



HAL
open science

Modeling Inequalities in Geographical Space

Clémentine Cottineau

► **To cite this version:**

Clémentine Cottineau. Modeling Inequalities in Geographical Space. Clémentine Cottineau; Julie Vallée. Inequalities in Geographical Space, Wiley; ISTE, pp.151-182, 2022, SCIENCES – Geography and Demography, 9781789450880. 10.1002/9781394188338.ch6 . halshs-04828670

HAL Id: halshs-04828670

<https://shs.hal.science/halshs-04828670v1>

Submitted on 10 Dec 2024

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

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



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

6

Modeling Inequalities in Geographical Space

Clémentine COTTINEAU^{1,2}

¹ Centre Maurice Halbwachs, CNRS, Paris, France

² Technische Universiteit Delft, The Netherlands

6.1. Introduction: different modeling formalisms for different purposes

Neither the level of inequalities nor the trend of their evolution are random or accidental. They are the result of economic, political, social and spatial processes that build up over time, fueled by the interactions and strategies of multiple actors, not least collective institutions, whether public or private. As a result, there are multiple “explanations” for inequalities, and at least as many ways to detect them through quantitative analysis. Modeling inequalities, from the most abstract to the most descriptive, can thus account for observed trends, illustrate the (expected or unexpected) effect of a particular policy measure or even include interactions between agents or between agents and their geographical environment in the production and reproduction of inequalities. In this chapter, it is neither a question of reviewing all the models dedicated to inequalities in geographical space nor a question of producing a manual for modeling inequalities. Rather, the ambition is to present, on the one hand, different modeling approaches (statistical, spatial and agent-based), and what they bring to different inequality modeling objectives; on the other hand, it is to present some salient examples and important results from existing models dedicated to inequalities in geographical space.

Since a model is defined as a simplified representation of reality for a particular purpose, the question arises, as with other models: “why model?” (van der Leeuw

Inequalities in Geographical Space,
coordinated by Clémentine COTTINEAU and Julie VALLÉE. © ISTE Ltd 2022.

2004; Epstein 2008; Edmonds et al. 2019) inequalities among individuals in geographical space? The purpose of the model – whether it is to explain, predict or illustrate a theory, for example – conditions the formalism and precision with which the system in question is represented, as well as the choice of entities and interactions present in the model. This is similar to generalization in cartography, which corresponds to the selection and simplification of the elements chosen to appear on a map (Cottineau et al. 2019). With regard to the modeling of inequalities in geographical space more specifically, the first decision is based on the type of inequality to be studied: economic, health, education, etc. The second branch on the decision tree depends on the place occupied by space in the description and explanation of these inequalities. Thus, while a series of mathematical models attempt to simplify the description of inequalities (particularly income inequalities) in order to be able to map them and eventually draw conclusions in terms of spatial distribution (section 6.2), multilevel and geographically weighted regression (GWR) models combine the effect of the relative location and the geographical affiliation of individuals in the estimation of inter-individual inequalities (section 6.3). With simulation models, the aim is to model the process by which inequalities are created, by including geographical space in the rules of action and change of agents' attributes (section 6.4). There are, of course, a number of other models, especially equilibrium models from economics (see Chapter 4), which we felt were too specific and insufficiently spatialized to be included in this chapter. Here, we shall reduce the focus to the graphical, statistical and generative modeling of inequalities in geographical space.

6.2. Inequality in the distribution of economic resources and in its spatial distribution

In the introduction, we defined inequality as the existence of systematic, significant and perceived differences in the distribution of social resources (e.g. wealth, health, education) between individuals because of certain characteristics (e.g. gender, age, ethnicity). These inequalities between individuals can be translated into space when privileged individuals are concentrated in certain places and disadvantaged individuals in a distinct set of places. However, these spatially translated inequalities are distinct from spatial inequalities, which would concern, for example, unequal exposure to certain nuisances or unequal access to certain resources (resulting from the distribution of facilities and services over a territory). These two types of inequalities are obviously linked to the presence of rare and valuable facilities (historical monuments, blue and green spaces, renowned hospitals) or, on the contrary, to infrastructures with negative externalities (wastewater treatment plant, motorway, cemetery) which condition the location

(residential, for example) of favored or disadvantaged individuals, while modifying the potential access of populations to different resources. However, it seems more judicious to distinguish between them in the analysis in order to better understand the dynamics at work.

We therefore first discuss models for describing inequalities in the distribution of economic resources (section 6.2.1), before detailing a method for describing inequalities in space, namely through mapping (section 6.2.2). Third, we evoke the models of spatial inequality and social discontinuity in space (section 6.2.3).

6.2.1. Lorenz/Gini, Pareto and Theil in economics: three distinct models

In the field of economic inequality, and typically income inequality, the most widely used mathematical model is the Lorenz curve and its associated index: the Gini coefficient. The model implemented by Lorenz in 1905 is graphic and aims to represent the concentration of wealth in relation to the concentration of the population within a given area. By indicating the percentage of the population classified by increasing wealth on the x-axis, and the percentage of wealth held by the corresponding population on the y-axis, the Lorenz curve is a convex curve (or straight line) between the origin (0;0) and the coordinates (1;1). Indeed, 0% of the population always holds 0% of the wealth and 100% of the population always holds 100% of the wealth. Between these two known points, the convexity of the curve indicates the extent of the concentration of wealth in a small number of hands, and thus the degree of (mathematical) inequality in the distribution of income. A straight curve indicates a perfectly egalitarian distribution, since each proportion of the population has an equivalent proportion of the total wealth, whereas an extremely convex curve indicates a very high concentration of wealth. In the example in Figure 6.1, which has a particularly convex curve, we find that the poorest 80% of the population hold only 20% of the total wealth. Conversely, the richest 20% of individuals hold 80% of the wealth.

In 1914, Gini found the Lorenz curve and made it famous (Alacevich and Soci 2017) by creating a concentration index bounded between 0 and 1, which corresponds graphically to the area marked in yellow in Figure 6.1, between the Lorenz curve and the blue dotted line indicating equality, divided by the total area of the right-angled triangle below the line of equality (i.e. 1/2). The higher the Gini index, the greater the concentration of wealth, and therefore the greater the economic inequality. The modeling of inequality is unidimensional here. It does not account for possible asymmetries in the distribution, is not decomposable and gives more weight to transfers around the mode of the distribution than to transfers

between extremes. These characteristics make the Gini index, according to general consensus among economists (Alacevich and Soci 2017), a very imperfect indicator for summarizing empirical distributions of wealth and income. Yet it is one that is indispensable because of its simplicity and its adoption by the majority of economic and social scientists, as well as in reports by international institutions on economic inequality.

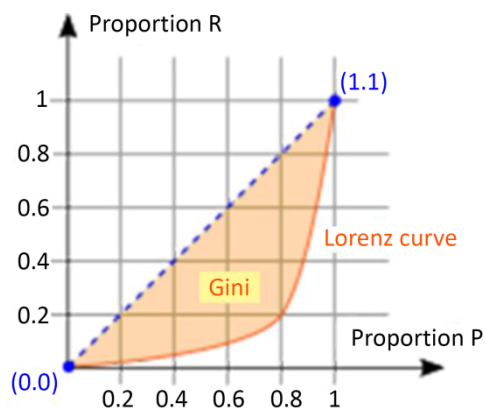


Figure 6.1. A graphic model of wealth inequality: the Lorenz curve, where R represents wealth and P represents the population. For a color version of this figure, see www.iste.co.uk/cottineau/inequalities.zip

Concentration indicators constructed from tax and wealth data have been revived by Piketty (2001). They correspond to particular points on the Lorenz curve: for example, the share of wealth held by the richest 1% or by the poorest 50%. These indices are simple to use and more intuitive than other aggregate measures, while describing the same trends in inequality. In this case, the concentration indices covary positively and significantly with the Gini index, the Atkinson index¹ and the 90:10 ratio, that is, the ratio between the value of the last and the first decile of wealth or income (Leigh 2007). In contrast, the use of the share of income captured

1. The Atkinson index (1970) is a normative inequality index that is based on a utility function representing the degree of society's aversion to inequality. This function makes it possible to give a different weight to the effect of a wealth transfer between rich and poor on the inequality index, depending on the direction of the transfer. The Atkinson index varies between 0 and 1 and is written as follows: $A = 1 - (y_e/\mu)$, where y_i = the income of an individual i ; $\mu = (\sum_i y_i) / n$; and $y_e = y | [nU(y) = \sum_i U(y_i)]$.

by the top 1% is questionable, according to Reynolds (2012), because these incomes are elastic (i.e. sensitive to economic conditions) and are calculated before redistribution². Their share in overall income therefore decreases sharply during crises, which does not reduce structural inequalities.

While the Lorenz curve and the Gini coefficient reflect the concentration of wealth in relation to an egalitarian reference model (the equivalence line between % population and % wealth), the Pareto model, which precedes them, is quite different. Resulting from the mathematical generalization of a series of empirical observations, the Pareto model for describing the distribution of wealth is that of a power law. Indeed, by collecting individual wealth figures in different countries of the world, Pareto showed that there is a decreasing and scalar mathematical relationship between the probability P for an individual to hold a wealth X greater than x , and the ratio between this level x and the minimum level of wealth x_m , such that: $P(X > x) = (x_m / x)^\alpha$, where α is a scale parameter interpreted by Pareto as an indicator of the level of inequality. Here, the geometric progression of wealth levels in the distribution is thus the empirical model used: a far cry from the egalitarian model of Lorenz and Gini. Paretian distributions are only compared to each other synthetically with α to determine an order of levels of inequality (much like the rank-size distribution in urban studies).

The Theil index applies the physical model of entropy and information theory to the distribution of economic quantities. This measure, which is broad enough to reflect inequality, segregation and information redundancy, is mainly used to decompose inequality into inter-group and intra-group variations. Thus, analyses show that national affiliation “explains” about two-thirds of inequalities between individuals in the world, but that this share has recently been decreasing with the considerable rise of giants such as China, Brazil and India (Sala-i-Martin 2006).

Compared to the study of economic inequalities resulting from the distribution of a single variable (wealth or income) within a population, the analysis of social, health or education inequalities, for example, involves putting two variables in relation to each other. The difference in means or concentration indices between groups are often used to illustrate differences in resources or access by social or demographic group, using statistical tests of significance, such as t-tests. For example, Kakwani et al. (1997) measure health inequalities by comparing the level of perceived health according to the income level in the Netherlands in 1980–1981.

2. This argument can be applied to other indices, although synthetic indices and reports are less sensitive to it.

They observe, by studying averages and standardized averages, that individuals report better health as their income decile increases. Based on a t-test between concentration indices similar to the Gini index (crossing the cumulative proportion of income and an ordinal health indicator), they confirm the existence of health inequalities to the benefit of the richest, even when the demographic structure of the sample is taken into account, and all the more so when the health variable is subjective.

Developed mainly by geographers, statisticians and sociologists to describe sociospatial segregation (Apparicio 2000), spatial inequality indices can be classified into three groups.

In the first group, we find spatialized versions of the inequality indices presented above, such as the Gini or Theil index, in which we replace the distribution of wealth population groups with the distribution of different social groups between geographical units.

In the second group, we find variations on Duncan and Duncan's (1955) dissimilarity index, which expresses the proportion of individuals who have to change geographical unit (by moving to another neighborhood, for example) in order to move from the observed situation to a situation of equitable distribution of the social groups considered in a given space (groups of different levels of education in a city, for example). This indicator of spatial inequality is the simplest and most frequently used measure, particularly for measuring ethnoracial segregation in American literature.

In the last group, there are indices developed by geographers such as Morrill (1991), Wong (1993), Reardon and O'Sullivan (2004), which include "spatial parameters such as distance between spatial units, area" (Apparicio 2000, §17) or the contiguity of spatial units in measuring inequalities. Although more complicated to calculate and more demanding in terms of spatial data and analysis, these indicators provide a more accurate account of the spatial structure of inequalities.

Box 6.1. *Spatial inequality indices*

All of the indices presented so far are a-spatial but they can be used to compare geographical units between them. Thus, inequality between regions and countries is most often analyzed by comparing the values of their indices (George 1981), possibly over time, or by breaking down the global Theil index into two components: international and intranational inequality (Sala-i-Martin 2006). Geographical space can be taken into account in inequality indices in two ways:

either by using spatial indices (Box 6.1) or by mapping inequalities and modeling them graphically (section 6.2.2).

6.2.2. Mapping socioeconomic inequalities and graphical modeling

The most basic form of spatialization of inter-individual inequalities is to map the aggregate values (e.g. mean, median, sum) of a social resource and to consider, in the manner of social inequalities, geographic units as groups whose (significant) differences in values indicate the presence of location-related inequalities. Examples of this form of representation can be found in most atlases dedicated to inequalities (e.g. Le Bras 2014), whether they are economic, gender, or health related. However, depending on the adequacy of the discretization choice for quantitative variables, the impression offered by the color patches on the map can be misleading or uninformative about the level and dispersion of inequalities represented (Bell et al. 2019). A second option for spatializing inequality is to map not the aggregate values but the value of an inequality index of the social resource in question, its distribution pattern, or the result of a typology within the geographic units represented. This is the case, for example, with the mapping of national Gini indices or the share of the wealthiest 1% (Figure 6.2), or with maps of socio-residential inequalities (e.g. François and Ribardi re 2004). In the latter case, the use of graphical models of spatial inequality is explicitly mobilized to interpret the maps. In particular, the authors describe the specialization of  le-de-France communes in the residence of more or less well-off households according to the sectoral model, identifying a “sector” of well-off communes from the center of Paris to the southwest of the region, and several sections specialized in the residence of poor households in the inner suburbs. The center-periphery model is used to distinguish between communes specialized in middle-class residence (pericentral areas), and the poor periphery of the eastern margins of Seine-et-Marne. Local discontinuities are also highlighted by changing the scale of analysis from the commune to the IRIS (*ilots regroup s pour l’information statistique*, that is, a homogeneous sub-municipal territorial unit with approximately 2,000 inhabitants).

As this example suggests, the description of spatial patterns lends itself to the use of spatial models of inequality, as well as to the analysis of spatial discontinuities. In the following section, we discuss the classical models of the organization of urban societies, popularized by American sociologists and geographers in the last century, to describe the process of segregation, before turning to the identification and interpretation of discontinuities in homogeneous areas.

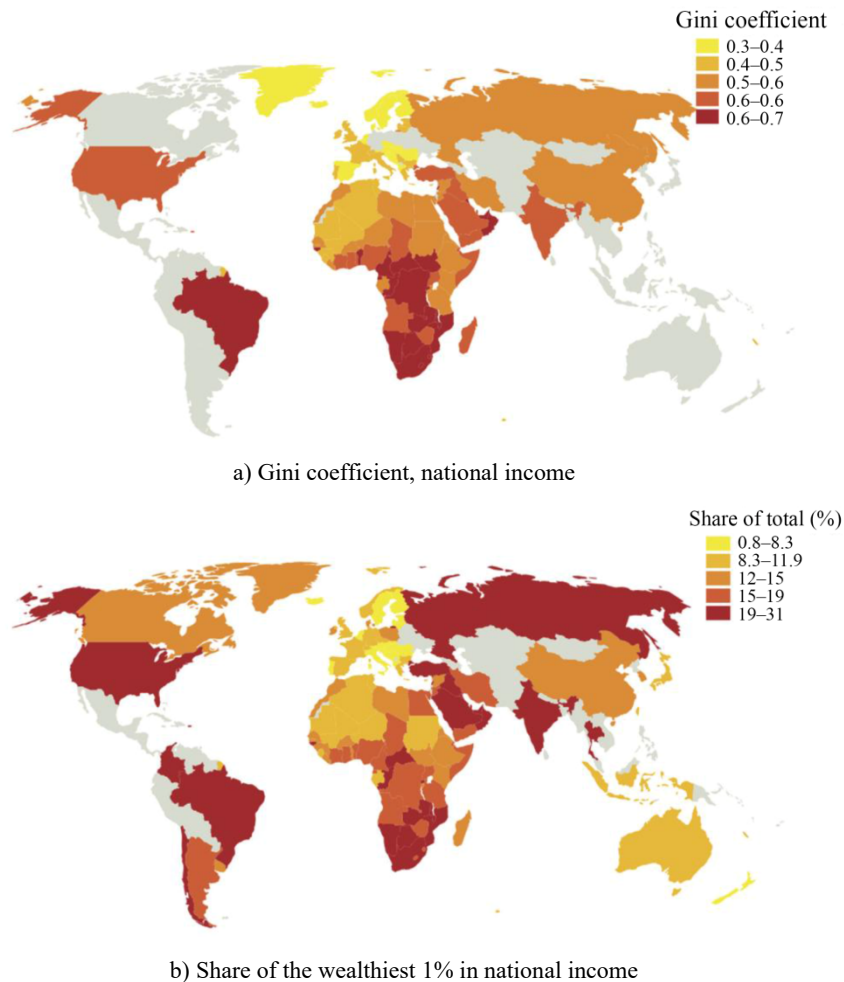


Figure 6.2. Global mapping of economic inequality indices (source: *World Inequality Database, latest available year 2020*³). For a color version of this figure, see www.iste.co.uk/cottineau/inequalities.zip

The zonal, or concentric, model describes the spatial distribution of a phenomenon in terms of distance to a center. In the original version by Park and Burgess (1925), the concentric zones represent relatively homogeneous areas in terms of population (by social and migratory origin) in Chicago. This model is used

3. Available at: wid.world.fr.

to represent the urban structure in which migrants circulate, from the center to the suburbs as they integrate into American and Chicago society. It has since been reused to describe the spatial distribution of urban functions (production in the center vs. residence in the suburbs) and urban forms (high-rise buildings in the center vs. detached houses in the suburbs). The sectoral model, attributed to Hoyt (1939), combines an axial dimension with the concentric model to represent the historical dependence of certain urban sectors on the presence of industries or wealthy residences along a transport axis (initially the railroad) during the development of a city. The multiple nuclei model proposed by Harris and Ullman (1945), suspending the assumption of isotropy of urban development, allows for more complex and variable patterns than Burgess's initial model, including secondary centers and areas of varying size.

The contribution of factorial ecology by Bell and Shevky (1955) was to combine these models to represent different dimensions of spatial organization of the population in a city. The concentric model would represent the different stages of the life cycle, with young adults living closer to the center and activity, families moving away from the center as the number of children increases, and the elderly returning to more central locations as their households shrink. The sectoral model would correspond to economic specialization, while the multiple nuclei model would describe concentrations of populations on an ethnic basis. This approach was operationalized by Berry and Kasarda (1977) on metropolitan areas by means of factor analysis, which also made it possible to rank the importance of these urban organization factors in the various cities.

In terms of spatial inequalities, these models aim to show that beyond the apparent complexity of cities and individual experiences, there are recurrent structural configurations in urban organizations. The proximity of industry to the homes of disadvantaged households, for example, indicates unequal access to a green environment as well as to the city as a whole (industrial and transport infrastructures often create barriers). These models are based on the presence of discontinuities between areas, but do not acknowledge them. On the contrary, some works (Grasland et al. 1997; François 2002; Roberto and Hwang 2015) put the notion of discontinuity at the heart of the analysis of spatial distributions of society. This notion allows for:

[...] understanding both the intensity of the separation that is forged in the neighborhood and the factors of social division of the space, because of the links that articulate them. Discontinuities, when they are defined in a multidimensional way (administrative and political limits, morphological or accessibility breaks, cognitive categories) and

when their trajectories in space and time are taken into account, therefore become a rich concept to read the urban and segregation in cities [...] At least two theoretical configurations can be identified: polarization and fragmentation. Polarization refers to simple partitions of the city where large homogeneous areas are separated by gradients. Taken to the extreme, this situation leads to a dual structure as described for many large cities and transposed without nuance to other cities. Fragmentation corresponds to a more complex spatial structure in which group concentrations are fragmented and dispersed in the city. Yet conventional measures of segregation cannot capture these spatial configurations, which are nonetheless decisive because of the contextual effects they generate. (Duroudier 2014, pp. 136–137)

These spatial models (polarization, sectorization, fragmentation, etc.) of inequality thus make it possible to describe the economic organization of societies in geographical space in a simplified manner, to identify recurring patterns – the center–periphery relationship in cities, a North–South gradient on an international scale between egalitarian and inegalitarian countries – as well as strong local discontinuities (between regions, neighborhoods, or residential blocks).

Although useful for describing the distribution of resources in a population and in geographical space, the analysis of (a-spatial) inequality indicators and spatial models of their distribution do not account for the effect of individuals' location on their economic and social characteristics. One way to model these effects is through multilevel and GWR statistical models.

6.3. Statistical regression models: estimating the effects of geographic location on inequality

We present the functioning of each regression model in detail before discussing its contribution and limitations for the study of inequalities in geography.

6.3.1. Description of the models

Two families of statistical models are discussed below: multilevel models and geographically weighted models. These are the most suitable for taking into account the geographical organization of individuals in the statistical explanation of their inequalities.

6.3.1.1. Multilevel models

Multilevel models are regression models applied to hierarchical data. They are increasingly used in social and educational sciences (Goldstein 1987). Unlike classical ordinary least squares (OLS) regression models, multilevel models have two variance components, one generally at the level of the individual (level 1) and the other at a higher level (level 2) in which the level 1 individuals are contained. These models combine individual explanatory variables (level 1) with higher level variables (level 2), whose effect applies to individuals belonging to the same higher level group. This higher level classically corresponds to a geographical level (the neighborhood, the community or the country) or to an institutional grouping (e.g. the class or the school in models of educational inequalities). The equation of a multilevel model is written as follows:

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

where i refers to a level 1 individual (e.g. a student), j refers to the level 2 set to which they belong (e.g. their school), y refers to the characteristic whose distribution we seek to explain (e.g. income, level of academic achievement, life expectancy), and x refers to the vector of lower level explanatory variables introduced into the model (e.g. the student's age, income, and parents' education). β_0 is the *intercept* at the origin, u_{0j} is the level 2 residual for set j and ε_{ij} the level 1 residual for individual i , with these two types of residuals being uncorrelated and each following a normal distribution. Finally, the vector of coefficients β_1 can be associated with a vector u_{1j} to differentiate the slope of the regression according to the membership of the individual i to a set j .

Depending on the implementation chosen, a multilevel model can thus allow you to vary the *intercepts* of the regression according to the groups j ($u_{0j} \neq 0$), to vary the slopes of the regression lines according to the groups j ($u_{1j} \neq 0$), or both. When applied to political, school and health data, these models can suggest neighborhood effects on important individual characteristics, including the intraclass correlation (ICC), which is the share of inter-individual variance attributable to top-level membership. Indeed, when the residuals of the geographical groups are significantly different, they make it possible to isolate a (geographical) belonging effect in the variation of the levels of school or health performance, thus suggesting that, all else being equal regarding individual variables, the place of residence plays a distinct role in the lives of individuals, through a composition effect, the action of peers or the more general physical and social environment (see Chapter 1). Examples of such results are given in section 6.3.2.

A variant of multilevel models estimates cross-level interactions. These cross-level interaction effects can measure the influence of higher level variables in the variation of regression slopes between higher level groups. The equation of a multilevel model with cross-level interactions is written as follows:

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

with:

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

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

where w_j represents the vector of top-level explanatory variables j (e.g. school size and specialization in the example of school inequality), γ_{00} corresponds to the intersection at the origin (*intercept*) common to all observations in the model. γ_{01} is the change in *intercept* of group j induced by a variation in the variable w . γ_{10} corresponds to the change in slope of the variable w induced by a variation in *intercept* at the group level. Finally, γ_{11} is the inter-level interaction effect, i.e., the change in slope of the individual variable induced by a change in slope of the higher level variable. Like the interactions between variables of the same level, the inter-level interactions allow us to identify the effects of variables that depend on other variables. In the field of inequalities, these interactions can be presented, for example, in the form of variable relationships between the level of education for women and men, for residents of areas with high unemployment rates and for others. Multilevel models with interlevel interactions – which are quite similar to variable slope models – therefore require working with sufficiently large samples, and with a sufficient proportion of total variance attributable to differences between higher level entities (ICC), to obtain significant effects (Mathieu et al. 2012).

6.3.1.2. Geographically weighted models

GWR models are regression models applied to localized data in which spatial autocorrelation is assumed. “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” (Brunsdon et al. 1996, p. 284). The value of a variable at a given point in space is thus assumed to vary systematically with the characteristics of the given point, as well as with the characteristics of neighboring points in geographical space. In other words, what happens at a point depends on the characteristics of the surrounding spaces. The technical implementation of geographically weighted

models aims precisely at delineating the “surrounding” space and quantifying its weight in the explanation. The equation of a GWR model is written as follows:

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

where y_i represents the characteristic y , whose distribution is sought to be explained (e.g. income, educational attainment, life expectancy) for an individual located at point i . x_i representing the vector of explanatory variables, β_{i0} represents the intersection at the origin for point i and ε_i , the residual at point i (the residuals being independent and distributed according to a normal distribution with a mean of zero). The central element of this model is the continuous neighborhood function $\beta(i)$. This function determines which observations are included in the regression for each point. Most applications of this model use a distance decay function around i to select these observations, but other functions are possible (Brunsdon et al. 1996). They share defining weight matrices between localized individuals. When the function restricts the neighborhood to very close observations, it generates a high variability of the estimated coefficients. On the contrary, a function that would systematically include all observations would not differ from an OLS regression model without a spatial component.

A GWR model will therefore produce as many estimates of *intercepts* and coefficients associated with the explanatory variables as there are localized individuals i . This model thus makes it possible to analyze the cases in which the spatial autocorrelation of the observations makes it possible to better explain the observed unequal patterns. They are particularly useful for revealing situations in which the direction and intensity of the relationship between two variables vary in different regions of the area. In addition, these continuous neighborhood function models make it possible to extrapolate the value of the coefficients associated with the explanatory variables to points not covered by empirical observations (Brunsdon et al. 1996). “To calibrate a GWR model at any one location, data is ‘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 – thus tying together the two main distinguishing features of spatial analysis” (Fotheringham et al. 2017, p. 1248). Their implementation, however, is more expensive (in computation, choice⁴, data, and interpretation), but is justified in

4. “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.

complex situations. The principles and differences between the classical, multilevel and geographically weighted models are summarized in Figure 6.3.

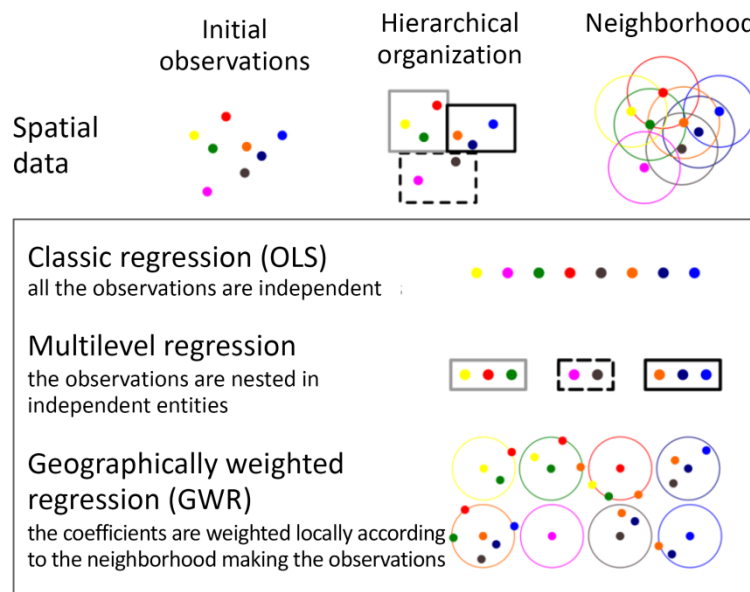


Figure 6.3. Data organization and principles of classical, multilevel and geographically weighted regression. For a color version of this figure, see www.iste.co.uk/cottineau/inequalities.zip

A multiscalar variant of GWR models has been proposed by Yang (2014). The MGWR method allows for a different scale to be associated with each explanatory variable, by varying the parameters of the neighborhood function (thereby expanding or contracting the area of influence of the neighbors on the estimate of the locally regressed quantity) for each variable. “That is, the bandwidths indicating the data-borrowing range can vary across parameter surfaces” (Fotheringham et al. 2017, p. 1249). The equation for an MGWR model is written as follows:

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

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).

where bw_j is the *bandwidth* parameter of the associated neighborhood function for explanatory variable j , and β_{i,bw_j} , the parameterized neighborhood function for the localized observation i . This model implements more degrees of freedom and is slower to calibrate (five to ten times slower in the simple simulations of Fotheringham et al. (2017)). However, it allows variables that represent processes acting at different scales to be included in the same regression by estimating the optimal spatial magnitude for each explanatory variable. In order to choose the most appropriate regression for each situation, Comber et al. (2020) recommend starting the analysis of a spatial dataset with a simple regression model such as OLS and then using a multiscale geographically weighted regression (MGWR) model. If the residuals of this regression appear not to be spatially autocorrelated, OLS is sufficient. If the estimated *bandwidths* for each variable are similar, then a geographically weighted regression (GWR) is more appropriate. If the residuals are spatially autocorrelated and the amplitudes are varied, then MGWR is the formalism that will give the best results.

6.3.2. Important results of the analysis of inequalities by statistic models

Using data from the well known U.S. Equality of Educational Opportunity survey, which stemmed from the 1966 Coleman Report, Borman and Dowling (2010) show, using a hierarchical model in which level 1 is the student and level 2 is the school, that 40% of the variance in student academic outcomes can be explained by the differences between schools. The ethnoracial composition (as a percentage of African Americans) of the school has a 1.3 times greater effect than being African American or not on academic outcomes, while the effect of average financial resources of the pupils' parents in a school are about three times greater than the resources of the student's own parents in predicting individual academic success. Because attendance at one school compared to another is generally based on residential location (see Chapter 1), these results suggest that the residential choices of affluent households (as well as the lack of choice among poorer households) significantly influence the level of societal and ethnic inequalities in education in the United States.

In a very comprehensive protocol including 400,000 students in the English public system in 2006, the different schools they attended during their schooling and the neighborhoods in which the schools are located, Leckie (2008) implements a multilevel model with random slopes. He finds that the strength of the negative relationship between school mobility and final exam scores depends strongly on the

timing and nature of school changes (compulsory between primary and secondary school, or non-compulsory). In this case, it is mainly students changing schools outside the primary–secondary transition, outside the summer vacations or without changing residence, who progress less quickly than others. “Accounting for pupil mobility also reveals that schools and neighbourhoods are more important than shown by previous analysis” (Leckie 2008, p. 537). The correlations between the top-level regression lines, on the other hand, help to qualify the synthetic results used to evaluate or rank schools, whether in terms of academic achievement or reduction in inequalities (social, gender, need) among students. Indeed, it shows that a school’s performance is not uniform for all types of school audiences. In this case, the correlation between the “school effect” of non-disadvantaged girls with good academic results and the “school effect” of disadvantaged boys with poor academic results is only 0.22.

By more explicitly spatializing their use of multilevel models, Roscigno et al. (2006) show, using a fixed-slope multilevel model, that students living in inner-city and rural areas of the United States have a systematically lower level of academic performance and a systematically higher probability of dropping out of school than their compatriots residing in peri-urban areas, given the same individual characteristics. The authors attribute these geographical differences to deficits in educational investment by families and schools lacking economic resources in the inner cities and the countryside, but also to the political and cultural orientations conveyed by the political and economic actors present, which would be less conducive to the educational success of all students.

In the field of health, Diez-Roux et al. (2000, p. 684) show that high levels of inequality in US states are detrimental to lower income individuals with respect to their risk of blood pressure and obesity (BMI index), but these risks are negatively and weakly associated with for individuals with high incomes. In other words, all else being equal, the economic inequalities measured in the different states reinforce the inequalities in health, with poor individuals more affected by the risks of a sedentary lifestyle, obesity and stress in the more unequal states, and on the contrary, wealthy individuals less at risk in these same states. These results reinforce the conclusions of Wilkinson and Pickett (2009), who conclude using simple multilevel models that individual incomes do not explain all the differences in health between individuals, but that economic inequalities (adversely) affect the health of all the geographical communities concerned (see Introduction). By associating the place of residence with the space used daily by more than 1,500 Île-de-France women, Vallée et al. (2010) show, using a multilevel model with interaction effects,

that women residing in neighborhoods with characteristics that do not favor the use of health care services (low average income and low density of practitioners, for example) tend to neglect cervical cancer screening more systematically than women from other neighborhoods, especially since their daily activity space is limited to their single neighborhood of residence.

Finally, in a paper discussing the incidence of the first wave of Covid-19 in the United States, Mollalo et al. (2020) draw on the results of GWR and MGWR models to show that the positive relationship between the Covid-19 incidence rate and the level of population inequality in US counties is most apparent “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” (Mollalo et al. 2020, p. 5), while the relation is more tenuous, if not absent, in the rest of the country. They find the same result for median household income level, which has a strong positive correlation with greater Covid-19 prevalence only in these areas of the country.

6.3.3. Limitations of the analysis of inequalities in geographical space by statistical models

Despite the type of strong results that statistical modeling of inequalities in geographical space produces, there are a number of criticisms of the use of these methods to highlight the effects of spatial and geographical context on educational or health inequalities between individuals.

First, although multilevel modeling can take into account the common membership of some individuals in neighborhoods, regions and schools, statistical regression is still a modeling framework that assumes the independence of observations considered at a given level (Figure 6.3). Thus, even when the hierarchical aspect of the data is taken into account, the higher level units are still considered independent. In particular, their spatial organization (the proximity between two neighborhoods, for example) remains excluded from the model. Multiscale geographically weighted modeling can compensate for this weakness when the residuals are effectively autocorrelated. For this type of modeling, however, the cost of implementation is sometimes prohibitive (Fotheringham et al. 2017), while the level of expertise required to evaluate the model and interpret its results leads some authors to publish incomplete, inadequate or erroneous analyses (Comber et al. 2020).

Second, there is strong endogeneity between the top and bottom levels, which is not accounted for by the model (Manski 1993). “Individual characteristics are often considered to be the property of individuals, whereas they are also determined by where they live. Similarly, the context whose characteristics can be measured at a given time, in a given territory, is strongly associated with the composition of the population residing there. It is therefore illusory (or somewhat artificial) to hope to distinguish perfectly between individual and contextual characteristics and to dissociate compositional effects from contextual effects” (Vallée 2009, §32). This endogeneity is added to the risk of attributing to one level a share of the effects of the other level by over- or underestimating the model (*ibid.*).

Third, Salgado et al. (2014) mention the homogeneity of the effects estimated by multilevel models, with level 2 variables applying uniformly across all level 1 individuals in the same way, whereas the group in which a peer effect applies – for example, in the school setting – would be the group of friends rather than the class or even the school as a whole. The use of regression models with varying slopes or with cross-level interactions may mitigate this limitation. However, the issue of multiple scales at which context effects operate is crucial. Models with more than two levels are certainly possible, but particularly costly and demanding. Moreover, the levels at which data are available are often arbitrary: “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” (Harris 2017, p. 442).

Finally, Salgado et al. (2014) emphasize that the choice of *proxies* used as variables at the neighborhood, regional, or school level can obscure the causal processes linking individuals to their geographic unit or institutional grouping. For example, when the school is introduced as a top-level variable in a model of individual school performance, it is difficult to know whether the effect being modeled is that of peers within the school, of teachers’ and families’ educational investment, or of economic and cultural resources available to students outside the school. Identifying specific variables for each mechanism and collecting corresponding data is often a challenging task (Hanushek et al. 2003). This limitation is not unique to multilevel or geographically weighted modeling, but rather to statistical modeling in general.

While statistical models describe the reality captured by empirical data and make it possible to identify privileged associations between variables without presuming

causality between them, simulation (in particular agent-based), by generating causal configurations, makes it possible to produce demonstrations of causal possibility, without it being possible to prove the identity between the simulated process and the empirical process having resulted in a similar configuration. These two approaches are therefore more complementary than alternative, as suggested by the studies that combine them. For example, Salgado et al. (2014) use the estimates from multilevel models on educational inequalities as a standard from which to estimate the quality of the agent-based model. In Cottineau's thesis (2014), variations in urban growth in the different regions of the former Soviet Union are identified by multilevel analysis, in order to build the generative mechanism that will be introduced into the multilevel model to represent the (causal) effect of different regional demographic regimes (level 2) on the demography of cities (level 1).

6.4. Simulation models to explain and illustrate the dynamics of inequalities in geographical space

Thus, whereas the models of inequality that we have seen so far have served to organize empirical data in such a way as to describe inequality or to isolate its factors and geographical scales, the models that we present here are designed to “recreate” in silico situations of inequality on the basis of causal processes applied to synthetic societies (i.e. artificial societies that exist only in the simulation). We restrict the type of simulation to two formalisms that are particularly well adapted to the consideration of geographical space: cellular automata and agent-based models, whose differences and common points are presented in Box 6.2.

In other words, we simulate a society of agents subject to rules of action based on hypotheses chosen by the modeler (according to empirical regularities and theoretical propositions). We then analyze the consequences of the interactions between agents and between agents and their environment on the level of inequalities, and their distribution in space, in order to identify which mechanisms are “candidates for explaining” the inequalities observed in real societies. These candidate mechanisms correspond to the rules of action that allow us to simulate the artificial society closest to real societies (according to a series of pre-determined indicators). We present three examples of agent-based models simulating inequalities in geographical space: a model of segregation (Schelling), two models of unequal access to localized resources and three models of eviction/gentrification.

Cellular automata (CA) were developed in the 1940s in Los Alamos by researchers in mathematics and computer science (Ulam, Von Neumann). Considering the computing power of computers at the time, their operation is quite simple. They are presented as a grid of cells (typically squares) with two properties: each cell contains a value of its own (whether it is a color, a type or a number) and is affected by the value of the cells in its neighborhood (whether it is the four cells located above, below and on each side – Von Neumann’s neighborhood – or the eight cells located surrounding the cell – Moore’s neighborhood). The animation of a cellular automaton comes from the implementation of two types of action rules: the rules of change of the cell according to its own value, and the rules of change related to the values of the neighboring cells. Although simple in their principle and implementation, cellular automata allow the simulation of a wide range of geometric patterns and spatial dynamics. For example, with only two possible states as cell values and five change rules, Conway’s (1970) “game of life” produces a range of patterns, some of which are stable, some oscillatory, some dynamic, and some ephemeral, depending on the starting spatial configuration (Gardner 1970). In geography, these simulations have been popularized by Batty (1997), White (1998) and Frankhauser (1998) for their ability to simulate certain structures (notably fractals) of urban growth.

Agent-based models (ABMs) are extensions of cellular automata. Indeed, the “agents” of these models can be immobile cells arranged on a grid, but can also be represented by point, linear or surface mobile entities. In most of the agent-based models developed in geography since the 1990s (Bura et al. 1996; Batty et al. 1998), mobile agents interact with each other and with a spatial environment that can itself be a cellular automaton. Rules of change are then defined for mobile agents: (1) according to their own dynamics; (2) according to their interactions with other agents; (3) according to their spatial situation (interaction with cells, for example). The main advantage of agent-based modeling compared to other forms of simulation is that it allows the representation of heterogeneous, mobile agents with objectives and potentially with memory. Its main disadvantage results from this flexibility: it is indeed expensive to evaluate the simulated results, in terms of robustness and sensitivity.

Box 6.2. Cellular automata and agent-based models

6.4.1. Agent-based models of spatial segregation: Schelling and its variants

Schelling’s model (1971, 1978) is one of the most well known examples of agent-based simulation in the social sciences. Designed by Schelling in the 1960s using a checkerboard, and then transcribed by his numerous successors into digital form, this very simple model involves two qualitatively different populations (often symbolized by a different color or shape) moving in a very abstract housing market.

The objective of the agents is to “reside in a dwelling” (i.e. to settle on a location corresponding to a cell of the checkerboard or the numerical environment) whose neighbors satisfy a diversity criterion chosen for the whole population. In this case, the agents are given a percentage that corresponds to the maximum share of their environment that neighbors of different color or shape can represent without the agents getting upset (i.e. they do not change their state in the simulation). If the proportion of different neighbors exceeds this percentage, then the agent “decides” to change its cell of residence and looks for one that is free, either randomly in most computer-implemented simulation models (e.g. the model available by default in the NetLogo software example library; Tisue and Wilensky (2004)), or as close as possible to the cell it leaves in the original version of the model. The simplest version of Schelling’s model consists of a square-shaped, small “city” (a few dozen cells organized in a checkerboard pattern). The capacity of the cells is generally uniform, with a maximum of one agent per cell, and only part of them is occupied at the beginning of the simulation (according to a percentage that is an important parameter of the model), which allows disgruntled agents to move to unoccupied cells during the simulation. The main conclusion of this very simple model in the field of modeling inequalities in geographical space is that the interplay of local interactions and the reduced perception of the agents lead to the production of a collective situation of spatial segregation, even when the agents are individually relatively tolerant (they can, for example, accept being surrounded by neighbors who are 60% or 70% different from them). Indeed, as soon as 30% intolerance is reached, the model converges towards segregated patterns (Gauvin et al. 2009). The presence of many vacant spaces further favors this separation of groups in space. Although Schelling’s model does not claim to represent the mechanism by which the housing market promotes urban segregation, it does tend to prove that it is not necessary to introduce conscious segregation or discriminatory mechanisms between populations to produce patterns of inequality in the distribution of individuals in space.

This model has been the subject of numerous criticisms and as many variations in implementation. Note, for example, the variations introduced in the 1990s by Portugali et al. (1994), who vary the maximum tolerance levels between the two simulated population groups, or by Benenson (1999), who introduces differences in economic status between agents, which accentuates or balances the effects of individual preferences in the location of individuals. Flache and Hegelsmann (2001) introduce irregularity into the grids and partitions of the model space, without leading to significant changes in the behavior of the model. Benenson and Torrens (2002) introduce the geometry of residential buildings and their capacity in the model, in order to better take into account the proximities between agents and the morphology of the city, whereas Laurie and Jaggi (2003) introduce differences in

scope in the perception of agents. Fossett and Dietrich (2009) modify Schelling's classic model to measure the sensitivity of the shape and size of city cells on the level of segregation obtained. In 2012, Hatna and Benenson (2012) produced a version of Schelling's model allowing for the simulation of a population composed of a majority and a minority, rather than two equal-sized groups, while Banos (2012) introduced reticular structures between cells into the model to compare the effect of network shapes on the observed levels of segregation at the end of the simulation. In this case, the author shows that simulation environments following fractal networks favor group segregation in the space. Spielman and Harrison (2014) vary the urban density and location preferences of agents as a function of their individual characteristics, using empirical data from the 19th century. Raimbault et al. (2019) focus on the sensitivity of the model to initial spatial conditions. They implement Schelling's model on "cities" whose spatial distribution of densities corresponds to configurations close to those of European cities, that is, non-uniform and varied (monocentric, polycentric, discontinuous cities). They show that the effect of the model's parameters (i.e. the individual tolerance level of agents and the city's vacancy rate) is the same regardless of the city's density profile, although discontinuous and polycentric cities favor systematically higher final segregation levels. Finally, Forsé and Parodi (2019) reveal the fragility of the model's results by revealing an implicit assumption: "Schelling's model neglects [...] moves made for reasons other than intolerance [...] Taking these other moves into account, we would spontaneously observe the disappearance of homogeneous neighborhoods that are only transitory formations. Schelling's conclusion is therefore very fragile because, basically, he stops the film at the moment when, by accident, homogeneous neighborhoods have formed" (Forsé and Parodi 2019, p. 101).

6.4.2. Two agent-based models of the emergence of inequalities linked to the distribution of resources in space

The Sugarscape model is one of several environment-society interaction models that help explain the formation of inequalities between individuals and between settlement nuclei, based on artificial societies embedded in their physical environment (Schmitt and Pumain 2013). Initially proposed by Epstein and Axtell (1996), who were pioneers in social simulation, agent-based models are designed to study the interactions between plausible mechanisms of emergence of the economy and differentiation between individuals in a society. Its configuration is very simple and can be completed by more elaborate versions of the model. Agents evolve in a landscape characterized by the non-uniform presence of a renewable resource (sugar). They can collect this resource, store it or use it for their survival. This collection depends on their perception, which varies from one individual to another

(some see further than others). The amount of sugar needed for survival (or metabolism) also varies from one individual to another. The uneven distribution of sugar across the landscape, combined with differences in individual vision and metabolism, is enough to produce an economically unequal society, in which levels of wealth (measured in sugar) vary greatly between individuals, some of whom have a very large amount of sugar and others not enough to survive. The measure on which the evaluation and comparison of the model outputs is based on is therefore generally, as in the NetLogo implementation of Li and Wilensky, the Gini index, the measure which here summarizes the inequalities of sugar-wealth between agents.

As with Schelling, the simplicity of the initial model has led to a significant number of additions and complementary mechanisms. This is particularly true of the authors of the original model, who have adapted Sugarscape from its most simple form to a model in which the behavior of agents approximates the empirical diversity of interactions between members of a society. “Over time, the repertoire of agent behaviors grows to include movement, resource harvesting, sexual reproduction, combat, culture transmission, exchange, trade, inheritance, credit, pollution, disease spread, and immune resistance” (Schmitt and Pumain 2013, §8). Comparing the impact of the initial spatial configuration in Schelling’s and Sugarscape’s models, Raimbault et al. (2019) show that Sugarscape’s simple implementation is highly sensitive to the initial distribution of the resource in the emergence of inequalities between individuals, to the point that the model outputs can be modified more strongly by changing the initial spatial conditions than by changing the values of the model parameters (minimum and maximum sugar production capacity of the cells, number of agents, etc.).

In the Auchincloss et al. (2011) model, the inequalities modeled between individuals are those related to their diet. They emerge as a result of income differences among individuals in a segregated environment, through the supply of healthy food in shops which adapt to the local residents’ demand. The model effectively simulates two types of agents: households and food stores. Households are distributed in two categories of equal size: rich households and poor households. They have a preference (continuous between 0 and 1) for healthy food, which is distributed either randomly or conditionally to their category (rich/poor), depending on the simulated scenarios. The food stores are distributed in two equally sized categories: those selling healthy food and those selling unhealthy food. Half of the stores are simulated as selling their food cheaply, and the other half selling their food at high prices, with the price being assigned either randomly or conditionally to their category (healthy/unhealthy) depending on the simulated scenarios. The space is simulated in a very abstract way, by a torus (i.e. a space folded on itself to remove edge effects) of 50×50 square cells. Households are randomly distributed or

isolated by income level in two distinct zones of the space according to the scenarios, while stores are randomly distributed or isolated by the type of food sold (healthy or not) in two distinct zones of the space according to the scenarios. No agent moves between these zones during the simulation. The dynamics of the model revolve around the relationship between households and stores. Households evaluate the stores they visit based on price, accessibility and the match between the food sold and their preference, and distribute their spending among the stores that score highest. Stores respond to household spending by closing if the number of customers is too low, while new stores (possibly with different prices and food) may replace them at the same locations. Auchincloss et al. (2011) show, using this model and different scenarios, that in the absence of other factors, income-related nutrition inequality results from the segregation of wealthy households and healthy food stores in separate areas from poor households and unhealthy food stores. When households have a preference for healthy food, regardless of income level, the diets of the poorest appear consistently more healthy than in the other scenarios, although inequality persists between rich and poor. Finally, the authors show that only favorable preferences and accessible cheap stores can overcome the differential introduced by household segregation, thus demonstrating the structural importance of agents' inscription in their environment on food inequalities in relation to possible changes in their behavior.

6.4.3. Three models simulating spatial relegation dynamics

The last type of simulation model of inequalities between individuals in geographical space that we wish to present in this chapter concerns the process of spatial relegation, that is, the forced displacement of poor populations from certain urban neighborhoods, following the chosen displacement of more affluent social fractions who, by investing in mixed, poor and/or degraded neighborhoods, contribute to the physical urban renewal, and subsequently to the social replacement of these neighborhoods. Three models are focused on this dynamic that modifies urban space through the interaction between economically unequal individuals.

The first model (O'Sullivan 2002) is a cellular automaton, that is, a model in which the geographical location of the agents is fixed and resembles a cell. The author models the spatial relegation process from the point of view of buildings and housing, focusing on the *rent gap* hypothesis (Smith 1979). The rent gap, or the gap between potential land rent and current land revenues, is one of the elements of economic theory involved in explaining the urban and social renewal of urban neighborhoods. O'Sullivan explores the geographical and temporal aspects of these economic mismatches between the use value of housing and its potential value, due

to its location. His model is more refined in terms of spatial representation than most conventional cellular automata, which are constructed as regular grids of identical square cells. The cells it represents are the centroids of buildings, whose neighborhoods are constructed according to a proximity graph rather than by contiguity. Each cell has four possible states: for sale, not for sale, looking for a tenant or rented. The transition between these four states is conditioned by the characteristics of the building (particularly its degree of deterioration) and the income of the occupants, the demand available for buying and renting, as well as by an indicator of the status of the neighborhood, but also contains an element of stochasticity (i.e. chance). The model is able to reproduce the dynamics of physical and economic deterioration of the neighborhood, followed by gentrification, using the example of Hoxton in East London. Its representation of inter-individual inequalities in geographical space is therefore rather mechanical and impersonal, since the dynamics of the model are played out at the level of the buildings and their objective characteristics. The author suggests, however, that the neighborhood status parameter is crucial in determining the dynamics of the model, since it updates at each point in time the composition of the neighborhood according to the income of the residents present at a given moment, and thus determines the income of potential new owners and renters. According to O'Sullivan himself, this parameter should rather be treated as an exogenous factor in connection with the representation of other neighborhoods in the city for a more realistic simulation: "A fairer application would imply that the neighborhood presented here in a larger model of the urban system" (O'Sullivan 2002, p. 270). Indeed, this parameter can be interpreted as an element of neighborhood reputation, that is, an important element of the social dynamics of gentrification in theory.

The second model is a hybridization between a cellular automaton representing the dwellings of a neighborhood according to a regular grid of square cells, and an agent-based model representing individuals moving between these dwellings and "animating" the social, economic and urban space. Torrens and Nara (2007) focus their attention on the functioning of the housing market and represent four scenarios of gentrification that they compare with data from a Salt Lake City neighborhood. The first scenario is that of no change (business as usual), the second represents a change from the point of view of demand (in modeling, the arrival of a group of "gentrifiers" with different economic and social behaviors from those of the initial population), the third is a change from the point of view of supply (by taking into account the rent gap and the undervaluation of the real estate stock at a given time), and the fourth illustrates a combination of scenarios two and three. The rule of action for agents is to decide whether or not to move, to select one of the three housing markets in the city if necessary and then choose a property to move into. The factors influencing this decision depend on their preferences (for a type of

dwelling, its size, etc.), the characteristics of the home, the characteristics of the neighborhood (accessibility to a shopping mall, a freeway off-ramp, the city center; as well as the surrounding vacancy rate) and their own resources. Torrens and Nara (2007, p. 352) show that only the last scenario, including supply and demand dynamics, can reasonably explain the gentrification dynamics observed in Salt Lake City. Here, the modeling of inequalities in geographical space incorporates competition for neighborhood housing by households, both local and newcomers, with unequal resources. The attractiveness of the housing itself depends on its location with respect to fixed resources (such as downtown and transport routes).

The last model, devoted to the forced displacement of populations in an urban space, is an agent-based model developed by Boeing (2018). The author is interested in the effect of the correlation between individual preferences and economic status in the displacement of low-income groups from areas that are suitable for them. Despite its 13 parameters, the model is much simpler than the previous two, and is designed as a theoretical model rather than as one that can be applied to a particular case. Agents are categorized into three levels of wealth (poor, medium, rich), while the cells available to them are characterized by two dimensions – on the one hand, the presence of abstract amenities functioning as club goods (i.e. whose benefit applies equally to all those who have access to them, without decreasing its stock through the increase in the number of members) and, on the other hand, their level of density. The preference of poor agents is to reside near the amenity they exploit for a living, while the preference of rich agents is to reside near densely populated areas. The land rent increases with the number and income of agents present in a cell, leading occupants who cannot afford to pay the rent to change location. The model leads poor households to form dense clusters around amenities that attract rich households. The dense presence of rich households in turn pushes poor households to the periphery of the cluster, and away from the amenities on which they depend, resulting in an inefficient state of the system. Much like Schelling's model, Boeing's model allows for the reproduction of a stylized fact of inequality between individuals in and through geographical space by simulating a mechanism sufficient to reproduce an empirically observed dynamic, although globally unrealistic by itself.

6.5. Conclusion

This overview of the modeling of inter-individual inequalities in geographical space has made it possible to take into account the diversity of the objectives of the possible models, the diversity of the methodological approaches implemented to meet these objectives, as well as the multiplicity of points of view on the same

question (spatial relocation, for example). We have noted that the different families of modeling (cartographic, statistical, generative) should be used in a complementary rather than competing manner. Indeed, the cartographic models synthesize the empirical regularities of the geographical distribution of individual inequalities and suggest factors to be introduced into the statistical models. These describe the relative influence of measurable factors on the distribution of inequalities, and their results feed into the construction and plausibility assessment of generative models. The theoretical results of the simulation can in turn enrich the set of hypotheses to be tested empirically. According to Manzo (2005, p. 39), “the recent literature is beginning to outline a type of quantitative empirical sociology in which variable-based analysis describes, mechanism-based modeling (constructed in terms of methodological individualism) explains, and the simulation dynamically animates (and tests) the supposed mechanisms underlying the observed statistical relationships”. In the same way, the variables and mechanisms selected by the different studies are often more complementary than competitors in explaining social and economic inequalities. Thus, some models emphasize the physical aspect of the spatial environment, whereas some other models emphasize its demographic, economic or social composition, to explain inequalities in geographical space. It seems that the diversity of social reality favors a combination of these explanatory registers, which vary according to the case and the time.

6.6. References

- Alacevich, M. and 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 [Online]. Available at: <https://discovery.ucl.ac.uk/id/eprint/225/1/paper4.pdf>.

- Bell, W. and Shevky, E. (1955). *Social Area Analysis: Theory, Illustrative Application, and Computational Procedures*. Stanford University Press, Stanford.
- Bell, A., Hartman, T., Piekut, A., Rae, A., Taylor, M. (2019). *Making Sense of Data in the Media*. Sage Publishing, Thousand Oaks.
- 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.
- Benenson, I. and Torrens, P.M. (2004). Geosimulation: Object-based modeling of urban phenomena. *Computers, Environment and Urban Systems*, 28(1/2), 1–8.
- Berry, B.J.L. and 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) [Online]. Available at: <https://doi.org/10.3390/urbansci2030076>.
- Borman, G. and 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., et al. (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 Thesis, Université Paris 1 Panthéon-Sorbonne, Paris.
- Cottineau, C., Chapron, P., Le Texier, M., Rey-Coyrehourcq, S. (2020). Modélisation territoriale incrémentale [Online]. Available at: <http://doi.org/10.1002/9781119687290> [see Chapter 4 for the English version].
- 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. and 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 [Online]. Available at: <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 [Online]. Available at: <https://doi.org/10.18564/jasss.3993>.
- Epstein, J.M. (2008). Why model? *Journal of Artificial Societies and Social Simulation*, 11(4), 12 [Online]. Available at: <http://jasss.soc.surrey.ac.uk/11/4/12.html>.
- Epstein, J.M. and Axtell, R. (1996). *Growing Artificial Societies: Social Science from the Bottom Up*. MIT Press, Cambridge.
- Flache, A. and 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 [Online]. Available at: <http://www.soc.surrey.ac.uk/JASSS/4/4/6.html>.
- Forsé, M. and Parodi, M. (2019). Retour critique sur le modèle de ségrégation urbaine de Schelling. *Émulations-Revue de sciences sociales*, 31, 91–104.
- Fossett, M. and 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.
- 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.
- 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. and Ribardière, A. (2004). Qu’apporte l’échelon infracommunal à la carte des inégalités de richesse en Île-de-France ? *MappeMonde*, 3(75), 1–8 [Online]. Available at: <http://mappemonde-archive.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, 120–123.
- 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.
- Goldstein, H. (1987). *Multilevel Models in Educational and Social Research*. Oxford University Press, Oxford.

- Grasland, C., François, J.-C., Brunet, R. (1997). La discontinuité en géographie : origines et problèmes de recherche. *L'Espace géographique*, 26(4), 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.
- Harris, C.D. and 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. and 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.
- 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. and Jaggi, N.K. (2003). Role of “vision” in neighbourhood racial segregation: A variant of the Schelling segregation model. *Urban Studies*, 40(13), 2687–2704.
- 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, Bristol.
- van der Leeuw, S.E. (2004). Why model? *Cybernetics and Systems*, 35(2/3), 117–128.
- 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 [Online]. Available at: <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. and 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 XX^e 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.
- 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.
- Reardon, S.F. and 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.
- Roberto, E. and 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*, 121(2), 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.
- 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. and 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.
- 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. and 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.

- Tisue, S. and Wilensky, U. (2004). Netlogo: A simple environment for modeling complexity. In *International Conference on Complex Systems*, 21, 16–21.
- Torrens, P.M. and 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). *Cybergeo: European Journal of Geography*, 477. doi:10.4000/cybergeo.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 healthcare activities: The case of cervical screening in the Paris metropolitan area (France). *Health & Place*, 16(5), 838–852.
- White, R. (1998). Cities and cellular automata. *Discrete Dynamics in Nature and Society*, 2(2), 111–125.
- Wilkinson, R. and Pickett, K. (2009). *The Spirit Level: Why Greater Equality Makes Societies Stronger*. Penguin Books, London.
- Wong, D.W.S. (1993). Spatial indices of segregation. *Urban Studies*, 30(3), 559–572.
- Yang, W. (2014). An extension of geographically weighted regression with flexible bandwidths. PhD Thesis, School of Geography and Geosciences, University of St. Andrews, Fife [Online]. Available at: <http://hdl.handle.net/10023/7052>.