d'autre part les effets du partage de contraintes communes propres au lieu de résidence [Marpsat, 1999]. C'est dans cette dynamique qu'a été lancé le projet Gautreaux aux États-Unis dans les années 1960, initiant lui-même une série d'autres expérimentations telle que le projet fédéral *Moving to Opportunity* (MTO) dans les années 1990. Toutes ces expérimentations de déségrégation visaient à mesurer les effets du quartier de résidence sur différents critères tels que l'insertion professionnelle, les résultats scolaires ou l'insertion sociale, le tout en conditions naturelles, avec des groupes témoins et en analyse longitudinale. Dans le projet Gautreaux (terminé en 1998), ce sont 7100 familles afro-américaines fortement désavantagées qui ont été relogées dans des quartiers de banlieue aisés composés majoritairement de blancs, avec une aide au logement. Dans le projet MTO, à nouveau 4500 familles noires défavorisées ont été divisées en trois groupes, l'un relogé dans des quartiers aisés, l'autre dans le quartier de leur choix et le troisième non-relogé pour servir de groupe témoin. Les résultats ont été contrastés, en fonction des quartiers et des individus. L'insertion a parfois été difficile, en raison

de la perte des liens sociaux et du mauvais accueil réservé aux familles, menant par exemple 30 % des ménages de Gautreaux à revenir en centre-ville [Bacqué et Fol, 2005]. Des analyses récentes ont néanmoins montré que dans le programme MTO, les résultats avaient été positifs pour les enfants qui avaient moins de 13 ans au début du programme, en termes d'augmentation du revenu à l'âge adulte [Chetty *et al.*, 2016]. Quoi qu'il en soit de ces contrastes, ces expériences montrent l'importance des contextes locaux dans la géographie des inégalités sociales, et la réalité de leur prise en compte dans la communauté scientifique.

Toujours dans la littérature anglo-saxonne, l'analyse des effets de quartier a également été largement développée à partir des années 1970 dans le cadre de la géographie électorale. Impulsé par John Cox et Ron Johnston, cette voie de recherche est partie du constat que les comportements électoraux tendent à être spatialement structurés, au-delà de ce qui serait attendu en ne tenant compte que de la composition sociale de la population, et que cette structuration répond à des effets de quartier. Ces études tendent donc à mettre en relation les comportements électoraux avec le contexte spatial de l'électeur, et à considérer que ce contexte local influence sa décision à travers différents mécanismes. Le premier de ces mécanismes proviendrait des interactions sociales, comme l'illustrent certaines expressions des précurseurs de ces hypothèses : "people who talk together vote together" ou "conversion through conversation" [Cox, 1969; Miller, 1977; Johnston, 1983]. Plusieurs études empiriques ont en effet montré, à partir de moyens statistiques plus ou moins sophistiqués, que la propension à changer de vote est influencée par l'entourage de l'électeur, famille et réseau social (voir par exemple Johnston [1983]; Pattie et Johnston [2000]), entourage qui est, en tout cas avant l'avènement des réseaux sociaux sur Internet, essentiellement contraint à un niveau local, typiquement au sein du quartier de résidence. D'autres études ont démontré l'importance de la composition socio-économique du quartier ou des institutions locales [David et Van Hamme, 2011]. Il convient de noter néanmoins que certains auteurs ont aussi dénoncé une surestimation de ces effets de quartier sur les comportements électoraux, en raison de biais autant théoriques (mauvaise spécification des mécanismes en jeu) qu'empiriques (manque de données pertinentes, notamment au niveau individuel, voir par exemple Gallego *et al.* [2016] et Blok *et al.* [2018]).

2.1.3 Au-delà du quartier : l'apport récent de la géographie de la santé

Les liens entre la santé et l'espace sont explorés depuis longtemps et ce domaine scientifique constitue plus que jamais aujourd'hui un champ investi et prometteur pour lutter contre les inégalités socio-spatiales en tout genre, révélées de façon redoutable lors de la crise du coronavirus qui touche en priorité les ménages et les territoires les plus précaires. De nombreux chercheurs s'y investissent, et la question

des effets de contexte y est centrale.

Pour comprendre ces liens entre géographie et santé et la façon dont ils sont étudiés, il convient en premier lieu de distinguer deux grands types de domaines sanitaires : celui des maladies infectieuses contagieuses, et celui des maladies chroniques non transmissibles (cancers, dépression, obésité, etc.) et des comportements associés (tabagisme, alimentation, activité physique, etc.). En termes épidémiologiques, l'objectif n'est pas le même : dans le cas des maladies infectieuses, on cherche à comprendre la géographie des vecteurs directs (virus ou bactéries) ou indirects (habitats des hôtes) et les dynamiques de diffusion spatiale des épidémies par contagion, alors que dans le cas des maladies chroniques non transmissibles, on cherche principalement à explorer la géographie des comportements humains qui en sont au moins partiellement à l'origine, ou les caractéristiques de l'environnement (social, bâti ou naturel) susceptibles de les affecter. De la cartographie du choléra à Soho en 1854 par Jon Snow aux contextes pathogènes de Max Sorre [Sorre, 1933], l'apport des analyses contextuelles en épidémiologie des maladies infectieuses est ancien et reconnu. En revanche, le rôle de l'environnement et des contextes locaux dans l'étude des comportements associés aux maladies chroniques non transmissibles est moins évident et toujours débattu [Von Korff *et al.*, 1992], bien que ce champ d'étude constitue aujourd'hui le terrain scientifique de bon nombre de géographes de la santé dans le monde occidental.

Jusqu'au début des années 1990, la majorité des modèles théoriques ne considérait que les facteurs individuels (tels que les facteurs biologiques, cognitifs ou psychologiques) et/ou les facteurs sociaux (interpersonnels) comme déterminants des comportements liés à la santé (modèle des croyances relatives à la santé [Hochbaum, 1958], théorie sociale cognitive [Bandura, 1977], théorie du comportement planifié [Ajzen, 1985], etc). Pendant longtemps, en épidémiologie, les variables contextuelles n'ont été vues que comme des proxies de variables individuelles impossibles à obtenir, avec la position sous-jacente que les déterminants des maladies étaient mieux conceptualisés au niveau individuel qu'au niveau contextuel [Diez-Roux, 2000]. Les années 1970 voient apparaître dans le champ de la psychologie environnementale des théories stipulant que l'environnement (naturel, bâti ou social) dans lequel évolue l'individu pouvait avoir une influence sur ses comportements [Barker, 1968; Bronfenbrenner, 1979]. C'est dans la continuité de ce courant de pensée que des chercheurs en épidémiologie et en santé publique ont défini les principes du modèle socio-écologique [McLeroy *et al.*, 1988; Stokols, 1992], qui reste aujourd'hui le cadre théorique le plus approprié – même si son appellation semble être un peu passée de mode – pour appréhender la complexité des liens entre l'environnement et les comportements liés à la santé (figure 2.1). Ce modèle considère (i) que les déterminants des comportements de santé sont multifactoriels (caractéristiques individuelles, environnement social, environnement physique et environnements politique et organisationnel), (ii) que ces dimensions interagissent, autrement dit que les effets de l'une sont dépendants des caractéristiques de

l'autre et (iii) que les types de déterminants varient selon les comportements et les personnes.

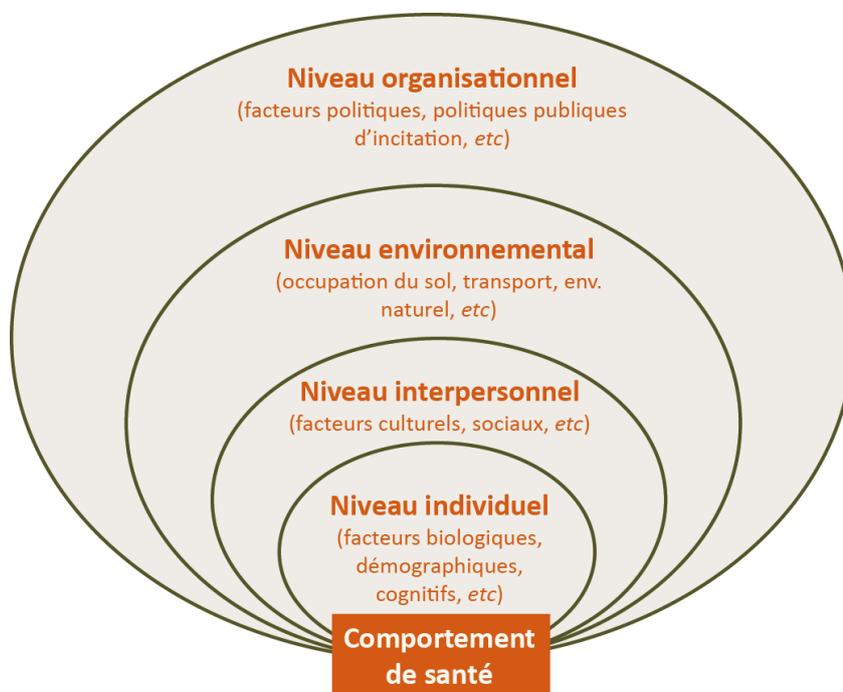


FIGURE 2.1 – Vision schématique du modèle socio-écologique des déterminants de la santé, attribué à McLeroy *et al.* [1988].

C'est dans ce cadre théorique que les premières études écologiques mettant en relation les caractéristiques du quartier de résidence et la santé vont commencer à se multiplier, à partir des années 1990. On pré-suppose alors que les inégalités dans les conditions à la fois sociales et physiques des environnements de résidence (qualité et quantité des ressources disponibles, offres alimentaires, criminalité, pollution de l'air, etc.) engendrent des inégalités spatiales de santé. Diez-Roux et Mair [2010] identifient deux générations d'études. La première met en relation des états de santé avec des variables socio-économiques issues du recensement de la population et agrégées au niveau de l'unité administrative de résidence des participants. Elles démontrent en général que vivre dans des quartiers défavorisés est associé à une plus grande mortalité et à une plus grande prévalence de maladies chroniques (cancers, dépressions, diabète, obésité, maladies cardiovasculaires). Néanmoins, elles ne permettent pas d'identifier des mécanismes causaux, puisque l'environnement socio-économique à un niveau agrégé n'est qu'un proxy des déterminants réels. La deuxième génération d'études va ainsi s'appliquer à mesurer directement et plus finement ces caractéristiques sociales et physiques des espaces vécus. On équipe ainsi les personnes de capteurs pour mesurer l'exposition réelle à la pollution de l'air, on caractérise l'environnement bâti (accessibilité aux transports en commun,

état des infrastructures, offres alimentaires, etc.) et l'environnement social (normes sociales, organisation sociale des lieux, etc.) autour du lieu de résidence. Dans le projet Spotlight visant à identifier les déterminants environnementaux de l'obésité en Europe, l'environnement bâti des quartiers de résidence des participants avaient par exemple été caractérisé à partir des images de Google Street View dans chaque segment de rue (état des trottoirs, présence de graffitis ou de déchets, présence de passages piétons, etc.), ce qui avait permis d'établir une typologie des quartiers en termes de potentiel obésogénique [Feuillet *et al.*, 2016a], puis de la mettre en relation avec les niveaux d'obésité [Roda *et al.*, 2016b].

Il se dégage depuis une dizaine d'années ce que l'on pourrait considérer comme la troisième génération d'études contextuelles en géographie de la santé. Jusque-là, la géographie des contextes était vue comme statique, ces contextes étant très principalement centrés sur le quartier de résidence, en partie en raison des contraintes liées à la disponibilité des données administratives. Plusieurs auteurs dénoncent alors le caractère partiel de cette position – ce *residential trap* [Chaix, 2009] – qui échoue à prendre en compte les expositions à des contextes multiples inhérents à la mobilité quotidienne des individus [Macintyre *et al.*, 2002; Cummins *et al.*, 2007; Chaix, 2009]. Le titre de l'article de Petrović *et al.* [2020] est très évocateur à ce sujet : "Freedom from the tyranny of neighbourhood : Rethinking sociospatial context effects". Au-delà de la question de la multiplicité des contextes, ce sont aussi leurs limites, souvent floues et superposées, qui y sont discutées. Selon Kwan [2012], l'incertitude qui règne autour de la délimitation géographique des contextes exerçant une influence réelle sur les individus mène à des biais analytiques importants, problème qu'elle définit par l'expression "uncertain geographic context problem", conceptuellement assez proche du MAUP d'ailleurs, auquel il fait référence par l'appellation. Ces réflexions ont donné naissance à la conceptualisation d'un raisonnement par espaces d'activités (combinaison des lieux de résidence, de travail, de loisirs, etc.) plutôt que par quartier de résidence seulement, dans une dimension dynamique et non statique, et centré sur l'individu et non l'unité spatiale [Chaix *et al.*, 2013; Perchoux *et al.*, 2013]. Le contexte devient donc un objet mobile et idiosyncrasique, aux limites complexes, floues ou chevauchantes. Cette perspective nécessite de concevoir les expositions contextuelles quotidiennes en trois dimensions (aquarium spatio-temporel), tel que le permet le cadre de la time-geography par exemple [Hägerstrand, 1970; Kwan et Lee, 2004].

Cette troisième génération d'études, qui fait de la géographie des contextes le point crucial, constitue donc un basculement conceptuel important, passant de la considération de contextes fixes (typiquement les quartiers administratifs de résidence) à des contextes dynamiques et mobiles (espaces d'activité), plus proches de l'exposition réelle des individus à leur environnement de vie quotidien. Alléchantes en théorie, ces hypothèses soulèvent en revanche un certain nombre de difficultés méthodologiques en termes d'acquisition et d'analyse de données [Perchoux *et al.*, 2019]. Des solutions innovantes sont actuellement proposées dans

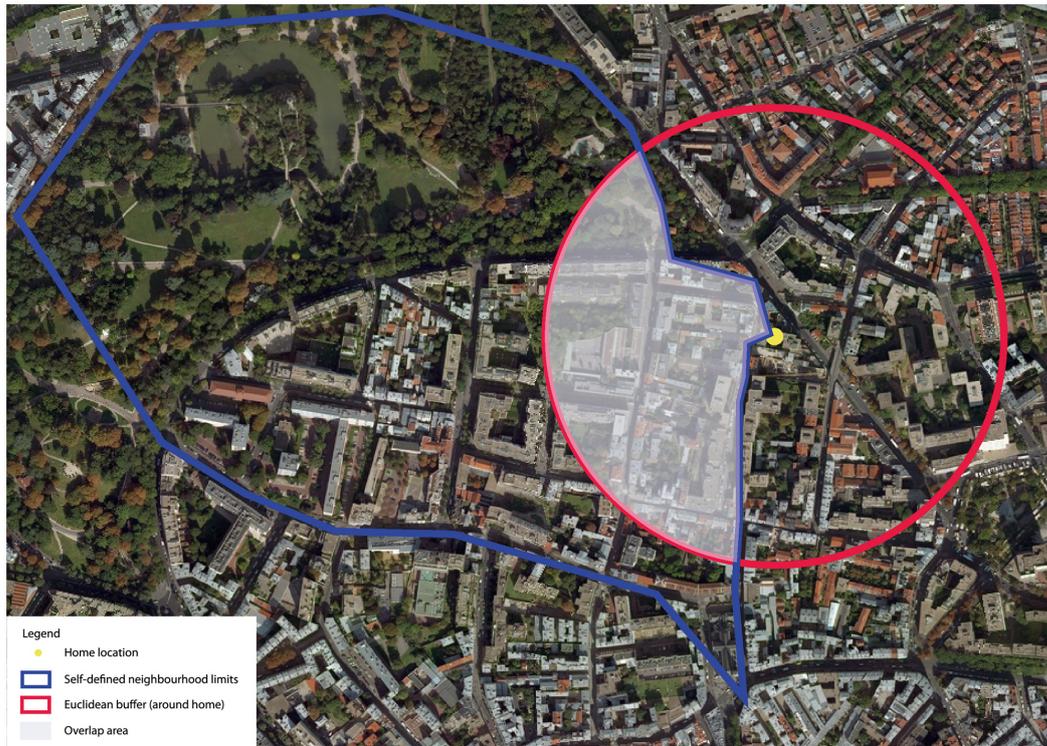


FIGURE 2.2 – Exemple de délimitation du quartier de résidence sur la base de la perception individuelle. Le point jaune représente le lieu de résidence du participant. Souvent, le contexte résidentiel supposé affecté les comportements individuels est défini par l’unité administrative d’appartenance ou par un buffer centré sur l’adresse de résidence, mais on constate dans cet exemple que le quartier perçu (en bleu) présente des limites singulières, qui différeront selon les pratiques spatiales de chaque individu. Ce constat encourage à privilégier une vision idiosyncrasique des effets contextuels en géographie humaine. Exemple tiré de Charreire *et al.* [2016].

la communauté des géographes de la santé, comme le suivi individualisé par des GPS et des capteurs intelligents [Chaix *et al.*, 2016] ou la délimitation des espaces d’activité sur la base de la perception individuelle, à travers des questionnaires cartographiques [Charreire *et al.*, 2016; Vallée, 2017]. Si des données individuelles pertinentes réussissent effectivement à être acquises à travers ces nouvelles solutions (figure 2.2), la question de leur analyse statistique reste centrale pour confronter les hypothèses théoriques à la réalité et démontrer des effets contextuels. La section suivante a donc pour objet de présenter un rapide historique et un état des lieux de la modélisation statistique des effets de contexte, qui sera complété dans la partie 3.

2.2 La modélisation statistique des effets contextuels en analyse transversale : modèles contextuels, multiniveaux et spatiaux

L'analyse empirique des effets contextuels, en géographie comme ailleurs, a été au cours de ces dernières décennies directement associée à l'évolution des moyens de mesure du contexte d'une part (notamment grâce au développement des logiciels SIG), mais aussi et surtout à l'augmentation des capacités computationnelles et des logiciels de traitement de données. Cette explosion des ressources, qui rendent aujourd'hui possibles des estimations complexes (modèles bayésiens ou simulations de Monte Carlo sur de larges échantillons, par exemple) en quelques secondes et à la portée de presque tous/tes, a considérablement accéléré la capacité de la communauté scientifique à confirmer ou infirmer des hypothèses théoriques jusque-là quasi indémontrables.

Dans cette dynamique, la modélisation statistique des effets de contexte a considérablement progressé, comme la partie qui vient a pour but de le démontrer. Mais avant cela, il convient de justifier en un mot l'intérêt même de cette modélisation. Généralement, la modélisation statistique peut servir deux desseins : (i) décrire des ressemblances, des corrélations dans un jeu de données (pour par exemple mettre en évidence des grands facteurs explicatifs) et (ii) prédire des évolutions ou des localisations non observées. C'est le premier point qui nous intéresse ici, et plus particulièrement l'effet de l'agrégation des données (dans des contextes) sur des variables d'intérêt (par exemple un état ou un comportement de santé). Les modèles statistiques (ou stochastiques) ont la particularité et la force d'intégrer intrinsèquement l'incertitude (due à d'éventuelles erreurs de mesure ou de mauvaise spécification) à travers le terme d'erreur (les résidus), contrairement aux modèles physiques (ou déterministes). L'autre force des modèles statistiques est le cadre inférentiel qui y est associé, basé lui-même sur différentes lois statistiques (notamment la loi des grands nombres). Ce cadre permet de mesurer l'incertitude des estimations à travers les probabilités, et va finalement nous autoriser à annoncer, avec une certaine chance de se tromper, l'existence éventuelle d'effets contextuels dans nos données. Sans ces modèles, nous serions contraints d'en rester aux suppositions, sans pouvoir distinguer ce qui relèverait du hasard, des erreurs de mesure ou de la mauvaise qualité de l'échantillon.

Pour synthétiser, la modélisation statistique des effets de contexte a été divisée en trois familles de modèles : les modèles contextuels, les modèles multiniveaux et les modèles spatiaux, qui vont être présentés plus en détail ci-dessous.

2.2.1 Les modèles contextuels

Les modèles contextuels sont des modèles de régression multiple (moindres carrés ordinaires - MCO) qui incluent parmi ses prédicteurs (variables explicatives) à la fois des variables individuelles et des variables contextuelles, c'est-à-dire des variables caractérisant des groupes d'individus [Diez-Roux, 2002]. Ce choix de modélisation passe donc par un processus de désagrégation des caractéristiques des groupes, permettant de rapporter toutes les mesures au niveau micro. Typiquement, ce sont ces modèles qui ont été utilisés lors de la première génération d'études contextuelles, dans laquelle des indicateurs de niveau individuel (comportements, états de santé, etc.) sont mis en relation avec des variables issues du recensement de la population, au niveau de l'unité administrative donc. En associant simultanément variables individuelles et contextuelles dans le modèle, on cherche à identifier un effet contextuel significatif en contrôlant les facteurs individuels supposés jouer un rôle potentiel de confusion, c'est-à-dire en assurant un raisonnement toutes choses égales par ailleurs. Ce type de modèles pose problème dans la mesure où on place au même niveau des processus opérant à des échelles différentes. En outre, la dépendance des observations au sein des contextes (plusieurs individus se voient attribués exactement les mêmes caractéristiques contextuelles) n'est pas prise en compte dans l'estimation des paramètres. Enfin, d'un point de vue statistique, ce choix a pour conséquence d'exagérer considérablement les tailles d'échantillons pour les macro-unités, augmentant ainsi le risque d'erreur de première espèce [Bressoux, 2010].

Les modèles à régimes spatiaux peuvent aussi être considérés comme un type de modèles contextuels. Ils consistent à intégrer dans un modèle MCO (ou éventuellement un modèle spatial) les contextes spatiaux en tant que variables dichotomiques, et de tester les interactions entre les variables explicatives du modèle et chaque terme dichotomique, par exemple avec le test de Chow [Anselin, 1989]. Une variation significative des termes d'interaction indique un effet contextuel (voir un exemple dans Myers *et al.* [2015]). Ce type de modèles est intéressant et se rapproche conceptuellement des modèles multiniveaux, mais présente l'inconvénient de considérer les contextes comme des boîtes noires, et de multiplier de façon importante les coefficients à estimer (d'autant plus si les contextes sont nombreux), limitant la parcimonie du modèle.

2.2.2 Les modèles multiniveaux

Ce n'est réellement qu'à partir des années 80-90 que les études contextuelles s'intéressant aux relations entre individus et lieux vont connaître un essor remarquable, avec l'adoption de la modélisation multiniveau comme cadre méthodologique privilégié. D'abord appliqué en sciences de l'éducation [Goldstein, 1987], ce cadre méthodologique va permettre de dépasser plusieurs verrous propres aux données

géographiques et que l'on croyait jusque-là difficilement surmontables. Comme en science de l'éducation où les élèves appartiennent à des classes, ces classes appartenant elles-mêmes à des établissements, nous vivons dans des lieux, ces lieux appartenant à des villes, ces villes à des régions, etc. Cette structure hiérarchique, multi-échelles, pose des problèmes statistiques dans les analyses quantitatives, puisque (i) l'agrégation des données dans des groupes remet en cause l'indépendance des observations¹, (ii) les relations statistiques sont susceptibles de changer en fonction des échelles (principes du MAUP et de l'erreur écologique définis dans l'introduction générale) et (iii) ces relations sont aussi susceptibles de varier selon les contextes. La force de la modélisation multiniveau, et la raison pour laquelle elle est si appropriée à l'étude des effets de contexte ou de quartier, est qu'elle permet précisément de spécifier des modèles au niveau micro (individus) et au niveau macro (contextes) de façon simultanée [Owen *et al.*, 2016] tout en permettant la non-indépendance des observations au sein des contextes. Il faut insister sur le progrès méthodologique que cela représente en comparaison des modèles contextuels utilisés jusqu'alors, dans lesquels des variables contextuelles (c'est-à-dire agrégées) étaient ajoutées aux variables individuelles dans des régressions linéaires classiques, ne tenant pas compte de la hiérarchie et de la dépendance des données et n'étant donc pas à même de démontrer de façon pertinente des effets de contextes. La modélisation multiniveau a été très rapidement employée en géographie sociale quantitative [Jones, 1991; Jones *et al.*, 1992], mais surtout en géographie de la santé pour mettre en évidence les effets du contexte sur les comportements et les états de santé. Chaix *et al.* [2005] la qualifiait ainsi de *gold standard* des analyses contextuelles. De nombreuses études ont trouvé grâce à elle les moyens de démontrer statistiquement – de façon distincte des effets individuels – l'effet d'appartenance à des contextes d'échelles variées (pays, régions, villes, quartiers) sur des états de santé (par exemple Von Korff *et al.* [1992]; Jones et Duncan [1995]; Ecob [1996]; Rundle *et al.* [2007]) ou des comportements liés à la santé tels que la consommation de drogue ou d'alcool (par exemple Duncan *et al.* [1997]; Rountree et Clayton [1999]). Cette modélisation multiniveau a indéniablement redonné un second souffle à l'analyse des effets contextuels, et est indissociable de sa progression.

La force de la modélisation multiniveau réside donc dans sa capacité à décomposer la structure résiduelle à tous les niveaux d'observation des données, du niveau individuel aux niveaux agrégés, pour tenir compte de cette dépendance et évaluer un possible effet contextuel. Typiquement, il est conseillé de débiter les analyses par ce qu'on appelle un modèle vide, c'est-à-dire un modèle sans variable explicative.

1. La proximité des observations au sein des contextes spatiaux a en effet toutes les chances de résulter en une dépendance spatiale, selon la loi de Tobler. Or, en modélisation statistique, la dépendance des observations viole une des hypothèses de Gauss-Markov et ne permet donc pas d'estimer des effets non biaisés. La dépendance a en effet pour conséquence de surestimer les effets (car l'estimation des intervalles de confiance des coefficients de régression dépend du nombre d'observations).

Prenons un exemple s'approchant des applications qui seront présentées en détail dans le chapitre 3. On cherche à étudier l'indice de masse corporelle (IMC) de personnes habitant dans Paris et sa petite couronne, pour déterminer dans quelle mesure la variation de cet IMC est due à des effets individuels ou contextuels. Chaque individu est donc caractérisé par une valeur d'IMC, et par une appartenance à un contexte spatial, disons par exemple la commune de résidence. Dans un modèle MCO classique, la variable d'appartenance à la commune serait introduite comme variable explicative, et les résidus du modèle ne serait estimé qu'au niveau individuel, malgré la présence de cette agrégation des observations au sein des communes. La force du multiniveau va être d'estimer également, en plus des résidus individuels (différences entre l'IMC des individus et l'IMC moyen de leur commune de résidence), des résidus au niveau des communes. Concrètement, on va estimer la moyenne des IMC par commune, et observer la manière dont cette moyenne communale varie par rapport à la moyenne générale. Parallèlement, on mesure la variation individuelle au sein de chaque commune (figure 2.3). Un principe statistique fondamental est que la somme de ces deux variances (variance entre individus dans chaque commune et variance entre les communes) égale la variance totale des observations. Ainsi, il devient très simple de calculer la part de la variance totale qu'on peut attribuer aux communes : il s'agit de l'indice de corrélation intraclasse (ICC). Cet indice se situe nécessairement entre 0 à 1. Une valeur de 0 indiquerait une absence d'effet contextuel (il n'y a pas du tout de variation entre les communes) tandis qu'une valeur de 1 impliquerait une absence de variation individuelle au sein de chaque commune (la totalité de la variance serait observée entre les communes). Cette distinction entre variance inter- et intracontextuelle dans les modèles vides fait songer à une classique analyse de la variance (ANOVA). La différence est qu'en modélisation multiniveau, on applique une pénalité à la variance contextuelle, qui est fonction du nombre d'observations dans chaque contexte. Ainsi, s'il y a très peu d'individus dans une commune, la variance dans ce contexte va tendre vers 0. En modélisation multiniveau, il n'est donc pas nécessaire d'avoir un nombre minimum d'observations dans les contextes.

Ce modèle vide est dit à constante aléatoire, car la constante du modèle (c'est-à-dire l'IMC moyen) est libre de varier dans chaque contexte. Après avoir estimé ce modèle vide, l'étape suivante consiste à introduire des variables explicatives, d'abord au niveau individuel puis, potentiellement, au niveau contextuel.

Surtout, la modélisation multiniveau offre la possibilité de faire varier les effets individuels par contexte (modèle multiniveau à pentes aléatoires, *cf.* figure 2.4), ce qui constitue un outil très puissant. Quand les effets individuels (par exemple le sexe et l'âge) sont laissés fixes, cela équivaut à considérer qu'ils sont identiques, quels que soient les contextes spatiaux. Ainsi, l'effet de l'âge sur l'IMC serait le même dans chaque commune, même si la composition de ces communes varie fortement. Les effets contextuels peuvent jouer sur le niveau moyen du phénomène étudié (modèle à constante aléatoire), mais tout laisse à penser qu'ils

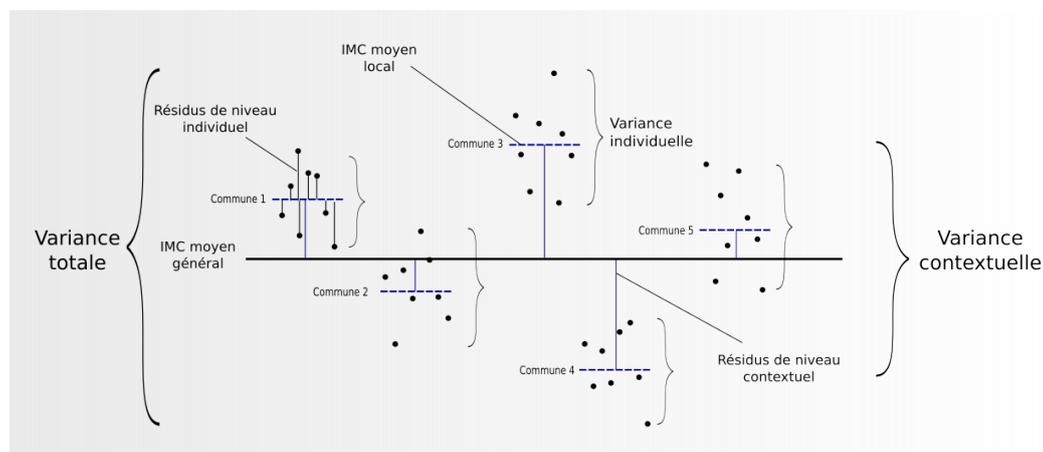


FIGURE 2.3 – Décomposition de la variance et des résidus au niveau individuel et au niveau contextuel : un principe essentiel de la modélisation multiniveau. Application théorique à l’IMC (indice de masse corporelle) au sein de communes. Inspiré de Merlo *et al.* [2005]

peuvent jouer également sur la relation entre ce phénomène et ses déterminants (le contexte spatial peut être un modérateur de ces relations, *cf.* figure 1.10). Le modèle multiniveau dit à pentes aléatoires permet d’estimer un coefficient de régression par contexte spatial et ainsi d’embrasser toutes les configurations possibles d’hétérogénéité territoriale des relations statistiques (figure 2.4), ou en d’autres termes la variété des effets contextuels. De fait, une nouvelle composante aléatoire est intégrée au modèle : la variance des pentes. Sa valeur est une source d’information sur l’existence d’effets contextuels. En effet, il est possible que les effets contextuels ne jouent pas sur la moyenne de la variable dépendante, mais sur les relations qu’elle entretient avec ses déterminants ou l’intensité de leur relation. Dans ce cas, le modèle indiquera un ICC faible, mais une variance des pentes fortes. S’arrêter à l’ICC peut ainsi conduire à conclure à tort à l’absence d’effets contextuels.

Comme pour l’estimation des résidus au niveau contextuel, la méthode applique une pénalité à ces coefficients (ou pentes) aléatoires, qui est fonction du nombre d’observations dans les contextes. Cette pénalité « force » le coefficient local vers la moyenne de tous les coefficients quand le nombre d’observations est petit. Par exemple, si le coefficient moyen associé à l’âge dans notre modèle est égal à 0,3 (signifiant qu’une augmentation de l’âge d’un an est associée à une augmentation de 0,3 de l’IMC), alors même dans un modèle à pentes aléatoires, cet effet sera très proche de 0,3 dans les contextes comptant peu d’individus (mais pourra varier fortement dans les autres contextes sur-représentés).

Ces modèles à pentes aléatoires restent étonnamment sous-utilisés dans la littérature scientifique, malgré leur intérêt. Cela peut être dû à des problèmes de convergence des estimations, car ces modèles intègrent un nombre important de

paramètres aléatoires à estimer – surtout si on intègre plus de deux niveaux, ce qui est tout à fait possible – nécessitant de disposer de suffisamment d’observations pour compenser le manque de parcimonie.

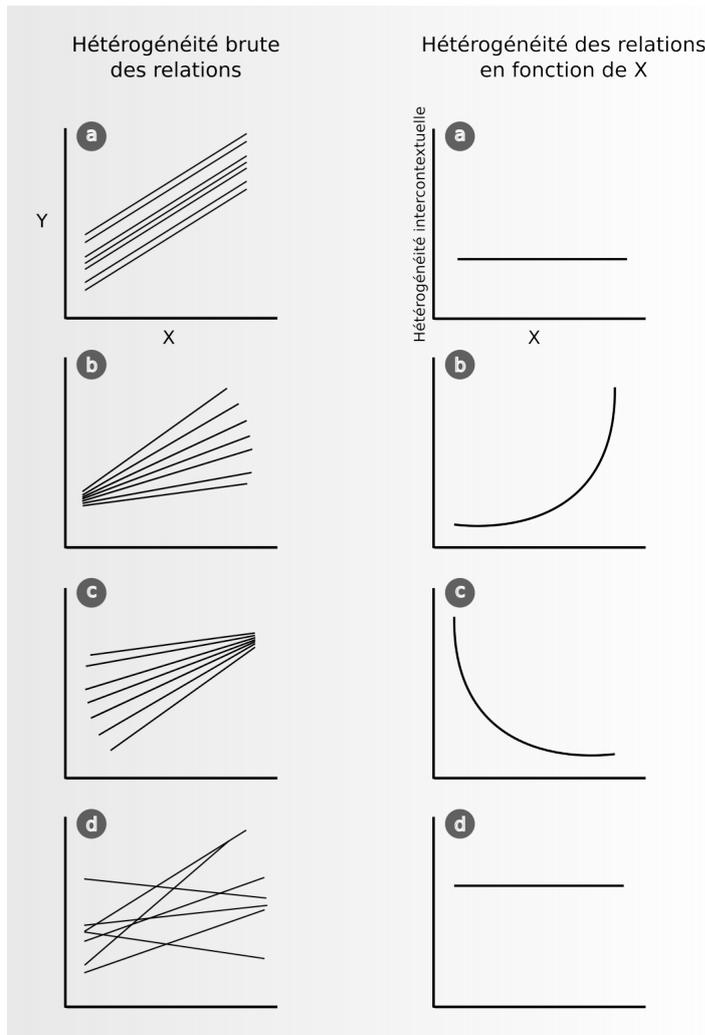


FIGURE 2.4 – Illustration théorique de la capacité des modèles multiniveaux à modéliser des relations statistiques territorialement hétérogènes. En A, le niveau moyen de la variable dépendante varie selon les contextes, mais pas la pente (c’est-à-dire la relation avec la variable indépendante), qui reste identique. En B et en C, le niveau moyen et la pente varient, mais dans le même sens et en fonction des valeurs de X (par exemple en B, l’hétérogénéité des valeurs moyennes de Y est plus grande pour les valeurs élevées de X, et inversement en C). En D, les pentes varient dans des sens parfois opposés et indépendamment des valeurs de X. Inspirée de [Duncan *et al.* \[1998\]](#).

Les principales forces de la modélisation multiniveau résident donc dans son cadre inférentiel solide, basé sur la décomposition résiduelle et l’estimation de

composantes à la fois fixes et stochastiques, ainsi que dans sa capacité à isoler les effets contextuels (c'est-à-dire les effets d'agrégation des observations dans des contextes spatiaux) des effets individuels. Ce dernier point est très important : dans les modèles classiques où les résidus ne sont estimés qu'au niveau individuel (c'est-à-dire au niveau des observations micro), il n'est pas possible de conclure de façon certaine à des effets contextuels - même si les coefficients associés à des variables contextuelles sont significatifs - car les effets individuels (ou effets de composition, dans le langage des épidémiologistes) ne sont pas distingués des effets contextuels d'un point de vue statistique.

Néanmoins, la modélisation multiniveau présente aussi deux principales limites² :

1. Tout d'abord, elle nécessite de considérer les contextes spatiaux comme des entités pré-définies, et donc potentiellement déconnectées des processus réels. Typiquement, ces entités sont des maillages administratifs. Cette limite peut avoir des implications majeures. Ainsi, le modèle peut amener à conclure en l'absence d'effets contextuels, quand en réalité les effets contextuels opèrent bien, mais à un autre niveau que celui de l'unité administrative :

"If the ICC is 0, it does not necessarily mean that the neighbourhood context is not important compared with individuals' factors. Rather, an alternative reason could be that the geographical boundaries we use to define the actual neighbourhoods do not correspond with the boundaries that shape the relevant environment for the concrete individual health outcome" *Merlo et al. [2005]*.

Ce problème a déjà été souligné dans la littérature, mais il n'en reste pas moins que la plupart des études continuent d'utiliser des maillages administratifs pré-définis sans garder le recul nécessaire quant aux biais potentiels qui en découlent.

2. Ensuite, la modélisation multiniveau considère l'espace de façon discrète, et ne tient ainsi pas compte des distances et de la non-stationnarité au sein des contextes, ni des interactions spatiales entre les contextes. Ils sont donc vus comme des boîtes homogènes, contiguës et exclusives, ce qui manque évidemment de réalisme dans bien des configurations.

2.2.3 Les modèles spatiaux (et locaux)

2.2.3.1 Les modèles spatiaux

Parallèlement à la modélisation multiniveau qui considère donc l'espace de façon discrète, sous forme d'unités spatiales, il existe d'autres familles de méthodes de modélisation statistique intégrant l'espace cette fois de façon explicite, c'est-à-dire

2. Une discussion complète et très intéressante des limites théoriques et méthodologiques de la modélisation multiniveau est proposée dans l'article d'*Owen et al. [2016]*.

en impliquant les coordonnées géographiques de chaque observation. Dans ces familles, les notions de distances et de proximité sont centrales.

La première famille de méthodes englobe les modèles spatiaux (modèle spatial autorégressif, modèle à erreur spatiale, modèle de Durbin spatial, etc.). Ces modèles consistent à intégrer des variables spatiales décalées (dérivées soit de la variable endogène, soit des variables exogènes, soit des résidus) comme variables exogènes pour corriger l'autocorrélation spatiale des résidus. Cette autocorrélation viole à nouveau l'hypothèse d'indépendance des observations. L'autocorrélation est également en général l'indication de facteurs inobservés (dans le modèle) étant eux-mêmes spatialement structurés. En intégrant des variables décalées créées *ad hoc*, on cherche à combler l'absence de ces facteurs latents inobservés : "Latent unobservable influences related to culture, infrastructure, recreational amenities and a host of other factors for which we have no available sample data can be accounted for by relying on neighboring values taken by the dependent variable" [LeSage, 2008, p. 20]. D'une manière un peu différente, les méthodes de filtrage spatial (*eigenvector spatial filtering*) permettent d'aboutir au même résultat [Griffith, 2003].

Les régressions spatiales sont surtout employées par les économètres, pour qui l'espace n'est qu'un générateur de problèmes statistiques, quand il représente une source d'informations pour les géographes. L'intégration de ces variables permet en effet d'estimer des coefficients de régression sans biais, mais la limite de ces méthodes est d'ordre théorique : ces variables spatiales décalées sont supposées remplacées des facteurs spatiaux inobservés mais, étant vidées de tout sens géographique réel, le modélisateur n'est pas beaucoup plus avancé sur les processus spatiaux réels ayant engendré une telle structure spatiale.

2.2.3.2 Les modèles locaux à coefficients spatialement variables

Depuis les années 1990, de nouvelles méthodes de modélisation spatiale non-paramétriques ont été proposées et ne cessent depuis de se développer dans de nombreux champs disciplinaires. Toutes ces méthodes consistent à estimer des coefficients de régression (c'est-à-dire des relations statistiques) spatialement variables et ainsi à mettre en évidence la non-stationnarité spatiale, elle-même potentiellement révélatrice d'effets contextuels. Il existe plusieurs méthodes aboutissant à ce résultat, comme la méthode d'expansion des paramètres [Casetti, 1972] ou les approches bayésiennes [Gelfand *et al.*, 2003]. Une méthode en particulier est sortie du lot, notamment pour son caractère très accessible. Il s'agit des régressions géographiquement pondérées [Brunsdon *et al.*, 1996; Fotheringham *et al.*, 2002]. Le principe de cette méthode est d'estimer les relations statistiques sur la base du voisinage géographique des observations uniquement et non sur l'échantillon complet.

En d'autres termes, pour chaque observation – que ce soit un point ou un polygone, mais qui doit nécessairement être localisée – on estime un modèle de régression locale en ne tenant compte que de ses voisins. Ces voisins sont définis selon un schéma pré-établi (cela peut être un nombre de voisins les plus proches ou une distance – il s'agit de la fenêtre (*bandwidth*) dans la figure 2.5) et sont éventuellement pondérés en fonction de la distance selon des fonctions au choix, par exemple exponentielle, gaussienne, etc., cf. figure 2.6).

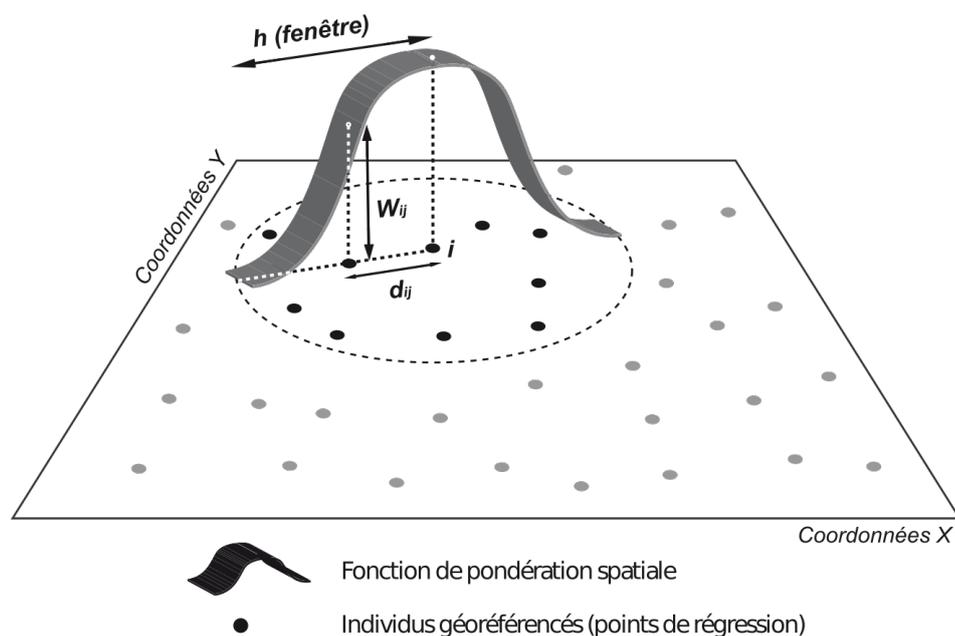


FIGURE 2.5 – La modélisation GWR est une méthode de régression locale basée sur la définition d'un noyau. Ce noyau (ou fenêtre) détermine l'emprise des observations locales utilisées pour l'estimation de chaque coefficient de régression. Éventuellement, une pondération (souvent inversement proportionnelle à la distance) est appliquée au voisinage.

Cette méthode d'analyse spatiale locale incorpore donc explicitement la géographie (à travers les coordonnées géographiques des observations), contrairement aux modèles multiniveaux, et représente en cela un outil exploratoire puissant pour mettre en valeur l'hétérogénéité spatiale des relations statistiques (non-stationnarité spatiale). Une relation étant estimée pour chaque observation, des cartes de relations peuvent être dessinées. Si le modèle est correctement spécifié et qu'il révèle malgré tout une hétérogénéité spatiale des relations, alors celle-là peut être assimilée à des contextes spatiaux inobservés. En effet, comme le notent [Janko et al. \[2019\]](#), "unmeasured covariates can confound, modify, or mediate the effects of an exposure on a health outcome, and do so in ways that induce spatial variability in the exposure-outcome relationship" [[Janko et al., 2019](#)]. Si cette variabilité spatiale est structurée, on peut considérer que ces facteurs inobservés se superposent à des contextes spatiaux dont les effets propres modifient la relation

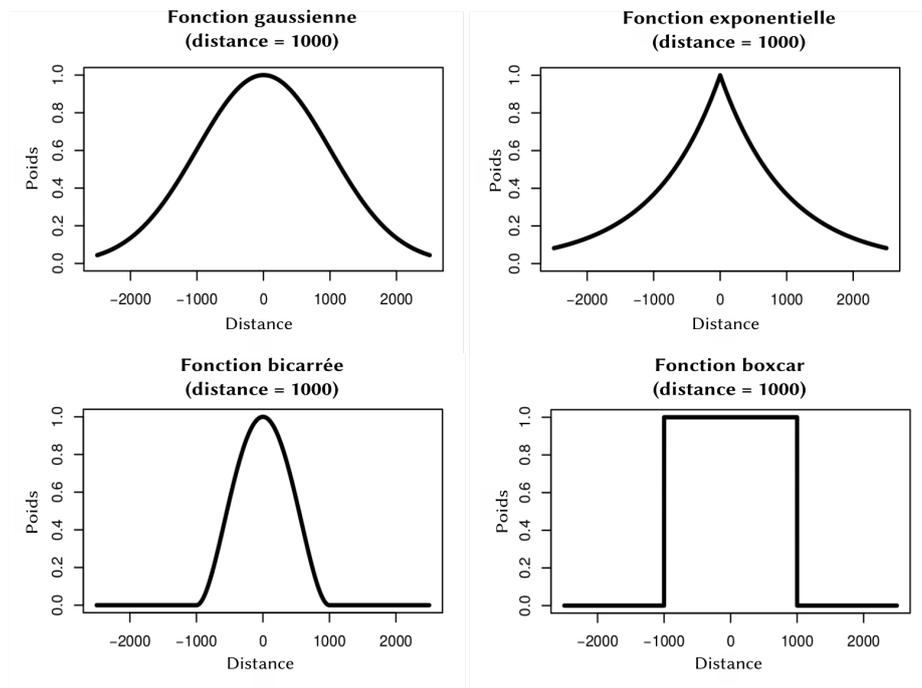


FIGURE 2.6 – Illustration du schéma de paramétrage de la GWR. La méthode repose sur le choix préalable d’une matrice de voisinage (le noyau, ou la fenêtre), qui peut être fonction de la distance, de la contiguïté ou du nombre de plus proches voisins. Une fois le voisinage défini, il peut être pondéré selon différentes fonctions de la distance.

principale étudiée. Dès lors, il devient particulièrement intéressant de tenter de délimiter et de caractériser ces contextes. Dans une étude sur les déterminants environnementaux (densités de routes, de pistes cyclables, d’espaces verts, etc.) de la marche en région parisienne à partir de la cohorte NutriNet-Santé [Feuillet *et al.*, 2016b], nous avons ainsi régionalisé le territoire en fonction des structures spatiales des coefficients GWR, puis avons réalisé une analyse discriminante des contextes révélés pour identifier leurs principaux caractères distinctifs parmi les variables environnementales initiales. Cela avait finalement permis de caractériser les contextes spatiaux au sein desquels les déterminants environnementaux de la marche étaient fortement significatifs, pouvant ensuite potentiellement servir de levier pour des politiques de santé publique ciblées. Dans la même logique, mais à l’échelle supra-régionale cette fois, nous avons répété cette stratégie sur la France entière [Feuillet *et al.*, 2018] pour aboutir à 9 clusters sur l’ensemble du territoire métropolitain. Chaque cluster avait ensuite été caractérisé par un vecteur de coefficients GWR pour établir une hiérarchie spécifique et spatialisée des déterminants environnementaux de la marche. La méthode hybride proposée dans le chapitre 3 est fondée sur cette stratégie de classification des coefficients GWR, mais cette fois en tant que première étape dans un *workflow* menant à une

modélisation multiniveau des effets contextuels.

Pour synthétiser, le point fort de la GWR est de tenir compte de la non-stationnarité spatiale dans un espace continu. Sa principale faiblesse relève de son cadre inférentiel – en lien avec les problèmes de tests multiples et de multicollinéarité locale [Wheeler et Tiefelsdorf, 2005] – ce qui l’a amenée à ne devoir être considérée par certains que comme un outil exploratoire. Son autre point faible, en tout cas dans l’optique de la modélisation des effets de contexte, est de ne pas tenir compte de l’agrégation des données dans ces contextes, comme le fait la modélisation multiniveau. Les effets contextuels potentiels ne peuvent ainsi pas être strictement distingués des effets individuels.

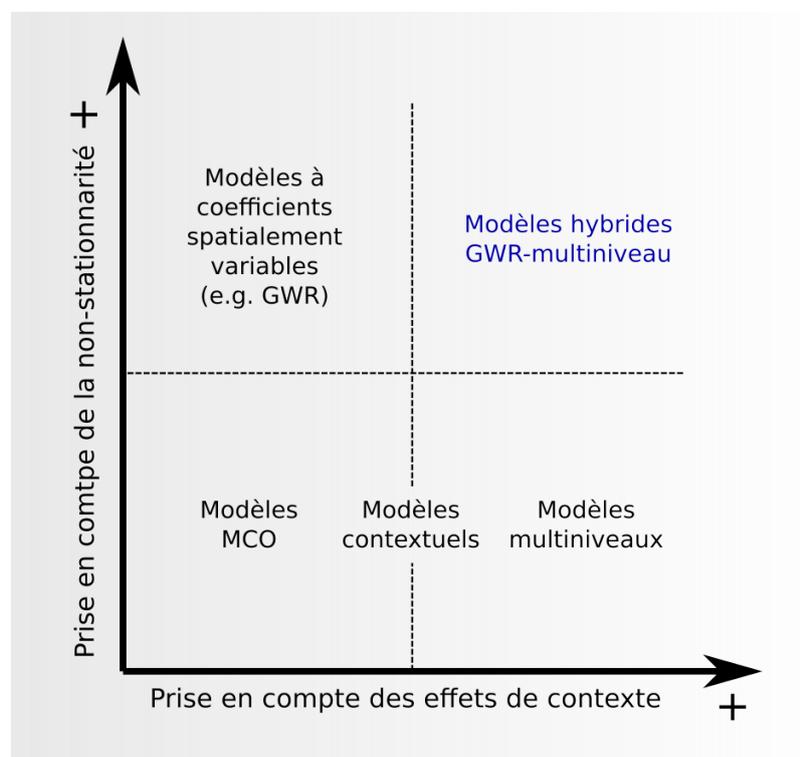


FIGURE 2.7 – Forces et faiblesses des principales méthodes actuelles de modélisation statistique des effets de contexte.

Les études contextuelles en géographie ont ainsi suivi un cheminement lent mais serein, impliquant dans un premier temps des chemins de causalité simples et unidirectionnels dans un cadre déterministe, puis articulant progressivement des mécanismes plus complexes, relationnels et systémiques. L’autre constat qui peut être fait au regard de l’évolution de ces études en général est que malgré les cloisonnements disciplinaires, on retrouve des similarités d’un point de vue théorique, que ce soit en géographie, en épidémiologie, en sociologie, en écologie :

on cherche à estimer les effets de l'appartenance à un espace commun, soit sur les comportements, soit sur la santé, ou sur tout autre état individuel observé à une échelle infra-contextuelle.

2.3 Typologie des erreurs associées à la modélisation statistique des effets contextuels

Démontrer une quelconque causalité dans les effets de contexte à l'aide de modèles statistiques relève quasiment de la mission impossible, même si des méthodes tentent de s'en approcher, comme les modèles à équations structurales [Bovaird, 2007] ou les modèles à variables instrumentales [Loney et Nagelkerke, 2014] sur données longitudinales. La simplification de ces effets y est donc inévitable – d'autant plus si l'on considère la difficulté parfois d'acquérir des données contextuelles pertinentes en regard des hypothèses théoriques posées, et ce à la bonne échelle – et a pour conséquence que les études contextuelles quantitatives incorporent un nombre important d'imprécisions et d'erreurs, autant théoriques que méthodologiques. La partie qui suit présente une tentative de typologie des erreurs régulièrement rencontrées dans la littérature scientifique sur la modélisation statistique des effets contextuels, avec une attention particulière apportée aux erreurs géographiques, c'est-à-dire celles qui concernent explicitement l'espace ou les processus spatiaux, comme déjà évoqué partiellement dans la partie 2.1. Cette typologie nourrira elle-même la troisième partie de ce travail, qui trouvera sa justification dans sa volonté de corriger une partie de ces erreurs par la définition d'un protocole de modélisation statistique spécifique.

Un biais peut être défini à la fois comme un écart entre une estimation et la réalité, ou comme une erreur méthodologique menant à des résultats et des interprétations erronées. Dans le contexte de ce travail, les biais définis ci-dessous représentent donc des erreurs communément faites dans les études contextuelles empiriques, ayant pour conséquence des interprétations partielles ou incorrectes des causalités réelles. Ces erreurs ont été regroupés en trois groupes : les erreurs géographiques (comprenant les biais spatiaux et contextuels), les erreurs causales et les erreurs analytiques (illustrés dans la figure 2.8).

La typologie qui suit est très principalement inspirée des travaux de géographie de la santé et d'épidémiologie sociale, et s'applique donc en premier lieu aux sciences sociales. Le domaine de la santé publique constitue un terrain parfaitement adapté à l'application d'études contextuelles, ce qui justifie la production d'un nombre considérable d'analyses variées au cours des dernières décennies. Les essais théoriques visant à décrire certaines de ces erreurs figurent dans le corpus, et ont évidemment inspiré cette partie. Néanmoins, une partie des éléments décrits, notamment des erreurs géographiques, peuvent tout à fait s'appliquer à des objets

inertes de l'environnement naturel typiquement étudiés en biogéographie, en géomorphologie ou en écologie.

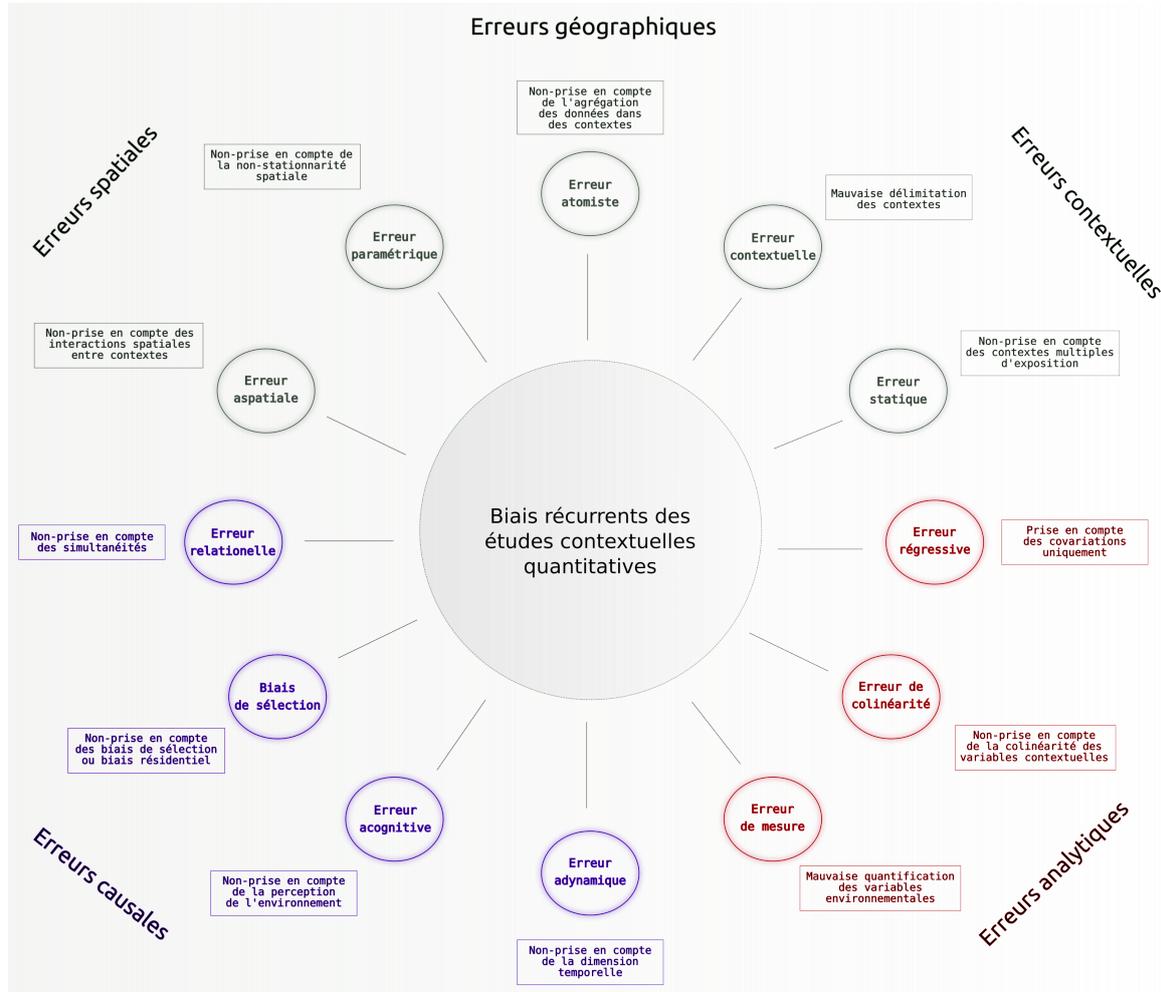


FIGURE 2.8 – Typologie des erreurs récurrents en modélisation statistique des effets contextuels.

2.3.1 Erreurs géographiques

Déjà évoquées dans la partie 2.1, les erreurs géographiques, qui nous intéressent tout particulièrement ici, peuvent être divisés en deux types : les erreurs spatiales, qui concernent l'omission des processus spatiaux, et les erreurs contextuelles, qui concernent la mauvaise spécification contextes spatiaux.

2.3.1.1 Erreurs spatiales

Les erreurs spatiales concernent l'omission des deux processus qui caractérisent les données spatiales : la dépendance spatiale et la non-stationnarité spatiale. L'omission de la dépendance spatiale, qualifiée d'erreur aspatiale, se produit quand les interactions spatiales entre les contextes spatiaux ne sont pas prises en compte. C'est typiquement une erreur de la littérature mobilisant la modélisation multiniveau standard. En effet, dans ce type de modélisation, les contextes spatiaux sont vus comme indépendants les uns des autres, même s'ils sont contigus [Owen *et al.*, 2016]. En outre, si la dépendance spatiale y est bien considérée au sein des contextes à travers la décomposition de la variance inter- et intraclasse, la méthode standard n'intègre en revanche aucune information spatiale explicite sur la position géographique des observations au sein des contextes. Cela revient à dire que la distance entre les observations n'est pas prise en compte dans les contextes, alors qu'elle peut varier significativement. C'est là le prix à payer pour la considération du territoire sous la forme de lieux (vision discrète) et non d'espace (vision continue).

L'omission de la non-stationnarité spatiale constitue la seconde erreurs spatiale, et est qualifiée ici d'erreur paramétrique. Erreur paramétrique dans la mesure où elle consiste à adopter une vision moyenne des relations entre environnement et observations micro sur l'ensemble d'une région d'étude, sans considération des variations locales (elles-mêmes dues aux effets de contexte, *cf.* § 1.3.2), comme c'est typiquement le cas dans les modèles paramétriques, à effets fixes. D'un point de vue de la statistique inférentielle, la non-stationnarité spatiale n'est autre que ce qu'on pourrait considérer comme la variante spatiale du paradoxe de Simpson [Simpson, 1951]. Ce paradoxe stipule qu'une relation observée dans plusieurs groupes peut s'inverser quand les groupes sont combinés. En termes spatiaux, cela se traduirait par exemple par le fait que la relation entre X et Y est négative quand elle est observée dans deux contextes spatiaux pris séparément, mais positive dans l'échantillon total. Cela correspond à ce que Jones [1993] et Duncan *et al.* [1996] ont appelé l'effet everywhere-is-nowhere. Janko *et al.* [2019] en donne un exemple théorique limpide : dans deux zones comprenant chacune 100 individus, la prévalence d'une maladie quelconque est de 30 %. Une intervention est mise en place, qui provoque une augmentation de la prévalence de 50 % dans la zone A, mais une diminution de 50 % dans la zone B, en raison de facteurs locaux inobservés. Au final, on observe 15 individus malades dans la zone A et 45 dans la zone B, mais la prévalence globale reste de 30 %. Un modèle paramétrique à effet fixe aurait estimé un coefficient nul pour cette intervention, c'est-à-dire une estimation erronée dans une zone comme dans l'autre.

Cette variante spatiale du paradoxe de Simpson est d'ailleurs elle-même une conséquence du *modifiable areal unit problem* (MAUP, Openshaw [1977]; Wilson [2013]) – l'échelle d'agrégation des données a un impact sur les associations

statistiques – et peut être considérée comme un cas particulier³ d’erreur écologique [Robinson, 1950], qui consiste à confondre les corrélations entre différents niveaux d’agrégation des données.

2.3.1.2 Erreurs contextuelles

Les erreurs contextuelles ont été divisés en trois types d’erreurs : l’erreur atomiste, l’erreur contextuelle et l’erreur statique. L’erreur atomiste peut être définie d’un point de vue théorique, ou d’un point de vue statistique. D’un point de vue théorique, elle se réfère à l’omission du contexte spatial en tant que facteur explicatif du phénomène étudié, point déjà évoqué dans le § 1.4.2, et illustré dans la figure 1.10, sur lequel on ne reviendra donc pas. Le point de vue statistique, en particulier issu du vocabulaire de la régression multiniveau, est plus strict. Elle désigne précisément dans ce cas le fait d’inférer des conclusions à une échelle agrégée à partir de données individuelles uniquement, et correspond donc à l’inverse de l’erreur écologique. Diez-Roux [2002] fournit l’exemple théorique suivant : on constate, à partir de données individuelles, que l’augmentation du revenu est associée à une diminution de la mortalité due aux maladies coronariennes. L’erreur atomiste est commise si le chercheur conclut que cette relation négative s’observe au niveau national. En effet, ce lien peut être positif quand il est estimé à un niveau agrégé. On comprend donc le lien fort qui lie cette erreur atomiste à l’erreur écologique et donc à l’erreur paramétrique mentionnée préalablement, puisque dans tous les cas il est question des conséquences de l’agrégation des données sur la relation étudiée⁴.

L’erreur contextuelle a été nommée ainsi pour désigner les conséquences d’une mauvaise délimitation des contextes spatiaux sur les inférences statistiques. Déjà évoquée dans le § 2.1 au moment de la description de la troisième génération d’études contextuelles en géographie de la santé, cette erreur fait référence au fait que très souvent dans les études quantitatives, les limites des contextes d’exposition sont prises par défaut, contraintes par la disponibilité des données et non par des hypothèses théoriques. Si cette faille est oubliée au moment de l’interprétation des résultats, le risque d’arriver à une conclusion biaisée est forte. C’est donc au chercheur d’estimer l’écart entre le contexte spatial disponible et le contexte spatial réel, comme l’expriment Diez-Roux et Mair [2010] à propos des effets de quartier mais qui pourrait s’appliquer à tout type d’effets contextuels :

"The relevant question in interpreting the results of studies becomes not

3. Cas particulier seulement – extrême en fait – car le paradoxe de Simpson implique une inversion des signes, contrairement à l’acceptation plus large de l’erreur écologique.

4. A noter que la vision statistique est volontairement privilégiée dans cette partie dédiée aux biais des études contextuelles, et que la démarche n’est donc pas explicative. Pour autant, les causes théoriques de ces biais jusque là mentionnés sont bien tous relatifs aux effets de contexte non ou mal appréhendés, comme spécifié dans le § 1.4.2.

whether the definition of neighborhood used in the research is the right one, but whether the measure for the spatial context available is likely to be a reasonable proxy for, or highly correlated with, the true causally relevant context" [Diez-Roux et Mair, 2010, p. 134].

Enfin, l'erreur statique, conceptuellement assez proche de la précédente, consiste à ne pas tenir compte des contextes multiples d'exposition. Cette erreur ne concerne bien évidemment que les études dont l'objet est mobile, comme les humains ou les animaux. Elle est particulièrement fréquente dans les études de modélisation des effets de quartier et a déjà été décrite, en lien avec la notion d'espace d'activités, par de nombreux auteurs. Ceux-là dénoncent la vision réductrice consistant à penser que les effets de contexte n'opèrent qu'à travers les interactions sociales au sein du quartier de résidence, comme si les individus n'étaient jamais confrontés à d'autres contextes spatiaux (par exemple Kwan [2012, 2018]; Petrović *et al.* [2020]; Vallée *et al.* [2011]). Il est en effet rare que les études empiriques sur les effets de quartier prennent en compte la variabilité de ces expositions individuelles au fil de la journée, suivant la vision idiosyncrasique des effets contextuels recommandés par Kwan [2018], pour des raisons méthodologiques principalement (*cf.* § 2.1).

2.3.2 Erreurs causales

Les erreurs impliquant un biais dans l'interprétation des relations causales entre contexte spatial et objet d'étude ont été regroupées sous l'appellation d'erreurs causales, et comportent quatre types d'erreurs.

La première, essentielle tant elle représente presque un paradigme, est ce que nous avons qualifié d'erreur relationnelle, selon la terminologie proposée par Cummins *et al.* [2007]. Cette approche remet en cause la pertinence d'une opposition entre effet contextuel et effet compositionnel sur la santé. Les auteurs considèrent que les mécanismes causaux liant les caractéristiques des populations et celles des contextes spatiaux avec la santé sont interconnectés, c'est-à-dire bidirectionnels (ou simultanés, dans le langage des économètres, qui englobent plus largement ce problème sous le concept d'endogénéité). L'exemple donné dans cette même étude est le suivant : la vie d'enfants grandissant dans un quartier spécifique peut être influencée par les aspects sociaux et matériels de ce quartier (valeurs et normes sociales, environnement bâti, offre éducative), mais les interactions sociales et les comportements de ces enfants, et la façon dont ils se comportent dans ce quartier une fois adultes, façonnent également l'environnement physique et social et contribuent à créer un contexte pour leurs voisins. En langage statistique, il y a alors simultanéité des relations de X vers Y et de Y vers X. Ce problème d'endogénéité est en théorie incompatible avec la régression. Comme Y est susceptible d'influencer X, il n'est de fait pas possible de mesurer l'effet de la variation de X sur Y. La relation estimée dans un modèle de régression sera donc biaisée.

La deuxième erreur est un biais de sélection, le biais résidentiel, très fréquent

dans les études transversales sur les effets de quartier (contrairement aux études expérimentales et longitudinales, où il est mieux contrôlé). Ce biais se produit quand les facteurs de sélection des contextes de vie par les individus étudiés (typiquement le lieu de résidence) ne sont pas pris en compte dans les modèles, alors qu'ils influencent à la fois la variable dépendante (la santé, les comportements, etc.) et les variables indépendantes [Handy *et al.*, 2006; Boone-Heinonen *et al.*, 2010]. Par exemple, dans la littérature sur les déterminants des choix modaux où ce biais est souvent rencontré, on pourrait conclure que les préférences individuelles pour la voiture ou les modes actifs sont expliquées par les caractéristiques des quartiers de résidence (éloignement aux centre-villes, densité de pistes cyclables, etc.) si les relations statistiques vont dans ce sens. Néanmoins, si au moins une partie des individus étudiés ont choisi leur quartier de résidence pour des raisons liées à leurs habitudes de transport, et que cette information n'a pas été intégrée dans le modèle statistique d'une manière ou d'une autre, les estimateurs seront biaisés. En d'autres termes, l'influence réelle de l'environnement sur les comportements sera mal quantifiée. En comparant 7 différentes alternatives méthodologiques pour contourner ce biais résidentiel (par exemple via l'introduction de variables instrumentales, d'un contrôle statistique, d'un design longitudinal, etc.), Mokhtarian et Cao [2008] recommandent finalement l'usage d'une modélisation longitudinale à équations structurelles.

La notion d'estimateur en régression

Dans le langage de la régression, les estimateurs font référence aux coefficients de régression (c'est-à-dire aux effets supposés des variables explicatives sur la variable que l'on cherche à expliquer), estimés sur la base d'échantillons de données. On leur donne ce nom, car ils ne peuvent être que des estimations des effets réels, ceux qui seraient observés si l'on disposait de l'ensemble des données de la population, sans aucune erreur de mesure. Comme c'est par définition impossible, on se suffit donc d'estimateurs. L'écart entre l'estimateur et l'effet réel définit le biais.

L'erreur acognitive dérive de l'omission, dans les modèles statistiques, de la perception qu'ont les individus de leur contexte d'exposition. Cette perception est supposée être, dans de nombreux modèles théoriques, un effet de modulation ou de médiation dans la relation entre l'environnement réel et la santé ou les comportements. Par exemple, l'effet de la présence d'espaces verts sur la pratique de la marche va varier selon que cette présence sera mesurée de façon objective (via une photographie aérienne par exemple) ou estimée sur la base de la déclaration des participants de l'étude, déclaration qui va varier subjectivement selon la perception de chacun. Plusieurs études ont effectivement montré un écart entre l'effet de l'environnement perçu et celui de l'environnement réel [Lin et Moudon, 2010; Boehmer *et al.*, 2007]. Si la perception de l'environnement est

bien un effet de médiation, alors les implications sont considérables puisque cela signifierait que l'environnement objectif (ou réel) n'est pas un déterminant direct des comportements individuels. Dans le projet Spotlight, nous avons par exemple mis en évidence que l'importance de cet écart entre les effets du perçu et du réel (ici sur les comportements obésogènes) dépendait du type de quartier (faible dans les quartiers denses) et du type de caractéristiques environnementales mesurées (faible écart pour le paysage alimentaire ou les aménités récréationnelles, mais fort pour les éléments relevant de l'esthétique du quartier) et qu'il convenait donc de combiner les deux dimensions dans les modèles statistiques pour limiter les biais d'interprétation [Roda *et al.*, 2016a].

L'erreur adynamique est due à la non-prise en compte de la dimension temporelle dans les modèles. Cette temporalité est présente à différentes échelles de temps. Tout d'abord, les contextes eux-mêmes ne sont pas statiques. Ils évoluent en permanence, par le biais de différents processus de transformation comme les aménagements, l'érosion, etc. Or, dans la plupart des études transversales, on caractérise le contexte à un instant t sans tenir compte de cette évolution. A plus court terme, les caractéristiques des contextes et des objets qui les habitent évoluent aussi au cours d'une même journée [Kwan, 2018; Le Roux *et al.*, 2017]. Par exemple, on caractérise souvent l'environnement socio-économique des contextes territoriaux sur la base des données du recensement de la population, ce qui ne le représente que partiellement. Ainsi, Le Roux *et al.* [2017] ont calculé des indices de ségrégation sociale en région parisienne à différents moments de la journée. Les résultats montrent qu'en fonction de l'heure, ces indices varient fortement dans l'espace, suivant la mobilité quotidienne des navetteurs. Des quartiers qui la nuit sont fortement ségrégués apparaissent socialement mixtes la journée. Là encore, les conséquences de cette omission sont la mauvaise estimation des effets réels du contexte.

La temporalité intervient aussi chez les individus, à travers les parcours de vie et les habitudes. Les contextes dans lesquels nous avons vécu, par le passé, ont forgé une partie de nos habitudes et de nos perceptions, ce qui n'est pas toujours considéré dans les études écologiques, où l'on ne tient souvent compte que des caractéristiques du quartier dans lequel on réside à un instant t , même si l'on y réside que depuis très peu de temps.

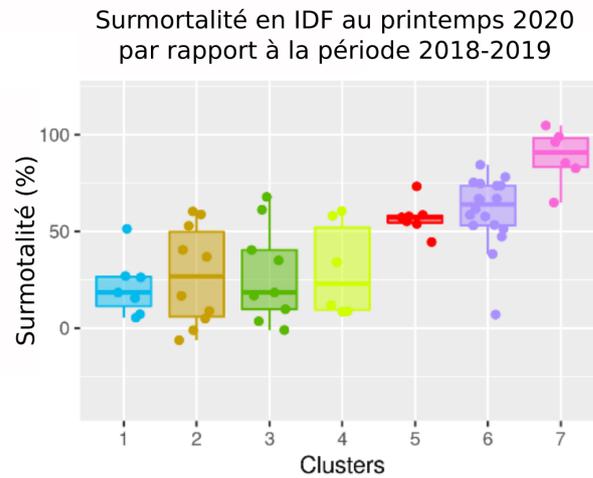
2.3.3 Erreurs analytiques

Les erreurs analytiques dérivent d'erreurs méthodologiques au sein même des modèles statistiques, que ce soit dans le choix de la méthode ou dans la quantification des variables. L'erreur régressive fait référence au caractère parfois inadapté de la régression pour mettre en valeur des effets de contexte. La régression reste la méthode statistique la plus utilisée dans ce domaine, en particulier en épidémiologie environnementale où son recours est presque systématique, or elle

n'est pas toujours appropriée pour modéliser ces effets. La régression est basée sur l'analyse des covariations de variables. Si l'une des variables étudiées varie peu ou pas, alors par définition la relation estimée avec une autre variable sera faible ou nulle, même s'il existe une structuration spatiale dans les processus étudiés (par exemple sous forme de hot-spots). Il peut être parfois plus pertinent de comparer les moyennes des variables étudiées au sein des contextes spatiaux, ou les co-occurrences d'événements, plutôt que les covariations. C'est ce type d'approches que nous avons privilégié dans le cadre d'une étude épidémiologique de la covid-19 au cours de la première vague du printemps 2020, en collaboration avec l'Observatoire régional de la santé d'IDF (figure 2.9). Par le biais d'analyses factorielles incorporant un certain nombre de dimensions socio-environnementales agrégées au niveau des EPCI, nous avons identifié 7 contextes territoriaux relativement homogènes en termes de niveau socio-économique, d'accessibilité aux soins et de densité de population. Dans chacun de ces contextes, nous avons estimé la surmortalité totale moyenne (surmortalité attribuée à la covid-19), avant d'évaluer les différences entre contextes par des ANOVA et des tests de signification honnête de Tukey. Les résultats ont montré des différences significatives entre contextes – avec sans surprise des moyennes très élevées dans les secteurs défavorisés et denses comme la Seine-Saint-Denis (*cf.* figure 2.9) – montrant ainsi que le recours aux analyses de covariations n'est pas toujours indispensable, notamment quand les contextes sont complexes et peu nombreux.

L'erreur de colinéarité consiste à ne pas tenir compte de la redondance des variables contextuelles dans des modèles multivariés, ce qui peut mener à des estimateurs volatiles. En outre, quand les variables sont spatialisées comme c'est nécessairement le cas pour des variables contextuelles, la colinéarité peut être locale et créer des problèmes d'estimation dans les modèles non-paramétriques d'analyses locales (c'est le cas pour la GWR). Enfin, l'erreur de mesure désigne la mauvaise quantification des variables contextuelles, qui peut bien entendu mener à des conclusions biaisées. Ce problème de quantification est fréquent, d'autant plus quand les contextes en question ne suivent pas des mailles conventionnelles (unités administratives par exemple), ce qui nécessite de faire des choix sur la façon d'agréger, ou au contraire de ventiler, des données pré-existantes.

Cette typologie des erreurs, probablement encore incomplète, montre à quel point les écueils théoriques et méthodologiques sont nombreux lorsqu'il s'agit de modéliser les effets de contexte à l'aide de l'outil statistique. Le fait que les contextes spatiaux supposés agir sur les objets étudiés en leur sein ne soient pas nécessairement des maillages pré-existants, et puissent être également multiples, superposés, ou non délimités de façon stricte (contours flous), représentent des défis méthodologiques particulièrement excitants pour les géographes quantitativistes.



Clusters 1 et 2 : quartiers socialement favorisés et peu denses (ouest et sud francilien)
Clusters 3 et 4 : quartiers socialement défavorisés et peu denses (est francilien)
Cluster 5 : quartiers socialement défavorisés et denses (Paris)
Cluster 6 : quartiers moyennement défavorisés et moyennement denses
Cluster 7 : quartiers fortement défavorisés et denses (principalement Seine-Saint-Denis)

FIGURE 2.9 – Comparaison de la surmortalité totale (toutes tranches d'âge) au printemps 2020 (par rapport aux deux années précédentes) en fonction des contextes territoriaux (clusters) en IDF. Chaque point représente un EPCI. Ces contextes sont dérivés d'une classification hiérarchique sur composantes principales. La comparaison de moyennes peut dans certains cas être plus pertinente que l'analyse de covariations, typique de la régression. C'est notamment le cas si les contextes sont complexes (caractérisés par de très nombreuses dimensions) et peu nombreux.

Dans la partie qui suit, notre contribution sera de tenter de remédier à une partie de ces erreurs géographiques, à l'aide d'une modélisation statistique hybride tenant compte à la fois des effets spatiaux et des effets contextuels. Cette solution sera testée sur des données réelles, relative à la géographie de l'obésité en région parisienne.

De la non-stationnarité spatiale à une géographie réaliste des contextes : vers un affranchissement des maillages pré-établis en modélisation multiniveau

La partie précédente a permis de mettre en évidence un certain nombre de limites théoriques et statistiques récurrentes dans le domaine de la modélisation des effets contextuels en géographie. La partie qui suit a pour but de mettre en œuvre un protocole méthodologique basé sur une hybridation de deux des principales méthodes de modélisation employées dans ce domaine, la GWR et le multiniveau. Ce protocole vise à utiliser les forces de chacune des méthodes pour combler les limites de l'autre, permettant notamment de s'affranchir des maillages administratifs pour aboutir à une géographie des contextes plus réaliste, directement dérivée de la non-stationnarité spatiale des données.

Bien que cette partie vise à déployer des modèles statistiques sur des données réelles, le choix a été fait de ne pas encombrer le texte d'équations. Ces équations ne sont en effet pas nécessaires à la compréhension du raisonnement théorique qui justifie la mise en place de cette méthodologie, ni sur les interprétations des résultats. En outre, ce protocole embarque des méthodes existantes et déjà démontrées mathématiquement par de nombreux auteurs. Plutôt que de recopier ici ces équations, nous invitons les lecteurs intéressés à consulter directement ces références, dont les principales sont citées dans le texte. En revanche, un effort a été fait pour décrire les principes de fonctionnement de chacune de ces méthodes, de façon littéraire et à l'aide d'illustrations schématiques.

La thématique d'application – la géographie de l'obésité – a été sélectionnée pour trois principales raisons. D'une part, elle se prête très bien à la question des effets de contextes spatiaux (on sait que les comportements obésogènes (activité physique et alimentation) sont en partie expliqués par les caractéristiques de l'environnement).

D'autre part, elle constitue une thématique sur laquelle des études contextuelles ont déjà été engagées, mais à une autre échelle [Feuillet *et al.*, 2020b]. Enfin, nous avons la chance de pouvoir disposer d'une base de données infiniment riche (données de la cohorte NutriNet-Santé), qui nous donne accès non seulement à des données individuelles (ce qui est presque essentiel à la modélisation multiniveau), et en outre géoréférencées à l'adresse de résidence (ce qui permet d'appréhender l'espace de façon continue et ainsi d'exploiter toute la puissance des modèles GWR).

3.1 Justification et présentation du protocole méthodologique combinant modèles GWR et multiniveau

Il est surprenant de constater que les études impliquant des modèles GWR s'arrêtent bien souvent à une analyse purement descriptive de la non-stationnarité spatiale, sans faire le lien avec les théories proposées dans la littérature sur les effets de contexte, sans interpréter de façon approfondie ni utiliser a posteriori les structures spatiales révélées. Cela équivaut à quitter la table après l'entrée. D'autre part, il est également manifeste que la quasi totalité des études impliquant des modèles multiniveaux dans le domaine de la géographie de la santé et de l'épidémiologie environnementale utilisent des unités administratives comme niveau d'agrégation des observations. Ainsi, la complémentarité de la GWR avec la modélisation multiniveau semble être sous-exploitée. On a d'une part une méthode au cadre inférentiel éprouvé et adopté unilatéralement par la communauté scientifique, mais qui par contre nécessite une délimitation des contextes spatiaux a priori qui ne correspond pas toujours à la réalité des faits géographiques (la modélisation multiniveau) et d'autre part une méthode au cadre inférentiel plus discuté et qui n'isole pas la variabilité résiduelle en fonction de l'agrégation des données, mais qui permet de révéler des contextes spatiaux de façon empirique (la modélisation GWR ou autres modèles à coefficients spatialement variables). En d'autres termes, les points faibles de l'une semblent pouvoir en théorie être compensés par les points forts de l'autre (figure 2.7). L'objectif est ici précisément de tenter de relier ces deux approches, en utilisant la GWR pour déterminer des contextes spatiaux qui seront subséquentement utilisés comme niveau hiérarchique dans un modèle multiniveau.

Cette stratégie part du principe que la non-stationnarité spatiale révélée par les coefficients locaux de la GWR peut être vue comme une indicatrice de contextes spatiaux inobservés qui agissent par effets propres (effets endogènes et exogènes, *cf.* figure 1.5). Exhibée de manière heuristique, directement dérivée des données, la structure spatiale de cette non-stationnarité nous informe en effet sur le bruit de fond de l'espace, bruit suffisamment fort pour affecter l'intensité ou le sens de relations statistiques. La position de Fotheringham *et al.* [2002] ou Oshan *et al.* [2020] est de considérer que les structures spatiales des coefficients GWR résultent d'un effet explicite de la géographie, dans le sens où, toutes choses étant égales par

ailleurs, un même stimulus peut provoquer une réponse différente sur le phénomène observé car il opère dans un lieu fondamentalement différent. Cet effet explicite de l'espace est parfois désigné par les mêmes auteurs comme un effet intrinsèque de l'espace. Cette interprétation peut toutefois induire en erreur : ce n'est pas l'espace qui est en soi différent d'un lieu à l'autre, ce sont ses composantes – et la façon dont elles sont localement arrangées – qui varient. On retombe sur la définition des contextes spatiaux qui a été donnée en première partie. Les deux raisonnements parviennent au même résultat (les relations statistiques varient d'un lieu à l'autre car elles sont affectées localement par des effets de contextes spatiaux inobservés), mais le second est différent car il n'implique pas les mêmes mécanismes de dynamiques de mise en place et d'évolution de ces lieux.

Quoi qu'il en soit, les cartes de coefficients GWR semblent pouvoir être utilisés comme un moyen de révéler des contextes spatiaux difficilement délimitables par la mesure directe. Ils sont difficilement délimitables car la nature et la disposition des composantes locales qui les caractérisent et qui leur donnent ce pouvoir de modulation des phénomènes étudiés sont par définition complexes, multidimensionnels et spatialement hétérogènes. Si ces contextes ainsi révélés par l'empirie agissent réellement sur les observations, alors on peut raisonnablement poser l'hypothèse que le modèle multiniveau qui les intégrerait comme niveau d'agrégation des données serait plus pertinent que le modèle multiniveau reposant sur des limites administratives ou autres délimitations construites a priori. C'est cette hypothèse que nous allons tester ici.

3.1.1 Aperçu des méthodes hybrides combinant modèles spatiaux et modèles multiniveaux

Avant de présenter la stratégie d'implémentation de la modélisation, il est intéressant de se renseigner sur les hybridations méthodologique déjà existantes. Un rapide état de l'art permet de constater que plusieurs études ont déjà tenté de combiner GWR (ou autres modèles à coefficients spatialement variables) et modèle multiniveau, de différentes manières. [Chen et Truong \[2012\]](#), dans leur étude des déterminants socio-économiques de l'obésité à Taïwan, ont d'abord appliqué un modèle multiniveau afin d'ajuster des odds ratios moyens d'obésité par unité administrative, qu'ils ont ensuite utilisés comme variable dépendante dans une GWR. D'autres auteurs ont combiné modèles spatiaux et multiniveaux, en incorporant de la corrélation spatiale dans les effets aléatoires. C'est le cas de [Hajat et al. \[2013\]](#) dans le cadre de leur étude sur la relation entre l'exposition à la pollution de l'air et le niveau socio-économique individuel. Dans un raisonnement assez similaire, [Park et Kim \[2014\]](#) ont développé une méthode de modélisation multiniveau spatialement filtrée (*spatially filtered multilevel model*) en appliquant un filtrage spatial aux résidus d'un modèle multiniveau, puis en incorporant les structures spatiales dérivées dans l'équation du modèle multiniveau. D'autres

extensions de modèles multiveaux intégrant des effets spatiaux ont été proposés par ailleurs [Dong *et al.*, 2016; Janko *et al.*, 2019].

Ces études apportent par leurs hybridations méthodologiques des solutions très utiles pour diminuer les biais d'estimations, notamment ceux dus à la dépendance spatiale, mais elles ont néanmoins toutes recours à des délimitations administratives pour définir l'agrégation des données, c'est-à-dire le niveau géographique dans lequel les chercheurs considèrent que les effets de contexte agissent. La seule étude (à notre connaissance bien évidemment) qui définit explicitement des niveaux géographiques sur une autre base que le découpage administratif, et qui les intègre ensuite comme niveaux dans un modèle multiniveau, est l'article d'Arcaya *et al.* [2012]. Dans cette étude, les auteurs ont cherché à modéliser les variations spatiales de l'espérance de vie aux États-Unis à travers une série de modèles multiniveaux, prenant en compte, comme niveaux géographiques d'agrégation des observations micro (qui sont en l'occurrence des *counties*, mais qui pourraient tout aussi bien être des individus), soit les États des États-Unis, soit des regroupements de *counties* contigus (régions) établis sur la base de la proximité spatiale uniquement (selon une matrice de pondération spatiale du voisinage définie au préalable). Chaque *county* peut faire partie de plusieurs régions (cross-classified structure). Cette méthode astucieuse permet au modèle d'intégrer des corrélations entre *counties* contigus, quels que soient les États auxquels ils appartiennent.

Néanmoins, la façon dont sont définies les régions dans cette étude (critère de proximité spatiale uniquement, selon un schéma de voisinage identique sur le tout le territoire) n'assure en rien que ces régions délimitent effectivement des contextes susceptibles d'affecter les observations micro par des effets propres. L'objectif affiché ici est justement d'établir des limites de contextes qui soient dérivées des observations, spécifiques donc au jeu de données utilisé.

3.1.2 Présentation de la stratégie d'analyse

La stratégie d'analyse se décompose en trois principales étapes :

- (i) Étape 1 : estimer une GWR multivariée ;
- (ii) Étape 2 : établir une classification spatiale multivariée des coefficients estimés par la GWR ;
- (iii) Étape 3 : estimer un modèle multiniveau intégrant la classification spatiale des coefficients GWR comme niveau d'agrégation des données.

Cette stratégie nécessite d'arrêter une série de choix de méthodes (méthodes de classification et de régionalisation des coefficients GWR) et de paramètres (schéma de voisinage de la GWR (*bandwidth*, abrégé ci-dessous en *bw*) et nombre k de contextes spatiaux) de la part du modélisateur, susceptibles d'affecter fortement les résultats. Par exemple, le changement du nombre de voisins considérés dans l'implémentation de la GWR a une implication en chaîne sur tous les traitements

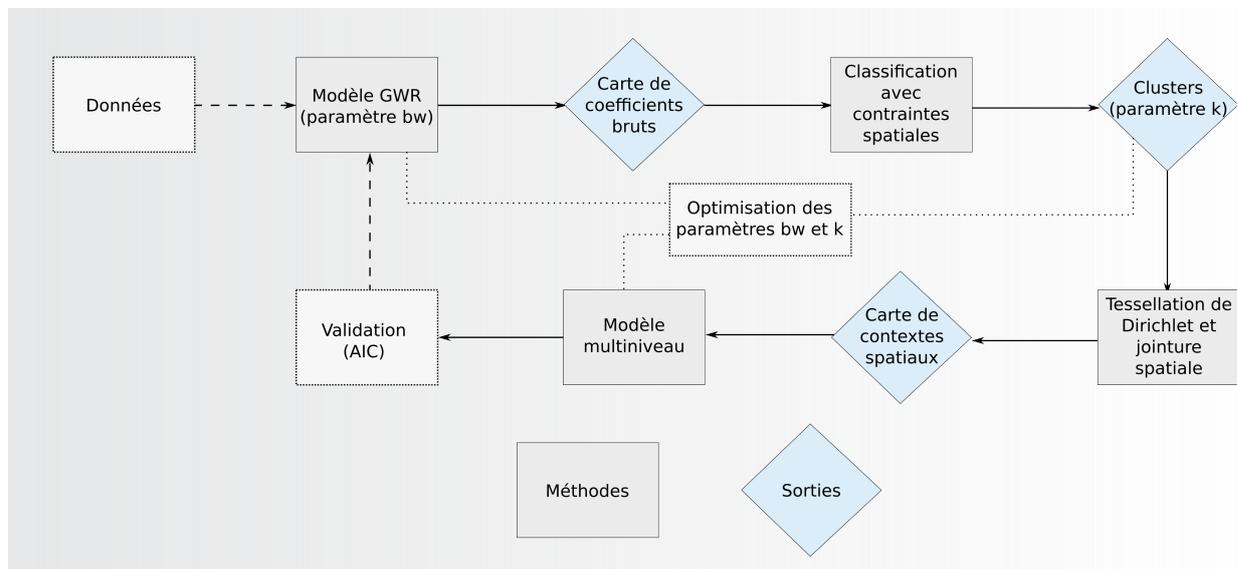


FIGURE 3.1 – Protocole méthodologique pour la mise en place du modèle hybride combinant modèle GWR et modèle multiniveau.

subséquents (nombre et localisation des contextes, performance global du multiniveau). Ces choix doivent donc être discutés et justifiés.

3.1.2.1 Optimisation de bw et de k

Le paramétrage de la matrice de pondération spatiale du voisinage dans la GWR repose sur deux éléments : (i) la définition d'une matrice de voisinage et (ii) la définition de sa pondération. Le voisinage d'une observation peut être lui-même défini de deux façons : selon un nombre de plus proches voisins ou selon une distance. Quand les observations ne sont pas régulièrement réparties dans l'espace, il est recommandé de préférer la méthode des plus proches voisins pour assurer à chaque estimation locale un nombre minimal suffisant d'observations et ainsi éviter des problèmes de non-convergence des calculs. Dans l'application qui sera traitée ci-dessous (géographie de l'obésité en petite couronne parisienne), nos observations sont des individus localisés à l'adresse de résidence (entités ponctuelles donc) dont l'espacement est approximativement fonction de la densité résidentielle. Pour cette raison, la méthode des plus proches voisins a été choisie.

Une fois la matrice de voisinage élaborée, il convient d'attribuer un poids aux voisins. Typiquement, ce poids est inversement proportionnel à la distance, selon une fonction choisie. Deux familles de fonctions existent : l'une permettant à l'ensemble des observations d'être pris en compte dans les estimations locales, mais attribuant un poids extrêmement faible aux observations éloignées (fonctions gaussienne et exponentielle) et l'autre attribuant un poids strictement nul aux observations

situées au-delà du voisinage préalablement défini (fonctions bicarrée, tricube et boxcar, cette dernière ayant la particularité d'être binaire, sans décroissance du poids avec la distance, donc). Ici, la fonction bicarrée a été choisie pour des raisons théoriques. Le but étant de définir des contextes spatiaux locaux, elle semble plus appropriée dans la mesure où elle permet aux estimations locales de ne pas être excessivement lissées par la prise en compte d'observations très éloignées. Avec une fonction gaussienne par exemple, le risque est d'aboutir à des cartes de coefficients présentant des tendances spatiales sous forme de gradients unidirectionnels, même dans le cas d'un nombre de voisins (bw) petit. Le choix entre fonction bicarrée et fonction tricube a un impact faible, les deux pondérations étant très proches.

Le choix du nombre de voisins (bw) est en revanche très impactant. Dans tous les outils permettant d'implémenter une GWR, il est proposé des algorithmes d'optimisation de ce nombre. Bien entendu, les critères d'optimisation sont basés sur la performance globale de la GWR entreprise (minimisation de l'AIC ou validation croisée par exemple). Néanmoins dans notre cas, ce n'est pas la performance de la GWR qui doit être maximisée, mais celle du modèle multiniveau intervenant à la fin de la chaîne, la GWR n'étant qu'une étape. En l'occurrence, c'est donc l'AIC du modèle multiniveau qui doit être minimisé. Ainsi, l'optimisation de bw dépend d'autres paramètres de la chaîne, qu'ils soient variables (nombre de clusters) ou fixes (contrainte spatiale de la classification des coefficients, cf. § 3.1.2.2), bornés ou non. Sachant en outre que la fonction objectif (l'AIC du modèle multiniveau) disposent potentiellement de plusieurs minima locaux, cela requiert, en langage mathématique, d'implémenter un algorithme d'optimisation sous contraintes d'un minimum global d'une fonction objectif complexe non-linéaire à plusieurs variables. Les algorithmes d'optimisation se sont énormément développés ces dernières années, en raison des progrès computationnels et du nombre considérable de domaines potentiellement concernés (en finance, en physique, en biologie, etc.). Celui qui été choisi ici est l'algorithme du recuit simulé généralisé (*generalized simulated annealing*), présentant l'avantage, outre le fait de répondre aux contraintes définies ci-dessus, de converger plus efficacement et plus rapidement que d'autres méthodes similaires [Xiang *et al.*, 2013]. Les détails de la méthode peuvent être consultés dans des articles dédiés [Kirkpatrick *et al.*, 1983; Tsallis et Stariolo, 1996].

Finalement, l'implémentation de cet algorithme d'optimisation nous a donc permis d'estimer simultanément le nombre de voisins dans le paramétrage de la GWR et le nombre de contextes spatiaux qui minimise l'AIC du modèle multiniveau final.

3.1.2.2 Régionalisation des coefficients GWR

L'autre étape essentielle consiste à établir un protocole de régionalisation des coefficients GWR. La classification désigne le regroupement des observations uniquement sur la base des valeurs des coefficients qui y sont associés, tandis que la régionalisation consiste à appliquer une contrainte spatiale à la classification

pour s'assurer que les regroupements finaux constitueront bien des contextes spatiaux contigus, spatialement exclusifs. Par essence, chaque surface de coefficient GWR présente une tendance spatiale relativement lissée, puisque la GWR est une méthode basée sur des noyaux. De fait, la classification d'une surface équivaut quasiment à une régionalisation. Néanmoins, si le modèle de régression intègre plusieurs variables explicatives – ce qui sera notre cas dans l'application développée ci-dessous – la classification sera multivariée. Dans ce cas, une contrainte spatiale doit être appliquée à la classification.

Il y a plusieurs façons de procéder à la régionalisation des coefficients GWR. Dans un article sur les déterminants de la marche [Feuillet *et al.*, 2016b], nous avons contourné la question de la contrainte spatiale en amont en appliquant une analyse factorielle aux variables indépendantes et ne régressant notre variable dépendante qu'avec la variable latente (première composante principale) qui en a résulté. On s'est ainsi trouvé dans le cas d'une GWR univariée, et une classification simple de la surface des coefficients a suffi pour régionaliser le territoire. Dans un article suivant [Feuillet *et al.*, 2018], nous avons appliqué l'algorithme des k-médoïdes au vecteur des coefficients, mais sans chercher ensuite à régionaliser les clusters. Dans une étude visant à délimiter des sous-marchés immobiliers en Autriche, Helbich *et al.* [2013] sont allés plus loin en soumettant à des variables latentes synthétisant leurs coefficients GWR un algorithme de classification spatiale. La méthode utilisée (SKATER, pour *Spatial Klustering Analysis by Tree Edge Removal*, Assunção *et al.* [2006]) consiste à construire un graphe à partir de la matrice de contiguïté des observations, à pondérer les liens du graphe en fonction d'une matrice de dissimilarités issue des données en jeu (ici les coefficients GWR ou leur variables synthétiques), à construire l'arbre portant minimal, avant de supprimer ses branches dans un objectif de maximisation de la variance inter-classes [Colin et Roussez, 2018]. Cette méthode serait adaptée à notre objectif, mais elle présente l'inconvénient d'être très gourmande en ressources computationnelles quand le nombre d'unités est grand ($n \sim 9000$ dans notre cas), ce qui complique assez sérieusement le processus d'optimisation des paramètres. Afin de contourner ce problème, une autre méthode de régionalisation à peu près équivalente – mais plus efficace – a été préférée. Cette approche, proposée par Chavent *et al.* [2018], est simple : elle consiste à intégrer dans une classification ascendante hiérarchique deux matrices. La première est une matrice de dissimilarités issues de la matrice de données (les coefficients GWR) et la seconde est une matrice de voisinage des observations (contiguïté ou distance, au choix), qui représente la contrainte spatiale. Un coefficient α , variant de 0 à 1, permet de pondérer la contrainte spatiale dans la classification. Ainsi, si $\alpha = 0$, alors la classification n'est faite qu'à partir de la première matrice (CAH classique). Si $\alpha = 1$, alors la classification n'est faite que sur la base des distances géographiques entre observations. Il s'agit donc de trouver un compromis dans la valeur de α qui permette d'appliquer à notre classification une contrainte spatiale suffisante pour rendre les contextes contigus, mais pas trop forte non plus afin que la classification tienne bien compte des coeffi-

cients GWR. Cette valeur de contrainte spatiale a été déterminée de façon manuelle.

Le dernier choix méthodologique à réaliser est relatif à un problème de géotraitement, lié au découpage de l'espace en aires (contextes spatiaux) à partir de données ponctuelles. En effet, dans l'application développée plus bas, les observations sont faites au niveau des adresses de résidence d'individus, représentés sous forme ponctuelle. Les estimations GWR et les clusters sont donc également réalisés à ce niveau ponctuel. A nouveau, il existe plusieurs options pour assurer cette transition des points aux polygones. La première consiste à procéder à une interpolation en amont de la classification, directement sur les coefficients GWR, pour produire des surfaces rasterisées. La classification se fait sur les différents rasters, et le pixel devient alors l'unité à classifier. C'est par exemple cette approche qui a été privilégiée par [Helbich et al. \[2013\]](#). Une autre façon de faire est de réaliser la classification au niveau des entités ponctuelles, puis de polygoniser ces points par une tessellation¹ de Dirichlet. On attribue alors à chaque polygone de Voronoï qui en résulte l'appartenance à un cluster, puis on unit les polygones sur la base de ces appartenances. C'est cette approche qui a été choisie, car elle permettait d'éviter le recours à la rasterisation, qui aurait elle-même nécessité le choix important d'une résolution spatiale optimale.

3.1.3 Validation des modèles : minimisation de l'AIC et maximisation des effets contextuels

L'objectif de l'implémentation de ce modèle hybride est de tester l'hypothèse selon laquelle, pour modéliser des effets de contexte, un modèle multiniveau intégrant des contextes spatiaux dérivés de la non-stationnarité spatiale des données est plus pertinent qu'un modèle multiniveau basé sur des limites administratives. Pour tester cette hypothèse, il faut donc comparer les deux modèles multiniveaux (le modèle hybride multiniveau-GWR et le modèle classique basé sur les limites administratives) selon des critères de pertinence. Un critère statistique largement utilisé pour comparer des modèles non emboîtés est le critère d'information d'Akaike (AIC), qu'il faut chercher à minimiser. Cet indice utilise la log-vraisemblance des modèles (reflétant leur qualité d'ajustement aux données) en la pénalisant par le nombre de paramètres. C'est déjà ce critère qui a été utilisé pour optimiser les paramètres de la GWR et de la classification.

Outre ce premier critère de comparaison des modèles, d'autres indices peuvent être utilisés. Rappelons que les effets de contexte sont susceptibles d'agir soit directement sur la variable dépendante, soit sur les relations entre les variables indépendantes et dépendantes (figure 1.10). Si le modèle hybride incorpore effectivement des contextes spatiaux plus pertinents que ceux basés sur les limites administratives, alors on peut s'attendre à ce qu'il présente des variances de constantes et/ou de pentes plus élevées que dans le modèle classique. Pour cette raison, en plus de l'AIC, l'ICC (qui rapporte la variance contextuelle à la variance totale) et la variance

1. La tessellation est l'action de découper une surface en parties régulières.

Tâche	Package	Source
Modélisation GWR	GWmodel	Gollini <i>et al.</i> [2015]
Régionalisation	ClustGeo	Chavent <i>et al.</i> [2018]
Algorithmes d'optimisation	genSA	Xiang <i>et al.</i> [2013]
Modélisation multiniveau	lme4	Bates <i>et al.</i> [2015]
Sorties de modèles (tables)	sjPlot	Lüdecke <i>et al.</i> [2021]
Cartographie	tmap	Tennekes [2018]

Tableau 3.1 – Récapitulatif des librairies R utilisées pour la mise en œuvre des analyses statistiques.

des pentes seront également comparés dans les deux modèles. Le modèle présentant les valeurs les plus élevées pourra être considéré comme celui intégrant les contextes spatiaux les plus pertinents (c'est-à-dire ceux présentant le plus d'effets soit sur la constante, soit sur les pentes).

3.1.4 Outils mobilisés

Toutes les analyses statistiques ont été réalisées avec R [R_Core_Team, 2021]. Le tableau 3.1 résume les packages utilisés en fonction des tâches.

3.2 Application du modèle hybride multiniveau-GWR à la géographie de l'obésité en région parisienne

3.2.1 La géographie de l'obésité reflète des effets de contexte

L'obésité est largement étudiée dans la communauté scientifique, au sein de disciplines variées. La première raison en est qu'elle représente un véritable enjeu public, et de société. En 2016, près de 2 milliards d'adultes dans le monde étaient en situation de surpoids ($IMC > 25 \text{ kg/m}^2$), dont 650 millions obèses ($IMC > 30 \text{ kg/m}^2$) selon l'OMS. En France, 17 % des adultes sont obèses (7 millions de personnes). Surtout, la courbe ne cesse d'augmenter depuis les années 70 principalement en lien avec le développement du système alimentaire industrialisé. Il s'agit d'un enjeu de santé publique, car l'obésité est associée à une série de problèmes de santé, incluant les maladies cardio-vasculaires, le diabète, ainsi que 8 types de cancers. C'est par ailleurs un enjeu de société car on sait que, dans les pays occidentalisés, l'obésité suit un net gradient social. En France en 2012, 16,7 % des ouvriers étaient obèses, contre seulement 8,7 % des cadres supérieurs. Ces inégalités sociales de santé ont été redoutablement révélées lors la pandémie de la covid-19, l'obésité en étant un facteur de comorbidité avéré du virus. Les chercheurs s'intéressent donc à l'identification de déterminants de l'obésité qui pourraient être potentiellement modifiables, pour s'en servir comme leviers.

La prise de poids menant à l'obésité résulte d'une balance énergétique déséquilibrée, dans laquelle les dépenses énergétiques (activité physique) sont inférieures aux apports (alimentation). Ainsi, les deux principaux déterminants directs de l'obésité sont les pratiques d'activité physique et les comportements alimentaires. Il s'agit donc de s'intéresser aux déterminants de ces comportements obésogènes, car ce sont eux qui, via une relation médiée par les comportements, peuvent être modulés pour tenter de lutter contre l'obésité. C'est là que les effets de contexte entrent en jeu. Depuis les années 1990, les géographes de la santé et les épidémiologistes multiplient les preuves empiriques de relations entre l'environnement de vie (bâti et/ou social) et les comportements obésogènes. On sait que les lieux présentant de fortes densités d'infrastructures de transport, de destinations potentielles et d'aménités urbaines favorisent la pratique de la marche et du vélo [Ewing et Certero, 2010; Giles-Corti *et al.*, 2016; Saelens et Handy, 2008]. De nombreuses études ont également montré des liens entre le paysage alimentaire (accessibilité/densités de fast-food ou de restaurants par exemple) et la qualité de l'alimentation. En ce qui concerne l'environnement social, les relations entre la précarité du lieu de résidence et l'obésité sont aussi connus. Nous avons par exemple montré récemment, sur un échantillon de 70 000 adultes français, une relation positive entre le niveau de désavantage socio-économique de son quartier de résidence (IRIS) et la probabilité d'être en situation de surpoids [Feuillet *et al.*, 2020b], *cf.* figure 1.6. Plus largement, dans ce même échantillon, le pourcentage de personnes en surpoids varie fortement selon les contextes urbains, confirmant l'existence très probable d'effets contextuels (figure 3.2).

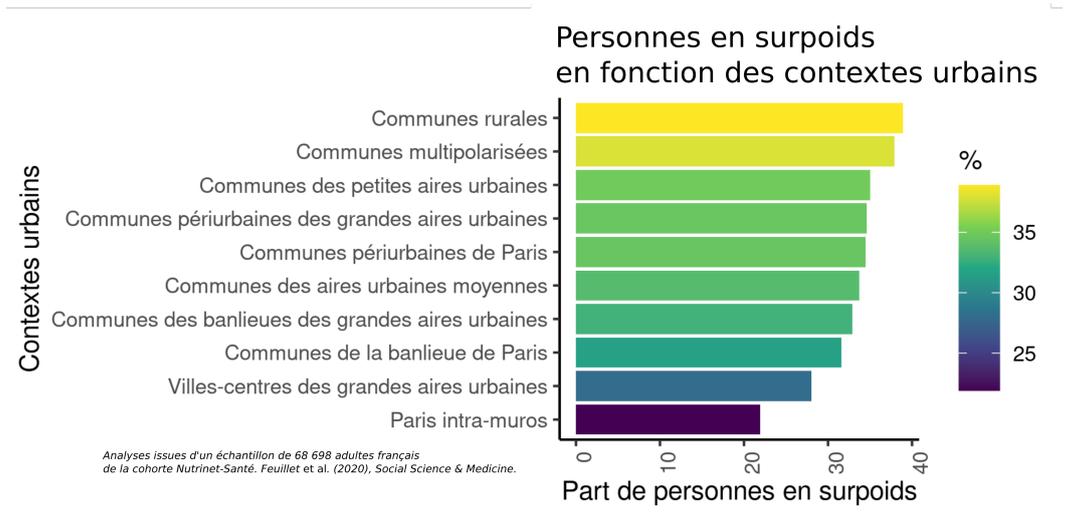


FIGURE 3.2 – Dans la cohorte NutriNet-Santé, le pourcentage de personnes en surpoids varie selon les contextes urbains (de façon inversement proportionnelle au gradient d'urbanité), laissant deviner l'existence d'effets contextuels.

Ainsi, une partie des déterminants des comportements obésogènes est spatialisée, agit au niveau local, et résulte en une géographie spécifique de l'obésité qu'il convient de déchiffrer. La complexité de cette géographie est due aux interactions locales entre les caractéristiques environnementales, traduisant le fait que l'effet d'une caractéristique dépend en réalité du contexte local dans lequel elle se trouve (si on aménage une piste cyclable le long de l'autoroute, il y a peu de chances qu'elle soit utilisée). Cette complexité est révélée dans la littérature par les nombreux résultats contradictoires selon les études, ou la mise en évidence de la non-stationnarité spatiale des déterminants (voir par exemple Oshan *et al.* [2020]). Dans notre étude sur l'obésité à l'échelle nationale [Feuillet *et al.*, 2020b], nous avons par exemple montré des variations de la relation précarité-obésité en fonction des contextes urbains (centres-villes des villes centres, banlieues d'aires urbaines moyennes, couronnes périurbaines de grandes aires urbaines, etc.). Ces regroupements locaux de caractéristiques environnementales spécifiques (qu'elle soient bâties ou sociales) affectent donc les relations environnement-obésité. Ce mécanisme n'est rien d'autre qu'un exemple d'effet de contexte tel que défini dans la première partie (effet de normes locales partagées par exemple, figure 1.5). Le fait d'appartenir, pour un individu et ses voisins, à tel ou tel contexte spatial, caractérisé par une nature et un agencement spécifiques de ses composantes, est donc susceptible d'affecter leurs comportements et leurs poids. Ces effets de contexte engendrent inévitablement une dépendance des données au sein des contextes spatiaux. En conséquence, la modélisation multiniveau apparaît comme une méthode particulièrement bien adaptée à l'analyse statistique de l'obésité et de ses déterminants.

3.2.2 Présentation du jeu de données et de la cohorte NutriNet-Santé

Pour mettre en œuvre le modèle hybride présenté dans la partie précédente, et tester l'hypothèse selon laquelle des contextes spatiaux dérivés de la non-stationnarité spatiale des données plutôt que des limites administratives est plus pertinente pour modéliser la géographie de l'obésité, nous avons sélectionné un sous-échantillon de la cohorte prospective NutriNet-Santé, dans la région de Paris et de sa petite couronne. Cette cohorte d'adultes volontaires (les *nutrinautes*) a été lancée en mai 2009 et incluait en 2020 plus de 170 000 personnes. Les *nutrinautes* répondent régulièrement à différents questionnaires de santé, portant sur l'alimentation, l'activité physique, les états de santé, et ont par ailleurs été localisés à leur adresse de résidence, ce qui nous permet de relier ces comportements et ces états de santé à l'environnement de résidence.

La web-cohorte prospective NutriNet-Santé

La web-cohorte NutriNet-Santé a été lancée en mai 2009 [Hercberg *et al.*, 2010]. Elle représente une des principales cohortes prospectives en France dans le domaine de la nutrition et impliquait, en 2020, plus de 171 000 adultes volontaires (il est d'ailleurs toujours possible de participer : <https://etude-nutrinet-sante.fr/>). NutriNet-Santé est gérée au quotidien par l'Équipe de recherche en épidémiologie nutritionnelle (EREN, équipe mixte basée à Bobigny et adossée à l'INSERM, l'INRAE, le CNAM et l'université Paris 13), dirigée par Serge Hercberg jusqu'en 2019, puis par Mathilde Touvier depuis cette date. La gestion et la préparation des bases de données issues de cette cohorte mobilisent des dizaines de techniciens, ingénieurs, biostatisticiens et chercheurs, dont l'implication quotidienne est nécessaire pour rendre possible l'application d'analyses telles que celles proposées dans ce volume.

3.2.2.1 Variables individuelles

Après nettoyage de la base (données manquantes ou aberrantes), l'échantillon dans la région de Paris et sa petite couronne comprend 9 089 individus géolocalisés (leur densité est illustrée dans la figure 3.3). La variable dépendante de tous les modèles estimés par la suite est l'indice de masse corporelle (IMC), calculé en divisant le poids des individus par leur taille (en m) au carré, selon la recommandation de l'OMS. Deux caractéristiques individuelles seront utilisées en variables exogènes : le sexe et l'âge (au moment de l'inclusion dans la cohorte), deux dimensions que l'on sait être associées à l'IMC. Sur cet échantillon de nutrinateurs, l'IMC moyen est de 23,4 kg/m². Il est composé de 76 % de femmes et l'âge moyen est de 42 ans.

3.2.2.2 Variable contextuelle

La densité de bâti a été sélectionnée pour représenter l'environnement bâti au sein des contextes. On sait que la densité de bâti est un bon indicateur de beaucoup d'autres composantes de l'environnement bâti (densités de routes, de commerces, de population etc.), et est régulièrement associée aux comportements liés à l'obésité, de façon négative (les niveaux d'obésité sont en moyenne plus faibles dans les zones denses). Elle a été estimée dans chaque contexte (administratifs et ceux dérivés de la GWR) comme le rapport entre la surface recouverte de bâtiments (provenant de la BD TOPO de l'IGN) et la surface totale du contexte spatial.

3.2.2.3 Contextes selon les limites administratives

Le niveau retenu pour délimiter les contextes administratifs est celui des intercommunalités (EPCI). L'EPCI de Paris a été divisé par arrondissement, pour

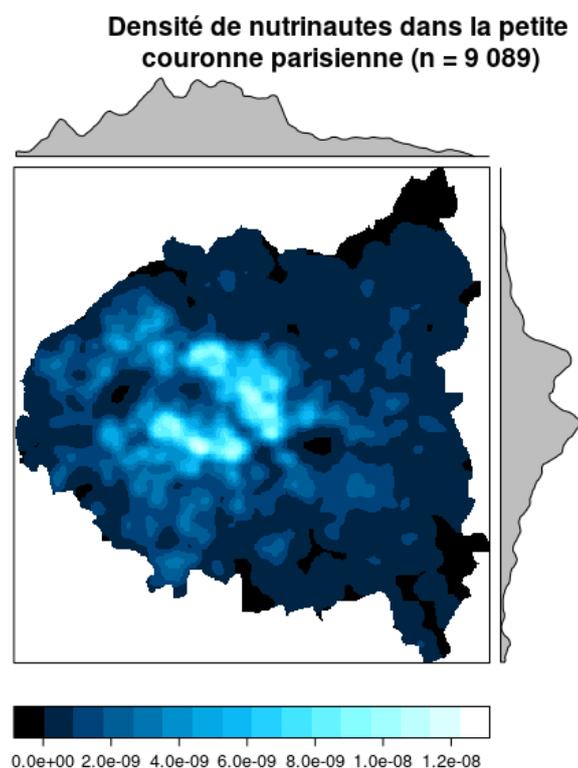


FIGURE 3.3 – Densité de noyau de l'échantillon de nutrinautes utilisé dans les analyses statistiques.

gagner en finesse dans cette zone très densément peuplée. Cela résulte en un total de 31 unités administratives (illustration en figure 3.5). Les EPCI ont été préférés aux communes car, étant trop nombreuses (143), un trop grand nombre d'entre elles ne regroupaient que trop peu de participants (la maille communale a néanmoins été testée dans les modèles pour servir de comparaison). La figure 3.4 illustre la répartition des nutrinautes selon les différentes mailles mentionnées.

3.2.3 Présentation des résultats

3.2.3.1 Régionalisation des coefficients GWR

L'algorithme d'optimisation des paramètres bw (nombre de plus proches voisins considéré dans le noyau de la GWR) et k (nombre de contextes issus de la régionalisation des coefficients GWR) basé sur la minimisation de l'AIC du modèle multiniveau complet a abouti à 66 voisins et 60 contextes. Précisons que la régionalisation est basée sur trois vecteurs de coefficients : les deux associés aux variables exogènes (sexe et âge), mais également le vecteur des constantes (figure 3.6). Dans la mesure où celles-là représentent l'IMC moyen quand les variables exogènes sont

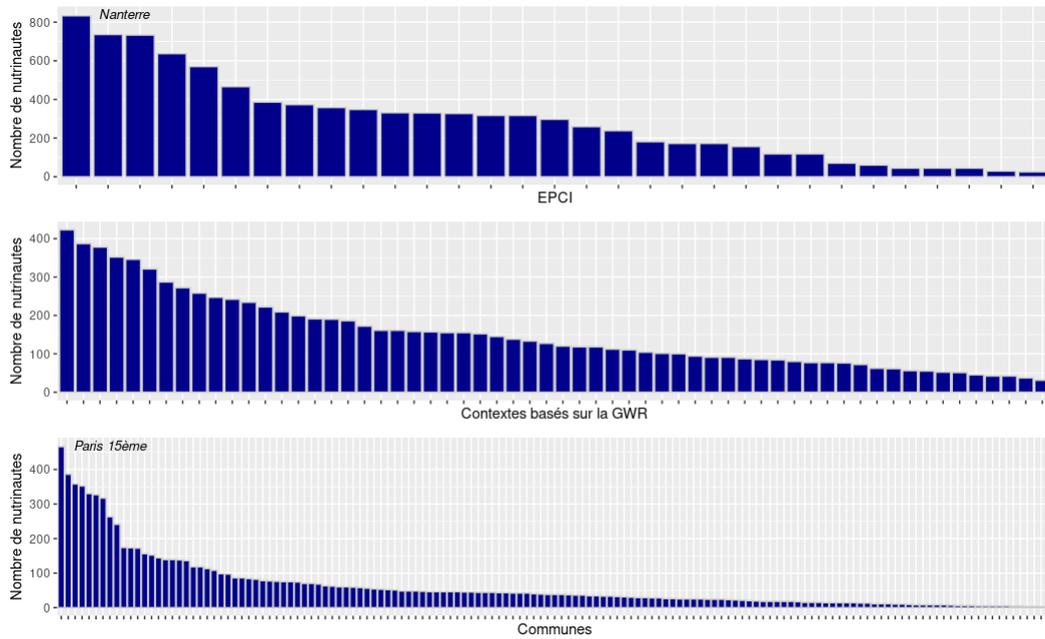


FIGURE 3.4 – Distribution des nutrinautes en fonction des mailles mentionnées ou utilisées dans les modèles.

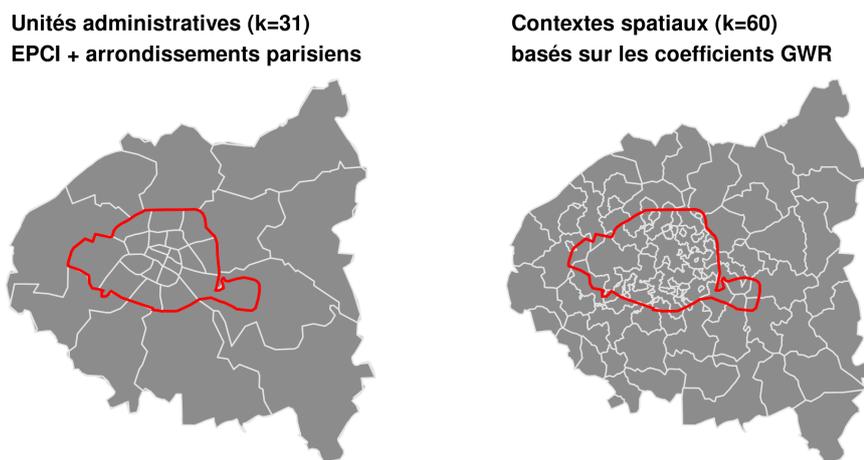


FIGURE 3.5 – Cartographie des contextes spatiaux selon les limites administratives (EPCI, carte de gauche) et selon la non-stationnarité spatiale des données, basées sur la régionalisation des coefficients de la GWR (carte de droite). Les limites de la ville de Paris sont représentées en rouge.

fixes (c'est-à-dire être un homme d'un âge moyen, l'âge ayant été au préalable standardisé), les constantes nous informent sur la répartition spatiale de l'IMC ajusté, lui-même susceptible d'être affecté directement par des effets de contexte. Il apparaît donc cohérent de les impliquer dans la construction empirique des contextes spatiaux (c'est-à-dire dans la régionalisation des coefficients GWR). La cartographie finale des contextes est illustrée dans la figure 3.5.

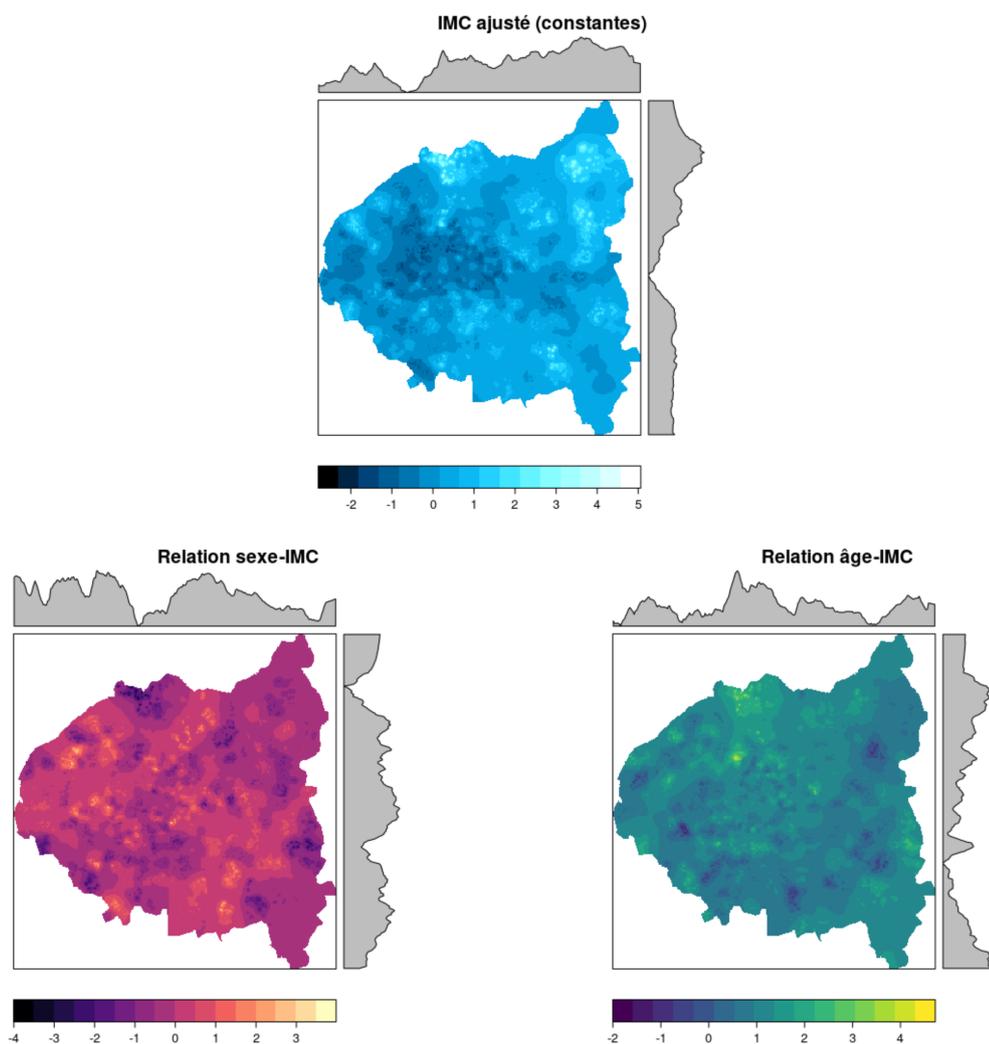


FIGURE 3.6 – Cartes de surfaces des coefficients GWR, après interpolation (méthode IDW). Ces trois séries de valeurs ont été classifiées et régionalisées pour aboutir aux 60 contextes spatiaux utilisés dans les modèles multiniveaux subséquents.

<i>Predictors</i>	Multiniveau contextes EPCI			Multiniveau contextes GWR		
	<i>Estimates</i>	<i>CI</i>	<i>p</i>	<i>Estimates</i>	<i>CI</i>	<i>p</i>
Constante	23.32	23.04 - 23.61	<0.001	23.77	23.51 - 24.02	<0.001
Random Effects						
σ^2	19.678			19.325		
τ_{00}	0.528 _{epci}			0.832 _{clustGwr1}		
ICC	0.026			0.041		
N	31 _{epci}			60 _{clustGwr1}		
Observations	9089			9089		
Deviance	52933.113			52821.266		
AIC	52941.130			52829.513		
log-Likelihood	-26467.565			-26411.757		

Tableau 3.2 – Résultat des estimations pour les modèles multiniveaux vides. A gauche, le modèle multiniveau EPCI (contextes spatiaux basés sur les limites administratives) et à droite le modèle multiniveau GWR (contextes spatiaux basés sur la régionalisation préalable des coefficients d'un modèle GWR). L'AIC du modèle utilisant les contextes spatiaux dérivés de la GWR est plus faible, indiquant sa plus grande pertinence.

3.2.3.2 Modèle vide à constantes aléatoires

Il est d'usage de débiter la modélisation multiniveau par un modèle vide à constantes aléatoires, c'est-à-dire un modèle sans aucune variable explicative, équivalant à une ANOVA à effets aléatoires. Les résultats de ce modèle sont donnés dans le tableau 3.2. La première chose que l'on constate est que l'AIC est légèrement plus faible dans le modèle à contextes GWR que dans le modèle à contextes EPCI (52 821 contre 52 933), indiquant donc un meilleur ajustement aux données pour le premier. Dans les composantes aléatoires des modèles, σ^2 désigne la variance individuelle, et τ_{00} la variance des constantes (entre les contextes). L'ICC est donc égal à τ_{00}/σ^2 . L'ICC est égal à 0,026 dans le modèle à contextes EPCI, signifiant que 2,6 % de la variation totale de l'IMC peut être attribuée à la variation entre les EPCI. Ce pourcentage monte à 4,1 % pour ce qui est de la variation entre les contextes basés sur la GWR. C'est certes supérieur, mais dans les deux cas, ces pourcentages restent relativement faibles, indiquant que les différences entre individus, quel que soit leur contexte de résidence, sont l'élément essentiel pour expliquer la variation de l'IMC. Cette variance inter-contextuelle existe néanmoins et reflète donc un léger effet contextuel, en particulier quand ces contextes sont basés sur la non-stationnarité spatiale.

Le deuxième modèle consiste à ajouter aux modèles les variables individuelles (sexe et âge), dont une (l'âge) en effet aléatoire (c'est-à-dire avec des coefficients variables selon les contextes).

On constate tout d'abord (tableau 3.3) qu'à nouveau, le modèle à contextes GWR minimise l'AIC (52 156 contre 52 326 pour le modèle EPCI). Ces deux AIC sont logiquement plus faibles que dans les modèles vides, puisque des variables explicatives ont été ajoutées. En ce qui concerne les effets fixes (effets moyens), on conclut que le fait d'être une femme et d'être plus âgé est positivement associé à l'IMC (pour les deux modèles), relations par ailleurs déjà bien connues. L'ICC a également évolué, passant de 4,1 % à 5,9 % dans le modèles à contextes GWR, alors que celui du modèle EPCI n'a quasiment pas évolué. Mais le résultat le plus intéressant est révélé par la variance des pentes, notée τ_{11} , qui présente une valeur considérablement plus forte dans le second modèle (0,212 contre 0,032 pour le modèle EPCI). Cela signifie, en toute logique étant donné la manière dont les contextes ont été construits, que les effets contextuels agissent davantage sur la relation entre l'âge et l'IMC que sur l'IMC lui-même.

<i>Predictors</i>	Multiniveau contextes EPCI			Multiniveau contextes GWR		
	<i>Estimates</i>	<i>CI</i>	<i>p</i>	<i>Estimates</i>	<i>CI</i>	<i>p</i>
Constante	23.04	22.76 - 23.32	<0.001	23.48	23.21 - 23.75	<0.001
Sexe (ref=homme)	1.23	1.02 - 1.44	<0.001	1.25	1.04 - 1.46	<0.001
Age	0.99	0.88 - 1.11	<0.001	1.05	0.90 - 1.21	<0.001
Random Effects						
σ^2	18.344			17.804		
τ_{00}	0.504 <small>epci</small>			0.912 <small>clustGwr1</small>		
τ_{11}	0.032 <small>epci.scale(age_incl)</small>			0.212 <small>clustGwr1.scale(age_incl)</small>		
ρ_{01}	0.387 <small>epci</small>			0.395 <small>clustGwr1</small>		
ICC	0.028			0.059		
N	31 <small>epci</small>			60 <small>clustGwr1</small>		
Observations	9089			9089		
Deviance	52303.804			52134.225		
AIC	52326.392			52156.402		
log-Likelihood	-26156.196			-26071.201		

Tableau 3.3 – Résultat des estimations pour les modèles multiniveaux à pentes aléatoires. À gauche, le modèle multiniveau EPCI (contextes spatiaux basés sur les limites administratives) et à droite le modèle multiniveau GWR (contextes spatiaux basés sur la régionalisation préalable des coefficients d'un modèle GWR). L'AIC du modèle utilisant les contextes spatiaux dérivés de la GWR est plus faible, indiquant sa plus grande pertinence.

La figure 3.7 illustre de façon graphique l'hétérogénéité des relations entre l'âge (standardisé) et l'IMC en fonction des contextes spatiaux, sur une même échelle afin de faciliter les comparaisons. On constate que dans le modèle à contextes EPCI, les droites sont davantage regroupées dans une fourchette d'IMC réduite (révélant une variance des constantes plus faible) et subparallèles (indiquant une variance des pentes également plus faible) que dans le modèle à contextes GWR. Il faut noter

que même dans ce dernier, on ne remarque pas non plus de très forte hétérogénéité entre contexte – aucune inversion de signes notamment qui aurait pu illustrer un paradoxe de Simpson – même si la variation est bien réelle. Cela est dû à l’effet d’écrasement (*shrinkage effect*) propre à la modélisation multiniveau, qui a tendance à pénaliser les coefficients vers la moyenne générale de tous les coefficients quand les observations ne sont pas assez nombreuses dans les contextes.

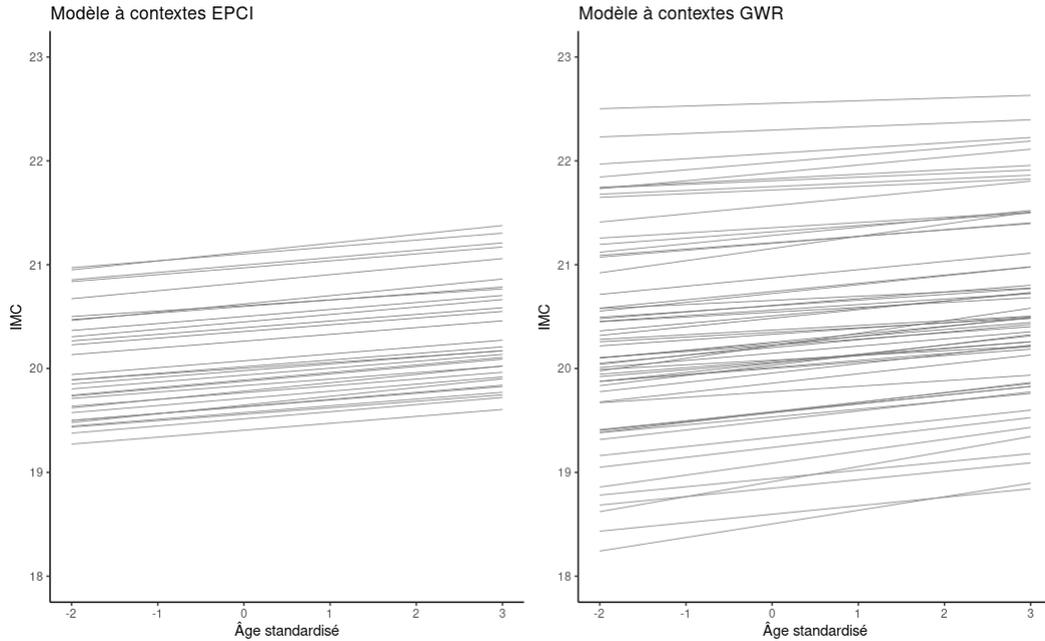


FIGURE 3.7 – Représentation graphique de l’hétérogénéité des relations entre l’âge et l’IMC (ajustées sur le sexe) en fonction des contextes spatiaux. A gauche, le modèle à contextes EPCI (contextes spatiaux basés sur les limites administratives) et à droite le modèle à contextes GWR (contextes spatiaux basés sur la régionalisation préalable des coefficients d’un modèle GWR).

3.2.3.3 Modèle complet (constantes et pente aléatoires avec variable contextuelle)

Dans ce dernier modèle, la densité de bâti, agrégée au niveau des contextes, a été ajoutée comme variable explicative (figure 3.8). Sans surprise, cette variable exhibe un gradient centre-périphérie. Le modèle hybride reste toujours le plus pertinent (AIC plus faible). Surtout, on constate que le modèle EPCI présente une covariance constantes-pentes (notée ρ_{01}) strictement égale à 1. Cela traduit un problème de convergence du modèle, qui s’explique en général par un nombre insuffisant d’observations ou de contextes. Plus que le nombre de contextes ici (31 est une valeur tout à fait acceptable et théoriquement suffisante pour des modèles multiniveaux – nombre de bases de données en géographie de la santé n’en comportent pas autant), le problème peut être provoqué par le manque de

variabilité de la variable de densité de bâti. Cela reflète un problème intrinsèque aux entités administratives telles que les EPCI, qui embrassent des lieux hétérogènes. Comme les caractéristiques environnementales mesurées au sein de ces entités sont moyennées, cela tend à homogénéiser des contextes qui pourtant, à une échelle intra-contextuelle, présentent de réels contrastes.

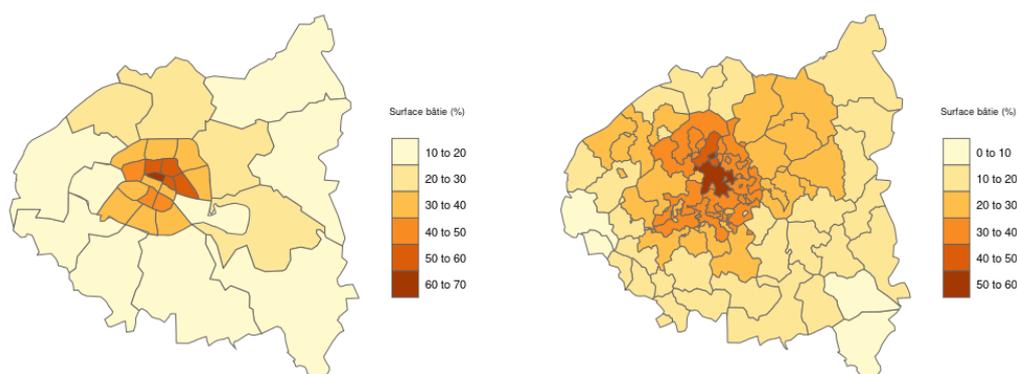


FIGURE 3.8 – Cartographie de la densité de bâti (BD TOPO) en fonction des contextes (à gauche les EPCI, à droite les contextes basés sur la GWR).

3.2.4 Robustesse du modèle aux paramétrages alternatifs

3.2.4.1 Sensibilité au nombre de contextes

On pourrait poser l'hypothèse que la meilleure performance du modèle hybride par rapport au modèle classique, en termes de minimisation de l'AIC, n'est due qu'au nombre de contextes. En d'autres termes, la performance de la méthode ne relèverait pas de la pertinence géographiques des contextes spatiaux, mais simplement de leur nombre. Cette hypothèse est possible, étant entendu qu'en général en modélisation multiniveau, un nombre plus important de macro-unités tend à augmenter la variance des composantes aléatoires du modèle et par là même certains indicateurs de qualité d'ajustement. Une analyse de sensibilité a donc été réalisée, consistant à faire varier le nombre de contextes spatiaux dans le modèle hybride de 20 à 60, puis à comparer l'AIC pour chaque k avec celui du modèle EPCI (comprenant donc 31 macro-unités). Le résultat (figure 3.9) est sans appel : même pour un nombre de contextes inférieur à 31, le modèle hybride reste plus performant. Le modèle hybride est donc robuste au nombre de macro-unités qu'il intègre. On constate par ailleurs que la performance se stabilise à partir de 45 contextes.

<i>Predictors</i>	Multiniveau contextes EPCI			Multiniveau contextes GWR		
	<i>Estimates</i>	<i>CI</i>	<i>p</i>	<i>Estimates</i>	<i>CI</i>	<i>p</i>
Constante	23.23	22.99 - 23.48	<0.001	23.29	23.03 - 23.55	<0.001
Sexe (ref=homme)	1.23	1.02 - 1.44	<0.001	1.25	1.05 - 1.46	<0.001
Age	1.01	0.89 - 1.12	<0.001	1.04	0.89 - 1.19	<0.001
Dens. bati (epci)	-0.48	-0.65 - -0.31	<0.001			
Dens. bati (gwr)				-0.48	-0.71 - -0.25	<0.001
Random Effects						
σ^2	18.338			17.813		
τ_{00}	0.332	epci		0.741	clustGwr1	
τ_{11}	0.039	epci.scale(age_incl)		0.210	clustGwr1.scale(age_incl)	
ρ_{01}	1.000	epci		0.598	clustGwr1	
ICC				0.051		
N	31	epci		60	clustGwr1	
Observations	9089			9089		
Deviance	52282.129			52120.325		
AIC	52310.518			52147.242		
log-Likelihood	-26147.259			-26065.621		

Tableau 3.4 – Résultats des estimations des modèles multiniveaux à constantes et pente aléatoires avec variable contextuelle (de niveau 2). A gauche, le modèle modèle multiniveau EPCI (contextes spatiaux basés sur les limites administratives) et à droite le modèle multiniveau GWR (contextes spatiaux basés sur la régionalisation préalable des coefficients d'un modèle GWR). L'AIC du modèle utilisant les contextes spatiaux dérivés de la GWR est plus faible, indiquant sa plus grande pertinence.

3.2.4.2 Sensibilité à la nature de l'échelon administratif

Si le modèle hybride est robuste au nombre de contextes, il est légitime de s'interroger également sur la sensibilité du modèle « administratif » à l'échelle du maillage. Les EPCI avaient été choisis plutôt que les communes en raison de leur effectif plus petit, ce qui assurait un nombre important d'observations par maille. En prenant en considération les communes au lieu des EPCI, le nombre d'unités représentées par au moins une observations monte à 143 (figure 3.4). Néanmoins, les mêmes modèles ont été estimés avec les communes comme macro-unités, pour tester l'hypothèse selon laquelle le manque de pertinence des modèles EPCI serait dû au nombre trop restreint d'unités (31). On peut en effet penser que l'échelon communal est pertinent en regard des états de santé liés à des comportements individuels – comme l'obésité – dans la mesure où un certain nombre de politiques d'aménagements susceptibles d'affecter ces comportements sont décidées à ce niveau local. En outre, l'échelle communale autorise une plus grande homogénéité intra-contextuelle des caractéristiques environnementales et sociales que l'échelle des EPCI, ce qui pourrait se traduire par une augmentation des variances des composantes inter-contextuelles du modèle (ICC et variance des pentes, indicateurs

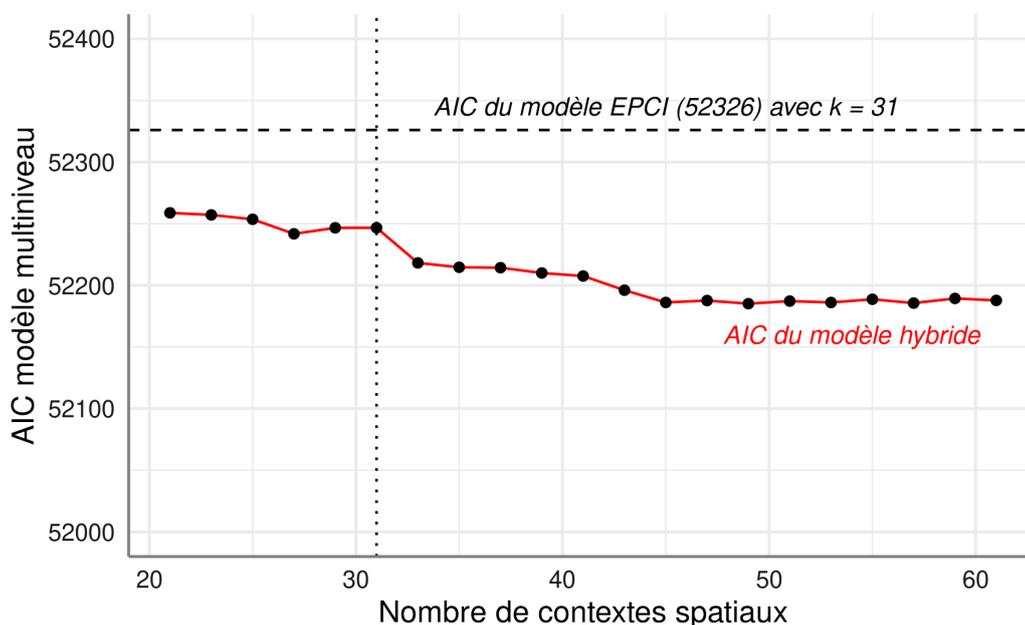


FIGURE 3.9 – Analyse de sensibilité au nombre de contextes spatiaux dans le modèle hybride. Quel que soit ce nombre, l’AIC du modèle hybride reste inférieur à celui du modèle EPCI, reflétant sa robustesse.

des effets contextuels).

Malgré ces éléments, les résultats révèlent finalement des AIC plus élevés dans les modèles intégrant les communes, non seulement par rapport aux modèles hybrides, mais également par rapport aux modèles EPCI (tableau 3.5). Ces résultats confirment donc à nouveau la pertinence du modèle hybride.

3.2.4.3 Modèles aléatoires croisés

Si les limites administratives sont si souvent utilisées dans la modélisation multiniveau en sciences sociales, c’est en grande partie parce que les unités administratives correspondent aux niveaux d’agrégation des données du recensement de la population. Or, ces données sont régulièrement utilisées comme variables explicatives dans les modèles statistiques. Si les contextes spatiaux sont construits *ad hoc* et que l’on souhaite néanmoins intégrer des variables socio-démographiques au niveau populationnel, il faut alors recourir à un processus de désagrégation-réagrégation des données, qu’il convient en général d’éviter. Il existe maintenant certaines variables disponibles en France au niveau des carreaux INSEE – ce qui permettrait d’éviter la désagrégation – mais cela reste limité. L’autre possibilité est d’intégrer des niveaux hiérarchiques croisés dans les modèles multiniveaux, ce qui est tout à possible. Dans ces modèles croisés, on intègre en niveau 2 les contextes

Predictors	Modele vide			Modele a pente aleatoire			Modele complet		
	Estimates	CI	p	Estimates	CI	p	Estimates	CI	p
Constante	23.75	23.58 - 23.92	<0.001	23.43	23.25 - 23.61	<0.001	23.20	23.05 - 23.35	<0.001
Sexe (ref=homme)				1.23	1.02 - 1.44	<0.001	1.23	1.02 - 1.44	<0.001
Age				0.98	0.87 - 1.08	<0.001	0.99	0.89 - 1.09	<0.001
Dens. bati							-0.52	-0.64 - -0.40	<0.001
Random Effects									
σ^2	19.633			18.278			18.271		
τ_{00}	0.526	communes		0.528	communes		0.257	communes	
τ_{11}				0.031	communes.scale(age_incl)		0.027	communes.scale(age_incl)	
ρ_{01}				-0.032	communes		0.993	communes	
ICC	0.026			0.030			0.015		
N	143	communes		143	communes		143	communes	
Observations	9089			9089			9089		
AIC	52975.112			52356.306			52306.817		
log-Likelihood	-26484.556			-26171.153			-26145.408		

Tableau 3.5 – Résultats des modèles intégrant les communes (k=143) en macro-unités. L'AIC est plus élevé que pour tous les autres modèles, indiquant que cette solution est moins pertinente.

spatiaux *ad hoc* et les unités administratives simultanément. Ils sont dits croisés car chaque observation fait nécessairement partie de deux macro-unités. Néanmoins il n'y a pas de hiérarchie stricte entre ces deux macro-unités. C'est par exemple ce type de modèle qui a été utilisé par [Arcaya et al. \[2012\]](#), cf. § 3.1.1).

En appliquant cette solution à nos propres données, les résultats sont mitigés (tableau 3.6). Quelles que soient les variables intégrées dans le modèle, les AIC sont plus faibles que dans les modèles EPCI, mais légèrement plus élevés néanmoins que dans les modèles n'intégrant que les contextes GWR (par exemple pour le modèle complet : 52153 contre 52147 respectivement). Il faut noter que certaines composantes aléatoires du modèle sont strictement égales à 0, indiquant potentiellement des problèmes de convergence qui pourraient être dus au trop faible nombre d'EPCI. Cette solution reste donc une option à envisager, mais semble conditionnelle aux effectifs des macro-unités.

3.3 Implications des résultats en géographie et en santé publique

Le principal résultat de ces analyses a été de montrer une plus grande pertinence (AIC plus faible et variances des pentes plus élevées) des modèles multiniveaux intégrant des contextes spatiaux *ad hoc* basés sur la non-stationnarité spatiale des données par rapport aux modèles intégrant des limites administratives, que ce soient les communes ou les EPCI. Ils surpassent également les modèles croisés intégrant deux mailles au niveau 2, l'une administrative et l'autre les contextes GWR. La figure

Predictors	Modele vide			Modele a pente aleatoire			Modele complet		
	Estimates	CI	p	Estimates	CI	p	Estimates	CI	p
Constante	23.77	23.51 - 24.02	<0.001	23.48	23.21 - 23.75	<0.001	23.29	23.03 - 23.55	<0.001
Sexe (ref=homme)				1.25	1.04 - 1.46	<0.001	1.25	1.05 - 1.46	<0.001
Age				1.05	0.90 - 1.21	<0.001	1.04	0.89 - 1.19	<0.001
Dens. bati							-0.48	-0.71 - -0.25	<0.001
Random Effects									
σ^2	19.325			17.804			17.812		
τ_{00}	0.832	clustGwr1		0.912	clustGwr1		0.741	clustGwr1	
	0.000	epci		0.000	epci		0.000	epci	
τ_{11}				0.212	clustGwr1.scale(age_incl)		0.210	clustGwr1.scale(age_incl)	
				0.000	epci.scale(age_incl)		0.000	epci.scale(age_incl)	
ρ_{01}				0.395	clustGwr1		0.598	clustGwr1	
N	31	epci		31	epci		31	epci	
	60	clustGwr1		60	clustGwr1		60	clustGwr1	
Observations	9089			9089			9089		
AIC	52831.513			52162.402			52153.242		
log-Likelihood	-26411.757			-26071.201			-26065.621		

Tableau 3.6 – Résultats des modèles aléatoire croisés combinant les contextes administratifs et les contextes GWR en niveau 2.

3.10 récapitule tous les résultats. Ces résultats ont des implications de fond, autant sur les analyses contextuelles en géographie que dans le domaine plus appliqué de la santé publique. Il convient néanmoins de garder en tête, à la lecture de ces implications, en particulier en santé publique, que l'échantillon qui a été utilisé dans le cadre de ces analyses n'est pas représentatif de la population générale (surreprésentation de femmes et de personnes diplômées) et que l'approche est strictement observationnelle et transversale, limitant les interprétations causales.

3.3.1 Implications en analyse spatiale : vers un affranchissement des maillages administratifs ?

Nous avons affiché un des problèmes majeurs de la modélisation multiniveau dès l'introduction, problème relatif au fait que la méthode nécessite de regrouper les observations *a priori*, et que ce regroupement se fait, dans la plupart des cas, selon des maillages pré-établis par défaut (en géographie humaine, il s'agit principalement des unités administratives). Néanmoins, ces maillages ne reflètent pas toujours la géographie des processus réels, et peuvent aboutir à des interprétations biaisées quant à l'existence et à la nature d'effets contextuels. Cette contrainte des maillages pré-établis vaut d'ailleurs pour toute analyse spatiale quantitative, au-delà des modèles multiniveaux. Elle vaut également au-delà de la géographie humaine. En géographie physique, on utilise souvent des massifs ou des bassins-versants pour délimiter des secteurs qui ne sont pas nécessairement les plus pertinents en regard des phénomènes étudiés.

Dans ce travail, nous avons montré qu'une solution possible pour contourner

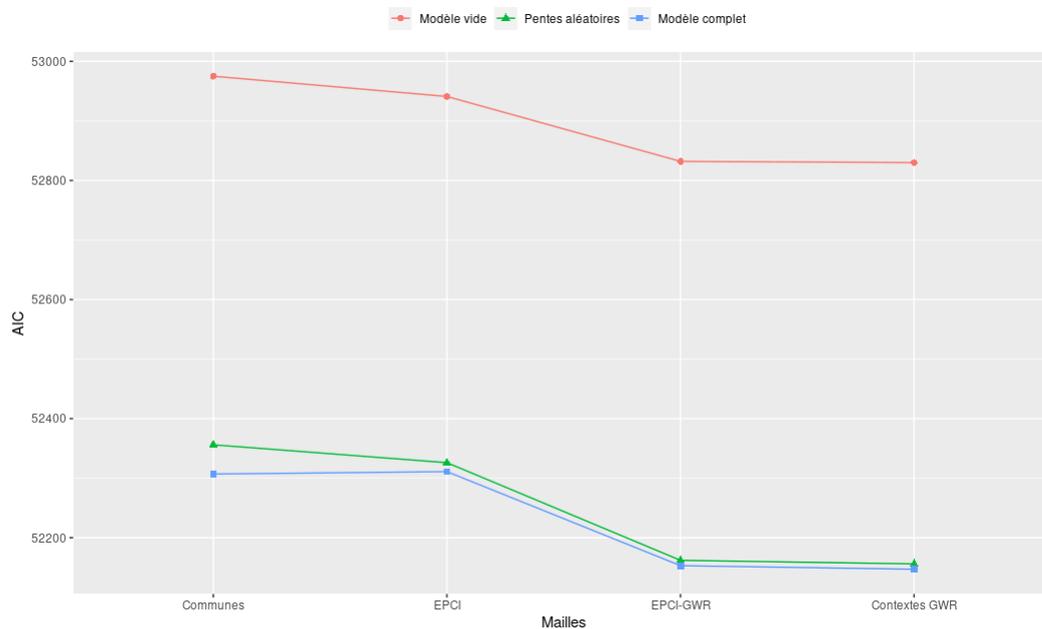


FIGURE 3.10 – Synthèse des performances de l’ensemble des modèles estimés. Le modèle multiniveau intégrant les contextes GWR est celui qui minimise l’AIC.

cette contrainte consiste à délimiter des contextes spatiaux sur la base de la non-stationnarité spatiale des données, elle-même révélée par la régionalisation des coefficients GWR. Certes, les contextes en résultant préservent les défauts inhérents à un découpage spatial en polygones exclusifs et contigus (contrairement à des polygones qui seraient parfaitement individualisés et chevauchants), mais permettent en revanche d’approcher une géographie des contextes plus pertinente et non tributaire de maillages existants déconnectés des données.

Ce protocole d’analyse respecte le contrat énoncé en introduction, consistant à laisser davantage parler les données (ce qui a été fait en délimitant des contextes spatiaux *ad hoc*) tout en inscrivant les analyses dans un cadre théorique solide, celui défini dans les deux premières parties de ce mémoire. En conséquence, les effets de contexte finalement modélisés gagnent en fiabilité, ce qui a bien entendu des implications sur l’interprétation des mécanismes en jeu, dans le domaine ici étudié de la santé publique.

3.3.2 Implications en santé publique : vers un ciblage géographique optimal des déterminants de santé

L’épidémiologie a principalement pour but d’identifier des déterminants de santé ou de comportements liés à la santé, dans l’optique d’orienter efficacement

des interventions ciblées pour contrôler et prévenir les inégalités de santé dans la société. Dans ce cadre, une partie des études s'intéresse aux déterminants individuels, afin de cibler des populations spécifiques. Une autre partie s'intéresse par ailleurs aux déterminants environnementaux, qui permet cette fois de cibler des territoires locaux. Comme le précise V. Chasles, "le territoire, dans toutes ses composantes, a une part de responsabilité dans la production et l'aggravation des inégalités de santé" [Chasles, 2020, p. 57]. Des études de plus en plus nombreuses recommandent en effet de décliner localement les interventions de santé publique pour mieux respecter l'hétérogénéité spatiale des territoires et des déterminants de santé, et ainsi gagner en efficacité [Rican et Vaillant, 2015]. Dans un article de 2015 portant sur les facteurs de la mobilité active en région parisienne, nous avons ainsi volontairement affiché le sous-titre évocateur "Towards a locally varying social ecological model" [Feuillet *et al.*, 2015], afin d'insister sur la pertinence d'une déclinaison locale de la hiérarchie des déterminants de santé.

Cette déclinaison locale est souvent prisonnière des découpages administratifs, pour les raisons déjà évoquées ci-dessus. Or, les niveaux administratifs ont fréquemment la particularité d'incorporer des lieux hétérogènes. C'est le cas pour les communes ou les EPCI par exemple, qui intègrent au sein d'une même unité des contextes urbains contrastés à bien des niveaux (centre-ville et banlieue par exemple). D'un côté, ces niveaux administratifs correspondent aux échelons décisionnels pour ce qui relève d'une partie de l'aménagement urbain et de la mise en place de politiques d'urbanisme favorables à la santé. En cela, l'utilisation du maillage administratif dans les analyses spatiales peut faciliter l'articulation entre la recherche académique et les mesures interventionnelles.

Néanmoins, deux autres configurations sont possibles :

- (i) Les aménagements urbains visant à favoriser les comportements favorables à la santé peuvent tout d'abord concerner des échelles plus fines que la maille administrative : un secteur précis d'une commune ou même d'un quartier, une rue, etc. C'est d'autant plus probable que le secteur en question est densément peuplé, comme à Paris *intra muros*. En cela, notre découpage en contextes *ad hoc* peut être utile pour cibler ces microzones. On voit en effet sur la figure 3.5 que les contextes dans Paris sont nettement plus complexes et nombreux que les arrondissements. Un exemple clair en est le 12^{ème} arrondissement qui comprend, pour des raisons historiques, le bois de Vincennes. Dans notre classification, le bois est divisé en différents contextes rattachés aux communes de proche banlieue. On constate en revanche que le 16^{ème} arrondissement, plus homogène dans sa composition sociale et bâtie, est logiquement représenté par un contexte qui se rapproche des limites administratives.
- (ii) Les aménagements urbains peuvent au contraire concerner des zones chevauchant plusieurs unités administratives, comme c'est le cas le long d'axes

structurants par exemple. Dans la figure 3.5, carte de droite, on constate la présence d'un contexte, juste au sud du bois de Vincennes, qui chevauche ainsi trois EPCI. Cette géographie pourrait être mise en relation avec la vallée de la Seine, de direction méridienne, qui par sa présence même tend à homogénéiser ces lieux, quelle que soient leurs communes/EPCI d'appartenance. Il est en effet cohérent de penser que des aménagements favorisant l'activité physique puissent être mis en place le long du cours d'eau, plutôt que de façon différenciée selon les EPCI.

L'autre résultat important en géographie de l'obésité, au-delà des questions de délimitation contextuelle, est que les effets de contexte que nous avons mis en évidence semblent agir davantage sur les relations entre l'IMC et ses déterminants, que sur l'IMC lui-même. Cela peut se comprendre dans la mesure où les états individuels de santé, tels que l'IMC, sont très variables d'un individu à l'autre, y compris au sein d'un même contexte spatial, si bien que la part de variance inter-contextuelle reste toujours relativement faible par rapport à la variance inter-individuelle, quelle que soit la géographie du contexte spatial. Il en serait différemment si l'on étudiait les prix immobiliers par exemple, qui présentent une dépendance spatiale très forte au sein des sous-marchés régionalisés [Helbich *et al.*, 2013]. Néanmoins, nous avons donc montré que les effets de contexte agissaient sur l'intensité des relations entre les variables individuelles et l'IMC, révélant leur hétérogénéité territoriale et confirmant la pertinence d'une approche localement différenciée des interventions publiques visant à encourager les comportements favorables à la santé.

3.3.3 Discussion méthodologique : les limites intrinsèques de la régression en analyse contextuelle

Un autre point mérite d'être discuté. Nous avons évoqué dans la section 2.3.3 « l'erreur régressive » dans les analyses contextuelles, consistant au recours trop systématique à la régression pour mettre en évidence des associations, qui peuvent pourtant parfois exister sans covariation à certaines échelles. La question est de savoir si les régressions locales (telles que la GWR) sont réellement adaptées à la détection de contextes spatiaux homogènes, quelle que soit l'échelle d'analyse. La régression nécessite de l'hétérogénéité pour détecter quoi que ce soit, puisqu'elle basée sur les covariations. Si, à certaines échelles et dans certains contextes, les phénomènes étudiés ne varient pas, alors les régressions locales seront inefficaces pour les caractériser. Or, on peut parfois avoir à chercher l'homogénéité. Si dans un quartier, la valeur d'un état de santé et de son déterminant principal sont tous les deux élevés mais homogènes (variances faibles), alors la GWR ne le détectera pas ou mal (en fonction du paramétrage du noyau). Pourtant, il existe bien une association locale. Elle ne révélera qu'une certaine structure si (i) l'échelle d'analyse est largement supérieure au quartier et (ii) si les quartiers sont hétérogènes. On retombe ainsi sur le constat que la notion d'hétérogénéité est indissociable de celle d'échelle. Pour en revenir à nos analyses, la pertinence de la GWR à dessiner des contextes spatiaux

pour une modélisation multiniveau subséquente est donc fonction de l'échelle d'analyse. Si elle est trop fine, la méthode peut passer à côté de structures spatiales réelles dans les associations de phénomènes. Une solution alternative pourrait alors être de construire des contextes spatiaux à partir d'associations entre variables environnementales uniquement, en fondant par exemple les analyses sur des versions locales de tests de Student plutôt que sur la variance, ou sur la régionalisation d'indicateurs multivariés issues d'analyses factorielles [Feuillet *et al.*, 2016a].

Conclusion

Ce volume a d'abord été l'occasion de démontrer du point de vue théorique que les localisations et leurs logiques peuvent être appréhendées de façon convergente et cohérente à travers une vision contextuelle - c'est-à-dire discrète - de l'espace. La localisation de toutes choses dépend de la géographie et de la structure des contextes spatiaux dans lesquels elles s'inscrivent. Pris généralement, cette position englobante permet de faire converger l'objet même de travaux géographiques variés, comme l'illustre la figure 3.11, et peut contribuer en cela à renforcer l'identité de la discipline géographique. La dépendance des localisations aux contextes spatiaux est même probablement si souvent vérifiée qu'elle devrait pouvoir être considérée comme une loi géographique.

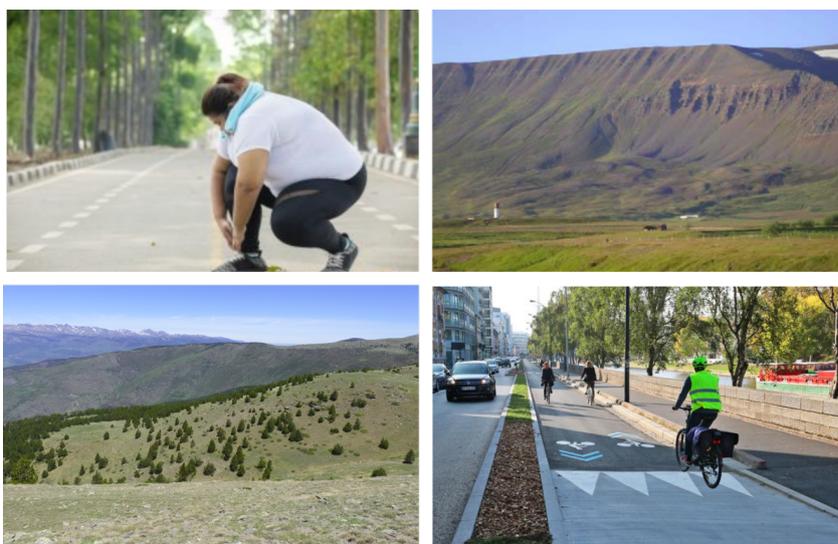


FIGURE 3.11 – Quel est le rapport entre une personne en situation d'obésité, un glissement de terrain paraglacière en Islande, une *treeline* pyrénéenne et une piste cyclable? Dans tous ces exemples, la localisation et la dynamique de l'objet étudié dépendent du contexte spatial dans lequel il s'inscrit. Délimiter ces contextes et modéliser leurs effets constitue les éléments essentiels de leur analyse géographique.

Sources : en haut à gauche : *La Gazette des communes*, 2021. En haut à droite : photographie prise dans le Skagafjörður [Feuillet et al., 2014]. En bas à gauche : massif du Puigmal, Pyrénées-Orientales, photo de l'auteur. En bas à droite : le Réseau express vélo à Rennes.

D'ailleurs, les mécanismes de genèse et d'entretien de ces contextes spatiaux sont directement dérivés des deux principales lois géographiques. Ils naissent de l'hétérogénéité spatiale (effet vertical), elle-même associée aux processus de long terme, physiques ou sociaux, qui façonnent l'espace en profondeur. Ils se nourrissent par ailleurs de la dépendance spatiale (effet horizontal), qui regroupe les interactions entre et au sein des contextes. Les effets contextuels structurent

l'espace à travers des processus intrinsèques, propres à chacun d'eux.

En s'inspirant des typologies existantes dans les disciplines connexes, notamment sur les effets de quartier, il a également été proposé dans ce volume une classification des effets contextuels selon des critères avant tout géographiques. Cette classification a été élaborée pour pouvoir servir d'assise théorique aux interprétations d'analyses géographiques empiriques diverses, analyses trop souvent déconnectées de cette dimension théorique. Les effets de contexte ont ainsi été regroupés en deux grands types : les effets endogènes et les effets exogènes. Les effets endogènes sont ceux agissant au sein même du contexte spatial. Ils impliquent deux sous-types, les effets d'interactions et les effets de normes locales partagées (cette dernière n'impliquant pas d'interactions entre agents). Les effets exogènes dérivent de la position géographique du contexte spatial dans son environnement : soit dans son environnement immédiat (effets de voisinage), soit dans une hiérarchie territoriale plus large (effets de position). Ainsi, les effets de contexte peuvent être considérés comme agissant à deux échelles : intracontextuelle et supracontextuelle.

Ce volume a également apporté les bases d'une solution méthodologique pour traiter la question centrale de la délimitation géographique des contextes spatiaux. Comment passer d'une géographie des surfaces à une géographie des contextes? Comment s'affranchir des maillages pré-établis, parfois déconnectés des processus réels? Après avoir passé en revue les limites et les faiblesses de certaines des méthodes de modélisation contextuelle les plus communes, un protocole inédit a été mis en place et appliqué à la géographie contextuelle de l'obésité en Île-de-France. Ce protocole repose sur l'idée de régionaliser les coefficients d'un modèle GWR afin d'identifier des contextes qui seraient basés uniquement sur la non-stationnarité spatiale des données. Ces contextes peuvent ensuite servir de niveaux d'agrégation dans un modèle multiniveau au cadre inférentiel solide et adapté à la structure hiérarchique des données contextuelles. L'application de ce protocole méthodologique à la géographie de l'obésité a révélé d'une part une plus grande performance statistique que les modèles intégrant des niveaux administratifs, et d'autre part un signal statistique plus fort de la présence d'effets contextuels, que ce soit sur l'obésité elle-même, ou sur la relation entre l'obésité et deux de ses principaux déterminants individuels (âge et sexe). La principale implication de ce résultat est la possibilité de s'affranchir des maillages pré-existants pour s'approcher au mieux de la géographie réelle des contextes spatiaux œuvrant à la construction de l'espace. En santé publique, cela peut impliquer une meilleure lecture de la géographie des déterminants de santé, et finalement un ciblage territorial optimisé pour mettre en place des politiques d'aménagement local favorables à la santé.

Si une partie des biais inhérents aux études contextuelles quantitatives a pu être prise en compte dans cet essai (prise en compte de la non-stationnarité spatiale, délimitation sur mesure des contextes spatiaux), il en reste encore de nombreux à contourner, constituant autant de pistes de recherche futures pour

tout amateur d'analyse contextuelle. En particulier, les biais liés aux temporalités et aux dynamiques n'ont pas été traités ici, nos analyses restant transversales et descriptives. Il est probable que la simulation puisse apporter à ce titre de précieux éléments sur la façon dont les contextes répondent à différents forçages, autant dans leurs mécanismes d'action sur les objets spatialisés, que dans leur délimitation géographique. La simulation offre en effet cette richesse de pouvoir approcher des causalités en observant les effets de la variation de différents paramétrages sur un objet d'intérêt donné. Ces pistes représentent pour l'avenir de futures collaborations fructueuses avec des spécialistes de l'analyse des systèmes complexes.

Les réflexions et les méthodes décrites dans ce volume inédit reflètent, je l'espère, ma volonté de considérer la démarche scientifique comme un continuum, reliant sans rupture les hypothèses théoriques à la mise en œuvre des méthodes et des techniques permettant de les confronter à la réalité des données. Il y a différentes catégories de scientifiques : certains se positionnent comme thématiciens-théoriciens, laissant à d'autres le soin de développer les méthodes et les outils, quand d'autres se concentrent sur ces méthodes en délaissant partiellement le terrain des théories et des hypothèses. Les deux aspects sont pourtant indissociables, et se nourrissent. Une hypothèse n'est pas valable si elle n'est pas en mesure de pouvoir être vérifiée par des méthodes solides, elles-mêmes mises en œuvre grâce à des outils adaptés. Dans l'autre sens, il faut aussi avoir conscience que la connaissance des méthodes et des techniques, même quand elles ont été conçues dans d'autres disciplines que la sienne (ce qui est souvent le cas, et ce qui justifie d'autant plus de ne pas rester prisonnier de sa discipline), contribue à suggérer de nouvelles idées et de nouvelles hypothèses. Surtout, la maîtrise de ces aspects plus techniques est un moyen de pouvoir s'affranchir des limites méthodologiques susceptibles d'entraver la formalisation de nouvelles hypothèses théoriques. L'autonomie du chercheur le long de ce continuum, sa capacité à choisir, ajuster et optimiser les méthodes et les outils à ses questionnements scientifiques, est le gage d'une parfaite adéquation entre toutes les étapes de la démarche scientifique. Cette position représente le message que je chercherai à transmettre à mes étudiants et doctorants dans les années qui viennent, autant à travers la formation que la recherche : les connaissances théoriques, les méthodes et les outils sont indissociables. Leur maîtrise combinée a les mêmes propriétés qu'un système complexe de contextes spatiaux : le tout surpasse la somme des parties.

Bibliographie

- AJZEN, I. (1985). From intentions to actions : A theory of planned behavior. *In* KUHL, J. et BECKMANN, J., éditeurs : *Action control : From cognition to behavior*, SSSP Springer Series in Social Psychology, pages 11–39. Springer, Berlin, Heidelberg. (Cité en page 37.)
- ALKER, H. (1969). A typology of ecological fallacies. *In* DOGAN, H. et ROKKAN, S., éditeurs : *Quantitative ecological analysis in the social sciences*, pages 69–86. MIT Press, London. (Cité en page 30.)
- ANSELIN, L. (1989). What is special about spatial data? alternative perspectives on spatial data analysis. Rapport technique. <https://escholarship.org/uc/item/3ph5k0d4>. (Cité en pages 12 et 42.)
- ARCAYA, M., BREWSTER, M., ZIGLER, C. M. et SUBRAMANIAN, S. V. (2012). Area variations in health : A spatial multilevel modeling approach. *Health & Place*, 18(4):824–831. (Cité en pages 64 et 82.)
- ASSUNÇÃO, R. M., NEVES, M. C., CÂMARA, G. et FREITAS, C. D. C. (2006). Efficient regionalization techniques for socio-economic geographical units using minimum spanning trees. *International Journal of Geographical Information Science*, 20(7):797–811. (Cité en page 67.)
- BACQUÉ, M.-H. et FOL, S. (2005). Ségrégation et politiques de mixité sociale aux états-Unis. *Informations sociales*, n° 125(5):82–93. (Cité en page 36.)
- BANDURA, A. (1977). *Social learning theory*. Prentice Hall, Englewood cliffs. (Cité en page 37.)
- BANOS, A. (2010). À la recherche d’effets réseaux dans la dynamique du modèle de ségrégation de Schelling. *Cybergeo : European Journal of Geography*. (Cité en page 17.)
- BARKER, R. G. (1968). *Ecological psychology*. Stanford University Press, Stanford. (Cité en page 37.)
- BARROWS, H. H. (1923). Geography as human ecology. *Annals of the Association of American Geographers*, 13(1):1–14. (Cité en page 34.)
- BATES, D., MÄCHLER, M., BOLKER, B. et WALKER, S. (2015). Fitting linear mixed-effects models using **lme4**. *Journal of Statistical Software*, 67(1). (Cité en page 69.)
- BAUMAN, A. E., SALLIS, J. F., DZEWALTOWSKI, D. A. et OWEN, N. (2002). Toward a better understanding of the influences on physical activity : the role of determinants, correlates, causal variables, mediators, moderators, and confounders. *American Journal of Preventive Medicine*, 23(2 Suppl):5–14. (Cité en page 30.)
- BLOK, D. J., van LENTHE, F. J. et de VLAS, S. J. (2018). The impact of individual and environmental interventions on income inequalities in sports participation :

- explorations with an agent-based model. *International Journal of Behavioral Nutrition and Physical Activity*, 15(1):107. (Cit  en page 36.)
- BOEHMER, T. K., HOEHNER, C. M., DESHPANDE, A. D., BRENNAN RAMIREZ, L. K. et BROWNSON, R. C. (2007). Perceived and observed neighborhood indicators of obesity among urban adults. *International Journal of Obesity (2005)*, 31(6): 968–977. (Cit  en page 57.)
- BOONE-HEINONEN, J., GUILKEY, D. K., EVENSON, K. R. et GORDON-LARSEN, P. (2010). Residential self-selection bias in the estimation of built environment effects on physical activity between adolescence and young adulthood. *International Journal of Behavioral Nutrition and Physical Activity*, 7(1):70. (Cit  en page 57.)
- BOVAIRD, J. A. (2007). Multilevel structural equation models for contextual factors. In LITTLE, T. D., BOVAIRD, J. A. et CARD, N. A.,  diteurs : *Modeling contextual effects in longitudinal studies*, pages 149–182. Lawrence Erlbaum Associates Publishers, Mahwah, NJ, US. (Cit  en page 52.)
- BRESSOUX, P. (2010). *Mod lisation statistique appliqu e aux sciences sociales*. De Boeck Sup rieur, Louvain-la-Neuve, Belgique. (Cit  en page 42.)
- BRIEND, A.-M. et JOSEPH, B. (1997). Th ses de g ographie soutenues en France de 1872   1972. *Cybergeo : European Journal of Geography*. (Cit  en page 25.)
- BRONFENBRENNER, U. (1979). *The ecology of human development : Experiments by nature and design*. Harvard University Press. (Cit  en page 37.)
- BRUNET, R. (1976). Rapport sur la “New geography” en France. In CLARK, J. I. et PINCHEMEL, P.,  diteurs : *Geography in France and in Britain*, pages 40–44. Institute of British Geographers, SSRC. (Cit  en page 26.)
- BRUNSDON, C., FOTHERINGHAM, A. S. et CHARLTON, M. E. (1996). Geographically weighted regression : a method for exploring spatial nonstationarity. *Geographical Analysis*, 28(4):281–298. (Cit  en pages 2 et 48.)
- BUNGE, W. (1962). *Theoretical geography*. Dept. of Geography, Royal University of Lund. (Cit  en page 6.)
- CASETTI, E. (1972). Generating models by the expansion method : Applications to geographical research. *Geographical Analysis*, 4(1):81–91. (Cit  en page 48.)
- CHAIX, B. (2009). Geographic life environments and coronary heart disease : a literature review, theoretical contributions, methodological updates, and a research agenda. *Annual Review of Public Health*, 30:81–105. (Cit  en page 39.)
- CHAIX, B., KESTENS, Y., DUNCAN, D. T., BRONDEEL, R., M LINE, J., EL AARBAOUI, T., PANNIER, B. et MERLO, J. (2016). A GPS-based methodology to analyze environment-health associations at the trip level : Case-crossover analyses of built environments and walking. *American Journal of Epidemiology*, 184(8):579–589. (Cit  en page 40.)
- CHAIX, B., MERLO, J., SUBRAMANIAN, S. V., LYNCH, J. et CHAUVIN, P. (2005). Comparison of a spatial perspective with the multilevel analytical approach

- in neighborhood studies : The case of mental and behavioral disorders due to psychoactive substance use in Malmö, Sweden, 2001. *American Journal of Epidemiology*, 162(2):171–182. (Cit  en page 43.)
- CHAIX, B., M LINE, J., DUNCAN, S., JARDINIER, L., PERCHOUX, C., VALL E, J., MERRIEN, C., KARUSISI, N., LEWIN, A., BRONDEEL, R. et KESTENS, Y. (2013). Neighborhood environments, mobility, and health : towards a new generation of studies in environmental health research. *Revue D'epidemiologie Et De Sante Publique*, 61 Suppl 3:S139–145. (Cit  en page 39.)
- CHARREIRE, H., FEUILLET, T., RODA, C., MACKENBACH, J. D., COMPERNOLLE, S., GLONTI, K., B RDOS, H., LE VAILLANT, M., RUTTER, H., MCKEE, M., DE BOURDEAUDHUIJ, I., BRUG, J., LAKERVELD, J. et OPPERT, J.-M. (2016). Self-defined residential neighbourhoods : size variations and correlates across five European urban regions : Self-defined residential neighbourhoods. *Obesity Reviews*, 17:9–18. (Cit  en page 40.)
- CHASLES, V. (2020). *Territoire et  galit  des chances, que nous apprennent les in galit s environnementales de sant  ?* Habilitation   diriger des recherches, Universit  Lyon 3. (Cit  en page 85.)
- CHAVENT, M., KUENTZ-SIMONET, V., LABENNE, A. et SARACCO, J. (2018). Clust-Geo : an R package for hierarchical clustering with spatial constraints. *Computational Statistics*, 33(4):1799–1822. (Cit  en pages 67 et 69.)
- CHEN, D.-R. et TRUONG, K. (2012). Using multilevel modeling and geographically weighted regression to identify spatial variations in the relationship between place-level disadvantages and obesity in Taiwan. *Applied Geography*, 32(2):737–745. (Cit  en page 63.)
- CHETTY, R., HENDREN, N. et KATZ, L. F. (2016). The effects of exposure to better neighborhoods on children : New evidence from the Moving to Opportunity experiment. *American Economic Review*, 106(4):855–902. (Cit  en page 36.)
- CLAVAL, P. (1998). *Histoire de la g ographie fran aise : De 1870   nos jours*. Nathan Universit , Paris. (Cit  en page 34.)
- CLERC, P. (2004). Lieu. *Hypergeo*. (Cit  en page 8.)
- CLOZIER, R. (1940). *La gare du Nord : Th se pour le doctorat pr sent e   la facult  des lettres de Paris*. Th se de doctorat, Paris. (Cit  en page 25.)
- COLIN, S. et ROUSSEZ, V. (2018). Elaboration d'une maille g ographique pour l'habitat. Paris. Insee. (Cit  en page 67.)
- COSSART, E. et FRESSARD, M. (2017). Assessment of structural sediment connectivity within catchments : insights from graph theory. *Earth Surface Dynamics*, 5(2):253–268. (Cit  en page 17.)
- COX, K. R. (1969). The voting decision in a spatial context. *Progress in Geography*, 1:81–117. (Cit  en page 36.)
- CUMMINS, S., CURTIS, S., DIEZ-ROUX, A. V. et MACINTYRE, S. (2007). Understanding and representing 'place' in health research : A relational approach. *Social Science   Medicine*, 65(9):1825–1838. (Cit  en pages 3, 4, 39 et 56.)

- CUYALA, S. (2015). L'affirmation de la géographie quantitative française au cœur d'un moment d'ébullition disciplinaire (1972-1984). *Bulletin de l'association de géographes français. Géographies*, 92(92-1):67–83. (Cité en page 26.)
- DALE, M. R. T. et FORTIN, M.-J. (2014). *Spatial analysis : A guide for ecologists*. Cambridge University Press, Cambridge. (Cité en page 17.)
- DARWIN, C. (1859). *L'origine des espèces au moyen de la sélection naturelle ou la préservation des races favorisées dans la lutte pour la survie*. Murray, Londres. (Cité en page 34.)
- DAVID, Q. et VAN HAMME, G. (2011). Pillars and electoral behavior in Belgium : The neighborhood effect revisited. *Political Geography*, 30(5):250–262. (Cité en page 36.)
- DIEZ-ROUX, A. V. (2000). Multilevel analysis in public health research. *Annual Review of Public Health*, 21(1):171–192. (Cité en page 37.)
- DIEZ-ROUX, A. V. (2002). A glossary for multilevel analysis. *Journal of Epidemiology & Community Health*, 56(8):588–594. (Cité en pages 42 et 55.)
- DIEZ-ROUX, A. V. et MAIR, C. (2010). Neighborhoods and health. *Annals of the New York Academy of Sciences*, 1186:125–145. (Cité en pages 38, 55 et 56.)
- DOGAN, M. et ROKKAN, S. (1967). Quantitative ecological analysis : Contexts, trends, tasks. *Social Science Information*, 6(6):35–47. (Cité en page 30.)
- DONG, G., MA, J., HARRIS, R. et PRYCE, G. (2016). Spatial random slope multilevel modeling using multivariate conditional autoregressive models : A case study of subjective travel satisfaction in Beijing. *Annals of the American Association of Geographers*, 106(1):19–35. (Cité en page 64.)
- DUNCAN, C., JONES, K. et MOON, G. (1996). Health-related behaviour in context : A multilevel modelling approach. *Social Science & Medicine*, 42(6):817–830. (Cité en page 54.)
- DUNCAN, C., JONES, K. et MOON, G. (1998). Context, composition and heterogeneity : Using multilevel models in health research. *Social Science & Medicine*, 46(1):97–117. (Cité en page 46.)
- DUNCAN, T. E., DUNCAN, S. C., ALPERT, A., HOPS, H., STOOLMILLER, M. et MUTHEN, B. (1997). Latent variable modeling of longitudinal and multilevel substance use data. *Multivariate Behavioral Research*, 32(3):275–318. (Cité en page 43.)
- DURAND-DASTÈS, F. (1984). Systèmes et localisations : Problèmes théoriques et formels. *Géopoint*, 84:19–44. (Cité en pages 9 et 16.)
- DURAND-DASTÈS, F. (1991). La mémoire de Gaïa. *Géopoint*, 90:141–146. (Cité en page 16.)
- DURAND-DASTÈS, F. (1992). Le particulier et le général en géographie. *In Actes du 6e Colloque de Didactique de l'histoire et de la géographie*, pages 209–219, Institut National de Recherche Pédagogique : Paris. (Cité en pages 9 et 16.)

- DURAND-DASTÈS, F. (2001). Le temps, la géographie et ses modèles. *Bulletin de la Société géographique de Liège*, 40:5–13. (Cité en pages 14, 15, 16 et 107.)
- DURKHEIM, E. (1893). *De la division du travail social*. Presses Universitaires de France (ré-impression), Paris. ISSN : 0291-0489. (Cité en page 20.)
- DUTILLEUL, P. et LEGENDRE, P. (1993). Spatial heterogeneity against heteroscedasticity : An ecological paradigm versus a statistical concept. *Oikos*, 66(1):152–171. (Cité en page 13.)
- ECOB, R. (1996). A Multilevel modelling approach to examining the effects of Area of residence on health and functioning. *Journal of the Royal Statistical Society : Series A (Statistics in Society)*, 159(1):61–75. (Cité en page 43.)
- EWING, R. et CERVERO, R. (2010). Travel and the Built Environment. *Journal of the American Planning Association*, 76(3):265–294. (Cité en page 70.)
- FEVRE, L. (1922). *La terre et l'évolution humaine. Introduction géographique à l'histoire*. Albin Michel, Paris. (Cité en page 34.)
- FEUILLET, T., BIRRE, D., MILIAN, J., GODARD, V., CLAUZEL, C. et SERRANO-NOTIVOLI, R. (2020a). Spatial dynamics of alpine tree lines under global warming : What explains the mismatch between tree densification and elevational upward shifts at the tree line ecotone? *Journal of Biogeography*, 47(5):1056–1068. (Cité en page 27.)
- FEUILLET, T., CHARREIRE, H., MENAI, M., SALZE, P., SIMON, C., DUGAS, J., HERCBERG, S., ANDREEVA, V. A., ENAUX, C., WEBER, C. et OPPERT, J.-M. (2015). Spatial heterogeneity of the relationships between environmental characteristics and active commuting : towards a locally varying social ecological model. *International Journal of Health Geographics*, 14(1). (Cité en page 85.)
- FEUILLET, T., CHARREIRE, H., RODA, C., BEN REBAH, M., MACKENBACH, J. D., COMPERNOLLE, S., GLONTI, K., BÁRDOS, H., RUTTER, H., DE BOURDEAUDHUIJ, I., MCKEE, M., BRUG, J., LAKERVELD, J. et OPPERT, J.-M. (2016a). Neighbourhood typology based on virtual audit of environmental obesogenic characteristics : Virtual audit and neighbourhood typology. *Obesity Reviews*, 17:19–30. (Cité en pages 39 et 87.)
- FEUILLET, T., COMMENGES, H., MENAI, M., SALZE, P., PERCHOUX, C., REUILLON, R., KESSE-GUYOT, E., ENAUX, C., NAZARE, J. A., HERCBERG, S., SIMON, C., CHARREIRE, H. et OPPERT, J. M. (2018). A massive geographically weighted regression model of walking-environment relationships. *Journal of Transport Geography*, 68:118–129. (Cité en pages 50 et 67.)
- FEUILLET, T., COQUIN, J., MERCIER, D., COSSART, E., DECAULNE, A., JÓNSSON, H. P. et SÆMUNDSSON, T. (2014). Focusing on the spatial non-stationarity of landslide predisposing factors in northern Iceland : Do paraglacial factors vary over space? *Progress in Physical Geography : Earth and Environment*, 38(3):354–377. (Cité en pages 2, 3 et 89.)
- FEUILLET, T., SALZE, P., CHARREIRE, H., MENAI, M., ENAUX, C., PERCHOUX, C., HESS, F., KESSE-GUYOT, E., HERCBERG, S., SIMON, C., WEBER, C. et

- OPPERT, J.-M. (2016b). Built environment in local relation with walking : Why here and not there? *Journal of Transport & Health*, 3(4):500–512. (Cité en pages 50 et 67.)
- FEUILLET, T., VALETTE, J. F., CHARREIRE, H., KESSE-GUYOT, E., JULIA, C., VERNEZ-MOUDON, A., HERCBERG, S., TOUVIER, M. et OPPERT, J. M. (2020b). Influence of the urban context on the relationship between neighbourhood deprivation and obesity. *Social Science & Medicine*, 265:113537. (Cité en pages 21, 23, 62, 70 et 71.)
- FLAKE, G. W. (1998). *The computational beauty of nature : Computer explorations of fractals, chaos, complex systems, and adaptation*. MIT Press, Cambridge, Mass. (Cité en page 16.)
- FOTHERINGHAM, A. S., BRUNSDON, C. et CHARLTON, M. (2002). *Geographically weighted regression : The analysis of spatially varying relationships*. Wiley-Blackwell, Chichester, England ; Hoboken, NJ, USA. (Cité en pages 2, 48 et 62.)
- FOTHERINGHAM, A. S., CHARLTON, M. et BRUNSDON, C. (1997). Measuring spatial variations in relationships with geographically weighted regression. In FISCHER, M. M. et GETIS, A., éditeurs : *Recent developments in spatial analysis : Spatial statistics, behavioural modelling, and computational intelligence*, Advances in Spatial Science, pages 60–82. Springer, Berlin, Heidelberg. (Cité en page 2.)
- FRIEDMAN, M. (1955). *The role of government in education*. Rutgers University Press, New Brunswick, N.J. OCLC : 80693211. (Cité en page 35.)
- GALLEGO, A., BUSCHA, F., STURGIS, P. et OBERSKI, D. (2016). Places and preferences : A longitudinal analysis of self-selection and contextual effects. *British Journal of Political Science*, 46(3):529–550. (Cité en page 36.)
- GALSTER, G. C. (2012). The mechanism(s) of neighbourhood effects : Theory, evidence, and policy implications. In van HAM, M., MANLEY, D., BAILEY, N., SIMPSON, L. et MACLENNAN, D., éditeurs : *Neighbourhood effects research : New perspectives*, pages 23–56. Springer Netherlands, Dordrecht. (Cité en page 18.)
- GAN, K. et BROU, G. (2018). *Introduction à la modélisation par équations structurales : Manuel pratique avec lavaan*. ISTE Group, Londres. (Cité en pages 26 et 27.)
- GELFAND, A. E., KIM, H.-J., SIRMANS, C. F. et BANERJEE, S. (2003). Spatial modeling with spatially varying coefficient processes. *Journal of the American Statistical Association*, 98(462):387–396. (Cité en page 48.)
- GEORGE, P. (1935). *La région du Bas Rhône : étude de géographie régionale*. J.-B. Baillièrre et Fils, Paris. (Cité en page 25.)
- GETIS, A. (1994). Spatial dependence and heterogeneity and proximal databases. In FOTHERINGHAM, A. S. et ROGERSON, P., éditeurs : *Spatial analysis and GIS*, pages 105–120. CRC Press, Londres, Taylor & Francis édition. (Cité en page 13.)

- GILES-CORTI, B., VERNEZ-MOUDON, A., REIS, R., TURRELL, G., DANNENBERG, A. L., BADLAND, H., FOSTER, S., LOWE, M., SALLIS, J. F., STEVENSON, M. et OWEN, N. (2016). City planning and population health : a global challenge. *The Lancet*, 388(10062):2912–2924. (Cité en page 70.)
- GOLDSTEIN, H. (1987). *Multilevel models in education and social research*. Multilevel models in education and social research. Oxford University Press, New York, NY, US. (Cité en page 42.)
- GOLLEDGE, R. et AMEDEO, D. (1968). On laws in Geography. *Annals of the Association of American Geographers*, 58(4):760–774. (Cité en page 26.)
- GOLLINI, I., LU, B., CHARLTON, M., BRUNSDON, C. et HARRIS, P. (2015). GW-model : An R package for exploring spatial heterogeneity using geographically weighted models. *Journal of Statistical Software*, 63(1):1–50. (Cité en page 69.)
- GOODCHILD, M. (2003). The fundamental laws of GIScience. Pacific Grove, CA. (Cité en page 24.)
- GOODCHILD, M. F. (2004). GIScience, geography, form, and process. *Annals of the Association of American Geographers*, 94(4):709–714. (Cité en pages 12 et 27.)
- GOODCHILD, M. F. (2008). Theoretical geography (1962) : William Bunge. In HUBBARD, P., KITCHIN, R. et VALENTINE, G., éditeurs : *Key texts in human Geography*, pages 9–16. SAGE Publications Inc., Londres. (Cité en page 17.)
- GOODCHILD, M. F. (2009). Geographic information systems and science : today and tomorrow. *Annals of GIS*, 15(1):3–9. (Cité en page 12.)
- GOODCHILD, M. F. (2016). GIS in the era of big data. *Cybergeog : European Journal of Geography*. (Cité en page 5.)
- GRASLAND, C. (2001). Agrégation de géographie. Préparation à l'épreuve écrite de commentaire de documents. Module n°7 : contextes spatiaux et contextes territoriaux. (Cité en pages 8 et 9.)
- GRIFFITH, D. A. (2003). *Spatial autocorrelation and spatial filtering : Gaining understanding through theory and scientific visualization*. Springer Science & Business Media. (Cité en page 48.)
- HAGGETT, P. (1965). *Locational analysis in human geography*. Edward Arnold, Londres. (Cité en pages 26 et 34.)
- HAJAT, A., DIEZ-ROUX, A. V., ADAR, S. D., AUCHINCLOSS, A. H., LOVASI, G. S., O'NEILL, M. S., SHEPPARD, L. et KAUFMAN, J. D. (2013). Air pollution and individual and neighborhood socioeconomic status : evidence from the Multi-Ethnic Study of Atherosclerosis (MESA). *Environmental Health Perspectives*, 121(11-12):1325–1333. (Cité en page 63.)
- HANDY, S., CAO, X. et MOKHTARIAN, P. L. (2006). Self-selection in the relationship between the built environment and walking : Empirical evidence from Northern California. *Journal of the American Planning Association*, 72(1):55–74. (Cité en page 57.)

- HARTSHORNE, R. (1939). The nature of geography : A critical survey of current thought in the light of the past. *Annals of the Association of American Geographers*, 29(3):173–412. (Cité en page 34.)
- HARVEY, D. (2011). Roepke lecture in economic Geography—Crises, geographic disruptions and the uneven development of political responses. *Economic Geography*, 87(1):1–22. (Cité en page 34.)
- HELBICH, M., BRUNAUER, W., HAGENAUER, J. et LEITNER, M. (2013). Data-driven regionalization of housing markets. *Annals of the Association of American Geographers*, 103(4):871–889. (Cité en pages 67, 68 et 86.)
- HERCBERG, S., CASTETBON, K., CZERNICHOW, S., MALON, A., MEJEAN, C., KESSE, E., TOUVIER, M. et GALAN, P. (2010). The Nutrinet-Santé Study : a web-based prospective study on the relationship between nutrition and health and determinants of dietary patterns and nutritional status. *BMC Public Health*, 10(1):242. (Cité en page 72.)
- HOCHBAUM, G. M. (1958). *Public participation in medical screening programs : A socio-psychological study*. U.S. Department of Health, Education, and Welfare, Public Health Service, Bureau of State Services, Division of Special Health Services, Tuberculosis Program. (Cité en page 37.)
- HÄGERSTRAND, T. (1970). What about people in regional science? *Papers of the Regional Science Association*, 24(1):6–21. (Cité en page 39.)
- JANKO, M., GOEL, V. et EMCH, M. (2019). Extending multilevel spatial models to include spatially varying coefficients. *Health & Place*, 60:102235. (Cité en pages 49, 54 et 64.)
- JENCKS, C. et MAYER, S. (1990). The social consequences of growing up in a poor neighborhood. In LYNN, L. et MCGEARY, M., éditeurs : *Inner-city poverty in the United States*, pages 111–186. Washington DC, national academy press édition. (Cité en page 19.)
- JIANG, B. (2015). Geospatial analysis requires a different way of thinking : the problem of spatial heterogeneity. *GeoJournal*, 80(1):1–13. (Cité en page 14.)
- JOHNSTON, R. J. (1983). The neighbourhood effect won't go away : Observations on the electoral geography of England in the light of Dunleavy's critique. *Geoforum*, 14(2):161–168. (Cité en page 36.)
- JOHNSTON, R. J. (1991). *Geography and geographers : Anglo-American human geography since 1945*. Edward Arnold, Londres. (Cité en page 34.)
- JONES, K. (1991). Specifying and estimating multilevel models for geographical research. *Transactions of the Institute of British Geographers*, 16(2):148–159. (Cité en page 43.)
- JONES, K. (1993). Everywhere is nowhere : multilevel perspectives on the importance of place ; inaugural lecture. (Cité en page 54.)
- JONES, K. et DUNCAN, C. (1995). Individuals and their ecologies : analysing the geography of chronic illness within a multilevel modelling framework. *Health & Place*, 1(1):27–40. (Cité en page 43.)

- JONES, K., JOHNSTON, R. et PATTIE, C. J. (1992). People, places and regions : Exploring the use of multi-level modelling in the analysis of electoral data. *British Journal of Political Science*, 22(3):343–380. (Cité en page 43.)
- JUILLARD, E. (1953). *La vie rurale dans la plaine de Basse-Alsace : Essai de géographie sociale*. F.X. Le Roux. (Cité en page 25.)
- KIRKPATRICK, S., GELATT, C. D. et VECCHI, M. P. (1983). Optimization by simulated annealing. *Science*, 220(4598):671–680. (Cité en page 66.)
- KWAN, M.-P. (2012). The uncertain geographic context problem. *Annals of the Association of American Geographers*, 102(5):958–968. (Cité en pages 39 et 56.)
- KWAN, M.-P. (2018). The limits of the neighborhood effect : Contextual uncertainties in geographic, environmental health, and social science research. *Annals of the American Association of Geographers*, 108(6):1482–1490. (Cité en pages 56 et 58.)
- KWAN, M. P. et LEE, J. (2004). Geovisualization of human activity patterns using 3D GIS : A time-geographic approach. In GOODCHILD, M. et JANELLE, D., éditeurs : *Spatially Integrated Social Science*, pages 48–66. Oxford University Press, New York. (Cité en page 39.)
- LE ROUX, G., VALLÉE, J. et COMMENGES, H. (2017). Social segregation around the clock in the Paris region (France). *Journal of Transport Geography*, 59:134–145. (Cité en page 58.)
- LESAGE, J. P. (2008). An introduction to spatial econometrics. *Revue d'économie industrielle*, (123):19–44. (Cité en page 48.)
- LIN, L. et MOUDON, A. V. (2010). Objective versus subjective measures of the built environment, which are most effective in capturing associations with walking? *Health & Place*, 16(2):339–348. (Cité en page 57.)
- LONEY, T. et NAGELKERKE, N. J. (2014). The individualistic fallacy, ecological studies and instrumental variables : a causal interpretation. *Emerging Themes in Epidemiology*, 11(1):18. (Cité en page 52.)
- LÜDECKE, D., BARTEL, A., SCHWEMMER, C., POWELL, C., DJALOVSKI, A. et TITZ, J. (2021). sjPlot : Data visualization for statistics in social science. (Cité en page 69.)
- MACINTYRE, S., ELLAWAY, A. et CUMMINS, S. (2002). Place effects on health : how can we conceptualise, operationalise and measure them? *Social Science & Medicine*, 55(1):125–139. (Cité en pages 3, 18 et 39.)
- MANDELBROT, B. B. (1982). *The fractal geometry of nature*. Freeman & Co Ltd, San Francisco. (Cité en page 24.)
- MANSKI, C. F. (1993). Identification of endogenous social effects the reflection problem. *Review of Economic Studies*, 60(3):531–542. (Cité en page 19.)
- MARPSAT, M. (1999). La modélisation des effets de quartier aux États-Unis : une revue des travaux récents. *Population*, 54(2):303–330. (Cité en page 35.)

- MATHIAN, H. et SANDERS, L. (2014). *Objets géographiques et processus de changement : Approches spatio-temporelles*. ISTE Group, Londres. (Cité en page 17.)
- MCCLEROY, K. R., BIBEAU, D., STECKLER, A. et GLANZ, K. (1988). An ecological perspective on health promotion programs. *Health Education Quarterly*, 15(4): 351–377. (Cité en pages 37 et 38.)
- MERLO, J., CHAIX, B., YANG, M., LYNCH, J. et RÅSTAM, L. (2005). A brief conceptual tutorial of multilevel analysis in social epidemiology : linking the statistical concept of clustering to the idea of contextual phenomenon. *Journal of Epidemiology & Community Health*, 59(6):443–449. (Cité en pages 20, 45 et 47.)
- MILLER, H. J. (2004). Tobler's first law and spatial analysis. *Annals of the Association of American Geographers*, 94(2):284–289. (Cité en page 11.)
- MILLER, H. J. et GOODCHILD, M. F. (2015). Data-driven geography. *GeoJournal*, 80(4):449–461. (Cité en page 16.)
- MILLER, W. L. (1977). *Electoral dynamics in Britain since 1918*. Palgrave Macmillan UK. (Cité en page 36.)
- MOKHTARIAN, P. L. et CAO, X. (2008). Examining the impacts of residential self-selection on travel behavior : A focus on methodologies. *Transportation Research Part B : Methodological*, 42(3):204–228. (Cité en page 57.)
- MYERS, C. A., SLACK, T., MARTIN, C. K., BROYLES, S. T. et HEYMSFIELD, S. B. (2015). Regional disparities in obesity prevalence in the United States : A spatial regime analysis. *Obesity*, 23(2):481–487. (Cité en page 42.)
- NSF (2020). The State of U.S. Science and Engineering 2020 | NSF - National Science Foundation. Rapport technique. (Cité en page 5.)
- OAKES, J. M. (2004). The (mis)estimation of neighborhood effects : causal inference for a practicable social epidemiology. *Social Science & Medicine*, 58(10):1929–1952. (Cité en page 35.)
- OPENSHAW, S. (1977). A geographical solution to scale and aggregation problems in region-building, partitioning and spatial modelling. *Transactions of the Institute of British Geographers*, 2(4):459–472. (Cité en pages 30 et 54.)
- OSHAN, T. M., SMITH, J. P. et FOTHERINGHAM, A. S. (2020). Targeting the spatial context of obesity determinants via multiscale geographically weighted regression. *International Journal of Health Geographics*, 19(1):11. (Cité en pages 62 et 71.)
- OWEN, G., HARRIS, R. et JONES, K. (2016). Under examination : Multilevel models, geography and health research. *Progress in Human Geography*, 40(3):394–412. (Cité en pages 43, 47 et 54.)
- PARK, R. E. et BURGESS, E. W. (1925). *The city*. University of Chicago Press. (Cité en page 34.)
- PARK, Y. M. et KIM, Y. (2014). A spatially filtered multilevel model to account for spatial dependency : application to self-rated health status in South Korea. *International Journal of Health Geographics*, 13(1):6. (Cité en page 63.)

- PATTIE, C. et JOHNSTON, R. (2000). 'People who talk together vote together' : An exploration of contextual effects in Great Britain. *Annals of the Association of American Geographers*, 90(1):41–66. (Cité en page 36.)
- PELLETIER, P. (2016). Élisée Reclus : Théorie géographique et théorie anarchiste. *Terra Brasilis (Nova Série). Revista da Rede Brasileira de História da Geografia e Geografia Histórica*, (7). (Cité en page 34.)
- PERCHOUX, C., CHAIX, B., CUMMINS, S. et KESTENS, Y. (2013). Conceptualization and measurement of environmental exposure in epidemiology : Accounting for activity space related to daily mobility. *Health & Place*, 21:86–93. (Cité en page 39.)
- PERCHOUX, C., CHAIX, B. et KESTENS, Y. (2019). Activity spaces in place and health research : Novel exposure measures, data collection tools, and designs. *Health & Place*, 58:102130. (Cité en page 39.)
- PETROVIĆ, A., MANLEY, D. et van HAM, M. (2020). Freedom from the tyranny of neighbourhood : Rethinking sociospatial context effects. *Progress in Human Geography*, 44(6):1103–1123. (Cité en pages 8, 39 et 56.)
- PINCHEMEL, P. (1981). Langages de la géographie, langages des géographes. *Annales de Géographie*, 90(502):686–689. (Cité en page 16.)
- PINCHEMEL, P. et PINCHEMEL, G. (1997). *La face de la Terre : Éléments de géographie*. Armand Colin, Paris, 5e édition édition. (Cité en page 9.)
- PUMAIN, D. (2003). Une approche de la complexité en géographie. *Géocarrefour*, 78(Vol. 78/1):25–31. (Cité en page 32.)
- PUMAIN, D. et ROBIC, M.-C. (2002). Le rôle des mathématiques dans une « révolution » théorique et quantitative : la géographie française depuis les années 1970. *Revue d'Histoire des Sciences Humaines*, no 6(1):123–144. (Cité en page 26.)
- PUMAIN, D. et SAINT-JULIEN, T. (2010). *Analyse spatiale : Les localisations*. Armand Colin. (Cité en pages 1, 9 et 16.)
- R_CORE_TEAM (2021). R : A language and environment for statistical computing. R Foundation for Statistical Computing. (Cité en page 69.)
- REY, G., JOUGLA, E., FOUILLET, A. et HÉMON, D. (2009). Ecological association between a deprivation index and mortality in France over the period 1997 – 2001 : variations with spatial scale, degree of urbanicity, age, gender and cause of death. *BMC Public Health*, 9(1):33. (Cité en page 21.)
- RHEIN, C. (2003). L'écologie humaine, discipline-chimère. *Sociétés contemporaines*, no 49-50(1):167–190. (Cité en page 34.)
- RICAN, S. et VAILLANT, Z. (2015). Urbanisme et réduction des inégalités sociales de santé : les diagnostics locaux de santé. *La santé en action*, pages 24–25. (Cité en page 85.)
- ROBINSON, W. S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review*, 15:351–357. (Cité en pages 30 et 55.)

- ROCHE, S. (2016). Geographic information science II : Less space, more places in smart cities. *Progress in Human Geography*, 40(4):565–573. (Cité en page 8.)
- RODA, C., CHARREIRE, H., FEUILLET, T., MACKENBACH, J., COMPERNOLLE, S., GLONTI, K., BEN REBAH, M., BÁRDOS, H., RUTTER, H., MCKEE, M., DE BOURDEAUDHUIJ, I., BRUG, J., LAKERVELD, J. et OPPERT, J.-M. (2016a). Mismatch between perceived and objectively measured environmental obesogenic features in European neighbourhoods : Perceived and objectively measured obesogenic environment. *Obesity Reviews*, 17:31–41. (Cité en page 58.)
- RODA, C., CHARREIRE, H., FEUILLET, T., MACKENBACH, J. D., COMPERNOLLE, S., GLONTI, K., BÁRDOS, H., RUTTER, H., MCKEE, M., BRUG, J., DE BOURDEAUDHUIJ, I., LAKERVELD, J. et OPPERT, J.-M. (2016b). Lifestyle correlates of overweight in adults : a hierarchical approach (the SPOTLIGHT project). *International Journal of Behavioral Nutrition and Physical Activity*, 13(1). (Cité en page 39.)
- ROUNTREE, P. W. et CLAYTON, R. R. (1999). A contextual model of adolescent alcohol use across the rural-urban continuum. *Substance Use & Misuse*, 34(4-5):495–519. (Cité en page 43.)
- RUNDLE, A., DIEZ ROUX, A. V., FREE, L. M., MILLER, D., NECKERMAN, K. M. et WEISS, C. C. (2007). The urban built environment and obesity in New York City : a multilevel analysis. *American journal of health promotion : AJHP*, 21(4 Suppl):326–334. (Cité en page 43.)
- SAELENS, B. E. et HANDY, S. L. (2008). Built environment correlates of walking : A review. *Medicine and science in sports and exercise*, 40(7 Suppl):S550–S566. (Cité en page 70.)
- SANDERS, L. (1999). Permanences et ruptures, questions de mesure et d'interprétation. *Travaux de l'Institut de Géographie de Reims*, 26(101):25–62. (Cité en page 9.)
- SANDERS, L. (2001). *Modèles en analyse spatiale*. Hermes Science Publications, Paris. (Cité en page 27.)
- SANDERS, L. (2011). Géographie quantitative et analyse spatiale : quelles formes de scientificité? In *Les sciences humaines sont-elles des sciences ?*, pages 71–91. Vuibert, Paris. (Cité en page 26.)
- SAUER, C. O. (1927). *Recent developments in cultural geography*. s.n., S.l. OCLC : 800731248. (Cité en page 34.)
- SCHAEFER, F. K. (1953). Exceptionalism in Geography : A methodological examination. *Annals of the Association of American Geographers*, 43(3):226–249. (Cité en page 26.)
- SIMPSON, E. H. (1951). The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society : Series B (Methodological)*, 13(2):238–241. (Cité en page 54.)
- SORRE, M. (1933). Complexes pathogènes et géographie médicale. *Annales de Géographie*, 42(235):1–18. (Cité en page 37.)

- SORRE, M. (1948). Fondements de la géographie humaine. *Cahiers Internationaux de Sociologie*, 5:21–37. (Cité en page 34.)
- STOKOLS, D. (1992). Establishing and maintaining healthy environments : Toward a social ecology of health promotion. *American Psychologist*, 47(1):6–22. (Cité en page 37.)
- SUSSER, M. (1973). *Causal thinking in the health sciences : concepts and strategies of epidemiology*. Oxford University Press, New York. (Cité en page 30.)
- TENNEKES, M. (2018). **tmap** : Thematic maps in R. *Journal of Statistical Software*, 84(6). (Cité en page 69.)
- TOBLER, W. R. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography*, 46(sup1):234–240. (Cité en page 11.)
- TRICART, J. (1949). *La partie orientale du Bassin de Paris . Tome I . La genèse du Bassin : étude morphologique*. Société d'édition d'enseignement supérieur, Paris. (Cité en page 25.)
- TSALLIS, C. et STARIOLO, D. A. (1996). Generalized simulated annealing. *Physica A : Statistical Mechanics and its Applications*, 233(1):395–406. (Cité en page 66.)
- VALLÉE, J. (2017). The daycourse of place. *Social Science & Medicine*, 194:177–181. (Cité en page 40.)
- VALLÉE, J., CADOT, E., ROUSTIT, C., PARIZOT, I. et CHAUVIN, P. (2011). The role of daily mobility in mental health inequalities : The interactive influence of activity space and neighbourhood of residence on depression. *Social Science & Medicine*, 73(8):1133–1144. (Cité en page 56.)
- VARENNE, F. (2017). *Théories et modèles en sciences humaines, Le cas de la géographie*. Éditions Matériologiques, Paris. (Cité en pages 26 et 34.)
- VON KORFF, M., KOEPSSELL, T., CURRY, S. et DIEHR, P. (1992). Multi-level analysis in epidemiologic research on health behaviors and outcomes. *American Journal of Epidemiology*, 135(10):1077–1082. (Cité en pages 37 et 43.)
- WHEELER, D. et TIEFELSDORF, M. (2005). Multicollinearity and correlation among local regression coefficients in geographically weighted regression. *Journal of Geographical Systems*, 7(2):161–187. (Cité en page 51.)
- WILSON, R. (2013). Changing geographic units and the analytical consequences : An example of Simpson's paradox. *Cityscape*, 15(2):289–304. (Cité en page 54.)
- WOODS, R. A. (1914). The neighborhood in social reconstruction. *American Journal of Sociology*, 19(5):577–591. (Cité en page 35.)
- WU, J. (2004). Effects of changing scale on landscape pattern analysis : scaling relations. *Landscape Ecology*, 19(2):125–138. (Cité en page 12.)
- XIANG, Y., GUBIAN, S., SUOMELA, B. P. et HOENG, J. (2013). Generalized Simulated Annealing for Global Optimization : The GenSA Package. (Cité en pages 66 et 69.)

Table des figures

1	Influence non-stationnaire de la pente sur la probabilité d'occurrence des glissements de terrain paraglaciales en Islande du nord	3
1.1	Représentation schématique des systèmes de contextes spatiaux dans un espace hétérogène	10
1.2	Illustration théorique du rôle essentiel de l'échelle dans la distinction entre homogénéité et hétérogénéité spatiales	13
1.3	Représentation sagittale des mécanismes de structuration des systèmes de contextes spatiaux	14
1.4	Modèle d'explication des localisations selon Durand-Dastès [2001]	15
1.5	Typologie des effets contextuels	22
1.6	Relations entre le risque d'être en situation de surpoids et le niveau de défaveur sociale du quartier de résidence, en fonction du contexte urbain, dans un échantillon de 70 000 personnes de la cohorte Nutrinet-Santé	23
1.7	Représentation schématique de la composante fractale des systèmes de contextes spatiaux	24
1.8	Illustration théorique de l'importance des contextes spatiaux dans l'analyse de la dynamique de la <i>treeline</i>	29
1.9	Illustration photographique de deux contextes spatiaux d'évolution de la limite supérieure de la strate arborée en milieu montagnard	29
1.10	Illustration théorique de l'importance de la considération des contextes spatiaux dans les chaînes causales	31
2.1	Vision schématique du modèle socio-écologique des déterminants de la santé	38
2.2	Exemple de délimitation du quartier de résidence sur la base de la perception individuelle	40
2.3	Décomposition de la variance et des résidus au niveau individuel et au niveau contextuel : un principe essentiel de la modélisation multiniveau	45
2.4	Illustration théorique de la capacité des modèles multiniveaux à modéliser des relations statistiques territorialement hétérogènes	46
2.5	La modélisation GWR : une méthode de régression locale basée sur la définition d'un noyau	49
2.6	Illustration du schéma de paramétrage de la GWR	50
2.7	Forces et faiblesses des principales méthodes actuelles de modélisation statistique des effets de contexte	51
2.8	Typologie des erreurs récurrentes en modélisation statistique des effets contextuels	53
2.9	Comparaison de la surmortalité totale au printemps 2020 en fonction des contextes territoriaux en IDF	60

3.1	Protocole méthodologique pour la mise en place du modèle hybride combinant modèle GWR et modèle multiniveau	65
3.2	Variation du pourcentage de personnes en surpoids selon les contextes urbains dans la cohorte NutriNet-Santé	70
3.3	Densité de noyau de l'échantillon de nutrinautes utilisé dans les analyses statistiques	73
3.4	Distribution des nutrinautes en fonction des mailles mentionnées ou utilisées dans les modèles	74
3.5	Cartographie des contextes spatiaux selon les limites administratives et selon non-stationnairté spatiale des données	74
3.6	Cartes de surfaces des coefficients GWR, après interpolation (méthode IDW)	75
3.7	Représentation graphique de l'hétérogénéité des relations entre l'âge et l'IMC en fonction des contextes spatiaux	78
3.8	Cartographie de la densité de bâti en fonction des contextes	79
3.9	Analyse de sensibilité au nombre de contextes spatiaux dans le modèle hybride	81
3.10	Synthèse des performances de l'ensemble des modèles estimés	84
3.11	Quel est le rapport entre une personne en situation d'obésité, un glissement de terrain paraglaciale en Islande, une <i>treeline</i> pyrénéenne et une piste cyclable ?	89

Liste des tableaux

3.1	Récapitulatif des bibliothèques R utilisées pour la mise en œuvre des analyses statistiques	69
3.2	Résultat des estimations pour les modèles multiniveaux vides	76
3.3	Résultat des estimations pour les modèles multiniveaux à pentes aléatoires	77
3.4	Résultats des estimations des modèles multiniveaux à constantes et pente aléatoires avec variable contextuelle (de niveau 2)	80
3.5	Résultats des modèles intégrant les communes (k=143) en macro-unités	82
3.6	Résultats des modèles aléatoire croisés combinant les contextes administratifs et les contextes GWR en niveau 2	83

Table des matières

Introduction	1
1 Les effets de contexte dans leur cadre théorique et conceptuel	7
1.1 Définition d'un contexte spatial	8
1.2 Structuration de l'espace par systèmes de contextes spatiaux	9
1.3 Le rôle de l'hétérogénéité et de la dépendance spatiales dans la structuration des contextes spatiaux	16
1.3.1 Des effets verticaux : l'hétérogénéité spatiale et les processus de long terme	16
1.3.2 Des effets horizontaux : dépendance spatiale et effets de voisinage	17
1.4 Apports de la considération des systèmes de contextes spatiaux dans les analyses géographiques	25
1.4.1 Les effets de contexte en géographie : des conciliateurs épistémologiques	25
1.4.2 L'importance des effets de contexte dans les chaînes causales	28
2 Les études contextuelles, leur modélisation statistique et leurs limites	33
2.1 Le cheminement du déterminisme au contextualisme : historique de l'étude des liens entre l'humain et son environnement	33
2.1.1 Déterminisme, possibilisme et écologie humaine	33
2.1.2 De l'écologie humaine aux effets de quartier : le poids des contextes locaux	35
2.1.3 Au-delà du quartier : l'apport récent de la géographie de la santé	36
2.2 La modélisation statistique des effets contextuels en analyse transversale : modèles contextuels, multiniveaux et spatiaux	41
2.2.1 Les modèles contextuels	42
2.2.2 Les modèles multiniveaux	42
2.2.3 Les modèles spatiaux (et locaux)	47
2.3 Typologie des erreurs associées à la modélisation statistique des effets contextuels	52
2.3.1 Erreurs géographiques	53
2.3.2 Erreurs causales	56
2.3.3 Erreurs analytiques	58

3 De la non-stationnarité spatiale à une géographie réaliste des contextes	61
3.1 Justification et présentation du protocole méthodologique combinant modèles GWR et multiniveau	62
3.1.1 Aperçu des méthodes hybrides combinant modèles spatiaux et modèles multiniveaux	63
3.1.2 Présentation de la stratégie d'analyse	64
3.1.3 Validation des modèles : minimisation de l'AIC et maximisation des effets contextuels	68
3.1.4 Outils mobilisés	69
3.2 Application du modèle hybride multiniveau-GWR à la géographie de l'obésité en région parisienne	69
3.2.1 La géographie de l'obésité reflète des effets de contexte	69
3.2.2 Présentation du jeu de données et de la cohorte NutriNet-Santé	71
3.2.3 Présentation des résultats	73
3.2.4 Robustesse du modèle aux paramétrages alternatifs	79
3.3 Implications des résultats en géographie et en santé publique	82
3.3.1 Implications en analyse spatiale : vers un affranchissement des maillages administratifs ?	83
3.3.2 Implications en santé publique : vers un ciblage géographique optimal des déterminants de santé	84
3.3.3 Discussion méthodologique : les limites intrinsèques de la régression en analyse contextuelle	86
Conclusion	89