STATISTICAL COMMISSION and
ECONOMIC COMMISSION FOR EUROPE

Working Paper No. 11
English only

CONFERENCE OF EUROPEAN STATISTICIANS

UN/ECE Work Session on Methodological Issues Involving the Integration of Statistics and Geography (Neuchâtel, Switzerland, 10-12 April 2000)

Topic (iv): Spatial analysis in a statistical context and disclosure control procedures

SPATIAL HOMOGENEITY AND TERRITORIAL DISCONTINUITIES

Submitted by University Paris VII, France[1]

Invited paper

ABSTRACT

Historically, the development of spatial analysis has been related, in a first step, to the transposition of purely statistical or mathematical tools to the analysis of geographical information (1940-1970) and, in a second step, to the development of specific methods adapted to the particular nature of spatial distribution (1970-1990). Today, the research focuses on the implementation of previous discoveries into the GIS environment and many attempts are made to standardise the operations related to the spatial analysis of geographical information. This standardisation of the procedure of spatial analysis can be dangerous because, in many cases, it is related to a focus on a purely methodological approach without considerations of theoretical background and empirical applications.

The danger of "purely spatial" analysis is particularly clear when methods established for the distribution of physical phenomena (geology, biology, climate, etc.) are applied to the distribution of social or economical phenomena such as regional inequalities. Through various examples related to recent research on discontinuities and heterogeneity in the distribution of GNP on a European and a world scale, we would like to demonstrate how it is possible to improve spatial analysis through a deeper reflection on the implicit assumptions (social, political) of various techniques of spatial analysis. More precisely we would point out some crucial differences between discrete and continuous approach of geographical distributions of social facts on various scales in order to distinguish between territorial analysis and spatial analysis of human behaviour on various scales.

I. INTRODUCTION

1. The Hypercarte Research Group is an international network of geographers, statisticians and mathematicians established in 1996 to answer a call for tender proposed by EUROSTAT. The aim of this call for tender was to find solutions to the problem of spatial heterogeneity of territorial divisions and related bias in cartographical representation.

2. The answer of the Hypercarte Research Group was based on a general smoothing method derived from a probabilistic reformulation of the concept of population potential (The Hypercarte Project, Working Paper n°1). Our proposal was not the one accepted by EUROSTAT but it appeared in the final list of selected offers. Nevertheless, some of the research teams involved in the answer to the call for tender decided to maintain their cooperation and to develop the research without institutional subvention. This concerned mainly three research teams in geography (UMR Géographie-Cités, Paris), statistics (LABSAD, Grenoble) and informatics (LMC-IMAG, Grenoble) with special implications from three researchers : J.M. Vincent (LMC-IMAG), G. d'Aubigny (LABSAD) and C. Grasland (UMR Géographie-Cités) who is the actual coordinator of the project.

---

[1] Prepared by Claude Grasland.

3. The implications of the research developed in the Hypercarte network appear to be very important in the fields of cartography, spatial analysis and territorial planning. Theoretical solutions have been analysed and tested on small samples of spatial data with non-optimised algorithm of computation and cartography. But it is now necessary to build specific informatic software for an application of the proposed methods on a larger database. Accordingly, the Hypercarte Research Group is actually looking for national or international subventions in order to develop a full version of the software HYPERCARTE, which should be able to produce multiscalar representations of spatial distributions.

4. Taking the opportunity of an invitation to the UN-ECE conference of Neuchâtel, we have decided to summarise the basic results obtained by the Hypercarte research group and to present them for general discussion to the specialists of GIS, statistics and Spatial Analysis.

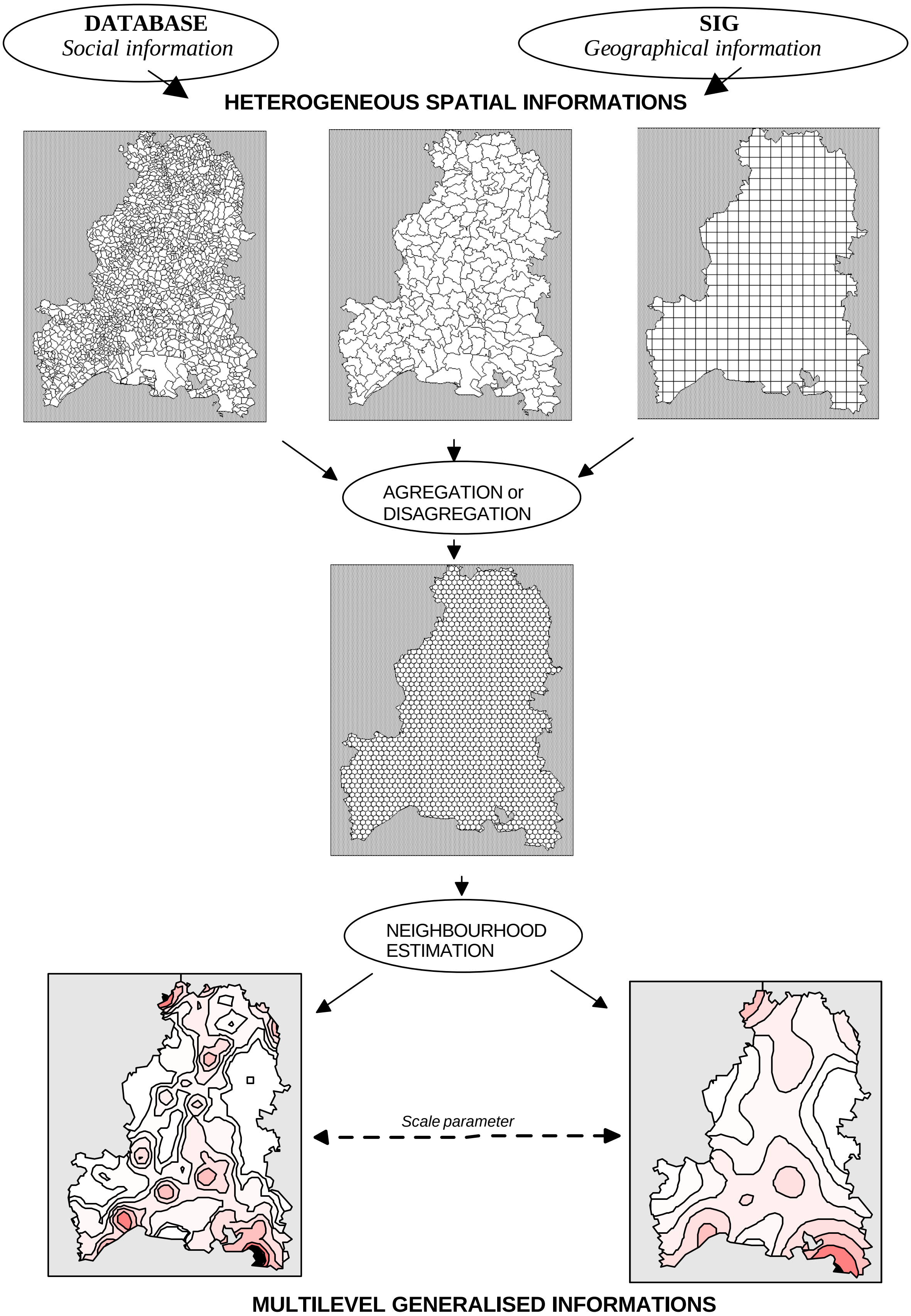
5. The initial aim of the hypercarte project was to to transform heterogeneous spatial information into multilevel generalised information. This result could be obtained in different ways but the proposed solution, based on the concepts of neighbourhood and population potential, seems to be the most general and the most useful from technical, methodological and theoretical points of view. With this solution, it is indeed theoretically possible to store intermediary results of multiscalar computation and to build a particular software coupled with a database structure which ensures the possibilities :

- to map the same information at different levels of generalisation without new computation of basic information. In other words, it should be possible for the user to interactively control the level of map generalisation with a scale button;
- to compare data originating from different statistical offices and to control the quality of the resulting information through the introduction of the database structure of specific parameters which define a minimum level of generalisation;
- to examine the relationship between data which can generally not be associated, like environmental and spatial databases. Through aggregative and disaggregative procedures, it should be possible to define a common grid of information.

6. These results were direct answers to the call for tender from EUROSTAT, but the member of the hypercarte project also planned to build a software and database structure which should be able to take other functions into account in the future. For example :

- to combine different territorial divisions of the same area at different time periods and to interactively examine their evolution with a time button;
- to compare the level of correlation of statistical indices on different scales, from local to global level, and to propose some solutions to the statistical problem of the Modifiable Area Unit Problem through the construction of curves of multiscalar correlation;
- to implement new spatial analysis tools on the database structure like, for example, the definition of homogeneous regions or spatial discontinuities on different scales.

Figure 1 : Expected results of the Hypercarte Software



II. THEORETICAL BASIS[2]

7. In this section, we briefly present the concept of population potential reformulated by the author and its potential application to the problem submitted by the Statistical Office of European Communities. Of course, it is only a working paper, which will be discussed and improved by the expert group of researchers from different teams involved in Workpackage 1.

8. A first version of this paper was presented by C. Grasland in May 1997 at the INSEE Séminaire de Géostatistique directed by C. Terrier.

II.1 Levels of information

II.1.1 Maximum information

9. In this theoretical example, a random sample of 100 individuals has been uniformly distributed inside a square of size 100x100. Let us define, and assume too, that the quantity of space occupied by an individual is equal to a small square of 1x1. Thus, Map 1 represents the finest (atomistic) level of disaggregation, the maximum information available on the distribution of the population of interest. A smaller grid would indeed generate pixel parts of individuals which is contradictory to our initial atomistic assumption. For example, if we want to analyse the distribution of people located on a beach, it is nonsense to use a grid smaller than 1 or 2 m2. We may also consider that the location of each individual is known through the coordinates (x,y) with an uncertainty (x= y=0.5) which is proportional to the size and mobility possibility of individuals during the period of observation.

10. We assume that the practical problem to be solved is the realisation of a map of the population density in the observed area. Since population density defines as the ratio of a quantity of population on a quantity of space, it is not possible to measure it directly on Map 1. In general, this information is not available at the individual level and one starts from a territorial division of space in cells where more than one individual may be located.

II.1.2 Regular grid information (carroyage)

11. In the most favourable case, the territorial partition is based on a regular division of space as in Map 2, where a network of 20x20 square-cells was established in practice. This sort of transformation implies a loss of information because, after transformation from Map 1 to Map 2, it is no longer possible to define the true location of people located in each territorial division. The configuration in a map of population density is related to the scale of the aggregation process (size of territorial divisions) but also to the shape of aerial units (squares, hexagons, triangles) and to their reference coordinates. In the present example the same grid starting from point (10,10) instead of point (0,0) would produce a distinct picture of population density.

II.1.3 Administrative grid information (maillage)

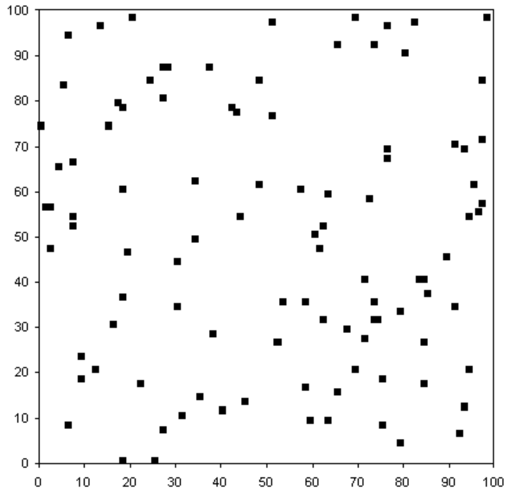
12. In classic empirical situations, the territorial partition worked on is based on administrative units which present great heterogeneity in shape and size. As an example, Map 3 results from a projection of one administrative division of France (départements around Paris) on the random distribution of Map 1. The number of territorial divisions in Map 2 and Map 3 is approximatively equal (25 and 27). But the loss of information on the location of individuals is more important on Map 3, because of the variation of size and shape of the administrative divisions. We shall now examine the following problem: how is it possible to produce equivalent maps of population density when starting from the information contained in Map 1, Map 2 or Map 3 ?

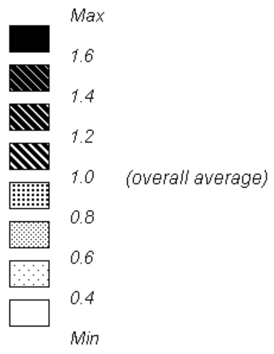
---

[2] Extract from the original answer to the call for tender from Eurostat. Hypercarte Working Paper No. 1.

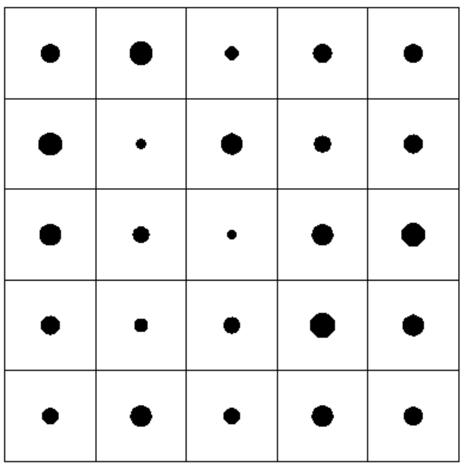
INFLUENCE OF THE AGREGATION LEVEL ON THE PERCEPTION OF POPULATION DENSITY

Map 1 : Maximum information

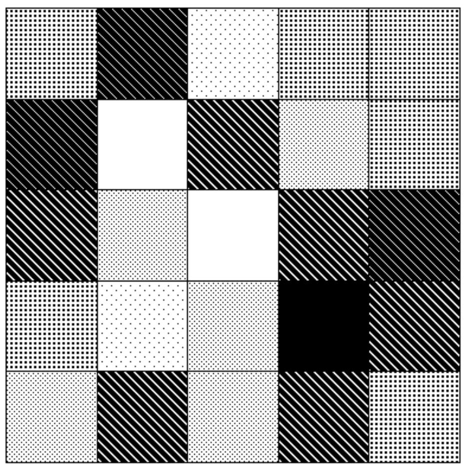




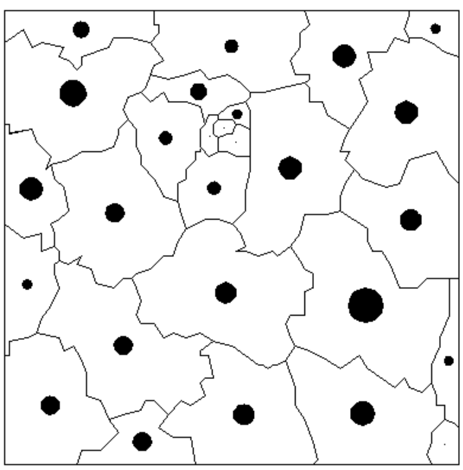
Map 2 : Agregation through regular grid



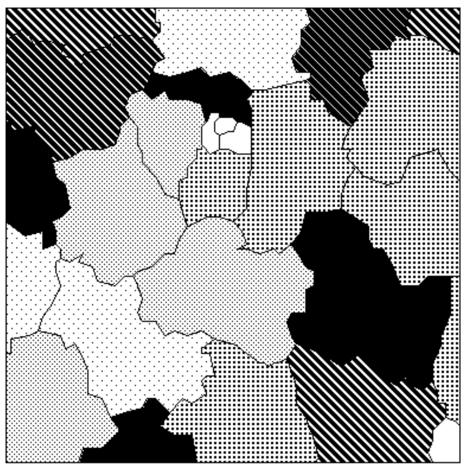
Map 2-b : Density derived from Map 2



Map 3 : Agregation through an administrative grid



Map 3-b : Density derived from map 3



Legend of all maps of population density

II.2 Neighbourhood and population potential concept

12. Before we analyse the problems induced by aggregated information, we will analyse a case where maximum information is available. Starting with Map 1, one needs to know how to produce a map of the population density.

II.2.1 A general formulation of population potential based on neighbourhood

13. In each of the 100x100 elementary cells of Map 1 the population is 0 or 1 and the surface area is 1. This information is detached and if we want to produce a continuous map of density, we are obliged to produce an estimation of the quantity of population and surface area located in the neighbourhood of each point of the observed area. An observer located in site i who wants to determine the density of population in his neighbourhood will select a spatial sample of places 1..j..N and then check whether or not they are occupied. The criterium of selection of the N places is called neighbourhood. Let K be the number of inhabited places in this neighbourhood (K<N). The ratio d=K/N represents the estimated density of population in the neighbourhood of place i. In mathematical terms, the neighbourhood induces a function p_{ij} which attributes to each couple of points (i,j) the probability of truth of the proposal « j is located in the neighbourhood of i ». In an isolated case the value of p_{ij} are 0 or 1 while in the continuous case, p_{ij} can be any real number in the interval [0 ; 1].

II.2.2 Territorial neighbourhood

14. Let us now examine what is the implicit neighbourhood function used for the estimation of the population density in Map 2-b and Map 3-b according to the initial information contained in Map 1. An administrative observer located in any of the 10,000 elementary pixels can use as an isolated neighbourhood the set of sites located in the same territorial unit as he himself. The neighbourhood used to determine the population density at any point can be expressed as a partition matrix A = (aij) of common territorial belonging defined as :

> *a_{ij} = 1 if i and j are located in the same territorial unit*
> *a_{ij} = 0 if i and j are located in different territorial units*

15. Let us now define a variable *X* which takes the value 1 for inhabited places and the value 0 for places which are not occupied; the estimation of the population density for administrative neighbourhoods can be calculated from *Map 1* as follows :

$$\overline{X}(i,A) = \frac{\sum_{j=1}^{10000} X_j . a_{ij}}{\sum_{j=1}^{10000} a_{ij}} = \sum_{j=1}^{10000} k_{ij} X_{ij} \quad when \ k_{ij} = \frac{a_{ij}}{\sum_{j=1}^{10000} a_{ij}} \qquad (1)$$

16. It now appears clearly that the population density is the ratio of a potential of population (i.e. the quantity of X located in the neighbourhood of i) to a potential of surface area (the quantity of space located in the neighbourhood of i).

II.2.3 Spatial neighbourhood

17. The problems encountered during the use of territorial neighbourhood information can be explained by the fact that formula (1) does not use homogeneous spatial units (in shape and size) for the measurement of the population density. Some of the discontinuities in the distribution of density are related to the fact that all the points located in the same territorial unit have the same neighbourhood, whatever their position in the territorial units (center, periphery, etc.). The territorial observer of population density thus uses a varying technique to observe and measure, which may result in an artificial heterogeneity of the map of density.

18. The concept of spatial neighbourhood basically results from the assumption that the neighbourhood used by the observer is a function f of a given distance d_{ij} whose associated mathematical properties are :

(a) $d_{ii}=0$	
(b) $d_{ij} = d_{ji}$	$\forall\ i, j, k :\quad i, j, k \in \{1,2, ..., n\}$
(c) $d_{ij} + d_{ik} \geq d_{ik}$	

19. The properties of the spatial neighbourhood function f are induced by the assumption made above the information field of the observer. In the spatial neighbourhood approach, we assume that the observer has full information about the place where he is located and null information about places located at an infinite distance. This model assumes too that the level of information of the spatial observer decreases with distance. Accordingly, the function of spatial neighbourhood should have the following properties :

(a') $f(0) = 1$
(b') $f(\infty) = 0$
(c') if $d_1 < d_2$ then $f(d_1) \geq f(d_2)$

A great number of mathematical functions satisfy these conditions. The most interesting are parametric functions which allow the possibility to control the size and shape of the neighbourhood.

Circular neighbourhood

20. The simplest way to define spatial neighbourhoods is to assume that the observer has full information on places located at a distance lower than or equal to a threshold R and null information on places located at distances greater than R. The potential associated to this hard core function is the quantity of population located in a disk of radius R. It can be calculated at each point of the observed area or on a sample grid as in Map 4. Differing maps of the population potential can be produced according to the value of R retained which defines the span of the neighbourhood.

21. One advantage of circular neighbourhoods is the simplicity of their definition. But the resulting kernel function has many weaknesses when applied to aggregated data. When the exact location of the population is unknown (as in Map 2 or 3) the use of circular neighbourhood will produce discontinuities if one assumes a location of the whole population of territorial units at a single representative point. A small variation of location can indeed produce a dramatic increase of the population potential and the map of population density will be biased (with the apparition of circular areas of high density). Furthermore, it has been observed in most empirical situations that the perception of space by people is not a discrete but rather a continuous function decreasing with distance. Thus, maps which use circular neighbourhoods do not reflect the actual perception of space by human beings. It can also be observed that the position of people given by census information is conventional and represents only a point of maximal probability to find an individual during a given period of time. People move (for work, holidays and so on) and this mobility is also, in general, a function which decreases with distance.

Exponential neighbourhood - Gaussian neighbourhood

22. The interest of the family of negative exponential function $exp\ (-a.d_{ij}^{\,b})$ is to be related to empirical results referring both to the perception of space and to the mobility of populations. The value of the parameter *b* helps to define the shape of the neighbourhood (decrease of information rate with distance) and the value of the parameter *a* controls the span of the neighbourhood. Conventionally, we propose to define the span of an exponential neighbourhood as the distance *R* where $exp\ (-a.R_j^{\,b})$= 0.5. The Gaussian neighbourhood *(b=2)* presents specific advantages from both mathematical and theoretical points of view. As demonstrated by P. Chataignon (INSEE) and A. Guerreau (C.R.H.-Equipe P.A.R.I.S.), the smoothing methods based on gaussian neighbourhood can be derived from the theory of kernel and non-parametric estimators of density. C. Grasland (Equipe P.A.R.I.S.) also demonstrated that it can be related to the theory of Spatial Information Fields

developed by Swedish time-geographers such as T. Hägerstrand and to the theory of intervening opportunities of the American sociologist S. Stouffer. Map 5 presents the population potential associated to a gaussian neighbourhood function of span 20 applied to data from Map 1. As in the circular neighbourhood case, it is possible to modify this span of neighbourhood, according to the value of the parameter *a*.

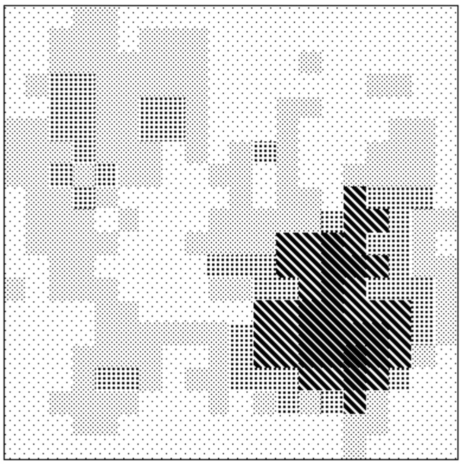
Power Neighbourhood (Modified Pareto Function)

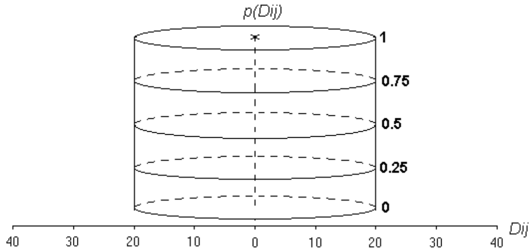
23. According to the general theory of spatial interaction, many researchers (e.g. H. Le Bras) propose definition of the spatial neighbourhood as an inverse function of distance, called Pareto function : d_{ij}^{-b}. This solution is consistent with empirical observations on the population mobility but not necessary with observations on information fields. However, it has the disadvantage of producing an infinite value of the potential for places where people are located. The Pareto function does not strictly define a neighbourhood function because the condition (a') is not available. Accordingly, we propose using the family $(1+a.d)^{-b}$ which has equivalent properties and where the parameters *a* and *b* can be analysed in the same way as in the case of negative exponential family of neighbourhood. For a given value of shape of the neighbourhood (parameter *b*), it is possible to control the span of the neighbourhood through the parameter *a*.

VARIATION OF POTENTIAL ACCORDING TO SHAPE AND SPAN OF NEIGHBOURHOOD

Map 4 : Circular neighbourhood (span 20)

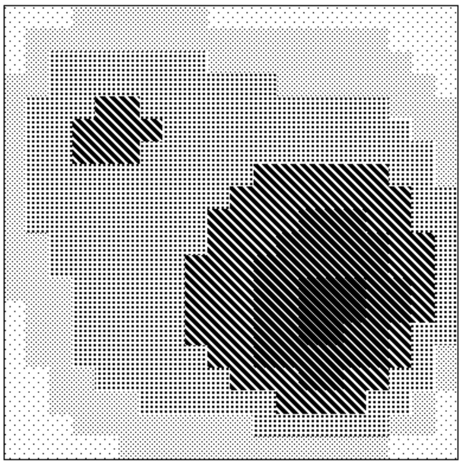
f(d) = 1 if d £ 20 and f(d)=0 if d >20

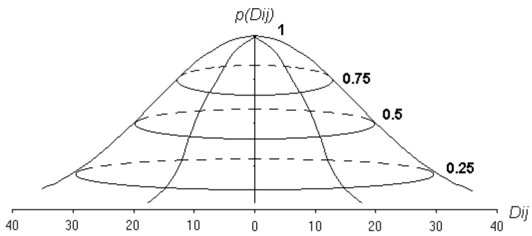




Map 5 : Gaussian neighbourhood (span 20)

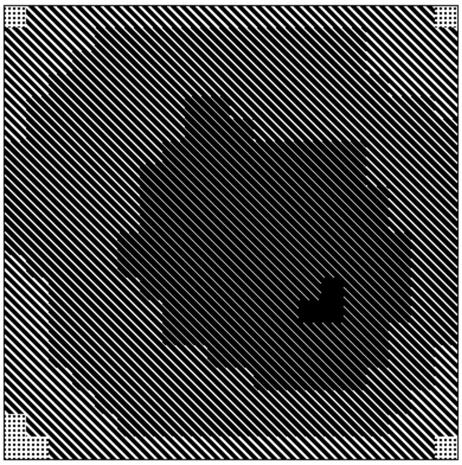
$f(d) = exp(-0.00693\ d^2)$

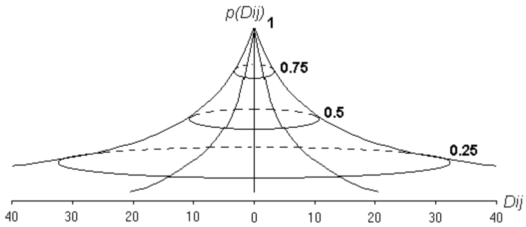




Map 6 : Pareto-1 neighbourhood (span 10)

$f(d) = (1+0.1\ d)^{-1}$





II.3 Estimation of density and control of border effects

24. Maps 4 to 6 illustrate clearly that the population potential associated with various definitions of the spatial neighbourhood does not reflect exactly the level of population density, because of border effects. Points located near the boundary of the observed area have generally lower levels of population potential, simply because their neighbourhood is censored. This effect could be interesting if the observed area was an island and if no individuals were in fact located outside the observed area. Otherwise, it must be considered as a bias factor. In our setting, it is easy to measure the bias induced by the boundary effect through the computation of a potential of surface area which can be considered as the potential of population which would be observed with a homogeneous and uniform distribution of population. The potential of surface area also measures the size of the spatial sample used for the estimation of population density at a given point of the observed area and it indicates the level of confidence of the final result.

<u>II.3.1 Standardisation of the potential of population and superficy</u>

25. For a given function of spatial neighbourhood *f*, and a given measure of distance *d*, the potential of population at a point *i* (inhabited or not) is equal to $\overline{X}(i, f, d)$

$$\overline{X}(i, f, d) = \sum_{j=1}^{N} f\left(d_{ij}\right).\ X_j \qquad (2)$$

26. *N* is the number of pixels in the observed area and *Xj* counts the number of members of the population *X* located in each pixel. This equation can thus be applied to an individual as well as to an aggregated distribution of population. If *f* is a neighbourhood function satisfying the properties defined above, the maximum value of $\overline{X}(i, f, d)$ is the total population located in the observed area denoted X_T. We can thus define a standardized value of population potential which is the frequency of the total population located in the neighbourhood of any point.

$$X^*(i,f,d) = \overline{X}(i, f, d) / X_T \qquad \text{with } X^*(i,f,d) \in [0,1] \qquad (3)$$

27. The same formulation can be used for the estimation of the potential of surface area and the standardised potential of surface area as well:

$$\overline{S}(i, f, d) = \sum_{j=1}^{N} f\left(d_{ij}\right).S_j \qquad (4)$$

$$S^*(i,f,d) = \overline{S}(i, f, d) / S_T \qquad \text{with } S^*(i,f,d) \in [0,1] \qquad (5)$$

28. In the case of the maximal information grid (*Map 1*), all the values of S_j are equal to 1 and N is equal to S_T which is the surface area of the observed area. Moreover this formulation can also be applied to aggregated data (*Map 2* and *Map 3*).

29. After standardisation, the potential of surface area *(Map 8-a)* may be interpreted as an indicator of *spatial accessibility* (i.e. the number of sites, inhabited or not, which can be reached from a given observed point of observation under the neighbourhood assumption), and the potential of population *(Map 7-a)* as a measure *of social accessibility* (i.e. the number of inhabited sites which can be in relation with the observed point under the same neighbourhood assumption). In both cases, a standardised measure of spatial anisotropy of the potential function *(Map 7-b and 8-b)* can be calculated through a summation of contributions to the potential multiplicated by a unit vector :

$$\vec{X}^*(i,f,d) = \frac{1}{X_T}\sum_{j=1}^{N} X_j \,.\, f\left(d_{ij}\right).\,\vec{1}_{ij} \qquad (6)$$

$$\vec{S}^*(i,f,d) = \frac{1}{S_T}\sum_{j=1}^{N} f\left(d_{ij}\right).\,\vec{1}_{ij} \qquad (7)$$

where $\vec{1}_{ij}$ *is a vector directed from i to j with a norm equal to* 1

30. The resulting vectors have a norm between 0 and 1 which indicates the orientation and the level of polarisation of contributions to the potential. If the contributions (population or surface area) to the potential of a given observed point are located in each direction of space - or in opposite directions - around this observed point, this norm will be more or less equal to 0. If all the contributions originate from the same direction of space, the norm will be equal to 1 (for the demonstration of this result, see Grasland C., Mathian H., 1993).

II.3.2 Determination of neighbourhood density

31. The neighbourhood density can now be defined as the ratio of the potentials of population and surface area :

$$\overline{Z}(i,f,d) = \frac{\overline{X}(i,f,d)}{\overline{S}(i,f,d)} \qquad (8)$$

32. It is also possible to define a standardised measure of neighbourhood density *(Map 9-a)* as the ratio of the standardised values of potential of population and surface area. This standardised density yields an index of relative concentration whose reference value is 1 (mean density of the observed area).

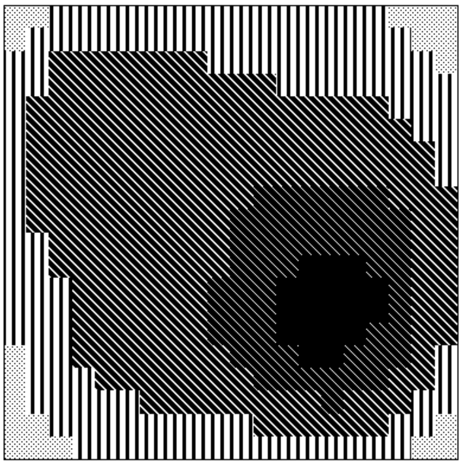
$$Z^*(i,f,d) = \frac{X^*(i,f,d)}{S^*(i,f,d)} = \frac{S_T}{X_T}\,.\,\overline{Z}(i,f,d) \qquad (9)$$

33. Finally, estimations of both the direction and the intensity of the gradient of the density of population can be obtained through the vector substraction of the standardised vectors associated to the population potential and the surface area potential *(Map 9-b)*.

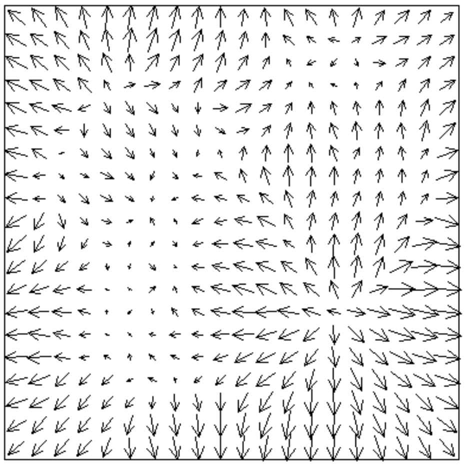
$$\vec{Z}^*(i,f,d) = \frac{1}{2}\left[\,\vec{P}^*(i,f,d) - \vec{S}^*(i,f,d)\right] \qquad (10)$$

ESTIMATION OF POTENTIAL AND DENSITY WITH GAUSSIAN NEIGHBOURHOOD (span 20)

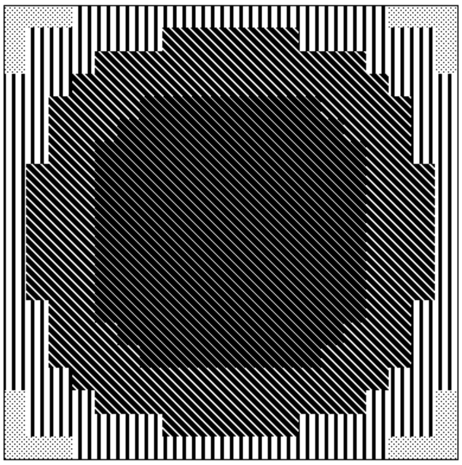
Map 7-a : Population potential



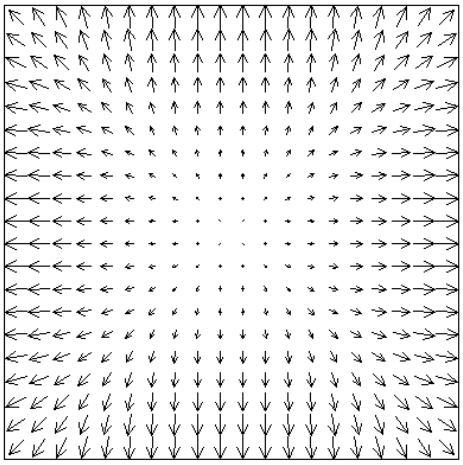
Map 7-b : Vector of population potential



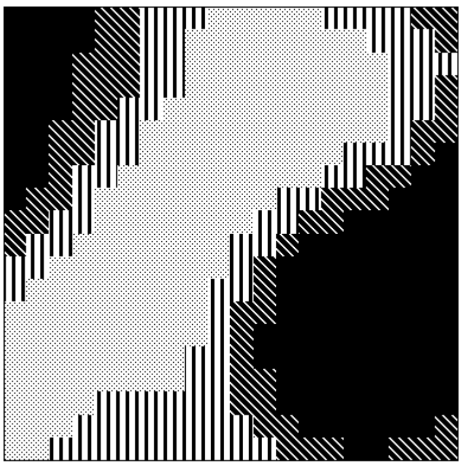
Map 8-a : Surface area potential



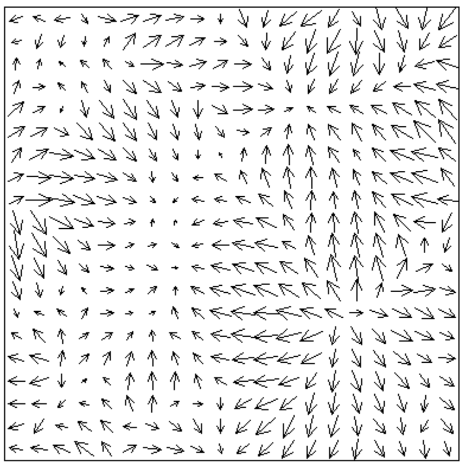
Map 8-b : Vector of surface area potential



Map 9-a : Neighbourhood density



Map 9-b : Vector of neighbourhood density



II.4 Stability of the results

34. The method proposed above may appear a little bit complicated to the layman user, who might ask if such a complex theoretical framework is mandatory to the elaboration of a generalised map of primary population density. Many commercial GIS's offer rough smoothing methods which produce a continuous map according to the density of each territorial unit. These methods generally rely on the assumption that the density z_i of each territorial unit can be located at the centroïds (x_i,y_i) of those units and they interpret them as a sample of points of a surface $z=f(x,y)$, the equation of which has to be estimated. Various methods have been proposed: triangulation with linear interpolation, kriegeage, polynomial estimation, all of which will be detailed in the first Report.

35. But these methods are unsatisfactory from the theoretical point of view because they do not take into account that the density at a given point can never be defined in objective terms without some consideration of scale. A density is necessarily related to the choice of a neighbourhood family and so it is dependant on their properties of shape and scale.

36. The problem with classical methods is that the sample $\{z_1,..z_i,..z_n\}$ defined as the density at the barycenter of territorial units *mixes heterogeneous neighbourhoods* because units have different sizes and shapes. Accordingly, it is not surprising to observe that these methods produce distinct maps of density when they are applied to varied territorial divisions while tentatively describing the same information (see for example *Map 2* and *Map 3*). Furthermore, these methods are generally unable to estimate density with full information *(Map 1)* because they imply a first step of aggregation for the estimation of a sample of points where density is available.

37. In our opinion the main quality criterium of a smoothing method should quantify the ability to obtain approximately the same map when starting from differing nested levels of information and different types of territorial division describing the same information (**d'Aubigny, 1996**). The question is how to make it possible?

II.4.1 An empirical experiment with gaussian neighbourhood.

38. The scale parameter involved in the definition of a spatial neighbourhood function, as presented above, yields the opportunity to obtain different pictures of the same phenomenon, when the assumption made on the information field of the observer or the mobility of population varies.

39. Starting with full information *(Map 1)*, one can produce a continuum of maps of density depending on the value of the span parameter. High values of the parameter *a (see Gaussian neighbourhood, p. 15*) will produce a low level of generalisation and a great number of peaks of density will appear in the resulting map. For example, in our simulation with a gaussian neighbourhood of span 10, we can distinguish 7 or 8 maxima in the distribution of the population density (*Map 10-a*). A lower value of the parameter *a* allows more generalisation and the number of extrema will decrease. With a neighbourhood span of 20, only 2 maxima of population density remain (*Map 10-b*). At one end of the domain of variation ($a=0$) the density is equal in all the observed areas because each point has both a maximal potential of population and a maximal potential of surface area. Accordingly, the level of density is equal to the mean density in each point of the observed area. At the other end of the domain of variation $(a = \infty)$ the density is maximal in each inhabited place and minimal in other cases.

40. This result is, of course, a theoretical one because, in general, one does not have full information on the location of people, and must begin with aggregated information. But in such a situation, the problem is to determine *the minimum span which guaranties that the resulting map of density approximates the one we would have obtained with full information.*

41. If one applies a gaussian neighbourhood of span 10 on the dataset used in *Maps 2 and 3,* generalised maps of density (*Maps 11-a and 12-a*) will result; these are significantly different from

the *Map 10-a* based on full information. The general configuration of high and low densities is more or less the same but the location and the number of extrema are different. These bias are related to the fact that we had to make assumptions on the location of people inside territorial units and that the geometric centroïd of a territorial unit is not necessarily the same as the barycenter of the population located inside this territorial unit.

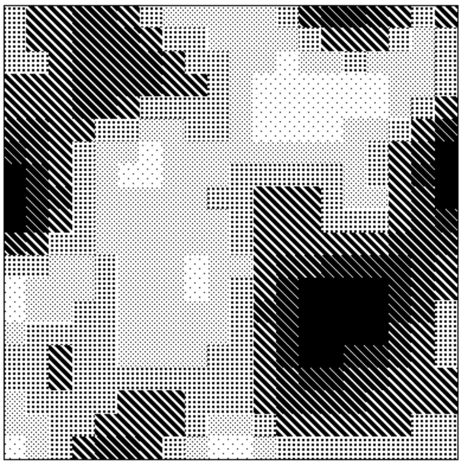
42. If one applies a gaussian neighbourhood of span 20, the differences between maps produced from distinct starting levels of information are fewer than in the previous case. *In Maps 10-b to 12-b*, the values of density are approximately the same, as are the number and the location of extrema. Even the map produced by administrative division, while show great differences in sizes and shapes of territorial units, appears similar to the true map based on full information.

43. Of course, this very simple experiment needs to be quantified and complemented by statistical tests of correlation between the various estimations of the density. But at the moment, it is sufficient to suggest that it is possible to define a minimum span of generalisation according to the empirical properties of territorial units and the precision of the measurement under scrutiny. In other words, the proposed method of map generalisation partly controls the loss of information (variation of span) and might suggest the minimum level of generalisation necessary to obtain a satisfactory estimation of the phenomenon of interest. The introduction of this function in the GIS is very important because it guarantees the scientific quality of the maps which will be realised according to the structure of initial aggregated information[3].

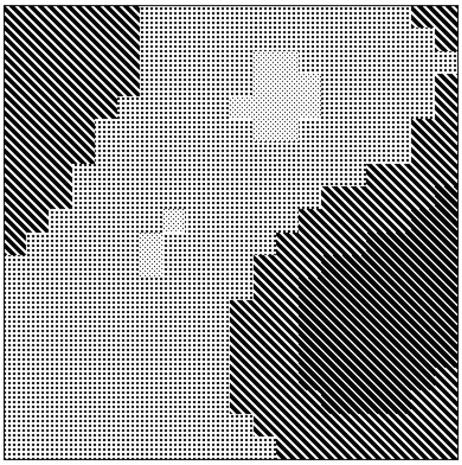
[3] For example, empirical tests indicate that with data from level NUTS 3, it is incorrect to use a neighbourhood gaussian function with a span lower than 80 km. The map of population density in 1980 presented on the front of the report represents the most detailed map, according to the available information which was used for the realisation. With information available at level NUTS 5 it is, of course, possible to produce the same map, but also to examine the location of extrema of population in the European Union at smaller levels of analysis. In experiments with the 36,000 French communes, it was possible to use a span of 5 km, but it is not necessarily possible in other states such as Belgium where those units are greater. Thus, the minimum span might vary according to the shape of the area to be mapped (the whole European Union or some parts of it). This is the reason why the minimum span should be estimated interactively by the software to be implemented in Luxembourg.

STABILITY OF THE RESULTS ACCORDING TO LEVEL OF AGREGATION (Gaussian method)

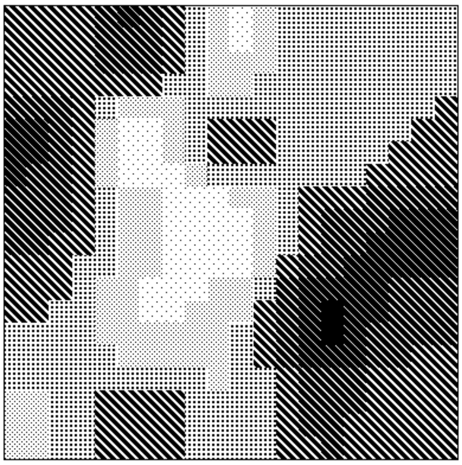
Map 10-a : Maximum Information / Span 10



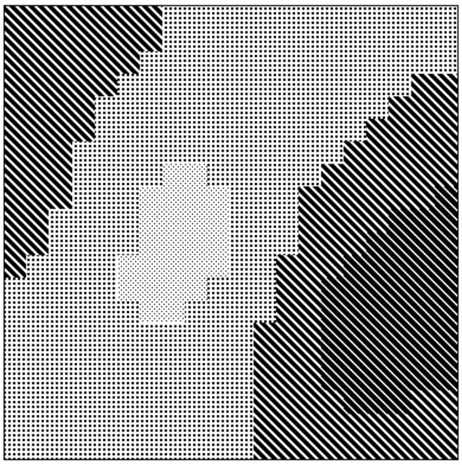
Map 10-b : Maximum Information / Span 20



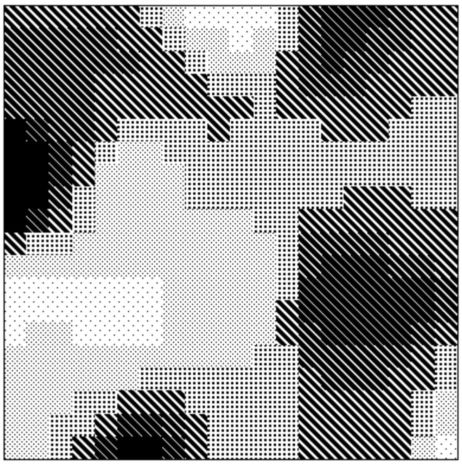
Map 11-a : Regular Grid / Span 10



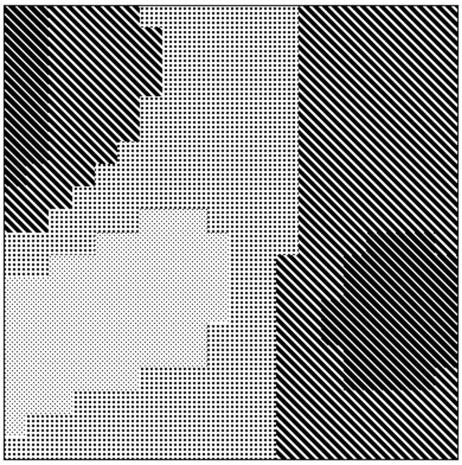
Map 11-b : Regular Grid / Span 20



Map 12-a : Administrative Grid / Span 10



Map 12-b : Administrative Grid / Span 20



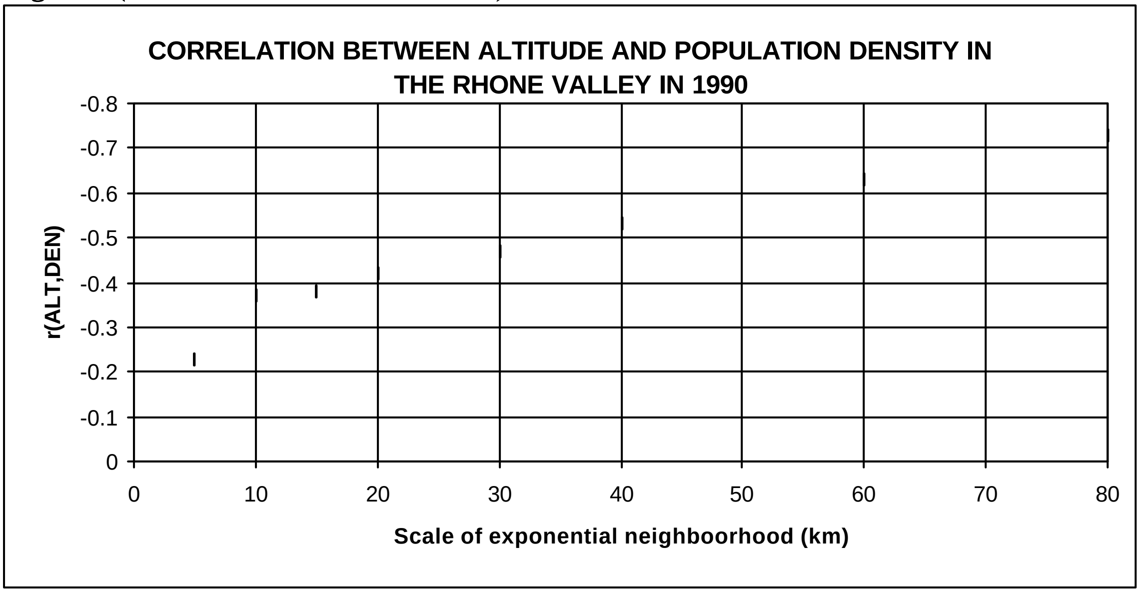
II.4.2 Multiscalar Analysis and the Modifiable Area Unit Problem

44. The problem induced by the level of aggregation of information is not only to define a minimum level of generalisation but also to examine the effect of this generalisation on the terms of statistical analysis. The Modifiable Area Unit Problem (MAUP in abstract) designates the fact that the correlation between two phenomena distributed over space can change quite a lot according to the size and shape of territorial units on which those phenomena are measured. Accordingly, all the statistical methods based on correlation or covariation (regression analysis, cluster analysis, factor analysis, ANOVA, etc.) produce results which vary according to the choice of territorial divisions. Many authors (e.g. Oppenshaw) think that the bias resulting from the MAUP can be reduced by some procedures of aggregation or disaggregation of territorial units which take into account the local levels of spatial autocorrelation. But these methods generally rely on the assumption that it is possible to define an objective optimal division of space or an optimal scale for statistical analysis.

45. In our opinion, this assumption is overoptimistic. The degree of association between two phenomena distributed in space is necessarily dependent on a particular scale, a particular choice of neighbourhood. Two sub-populations distributed in a state can be located in the same region at the macroscopic level but in different villages at the microscopic level. Thus, the spatial correlation (or association) between these sub-populations will be objectively positive at one resolution scale of analysis and negative at another one. The only way to take care of this specificity is to build a *curve* which describes the variation of the correlation intensity according to scale levels.

46. As an example, **C. Grasland & H. Mathian (1993)** analysed the relationship between density of population and altitude in the Rhone Valley *(Figure 5).* At the local level, the analysis of a sample of communes showed a significant but not very important correlation (-0.23). But when using this same sample with density of population and altitude calculated with the help of a gaussian neighbourhood of varying span, they demonstrate that the strength of the correlation increased (rather smoothly) with the span of neighbourhood:

Figure 5 *(Grasland C., Mathian H., 1993)*



47. It is not a general law, but Fig.5 shows that the location of population is related to altitude at a general level of observation, while there are naturally many exceptions to these statistical findings at local level. This happens, for exammple, in low populated areas in the Rhone Valley (e.g. Plaine de la Crau, Camargue) and in the relatively highly populated area in the surrounding mountains and hills (e.g. Alès, Aubenas). But the most important towns or concentrations of population (Marseille, Montpellier, Avignon) are located in the regions of low altitude, and this is the reason why the correlation increases when the neighbourhood increases.

48. The fact that this is not a general law is a crucial problem for multivariate analysis because the order of correlation between a set of variables can change greatly when the span varies. Thus, the factor and the class which will be revealed by multivariate analysis can be very different and their rank or power can also be modified. As an example, just consider the correlation between four indices describing the Rhone Valley at communal level and in a gaussian neighbourhood of span 20 km (*Figure 6*). After neighbourhood transformation we can observe a dramatic increase of the level of correlation between the variation of population and the altitude (-0.06 to -0.57) but a reduction of the level of correlation between the proportion of young and the altitude (-0.30 to -0.16). These facts indicates clearly that multivariate analysis applied to communal information or neighbourhood generalised information would produce very different results and different empirical interpretations of the relations between variables.

Figure 6 : Correlation between four variables according to the span of neighbourhood in the Rhone Valley *(Grasland C., Mathian H., 1993)*

Definition of variables

ALT	Altitude
YOU	% of young (0-14) in 1990
DEN	Density of population in 1990
VPO	Variation of population 1936-1990

Pearson correlation at communal level

	ALT	*YOU*	*DEN*	*VPO*
ALT	1			
YOU	-0.30	1		
DEN	-0.23	0.15	1	
VPO	-0.06	0.07	0.21	1

Pearson correlation with gaussian neighboorhood of span 20 km

	ALT	*YOU*	*DEN*	*VPO*
ALT	1			
YOU	-0.16	1		
DEN	-0.41	0.14	1	
VPO	-0.57	0.56	0.22	1

II.5 GENERALISATION OF THE METHOD

II.5.1 The case of a ratio (Z = A/X)

49. The method presented above for the generalisation of a population density can be applied to any kind of spatial distribution related to social information, natural environment and landcover. As an example, we present the case of the generalisation of a ratio which expresses the proportion of a subpopulation A in the whole population.

50. Starting again from the distribution of *Map 1*, 50 % of individuals were randomly attributed to population *A* and 50% to population *B* (*Map 13-a*). The usual way to map the relative strength of *A* in the total population *X=AÈB* is to define a territorial partition and to count the numbers A_i and B_i of representatives of *A* and *B* (respectively) in each cell *i* (*Map 13-b*). It is now possible to calculate for each cell *i* the value Z_i of the index *Z* defined as :

$$Z_i = A_i / (A_i + B_i) = A_i / X_i \qquad (11)$$

51. The results are presented in *Map 14-b* and show important discontinuities between contiguous cells (*Map 14-c*). These are artefacts created by the choice of a territorial neighbourhood used in the computation of the ratio.

52. Starting from *Map 13-a* or *Map 13-b*, an alternative solution is to compute both the potential of population of tye A and the potential of total population X within the same spatial neighbourhood. Now it is possible to define a local neighbourhood value of the ratio *Z* as :

Z(i,f,d) = A(i,f,d) / X(i,f,d)	*(12)*

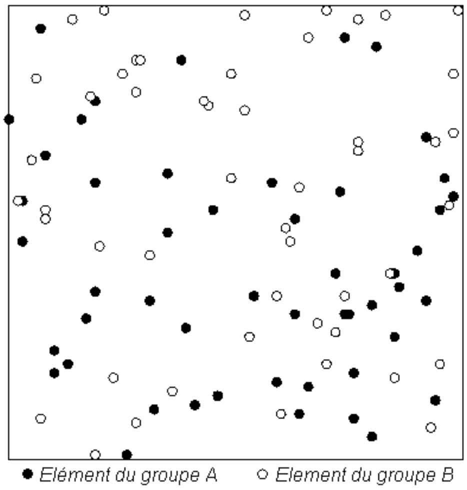
53. Using the samea gaussian neighbourhood function as before (span 20), one obtains a continuous representation of the proportion of *A* in the total population (*Map 14-a*). When substracting the standardised vector associated with the potential of **A** from the standardised vector associated with the potential of *X*, a representation of the spatial variation (gradient) of *Z* (*Map 15-a*) is obtained. The application to real data (proportion of youth in the Rhone Valley) is presented on *Figure 6*.

$$\vec{Z}*(i,f,d) = \frac{1}{2}\left[\vec{A}*(i,f,d) - \vec{X}*(i,f,d)\right] \qquad (13)$$

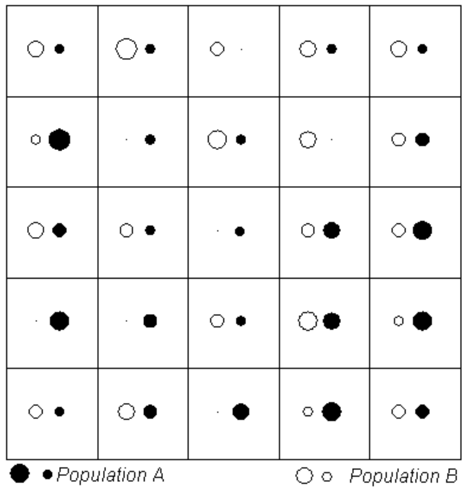
54. This map of the most important gradient gives a better image of the area of spatial transition of the phenomenum Z than the map of territorial discontinuities established on the base of the limits of territorial units (*Map 15-b*). For more details on the advantages and weaknesses of both methods of analysis of spatial transitions and spatial discontinuities, see **C. Grasland, 1997.** As in the case of density, the choice of a span of neighbourhood must account for the level of aggregation of the initial information. Moreover, the use of several spans of neighbourhood will show varying levels of map generalisation of the variable *Z*.

ESTIMATION OF A RATIO Z (proportion of A in a population

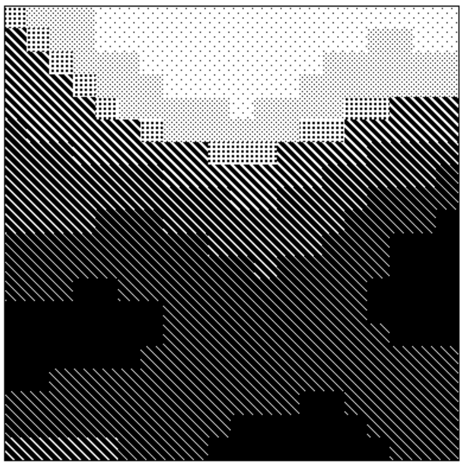
Map 13-a : Maximum Information



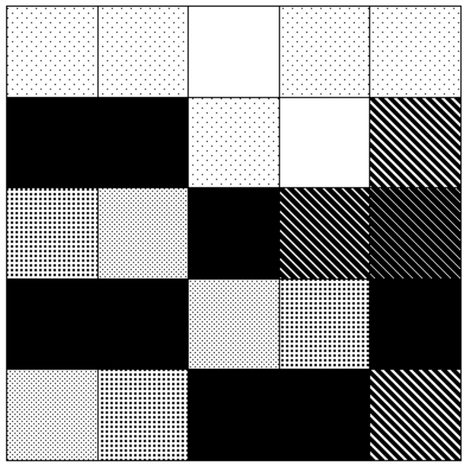
Map 13-b : Grid Information



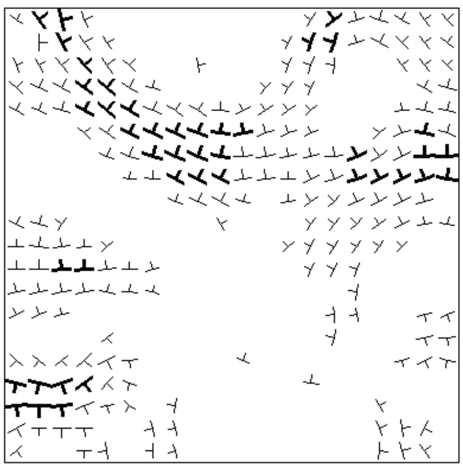
Map 14-a : Spatial estimation of Z (Gaus. / 20)



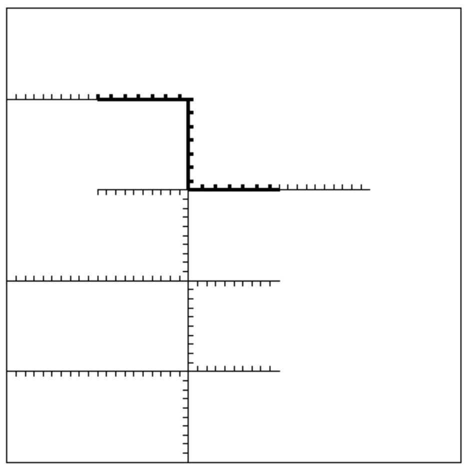
Map 14-b : Territorial estimation of Z



Map 15-a : Spatial gradients of Z (Gauss. / 20)



Map 15-b : Territorial discontinuities of Z



II.5.2 The case of a variation of population

55. When one wants to analyse the evolution of a population *X* at several periods of time, one need only calculate the potential associated with the series $X_1..Xn$ which denotes the observation of the index *X* at time points $t_1..t_n$. An annual rate of evolution of the potential of X can be calculated in any point of the observed area; this produces an estimation of the evolution of the population X in the neighbourhood defined by the observer. As an example, see the variation of population in the Rhone Valley from 1930 to 1990 (*Figure 6*).

56. It is very important to observe that, *if the span of neighbourhood is large enough*, the territorial division used for the definition of X at different periods may have changed. In other words, it is possible with this method to map the global evolution of a distribution known to have changing territorial division. The problem is simply to define the minimum span of neighbourhood which guarantees full compatibility of the various territorial divisions. This solution will be tested by the *CEG (Lisboa)* through an application of changing territorial divisions at the communal level in Portugal.

II.5.3 The case of complex indices

57. Many social indices can not be easily generalised because the value of an aggregate is not the mean or the weighted mean of this index at a lower level of aggregation. Examples of this problem are the median age of a population and the synthetic index of fertility.

58. But these indices can generally be expressed as a function *z* of several populations or subpopulations $P^1..P^k$: $Z_i = z(P^1_i..P^k i)$. As it is possible to compute the potential of each of those populations or subpopulations for a given spatial neighbourhood, it is also possible to obtain the neighbourhood estimation of the indice through substitution of potential to population or subpopulation in the function *z*:

$$Z(i,f,d) = z\ [P^1(i,f,d)\ ...\ P^k(i,f,d)]$$

59. In this case, the computation time can be very long but it is important to notice that, generally, a great number of indices are derived from a small number of population variables. Thus, when the potential of all those population variables is computed and stored (for a given assumption of neighbourhood), all the indices derived from them can be very quickly estimated without new computations, like a standard database based on administrative division. This is a great advantage, compared to usual methods of interpolation of the value of indices. This solution will be tested by the *IGEAT-ULB (Bruxelles)* in the case of demographic indices decribing the France-Belgium border area.

III. APPLICATION EXAMPLE : «THE OBJECTIVE 13-BIS »[4]

60. The aim of this working paper is to demonstrate the interest of the project through its application to a particular problem: *the determination of subvention granted to territorial area according to selected criteria*. As an example, we discuss the possible solutions for a fictive example called "**Objective 13-bis**", applied to the case of Belgium territory.

III.1 A territorial approach

61. Imagine that an international institution decided in 1991 *to grant money to an area where a decrease of population has been observed during the last 13 years (1978-1991).* This "objective 13-bis" can be established on the basis of existing territorial division but the total amount of subvention obtained by Belgium will be different according to the choice of the level of territorial organisation used for the

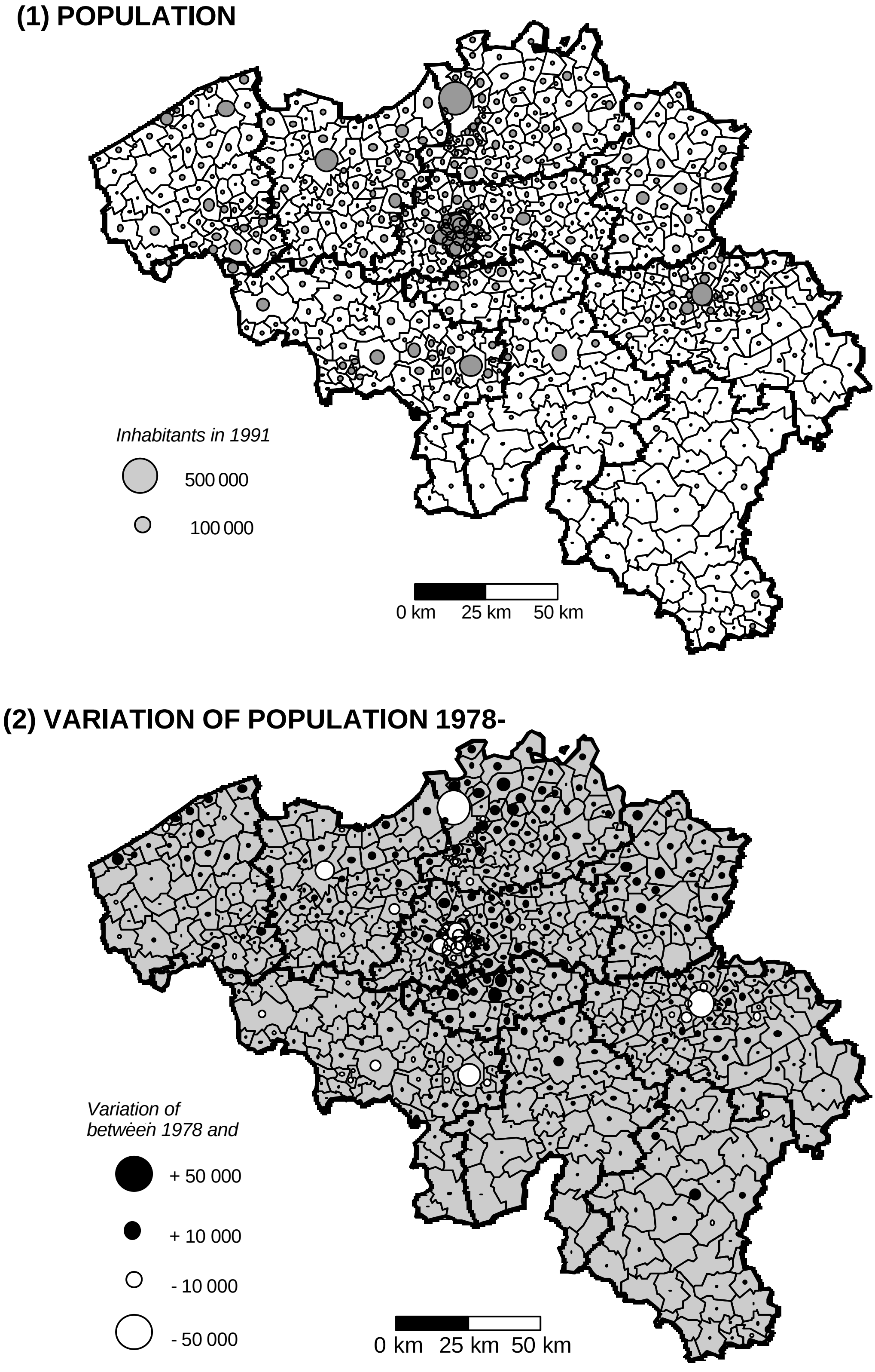
4 Extract from a conference paper presented to the 2nd Meeting of National Focal Points of the Study Program on European Spatial Planning (Stockholm, Febr. 1999).

computation of population decrease. The results will also be different according to the nature of the subvention (equal for each territorial unit or proportional to their population in 1991).

62. The example of Belgium is very interesting because the increase of population during the period 1978-1991 was slightly positive but very near to 0. Thus, a slow change in territorial divisions can produce dramatic change in the amount of subvention.

- At the most local level (*589 communes*), it is possible to observe a decrease of population in 165 communes with a related population of 4540000 inhabitants. Accordingly, 28% of the communes and 46% of the population could be concerned by the "Objective 13 bis".
- At the level of the *43 arrondissements*, the results are more or less the same : 13 arrondissements (30%) and 40,46000 inhabitants (41%) would be concerned by the objective. But their location is rather different and some communes will gain or lose subventions according to their location in increasing or decreasing arrondissements, whatever their local situation of population variation.
- At the level of the *11 provinces*, the results are not so interesting for Belgium because only 3 territorial units (27%) and 3,240,000 inhabitants (32%) are now concerned by the objective. And the situation could be worse if Belgium has maintained the older division in *9 provinces* because, in this case, there would be no subventions for the decreasing unit of Brussels, aggregated with increasing provinces of north and south-Brabant. At the level of the 9 former provinces, only 23% of the inhabitants would be concerned by the objective.
- At the level of the *3 regions*, only Brussels is concerned with the objective (10% of the inhabitants). But if Brussels was aggregated with Wallonia, the new region would be decreasing and 42% of the inhabitants would be concerned by the objective. In the reverse case (Brussels aggregated with Flanders), nobody in Belgium would be concerned by the objective.
- Finally, if Belgium is considered as a single territorial unit (*1 state*), the increase of population is slightly positive (+2%) and the whole state is no longer concerned with the objective 13-bis.

Figure 1 : Variation of population in Belgium at level Nuts 5 (1978-1991)

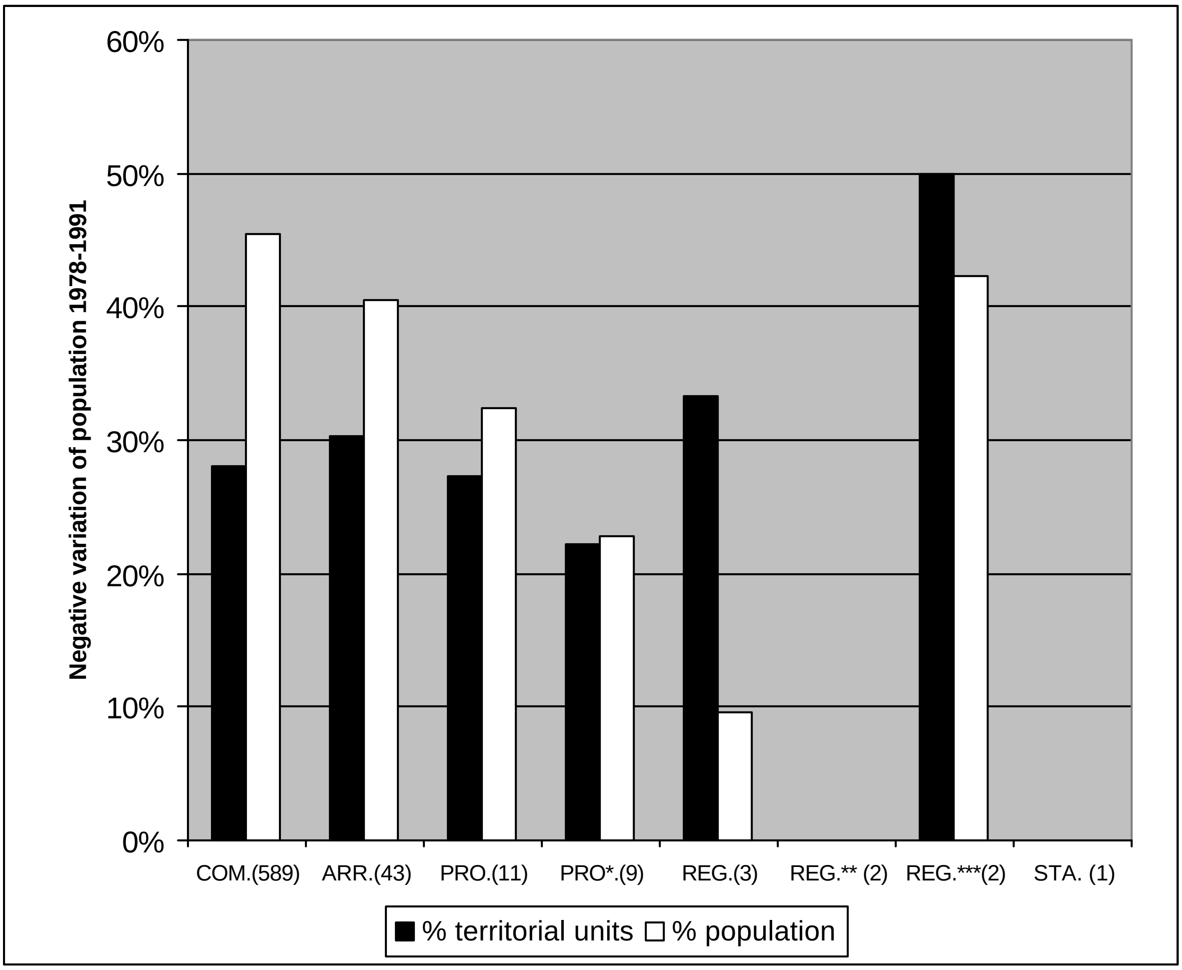


(c) GRASLAND C., UMR Géographie-Cités

Figure 2 : Results of Objective 13-bis according to various territorial divisions

Territorial division		*Administrative result*		*Population result*	
	n	*units*	*%*	*inhabit.*	*%*
Communes	*589*	165	***28%***	4540635	***46%***
Arrondissements	*43*	13	***30%***	4046031	***41%***
Provinces	*11*	3	***27%***	3241356	***32%***
*Provinces **	*9*	2	***22%***	2281032	***23%***
Regions	*3*	1	***33%***	960324	***10%***
*Regions***	*2*	0	***0%***	0	***0%***
*Regions ****	*2*	1	***50%***	4219119	***42%***
State	*1*	0	***0%***	0	***0%***

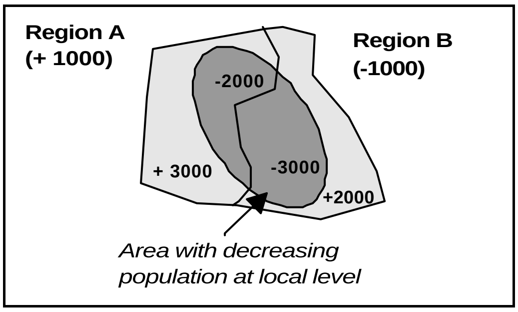
***	: former division in 9 provinces (Bruxelles + North-Brabant + South Brabant)**
** : Bruxelles aggregated with Flanders
*** : Bruxelles aggregated with Wallonia



III.2 A multiscalar spatial approach

63. Another approach to the problem is to answer the question *: "Why do we give money to people or territory?"* or *"At which scale of neighbourhood is it justified to provide subvention to inhabitants located in regions with decreasing population?"*. In Belgium, as in the rest of Europe, the centres of urban area have a negative variation of population but are surrounded by suburban areas with increasing population. The decrease of urban centres is only the consequence of a relocation of inhabitants and does not necessary imply a decrease of population at the level of whole functional urban areas. If the purpose of "Objective 13-bis" is to help people located in *regions* with negative increase of population, it is obvious that the level of communes is not adapted and that it is necessary to use Nuts 2 or Nuts 3 levels for the attribution of subvention. But if the purpose of "Objective 13-bis" is to help territorial communities with fiscal problems (decrease of resources related to taxes in central urban areas), the level of communes could also be justified.

64. But whatever the solution adopted, the choice of territorial units has a tendency to produce social and spatial inequities related to discontinuities in the attribution of subvention. If a continuous area of decreasing population is located on both sides of an administrative boundary, the attribution of subventions can be granted on one side of the boundary and refused on the other side according to the overall variation of population in the territorial units located on each side of the boundary :



65. The use of a multiscalar smoothing method appears very interesting in this framework because the situation of each place can be evaluated in a continuous approach based on a homogeneous criterium of neighbourhood, based in turn on a function of distance around each commune (**Figure 3**) . The experiments realised in Belgium with gaussian neighbourhood at different scales indicates clearly the dramatic variation of the criterium according to the span of neighbourhood which is used (**Figure 4**). We can observe that the amount of subvention does not simply decrease with the size of neighbourhood and that an "optimal solution" (maximisation of subvention) is obtained for a scale parameter equal to 20 km (44% of the communes and 57% of the inhabitants involved in decreasing area). With a lower scale parameter we observe a decrease of subvention (e.g. 31% of communes and 49% of population with scale parameter equal to 5 km). With a greater scale parameter we also observe a decrease and the amount of subvention is equal to 0 for each scale parameter greater or equal to 85 km. A precise analysis of the maps related to each scale parameter (**Figure 4**) indicates important variations in the territorial location of subvention and provides solutions for an "optimal territorial division" according to political or strategic decisions. To summarise, these maps are probably one of the best tool for a state which decides to maximise its benefits and/or to proceed towards an equitable distribution of subventions over its whole territory (e.g. computation of the amount of subventions obtained by Flanders and Wallonia according to the scale parameter).

Figure 3 : Variation of population in Belgium (1978-1991) according to different scales of spatial neighbourhood (gaussian)

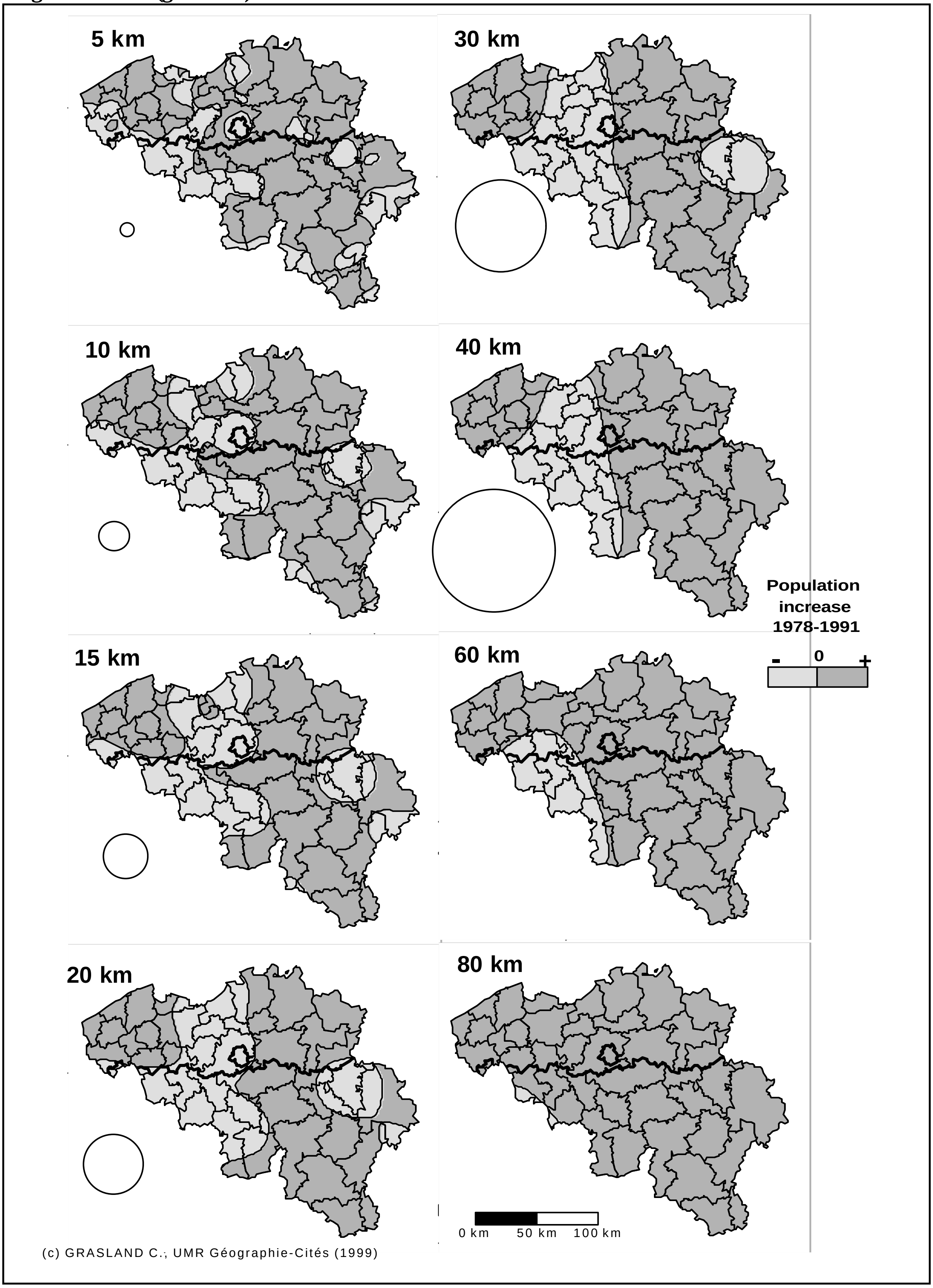
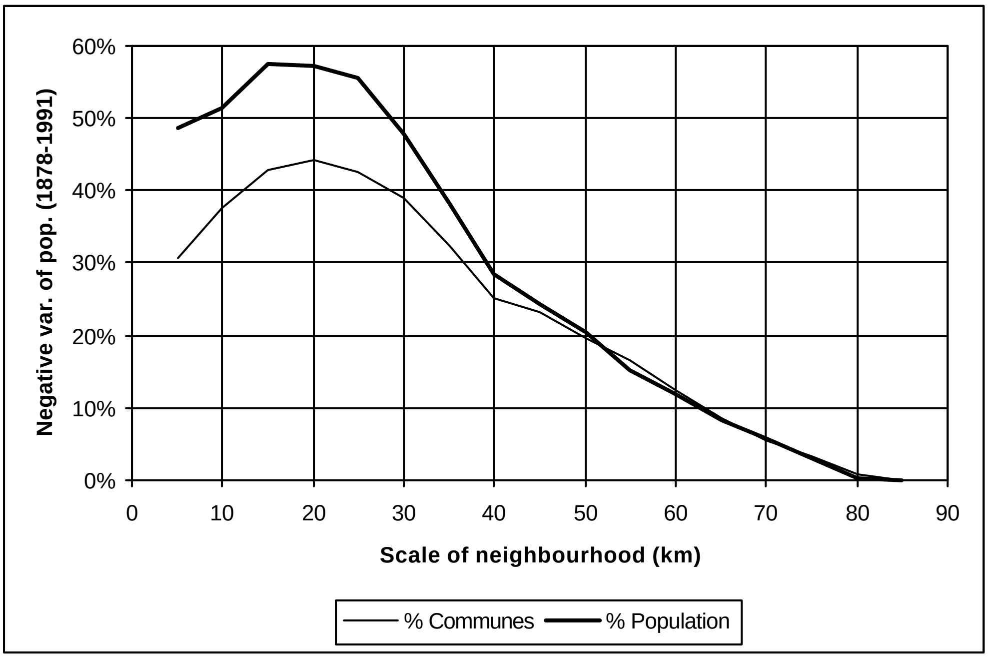


Figure 4 : Results of Objective 13-bis according to various scales of spatial neighbourhood

Spatial Neighbourhood (gaussian function)	*Administrative result*		*Population result*	
	communes	*%*	*inhabit.*	*%*
0 km (communes)	165	***28%***	4540635	***46%***
5 km	180	***31%***	4856867	***49%***
10 km	221	***38%***	5140270	***52%***
15 km	252	***43%***	5732949	***57%***
20 km	260	***44%***	5697571	***57%***
25 km	251	***43%***	5551928	***56%***
30 km	229	***39%***	4776694	***48%***
35 km	191	***32%***	3811093	***38%***
40 km	148	***25%***	2847600	***29%***
45 km	137	***23%***	2415950	***24%***
50 km	116	***20%***	2038452	***20%***
55 km	97	***16%***	1513407	***15%***
60 km	73	***12%***	1176705	***12%***
65 km	51	***9%***	822845	***8%***
70 km	32	***5%***	566429	***6%***
75 km	19	***3%***	298120	***3%***
80 km	5	***1%***	32219	***0%***
85 km	0	***0%***	0	***0%***



III.3 Political consequences of GIS innovations

66. The aim of this working paper is not to criticise actual practices of territorial planning or to propose a so-called "optimal solution" to the very difficult question of territorial planning and territorial subventions. We simply assume that the solutions to those very difficult questions (which have to take into account many non-scientific parameters) could be ameliorated with new tools of spatial analysis and cartography.

67. The spatial neighbourhood approach (based on euclidian distance) is only a preliminary approach to the problem which could be improved through the consideration of social neighbourhoods (based on time, cost or perceived distances). What is important is to identify the time-space budget of people who need subventions and to examine whether or not they are able to find resources in the territory where they are living.

68. In the case of unemployment, for example, the question is to define the maximum area where people are able to find a job without their being obliged to migrate. If we observe that this "job-area" is equal to a radius of 2 hours around the place where they live, the rate of unemployment in a 2 hours neighbourhood is probably a good criterium for the attribution of subventions to areas with a high unemployment rate.

69. According to Max Weber, we think that the work of scientists is not to take decisions in the place of policy makers, but rather to help them make their decisions, to provide them with clear explanations on which they can base their decisions and foresee the consequences of those decisions.

IV. CONCLUSION

70. The aim of this paper is only to suggest the basic concept of the research developed within the framework of the Hypercarte Network. More details on theoretical and methodological solutions will be presented during the conference (with various empirical examples of applications).