
**Towards a Multilevel Analysis
in Social Sciences/**
*Vers une analyse multi-niveaux
en sciences sociales*

Daniel COURGEAU

The research on relationships between different biodemographic characteristics, seems *a priori* easier to undertake when using individual data, obtained from surveys or specific population registers. In this case, the analysis is usually undertaken by using logit, probit, log linear or regression models with cross-sectional data, or by using event history analysis with longitudinal data. However, the relative scarceness of such surveys and population registers or the wish to put into evidence relationships at a different level of aggregation, may lead the researcher to work on aggregate data sets, considered at different levels: parishes, communes, regions, etc.

This occurs particularly when, in order to get the essential data for the analysis, it is necessary to work on different data sets: vital registration, censuses, and specific surveys. In this case it is no longer possible to reconstruct individual data, but it is always possible to analyse the relationships between rates or indexes which are estimated on each of the considered areas, using different regression models.

The use of one or another kind of data is grounded on different hypotheses, which we now have to take into consideration before approaching a multilevel analysis.

Distinction between individual and aggregate level data

When the analysis is undertaken at the individual level, its aim is to give a model for a demographic or biologic behaviour related to different characteristics of individuals which are supposed to affect such behaviour. Underlying this aim is the hypothesis that only individual characteristics will affect personal behaviour. It is easy to see how such an hypothesis can be found guilty of the *atomic fallacy*, as we miss the context in which individual behaviour occurs. In fact, such a context may play a role on individual behaviour and it seems misleading to consider an individual who is acting according to his own characteristics without consideration of the restraints imposed by the society and the environment in which he is living.

Conversely, when the analysis is undertaken at the aggregate level, such an approach at least permits one to take into account the effect the characteristics of the studied groups may have on group behaviour. The aggregate characteristics that are measured, can approximately represent different restraints that the diverse groups may impose on their members. On the other hand, such an approach lays one open to the charge of what is usually called *ecological fallacy* which fails to recognize that individuals do act, not aggregates. In this way an ecological correlation in which the statistical object is a group of persons, is almost certainly not equal to an individual correlation in which the statistical object is indivisible. This leads to an aggregation bias of

La recherche des relations qui existent entre diverses caractéristiques biodémographiques, semble *a priori* plus aisée si l'on dispose de *données individuelles*, recueillies par des enquêtes ou des registres spécifiques. L'analyse de ce type de données est dans ce cas parfaitement réalisable à l'aide de modèles logit, probit, log-linéaires ou de régression, lorsque l'on dispose de données du moment, et à l'aide de modèles biographiques lorsque l'on dispose de données longitudinales. Cependant, la relative rareté de telles enquêtes ou le désir de mettre en évidence des relations à un niveau différent, peuvent conduire le chercheur à travailler sur des *données agrégées*, à différents niveaux possibles : communes, paroisses, régions, etc. Cela peut se produire en particulier lorsque, pour disposer des données nécessaires à l'analyse, il faut travailler sur des sources différentes : état civil, recensements, enquêtes particulières. Il n'est plus possible, dans ce cas, de reconstituer les données individuelles, mais on peut encore analyser les relations entre les taux ou indices calculés sur chacune des zones considérées, grâce à divers modèles de type régression.

L'utilisation de l'un ou l'autre type de données est basée sur des hypothèses différentes qu'il nous faut maintenant considérer avant de passer à une analyse multi-niveaux.

Distinction entre données individuelles et données agrégées

Lorsque l'on travaille au niveau individuel, l'objectif va être de modéliser un comportement démographique ou biologique en fonction des diverses caractéristiques de l'individu qui sont supposées jouer sur ce comportement. Sous-jacente à cet objectif, se trouve l'hypothèse selon laquelle seules les caractéristiques propres à chaque individu vont jouer sur son comportement. On voit facilement comment cette hypothèse risque de conduire à ce que l'on appelle *l'erreur atomiste*, car on ignore dans ce cas le contexte dans lequel les conduites humaines se produisent. En fait, ce contexte doit certainement jouer sur les comportements individuels et il paraît fallacieux d'isoler l'individu agissant en fonction de ses propres caractéristiques sans faire intervenir les contraintes imposées par la société et le milieu dans lequel il vit.

A l'inverse, choisir de travailler au niveau agrégé permet, en partie du moins, de tenir compte de l'effet que les groupes étudiés peuvent avoir sur les comportements. Les caractéristiques agrégées que l'on mesure, peuvent en effet représenter de façon approchée un certain nombre de contraintes que les diverses sous-populations vont imposer à leurs membres. En revanche, cela risque de conduire à ce que l'on a coutume d'appeler *l'erreur écologique* lorsque l'on cherche à déceler un effet des caractéristiques individuelles sur le comportement étudié. Ainsi, une corrélation calculée sur des données agrégées est presque sûrement différente de la même corrélation calculée sur des données individuelles. Cela conduit à un biais

individual level relationships. However, such a bias may be more especially reduced as the within-area variance of each characteristic is lower than the between-area variance, i.e. the ecological one. Usually, such an hypothesis is far from being verified, so that it is always difficult, to draw individual conclusions from aggregate data when individual data are not available (which is unfortunately often the case).

In view of this dilemma, some researchers recently explored the possibility of working simultaneously on different levels of aggregation (Jones, 1993; von Korff *et al.*, 1992; Courgeau, 1995). Such an approach will cause us to re-examine the problems we met when working at only one aggregation level and to display the hypotheses necessary to undertake a *multilevel analysis*. We will show, hereafter, some new tools which are being used to carry out such an analysis.

Problems and hypotheses for multilevel analysis

As soon as one tries to undertake an analysis introducing different levels of aggregation, a number of distinctions between response variables and predictor variables are to be introduced. A first kind of variable corresponds to *individual characteristics* – biological as well as demographical – blood pressure, height, birth of a child, death of an individual, etc. In this case, it is easy to see how to aggregate such characteristics to get a new variable at any aggregation level: mean blood pressure, mean height, number of births, or birth rate, number of deaths or death rate, etc., measured on any arbitrary area. However, there are other characteristics that have no corresponding attribute at the individual level as *contextual properties*: population density, for example. These characteristics however can be aggregated at different macro-levels: population density of communes, parishes, regions, etc. Finally, there are some other *collective characteristics* which are defined at a given level of aggregation. For example, the hardness of the water is identically defined in its whole drainage basin and affects the whole population living in this district. It constitutes a measure of the individual exposure of all the members of the district. If we try to change the geographical areas, it may always be possible to define a mean hardness of the water, on weighting by the different subpopulations submitted in each different basin. Yet such a measurement seems too arbitrary and of little interest to be considered, so that it seems necessary to consider only the level at which it is defined.

We have also to draw a distinction between *biological characteristics* of individuals, for which an individual measurement will be the best one, and other *socio-demographic characteristics*, for which there is no reason to choose between an individual measurement or a more aggregate one. As a matter of fact, to analyse socio-demographic characteristics, event

d'agrégation si l'on veut raisonner au niveau de l'individu. Ce biais est d'autant plus réduit, que la variabilité des caractéristiques entre individus d'un même groupe est inférieure à la variabilité entre les groupes considérés, cela pour toutes les caractéristiques. On voit facilement que cette condition est en général loin d'être vérifiée et qu'il est toujours délicat, lorsque l'on ne peut disposer de données individuelles (ce qui est malheureusement souvent le cas) de tirer des conclusions sur les individus à partir de données agrégées.

Devant ce dilemme, certains chercheurs ont exploré récemment la possibilité de travailler simultanément à divers niveaux d'agrégation (Jones, 1993 ; von Korf *et al.*, 1992 ; Courgeau, 1994). Une telle approche va nous conduire à revoir les problèmes que l'on rencontre en travaillant à un seul niveau d'agrégation et à montrer les hypothèses à faire pour réaliser une *analyse multi-niveaux*. Nous verrons ensuite un certain nombre de nouveaux outils qui se sont mis en place pour réaliser une telle analyse.

Problèmes et hypothèses pour une analyse multi-niveaux

Dès que l'on cherche à réaliser une analyse faisant intervenir plusieurs niveaux d'agrégation, un certain nombre de distinctions entre les traits à expliquer ou entre les traits explicatifs paraissent nécessaires. Un premier type de traits correspond aux *caractéristiques individuelles* tant biologiques que démographiques : la pression sanguine, la taille, la naissance d'un enfant, la mort de l'individu, etc. On voit qu'il est facile dans ce cas d'agrèger ces caractéristiques pour obtenir une mesure à n'importe quel niveau d'agrégation : pression sanguine moyenne, taille moyenne, nombre de naissances, nombre de décès, etc., mesurés sur une zone arbitraire. Cependant, d'autres caractéristiques sont plus *synthétiques* : densité de population, par exemple. Aucune caractéristique individuelle ne leur correspond, mais ces caractéristiques peuvent être agrégées à différents niveaux macro : communes, paroisses, régions, etc. Enfin d'autres *caractéristiques collectives* sont définies à un niveau d'agrégation donné. Ainsi, la dureté de l'eau locale, est définie de façon identique dans tout son bassin de distribution et affecte l'ensemble des résidents d'un même district. Elle constitue dès lors une mesure de l'exposition individuelle de tous les membres de ce district. Si l'on change de découpage géographique, on peut toujours définir une dureté moyenne de l'eau, pondérée par les diverses populations soumises au risque. Mais on voit facilement l'arbitraire et le peu d'intérêt d'une telle mesure.

Nous devons également opérer une distinction entre les *caractéristiques biologiques* des individus, pour lesquelles on peut penser que la mesure individuelle fournit la meilleure approche, et les *caractéristiques socio-démographiques*, pour lesquelles on peut penser qu'il n'y a pas lieu de privilégier une mesure individuelle par rapport à une mesure plus agrégée.

history analysis can explain a time dependant individual characteristic (migration rate, fertility rate for a given rank, mortality rate, etc.) by different other characteristics of the same individual (to be a farmer, a foreigner, etc.) which can also be time dependant. Similarly, econometric regression models can use aggregate characteristics. For example, regional out migration rates can be explained by the percentages of individuals in different situations (percentage working in agriculture, percentage of foreigners, etc.)

Consequently, it appears that an important problem in multilevel analysis will be to find what the relationships between these different models are and what new contribution can be taken from such a synthetic approach compared with event history analysis and regression analysis undertaken separately.

Under these conditions, it may be agreed to unify methods of analysing biological data and socio-demographic characteristics. In order to explain an individual characteristic, it seems better to simultaneously introduce independent variables measured at the individual level and characteristics measured at different levels of aggregation. It may be the same characteristic measured at individual and diverse aggregation levels (individual working in agriculture / proportion of farmers in each region). They may also be characteristics defined only at aggregate levels (population density in each parish / population density in each region). Finally, they may be characteristics which only exist at a given aggregation level (hardness of the water distributed in a given district).

Conversely, one may explain an aggregated characteristic by different variables measured at the same aggregation level or at different regrouping levels. If in order to explain aggregated data, it is not possible to come back to individual explanatory variables, other summaries than the regional proportions may be used to improve the fit of the model. In any case, such an approach introduces an important asymmetry between individual and more aggregate levels.

Now that we have better defined the field to explore, it is important to establish if the methods of measurement generally used in social sciences can give us the all of the useful information to undertake such an analysis.

Amongst the exhaustive measures, population register data are a serviceable source of multilevel data. When these population registers are able to cover a great number of fields in the social life, like the Danish registers, they are in a favourable position. Unfortunately, the number of countries having such population registers is very minimal and the field covered is often restricted in many of these countries: births, marriages, internal migrations and deaths are the most frequently registered events. This will permit only limited studies with such a little number of registered events.

En effet, pour l'analyse de caractéristiques socio-démographiques, les modèles biographiques sont à même d'expliquer une caractéristique individuelle dépendant du temps (quotient de migration, de fécondité de rang donné, de mortalité, etc.) par diverses caractéristiques de ces mêmes individus (être agriculteur, de nationalité étrangère, etc.) pouvant elles-mêmes dépendre du temps. De façon semblable, les modèles de régression économétriques travaillent de façon permanente sur des caractéristiques agrégées. Ainsi par exemple, les taux d'émigration régionaux peuvent être expliqués par les parts d'individus dans les diverses situations (pourcentage d'agriculteurs, d'étrangers, etc.). On voit dès lors qu'un problème important dans une analyse multi-niveaux va être de préciser quels types de dépendance peuvent exister entre ces divers modèles et quel apport nouveau peut fournir une telle analyse comparée à des analyses biographiques et des régressions séparées.

Sous de telles conditions, on peut penser unifier les méthodes d'analyse tant des données biologiques que des caractéristiques socio-démographiques. Pour expliquer une caractéristique individuelle, il semble préférable de faire intervenir à la fois des variables indépendantes mesurées au niveau individuel et des caractéristiques mesurées à des niveaux d'agrégation différents. Il peut s'agir du même trait mesuré à divers niveaux (individu agriculteur/part d'agriculteurs dans la région). Il peut également s'agir de caractéristiques dont seules des mesures à divers niveaux agrégés existent (densité de population d'une commune/densité de population d'une région). Enfin, il peut s'agir de caractéristiques n'existant qu'à certains niveaux d'agrégation donnés (dureté de l'eau distribuée dans un district).

Inversement, on peut penser expliquer une caractéristique agrégée par diverses autres variables mesurées au même niveau d'agrégation ou à des niveaux encore plus regroupés. Si, pour expliquer les caractéristiques agrégées, il n'est pas possible de revenir au niveau individuel pour les variables explicatives, d'autres résumés que les taux régionaux considérés pourraient être utilisés pour améliorer l'ajustement. De toute façon, cela introduit une asymétrie importante entre les différents niveaux d'agrégation.

Maintenant que nous avons mieux défini le champ à explorer, il est important de voir si les mesures habituellement utilisées en sciences sociales peuvent nous fournir toutes les informations utiles pour effectuer une telle analyse.

Parmi les mesures exhaustives, les données de registres de population permettent de fournir une source utilisable pour des études multi-niveaux. Lorsque ces registres couvrent de nombreux champs de la vie sociale, comme les registres danois, ils sont en position favorisée. Malheureusement le nombre de pays disposant de ces données est très restreint et l'information recueillie n'est pas très détaillée dans la plupart de ces pays : naissances, mariages, migrations internes et décès sont les phénomènes les plus souvent enregistrés. Cela ne permet que des études faisant intervenir un nombre réduit de variables.

Under these conditions, it seems useful to undertake more detailed surveys, even in countries with population registers. Such surveys based on samples which are often of small size, will have to collect not only individual information, but also information on different characteristics of areas in which these people live: hardness of the water, proximity to a market or health center, etc. This will permit to set up the whole set of individual and aggregate characteristics in order to be able to undertake a multilevel analysis.

The usual methods of analysis can be applied to such data, with some changes: regression, logit, probit models and others, when the characteristics are not time dependant; event history analysis models when they are time dependant. However, more specific new multilevel statistical models need to be developed to analyse such data (Goldstein, 1995). It also seems necessary, as these aggregation problems occur in each social sciences to compare their different approaches in order to be able to determine human behaviour correctly.

Also, this multilevel approach raises new epistemological questions, especially about the status to be given to aggregate characteristics. Are they a reflection of the social organization in which we are living, whereas individual freedom would better appear under individual characteristics? If such is the case, what meaning can we give to a great number of intervening aggregation levels? Would it not be better to display privileged levels which would be integrated in a more general theory? Lastly, how can we link different levels such as the individual, the family, and the neighborhood, the city, the region, and the whole country?

Even if these problems are far from being solved, a reflection on this theme would be very important to undertake in the future.

Some new tools

A number of research workers in different social sciences (biology, demography, human geography, social statistics, etc.) met to take stock of recent progress in multilevel analysis.

Stewart Fotheringham *et al.*, apply themselves to analyse a synthetic characteristic, for example population density, measured at different aggregation levels. In this case, we could not have an individual measure, as density is only defined at an aggregate level. In return, some methods, that permit to explore spatial non-stationarity are explored here and compared to classical use of regression with spatial data investigating the relationships between population density and a series of attributes of the physical landscape.

Il est donc utile, même dans les pays disposant de registres, d'entreprendre des enquêtes plus détaillées. Pour ces enquêtes non exhaustives et portant souvent sur un échantillon restreint, il est important qu'elles recueillent non seulement des informations individuelles mais également des informations sur diverses caractéristiques des zones dans lesquelles les individus vivent : qualité de l'eau, présence d'un marché, etc. Il sera dès lors possible de reconstituer toutes les caractéristiques individuelles et agrégées pour réaliser une analyse multi-niveaux.

Les méthodes d'analyse classiques sont applicables, avec certaines modifications à de telles données : régressions, modèles logit, probit et autres, lorsque les caractéristiques ne dépendent pas du temps ; modèles biographiques lorsque l'on introduit la dimension temporelle. Cependant il serait préférable de mettre en place des modèles d'un type nouveau pour traiter de façon plus précise ces données (Goldstein, 1995). Dans la mesure où ces problèmes d'agrégation se posent dans toutes les sciences sociales, il paraît utile de confronter leurs diverses approches pour mieux cerner les comportements humains.

Enfin, cette approche multi-niveaux va poser de nouveaux problèmes épistémologiques. On peut en particulier se demander quelle signification accorder à des caractéristiques agrégées. S'agit-il d'un reflet de l'organisation sociale dans laquelle nous vivons, tandis que la liberté individuelle se montrerait sous les caractéristiques individuelles ? Dans ce cas quelle signification pouvons-nous donner à l'intervention d'un grand nombre de niveaux d'agrégation ? N'y a-t-il pas lieu d'essayer de mettre en évidence des niveaux privilégiés à intégrer dans une théorie plus générale ? Enfin, comment peut-on articuler entre eux les différents niveaux tels que l'individu, la famille, le voisinage, la ville, la région, le pays ?

Bien que ces problèmes soient encore loin d'être résolus, une réflexion sur ce thème serait très importante.

De nouveaux outils

Un certain nombre de chercheurs issus de différentes sciences sociales (biologie, démographie, géographie humaine, statistiques sociales, etc.) ont pu se rencontrer pour faire ici le point sur les avancées récentes dans ce domaine.

Stewart Fotheringham *et al.* s'attachent à l'analyse d'une caractéristique synthétique, ici par exemple la densité de population, vue à différents niveaux d'agrégation. Dans ce cas on ne peut disposer du niveau individuel, car la densité ne peut être définie qu'à un niveau agrégé. En revanche, des méthodes permettant de mettre en évidence une non-stationnarité spatiale sont développées ici et comparées à une estimation globale de la régression entre la densité de population et diverses autres caractéristiques agrégées.

On the contrary, Tim Holt *et al.* apply themselves to estimating individual level regression coefficients, when only aggregate data for geographic areas are available and, with a small amount of areas, an individual level information on a set of "grouping variables" which come from another data set. Their theoretical investigation provided new results on the problem of ecological fallacy. Also, their empirical investigation on biological data using blood pressure show the contribution such an approach brings to the comparison between individual and aggregate data.

Michel Poulain then tackles the problem of collecting multilevel data, in order to undertake socio-demographic studies. He shows the necessity to modify data input to permit such analyses. The use of population registers, which are centralized and computerized, is described here and their potentialities are shown with detailed example cases.

Brigitte Baccaïni and Daniel Courgeau approach the question of using the same socio-demographic data at an individual level and at aggregate level. They show first from a theoretical approach, the links between the estimated parameters of regression, logit and event history models which may be verified if no aggregation effect exists. They used Norwegian population register data. They gave an estimation of the different model parameters and concluded that the effect of individual characteristics is almost independent of the effect of aggregates characteristics. The two levels each brought an independent information on the characteristics studied here, the out-migration of Norwegian regions.

Towards a theory of multilevel models

The different advances explored in the papers from this session show the potential wealth of multilevel models and, simultaneously, the need of a coherent theory for these models. First, such a theory has to go beyond the concepts of ecological fallacy and of atomistic fallacy. In fact, the ecological fallacy is the result of an improper inference on individual-level associations based on associations measured only at the aggregate level. Similarly, the atomistic fallacy is the result of an incomplete specification of the different characteristics acting on individual behaviour. As soon as one acknowledges the importance of both individual and environmental variables in determining individual level behaviour, the ecological and atomistic fallacy no longer exists.

Furthermore, such a theory needs to go beyond the study of individual behaviour, which tries to explain it by using different characteristics measured at different aggregation levels: it may study the behaviour suited for each aggregation level which will simultaneously further explain each of them in relationship to the others. So that individuals acting in a given community can become aware of a problem

A l'opposé Tim Holt *et al.* vont s'attacher à estimer des relations au niveau individuel, alors que l'on dispose essentiellement de données agrégées avec, cependant, une information au niveau individuel sur certaines caractéristiques discriminantes qui peuvent d'ailleurs provenir d'une tout autre source. Leur démonstration théorique apporte des éléments tout à fait nouveaux sur le problème de l'erreur écologique et l'application à des données biologiques, ici la pression sanguine, montre l'apport de cette réflexion sur les liens entre données individuelles et données agrégées.

Michel Poulain aborde ensuite le problème de la collecte de données multi-niveaux, dans le cas d'études socio-démographiques. Il montre la nécessité de modifier les façons de saisir ces données pour rendre de telles analyses possibles. L'utilisation du registre de population centralisé et informatisé est ici décrite et ses potentialités sont montrées à l'aide d'exemples détaillés.

Brigitte Baccaïni et Daniel Courgeau abordent enfin le problème de l'utilisation des mêmes données socio-démographiques considérées soit au niveau individuel, soit au niveau agrégé. Ils montrent, de façon théorique, les liens entre les paramètres des modèles de régression, logit et biographique, qui devraient être vérifiés s'il n'y avait pas d'effet d'agrégation. Utilisant ensuite les données du registre de population norvégien ils estiment, de façon pratique, les paramètres des divers modèles et concluent que l'effet des données individuelles est pratiquement indépendant de celui des données agrégées. Les deux niveaux apportent chacun une information indépendante sur la variable ici étudiée, l'émigration des régions norvégiennes.

Vers une théorie des modèles multi-niveaux

Les diverses avancées explorées dans les communications de cette séance montrent la richesse potentielle des modèles multi-niveaux et, en même temps, le besoin de mettre en place une théorie cohérente de ces modèles. En premier lieu, une telle théorie doit dépasser les concepts d'erreur écologique et d'erreur atomiste. En effet, l'erreur écologique est le résultat d'une inférence incorrecte sur des relations au niveau individuel, basée sur des relations observées au niveau agrégé. De même, l'erreur atomiste viendrait d'une spécification incomplète des caractéristiques jouant sur le comportement d'un individu. Dès que l'on fait intervenir, non seulement des caractéristiques individuelles, mais aussi des caractéristiques agrégées, dans le même modèle, l'erreur atomiste ou écologique disparaît.

Ensuite, une telle théorie doit dépasser l'étude des comportements individuels, qui cherche à les expliquer par des caractéristiques mesurées à différents niveaux d'agrégation, par une étude des comportements propres aux divers niveaux d'agrégation, qui va chercher à les expliquer simultanément les uns en fonction des autres. Ainsi des actions individuelles dans une communauté donnée peuvent amener la prise de conscience d'un problème qui

which affects the collectivity and lead thus, to political action taken at a more aggregated level. Conversely, these political measures will affect individual behaviour, and so on.

Finally an important problem occurs, from the possible existence of a great number if not an infinity, of existing aggregate levels. It seems evident that some of these levels may be more important than others, but the precise criteria and the methods to detect them are, for the moment, waiting to be settled. Evidently, it is easy to consider some aggregation levels with an administrative definition (city, commune, region, etc.), but it is not obvious to say if they are the most relevant ones to be considered and if other levels, more difficult to define, such as a community, are not of greater interest. These aggregation levels may also change according to time and some new aggregation levels may appear which will have to be explained.

In conclusion, it may be said that such a theory is entirely pending and that its epistemological basis, its methods of measurements, and of analysis are not yet stated. In this paper we provoke thought about a theory of multilevel models, which will try to synthesize human and biological sciences and we hope that forthcoming research will bear out its importance.

affecte la collectivité et conduire à des mesures politiques prises à un niveau plus agrégé. A l'inverse ces mesures politiques vont, dans l'autre sens, affecter les conduites individuelles et ainsi de suite.

Un problème important réside, enfin, dans l'existence *a priori* d'un très grand nombre, sinon d'une infinité, de niveaux d'agrégation envisageables. Certains de ces niveaux sont, à l'évidence, plus importants que d'autres, mais les critères précis et les méthodes pour les déceler sont encore en pleine gestation. Bien sûr il est aisé de s'intéresser à certains niveaux d'agrégation définis de façon administrative (ville, commune, région, etc.) mais on peut se demander s'ils sont les plus pertinents à considérer et si d'autres niveaux, plus difficiles à définir, tels que la communauté, ne sont pas de plus grand intérêt. On peut également penser que ces niveaux peuvent changer au cours du temps et s'interroger sur les conditions d'apparition d'un nouveau niveau d'agrégation pertinent.

Comme on peut le voir, il s'agit là d'une théorie en pleine élaboration dont les bases épistémologiques, les méthodes de mesure et d'analyse ne sont encore pas fixées. Nous apportons ici un début de réflexion sur cette théorie des modèles multi-niveaux, qui cherchent à apporter une synthèse des sciences biologiques et des sciences humaines et espérons que les recherches à venir viendront confirmer son importance.

REFERENCES

- Courgeau, D., (1994), Du groupe à l'individu : l'exemple des comportements migratoires, *Population*, Vol.49, n°1: 7-26. Version anglaise : (1995), From the group to the individual: what can be learned from migratory behaviour, *Population, An English Selection*, 7, 145-162.
- Goldstein, H., (1995), *Multilevel statistical models*, Edward Arnolds, London, 180 p.
- Jones, K., (1993), "Everywhere is nowhere" : multilevel perspectives on the importance of place, *The University of Portsmouth Inaugural Lectures*, Portsmouth, 12 p.
- Von Korff, M., Koepsel, T., Curry, S., Diehr, P., (1992), Multilevel analysis in epidemiologic research on health behaviours and outcomes, *American Journal of Epidemiology*, Vol. 135, n°10: 1077-1082.