



**HAL**  
open science

# Modéliser les inégalités dans l'espace géographique

Clémentine Cottineau

► **To cite this version:**

Clémentine Cottineau. Modéliser les inégalités dans l'espace géographique. Clémentine Cottineau; Julie Vallée. Les inégalités dans l'espace géographique, ISTE éditions, pp.187 - 219, 2022, Encyclopédie des Sciences - Domaine Géographie et Démographie, 9781789480887. 10.51926/iste.9088.ch6 . halshs-03801388

**HAL Id: halshs-03801388**

**<https://shs.hal.science/halshs-03801388v1>**

Submitted on 6 Oct 2022

**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.

# Chapitre 6

## Modéliser les inégalités dans l'espace géographique

Clémentine Cottineau<sup>1,2</sup>

<sup>1</sup> CNRS, Centre Maurice Halbwachs, Paris.

<sup>2</sup> Technische Universiteit Delft, Pays-Bas.

### 6.1. Introduction : différents formalismes de modélisation pour différents objectifs

Ni le niveau des inégalités ni la tendance de leur évolution ne sont aléatoires ou accidentels. Ils résultent de processus économiques, politiques, sociaux et spatiaux accumulés sur le temps long, alimentés par les interactions et stratégies d'acteurs multiples, notamment les institutions collectives, qu'elles soient publiques ou privées. De ce fait, il existe de multiples « explications » des inégalités, et au moins autant de façons de les détecter par l'analyse quantitative. La modélisation des inégalités, du modèle le plus abstrait au plus descriptif, peut ainsi rendre compte d'évolutions observées, illustrer l'effet (attendu ou inattendu) de telle ou telle mesure politique, voire inclure les interactions entre agents et entre les agents et leur environnement géographique dans la production et la reproduction des inégalités. Dans ce chapitre, il n'est question ni de passer en revue l'ensemble des modèles dédiés aux inégalités dans l'espace géographique, ni de produire un mode d'emploi de la modélisation des inégalités. L'ambition est plutôt de présenter, d'une part, différentes approches de modélisation (statistique, spatiale et multi-agents), et ce qu'elles apportent à différents objectifs de modélisation des inégalités ; d'autre part, il s'agit de présenter quelques exemples saillants et résultats importants issus de modèles existants dédiés aux inégalités dans l'espace géographique.

Un modèle étant défini comme une représentation simplifiée de la réalité dans un but particulier, la question se pose de savoir, comme pour d'autres modèles, « pourquoi modéliser » (Van der Leeuw 2004 ; Epstein 2008 ; Edmonds *et al.* 2019) les inégalités entre individus dans l'espace géographique. L'objectif du modèle – que ce soit l'explication, la prédiction ou l'illustration d'une théorie, par exemple – conditionne le formalisme et la précision avec lesquels le système en question est représenté, ainsi que le choix des entités et interactions présentes dans le modèle, à la manière de la généralisation en cartographie, qui correspond à l'opération de sélection et de simplification des éléments retenus pour figurer sur la carte (Cottineau *et al.* 2019). En ce qui concerne la modélisation des inégalités dans l'espace géographique plus spécifiquement, un premier aiguillage dépend du type d'inégalités étudiées : économiques, de santé, d'éducation, etc. Un second aiguillage dans l'arbre de décision dépend de la

place qu'occupent l'espace dans la description et l'explication de ces inégalités. Ainsi, alors qu'une série de modèles mathématiques s'appliquent à simplifier la description des inégalités (de revenu, notamment) pour pouvoir les cartographier, et éventuellement tirer des conclusions en termes de répartition spatiale (section 6.2), les modèles de régression multiniveaux et à pondération géographique combinent l'effet de la localisation relative et de l'appartenance géographique des individus dans l'estimation des inégalités interindividuelles (section 6.3). Dans les modèles de simulation, au contraire, il s'agit de modéliser le processus par lequel se créent les inégalités, en incluant l'espace géographique dans les règles d'action et de changement des attributs des agents (section 6.4). Il existe, bien sûr, nombre d'autres modèles, notamment certains modèles d'équilibre issus des sciences économiques (voir chapitre 4), qui nous ont paru trop spécifiques et insuffisamment spatialisés pour figurer dans ce chapitre. Nous réduisons ici la focale à la modélisation graphique, statistique et générative des inégalités dans l'espace géographique.

## **6.2. Inégalité de la distribution des ressources économiques et inégalité de sa répartition spatiale**

Nous avons défini dans l'introduction les inégalités comme l'existence de différences systématiques, notables et perçues dans la distribution des ressources sociales (richesse, santé, éducation, par exemple) entre individus, en raison de certaines de leurs caractéristiques (sexe, âge, origine ethnique, par exemple). Ces inégalités entre individus peuvent se traduire dans l'espace lorsque les individus favorisés se concentrent dans certains lieux, et les individus défavorisés dans un ensemble distinct de lieux. Toutefois, ces inégalités traduites spatialement se distinguent d'inégalités spatiales, qui concerneraient par exemple l'inégale exposition à certaines nuisances ou l'accès inégal à certaines ressources (issu de la répartition des équipements et services sur un territoire). Ces deux types d'inégalités sont évidemment liées : la présence d'équipements rares et précieux (monuments historiques, espaces bleus et verts, hôpitaux renommés) ou, au contraire, d'infrastructures connotées négativement (traitement des eaux usées, cimetière) conditionnent en effet la localisation (résidentielle, par exemple) d'individus favorisés ou défavorisés, tout en modifiant l'accès potentiel des populations aux différentes ressources. Toutefois, il semble plus judicieux de les distinguer dans l'analyse pour mieux comprendre les dynamiques à l'œuvre.

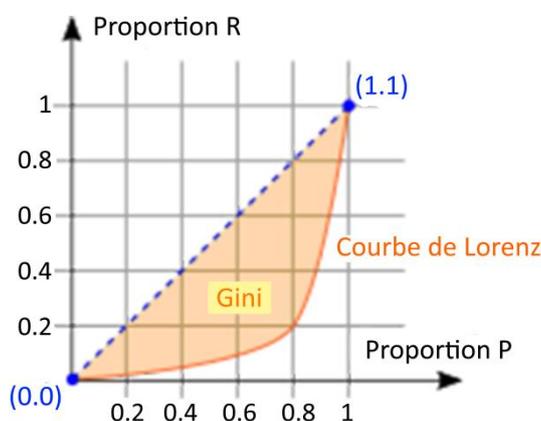
Nous abordons ainsi dans un premier temps les modèles de description des inégalités de distribution des ressources économiques (section 6.2.1), avant de détailler une méthode de description des inégalités dans l'espace, à savoir par la cartographie (section 6.2.2). Dans un troisième temps, nous évoquons les modèles d'inégalité spatiale et de discontinuité sociale dans l'espace (section 6.2.3).

### **6.2.1. Lorenz/Gini, Pareto et Theil en économie : trois modèles distincts**

Dans le domaine des inégalités économiques, et typiquement des inégalités de revenus, le modèle mathématique le plus largement mobilisé est celui de la courbe de Lorenz et son indice associé : le coefficient de Gini. Le modèle mis en œuvre par M. Lorenz en 1905 est graphique, et vise à représenter la concentration de la richesse par rapport à la concentration de la population au sein d'un périmètre donné. En indiquant en abscisse le pourcentage de la population classée par richesse croissante, et en ordonnée le pourcentage de la richesse détenue par la population correspondante, la courbe de Lorenz se présente comme une courbe convexe (ou droite) entre l'origine (0;0) et les coordonnées (1;1). En effet, 0 % de la population détient toujours 0 % de la richesse et 100 % de la population détient toujours 100 % de la richesse. Entre ces deux points connus, la convexité de la courbe indique l'ampleur de la concentration de la richesse entre un petit nombre de mains, et donc le degré d'inégalité (mathématique) de la distribution du revenu. Une courbe droite indique une distribution parfaitement égalitaire, puisqu'à chaque proportion de la population correspond une proportion équivalente de la richesse totale, tandis qu'une courbe extrêmement convexe indique une concentration très forte de la richesse. Dans l'exemple de la figure 6.1, dont la courbe est particulièrement convexe, on trouve notamment que les 80 % les

plus pauvres dans la population ne détiennent que 20 % de la richesse totale. À l'inverse, les 20 % d'individus les plus riches cumulent 80 % de la richesse.

En 1914, C. Gini retrouve la courbe de Lorenz et la rend célèbre (Alacevich, Soci 2017) en créant un indice de concentration borné entre 0 et 1, et qui correspond graphiquement à l'aire marquée en jaune sur la figure 7.1, comprise entre la courbe de Lorenz et la droite d'égalité en pointillés bleus, divisée par l'aire totale du triangle rectangle situé sous la droite d'égalité (soit 1/2). Plus l'indice de Gini est élevé, plus la concentration de la richesse est importante, et donc plus les inégalités économiques sont fortes. La modélisation des inégalités est ici unidimensionnelle. Elle ne rend pas compte de possibles asymétries dans la distribution, n'est pas décomposable et donne plus de poids aux transferts autour du mode de la distribution qu'aux transferts entre extrêmes. Ces caractéristiques font de l'indice de Gini, de l'accord général des économistes (Alacevich, Soci 2017), un indicateur très imparfait pour rendre compte des distributions empiriques de richesses et de revenus, mais incontournable en raison de sa simplicité et de son adoption encore aujourd'hui majoritaire en sciences économiques et sociales, ainsi que dans les rapports des institutions internationales sur les inégalités économiques.



**Figure 6.1.** Un modèle graphique des inégalités de richesses : la courbe de Lorenz ; R représente la richesse et P la population

Les indicateurs de concentration construits à partir de données fiscales et de patrimoine ont été remis à l'ordre du jour par T. Piketty (2001). Ils correspondent à des points particuliers de la courbe de Lorenz : par exemple, la part de richesses détenues par les 1 % les plus riches, ou encore par les 50 % les plus pauvres. Ces indices sont simples d'usage et plus intuitifs que d'autres mesures agrégées, tout en décrivant les mêmes tendances dans l'évolution des inégalités. En l'occurrence, les indices de concentration covarient positivement et significativement avec l'indice de Gini, d'Atkinson<sup>42</sup> et le rapport 90:10, c'est-à-dire le rapport entre la valeur du dernier et du premier décile de richesse ou de revenu (Leigh 2007). En revanche, l'utilisation de la part des revenus captés par le top 1 % est sujette à caution, selon A. Reynolds (2012), parce que ces revenus sont élastiques (à savoir sensibles à la conjoncture économique) et qu'ils sont calculés avant redistribution<sup>43</sup>. Leur part dans le revenu global décroît donc fortement lors de crises, ce qui ne réduit pas pour autant les inégalités structurelles.

<sup>42</sup> L'indice d'Atkinson (1970) est un indice d'inégalité normatif, c'est-à-dire basé sur une fonction d'utilité représentant le degré d'aversion de la société pour les inégalités. Cette fonction permet de donner un poids différent à l'effet d'un transfert de richesses entre riches et pauvres sur l'indice d'inégalité, selon la direction du transfert. L'indice d'Atkinson varie entre 0 et 1, et s'écrit :  $A=1-(y_e/\mu)$ , avec  $y_i$  = le revenu d'un individu  $i$  ;  $\mu = (\sum_i y_i) / n$  ; et  $y_e = y \mid [nU(y) = \sum_i U(y_i)]$ .

<sup>43</sup> Cet argument peut s'appliquer aux autres indices, même si les indices synthétiques et les rapports  $y$  sont moins sensibles.

Alors que M. Lorenz et C. Gini calculent la concentration de la richesse par rapport à un modèle de référence égalitaire (la droite d'équivalence entre % de population et % de richesse), le modèle de Pareto, qui les précède, est tout autre. Issu de la généralisation mathématique d'une série d'observations empiriques, le modèle de Pareto pour décrire la distribution des richesses est celui d'une loi de puissance. En effet, en collectant les chiffres de la richesse individuelle dans différents pays du monde, V. Pareto montre qu'il existe une relation mathématique décroissante et scalaire entre la probabilité  $P$  pour un individu de détenir une richesse  $X$  supérieure à  $x$ , et le rapport entre ce niveau  $x$  et le niveau minimum de richesse  $x_m$ , tel que :  $P(X > x) = (x_m/x)^\alpha$ , avec  $\alpha$  un paramètre d'échelle interprété par Pareto comme indicateur du niveau d'inégalité. Ici, la progression géométrique des niveaux de richesse dans la distribution est donc le modèle empirique utilisé, très loin de la référence au modèle égalitaire de Lorenz et Gini. Les distributions parétiennes sont seulement comparées entre elles de manière synthétique avec  $\alpha$  pour déterminer un ordre entre niveaux d'inégalité (à la manière de la distribution rang-taille dans les études urbaines).

L'indice de Theil, quant à lui, applique le modèle physique de l'entropie et la théorie de l'information sur la distribution des grandeurs économiques. Cette mesure, suffisamment large pour refléter à la fois inégalité, ségrégation et redondance de l'information, est principalement mobilisée pour obtenir une décomposition de l'inégalité entre ce qui relève de variations entre groupes et ce qui relève de variations au sein de ces groupes. Ainsi, les analyses montrent que l'appartenance nationale « explique » environ deux tiers des inégalités entre individus dans le monde, mais que cette part diminue récemment avec l'enrichissement considérable des géants comme la Chine, le Brésil et l'Inde (Sala-i-Martin 2006).

Lorsque l'on s'intéresse non plus aux inégalités économiques issues de la distribution d'une unique variable (richesse ou revenu) au sein d'une population, mais aux inégalités sociales, de santé ou d'éducation, par exemple, il s'agit alors de mettre deux variables en relation. La différence de moyennes ou d'indices de concentration entre groupes est souvent mobilisée pour illustrer les écarts de ressources ou d'accès par groupe social ou démographique, en utilisant les tests statistiques de significativité tels que les t-tests. Par exemple, N. Kakwani *et al.* (1997) mesurent des inégalités de santé en comparant le niveau de santé ressenti selon le niveau de revenus aux Pays-Bas en 1980-1981. Ils observent par l'étude des moyennes et des moyennes standardisées que les individus déclarent une meilleure santé à mesure que leur décile de revenus augmente. En se basant sur un t-test entre des indices de concentration similaires à l'indice de Gini (croisant proportion cumulée de revenu et d'un indicateur de santé ordinal), ils confirment l'existence d'inégalités de santé à l'avantage des plus riches, même lorsque la structure démographique de l'échantillon est prise en compte, et ce d'autant plus que la variable de santé est subjective.

Développés principalement par des géographes, statisticiens et sociologues dans le but de décrire la ségrégation sociospatiale (Apparicio 2000), les indices d'inégalité spatiale peuvent être rangés en trois groupes.

Dans le premier groupe, l'on trouve des versions spatialisées des indices d'inégalité présentés plus haut, tels que l'indice de Gini ou de Theil, pour lesquels l'on remplace la distribution de richesse entre groupes de population par la distribution de différents groupes sociaux entre unités géographiques.

Dans le second groupe, l'on trouve des variantes de l'indice de dissimilarité de Duncan et Duncan (1955), qui exprime la part d'individus devant changer d'unité géographique (en déménageant dans un autre quartier, par exemple) pour passer de la situation observée à une situation d'équirépartition spatiale des groupes sociaux considérés dans un espace donné (des groupes de différents niveaux d'éducation dans une ville, par exemple). Cet indicateur d'inégalité spatiale est la mesure la plus simple et la plus fréquemment mise en œuvre, notamment pour mesurer la ségrégation ethno-raciale dans la littérature américaine.

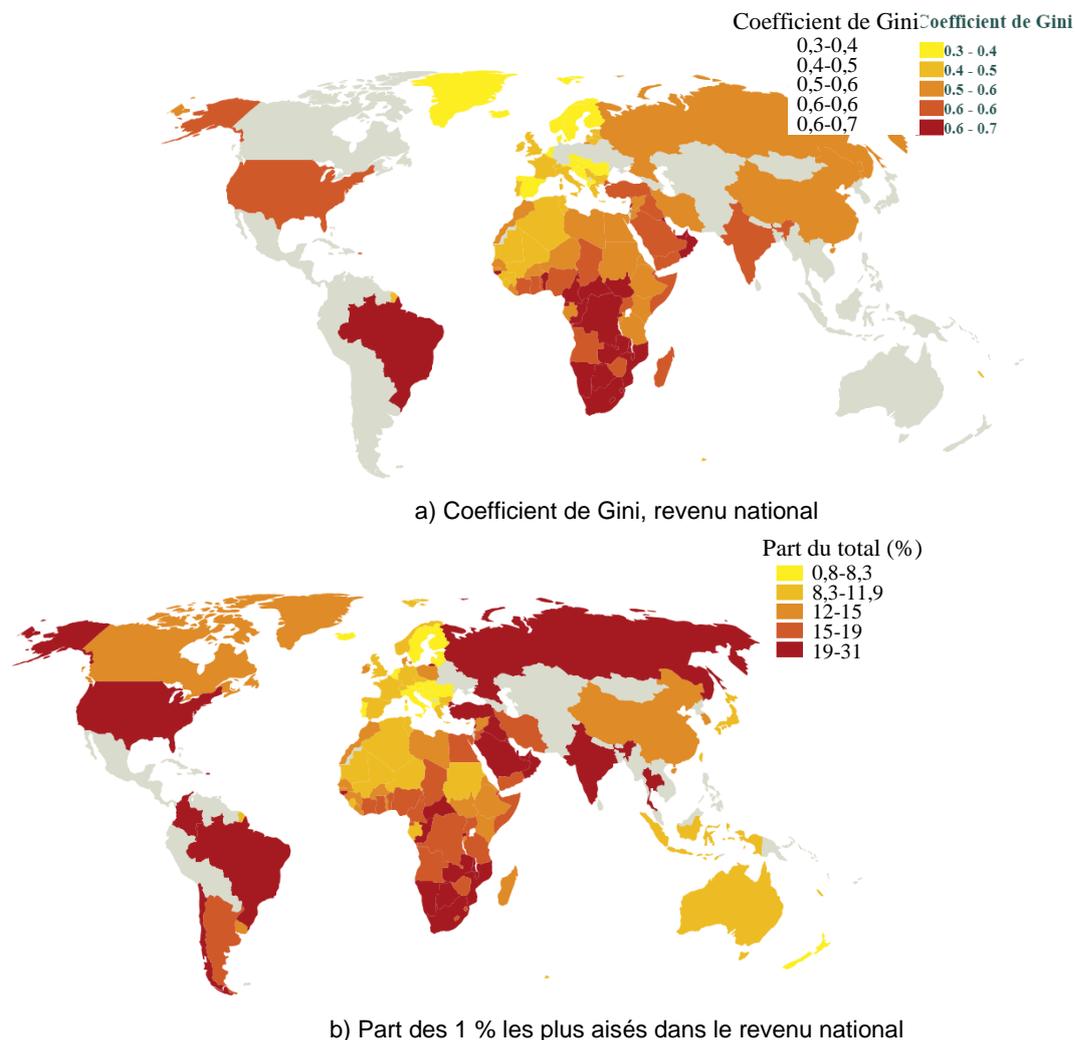
Dans le dernier groupe, l'on trouve des indices développés par des géographes tels que R. Morrill (1991), D. Wong (1993), S. Reardon et D. O'Sullivan (2004), qui incluent des « paramètres spatiaux comme la distance séparant les unités spatiales, la superficie » (Apparicio 2000, §17) ou la contiguïté des unités spatiales dans la mesure des inégalités. Bien que plus compliqués à calculer et plus demandeurs en données et analyses spatiales, ces indicateurs permettent de rendre compte de manière fidèle de la structure spatiale des inégalités.

#### Encadré 6.1. *Les indices d'inégalité spatiale*

Tous les indices présentés jusqu'ici sont a-spatiaux, mais peuvent être utilisés pour comparer des unités géographiques entre elles. Ainsi, les inégalités entre régions et entre pays sont le plus souvent analysées en comparant les valeurs de leurs indices (George 1981), éventuellement dans le temps, ou en décomposant l'indice de Theil mondial en deux composantes : l'inégalité internationale et l'inégalité intranationale (Sala-i-Martin 2006). La prise en compte de l'espace géographique dans les indices d'inégalité peut se faire de deux façons : soit par l'usage d'indices spatiaux (encadré 6.1), soit par la cartographie des inégalités et leur modélisation graphique (section 6.2.2.).

#### **6.2.2. Cartographie des inégalités socio-économiques et modélisation graphique**

La forme la plus basique de spatialisation des inégalités interindividuelles consiste à cartographier les valeurs agrégées (moyenne, médiane, somme, par exemple) d'une ressource sociale et à considérer, à la manière des inégalités sociales, les unités géographiques comme des groupes dont les différences (significatives) de valeurs indiquent la présence d'inégalités liées à la localisation. L'on trouve des exemples de cette forme de représentation dans la plupart des atlas dédiés aux inégalités (par exemple, Le Bras 2014), qu'elles soient économiques, genrées ou de santé. Toutefois, eu égard à la nécessité de discrétiser et au choix de discrétisation pour les variables quantitatives, l'impression offerte par les seules plages de couleur sur la carte peut s'avérer trompeuse ou peu informative sur le niveau et la dispersion des valeurs représentées (Bell *et al.* 2019). Une seconde option de spatialisation des inégalités consiste à cartographier non pas les valeurs agrégées, mais la valeur d'un indice d'inégalité de la ressource sociale en question, son profil de distribution ou le résultat d'une typologie au sein des unités géographiques représentée. Ainsi en est-il, par exemple, de la cartographie mondiale des indices de Gini nationaux ou de la part des 1 % les plus fortunés (figure 6.2), ou des cartographies sociorésidentielles (François, Ribardièrre 2004). Dans ce dernier cas, l'usage de modèles graphiques d'inégalité spatiale est mobilisé explicitement pour interpréter les cartes. Notamment, les auteurs décrivent la spécialisation des communes franciliennes dans la résidence de ménages plus ou moins aisés selon le modèle sectoriel, en identifiant une « section » de communes aisées allant du centre de Paris au sud-ouest de la région, et plusieurs sections spécialisées dans la résidence de ménages pauvres en petite couronne. Le modèle centre-périphérie est utilisé pour distinguer les communes spécialisées dans la résidence des classes moyennes (zones péricentrales), et la périphérie pauvre des marges orientales de Seine-et-Marne. Les discontinuités locales y sont aussi mises en valeur en changeant l'échelle d'analyse de la commune à l'IRIS (îlots regroupés pour l'information statistique, c'est-à-dire une maille territoriale infracommunale homogène regroupant environ 2 000 habitants).



**Figure 6.2.** Cartographie mondiale des indices d'inégalité économique (source : World Inequality Database, dernière année disponible en 2020<sup>44</sup>)

Comme le suggère cet exemple, la description des configurations spatiales se prête à l'usage de modèles spatiaux des inégalités, ainsi qu'à l'analyse des discontinuités spatiales. Nous abordons ci-après les modèles classiques d'organisation des sociétés urbaines, popularisés par les sociologues et géographes américains du siècle dernier pour décrire le processus de ségrégation, avant de nous intéresser à l'identification et à l'interprétation des discontinuités en zones homogènes.

Le modèle zonal, ou concentrique, décrit la distribution spatiale d'un phénomène comme un rapport de distance à un centre. Dans la version initiale de R. Park et E. Burgess (1925), les zones concentriques représentent des aires relativement homogènes en termes de population (origine sociale et migratoire) à Chicago. Ce modèle est mobilisé pour représenter la structure urbaine dans laquelle circulent les migrants, du centre vers la périphérie à mesure de leur intégration dans la société américaine et chicagolaise. Il a été réutilisé depuis pour décrire la distribution spatiale des fonctions urbaines (production au centre vs résidence en périphérie) et des formes urbaines (immeubles de grande hauteur au centre, pavillons en périphérie). Le modèle sectoriel, attribué à H. Hoyt (1939),

<sup>44</sup> wid.world/fr.

combine une dimension axiale avec le modèle concentrique pour représenter la dépendance historique de certains secteurs urbains à la présence d'industries ou de résidences fortunées le long d'un axe de transport (le chemin de fer, initialement) au cours du développement d'une ville. Le modèle polynucléaire proposé par C. Harris et E. Ullman (1945), en suspendant l'hypothèse d'isotropie du développement urbain, permet d'obtenir des configurations plus complexes et variables que le modèle initial de Burgess, avec notamment des centres secondaires et des zones de tailles variables.

L'apport de l'écologie factorielle par W. Bell et E. Shevky (1955) a ensuite été de combiner ces modèles pour représenter différentes dimensions d'organisation spatiale de la population dans une ville. Le modèle concentrique serait une représentation des différentes étapes du cycle de vie, les jeunes adultes vivant au plus proche du centre et de l'activité, les familles s'éloignant du centre à mesure que le nombre d'enfants s'accroît, tandis que les personnes âgées retournent vers des localisations plus centrales lorsque leur foyer se réduit. Le modèle sectoriel correspondrait à la spécialisation économique, tandis que le modèle polynucléaire décrirait plutôt des concentrations de populations sur base ethnique. Cette approche a été opérationnalisée par B. Berry et J. Kasarda (1977) sur les aires métropolitaines au moyen d'analyses factorielles, permettant en plus de hiérarchiser l'importance de ces facteurs d'organisation urbaine dans les différences villes.

En termes d'inégalités spatiales, ces modèles visent à montrer qu'au-delà de la complexité apparente des villes et du vécu individuel, il existe des configurations structurelles récurrentes dans l'organisation de la ville. La proximité entre industrie et résidences de ménages défavorisés, par exemple, donne à voir une inégalité d'accès à un environnement vert ainsi qu'à l'ensemble de la ville (les infrastructures industrielles et de transport créant souvent des barrières à la circulation). Ces modèles se basent sur la présence de discontinuités entre zones mais les sous-entendent implicitement. Au contraire, certains travaux (Grasland *et al.* 1997 ; François 2002 ; Roberto, Hwang 2015) mettent la notion de discontinuité au cœur de l'analyse des distributions spatiales de la société. Cette notion permet de :

« appréhender à la fois l'intensité de la séparation qui se noue dans le voisinage et les facteurs de division sociale de l'espace, en raison des liens qui les articulent. Les discontinuités, lorsqu'elles sont définies de manière multidimensionnelle (limites administratives et politiques, ruptures morphologiques ou d'accessibilité, catégories cognitives) et que sont prises en compte leurs trajectoires dans l'espace et dans le temps, deviennent donc une riche notion pour lire l'urbain et la ségrégation dans les villes [...] Au moins deux configurations théoriques peuvent être identifiées : la polarisation et la fragmentation. La polarisation renvoie à des partitions simples de la ville où de vastes zones homogènes sont séparées par des gradients. Poussée à l'extrême, cette situation aboutit à une structure duale telle qu'elle est décrite pour de nombreuses grandes villes et transposée sans nuance aux autres villes. La fragmentation correspond à une structure spatiale plus complexe où les ensembles sont fragmentés et dispersés dans les espaces urbains. Or les mesures classiques de la ségrégation ne peuvent saisir ces configurations spatiales pourtant déterminantes par les effets de contexte qu'elles suscitent. » (Duroudier 2014, p. 136-7)

Ces modèles spatiaux (polarisation, sectorisation, fragmentation, etc.) des inégalités permettent ainsi de décrire de manière simplifiée l'organisation économique des sociétés dans l'espace, d'identifier des motifs récurrents – le rapport centre-périphérie dans les villes, un gradient Nord-Sud à l'échelle internationale entre pays égalitaires et inégalitaires –, ainsi que des discontinuités locales fortes (entre régions, quartiers, ou blocs résidentiels).

Bien qu'utiles pour décrire la distribution des ressources dans une population et dans l'espace géographique, l'analyse des indicateurs d'inégalités (a-spatiaux) et les modèles spatiaux de leur

répartition ne rendent pas compte de l'effet de localisation des individus sur leurs caractéristiques économiques et sociales. Une façon de modéliser ces effets s'appuie sur les modèles statistiques de régression multiniveaux et à pondération géographique (GWR).

### 6.3. Modèles de régression statistique : estimer les effets de la position géographique sur les inégalités

Nous présentons en détail le fonctionnement de chaque modèle de régression avant de discuter son apport et ses limites quant à l'étude des inégalités dans l'espace géographique.

#### 6.3.1. Fonctionnement des modèles

Deux familles de modèles statistiques sont évoquées ci-dessous : les modèles multiniveaux et les modèles à pondération géographique. Ce sont les plus adaptés à la prise en compte de l'organisation géographique des individus dans l'explication statistique de leurs inégalités.

##### 6.3.1.1. Modèles multiniveaux

Les modèles multiniveaux sont des modèles de régression appliqués à des données hiérarchiques. Ils sont de plus en plus utilisés en sciences sociales et de l'éducation (Goldstein 1987). Contrairement aux modèles de régression classiques de types moindres carrées ordinaires (*Ordinary Least Squares* ou OLS), les modèles multiniveaux comportent deux composantes de la variance, l'un généralement au niveau de l'individu (niveau 1) et l'autre à un niveau supérieur (niveau 2) dans lequel sont contenus les individus de niveau 1. Ces modèles permettent de combiner des variables explicatives individuelles (niveau 1) avec des variables de niveau supérieur (niveau 2), dont l'effet s'applique aux individus appartenant à un même groupe de niveau supérieur. Ce niveau supérieur correspond classiquement à un niveau géographique (le quartier, la commune ou le pays) ou à un regroupement institutionnel (par exemple, la classe ou l'établissement scolaire dans les modèles d'inégalités scolaires). L'équation d'un modèle multiniveaux s'écrit :

$$y_{ij} = \beta_0 + (\beta_1 + u_{1j}) * x_{ij} + u_{0j} + \varepsilon_{ij}$$

Où  $i$  se réfère à un individu de niveau 1 (par exemple, un élève),  $j$  à l'ensemble de niveau 2 auquel il appartient (par exemple, son établissement scolaire),  $y$  représente la caractéristique dont on cherche à expliquer la distribution (par exemple, le revenu, le niveau de réussite scolaire, l'espérance de vie) et  $x$  représente le vecteur des variables explicatives de niveau inférieur introduites dans le modèle (par exemple, l'âge de l'élève, le revenu et le niveau d'éducation de ses parents).  $\beta_0$  correspond à l'intersection à l'origine (*intercept*),  $u_{0j}$  est le résidu de niveau 2 pour l'ensemble  $j$  et  $\varepsilon_{ij}$  le résidu de niveau 1 pour l'individu  $i$ , ces deux types de résidus étant non corrélés et suivant chacun une distribution normale. Enfin, au vecteur de coefficients  $\beta_1$  peut s'associer un vecteur  $u_{1j}$  pour différencier la pente de la régression en fonction de l'appartenance de l'individu  $i$  à l'ensemble  $j$ .

Selon l'implémentation retenue, un modèle multiniveaux peut donc permettre de faire varier les *intercepts* de la régression en fonction des groupes  $j$  ( $u_{0j} \neq 0$ ), de faire varier les pentes des droites de régression en fonction des groupes  $j$  ( $u_{1j} \neq 0$ ), ou les deux à la fois. Appliqués à des données politiques, scolaires et de santé, ces modèles peuvent révéler des effets du lieu de résidence sur des caractéristiques importantes des individus, notamment en calculant la corrélation intraclasse (ICC), c'est-à-dire la part de variance interindividuelle imputable à l'appartenance au niveau supérieur. En effet, lorsque les résidus de groupes géographiques sont significativement différents, ils permettent d'isoler un effet d'appartenance (géographique) dans la variation des niveaux de performance scolaire ou de santé, suggérant ainsi que, les variables individuelles étant contrôlées par ailleurs, le lieu de résidence joue un rôle à part dans la vie des individus, par effet de composition, l'action des pairs ou

l'environnement physique et social plus général (voir chapitre 1). Des exemples de ce type de résultats sont proposés dans la section 6.3.2.

Une variante des modèles multiniveaux estime des interactions entre niveaux. Ces effets d'interaction interniveaux (*Cross level interaction effects*) permettent de mesurer l'influence de variables de niveau supérieur dans la variation des pentes de régression entre groupes de niveau supérieur. L'équation d'un modèle multiniveaux à interaction entre niveaux s'écrit :

$$y_{ij} = \beta_{0j} + (\beta_{1j} + u_{1j}) * x_{ij} + u_{0j} + \varepsilon_{ij}$$

avec:

$$\beta_{0j} = \gamma_{00} + \gamma_{01} * w_j$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11} * w_j$$

Où  $w_j$  représente le vecteur des variables explicatives de niveau supérieur  $j$  (par exemple, la taille et la spécialisation de l'établissement dans l'exemple des inégalités scolaires).  $\gamma_{00}$  correspond à l'intersection à l'origine (*intercept*) commune à toutes les observations du modèle,  $\gamma_{01}$  est le changement d'*intercept* du groupe  $j$  induit par une variation de la variable  $w$ .  $\gamma_{10}$  correspond au changement de pente de la variable  $w$  induite par une variation d'*intercept* au niveau du groupe. Enfin,  $\gamma_{11}$  est l'effet d'interaction interniveaux, c'est-à-dire le changement de pente de la variable individuelle induite par un changement de pente de la variable de niveau supérieur. À l'instar des interactions entre variables d'un même niveau, les interactions inter-niveaux permettent d'identifier des effets de variables dépendant d'autres variables. Dans le domaine des inégalités, ces interactions peuvent se présenter par exemple sous la forme de relations variables entre niveau d'éducation pour les femmes et pour les hommes, pour les résidents de zones à fort taux de chômage et pour les autres. Les modèles multiniveaux avec interaction interniveaux – assez proches des modèles à pente variable – nécessitent donc de travailler sur des échantillons suffisamment larges, et dont la proportion de variance totale imputable aux différences entre entités de niveau supérieur (ICC) est suffisante pour obtenir des effets significatifs (Mathieu *et al.* 2012).

#### 6.3.1.2. Modèles à pondération géographique

Les modèles à pondération géographique, ou GWR (*Geographically Weighted Regression*), sont des modèles de régression appliqués à des données localisées, au sein desquelles l'on suppose une autocorrélation spatiale. « GWR est une technique relativement simple qui étend le champ de la régression traditionnelle [...] en permettant des variations locales des taux de variation de sorte que les coefficients du modèle sont spécifiques à une localisation plutôt que des estimations globales » (Brunsdon *et al.* 1996, p. 284)<sup>45</sup>. La valeur d'une variable en un point donné de l'espace est donc supposée varier systématiquement avec les caractéristiques du point donné, ainsi qu'avec les caractéristiques des points voisins dans l'espace géographique. Autrement dit, ce qu'il se passe à un endroit dépend des caractéristiques des espaces alentour. La mise en œuvre technique des modèles à pondération géographique vise précisément à délimiter l'espace « alentour » et à quantifier son poids dans l'explication. L'équation d'un modèle GWR s'écrit :

$$y_i = \beta_{i0} + \beta(i) * x_i + \varepsilon_i$$

Où  $y_i$  représente la caractéristique  $y$  dont l'on cherche à expliquer la distribution (par exemple, le revenu, le niveau de réussite scolaire, l'espérance de vie) d'un individu localisé en un point  $i$ .  $x_i$

---

<sup>45</sup> « GWR is a relatively simple technique that extends the traditional regression framework [...] by allowing local variations in rates of change so that the coefficients in the model rather than being global estimates are specific to a location. »

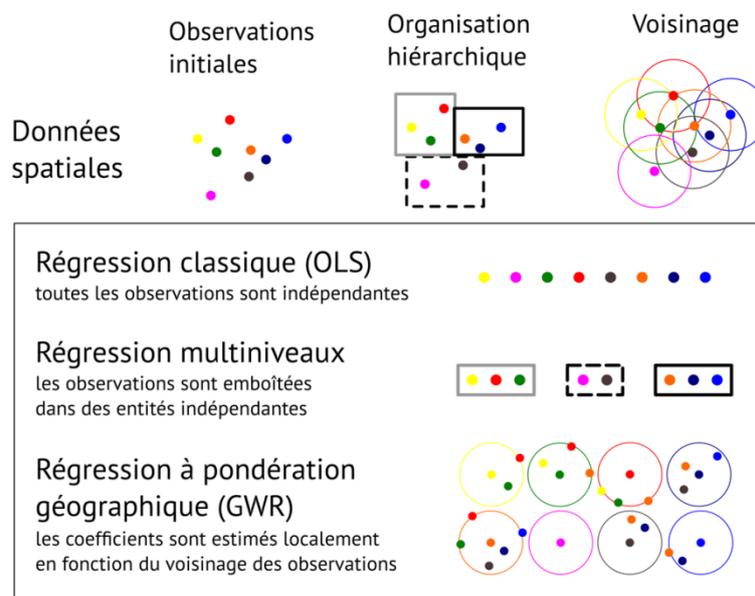
représente le vecteur des variables explicatives,  $\beta_{i0}$  l'intersection à l'origine pour le point  $i$  et  $\varepsilon_i$  le résidu au point  $i$  (les résidus étant indépendants et distribués selon une loi normale de moyenne nulle). L'élément central de ce modèle est la fonction continue de voisinage  $\beta(i)$ . Cette fonction détermine quelles observations sont incluses dans la régression pour chaque point. La plupart des applications de ce modèle ont recours à une fonction de pondération décroissante avec la distance (*distance decay*) autour de  $i$  pour sélectionner ces observations, mais d'autres fonctions sont possibles (Brunsdon *et al.* 1996). Elles ont en commun de définir des matrices de poids entre individus localisés. Lorsque la fonction restreint le voisinage aux observations très proches, elle engendre une forte variabilité des coefficients estimés. Au contraire, une fonction qui inclurait systématiquement toutes les observations ne se différencierait pas d'un modèle de régression par les moindres carrés ordinaires (OLS) sans composante spatiale.

Un modèle GWR va donc produire autant d'estimations d'*intercepts* et de coefficients associés aux variables explicatives qu'il y a d'individus localisés  $i$ . Ce modèle permet ainsi d'analyser les cas dans lesquels l'autocorrélation spatiale des observations permet de mieux expliquer les motifs inégaux observés. Ils sont particulièrement utiles pour révéler les situations dans lesquelles le sens et l'intensité de la relation entre deux variables varient en différentes régions de l'espace. De plus, ces modèles à fonction de voisinage continue permettent d'extrapoler la valeur des coefficients associés aux variables explicatives en des points de l'espace non couverts par des observations empiriques (Brunsdon *et al.* 1996). « Pour calibrer un modèle GWR en chaque localisation, les données sont "empruntées" aux localisations voisines et pondérées en fonction de la distance de chaque localisation voisine au point régressé. [...] De ce fait, GWR n'identifie pas seulement l'hétérogénéité spatiale dans les processus mais tire aussi parti de la dépendance spatiale dans les données - combinant ainsi les deux caractéristiques principales de l'analyse spatiale. » (Fotheringham *et al.* 2017, p. 1248)<sup>46</sup>. Leur mise en œuvre est toutefois plus coûteuse (en calcul, en choix<sup>47</sup>, en données et en interprétation), mais se justifie dans les situations complexes. Les principes et différences entre les modèles classique, multiniveaux et à pondération géographique sont synthétisés dans la figure 6.3.

---

<sup>46</sup> « To calibrate a GWR model at any one location, data are "borrowed" from nearby locations and weighted according to the distance each nearby location is from the regression point. [...] Hence, not only does GWR identify spatial heterogeneity in processes but it also takes advantage of the spatial dependence in data—so tying together the two main distinguishing features of spatial analysis. »

<sup>47</sup> « GWR's simplicity fuels its popularity, which is reflected by its implementation in a number of software packages including the ESRI ArcGIS suite of tools, five R packages: spgwr (Bivand *et al.* 2013), gwrr (Wheeler 2013), GWmodel (Lu *et al.* 2014; Gollini *et al.* 2015), McSpatial (McMillen 2013) and lctools (Kalogirou 2019), two Python packages: PySal (Rey, Anselin 2010) and mgwr (Oshan *et al.* 2019) and standalone implementations such as GWR3 (Charlton *et al.* 2003), GWR4 (Nakaya 2015) and MGWR 1.0 (Li *et al.* 2019). Each software package has a standard GWR option complemented by a variety of alternative GWR forms and associated tools. No single package provides a fully comprehensive choice to the user although the GWmodel package comes closest. » (Comber *et al.* 2020, p. 3)



**Figure 6.3.** Organisation des données et principes de la régression classique, multiniveaux et à pondération géographique

Une variante multiscale des modèles GWR a été proposée par Yang (2014). La méthode MGWR permet d'associer une échelle différente à chaque variable explicative, en faisant varier les paramètres de la fonction de voisinage (étendant ou réduisant ainsi l'aire d'influence des voisins sur l'estimation de la grandeur régressée localement) pour chaque variable. « C'est-à-dire, l'amplitude indiquant l'aire d'emprunt des données peut varier selon les surfaces de paramètre » (Fotheringham *et al.* 2017, p. 1249)<sup>48</sup>. L'équation d'un modèle MGWR s'écrit :

$$y_i = \beta_{i0} + \beta_{i,bwj} * x_{ij} + \varepsilon_i$$

Ou  $bw_j$  représente le paramètre d'amplitude (*bandwidth*) de la fonction de voisinage associée à la variable explicative  $j$ , et  $\beta_{i,bwj}$  la fonction de voisinage paramétrée pour l'observation localisée  $i$ . Ce modèle met en œuvre plus de degrés de liberté et est plus lent à calibrer (cinq à dix fois plus dans les simulations simples de Fotheringham *et al.* 2017). Il permet cependant d'inclure dans une même régression des variables représentant des processus agissant à différentes échelles, en estimant l'amplitude spatiale optimale pour chaque variable explicative. Afin de choisir la régression la plus adaptée à chaque situation, Comber *et al.* (2020) recommandent ainsi de démarrer l'analyse d'un jeu de données spatiales par une régression linéaire simple (OLS), puis d'utiliser un modèle à pondération géographique multiscale (MGWR). Si les résidus de cette régression apparaissent non autocorrélés spatialement, la régression linéaire simple suffit. Si les amplitudes (*bandwidth*) estimées pour chaque variable sont similaires, alors un modèle à pondération géographique simple (GWR) sera mieux adapté. En cas de résidus autocorrélés spatialement et d'amplitudes variées, alors MGWR est le formalisme qui permettra d'obtenir les meilleurs résultats.

### 6.3.2. Résultats importants de l'analyse des inégalités par les modèles statistiques

En utilisant les données de la célèbre enquête américaine *Equality of Educational Opportunity*, issue du « rapport Coleman » de 1966, Borman et Dowling (2010) montrent, à l'aide d'un modèle hiérarchique

<sup>48</sup> « That is, the bandwidths indicating the data-borrowing range can vary across parameter surfaces. »

dans lequel le niveau 1 correspond à l'élève et le niveau 2 à l'établissement, que 40 % de la variance des résultats scolaires des élèves peuvent s'expliquer par des différences entre établissements. La composition ethno-raciale (en % d'Afro-Américains) de l'établissement a un effet 1,3 fois plus important que le fait d'être Afro-Américain ou non sur les résultats scolaires, tandis que les ressources financières moyennes des parents d'élèves de l'école ont une importance environ trois fois supérieure à celle des ressources des parents de l'élève lui-même dans la prédiction de sa réussite scolaire individuelle. La fréquentation d'une école plutôt que d'une autre se faisant généralement sur la base de la localisation résidentielle (voir chapitre 1), ces résultats suggèrent que les choix résidentiels des ménages aisés (tout comme l'absence de choix des ménages les plus pauvres) influencent significativement le niveau des inégalités sociales et ethniques dans le domaine de l'éducation aux États-Unis.

Dans un protocole très complet incluant 400 000 élèves du système public anglais en 2006, les différents établissements fréquentés au cours de leur scolarité et les quartiers dans lesquels se situent les écoles, Leckie (2008) met en œuvre un modèle multiniveaux à pentes variables (*random slopes*). Il révèle que l'intensité de la relation négative entre mobilité scolaire et résultats aux examens terminaux dépend fortement du moment où ces changements d'école ont lieu, et de leur nature (obligatoire entre primaire et secondaire, ou non obligatoire). En l'occurrence, ce sont principalement les élèves changeant d'école en dehors de la transition primaire-secondaire, en dehors des vacances d'été ou sans changer de résidence qui progressent moins vite que les autres. « Prendre la mobilité des élèves en compte révèle aussi le fait que les écoles et les quartiers sont plus importants que ne le montraient les analyses précédentes » (Leckie 2008, p. 537)<sup>49</sup>. Les corrélations entre pentes des droites de régression au niveau supérieur permettent quant à elles de nuancer les résultats synthétiques utilisés pour évaluer ou classer les écoles, que ce soit en termes de résultats académiques ou de réduction des inégalités (sociales, de genre, de besoin) entre élèves. En effet, il montre que la performance d'un établissement n'est pas uniforme pour tous les types de publics scolaires. En l'occurrence, la corrélation entre l'« effet-établissement » des filles non défavorisées ayant de bons résultats scolaires et l'« effet-établissement » des garçons défavorisés ayant de faibles résultats scolaires n'est que de 0,22.

En spatialisant plus explicitement leur usage des modèles multiniveaux, Roscigno *et al.* (2006) montrent, à l'aide d'un modèle multiniveaux à pentes fixes, que les élèves vivant en centre-ville et dans les zones rurales des États-Unis ont un niveau de performance scolaire systématiquement plus bas, et une probabilité d'abandonner leurs études systématiquement plus importante que leurs compatriotes résidant dans des zones périurbaines, à caractéristiques individuelles égales. Les auteurs attribuent ces différences géographiques à des déficits d'investissement scolaire de la part des familles et des écoles manquant de ressources économiques dans les centres-villes et les campagnes, mais aussi aux orientations politiques et culturelles véhiculées par les acteurs politiques et économiques présents, qui seraient moins propices à la réussite scolaire de l'ensemble des élèves.

Dans le domaine de la santé, Diez-Roux *et al.* (2000, p. 684) montrent que les hauts niveaux d'inégalité dans les États américains sont préjudiciables aux individus aux plus petits revenus vis-à-vis de leur risque de pression artérielle et d'obésité (indice BMI), mais négativement et faiblement associés à ces risques pour les individus disposant de hauts revenus. En d'autres termes, toutes choses égales par ailleurs, les inégalités économiques mesurées dans les différents États renforcent les inégalités de santé, avec des individus pauvres plus touchés par les risques de sédentarité, d'obésité et de tension dans les États plus inégalitaires, et au contraire, des individus riches moins à risque dans ces mêmes États. Ces résultats renforcent les conclusions de Wilkinson et Pickett (2009), qui concluent à l'aide de modèles multiniveaux simples que les revenus individuels n'expliquent pas toutes les différences de santé entre individus, mais que les inégalités économiques affectent (négativement) la santé de l'ensemble des communautés géographiques concernées (voir introduction). En associant le lieu de

---

<sup>49</sup> « Accounting for pupil mobility also reveals that schools and neighbourhoods are more important than shown by previous analysis. »

résidence à l'espace pratiqué au quotidien par plus de 1 500 Franciliennes, Vallée *et al.* (2010) montrent à l'aide d'un modèle multiniveaux à effets d'interaction que les femmes résidant dans des quartiers avec des caractéristiques ne favorisant pas les recours aux soins (faibles revenus moyens et faible densité de praticiens, par exemple) ont tendance à négliger plus systématiquement le dépistage du cancer du col de l'utérus que les femmes des autres quartiers, et ce d'autant plus que leur espace quotidien d'activité se limite à leur seul quartier de résidence.

Enfin, dans un article traitant de l'incidence de la première vague du Covid-19 aux États-Unis, Mollalo *et al.* (2020) exploitent les résultats de modèles GWR et MGWR pour montrer que la relation positive entre le taux d'incidence du COVID-19 et le niveau d'inégalité dans la population dans les *counties* américains se manifeste principalement « dans la zone du tri-state (i.e. les États de New York, du Connecticut, et du New Jersey), dans le Massachusetts, et dans certaines parties de l'Ouest des États-Unis, en particulier dans le Nevada, l'Idaho et l'Utah » (Mollalo *et al.* 2020, p. 5)<sup>50</sup>, tandis que la relation est plus ténue, voire nulle, dans le reste du pays. Ils obtiennent le même résultat pour le niveau de revenu médian des ménages, qui n'est fortement positivement corrélé à une plus grande prévalence du COVID-19 que dans ces zones du pays.

### **6.3.3. Limites de l'analyse des inégalités dans l'espace géographique par les modèles statistiques**

Malgré le type de résultats forts que produit la modélisation statistique des inégalités dans l'espace géographique, il existe un certain nombre de critiques envers l'utilisation de ces méthodes pour mettre en évidence les effets du contexte spatial et géographique sur les inégalités scolaires ou de santé entre individus.

En premier lieu, même si la modélisation multiniveaux permet de prendre en compte l'appartenance commune de certains individus à des quartiers, des régions et des établissements scolaires, la régression statistique reste un cadre de modélisation qui postule l'indépendance des observations considérées à un niveau donné (figure 6.3). Ainsi, même lorsque l'aspect hiérarchique des données est pris en compte, les unités de niveau supérieur restent considérées comme indépendantes. Notamment, leur organisation spatiale (la proximité entre deux quartiers, par exemple) reste exclue du modèle. La modélisation à pondération géographique multiscalaire peut compenser cette faiblesse lorsque les résidus sont effectivement autocorrélés. Pour ce type de modélisation, en revanche, le coût d'implémentation est parfois prohibitif (Fotheringham *et al.* 2017), tandis que le niveau d'expertise requis pour évaluer le modèle et interpréter ses résultats conduit certains auteurs à publier des analyses incomplètes, inadaptées ou erronées (Comber *et al.* 2020).

Deuxièmement, il existe une forte endogénéité entre le niveau supérieur et le niveau inférieur, qui n'est pas prise en compte par le modèle (Manski 1993). « On considère souvent les caractéristiques individuelles comme la propriété des individus, alors que celles-ci sont également déterminées par leur lieu de résidence. De même, le contexte dont les caractéristiques peuvent être mesurées à un moment donné, dans un territoire donné, est fortement associé à la composition de la population qui y réside. Il est donc illusoire (ou quelque peu artificiel) d'espérer distinguer parfaitement les caractéristiques individuelles et contextuelles et dissocier les effets de composition des effets de contextes. » (Vallée 2009, §32) Cette endogénéité s'ajoute au risque d'attribuer à un niveau une part des effets de l'autre niveau, par sur ou sous-estimation du modèle (*ibid.*).

Troisièmement, Salgado *et al.* (2014) mentionnent l'homogénéité des effets estimés par les modèles multiniveaux : les variables de niveau 2 s'appliquant uniformément sur tous les individus de niveau 1 de

---

<sup>50</sup> « in the tri-state area (i.e., New York, Connecticut, and New Jersey states), Massachusetts, and in parts of the Western United States, particularly in Nevada, Idaho, and Utah »

la même façon, alors que le groupe dans lequel un effet de pair s'applique – par exemple, dans le cadre de l'école – serait celui du groupe d'amies et d'amis plutôt que de la classe, voire de l'établissement dans son ensemble. L'usage de modèles de régression à pentes variables ou avec interactions interniveaux peut nuancer cette limite. Toutefois, la question des échelles multiples auxquelles s'opèrent les effets de contexte est cruciale. Les modèles à plus de deux niveaux sont certes possibles, mais particulièrement coûteux et exigeants. Par ailleurs, les niveaux auxquels les données sont disponibles restent souvent arbitraires : « Une faiblesse de l'approche multiniveaux est qu'elle ne considère pas vraiment l'impact de l'échelle sur la ségrégation mais sur la part de ségrégation qui peut être attribuée aux différents niveaux du modèle. Le problème est qu' [...] un niveau correspond habituellement à une combinaison d'échelles parce que les unités surfaciques qui le comprennent varient en termes de forme et de taille » (Harris 2017, p. 442)<sup>51</sup>.

Enfin, Salgado *et al.* (2014) insistent sur le fait que le choix des *proxy* utilisées comme variables à l'échelle des quartiers, régions ou établissements scolaires peut masquer les processus causaux reliant les individus à leur unité géographique ou groupement institutionnel. Par exemple, lorsque l'établissement scolaire est introduit en variable de niveau supérieur dans un modèle des performances scolaires individuelles, il est difficile de savoir si l'effet modélisé est celui des pairs au sein de l'école, celui de l'investissement scolaire des enseignants, des familles ou des ressources économiques et culturelles accessibles aux élèves en dehors de l'école. L'identification de variables précises pour chaque mécanisme et la collecte de données correspondantes sont des tâches souvent compromises (Hanushek *et al.* 2003). Cette limite n'est pas propre à la modélisation multiniveaux ou pondérée géographiquement, mais plutôt à la modélisation statistique en général.

Alors que les modèles statistiques décrivent la réalité captée par des données empiriques et permettent de repérer des associations privilégiées entre variables sans présumer de la causalité entre elles, la simulation (notamment multi-agents), en générant de manière causale des configurations, permet quant à elle de produire des démonstrations de possibilité causale, sans que l'on puisse prouver l'identité entre le processus simulé et le processus empirique ayant abouti à une configuration semblable. Ces deux approches sont donc plus complémentaires qu'alternatives, comme le suggèrent les travaux qui les associent. Par exemple, Salgado *et al.* (2014) utilisent les estimations issues de modèles multiniveaux sur les inégalités scolaires comme un étalon à partir duquel estimer la qualité du modèle multi-agents. Dans la thèse de Cottineau (2014), les variations de croissance urbaine des différentes régions de l'ex-Union soviétique sont identifiées par l'analyse multiniveaux, afin de construire le mécanisme génératif qui sera introduit dans le modèle multi-agents pour représenter l'effet (causal) des différents régimes démographiques régionaux (niveau 2) sur la démographie des villes (niveau 1).

#### **6.4. Modèles de simulation pour expliquer et illustrer la dynamique des inégalités dans l'espace géographique**

Ainsi, alors que les modèles des inégalités que nous avons vus jusqu'à présent servaient à organiser des données empiriques de manière à décrire les inégalités ou à en isoler les facteurs et échelles géographiques, les modèles que nous présentons ici sont destinés à « recréer » *in silico* des situations d'inégalité à partir de processus causaux appliqués à des sociétés synthétiques (c'est-à-dire des sociétés artificielles, n'existant que dans la simulation). Nous restreignons le type de simulation à deux formalismes particulièrement bien adaptés à la prise en compte de l'espace géographique : les

---

<sup>51</sup> « A weakness of the multilevel approach is that it does not really look at the impact of scale upon segregation but upon the amount of segregation that can be attributed to the various levels of the model. The problem is that [...] one level is usually a mixture of scales because the areal units that comprise it vary in shape and size. »

automates cellulaires et les modèles multi-agents, dont les différences et points communs sont présentés dans l'encadré 6.2.

Autrement dit, l'on simule une société d'agents soumis à des règles d'action découlant d'hypothèses choisies par le modélisateur ou la modélisatrice (en fonction de régularités empiriques et de propositions théoriques). L'on analyse ensuite les conséquences des interactions entre agents et entre les agents et leur environnement sur le niveau des inégalités, et leur répartition dans l'espace pour identifier quels mécanismes sont « candidats à l'explication » des inégalités observées dans les sociétés réelles. Ces mécanismes candidats correspondent aux règles d'action qui permettent de simuler la société artificielle la plus proche de sociétés réelles (selon une série d'indicateurs déterminés au préalable). Nous présentons trois exemples de modèles multi-agents simulant des inégalités dans l'espace géographique : un modèle de ségrégation (Schelling), deux modèles d'accès inégal à des ressources localisées et trois modèles d'éviction/gentrification.

Les **automates cellulaires** (ou *Cellular Automata*, CA, en anglais) ont été développés dès les années 1940 à Los Alamos par des chercheurs en mathématiques et en informatique (Ulam, Von Neumann). Considérant le pouvoir de calcul des ordinateurs de l'époque, leur fonctionnement est donc assez simple. Ils se présentent comme une grille de cellules (typiquement, des carrés) dotées de deux propriétés : chaque cellule contient une valeur en propre (que ce soit une couleur, un type ou un nombre) et est affectée par la valeur des cellules de son voisinage (que ce soit les quatre cellules situées sur, sous et de chaque côté – voisinage de Von Neumann – ou les huit cellules situées encadrant la cellule – voisinage de Moore). L'animation d'un automate cellulaire provient de l'implémentation de deux types de règles d'action : les règles de changement de la cellule en fonction de sa valeur en propre, et les règles de changement liées aux valeurs des cellules voisines. Bien que simples dans leur principe et leur implémentation, les automates cellulaires permettent de simuler une large palette de motifs géométriques et de dynamiques spatiales. Par exemple, avec seulement deux états possibles comme valeurs de cellule et cinq règles de changement, le « jeu de la vie » de Conway (1970) produit une diversité de motifs, dont certains sont stables, d'autres oscillatoires, d'autres dynamiques, et d'autres encore éphémères, en fonction de la configuration spatiale de départ (Gardner 1970). En géographie, ces simulations ont été popularisées par Batty (1997), White (1998) et Frankhauser (1998) pour leur capacité à simuler certaines structures (notamment fractales) de la croissance urbaine.

Les **modèles multi-agents** (ou *Agent Based Models*, ABM, en anglais) constituent des extensions d'automates cellulaires. En effet, les « agents » de ces modèles peuvent être des cellules immobiles disposées sur une grille, mais peuvent aussi être représentés par des entités ponctuelles, linéaires ou surfaciques mobiles. Dans la plupart des modèles multi-agents développés en géographie depuis les années 1990 (Bura *et al.* 1996 ; Batty *et al.* 1998), des agents mobiles interagissent entre eux et avec un environnement spatial pouvant être lui-même un automate cellulaire. Des règles de changement sont alors définies pour les agents mobiles : 1) en fonction de leur dynamique propre ; 2) en fonction de leurs interactions avec d'autres agents ; 3) en fonction de leur situation spatiale (interaction avec des cellules, par exemple). L'intérêt principal de la modélisation multi-agents par rapport à d'autres formes de simulation est de permettre la représentation d'agents hétérogènes, mobiles, dotés d'objectifs et potentiellement de mémoire. Son inconvénient principal résulte de cette flexibilité : il s'avère en effet coûteux d'évaluer les résultats simulés, en termes de robustesse et de sensibilité.

#### Encadré 6.2. Automates cellulaires et modèles multi-agents

### 6.4.1. Modèle multi-agents de ségrégation spatiale : Schelling et ses variantes

Le modèle de Schelling (1971, 1978) est l'un des plus célèbres exemples de simulation multi-agents en sciences sociales. Imaginé par T. Schelling dans les années 1960 à partir d'un damier, puis retranscrit par ses nombreux successeurs sur support numérique, ce modèle très simple met en jeu deux populations qualitativement différentes (souvent symbolisées par une couleur ou une forme différente) se mouvant dans un marché du logement très abstrait. L'objectif des agents est de « résider dans un logement » (c'est-à-dire se fixer sur une localisation correspondant à une cellule du damier ou de l'environnement numérique) dont les voisins satisfont un critère de diversité choisi pour l'ensemble

de la population. En l'occurrence, les agents sont dotés d'un pourcentage qui correspond à la part maximale de leur entourage, que les voisins de couleur ou de forme différentes peuvent représenter sans que les agents ne s'en émeuvent (c'est-à-dire qu'ils ne changent pas d'état dans la simulation). Si la proportion de voisins différents dépasse ce pourcentage, alors l'agent « décide » de changer de cellule de résidence et en cherche une qui soit libre, soit aléatoirement dans la plupart des modèles de simulation implémentés sur ordinateur (par exemple, le modèle disponible par défaut dans la bibliothèque d'exemples du logiciel NetLogo ; Tisue, Wilensky 2004), soit au plus près de la cellule qu'il quitte dans la version originale du modèle. La version la plus simple du modèle de Schelling se compose d'une « ville » de forme carrée et de faible dimension (quelques dizaines de cellules organisées en damier). La capacité d'accueil des cellules est généralement uniforme, avec un agent maximum par cellule, et seule une partie d'entre elles sont occupées en début de simulation (selon un pourcentage qui constitue un paramètre important du modèle), ce qui permet aux agents mécontents de se déplacer vers des cellules inoccupées en cours de simulation. La conclusion principale de ce modèle très simple dans le domaine de la modélisation des inégalités dans l'espace géographique est que le jeu des interactions locales et la perception réduite des agents conduisent à produire une situation collective de ségrégation spatiale, même lorsque les agents sont individuellement relativement tolérants (ils peuvent, par exemple, accepter de s'entourer de voisins à 60 ou 70 % différents d'eux). En effet, dès 30 % d'intolérance, le modèle converge vers des motifs ségrégués (Gauvin *et al.* 2009). La présence de nombreux espaces vacants favorise d'autant plus cette séparation des groupes dans l'espace. Bien que le modèle de Schelling ne prétende pas représenter le mécanisme par lequel le marché du logement favorise la ségrégation urbaine, il tend à prouver qu'il n'est pas nécessaire d'introduire des mécanismes conscients de ségrégation ou discriminatoires entre les populations pour produire des configurations d'inégalités de répartition des individus dans l'espace.

Ce modèle a fait l'objet de nombreuses critiques et d'autant de variantes d'implémentation. Notons, par exemple, les variations introduites dans les années 1990 par Portugali *et al.* (1994), qui font varier les niveaux de tolérance maximale entre les deux groupes de population simulés, ou par Benenson (1999), qui introduit des différences de statut économique entre agents, ce qui accentue ou équilibre les effets des préférences individuelles dans la localisation des individus. Flache et Hegelsmann (2001) introduisent de l'irrégularité dans les grilles et les partitions de l'espace du modèle, sans aboutir à des changements significatifs du comportement du modèle. Benenson et Torrens (2002) incluent la géométrie des bâtiments résidentiels et leur capacité d'accueil dans le modèle, afin de mieux prendre en compte les proximités entre agents et la morphologie de la ville, tandis que Laurie et Jaggi (2003) introduisent des différences de portée dans la perception des agents. Fossett et Dietrich (2009) modifient le modèle classique de Schelling pour mesurer la sensibilité de la forme et de la taille des cellules de la ville sur le niveau de ségrégation obtenu. En 2012, Hatna et Benenson (2012) produisent une version de Schelling permettant de simuler une population composée d'une majorité et d'une minorité, plutôt que deux groupes de taille égale, tandis que Banos (2012) introduit dans le modèle des structures réticulaires entre cellules pour comparer l'effet des formes de réseaux sur les niveaux de ségrégation observés en fin de simulation. En l'occurrence, l'auteur montre que les environnements de simulation suivant des réseaux fractals favorisent la séparation des groupes dans l'espace. Spielman et Harrison (2014) font varier la densité urbaine et les préférences de localisation des agents en fonction de leurs caractéristiques individuelles, à partir de données empiriques du XIX<sup>e</sup> siècle. Raimbault *et al.* (2019) s'intéressent à la sensibilité du modèle aux conditions spatiales initiales. Ils implémentent le modèle de Schelling sur des « villes » dont la distribution spatiale des densités correspond à des configurations proches de celle des villes européennes, c'est-à-dire non uniformes et variées (villes monocentriques, polycentriques, discontinues). Ils montrent que l'effet des paramètres du modèle (c'est-à-dire le niveau de tolérance individuelle des agents et le taux de vacance de la ville) est le même quel que soit le profil de densité de la ville, bien que les villes discontinues et polycentriques favorisent des niveaux finaux de ségrégation systématiquement plus importants. Enfin, Forsé et Parodi (2019) révèlent la fragilité des résultats du modèle en révélant une hypothèse implicite : « Le modèle de Schelling néglige [...] les déménagements effectués pour d'autres raisons que l'intolérance [...] En

prenant en considération ces autres déménagements, on observerait spontanément la disparition de quartiers homogènes qui ne sont que des formations transitoires. La conclusion de Schelling est donc très fragile car, au fond, il arrête le film au moment où, par accident, des quartiers homogènes se sont formés. » (Forsé, Parodi 2019, p. 101)

#### **6.4.2. Deux modèles multi-agents de l'émergence des inégalités liées à la répartition de ressources dans l'espace**

Le modèle *Sugarscape* fait partie des modèles d'interaction environnement-société qui contribuent à expliquer la formation des inégalités entre individus et entre noyaux de peuplement, à partir de sociétés artificielles intégrées dans leur environnement physique (Schmitt, Pumain 2013). Proposé initialement par Epstein et Axtell (1996), des pionniers de la simulation sociale, ce modèle multi-agents est conçu pour étudier les interactions entre mécanismes plausibles d'émergence de l'économie et de la différenciation entre individus dans une société. Sa configuration est très simple et peut être complétée par des versions plus élaborées du modèle. Des agents évoluent dans un paysage caractérisé par la présence non uniforme d'une ressource (*sugar*) renouvelable. Ils peuvent collecter cette ressource, la stocker ou s'en servir pour leur survie. Cette collecte dépend de leur perception, qui varie d'un individu à l'autre (certains voient plus loin que d'autres). La quantité de sucre nécessaire à la survie (ou au métabolisme) varie aussi d'un individu à l'autre. La répartition inégale du sucre dans le paysage, combinée aux différences de vision et de métabolisme individuels, suffit à produire une société économiquement inégale, dans laquelle, très vite, les niveaux de richesse (mesurée en quantité de sucre) varient très fortement entre les individus, certains détenant une quantité très importante de sucre et d'autres une quantité insuffisante pour survivre. La mesure sur laquelle se base l'évaluation et la comparaison des sorties du modèle est donc généralement, comme dans l'implémentation NetLogo de J. Li et U. Wilensky, l'indice de Gini, qui synthétise ici les inégalités de richesse-sucre entre agents.

Comme pour Schelling, la simplicité du modèle initial a suscité un nombre important d'ajouts et de mécanismes complémentaires. C'est notamment le fait des auteurs du modèle initial, qui ont fait évoluer *Sugarscape* de sa forme la plus épurée à un modèle dans lequel le comportement des agents se rapproche de la diversité empirique des interactions entre membres d'une société. « Au fur et à mesure, le répertoire des comportements des agents grossit pour inclure le mouvement, la récolte de ressources, la reproduction sexuelle, le combat, la transmission de culture, l'échange, le commerce, l'héritage, le crédit, la pollution, la propagation de maladies et la résistance immunitaire. » (Schmitt, Pumain 2013, §8) En comparant l'impact de la configuration spatiale de départ dans les modèles de Schelling et de *Sugarscape*, Raimbault *et al.* (2019) montrent que l'implémentation simple de *Sugarscape* est très sensible à la distribution initiale de la ressource dans l'émergence des inégalités entre individus, au point que l'on peut modifier les sorties du modèle plus fortement en changeant la configuration spatiale de départ qu'en modifiant les valeurs des paramètres du modèle (capacité de production minimale et maximale de sucre des cellules, nombre d'agents, notamment).

Dans le modèle d'Auchincloss *et al.* (2011), les inégalités modélisées entre individus sont celles de leur alimentation. Elles émergent en raison des différences de revenus des individus dans un environnement ségrégué, par l'intermédiaire de l'offre de nourriture saine dans l'espace, qui s'adapte aux résidents présents localement. Le modèle simule en effet deux types d'agents : des ménages et des magasins de nourriture. Les ménages sont distribués en deux catégories de taille égale : les ménages riches et les ménages pauvres. Ils sont dotés d'une préférence (continue entre 0 et 1) pour la nourriture saine, qui est distribuée soit aléatoirement, soit conditionnellement à leur catégorie (riche/pauvre), selon les scénarios simulés. Les magasins de nourriture sont distribués en deux catégories de taille égale : les magasins vendant de la nourriture saine (*healthy*) et ceux vendant de la nourriture non saine (*unhealthy*). La moitié des magasins est simulée comme vendant leur nourriture bon marché, et l'autre moitié vendant leur nourriture à prix élevés, le prix étant attribué soit aléatoirement, soit conditionnellement à leur catégorie (sain/non sain) selon les scénarios simulés.

L'espace est simulé de manière très abstraite, par un tore (à savoir un espace replié sur lui-même pour supprimer les effets de bord) de 50 x 50 cellules carrées. Les ménages sont répartis aléatoirement ou isolés par niveau de revenus dans deux zones distinctes de l'espace selon les scénarios, tandis que les magasins sont répartis aléatoirement ou isolés par type de nourriture vendue (saine ou non) dans deux zones distinctes de l'espace selon les scénarios. Aucun agent ne se déplace entre ces zones au cours de la simulation. La dynamique du modèle s'articule autour de la relation entre les ménages et les magasins. Les ménages évaluent les magasins qu'ils fréquentent en fonction de leur prix, de leur accessibilité, de l'adéquation entre la nourriture vendue et leur préférence, et distribuent leurs dépenses entre les magasins qui obtiennent les meilleurs scores. Les magasins réagissent aux dépenses des ménages en fermant si le nombre de clients est trop bas, tandis que de nouveaux magasins (avec éventuellement des prix et de la nourriture différents) peuvent les remplacer aux mêmes localisations. Auchincloss *et al.* (2011) montrent, à partir de ce modèle et de différents scénarios, qu'en l'absence d'autres facteurs, les inégalités nutritionnelles liées au revenu résultent de la ségrégation des ménages riches et des magasins de nourriture saine dans des zones distinctes des ménages pauvres et des magasins de nourriture non saine. Lorsque les ménages ont une préférence pour l'alimentation saine, indépendamment de leur niveau de revenu, l'alimentation des plus pauvres apparaît systématiquement plus saine que dans les autres scénarios, bien qu'une inégalité persiste entre riches et pauvres. Enfin, les auteurs montrent que seuls des préférences favorables et des magasins bon marché accessibles peuvent permettre de surmonter le différentiel introduit par la ségrégation des ménages, démontrant ainsi l'importance structurelle de l'inscription des agents dans leur environnement sur les inégalités d'alimentation par rapport aux modifications possibles de leur comportement.

#### **6.4.3. Trois modèles simulant une dynamique de relégation spatiale**

Le dernier type de modèle de simulation des inégalités entre individus dans l'espace géographique que nous souhaitons présenter dans ce chapitre concerne le processus de relégation spatiale, c'est-à-dire le déplacement forcé de populations pauvres de certains quartiers urbains, suite au déplacement choisi de fractions sociales plus aisées qui, en investissant dans des quartiers mixtes, pauvres et/ou dégradés, contribuent au renouvellement urbain physique, puis au remplacement social de ces quartiers. Trois modèles s'intéressent à cette dynamique qui modifie l'espace urbain par l'interaction entre des individus économiquement inégaux.

Le premier de ces modèles (O'Sullivan 2002) est un automate cellulaire, c'est-à-dire un modèle dans lequel la localisation géographique des agents est fixe et s'apparente à une cellule. L'auteur modélise le processus de relégation spatiale du point de vue des bâtiments et des logements, en se concentrant sur l'hypothèse du *rent gap* (Smith 1979). Le *rent gap*, ou décalage entre la rente foncière potentielle et les revenus du foncier en cours, fait partie des éléments de théorie économique participant à l'explication du renouvellement urbain et social des quartiers urbains. D. O'Sullivan explore les aspects géographiques et temporels de ces décalages économiques entre la valeur d'usage du logement et sa valeur potentielle, du fait de sa localisation. Son modèle est plus raffiné en termes de représentation spatiale que la plupart des automates cellulaires classiques, construits comme des grilles régulières de cellules identiques et carrées. Les cellules qu'il représente sont les centroïdes de bâtiments, dont le voisinage est construit selon un graphe de proximité plutôt que par contiguïté. Chaque cellule a quatre états possibles : être à vendre, ne pas être à vendre, être à la recherche d'un locataire ou être loué. La transition entre ces quatre états est conditionnée par les caractéristiques du bâtiment (son degré de dégradation, notamment) et le revenu des occupants, la demande disponible pour l'achat et la location, de même que par un indicateur de statut du quartier, mais contient aussi un élément de stochasticité (soit de hasard). Le modèle est capable de reproduire une dynamique de dégradation physique et économique du quartier, suivie d'un mouvement de gentrification à partir de l'exemple du quartier de Hoxton dans l'Est londonien. Sa représentation des inégalités interindividuelles dans l'espace géographique est donc assez mécanique et impersonnelle, puisque la dynamique du modèle se joue à l'échelle des immeubles et de leurs caractéristiques objectives. L'auteur suggère toutefois que le

paramètre de statut du quartier est crucial pour déterminer la dynamique du modèle, puisqu'il actualise à chaque pas de temps la composition du quartier en fonction des revenus des habitants présents à un moment donné, et détermine ainsi le revenu des potentiels nouveaux propriétaires et locataires. De l'aveu de D. O'Sullivan lui-même, ce paramètre devrait plutôt être traité comme un facteur exogène en lien avec la représentation des autres quartiers de la ville pour une simulation plus proche de la réalité : « A much more satisfactory implementation would embed the neighborhood presented here in a larger model of the urban system. » (O'Sullivan 2002, p. 270) En effet, ce paramètre peut être interprété comme un élément de réputation du quartier, c'est-à-dire un élément important de la dynamique sociale de gentrification en théorie.

Le deuxième modèle est une hybridation entre un automate cellulaire représentant les logements d'un quartier selon une grille régulière de cellules carrées, et un modèle multi-agents représentant les individus se déplaçant entre ces logements et « animant » l'espace social, économique et urbain. Torrens et Nara (2007) focalisent leur attention sur le fonctionnement du marché du logement, et représentent quatre scénarios de gentrification qu'ils confrontent aux données d'un quartier de Salt Lake City. Le premier scénario est celui de l'absence de changement (*Business as Usual*), le second représente un changement du point de vue de la demande (en modélisation l'arrivée d'un groupe de « gentrificateurs » aux comportements économiques et sociaux différents de ceux de la population initialement présente), le troisième est un changement du point de vue de l'offre (par la prise en compte du *rent gap* et de la sous-évaluation du stock immobilier à un moment donné), et le quatrième illustre la combinaison des scénarios deux et trois. La règle d'action des agents consiste à décider s'ils déménagent ou non, à sélectionner le cas échéant l'un des trois marchés du logement de la ville, puis à y choisir une propriété où emménager. Les facteurs influençant cette décision dépendent de leurs préférences (pour un type de logement, sa taille, etc.), des caractéristiques du logement, de celles du quartier (accessibilité à un *shopping mall*, à une bretelle d'autoroute, au centre-ville ; ainsi que le taux de vacance alentour) et de leurs propres ressources. Les auteurs montrent (2007, p. 352) que seul le dernier scénario, incluant les dynamiques de l'offre et de la demande, permet d'expliquer raisonnablement les dynamiques de gentrification observées à Salt Lake City. Ici, la modélisation des inégalités dans l'espace géographique intègre la compétition pour les logements du quartier par des ménages, locaux ou nouveaux arrivants, disposant de ressources inégales. L'attractivité des logements eux-mêmes dépend de leur situation vis-à-vis de ressources fixes (comme le centre-ville et les axes de transport).

Le dernier modèle, consacré au déplacement forcé de populations dans un espace urbain, est un modèle multi-agents développé par Boeing (2018). L'auteur s'intéresse à l'effet de la corrélation entre préférences individuelles et statut économique dans l'éviction des catégories modestes des zones qui leur sont propices. Malgré ses treize paramètres, le modèle est beaucoup plus simple que les deux précédents, et conçu comme un modèle théorique plutôt que comme pouvant s'appliquer à un cas particulier. Les agents y sont catégorisés en trois niveaux de richesse (pauvres, moyens, riches), tandis que les cellules disponibles à leur installation se caractérisent par deux dimensions – d'une part, la présence d'aménités abstraites fonctionnant comme des biens de club (c'est-à-dire dont le bénéfice s'applique également à tous ceux qui y ont accès, sans diminution de son stock par l'augmentation du nombre de membres, à l'instar des biens de la connaissance, par exemple) et, d'autre part, leur niveau de densité. La préférence des agents pauvres est de résider près de l'aménité qu'ils exploitent pour vivre, tandis que la préférence des agents riches est de résider près des zones densément peuplées. La rente foncière augmente avec le nombre et le revenu des agents présents sur une cellule, conduisant les occupants qui n'ont pas les moyens suffisants pour payer la rente à changer de localisation. Le modèle conduit les ménages pauvres à former des agglomérations denses autour des aménités qui attirent les ménages riches. La présence dense des ménages riches expulse à son tour les ménages pauvres vers la périphérie du groupement, et loin des aménités dont ils dépendent, aboutissant ainsi à un état inefficace du système. À la manière de Schelling, le modèle de Boeing permet de reproduire un fait stylisé de l'inégalité entre individus dans et par l'espace géographique, en simulant un mécanisme

suffisant pour reproduire une dynamique observée empiriquement, bien que globalement peu réaliste par lui-même.

## 6.5. Conclusion

Ce tour d'horizon de la modélisation des inégalités interindividuelles dans l'espace géographique a permis de rendre compte de la diversité des objectifs des modèles possibles, de la diversité des approches méthodologiques mises en place pour remplir ces objectifs, ainsi que de la multiplicité des points de vue sur une même question (la relégation spatiale, par exemple). Nous avons relevé le fait que les différentes familles de modélisation (cartographique, statistique, générative) sont à utiliser de manière complémentaire plutôt que concurrente. En effet, les modèles cartographiques synthétisent les régularités empiriques de la distribution géographique des inégalités individuelles, et suggèrent des facteurs à introduire dans les modèles statistiques. Ceux-ci décrivent l'influence relative de facteurs mesurables sur la distribution des inégalités, et leurs résultats nourrissent la construction et l'évaluation de la plausibilité des modèles génératifs. Les résultats théoriques de la simulation peuvent en retour enrichir l'ensemble d'hypothèses à tester empiriquement. Selon Manzo (2005), « la littérature récente commence à dessiner un type de sociologie empirique quantitative selon lequel l'analyse par variables décrit la modélisation par mécanismes (construit en termes d'individualisme méthodologique) explique et la simulation anime (et teste) dynamiquement les mécanismes supposés être à la base des relations statistiques observées. » (p. 39) De la même manière, les variables et mécanismes retenus par les différentes études sont souvent plus complémentaires que concurrents dans l'explication des inégalités sociales et économiques. Ainsi, certains modèles mettent l'accent sur l'aspect physique de l'environnement spatial, d'autres sur sa composition démographique, économique ou sociale, pour expliquer les inégalités dans l'espace géographique. Il semble que la diversité de la réalité sociale plaide plutôt en faveur d'une combinaison de ces registres d'explication, variable selon les cas et les époques.

## 6.6. Bibliographie

- Alacevich, M., Soci, A. (2017). *A Short History of Inequality*. Agenda Publishing, Newcastle upon Tyne.
- Apparicio, P. (2000). Les indices de ségrégation résidentielle : un outil intégré dans un système d'information géographique. *Cybergeo*, 134. doi:10.4000/cybergeo.12063.
- Atkinson, A.B. (1970). On the measurement of inequality. *Journal of Economic Theory*, 2, 244–263.
- Auchincloss, A.H., Riolo, R.L., Brown, D.G., Cook, J., Diez-Roux, A.V. (2011). An agent-based model of income inequalities in diet in the context of residential segregation. *American journal of preventive medicine*, 40(3), 303–311.
- Banos, A. (2012). Network effects in Schelling's model of segregation: new evidences from agent-based simulation. *Environment and Planning B: Planning and Design*, 39(2), 393–405. doi:10.1068/b37068.
- Batty, M. (1997). Cellular automata and urban form: a primer. *Journal of the American Planning Association*, 63(2), 266–274.
- Batty, M., Jiang, B., Thurstain-Goodwin, M. (1998). Local movement: agent-based models of pedestrian flows. UCL Working Paper [En ligne]. Disponible à l'adresse : <https://discovery.ucl.ac.uk/id/eprint/225/1/paper4.pdf>.
- Bell, A., Hartman, T., Piekut, A., Rae, A., Taylor, M. (2019). *Making sense of data in the media*. Sage Publishing, Thousand Oaks.
- Bell, W., Shevky, E. (1955). *Social area analysis; theory, illustrative application, and computational procedures*. Stanford University Press, Stanford.

- Benenson, I. (1999). Modeling population dynamics in the city: from a regional to a multi-agent approach. *Discrete Dynamics in Nature and Society*, 3(2/3), 149–170. doi:10.1155/S1026022699000187.
- Benenson, I., Torrens, P.M. (2004). Geosimulation: object-based modeling of urban phenomena. *Computers, Environment and Urban Systems*, 28(1/2), 1–8. doi:10.1016/s0198-9715(02)00067-4.
- Berry, B.J.L., Kasarda, J.D. (1977). *Contemporary urban ecology*. Macmillan Publishing Company, New York.
- Boeing, G. (2018). The Effects of Inequality, Density, and Heterogeneous Residential Preferences on Urban Displacement and Metropolitan Structure: An Agent-Based Model. *Urban Science*, 2(76) [En ligne]. Disponible à l'adresse : <https://doi.org/10.3390/urbansci2030076>.
- Borman, G., Dowling, M. (2010). Schools and inequality: A multilevel analysis of Coleman's equality of educational opportunity data. *Teachers College Record*, 112(5), 1201–1246.
- Brunsdon, C., Fotheringham, A.S., Charlton, M.E. (1996). Geographically weighted regression: a method for exploring spatial nonstationarity. *Geographical analysis*, 28(4), 281–298.
- Bura, S., Guérin-Pace, F., Mathian, H., Pumain, D., Sanders, L. (1996). Multiagent systems and the dynamics of a settlement system. *Geographical analysis*, 28(2), 161–178.
- Comber, A., Brunsdon, C., Charlton, M., Dong, G., Harris, R., Lu, B., Lü, Y., Murakami, D., Nakaya, T., Wang, Y., Harris, P. (2020). The GWR route map: a guide to the informed application of Geographically Weighted Regression. arXiv preprint arXiv:2004.06070.
- Conway, J. (1970). The game of life. *Scientific American*, 223(4), 4.
- Cottineau, C. (2014). L'évolution des villes dans l'espace post-soviétique. Observation et modélisations. PhD dissertation, Université Paris 1 Panthéon-Sorbonne, Paris.
- Cottineau, C., Chapron, P., Le Texier, M., Rey-Coyrehourcq, S. (2019). Incremental Territorial Modeling. Dans *Geographical Modeling, Cities and Territories*, Pumain, D. (dir). ISTE Ltd, Londres et Wiley, New York, 95–123 [En ligne]. Disponible à l'adresse : <http://doi.org/10.1002/9781119687290.ch4>.
- Diez-Roux, A.V., Link, B.G., Northridge, M.E. (2000). A multilevel analysis of income inequality and cardiovascular disease risk factors. *Social science & medicine*, 50(5), 673–687.
- Duncan, O.D., Duncan, B. (1955). Residential distribution and occupational stratification. *American journal of sociology*, 60(5), 493–503.
- Duroudier, S. (2014). Les divisions socio-spatiales dans les villes intermédiaires des États-Unis. Perspectives de recherche à partir de la notion de discontinuité. *Espace géographique*, 43(2), 134–147 [En ligne]. Disponible à l'adresse : <https://doi.org/10.3917/eg.432.0134>.
- Edmonds, B., le Page, C., Bithell, M., Chattoe-Brown, E., Grimm, V., Meyer, R., Montañola-Sales, C., Ormerod, P., Root, H., Squazzoni, F. (2019). Different Modelling Purposes. *Journal of Artificial Societies and Social Simulation*, 22(3), 6 [En ligne]. Disponible à l'adresse : <https://doi.org/10.18564/jasss.3993>.
- Epstein, J.M. (2008). Why model?. *Journal of Artificial Societies and Social Simulation*, 11(4), 12 [En ligne]. Disponible à l'adresse : <http://jasss.soc.surrey.ac.uk/11/4/12.html>.
- Epstein, J.M., Axtell, R. (1996). *Growing Artificial Societies: Social Science from the Bottom Up*. MIT Press, Cambridge.
- Flache, A., Hegselmann, R. (2001). Do Irregular Grids make a Difference? Relaxing the Spatial Regularity Assumption in Cellular Models of Social Dynamics. *Journal of Artificial Societies and Social Simulation*, 4(4), 6 [En ligne]. Disponible à l'adresse : <http://www.soc.surrey.ac.uk/JASSS/4/4/6.html>.

- Forsé, M., Parodi, M. (2019). Retour critique sur le modèle de ségrégation urbaine de Schelling. *Émulations-Revue de sciences sociales*, 31, 91–104.
- Fossett, M, Dietrich, D.R. (2009). Effects of city size, shape, and form, and neighborhood size and shape in agent-based models of residential segregation: Are Schelling-style preference effects robust?. *Environment and Planning B: Planning and Design*, 36(1), 149–169. [doi:10.1068/b33042](https://doi.org/10.1068/b33042).
- Fotheringham, S.A., Yang, W., Kang, W. (2017). Multiscale Geographically Weighted Regression (MGWR). *Annals of the American Association of Geographers*, 107(6), 1247–1265. [doi:10.1080/24694452.2017.1352480](https://doi.org/10.1080/24694452.2017.1352480).
- François, J.-C. (2002). Ressemblance et proximités : un point de vue sur le contexte théorique de la notion de discontinuité géographique. *Cybergeo*, 214.
- François, J.-C., Ribardièrre, A. (2004). Qu'apporte l'échelon infracommunal à la carte des inégalités de richesse en Île-de-France ? *MappeMonde*, 3(75), 1–8 [En ligne]. Disponible à l'adresse : <http://mappemonde-archiv.mgm.fr/num3/articles/art04305.html>.
- Frankhauser, P. (1998). Fractal geometry of urban patterns and their morphogenesis. *Discrete Dynamics in Nature and Society*, 2(2), 127–145.
- Gardner, M. (1970). Mathematical Games. The fantastic combinations of John Conway's new solitaire game « life ». *Scientific American*, 223, 123–120
- Gauvin, L., Vannimenus, J., Nadal, J.P. (2009). Phase diagram of a Schelling segregation model. *The European Physical Journal B*, 70(2), 293–304.
- George, P. (1981). *Géographie des inégalités*. Presses Universitaires de France, Paris, 127.
- Goldstein, H. (1987). *Multilevel models in Educational and Social research*. Oxford University Press, Londres.
- Grasland, C., François, J.-C., Brunet, R. (1997). La discontinuité en géographie : origines et problèmes de recherche. *L'Espace géographique*, 297–308.
- Hanushek, E.A., Kain, J.F., Markman, J.M., Rivkin, S.G. (2003). Does peer ability affect student achievement ? *Journal of Applied Econometrics*, 18(5), 527–544. [doi:10.1002/jae.741](https://doi.org/10.1002/jae.741).
- Harris, C.D., Ullman, E.L. (1945). The Nature of Cities. *The Annals of the American Academy of Political and Social Science*, 242, 7–17
- Harris, R. (2017). Measuring the scales of segregation: Looking at the residential separation of White British and other schoolchildren in England using a multilevel index of dissimilarity. *Transactions of the Institute of British Geographers*, 42(3), 432–444.
- Hatna, E., Benenson, I. (2012). The Schelling model of ethnic residential dynamics: Beyond the integrated-segregated dichotomy of patterns. *Journal of Artificial Societies and Social Simulation*, 15(1), 6. [doi:10.18564/jasss.1873](https://doi.org/10.18564/jasss.1873).
- Hoyt, H. (1939). *The structure and growth of residential neighborhoods in American cities*. US Government Printing Office, Washington.
- Kakwani, N., Wagstaff, A., Van Doorslaer, E. (1997). Socioeconomic inequalities in health: measurement, computation, and statistical inference. *Journal of econometrics*, 77(1), 87–103.
- Laurie, A.J., Jaggi, N.K. (2003). Role of “vision” in neighbourhood racial segregation: a variant of the Schelling segregation model. *Urban Studies*, 40(13), 2687–2704. 00115. [doi:10.1080/0042098032000146849](https://doi.org/10.1080/0042098032000146849).
- Le Bras, H. (2014). *Atlas des inégalités. Les Français face à la crise*. Autrement, Paris.
- Leckie, G. (2008). Modelling the effects of pupil mobility and neighbourhood on school differences in educational achievement. Centre for Market and Public Organisation.

- Leigh, A. (2007). How closely do top income shares track other measures of inequality?. *The Economic Journal*, 117(524), F619–F633
- Manski, C.F. (1993). Identification of Endogenous Social Effects: The Reflection Problem. *The Review of Economic Studies*, 60(3), 531–542.
- Manzo, G. (2005). Variables, mécanismes et simulations : une synthèse des trois méthodes est-elle possible ? Une analyse critique de la littérature. *Revue française de sociologie*, 46(1), 37–74.
- Mathieu, J.E., Aguinis, H., Culpepper, S.A., Chen, G. (2012). Understanding and estimating the power to detect cross-level interaction effects in multilevel modeling. *Journal of Applied Psychology*, 97(5), 951.
- Mollalo, A., Vahedi, B., Rivera, K.M. (2020). GIS-based spatial modeling of COVID-19 incidence rate in the continental United States. *Science of The Total Environment*, 728, 138884 [En ligne]. Disponible à l'adresse : <https://doi.org/10.1016/j.scitotenv.2020.138884>.
- Morrill, R. (1991). On the measure of geographic segregation. *Geography research forum*, 11, 25–36.
- O'Sullivan, D. (2002). Toward micro-scale spatial modeling of gentrification. *Journal of Geographical Systems*, 4(3), 251–274.
- Park, R.E., Burgess, E.W. (1925). *The Growth of the City: An Introduction to a Research Project*. University of Chicago Press, Chicago.
- Piketty, T. (2001). *Les Hauts Revenus en France au 20 siècle*. Grasset, Paris.
- Portugali, J., Benenson, I., Omer, I. (1994). Socio-spatial residential dynamics: stability and instability within a self-organized city. *Geographical Analysis*, 26(4), 321–340. [doi:10.1111/j.1538-4632.1994.tb00329.x](https://doi.org/10.1111/j.1538-4632.1994.tb00329.x).
- Raimbault, J., Cottineau, C., Texier, M.L., Néchet, F.L., Reuillon, R. (2018). Space matters: extending sensitivity analysis to initial spatial conditions in geosimulation models. *Journal of Artificial Societies and Social Simulation*, 22(4), 10. [doi:10.18564/jasss.4136](https://doi.org/10.18564/jasss.4136).
- Reardon, S. F., O'Sullivan, D. (2004). Measures of spatial segregation. *Sociological methodology*, 34(1), 121–162.
- Reynolds, A. (2012). The misuse of top 1 percent income shares as a measure of inequality. *Cato working Paper*, 105. [doi:10.2139/SSRN.2226941](https://doi.org/10.2139/SSRN.2226941).
- Roberto, E., Hwang, J. (2015). Barriers to integration: Physical boundaries and the spatial structure of residential segregation. arXiv :1509.02574.
- Roscigno, V.J., Tomaskovic-Devey, D., Crowley, M. (2006). Education and the inequalities of place. *Social Forces*, 84(4), 2121–2145.
- Sala-i-Martin, X. (2006). The world distribution of income: falling poverty and... convergence, period. *The Quarterly Journal of Economics*, 351–397.
- Salgado, M., Marchione, E., Gilbert, N. (2014). Analysing Differential School Effectiveness Through Multilevel and Agent-Based Modelling. *Journal of Artificial Societies and Social Simulation*, 17(4), 3. [doi:10.18564/jasss.2534](https://doi.org/10.18564/jasss.2534).
- Schelling, T.C. (1971). Dynamic Models of Segregation. *Journal of Mathematical Sociology*, 1(2), 143–186.
- Schelling, T.C. (1978). *Micromotives and Macrobehavior*. Norton, New York.
- Schmitt, C., Pumain, D. (2013). Modélographie multi-agents de la simulation des interactions sociétés-environnement et de l'émergence des villes. *Cybergeo: European Journal of Geography*, 643. [doi:10.4000/cybergeo.25900](https://doi.org/10.4000/cybergeo.25900).

- Smith, N. (1979). Toward a theory of gentrification a back to the city movement by capital, not people. *Journal of the American planning association*, 45, 538–548.
- Spielman, S., Harrison, P. (2014). The Co-evolution of Residential Segregation and the Built Environment at the Turn of the 20th Century: A Schelling Model. *Transactions in GIS*, 18(1), 25–45. doi:[10.1111/tgis.12014](https://doi.org/10.1111/tgis.12014).
- Tisue, S., Wilensky, U. (2004). Netlogo: A simple environment for modeling complexity. Dans *International conference on complex systems*. 21, 16–21.
- Torrens, P.M., Nara, A. (2007). Modeling gentrification dynamics: A hybrid approach. *Computers, Environment and Urban Systems*, 31(3), 337–361.
- Vallée, J. (2009). Les disparités spatiales de santé en ville : l'exemple de Vientiane (Laos). *Cybergeog: European Journal of Geography*, 477. doi:[10.4000/cybergeog.22775](https://doi.org/10.4000/cybergeog.22775).
- Vallée, J., Cadot, E., Grillo, F., Parizot, I., Chauvin, P. (2010). The combined effects of activity space and neighbourhood of residence on participation in preventive health-care activities: The case of cervical screening in the Paris metropolitan area (France). *Health & place*, 16(5), 838–852.
- van der Leeuw, S.E. (2004). Why model?. *Cybernetics and Systems*, 35(2/3), 117–128.
- White, R. (1998). Cities and cellular automata. *Discrete dynamics in Nature and Society*, 2(2), 111–125.
- Wilkinson, R., Pickett, K. (2009). *The Spirit Level: Why greater equality makes societies stronger*. Penguin Books, Londres.
- Wong, D.W.S. (1993). Spatial Indices of Segregation. *Urban Studies*, 30(3), 559–572. doi:[10.1080/00420989320080551](https://doi.org/10.1080/00420989320080551).
- Yang, W. (2014). An extension of geographically weighted regression with flexible bandwidths. Thèse de doctorat, School of Geography and Geosciences, University of St. Andrews, Fife [En ligne]. Disponible à l'adresse : <http://hdl.handle.net/10023/7052>.