ESPON

EUROPEAN SPATIAL PLANNING OBSERVATION NETWORK

ESPON 3.4.3

THE MODIFIABLE AREAS UNIT PROBLEM

First Interim Report

15/01/06

UMR Géographie-cités & UMS RIATE (coordination teams)
NCG – University of UMEA – IGEAT – IMAG - CESA - BBR

Co-financed by the European Community through the Interreg III ESPON Programme

ESPON 3.4.3
THE MODIFIABLE AREAS UNIT PROBLEM

This report represents the interim results of a research project conducted within the framework of the ESPON 2000-2006 programme, partly financed through the INTERREG programme.

The partnership behind the ESPON programme consists of the EU Commission and the Member States of the EU25, plus Norway and Switzerland. Each partner is represented in the ESPON Monitoring Committee.

Information on the ESPON programme and projects can be found on www.espon.lu

The web side provides the possibility to download and examine the most recent document produced by finalised and ongoing ESPON projects.

ISBN number:
This basic report exists only in an electronic version.
Word version:

Table of contents

Figures

Tables

List of contributors

UMR Géographie-cités (FR)
Claude Grasland*
Hélène Mathian
Léna Sanders*

UMS RIATE (FR)
Nicolas Lambert
Malika Madelin**

NCGIA (IE)
Martin Charlton
Jianquan Cheng
Stewart Fotheringham

University of UMEA (SE)
Einar Holm
Kirsten Holme
Magnus Stromberg

IGEAT (BE)
Moritz Lennert
Pablo Medina Lockard
Christian Vandermotten

IMAG (FR)
Jean-Marc Vincent

CESA (FR)
Philippe Mathis
Kemal Serrhini

BBR (DE)
Wolf-Dieter Rase

*** Leaders of the project**
**** Scientific coordinator**

Executive summary

The aim of this Interim Report of ESPON project 3.4.3.is to present a clear and understandable presentation on the nature of the modifiable areas unit problem (MAUP) from general point of view (**Part. 1**) and to examine how ESPON is concerned (**Part.2**) before to propose recommendations (**Conclusion**). This presentation is completed by two case studies on Sweden (**Case study 1**) and Ireland (**Case Study 2**) and by various annex.

Part 1: What is the MAUP?

For planning purpose as well as for research topic, it is necessary to get an insight in the spatial variations of social and natural phenomena. It means to be able to have a representation of spatial inequalities, to identify specific zones, to identify discontinuities in space, and to understand the underlying principle of the spatial organization of a phenomenon as well as its correlation or independence relatively to other phenomena. Classical tools such as cartography, statistical analysis, and spatial modelling are used for that purpose. The fact is that the results of these analyses are dependent on the definition of the studied units, which depends on choices relatively to:

1) the ***scale*** or level at which the spatial entities are observed;

2) the kind of spatial ***aggregation*** which has been adopted

In other words, the results vary according to the spatial zoning which is used in the study. In the different sections of 1st Part, we explain and illustrate the different forms that this question can take. We first introduce the question from a theoretical and methodological point of view. Then we present the main cases of MAUP which occur in thematic research or in an operational perspective[1] and we illustrate them with a series of examples. From a methodological point of view the two first examples concern the spatial organisation of a single variable (sections 1 and 2), the third one has a temporal perspective (section 3), and the two last ones concern situations where multivariate analysis is involved. At last we will discuss in

[1] In this Interim Report the aim is not to provide a complete state of the art but rather to give a clear understanding of what MAUP is, and in what sense it is a problem to ESPON decision makers. A more complete scientific discussion with full bibliographic references will be done in the Final Report.

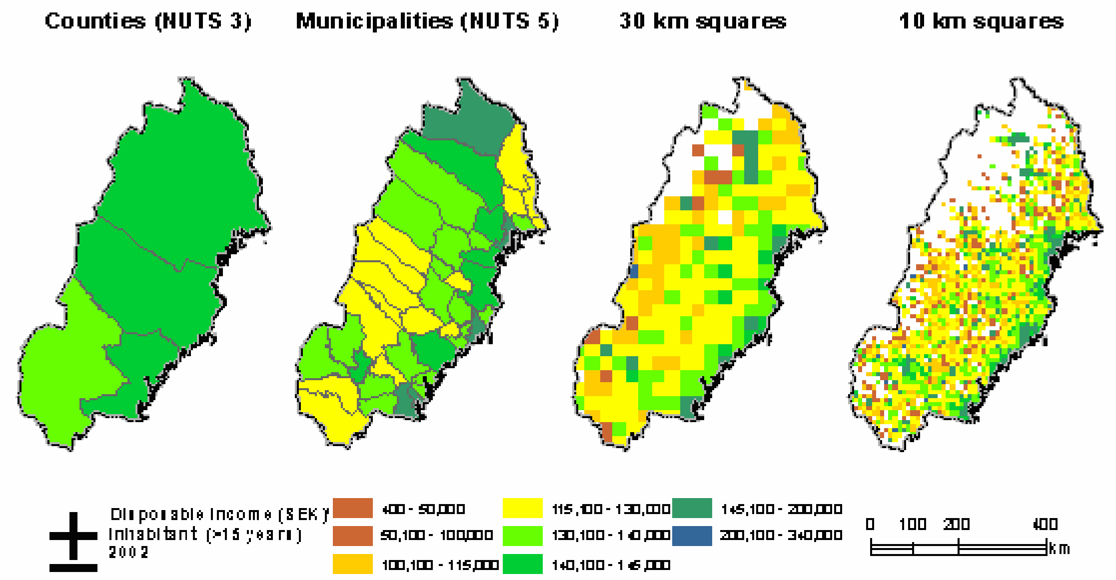
what sense such contradictory results is a problem to be solved but also in what sense it could produce knowledge about the phenomena studied.

1.1 Preliminary methodological consideration: Statistical distribution and spatial organisation

This section starts from the confusion between the descriptions of *statistical* and *spatial* distribution of a phenomenon, which s often due to the definition of the elementary units, and of the level of analysis. It explains that this scale problem is crucial from political point of view as it is the basis of the distinction between social cohesion (reduction of inequalities between individuals or households) and territorial cohesion (reduction of inequalities between territories). It demonstrates with one theoretical example that one statistical distribution can correspond to different spatial patterns depending on the underlying spatial processes of localisation.

1.2 With different zoning, spatial variations don't look the same

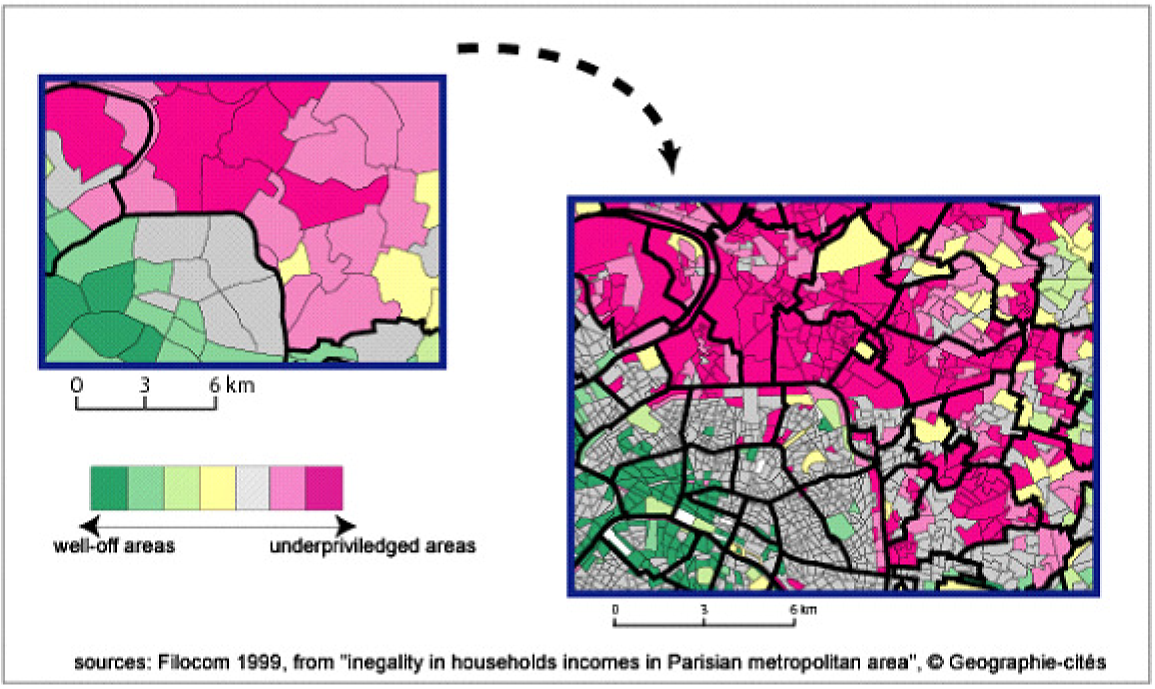
This section illustrates the problem of changing cartographic representations of the phenomena according to the choice of the zoning and the scale with different theoretical and empirical examples. The example of income distribution in northern Sweden provide a very good illustration of this:



Influence of the zoning on the perception of the phenomena (empirical example)

1.3 Correspondence between administrative limits and spatial discontinuities: a possibility, but seldom a rule

This section examines the question of the *position* of spatial discontinuities or barriers in the distribution of phenomena. In some cases an administrative boundary acts as a functional limit, it is the case for what is related to policy action for example (planning, social policy, taxations etc). But when administrative units are used as the basic spatial entities for representing socio-demographic phenomena, an apparent discontinuity could be artificial. A good illustration of this is provided by social segregation in the NE of the agglomeration of Paris where social discontinuity does not fit exactly with the limit of city-center and more generally with the limits of local authorities (communes).



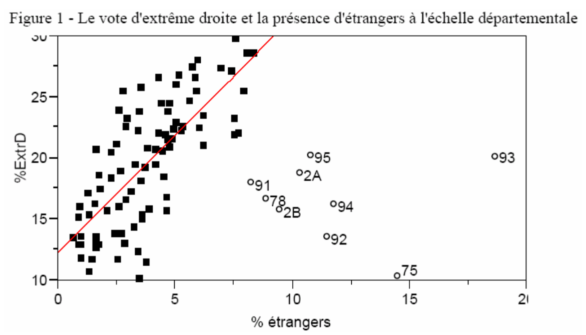
Social segregation to the test of scale

1.4 With different definitions of spatial entities, the forms of evolution appear different

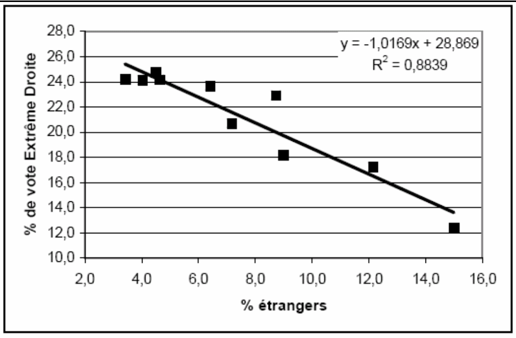
This section focuses on the analysis of time series with the example of changing delimitation of urban areas. It demonstrates that the urbanization evolution looks in France radically different depending on the adopted delimitation: urbanization seems either to decrease, to stagnate or to increase. From a thematic point of view, such differences can lead to an interpretation in terms of counter-urbanization or of increasing urban concentration.

1.5 The relations between phenomena seem different depending on the scale of observation and with the spatial extent of the study region

This section demonstrates that the statistical relation between two variables (a coefficient of correlation for example) can vary according to the spatial units at the level of which the variables have been observed. In certain cases, we can observe complete inversions of results:



(a) % of foreigners and % of extreme right votes for the 96 French departments



(b) % of foreigners and % of extreme right votes for rings around Paris (until 100km)

Relations between extreme right votes and foreigners rate: different filters

It implies that the generalization of a relation between variables to other scales or to other spatial zonings most often doesn't make sense. But, it means also that, when contradictory results appear, the study of the mechanisms underlying these differences between the results will give a better understanding of spatial differentiation.

1.6 And what about the effect of a boundary on the relation between phenomena

This section focus on the specific problem of transnational areas and examines to what extent do the observed differences in statistical analyses either side of a national border result from data on either side being drawn from incompatible spatial units or from the fact that the socio-economic processes which are operating on each side of the border are different. The problem is illustrated by the case study on Ireland which shows the sensitivity of the results to these variations: for example the effect of catholic religion on educational attainment is positive in Northern Ireland and negative in Ireland; and the effects of social class on education attainment are greater in Northern Ireland than in Ireland.

1.7 In what sense is the MAUP a problem?

This conclusive section of the first part indicate that MAUP is not necessary a problem if we consider that the differences in the results of statistical analyses according to the scale are creating knowledge about the observed phenomena. The identification of the scales at which segregation for example does or does not exist, learns a lot about the organization of the concerned society.

The most important problem is about international and historical comparisons: do the elementary spatial units which are used for the analysis have the same meaning in two different countries? At two different time periods? It is not easy to determine if a difference in the results is due to a difference in the processes which are underlying the observed phenomena, or simply to a difference in the meaning of the spatial entities that are used for the observation.

Part 2: ESPON and the MAUP

The question of the potential influence of changing territorial divisions on applied research in the field of spatial planning has been regularly discussed in the framework of the ESPON program and even before (SPESP). A brief summary of the discussion related to MAUP in the different ESPON seminar lead to the identification of three major questions

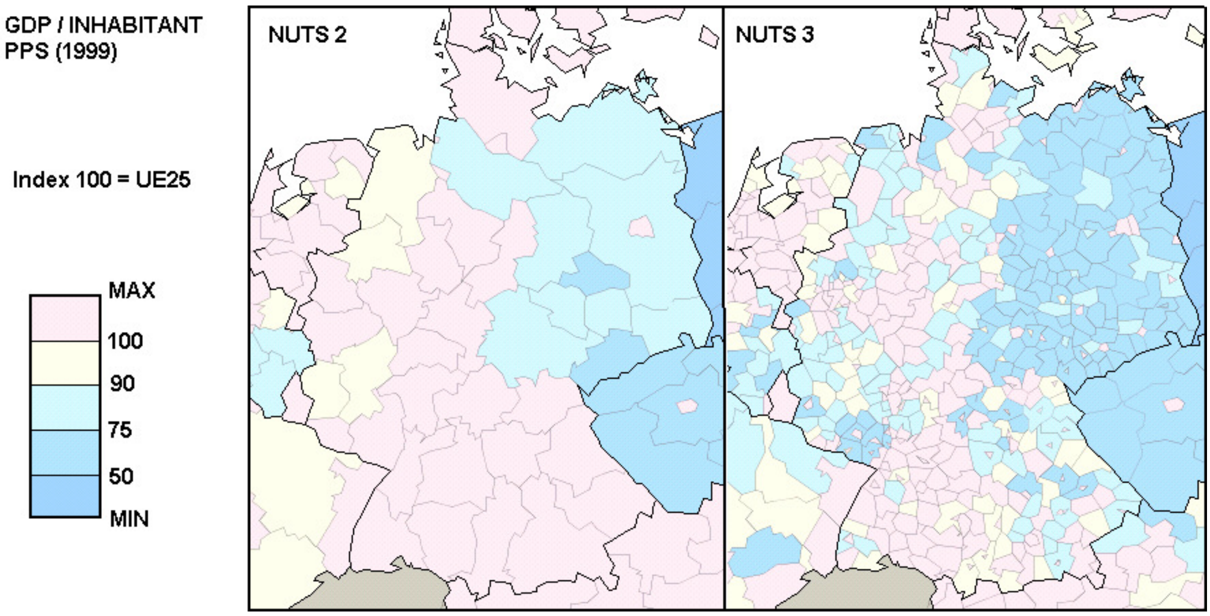
- *Question 1: What is the advantage of choosing NUTS 3 level instead of NUTS 2 as reference for ESPON results?*
- *Question 2: Is it possible to base political decision on territorial units which are officially relevant but display in practice a biased perception of reality?*
- *Question 3: What is the stability of correlation results established by the ESPON as regards to the level of territorial division? Would we have obtained different results with other territorial divisions? If yes, what are the consequences for political decision?*

The objective of this part is to clarify these questions through the analysis of selected examples of analysis of ESPON indicators.

2.1 How MAUP influences the univariate analysis

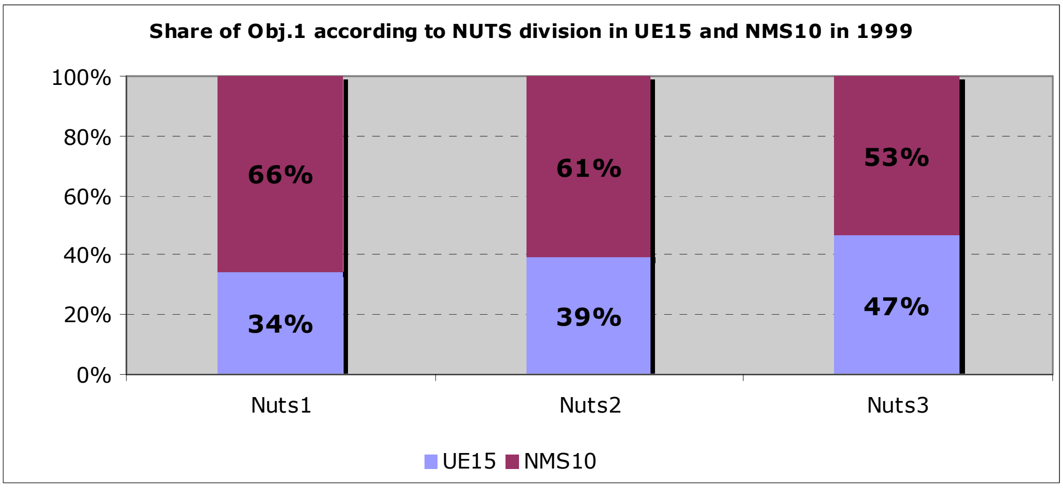
This section analyse the effect of MAUP on a single indicator, the GDP per inhabitant in 1999 expressed in parity power purchase (pps) and examine the cartographic variations and the statistical variations of results according to the choice of NUTS level 1, 2 and 3.

Cartography is a major tool for political decision in the ESPON program. But the choice of (1) territorial divisions and (2) statistical divisions has an important influence on the perception of results. As an example we propose to compare three maps presenting the distribution of regions with low level of GDP/inh. PPS in 1999 at NUTS 1, NUTS 2 and NUTS 3 levels and with three different statistical thresholds for the identification of low level (index 90, 75 and 50 with 100 equal to the mean of UE25). Dramatic cartographic changes are observed in the distribution of regions with low GDP, as it is clearly demonstrate by the example of Germany.



Changing cartographic perception of areas with low GDP/inh. pps (1999) in Germany

Statistical effects of territorial division are also crucial for the realisation of statistical tables presenting synthetic results useful for policy makers. As an example, the section analyse the amount of regions of each European state which are located under the level of 75% of the mean of EU25 for the criteria of GDP/inh. pps in 1999. The simulation demonstrate for example that the share of structural funds (Objective 1) allocated to New Member States would be different if it was based on NUTS 1 or NUTS 3 levels rather than NUTS 2. The choice of territorial units is not innocent and it is certainly not by chance that some states has chosen small or large territorial units when they have established their NUTS divisions...

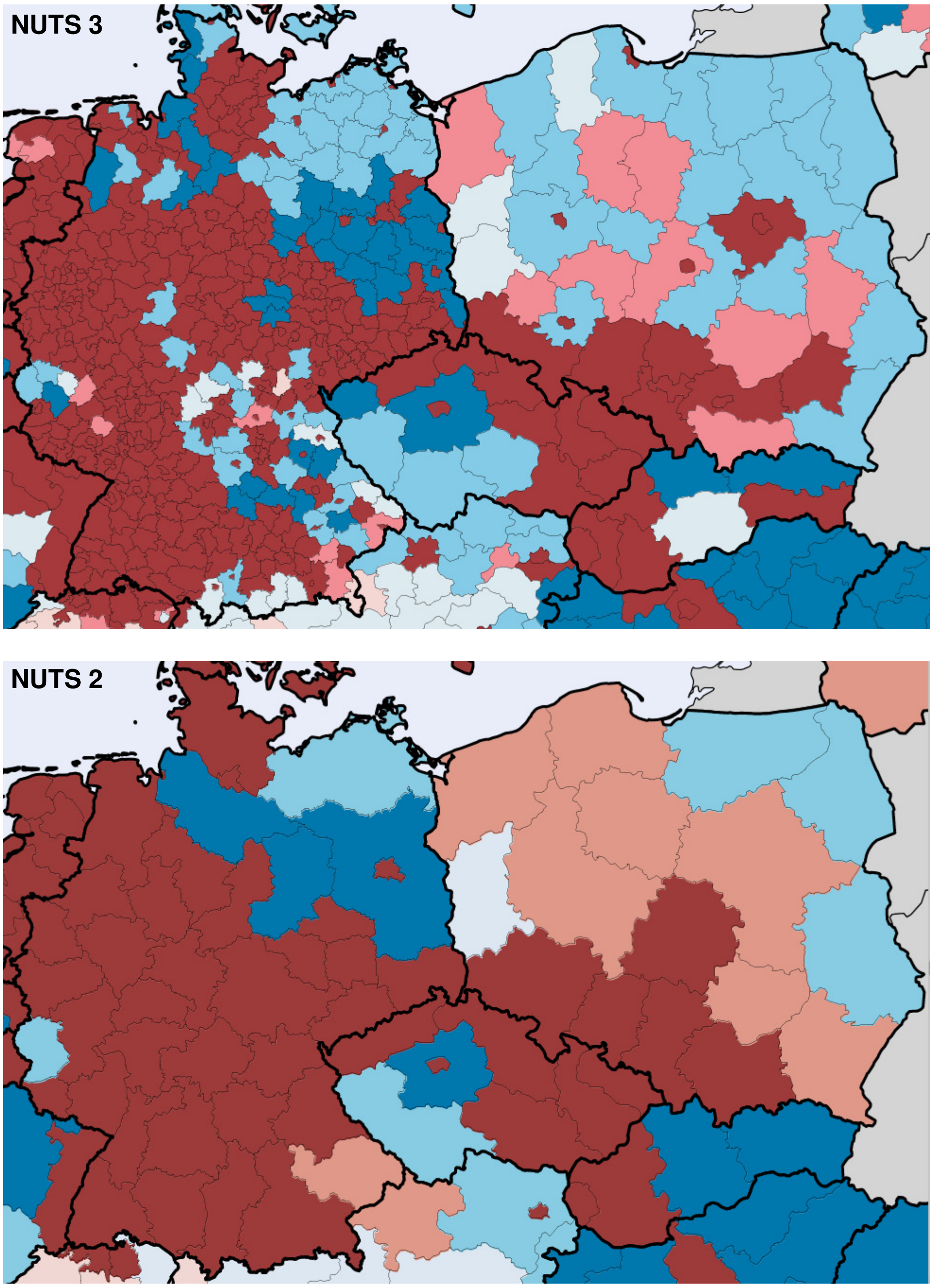


Influence of NUTS division on the allocation of structural funds in UE15 and NMC

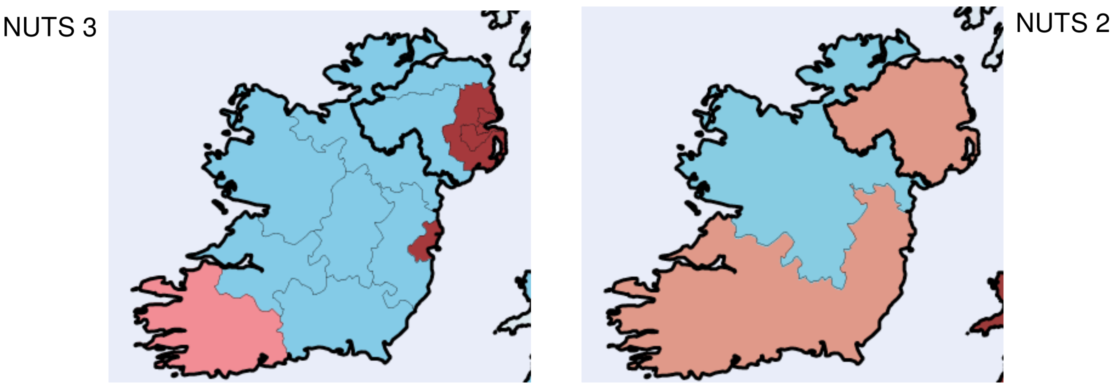
2.2 How MAUP can be seen for Cross Typologies?

In this section, we focus on the possible influence of the NUTS level on the results of cross typologies which are a major output of ESPON results. More precisely we have investigated the stability of these typologies to changes in the NUTS level used to establish them. To do so, we choose the urban-rural typology, based on population density, FUA ranking and land cover data, proposed in the project 1.1.2. and we replicated the initial typology (established at NUTS 3 level) at NUTS 2 level in order to evaluate the statistical and cartographical consequences.

On average, the same global spatial pattern of zones corresponding to the urban-rural typology is found for NUTS 2 and NUTS 3. But beyond these relative similarities observed in the global spatial structures from NUTS 3 and NUTS 2 typologies, important differences due the scale of the study can also be pointed. We examine in detail the differences in the typology based on the three criteria: population density, FUA ranking and land cover. As all of these three criteria are scale dependant, each of them can introduce different spatial modifications in the final typology. A typical and very frequent change can occur when one or several urban areas, characterised by a high population density and often by a high human intervention and defined individually at NUTS 3 level, are embedded in larger zones (Germany, Poland, …). But important differences are also related to the existence of FUA in territorial units or to the mean values of land cover (Ireland, Southern France, …).



Zoom of the urban-rural typologies on Germany, Poland, Austria and Hungary



Zoom of the urban-rural typologies on Ireland

Beside their synthetic interest, in particular at the cartographic level, the typologies are often used in the ESPON projects, in a second time, to compare the obtained classes to classical socio-economic indicators (GDP, population change and so on). As these results were based on the typology built at NUTS 3 in the case of Urban-Rural typology, our report examine how are these conclusions modified or not when the level is changed from NUTS 3 to NUTS 2. Differences can be very important in the case of area or rather limited in the case of GDP. This experiment indicates that ESPON results are generally not completely modified when changing scale from NUTS 3 to NUTS 2 and vice versa. But they prove that one should be very cautious and always mention the NUTS level which has been used for the elaboration of statistical results.

Area in km²

	Nuts 3 Total	as % from EU25+4	***Nuts 2*** Total	as % from EU25+4
Type 1	906 881	19,3%	1 070 885	22,8%
2	219 444	4,7%	445 664	9,5%
3	127 231	2,7%	460 217	9,8%
4	508 597	10,8%	566 394	12,1%
5	1 049 959	22,4%	740 891	15,8%
6	1 440 576	30,7%	975 898	20,8%
n.a.	442 103	9,4%	434 842	9,3%
Total	4 694 790	100,0%	4 694 790	100,0%

GDPpps in 1999

	Total	as % from EU25+4	Total	as % from EU25+4
Type 1	6 453 790	70,2%	6 136 414	66,8%
2	370 691	4,0%	525 665	5,7%
3	312 803	3,4%	876 917	9,5%
4	391 755	4,3%	450 045	4,9%
5	818 486	8,9%	564 942	6,1%
6	661 896	7,2%	484 163	5,3%
n.a.	182 581	2,0%	153 857	1,7%
Total	9 192 003	100,0%	9 192 003	100,0%

Urban–rural typology in relation to several indicators

2.3 MAUP and the use of multivariate analysis

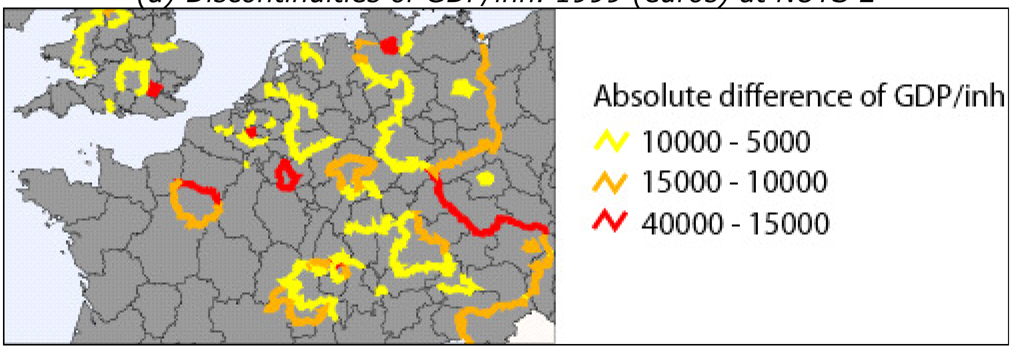
To go deeper into the MAUP question, we have completed the previous analysis by a multivariate analysis. The aim was to see how the territorial divisions can or cannot affect the results of a multiple linear regression. More precisely, we tried to account for the population variation by the other criteria and examined if the NUTS which register population decreases are characterised by a higher GDPpps per inhabitant, a lower unemployment rate or a higher population density.

The results of the analysis demonstrate important differences in statistical analysis according to the fact that the analysis of regression for the whole ESPON area is done at NUTS 0, NUTS 2 or NUTS 3. It reveals also that the factors which explains the variation of population are not the same and do not produce the same effect inside the different states of ESPON. For example, an increase of the population in a NUTS 3 can be accounted for firstly by a low density and then by a high GDP (independently of the density) for Belgium, whereas by a high density and by a low unemployment rate for Sweden. Furthermore, the NUTS 3 with an increased population are characterised by a high population density for Bulgaria, and by a low GDP for Latvia…

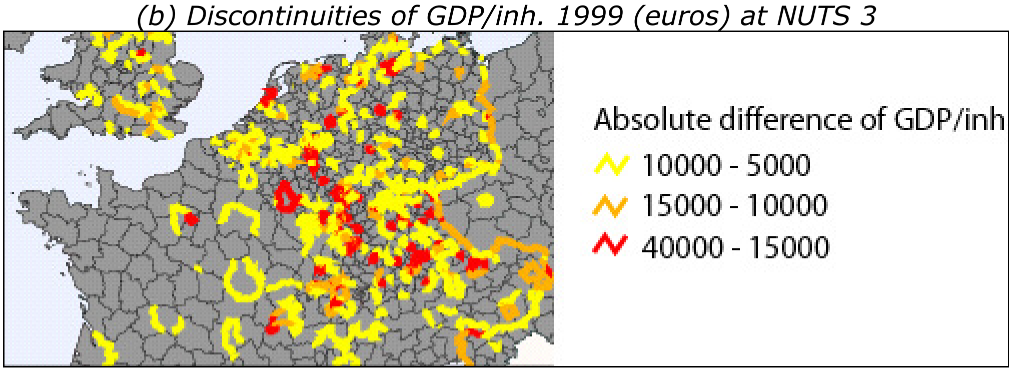
2.4 MAUP and the spatial analysis approaches

This final section examines how the MAUP has direct consequence on the results of advanced tools of spatial analysis and can completely change the cartographic and statistic outputs. All methods which are based on spatial contiguity (maps of discontinuities, local deviations, spatial autocorrelation, measure of spillover effects, …) are completely dependant from the choice of basic territorial units.

(a) Discontinuities of GDP/inh. 1999 (euros) at NUTS 2



(b) Discontinuities of GDP/inh. 1999 (euros) at NUTS 3



Impact of the choice of spatial units on the cartography of territorial discontinuities

Conclusion: "problem" or "progress"

The main conclusion of this interim report of project ESPON 3.4.3 is that Modifiable Areal Unit should not be considered as a **problem** for ESPON but rather as a challenge which could be the starting point of **progress** for territorial planning and political decisions. The project ESPON 3.4.3 could be better entitled Modifiable Areal Unit *Progress* if it succeeds in the development of new proposals which insure a better managing of various levels of territorial aggregation in future ESPON II (2007-2013). In the next months of the project, we will try to explore 4 main directions of research which do not intend to "solve" the questions that are related to changing territorial divisions but rather to "improve" the use of available information which has been yet collected by ESPON project and to help policymakers to gain better efficiency in the exploitation of results.

Combination of Nuts: The development of new levels of aggregation combining NUTS units of different levels (with minimum building blocks at NUTS 3 level) does not intend to replace actual use of NUTS 2 or NUTS 3 units in the ESPON projects but rather to complete them.

Multiscalar representations: ESPON is able to produce dynamics maps where it is possible to jump from one statistical or territorial division to another one (Hyper-Atlas, Web-GIS). They make easier the evaluation of complex situations which can not be analysed at one single level

Innovative cartography for political decisions: The question is not only to make better use of classical cartographical representations of spatial phenomena, but also to propose the implementation of new methods of representation of spatial phenomena like smoothing methods or cartograms (e.g. anamorphoses proportional to population or GDP). These innovative cartographic methods open the door for new policies at European scale.

Innovative statistical methods for territorial impact assessments: The same is true when considering the statistical dimension of the problem. The study of correlation between indexes is the basis of Territorial Impact Analysis developed in ESPON but the computation of correlation for all ESPON space is not necessary relevant and that better results would be obtained if we proceed to an analysis of correlation at different scale and in different regional contexts**.**

MAUP should not be considered as a problem but as a resource for improvement of spatial planning at European scale.

Introduction

According to the terms of reference, the IR of ESPON project 3.4.3. should be limited to a maximum number of 40-60 pages and provide to policymakers a short, clear and understandable presentation on the nature of the Modifiable Area Unit Problem from scientific point of view (**1st Part**) and set of example explaining how the MAUP can influence ESPON results and therefore how it can modify political decision (**2nd Part**). It should be derived from this a preliminary idea on tools and methods that can be proposed in order to improve ESPON results as regard to the MAUP problem (**Conclusion**).

The conclusions are supported by preliminary results of the two case studies on Northern Ireland (**Case study 2**) and Sweden (**Case study 1**) which provide very exciting illustration of the consequence of the MAUP on statistical analysis and cartographic representation of regional differentiation.

Finally, some annexes are provided in order to give more details on the experiments realised on ESPON area concerning the distribution of GDP/inh. in 1999 (**Annex. 1**), the Urban-Rural typology (**Annex. 2**) and a preliminary attempt of delimitation of new NUTS regions of equal size according to population (**Annex. 3**).

The coordinator of the project would like to thanks all project partners teams which agreed to participate with to this very exciting ESPON project despite the ridiculous[2] allocation of funds which has been allocated to a question of major interest for ESPON, especially in the perspective of future ESPON II (2007-2013).

We hope that the reading of this interim report will convince the ESPON Coordination Unit and the ESPON Monitoring Committee that additional funds could be allocated to the MAUP project in the next 3 months in order to go deeper in the exploration of promising directions described in the conclusion of this report.

[2] Each core partner team received 7 000 € which was just sufficient for paying travel expenses (TPG meeting and ESPON seminar) and elaboration of preliminary results of the Interim Report. Without an equivalent allocation of funds, it will not be possible to deepen the results elaborated in the Interim Report.

Part 1: What is the MAUP?

Introduction

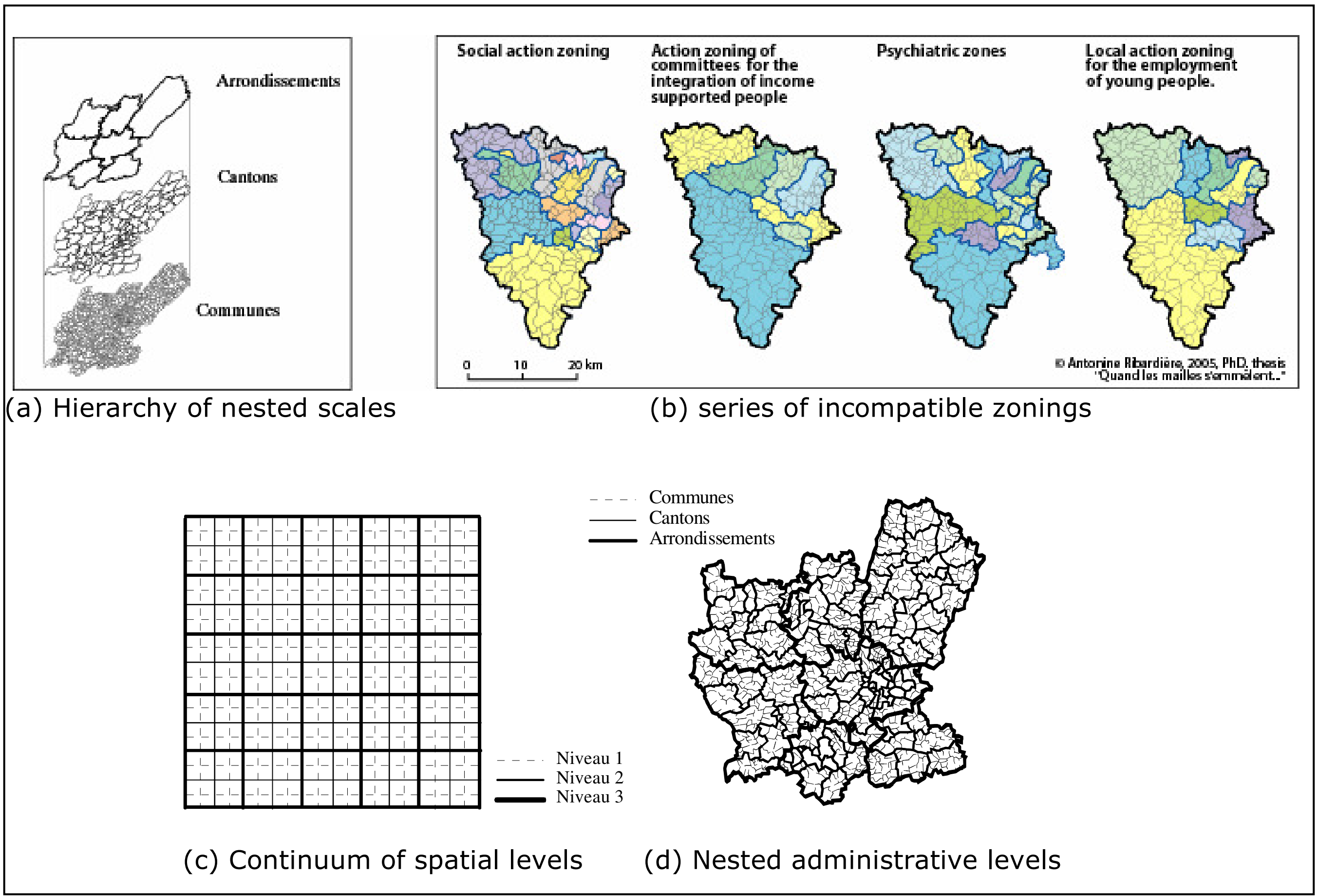
For planning purpose as well as for research topic, it is necessary to get an insight in the spatial variations of social and natural phenomena. It means to be able to have a representation of spatial inequalities, to identify specific zones, to identify discontinuities in space, and to understand the underlying principle of the spatial organization of a phenomenon as well as its correlation or independence relatively to other phenomena. Classical tools such as cartography, statistical analysis, and spatial modelling are used for that purpose. The fact is that the results of these analyses are dependent on the definition of the studied units, which depends on choices relatively to:

1) the ***scale*** or level at which the spatial entities are observed; **figure 1a** represents for example three nested levels of administrative units in France;

2) the kind of spatial ***aggregation*** which has been adopted; **figure 1b** represents an example of four incompatible zonings associated to public action in matter of fight against poverty.

In other words, the results vary according to the spatial zoning which is used in the study. For example, it can then happen that one analysis shows that wealth is negatively correlated with population density when another one will show the opposite. It can also happen that the mapping of a same indicator reveals a polycentric pattern with one spatial zoning, and a monocentric with another one. Such contradictory results could, quite naturally, be very disturbing for the decision maker…

Scale and aggregation define the resolution of the description of the phenomena. Whether the zoning is regular (**figure 1c**) or not (**figure 1d**), the spatial level which is chosen plays as a filter for the observation of the phenomena. In fact, one could compare these choices with the settings that are necessary for any instrument of measure in an experiment in physics. In a similar spirit, the choice of a spatial zoning on the one hand implies constraints, and on the other hand offers a tool of knowledge building.



(a) Hierarchy of nested scales (b) series of incompatible zonings

(c) Continuum of spatial levels (d) Nested administrative levels

Figure 1 Zonings and relations between zonings

In the following sections we will explain and illustrate the different forms that this question can take. We first introduce the question from a theoretical and methodological point of view. Then we present the main cases of MAUP which occur in thematic research or in an operational perspective[3] and we illustrate them with a series of examples. From a methodological point of view the two first examples concern the spatial organisation of a single variable (sections 1 and 2), the third one has a temporal perspective (section 3), and the two last ones concern situations where multivariate analysis is involved. At last we will discuss in what sense such contradictory results is a problem to be solved but also in what sense it could produce knowledge about the phenomena studied.

[3] In this Interim Report the aim is not to provide a complete state of the art but rather to give a clear understanding of what MAUP is, and in what sense it is a problem to ESPON decision makers. A more complete scientific discussion with full bibliographic references will be done in the Final Report.

1 Preliminary methodological consideration: Statistical distribution and spatial organisation

There is often confusion between the descriptions of *statistical* and *spatial* distribution of a phenomenon. This is often due to the definition of the elementary units, and of the level of analysis. Notice that this scale problem is crucial from political point of view as it is the basis of the distinction between social cohesion (reduction of inequalities between individuals or households) and territorial cohesion (reduction of inequalities between territories).

Let's illustrate this with an example starting at the micro level (individuals) described by their income (**figure 2**). The first statistical graphics (**case a**) illustrates a uniform distribution. This describes a *heterogeneous* situation from a social point of view, as all income categories are co-existing. There are as many households earning with lowest incomes as those with highest incomes, as those with middle incomes. In order to represent the spatial dimension of the income distribution, individuals have to be aggregate into spatial units. In that purpose, a function of aggregation has to be chosen. Part of the ambiguity comes from this operation.

First, when it is from individuals to spatial entities, the aggregation transforms totally the observed phenomena. Even if it is less evident when the aggregation concerns the passage from one spatial level to an upper one, all types of aggregation operate a transformation that one should be aware about (see section 2).

Figure 2 shows the multiple possible correspondences, between profiles of statistical distributions of a same phenomena at two different levels. These correspondences are also multiple between statistical distribution and spatial organisation of a phenomenon at a given scale: it is thus very important to precise the level of observation when studying localised information.

In the example of the income, a mixed situation (high heterogeneity) is characterized by a uniform distribution of the income when observed at the individual level. At the aggregate level of the district for instance, one could observe the average income per inhabitant (it could be the median, the heterogeneity, the range between minimum and maximum…). The social heterogeneity that is observed at the individual level may be associated either to spatial homogeneity (case a.1) or to spatial heterogeneity (case a.2). Case b is the same illustration, starting from a more egalitarian global situation from a social point of view, i.e. households are more concentrate around highest classes (case b) may lead theoretically to the same diversity of profiles of the statistical distributions at an upper level, and to the same diversity of spatial patterns. What is of interest for the question of the MAUP is also that to one statistical distribution can correspond to different spatial patterns depending on the underlying spatial processes of localisation. For instance the case a.1 corresponds to a segregation process taking place at upper spatial levels. Reversely the case b.1 illustrates a situation where the segregation process is very local. In this last case, at any upper levels the aggregation will smooth the local heterogeneity and produce a homogeneous pattern.

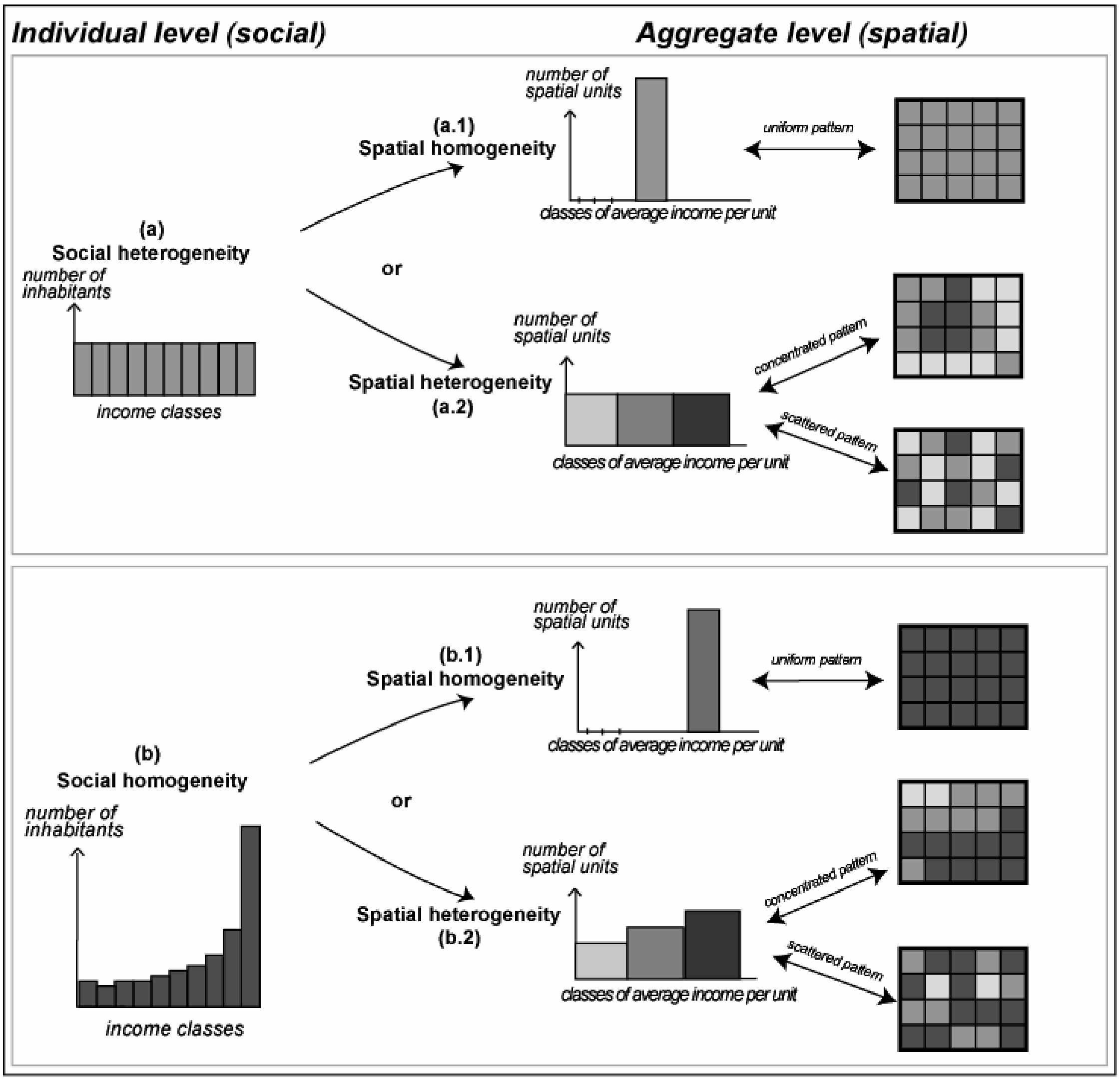


Figure 2 Statistical distribution, spatial organisation

These are really theoretical examples, and most of the time the combinatory is not as much extended. But this shows how the spatial dimension is complex even for the description of the organisation of one variable and how it is important to be explicit in order to understand the change in statistical relations.

2 With different zoning, spatial variations don't look the same

A statistical mapping is one possible representation of the phenomena among many. The perception of the global organisation is fully dependent on the zoning and the scale. **Figure 3** shows a theoretical construction of two different aggregations, with same resolution, starting from the same data. But a little move of the grid may lead to very different perception of the phenomenon.

Figure 4 shows a classical example of the "zoning effect" (change due to the zoning at a same scale) built up by Openshaw and Rao (1995): the two maps represent the spatial distribution of an indicator, the percentage of ethnic population in the county of Merseyside. The two maps represent the same phenomena, the relative importance of ethnic population in different places of the city, but they give two very different images of the associated spatial pattern:

- The map in **figure 4a** uses a zoning in 119 wards, which corresponds to conventional administrative zoning in Merseyside;
- The map in **figure 4b** uses the same number of spatial units but subdivided according to: the spatial entities are simply built from aggregation of the number of blocks necessary to reach a determined threshold of population. It means that all the spatial entities in that case have been built just to have the same population.

The two maps reveal different spatial organisation of the same phenomenon: a higher presence of ethnic population in the northern part in one case, in the south-eastern part of the municipality in the other one. This example shows clearly that the image of a phenomenon can't be interpreted and understood without having in mind the meaning of the spatial elementary units which are used for the representation.

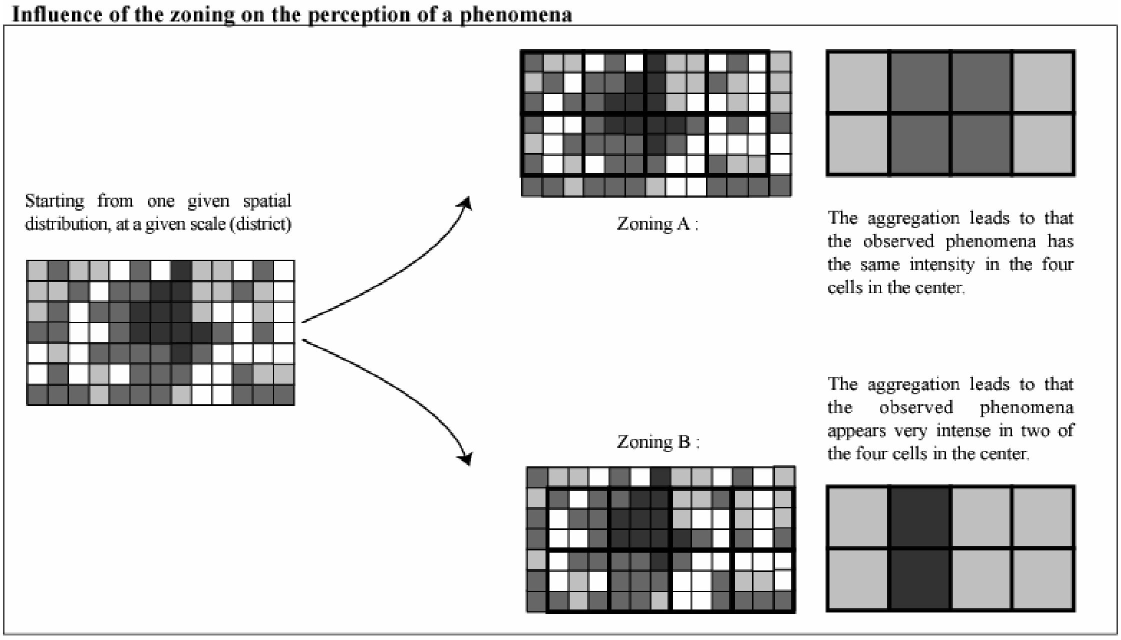


Figure 3 Influence of the zoning on the perception of the phenomena (theoretical example)

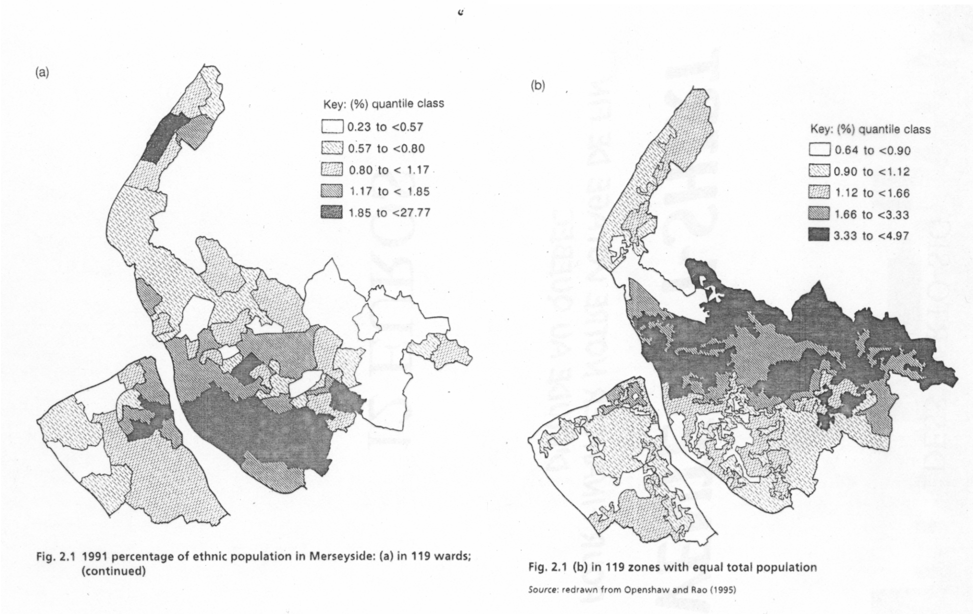


Figure 4 Influence of the zoning on the perception of the phenomena

(from Openshaw and Rao 1995)

In the Swedish case study (see Case Study 1), a systematic multiscale series of analysis have been made as well on administrative zoning as on regular square grids. **Figure 5** presents an empirical example mixing the "zoning effect" and the "scale effect". The studied region is northern Sweden, and the variable is the income per inhabitant. Four *resolution*s are compared as well as two types of zoning: two administrative ones (Counties – NUTS 3 and municipalities – NUTS 5), and two regular grids (30km squares and 10km squares). It shows obviously how scales smooth the phenomena. It shows also how interesting it is to have this multiscale representation, and how maps complement nicely each other, from level to level. Indeed, when the spatial organisation reveals juxtaposition of homogeneous regions at the upper levels, distinguishing the south, the centre and the north, the 30 kms grid map rather shows a lateral organisation with a gradient from coast to inner zones. This is very less typical in the administrative region, because of their shapes. Finally, the 10 kms grid map shows a more mosaic pattern.

This example is also a good illustration of the question of the meaning of a zoning. A zoning refers simultaneously to the delimitation of the zones and to the contents of the zones. The two first zonings of **Figure 5** are associated to a normative power, a decision-maker, when the two last ones are without any *a posteriori* meaning.

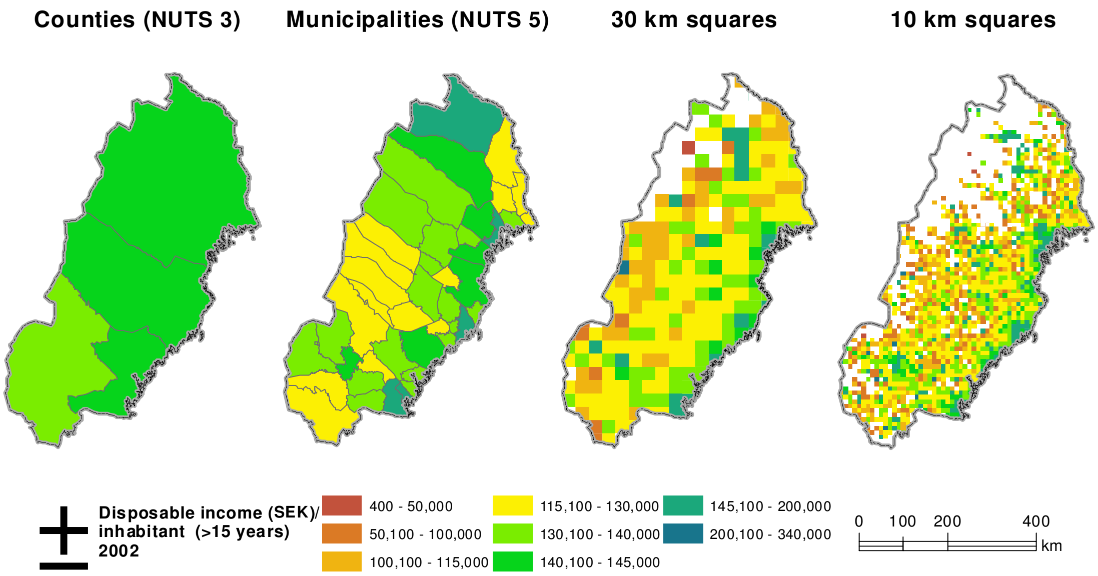


Figure 5 Influence of the zoning on the perception of the phenomena (empirical example)

(see Case Study 1, M. Strömgren, K. Holme, E. Holm)

3 Correspondence between administrative limits and spatial discontinuities: a possibility, but seldom a rule

As underlined in previous section, the spatial configuration of phenomena looks different depending on the adopted zoning. A related question is that of the *position* of spatial discontinuities or barriers in the distribution of phenomena. In some cases an administrative boundary acts as a functional limit, it is the case for what is related to policy action for example (planning, social policy, taxations etc). But when administrative units are used as the basic spatial entities for representing socio-demographic phenomena, an apparent discontinuity could be artificial.

Let's use a study about social cohesion in the Parisian metropole in order to precise this fact. **Figure 6** represents the distribution of low income household at: a) the level of the commune[4], lowest administrative entity (NUTS 5); b) the level of the IRIS[5], an equal populated division used in order to collect and represent data in intraurban areas. Two comments:

- a strong discontinuity between richer and poorer zones in northern Paris is clearly visible on **figure 6a**. It shows a gap between the city-centre and the neighbouring suburbs. Indeed the discontinuity exists but, as shown on **figure 6b**, its "real" position is moved a bit towards the city-centre. And we can notice that when ESPON uses NUTS 2 division to measure social and economic cohesion, the line of discontinuity between Paris and its suburbs is removed which is different from London where Inner/Outer London produces the most important difference of GDP/inh. in absolute terms (see. ESPON 3.1, FR).
- in some cases in the north-eastern suburbs, the mixing of different IRIS inside a commune smooth out discontinuities at the commune level, when there are strong inside the commune. For example, an opposition between a "richer" centre and a poorer periphery appears in a series of communes.

[4] The communes exist since around 1800. It is a very fine zoning as there are 36 000 communes in France, the mean area being about 12 km^2. These units correspond to old parishes and still make sense from an administrative and political point of view.

[5] IRIS are districts, divisions of the communes of more than 5000 inhabitants, defined in order to gather no more than 2500 inhabitants and to be socially quite homogeneous.

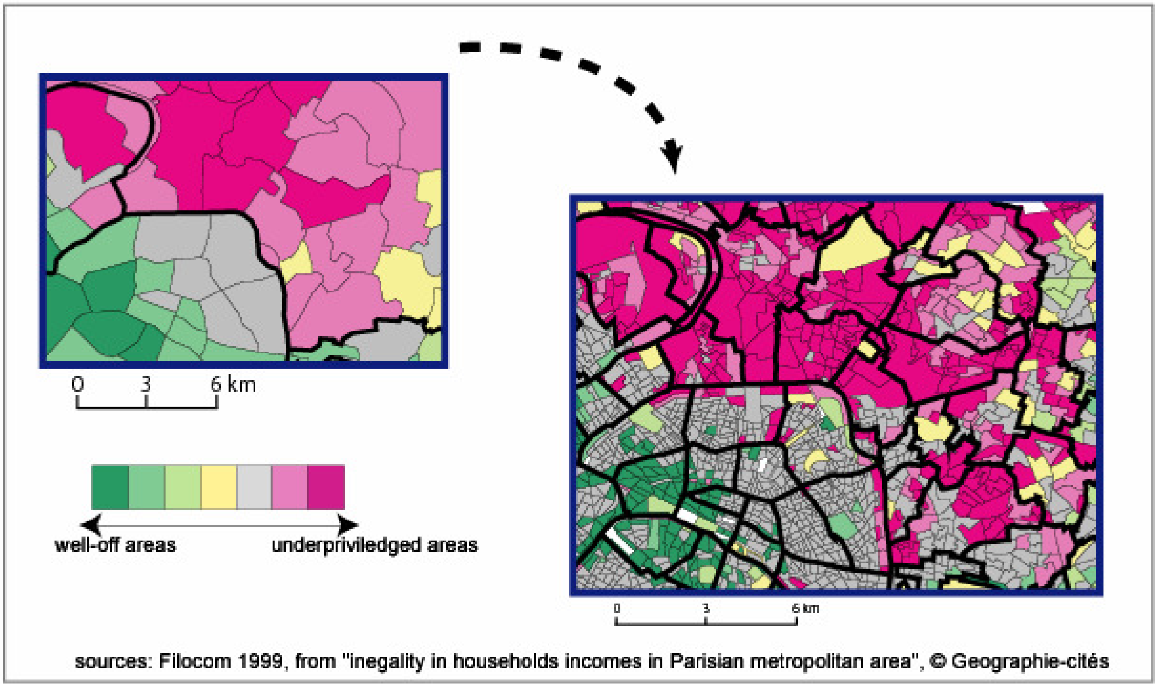


Figure 6 Social segregation to the test of scale

The scale at which spatial heterogeneity appears depends of course on the phenomena and on the geographical context. The study has shown that in the Parisian region, social mixing is observed in most communes, with the exception of extreme communes, that is the richest and the poorest ones. These communes evolve toward an increasing intra-communal homogeneity, and are the frame for the strongest mechanisms of segregation.

4 With different definitions of spatial entities, the forms of evolution appear different

Urbanization is a process about which planners and policy makers need to have a good perception of urban system as polycentrism is one main objective of ESDP. But in order to estimate the quantity of people living in cities and to visualize their spatial distribution and interlinked flows, it is of course necessary to be able to delimitate the cities. Such spatial entities are not associated with a complete and stable partition of space. And in a context of urban sprawl and development of polycentrism in metropolitan areas, the limits (borders) of the city are far from self evident. The question is even more complicated in a dynamic perspective.

Guérois and Paulus (2002) have shown how different choices in matter of definition of French urban units result in different images of urban evolution. In the French official statistics three definitions have emerged through time:

- the "city-center" which corresponds to the commune, the lowest administrative level in France (NUTS 5), which is the historical core of the city, which made sense until the first half of the 20th century;
- the "urban unit" (introduced in 1954) corresponds to a morphologic definition of the city and aggregate all contiguous built-up areas;
- the "urban area" (introduced in 1997) corresponds to a functional definition integrating the city and its periurban area. It is delimited according to the percentage of the labour force working in the city (with an iterative method of construction).

As shown in **figure 7**, the urbanization evolution looks in France radically different depending on the adopted delimitation: urbanization seems either to decrease, to stagnate or to increase. From a thematic point of view, such differences can lead to an interpretation in terms of counter-urbanization or of increasing urban concentration.

In the ESPON context, it means that very different patterns of FUA's could be established, producing very different vision of polycentric development of Europe. We can recall the discussion around a preliminary map of project 1.1.1. where Paris was reduced to 2 instead of 10 millions of inhabitant, as the author has chosen to use the city center instead of the agglomeration as criteria of delimitation.

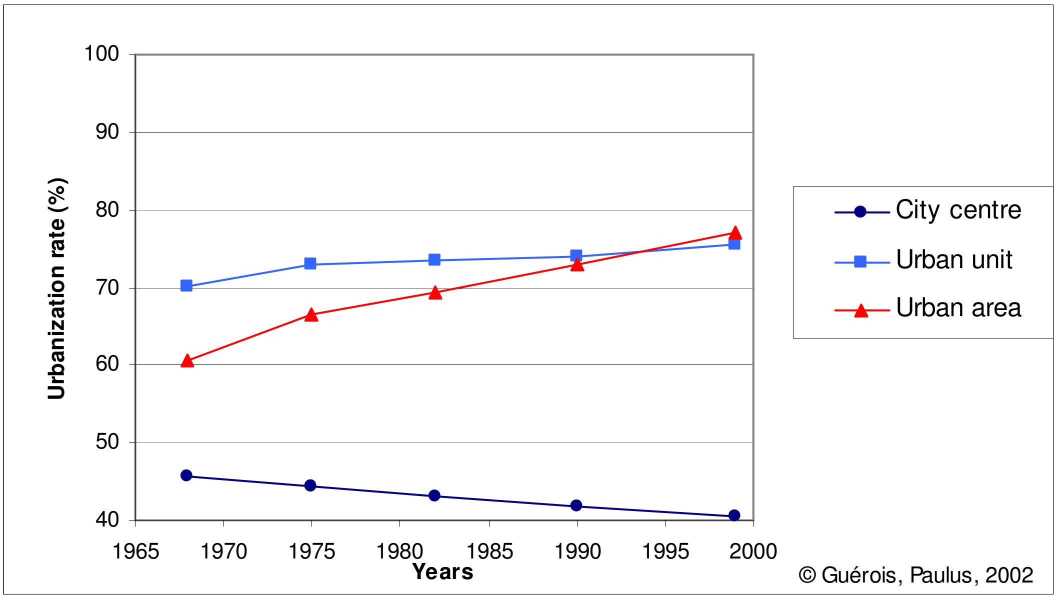


Figure 7 Evolution of the urbanization rate in France depending on the urban delimitation

Depending on the problematic, one definition will be more adapted than another one. And from a research point of view, the combination of the three curves gives the most complete information about the urbanization phenomena, showing a double process of concentration at a higher scale and deconcentration at the local level.

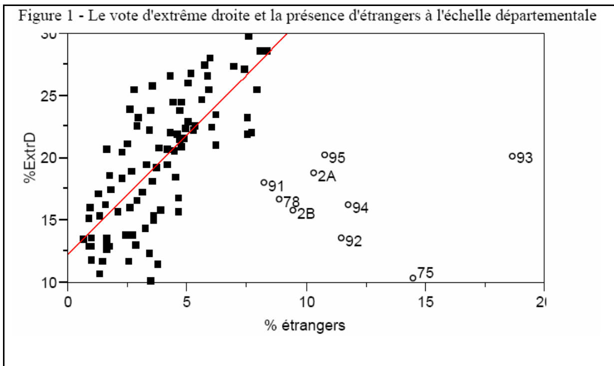
5 The relations between phenomena seem different depending on the scale of observation and with the spatial extent of the study region

Statistical analyses show the intensity, significance and sign of the relationships between variables. Such analyses help to explore the links between different phenomena and in the case where the statistical units are spatial entities, the links between different spatial patterns. But, the user should be aware that the value of a result (a coefficient of correlation for example) could vary according to the spatial units at the level of which the variables have been observed. It implies that the generalization of a result to other scales or to other spatial zonings most often doesn't make sense. But, it means also that, when contradictory results appear, the study of the mechanisms underlying these differences between the results will give a better understanding of spatial differentiation.

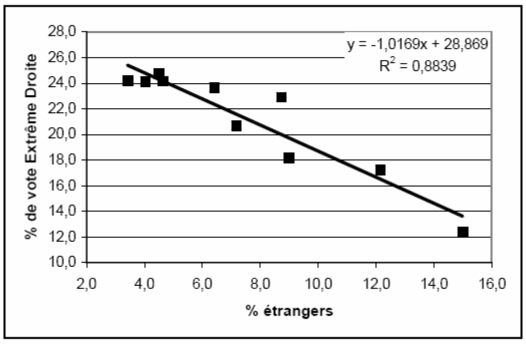
This may be illustrated by a study of Ravenel (2003). His aim was to analyse the relation between the importance of the foreign population and the vote for the extreme right. The variables are first observed at the level of the 95 French departments (NUTS 3). If one except from the departments of Corsica and the Parisian region, there is a strong positive relations between the two phenomena (**figure 8a**): the vote for the extreme right is more important in the departments where the part of the foreign population is higher. In a second analysis he chooses as spatial entities successive rings around Paris every 10 kms until 100 kms. The relation is this time strongly negative (**figure 8.b**), the vote for the extreme right registering its lower values in the circles with the highest level of foreign population.

Considering the same spatial area (a circle centred on Paris with a radius of 100 kms), but observing the variables at the level of the 3173 communes (NUTS 5) composing that area, the result is that there exist no significant relation between the vote for the extreme right and the presence of foreign population (**figure 8.c**). At last, the author concentrates on the communes located within 10 kms from the centre of Paris. The relation appears then clearly positive (**figure 8.d**).

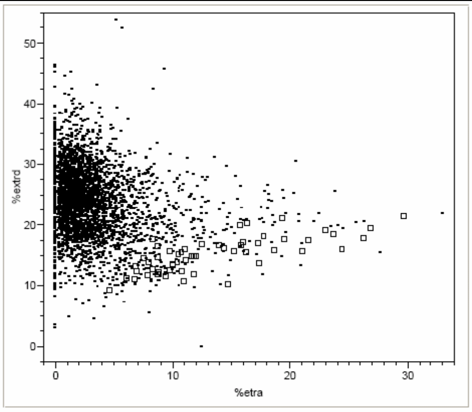
So the relations between the same couple of variables appear to be strongly positive, strongly negative, or without any significance depending on the spatial zoning and frame which are used in order to determine the spatial entities of observation. The author underlines that these variations are associated to different models of explanation which each refer to a specific spatial scale. A classical explanation is produced by a causality scheme claiming that more foreigners produce a xenophobic reaction. Another one is that a same phenomenon, as a certain socio-economic context, is simultaneously associated with high level of foreigners and high level of extreme right votes. And a socio-economic context has not either the same meaning at different scales. Indeed, in the Parisian region, there are strong gradients from centre to periphery for population of higher socio-economic groups AND for foreign population. The former ones are rarely voting for the extreme right, which explains the negative relation represented on **figure 8.b**. But at a local level, foreigners are overrepresented in the communes with lower socio-economic groups, who vote more often on the extreme right. The socio-economic context is fundamental at both levels but it has a different meaning, related to the scale at which such or such category is or is not concentrated.



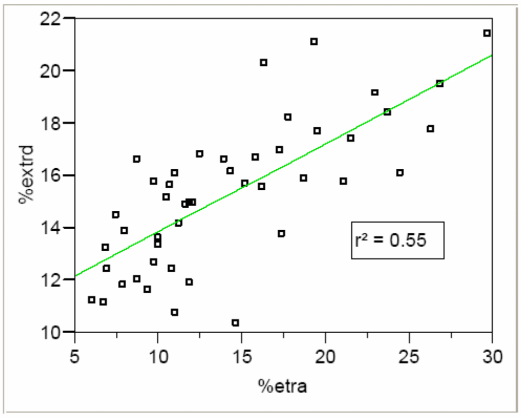
(a) % of foreigners and % of extreme right votes for the 96 French departments



(b) % of foreigners and % of extreme right votes for rings around Paris (until 100km)



(c) % of foreigners and % of extreme right for all communes at less than 100km from Paris



(d) % of foreigners and % of extreme right for communes close to Paris (<10km)

© Ravenel, 2003

Figure 8 Relations between extreme right votes and foreigners rate: different filters

Full examples have been developed on Swedish and Irish case studies (see Case studies 1 and 2). Indeed it is interesting to see how a correlation basically measured at the individual level (income and density for instance) is scale dependent, and change with the range and meaning of the neighbourhood which is taken into account.

6 And what about the effect of a boundary on the relation between phenomena

One case study explores systematically this question with a multivariate statistical analysis on socio-economic variables for Ireland and Northern Ireland (Case study 2). From a theoretical point of view, the question is "to what extent do the observed differences in statistical analyses on either side of a national border result from data on either side being drawn from incompatible spatial units" or from the fact that the socio-economic processes which are operating on each side of the border are different. The authors apply the same regression model (proportion of population having a university degree being regressed on a set of socio-economic indicators) in the case of 10 different data sets corresponding on the one hand to different scales of observation and on the other one on different extents of the area being observed. They show the sensitivity of the results to these variations: for example the effect of catholic religion on educational attainment is positive in Northern Ireland and negative in Ireland; and the effects of social class on education attainment are greater in Northern Ireland than in Ireland.

7 In what sense is the MAUP a problem?

Cartographic representation and statistical analyses of phenomena are giving different results according to the elementary spatial entities chosen for the study, that is a fact, but is it a problem? The denomination MAUP suggests it, and in some situations it obviously is. But it is not systematically the case and Openshaw, who is a reference for this question (MAUP), underlined that "according to one given domain, geographers would agree on what constitute objects of research and if these objects would be defined in a non arbitrary way, the problem of the spatial aggregation would disappear "(Openshaw, 1981).

- why it is a false problem

Let's do a parallel with the ecology fallacy which is an associated question. As it clearly comes out from **figure 8**, the inference of a correlation computed at the level of spatial entities to the individuals is in no case justifiable and could even be dangerous[6]. Such inference consists in the deny of the spatial dimension of people's settlement. As shown in section 3 spatial mixing (co-existence of people of different categories) or, at the opposite spatial segregation of populations of different categories, could be of very varied forms according to the context. Spatial mixing and segregation can then simultaneously characterize a same phenomenon but at different levels. Such variations are in themselves an object of study for the geographer. The differences in the results of statistical analyses according to the scale are, in that sense, creating knowledge about the observed phenomena. The identification of the scales at which segregation for example does or does not exist, learns a lot about the organization of the concerned society.

- why it is a real problem

The most important problem is about international and historical comparisons: do the elementary spatial units which are used for the analysis have the same meaning in two different countries? At two different time periods? It is not easy to determine if a difference in the results is due to a difference in the processes which are underlying the observed phenomena, or simply to a difference in the meaning of the spatial entities that are used for the observation. In that sense, to work on a grid is a good alternative. But, as shown by the Swedish case study (case study 1), the information given by analysis at a grid level or at an administrative entity level are complementary. One choice is not globally *better* than another one.

Conclusion

We tried in this first part to present all the ambiguity of the problem of the spatial aggregation. This characteristic is simultaneously the essence of the spatial data and its constraint.

In a *statistical* point of view, results to be reliable have to be stable and rather independent of point of measure. But as it has been often stressed, this is also a "*non geographical* position" (Openshaw, 1981) and denied of what is consistent in the spatial dimension. One progress in this question is to accept that results are zoning dependent.

[6] A very classical example from the US in the 1950s (Robinson, 1950) is that of the correlation between percentages of blacks and crime rate at the level of a city's quarters…

Then it implies to clearly identify the question and the chosen zoning. In zoning, one do not forget that there is either the delimitation (container) and the inside (contents). Thus two approaches may be adopted: to choose the best contents coherent with the question or to prefer the a priori choice of the delimitation independently with the question but perhaps more adapted to the methodology used for the analyse. None may solve the problem, but the conscience and the reflection on of the spatial zoning definition and these two components is certainly a necessary condition to advance in this problem.

Part 2: ESPON and the MAUP

Introduction

The question of the potential influence of changing territorial divisions on applied research in the field of spatial planning has been regularly discussed in the framework of the ESPON program and even before.

A working paper called "Objective 13-bis" was presented in 1998 by C. Grasland in the Study Program on European Spatial Planning (SPESP) which is the ancestor of ESPON and was published as an annex in the report of the working group I.4 Spatial integration[7]. This provocative paper was based on the fiction of new structural funds allocated to regions with decreasing population (the so-called "objective 13-bis") and tried to examine how Belgium could maximize its benefits according to the choice of various levels of territorial division from NUTS 0 to NUTS 5. It demonstrated that the choice of the optimal level was not obvious and that the maximum amount of population eligible to this fictive objective was obtained at a medium level of territorial division and not necessary with the larger or smaller territorial units. Furthermore, this paper proposed an alternative approach of identification of target areas based on multiscalar smoothing methods. The results of this research were further developed and published in a scientific journal[8].

This question was not immediately recognised as important by ESPON which defined NUTS 3 as target level of analysis for all projects in their terms of reference. NUTS 2 level was admitted as alternative target level only when data collection was obviously impossible at NUTS 3 level. Many lead partner complained that the obligation to develop analysis at NUTS 3 level had terribly hampered their research because of time necessary to elaborate regional data base which did not necessary exists in the European statistical systems, especially Eurostat where most data are available at NUTS 2 and, only a small proportion at NUTS 3. Typically, the teams in charge of elaboration of synthesis results (like ESPON projects 3.1, 3.2, 3.3 or 2.4.2.) has all decided to use NUTS 2 as reference level because they would have been obliged to eliminate more than half of the information available in the ESPON database if they had chosen the NUTS 3 level as reference. This lead to a first question:

[7] http://www.mcrit.com/SPESP/SPESP_REPORT/spatial_integration.pdf

[8] C. Grasland, H. Mathian, J.M. Vincent - « Multiscalar analysis and map generalisation of discrete social phenomena: Statistical problems and political consequences », *Statistical Journal of the United Nations,* ECE, 17, IOS Press, 1-32. 2, 2000

Question 1: What is the advantage of choosing NUTS 3 level instead of NUTS 2 as reference for ESPON results?

At the same time, a more general debate appears in the ESPON program about the opportunity to use systematically official NUTS divisions instead of building specific *ad hoc* division for each relevant topic of interest. The discussion on this point emerged during the ESPON seminar of 2003-2004, especially in the debates around the presentation of the results of projects 1.1.1. (Polycentrism) and 1.1.2 (Urban-Rural relations) because it was very clear that all variables related to the measure of urban phenomena were very sensitive to the question of territorial delimitation. Firstly, it appears that the delimitation of relevant urban units (FUA) implies the use of lower level of territorial units than in the other ESPON studies (NUTS 5). But at this level, information was very scarce and in many cases the economic attributes of MEGAS were based on NUTS 3 information, producing scale conflicts and debatable approximations. Secondly, the construction of typologies based on the presence of FUA or megas in European regions appeared clearly biased at NUTS 3 level because the size of territorial units is clearly correlated with the potential number of towns located in a given area. More generally, the population density and the variation of population are strongly correlated with the size of NUTS units because intra-urban, peri-urban and rural territories are generally strongly different in terms of territorial dynamics. In countries with large NUTS 3 divisions like France or Spain, the three categories of spatial structure (intra-urban, peri-urban, rural) are always "mixed" and the result is a "mean value" with small inter-regional differences. In countries with smaller NUTS 3 units like Germany or Belgium, each category of spatial structure can be isolated by the territorial division and the result is a very high statistical and cartographical heterogeneity of the distribution of population densities or population dynamics. It is not possible to assets that one scale of analysis is better than the other (depending on the target of the analysis) but it is obvious that the combination of units of different scales in the same table or the same map is a mistake because it produces confusion and give the false impression that some parts of the European territory are more differentiated and heterogeneous than the other. It is true that NUTS units are **politically relevant** because they are associated to political levels of decision but they are **not empirically relevant** because they introduce a confusion of scales, especially at NUTS 3 level where urban-rural differentiations are visible in certain states but not in the others. This leads to the second question:

Question 2: Is it possible to base political decision on territorial units which are officially relevant but display in practice a biased perception of reality?

Last but not least, it has been also underlined during several ESPON meetings, (especially the one of Matera) that MAUP is not only a cartographical problem but also a statistical one because correlation between units can be modified when territorial units are modified. This problem which is less "visible" appears as the major one because correlations are often a basis for policy recommendations and for studies of Territorial Impact Assessments. For example, the existence or non existence of a correlation between accessibility and economic competitiveness (growth of GDP/inh.) is a crucial input for the elaboration of political decisions related to the development of Trans European Network. The existence or non existence of correlation between Common Agriculture Policy funds (Pillar 1 or Pillar 2) and other structural criteria (GDP/inh., Unemployment, Population growth) is also of crucial importance for the evaluation of positive or negative effects of CAP on the objective of territorial cohesion. Major issues related to policy options proposed by the ESPON projects are really based on such identification of correlations between indicators established at NUTS level. This lead to the third question:

Question 3: What is the stability of correlation results established by the ESPON as regards to the level of territorial division? Would we have obtained different results with other territorial divisions? If yes, what are the consequences for political decision?

The objective of this part is to clarify these questions through the analysis of selected examples of analysis of ESPON indicators. We will use the most classical or emblematic variables (GDP/inh. PPS, Unemployment rate, ...) to provide examples which will be organised in an order of growing complexity, from simple univariate analysis to spatial analysis tools. Part of these results was firstly elaborated in ESPON project 3.1 (Dictionary of Tools, Final report) but have been here completed and presented in a more systematic way.

1 How MAUP influences the univariate analysis

We consider here the analysis of a single indicator, namely the GDP per inhabitant in 1999 expressed in parity power purchase (pps) for all ESPON units with the exception of ultra-peripheral regions of Spain, France and Portugal. We use the official NUTS division of 1999 which is different from the most recent one (2003), especially for new members states which had not established NUTS1 levels units at this time.

1.1 Influence of NUTS division on cartographic perception of results

Cartography is a major tool for political decision in the ESPON program. But the choice of (1) territorial divisions and (2) statistical divisions has an important influence on the perception of results. As an example we propose to compare three maps presenting the distribution of regions with low level of GDP/inh. PPS in 1999 at NUTS 1, NUTS 2 and NUTS 3 levels and with three different statistical thresholds for the identification of low level (index 90, 75 and 50 with 100 equal to the mean of UE25). The comparison of the maps established at level NUTS**,** NUTS 2 and NUTS 3 which are presented in Annex 1 reveals very different patterns of regions with low level of GDP/inh. which can be illustrated by the examples of Portugal and Germany.

- **Portugal (Figure 9)** is defined as a single territorial unit at level NUTS 1 with a GDP/inh. comprise between 75 and 90 % of European mean, mapped in light blue. At level NUTS 2, we have a very different pattern with the capital region of Lisbon above European mean (red) and three other regions with a low level of GDP/inh. comprise between 50 and 75% of European mean (medium blue). At level NUTS 3 a completely different picture is revealed with a mixture of units providing all possible levels (and all possible colors from deep blue to red). The capital region of Lisbon which appeared as a relative large red area at NUTS 2 is transformed into a mosaic of micro-regions generally above European mean, with the exception of the urban agglomeration of Lisbon which is now reduce to a very small area.
- **Germany (Figure 10)** is certainly one of the most spectacular examples of modification of the perception. At NUTS 1, all Länder from eastern Germany except Berlin have relative low values (light blue, between 75% and 90% of EU25 mean) and all other Länder from western part have relative high values (red, above 100% of EU25 mean) producing the vision of a complete West-East opposition. This feeling of opposition begins to change at NUTS 2

with the apparition of local differentiations in western part (in yellow or light blue) as well as in eastern part (with area in medium blue). At NUTS 3 level, a completely different picture appears with a mosaic of colors indicated by very different levels both in eastern part (where some urban Kreis are above EU mean) and western part (where many rural Kreis are in blue).

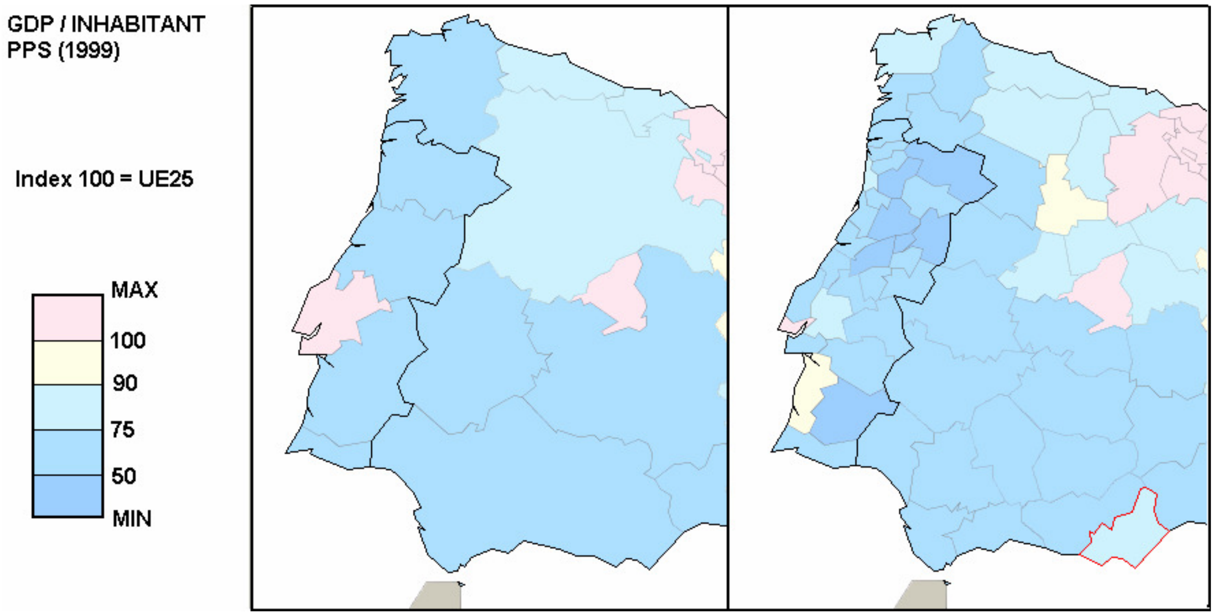


Figure 9 Changing cartographic perception of areas with low GDP/inh. pps (1999) in Portugal

(Maps established with the ESPON Hyper-atlas)

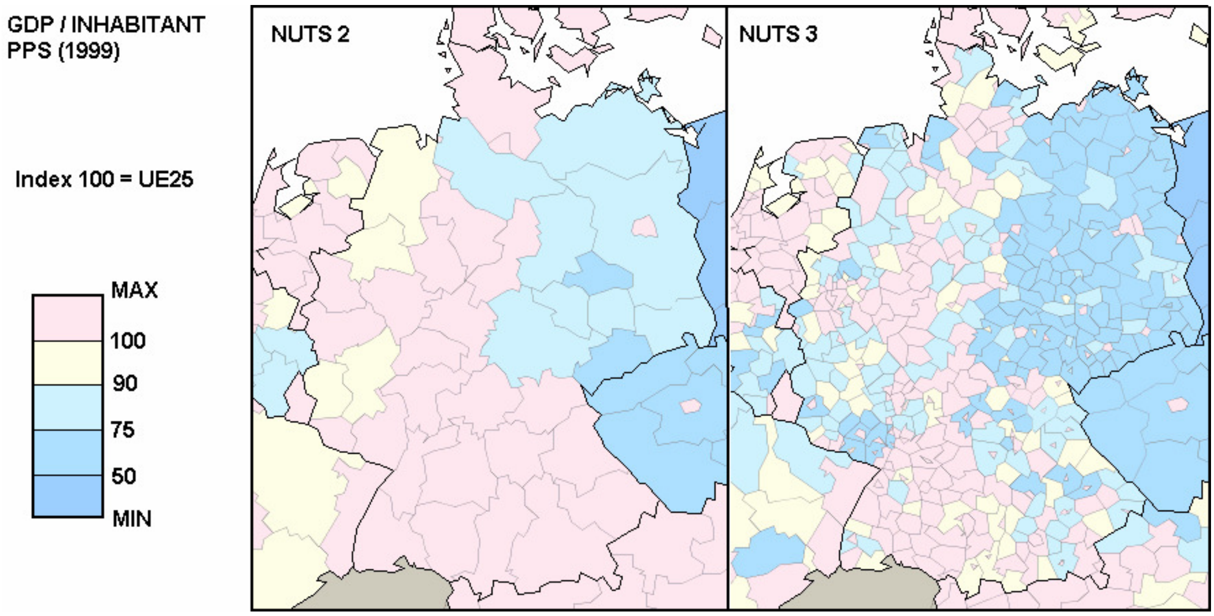


Figure 10 Changing cartographic perception of areas with low GDP/inh. pps (1999) in Germany

(Maps established with the ESPON Hyper-atlas)

Cartographic changes in visual pattern are not always as dramatic as in the cases of Portugal and Germany, but it is perfectly clear that **policymakers which try to analyse spatial patterns of regional development will receive completely different information according to the level of spatial aggregation.** The experiments on visual perception proposed by CESA (Tours) could be of major interest for the analysis of this phenomenon.

1.2 Influence of NUTS division on statistical analysis of results

Let us now consider the effect of territorial division on the realisation of statistical tables presenting synthetic results useful for policy makers. As an example, we have chosen to analyse the amount of regions of each european state which are located under the level of 75% of the mean of EU25 for the criteria of GDP/inh. pps in 1999. It is theoretically the criteria used for allocation of structural funds, but it is important to keep in mind that **this example is a fiction** as (1) some peripheral regions have not been included in the analysis and (2) the mean of UE25 is applied to 1999 at a moment where UE was restricted to 15 countries[9]. We do not discuss here the complete results which are presented in **Tables 3**, **4** and **5** of the Annex 1 and we focus on selected examples.

- **The shift of allocation of funds between old and new member countries** presented in **Figure 11** indicates a general tendency of reducing the share of new member countries when the number of territorial units increases. This is logical if we consider that increasing territorial divisions will produce the apparition of local peaks of poverty in richest countries of western Europe (increasing their share of eligible regions) and reversely produce the apparition of local peaks of wealth in poorest countries of Europe (reducing their share of eligible regions).

[9] We have intentionally chosen to present an example which is not directly applicable to actual debate on reform of structural funds in order to avoid reactions related to the defense of national interests. But the example is sufficiently close to actual "hot cakes" of political debate in order to show the importance of the MAUP. It would indeed be very easy to apply the same method on actual data...

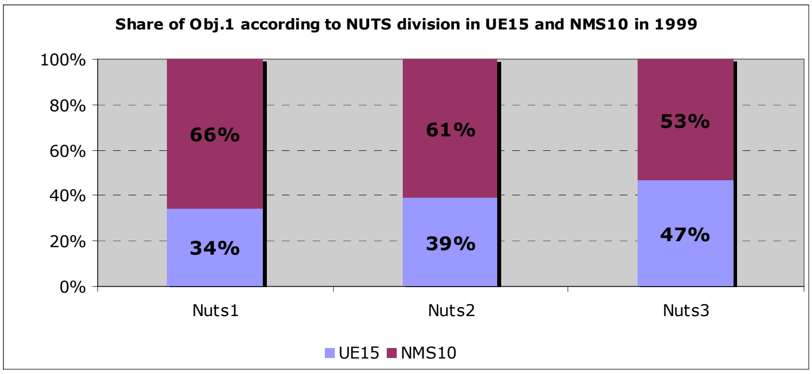
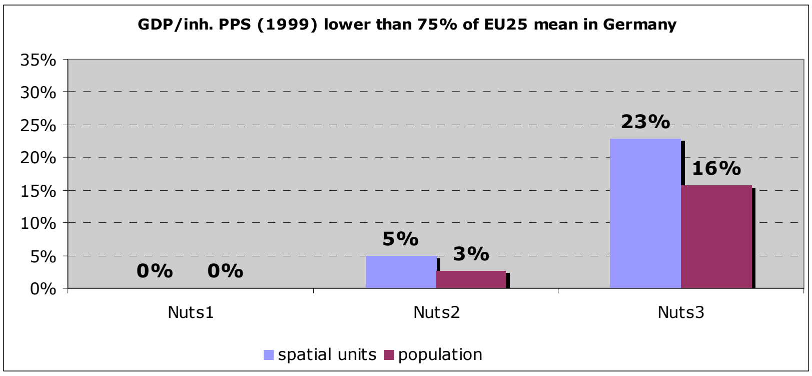


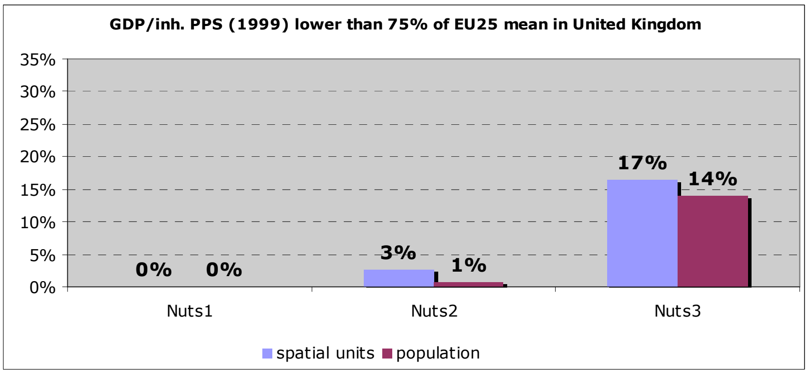
Figure 11 Influence of NUTS division on the allocation of structural funds in UE15 and NMC

- **The national examples** presented in **Figure 12** demonstrate clearly that the choice of the territorial level is an important challenge for European states as far as ESPON results can be used for the allocation of structural funds. In the case of Germany and United Kingdom, the number of regions eligible to Objective 1 is equal to 0% at NUTS 1 but rises to 1-5% at NUTS 2 level and could eventually reach 10-20% at NUTS 3 level. The situation is different in Italy where it is at NUTS 1 level that the most important proportion of regions would be eligible and where we observe a decrease at NUTS 2 and NUTS 3 levels. In Greece, we have a more complicated situation where the maximisation of benefits is obtained either at NUTS3 (82% of regions representing 49% of population) or at NUTS1 (only 50% of regions but representing 58% of population).

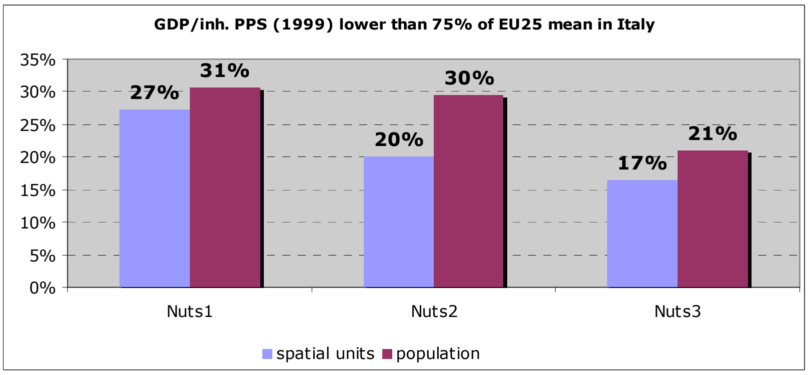
(a) Germany



(b) United Kingdom



(c) Italy



(d) Greece

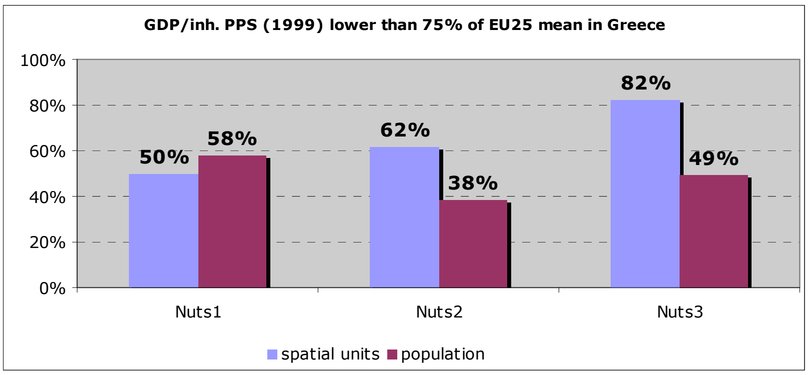


Figure 12 Simulation of the percentage of regions and population eligible to structural funds in different states according to the choice of NUTS division in 1999

This example is of course a fiction (see. Above) but it demonstrates clearly that the choice of territorial units is not innocent and that it is without any doubt an important topic for political negotiation. It is certainly not by chance that some states has chosen small or large territorial units when they have established their NUTS divisions...

2 How MAUP can be seen for Cross Typologies?

In the ESPON projects, typologies are often used, especially those combining of several phenomena (such as crossed typologies or other more complex clusters of regions). These typologies are often based on data at the NUTS 2 level and their results are then compared to several classical socio-economic indicators.

In order to explore further the possible influence of the NUTS level on the results, we have investigated the stability of these typologies to changes in the NUTS level used to establish them. To do so, we choose the urban-rural typology, based on population density, FUA ranking and land cover data, proposed in the project 1.1.2[10]. This choice is motivated by several reasons:

- Firstly, the latter typology, which is presented as a final goal in the project 1.1.2 summary, was used to produce one the emblematic maps of ESPON (see the Map 1 of Annex 2). This typology divides these European NUTS 3 data into six classes as a function of the urban influence (low/high) on one hand, and of the human intervention (low/medium/high) on the other hand (see Annex 2).
- Secondly, this typology is chosen because it is based on data at the NUTS 3 level and it was thus interesting to study the consequences of a change from NUTS 3 to NUTS 2 level on the cartographic renderings and on the associated statistical results.
- Finally, we choose this typology because their authors themselves tried to evaluate the consequences of a scale change on the robustness and on the validity of their analysis. For example, they slightly discuss the statistical biases that can arise from the time series they are dealing with on one hand, and they test their typology on the data from two different national contexts (i.e. Belgium and Austria; see Maps 2 and 3 of Annex 2[11]).

[10] See http://www.espon.lu/online/documentation/projects/thematic/thematic_7.html.
[11] They used NUTS 5 as reference territorial units; the criteria for urban influence and the degree of human intervention are based on the means of the considered countries.

For this first interim report, we re-used the methodology of the previous authors to build an urban-rural typology based on data at the NUTS 2 level. Let us first mention that in the same time, we redid their analysis at NUTS 3 level to check the coherence with their work and to allow for a precise comparison between typologies at NUTS 3 and NUTS 2 levels. For the 1329 NUTS 3 used we obtain the same typology except for 9 territorial units[12]. This discrepancy seems to be due to slight differences in the shares of land cover.

The spatial repartition of the urban-rural typology based on population density, FUA ranking and land cover is presented at NUTS 3 level on the **figure 13** and at NUTS 2 level on the **figure 14**. As pointed out for the NUTS 3 level in project 1.1.2, the distribution remains uneven at NUTS 3 level. In average, the same global spatial pattern of zones corresponding to the urban-rural typology is found for NUTS 2 and NUTS 3. As can be seen, the distribution of territorial units with high/low urban influence (red/blue) is as uneven whatever the base unit taken into account. Moreover, the high human intervention zones are mainly along an axis that goes from southern England to Bulgaria through Germany whereas low and medium human intervention zones are more characteristic of southern and northern Europe.

But beyond these relative similarities observed in the global spatial structures from NUTS 3 and NUTS 2 typologies, important differences due the scale of the study can also be pointed. The cartographic changes due to the use of different territorial units have been described before in the theoretical part of this FIR and in the part related to univariate analysis. We would like to insist here on the differences in the typology based on the three criteria: population density, FUA ranking and land cover. As all of these three criteria are scale dependant, each of them can introduce different spatial modifications in the final typology.

[12] Dinant (BE351), Philippeville (BE353), Schwandorf (DE239), Oberhavel (DE40A), Potsdam-Mittelmark (DE40E), Storstrøms amt (DK006), La Coruña (ES111), Buzau (RO022), Sibiu (RO076).

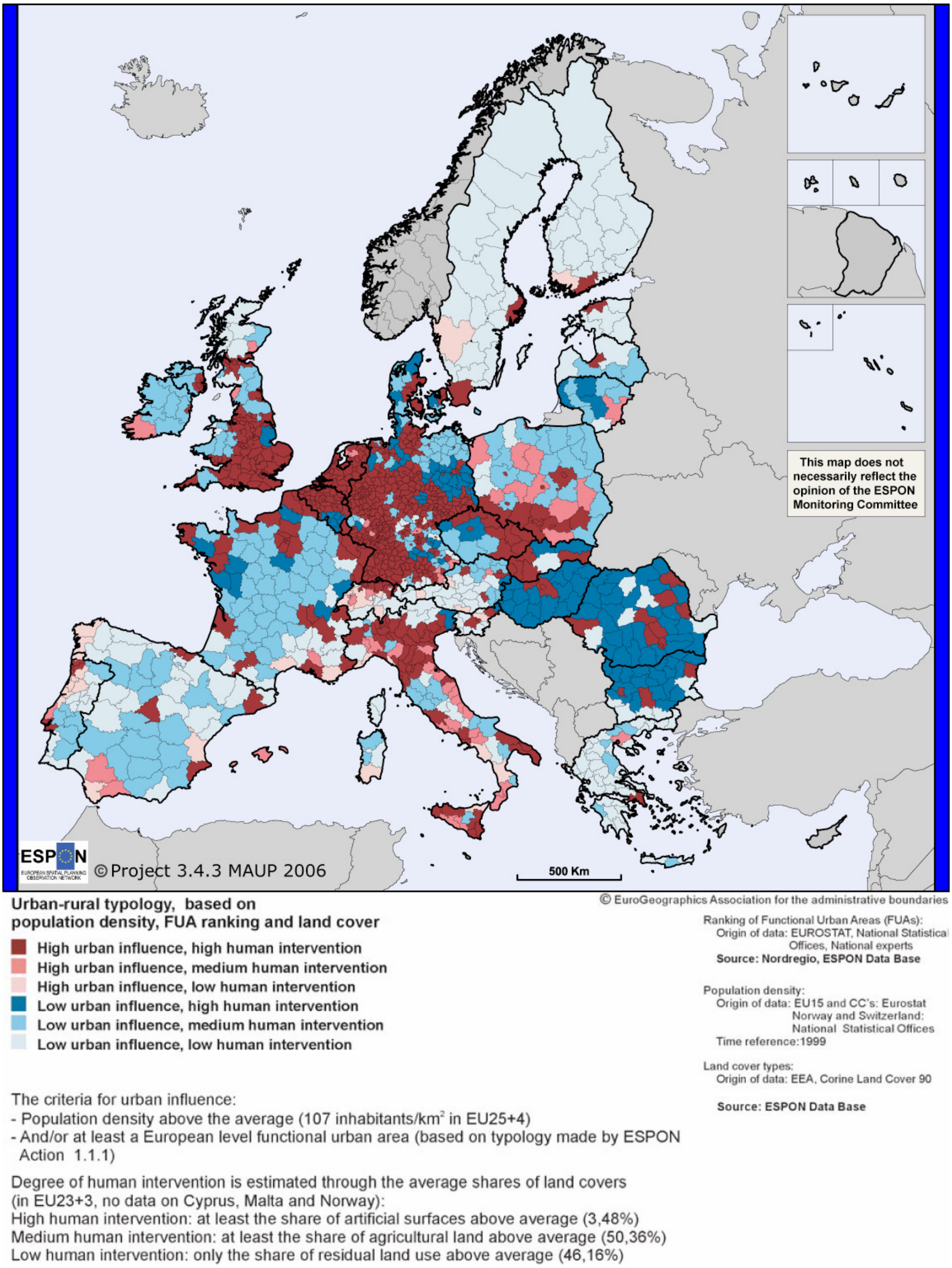


Figure 13 Urban-rural typology at NUTS 3 level

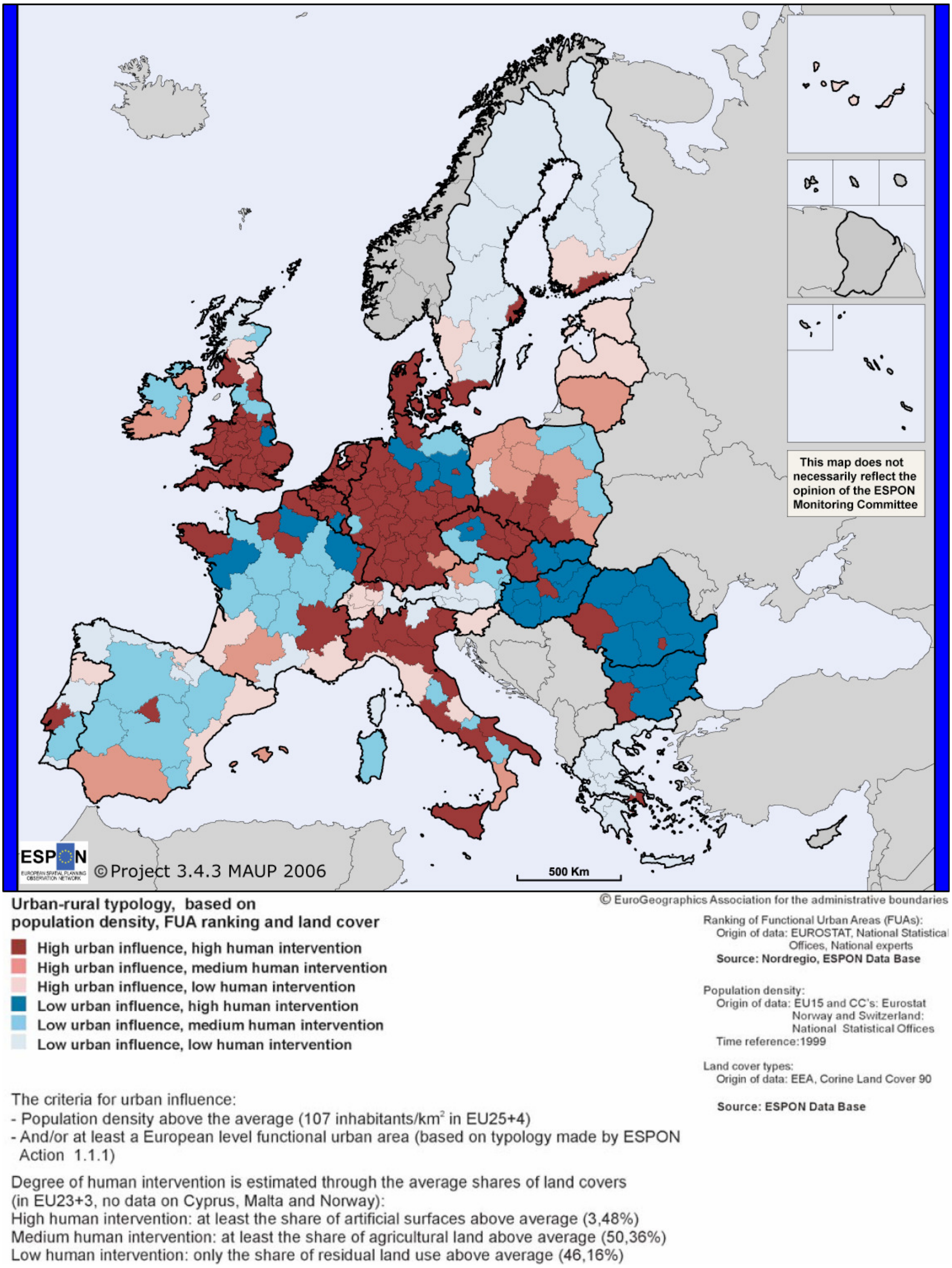


Figure 14 Urban-rural typology at NUTS 2 level

A typical and very frequent change can occur when one or several urban areas, characterised by a high population density and often by a high human intervention and defined individually at NUTS 3 level, are embedded in larger zones. For example, the south eastern part of Germany (eastern Bayern) shows a high variation in the urban-rural typology at NUTS 3 level while this is not the case at NUTS 2 level. These differences are due to an aggregation of the urban area at the NUTS 2 level while they remain individualised at NUTS 3 level. The effect of integration of isolated urban area (NUTS 3) in wider zones at NUTS 2 level is also clearly visible in Poland which in the cartographic rendering appears then as a continuous zone with a high urban influence at NUTS 2.

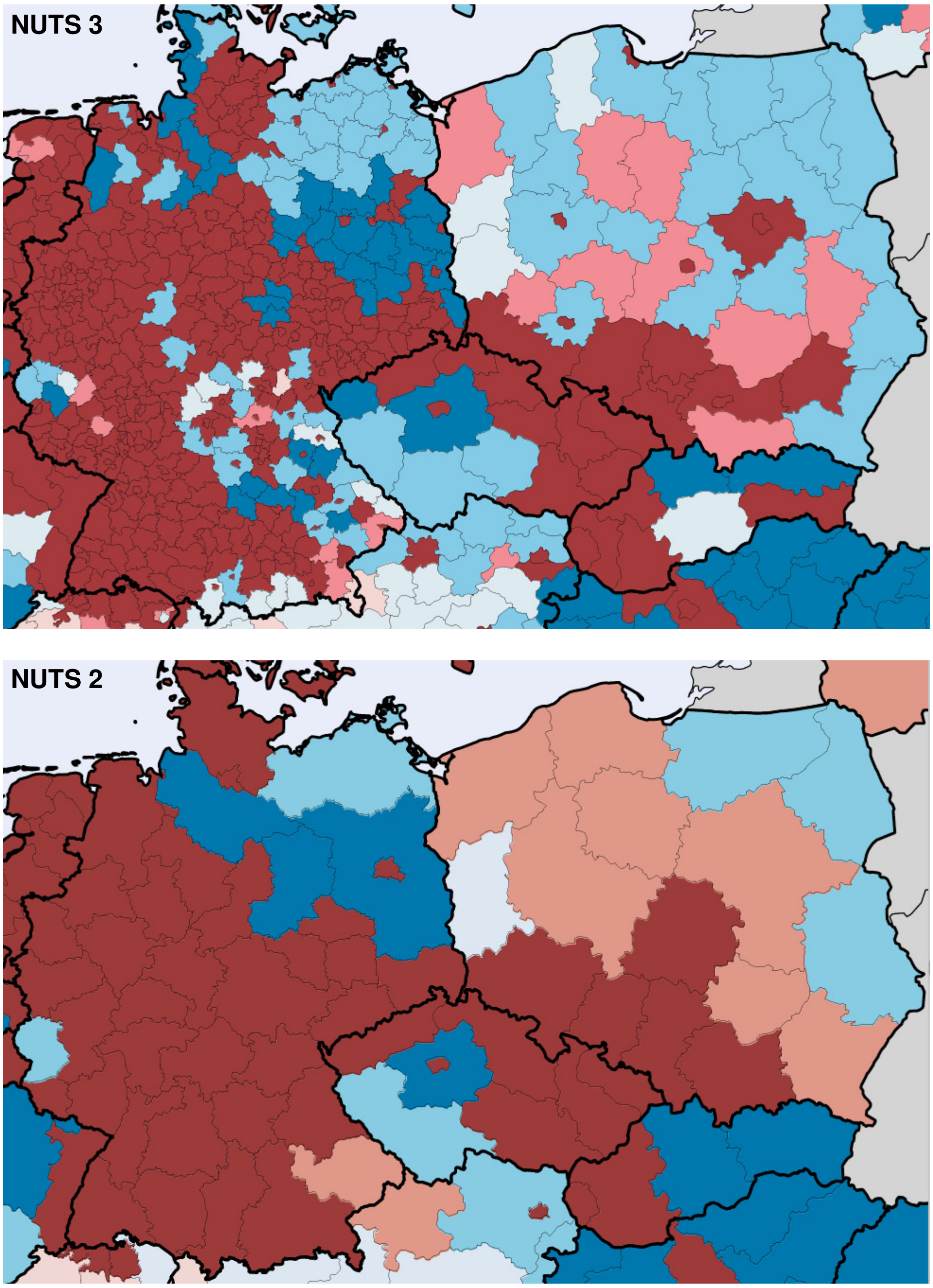


Figure 15 Zoom of the urban-rural typologies on Germany, Poland, Austria and Hungary

The criteria for urban influence are based on density population but also on the existence of a European level urban area[13]. Thus important differences between the typology at NUTS 2 and NUTS 3 levels can be expected. For example, in the south-western part of France (**Figure 16**), only the Gironde and Haute Garonne NUTS 3 are characterised by a high urban influence (and a high human intervention) due to the existence of a European level urban area (respectively Bordeaux and Toulouse) whereas neighbouring units are characterised by a low urban influence. In Ireland (**Figure 17**), the integration of the Dublin NUTS 3 unit, which is a European level urban area, in a wider zone at NUTS 2 changes the cartographic rendering and induces a division of the country into 2 zones according to the urban influence: a NUTS 2 zone with a high urban influence and another with a low urban influence (both being characterized by a medium human intervention). We could also quote the examples of Lituania, Estonia, Latvija or Norhern Irland which by switching from NUTS 3 to NUTS 2 show a much simpler typology due to the integration of European level urban area.

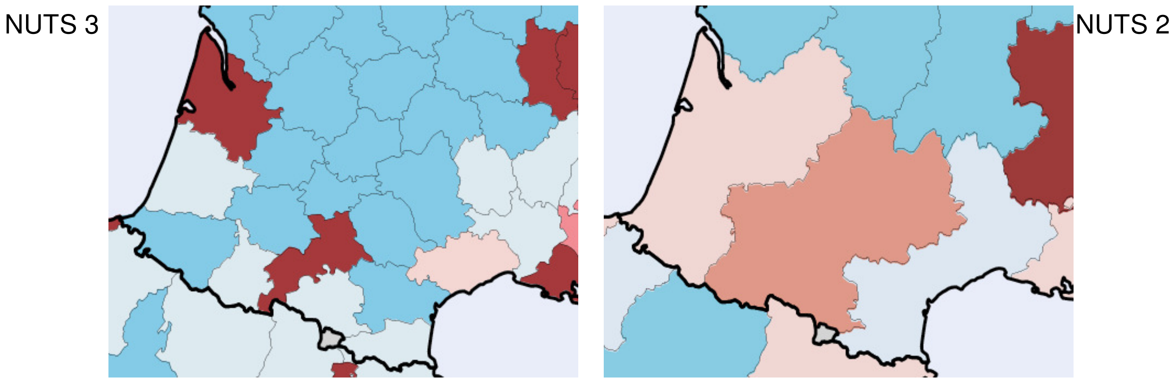


Figure 16 Zoom of the urban-rural typologies on the south-western part of France

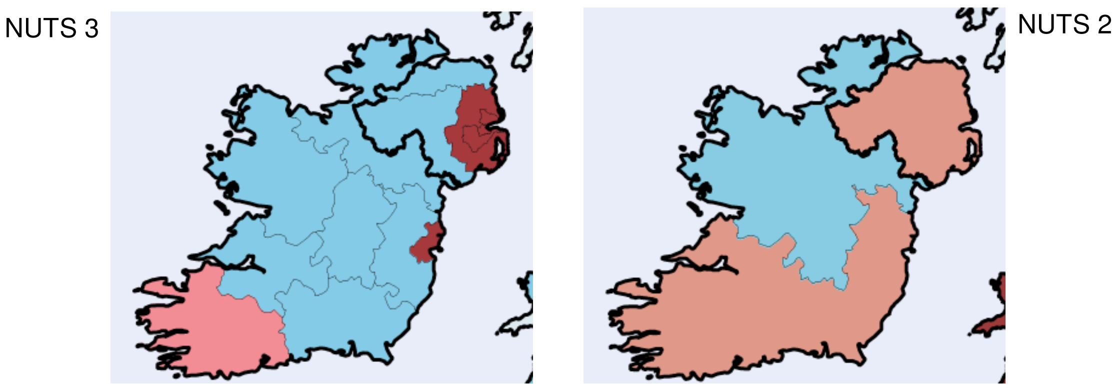


Figure 17 Zoom of the urban-rural typologies on Ireland

13 based on typology made by ESPON Action 1.1.1

Beyond these differences on the cartographic rendering that has been already discussed, we have also evidence of the influence of territorial divisions on the spatial repartitions and the typologies often used in the ESPON project. After the discussion of this problem in ESPON meetings, the authors of project 1.1.2 decided to deepen this point and proposed a typology at a finer level for Belgium and Austria (based on NUTS 5). They demonstrated that the results are different from the one obtained at NUTS3 (cf. annex 2) but considered, as we have done in the first part, that these contradictory results are not a problem but rather a progress in the analysis. They insisted on the flexibility of the method and on the comparisons it allows between different scales.

Beside their synthetic interest, in particular at the cartographic level, the typologies are often used in the ESPON projects, in a second time, to compare the obtained classes to classical socio-economic indicators (GDP, population change and so on). Thus, in their Final Report, the authors show that 4/5 of the wealth (from GDP) concerned a bit more than 1/4 of Europe where the urban influence is high. These results were based on the typology built at NUTS 3. How are these conclusions modified when the level is changed from NUTS 3 to NUTS 2? More generally, since we have seen that the spatial distributions were not exactly the same, how are the statistical comparisons with the socio-economic indicators affected?

Table 1[14] shows several indicators, in absolute and relative values, associated with NUTS 2 and NUTS 3 according to the typology:

- number of territorial units
- area
- population
- GDP

The proportion of units and overall their area vary significantly from NUTS 2 level to NUTS 3 level. For example, for the low human intervention (types 4, 5 and 6), the area of the units with high urban influence (type 6) decreases from 30.7% at the NUTS 3 level to 20.8% at the NUTS 2 level. The units with high human intervention gather 1/4 of the total area (EU25+4) at NUTS 3 and 2/5 at NUTS 2. This can be partly accounted for by the criteria for urban influence which are affected by the presence of a European level functional urban area, as illustrated by several examples previously.

[14] See *Table 3.6. Urban–rural typology in relation to core indicators*, in the Final Report 1.1.2 (fr-1.1.2_maps_chap3_revised_31-03-05.pdf, p. 60)

	Nuts 3		*Nuts 2*	
NUTSn regions				
	Number	*as % from EU25+4*	*Number*	*as % from EU25+4*
Type 1	691	52,0%	130	46,4%
2	58	4,4%	16	5,7%
3	34	2,6%	22	7,9%
4	125	9,4%	28	10,0%
5	184	13,8%	30	10,7%
6	201	15,1%	37	13,2%
n.a.	36	2,7%	17	6,1%
Total	1 329	100,0%	280	100,0%
Area in km²				
	Total	*as % from EU25+4*	*Total*	*as % from EU25+4*
Type 1	906 881	19,3%	1 070 885	22,8%
2	219 444	4,7%	445 664	9,5%
3	127 231	2,7%	460 217	9,8%
4	508 597	10,8%	566 394	12,1%
5	1 049 959	22,4%	740 891	15,8%
6	1 440 576	30,7%	975 898	20,8%
n.a.	442 103	9,4%	434 842	9,3%
Total	4 694 790	100,0%	4 694 790	100,0%
Population in 1999				
	Total	*as % from EU25+4*	*Total*	*as % from EU25+4*
Type 1	299 173	60,4%	284 426	57,5%
2	27 192	5,5%	41 154	8,3%
3	17 744	3,6%	47 059	9,5%
4	40 976	8,3%	49 214	9,9%
5	57 913	11,7%	36 042	7,3%
6	42 021	8,5%	28 843	5,8%
n.a.	9 925	2,0%	8 206	1,7%
Total	494 945	100,0%	494 946	100,0%
GDPpps in 1999				
	Total	*as % from EU25+4*	*Total*	*as % from EU25+4*
Type 1	6 453 790	70,2%	6 136 414	66,8%
2	370 691	4,0%	525 665	5,7%
3	312 803	3,4%	876 917	9,5%
4	391 755	4,3%	450 045	4,9%
5	818 486	8,9%	564 942	6,1%
6	661 896	7,2%	484 163	5,3%
n.a.	182 581	2,0%	153 857	1,7%
Total	9 192 003	100,0%	9 192 003	100,0%

Table 1 Urban–rural typology in relation to several indicators

However, the population and GDP variables are much less sensitive to the NUTS level switching. Especially, **figure 18** shows that the proportion of GDPpps per capita in 1999 remains almost the same at NUTS 3 and NUTS 2 levels.

This experiment indicates that ESPON results are generally not completely modified when changing scale from NUTS 3 to NUTS 2 and vice versa. But they prove that one should be very cautious and always mention the NUTS level which has been used for the elaboration of statistical results.

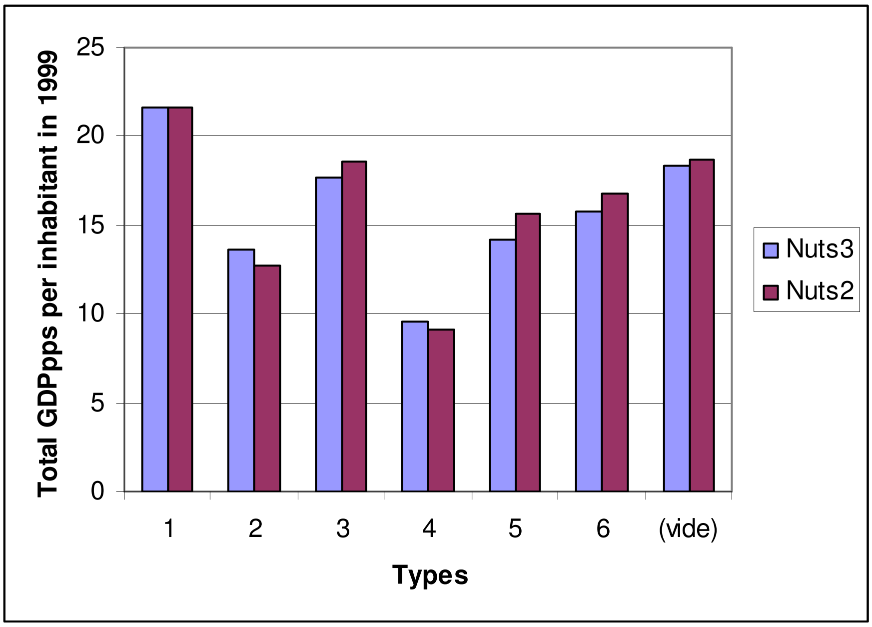


Figure 18   GDPpps per inhabitant according to the urban–rural typology

3 MAUP and the use of multivariate analysis

To go deeper into the MAUP question, we have completed the previous analysis by a multivariate analysis. The aim was to see how the territorial divisions can or cannot affect the results of a multiple linear regression. To do so, we used the classical ESPON dataset:

- GDP pps/Inh. (2000)
- density of population (1999; in log[15])
- unemployment rate (2000)
- and the variation population between 1995 and 2000.

More precisely, we tried to account for the population variation by the other criteria. Are the NUTS which register population decreases characterised by a higher GDPpps per inhabitant or a lower unemployment rate and so on? Are the phenomenon or the phenomena which account for the population variation the same according to the NUTS level considered?

Table 2 shows for NUTS 0, 2 and 3 levels the variables that play a significant role in the change of the population variation, their position in the multiple regression, the sign of the relationship (positive or negative) and the r-squared that defines the quality of the final models. Moreover, we performed multiple regressions at NUTS 3 level for each country (NUTS 0).

If we consider the results for the whole Europe, differences are observed at NUTS 0, 2 and 3 levels and especially in the quality of the models obtained. At NUTS 0 level, only the GDPpps/Inh. variable accounts for 58% of the change in the population variation: the higher the GDPpps/Inh. of a given country, the higher the increase of its population between 1995 and 2000. However, the two other models at NUTS 2 and 3 levels are of rather poor quality (respectively 8 and 12%).

Whereas at NUTS 3 level, on the whole Europe, the population variation between 1995 and 2000 is accounted for at only 12% by the unemployment rate, then by the population density (independently of the previous variable), the built models are very different according to the countries. This demonstrates the underlying high variability of the phenomenon in Europe. Firstly, the models are often better, even if for some countries no model can be built: the quality vary from 14% for France to 72% for Sweden.

[15] To avoid dealing with a strongly asymmetric distribution, the logarithm of population density data was chosen in order to improve the models

	GDPpps/Inh.	Log Density	Unemployment Rate	R-squared	Number
NUTS 0	+			58%	25
NUTS 2	+ (1)		- (2)	8%	257
NUTS 3		- (2)	- (1)	12%	1191
Belgium	+ (2)	- (1)		34%	43
Bulgaria		+		41%	28
Czech Rep.	-			32%	11
Germany		- (2)	- (1)	34%	438
Denmark		+ (1)	- (2)	51%	15
Spain		+		17%	50
Finland		+ (2)	- (1)	63%	20
France			+	14%	100
Italy	+			29%	103
Lt	-			30%	10
Latvia	-			63%	5
Poland		- (1)	- (2)	39%	42
Sweden		+ (1)	- (2)	72%	21
Slovenia	+			36%	12
United Kingdom	+ (2)	+ (3)	- (1)	35%	133

Table 2 Results of multiple linear regression analysis

Moreover, the built models show differences in the selected variables, in their rank but also in the sign of their contribution. For example, an increase of the population in a NUTS 3 can be accounted for firstly by a low density and then by a high GDP (independently of the density) for Belgium, whereas by a high density and by a low unemployment rate for Sweden. Furthermore, the NUTS 3 with an increased population are characterised by a high population density for Bulgaria, and by a low GDP for Latvia.

We could define in this manner different examples. The aim was here to show the differences observed in the multiple regressions according to the selected territorial division. This interesting question will be deeper explored in the final report.

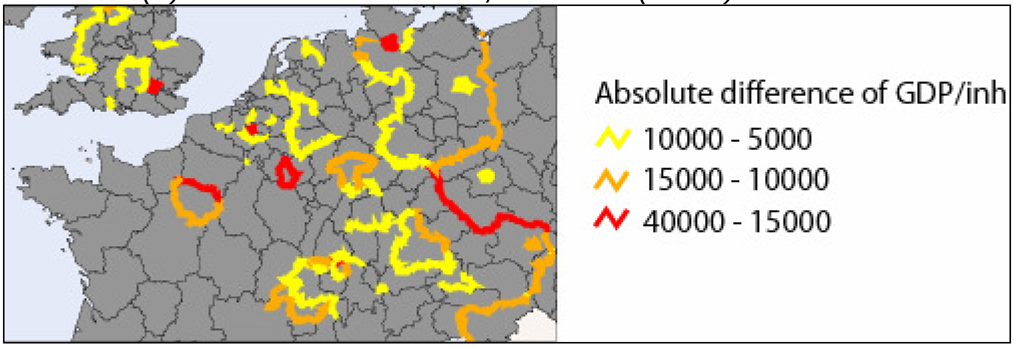
4 MAUP and the spatial analysis approaches

Last but not least, the MAUP has direct consequence on the results of advanced tools of spatial analysis and can completely change the cartographic and statistic outputs. A good example of this has been provided in the final report of ESPON project 3.1 in the part "dictionary of tools" where it was clearly demonstrated that all methods which are based on spatial contiguity (maps of discontinuities, local deviations, spatial autocorrelation, measure of spillover effects, …) are completely dependant from the choice of basic territorial units.

The **Figure 19** presents for example two maps of local economic heterogeneity (differences of GDP/inh. between contiguous regions) established at NUTS 2 and NUTS 3 levels.

- *At NUTS 2 level* (**19.a**), the attention of the reader is attracted (1) by spatial structures of upper level which are mainly related to international borders: the exceptional advantage of Luxembourg, the former political division of Europe, the difference with non-EU member countries like Switzerland and (2) by major metropolitan area, which are surrounded by circular discontinuities (Paris, München, Frankfurt, London, …).
- *At NUTS 3 level* (**19.b**) the picture is much more confused because all urban units located in Germany, Austria and Benelux are now able to produce differences with the surrounding peri-urban or rural areas. What the reader really perceives is local scale differences which are apparently present in certain states and not in other ones. But this perception is in fact false and equivalent urban-rural oppositions of GDP/inh would be observed in France, Spain or Italy if those states have used smaller territorial divisions at NUTS 3 levels.

(a) Discontinuities of GDP/inh. 1999 (euros) at NUTS 2



(b) Discontinuities of GDP/inh. 1999 (euros) at NUTS 3

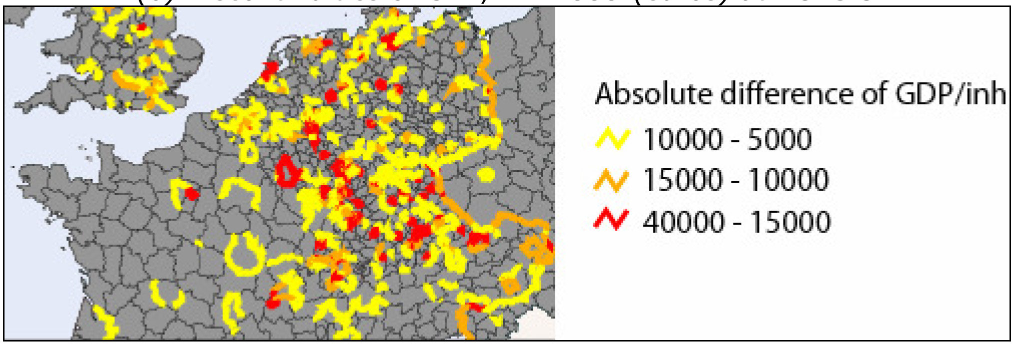


Figure 19 Impact of the choice of spatial units on the cartography of territorial discontinuities

(Source: ESPON 3.1, Final Report, Dictionary of tools)

=> The problem is not to decide which scale of analysis is the better one (both are interesting) but to avoid a *confusion of scale* on the same map.

Conclusion: "problem" or "progress"

The main conclusion of this interim report of project ESPON 3.4.3 is that Modifiable Areal Unit should not be considered as a **problem** for ESPON but rather as a challenge which could be the starting point of **progress** for territorial planning and political decisions. The project ESPON 3.4.3 could be better entitled Modifiable Areal Unit *Progress* if it succeeds in the development of new proposals which insure a better managing of various levels of territorial aggregation in future ESPON II (2007-2013).

In the next months of the project, we will try to explore 4 main directions of research which do not intend to "solve" the questions that are related to changing territorial divisions but rather to "improve" the use of available information which has been yet collected by ESPON project and to help policymakers to gain better efficiency in the exploitation of results. The objective is to give some expertises for decision aid for the choice of territorial zoning as well as the use of tools for interpretation aid.

Combination of Nuts

The development of new levels of aggregation combining NUTS units of different levels (with minimum building blocks at NUTS 3 level) does not intend to replace actual use of NUTS 2 or NUTS 3 units in the ESPON projects but rather to complete them. Indeed, the actual official divisions are necessarily of interest for policymakers as they are the basis of regional policy in Europe and ESPON results would be of limited for concrete action if they were not, at least partly, based on these official divisions. But, and it is the crucial point, the fact to maintain the production of maps and database at NUTS 2 and NUTS 3 levels does not means that ESPON should limit the production of results to this territorial breakdowns. If most relevant results can be achieved with other divisions (like mixture of NUTS units based on equal population, equal wealth or functional regions) it should be done in the same time and, why not, this new territorial divisions could be used as basis for a renewal of official territorial divisions in the future if they appears to be more relevant from political point of view.

Multiscalar representations

The classical objection to the production of new territorial divisions is that it would produce a multiplication of maps and tables which is expensive and make difficult the synthesis of results. But this objection can be rejected if ESPON is able to produce innovative cartographic tools which make possible the use easily multiple visualisation of the phenomena of interest. Dynamics maps where it is possible to jump from one statistical or territorial division to another one are precisely the solutions that have been developed in ESPON 3.1 with the Hyper-atlas and the WEB-GIS. Technical innovations can be considered as potential vector of political innovations if they make easier the evaluation of complex situations which can not be analysed at one single level. But what is striking is to examine if this complexity is compatible or not with actual practices. Therefore, we propose to evaluate the way that policymakers perceive and use maps in their usual practice, and to evaluate how they could adopt more complex tools for decision.

Innovative cartography for political decisions

Moreover, the question is not only to make better use of classical cartographical representations of spatial phenomena, but also to propose the implementation of new methods of representation of spatial phenomena like smoothing methods or cartograms (e.g. anamorphoses proportional to population or GDP). These innovative cartographic methods introduce a real change in the work of spatial planners as they are not only an alternative representation of the target phenomena but innovative tools which try to explore new dimensions of research which opens the door for new policies at European scale. As a very simple example, the realisation of a smoothed map of unemployment rate in a neighbourhood based on time distance (like 2 hours) is much more than a technical transformation. It is indeed the basis for the realisation of a political objective of territorial cohesion which could be the allocation of jobs for regions which are declining in order to avoid the out-migration of their inhabitants. The parameter used in the computation of the smoothing (2 hours) is typically a political parameter where many alternative dimensions could be discussed, like the fact to take into account the jobs available in neighbouring countries or not . It suggest that an improvement of labour market in a given region is not necessarily related to an increase of workplace inside this region but could also be based on changes in spatial accessibility (transport policy) or to a wider opening of the borders (market policy).

Innovative statistical methods for territorial impact assessments

The same is true when considering the statistical dimension of the problem. The study of correlation between indexes is the basis of Territorial Impact Analysis developed in ESPON projects under priority 2 and many important conclusions of ESPON has been derived from measure of correlation established for the whole ESPON area (in particular for CAP). But the study that we have developed in project 3.4.3 has clearly demonstrated that correlations are not necessary uniform according to scale level and can be different when we consider sub-parts of the European territory. It suggests strongly that the use of a mean value for all ESPON space is not necessary relevant and that better results would be obtained if we proceed to an analysis of correlation at different scale and in different regional contexts. With innovative statistical tools like Geographically Weighted Regression (GWR) developed by Fotheringham & Charlton, we can imagine to proceed to more in depth analysis of Territorial Impact, making possible to produce different recommendations for different part of the European territory. One more time, MAUP should not be considered as a problem but as a resource for improvement of spatial planning at European scale.

References

BBR (ed.), 2005, *ESPON project 3.1 – Integrated tools for European spatial development (2002-2004)*
http://www.espon.lu/online/documentation/projects/cross_thematic/index.html

CURS (ed.), 2005, *ESPON project 1.1.2 - Urban-rural relations in Europe (2002-2004)*
http://www.espon.lu/online/documentation/projects/thematic/thematic_7.html

FOTHERINGHAM A.S., BRUNSDON C.F., CHARLTON M.E., 1999, 'Geographically Weighted Regression: A Natural Evolution of the Expansion Method for Spatial Data Analysis', *Environment and Planning A*, 30, pp. 1905-1927.

FOTHERINGHAM A.S., BRUNSDON C.F., CHARLTON M.E., 2002, *Geographically weighted regression: the analysis of spatially varying relationships*, UK: Wiley and Sons, 269 p.

FRANCOIS J.C. *et al*., 2003, 'Les disparités des revenus des ménages franciliens en 1999: approches intercommunales et infracommunales et évolution des différenciations intercommunales 1990-1999', Paris, Direction Régionale de l'Equipement d'Ile-de-France, 108 p.

GRASLAND C., MATHIAN H., VINCENT J.M., 2000, 'Multiscalar analysis and map generalisation of discrete social phenomena: Statistical problems and political consequences', *Statistical Journal of the United Nations*, ECE, 17, IOS Press, pp. 1-32.

GUEROIS M. and PAULUS F., 2002, 'Commune-centre, agglomération, aire urbaine : quelle pertinence pour l'étude des villes?' , *Cybergeo : Revue Européenne de Géographie*, n°212 ; 18 p.
http://193.55.107.45/essoct/guerois/guerois.htm

HOLM E., HOLME K., MAKILA K., MATTSSON-KAUPPI M. and MORTVIK G., 2002, *The SVERIGE spatial microsimulation model. Content, validation, and example applications*, Sweden: GERUM, 4.

ITPS (ed.), 2005, *ESPON project 1.1.2 – The spatial effects of demographic trends and migration*
http://www.espon.lu/online/documentation/projects/thematic/thematic_75.html

OPENSHAW S., 1981, 'Le problème de l'agrégation spatiale en géographie', *L'Espace Géographique*, n°1, pp. 15-24

OPENSHAW S., 1984, 'The modifiable areal unit problem', *Concepts and Techniques in Modern Geography*, 38, 41.

OPENSHAW S. and RAO L., 1995, ‘Algorithms for re-aggregating 1991 Census geography’, *Environment and Planning A*, 27, pp. 425-446.

RASE W.-D., 2001, ‘Kartographische Anamorphosen und andere nichtlineare Darstellungen’, *Kartographische Bausteine*, Band 19, TU Dresden, Institut für Kartographie (See Annex 3)

RAVENEL L., 2003, ‘Dossier pédagogique : la présence d’étrangers entraîne-t-elle le vote pour l’extrême droite’, *Espace, populations, sociétés*, n°3, pp. 541-547.

ROBINSON W.S., 1950, ‘Ecological correlations and the behavior of individuals’, *American Sociological Review*, Vol. 15, pp. 351-357.

Case Studies

Case study 1: Case study on Sweden

Case study 2: Case study on Northern Ireland

ESPON 3.4.3

"THE MODIFIABLE AREAS UNIT PROBLEM"

Comparisons of Region and Grid Maps with Various Resolutions in Sweden

Submitted by:
Dr. Magnus Strömgren
Dr. Kirsten Holme
Professor Einar Holm

SMC, Kiruna
Department of Social and Economic Geography
Umeå University, Umeå
Sweden

Introduction

An administrative border will, in certain cases, act as a functional border. Taxation levels, local regulations, and other politically decided parameters often give rise to geographical differences that coincide with the boundaries of the administrative units. On the other hand, when summarizing other phenomena by administrative regions, artificial dissimilarities that are due to aggregation may appear. Discontinuities may exist in the area, but not where they are found on the map. The most tangible discontinuities may arise between coast/inland, urban regions/rural areas, or city centers/suburbs and not where the administrative borders happen to be situated. Although grid maps do not solve this problem, they are perhaps likely to provide a higher degree of homogeneity within each unit since the surfaces are of equal size. Still, aggregation of data into squares is a rigid method that doesn't take into consideration the distribution of the underlying process or phenomenon.

No matter whether data are shown as grids or summarized by administrative regions, it is difficult to compare large sparsely populated areas with small densely populated ones. Also use of uniform geographical areas will inevitably in most cases generate undesirable variation in the size of the population. In addition, the level of aggregation has a vast influence on the appearance of an indicator. A high-resolution map with a minimum of aggregation will often be fuzzy and hard to interpret. Increased aggregation levels out variation**.** Conspicuous outliers will disappear within the increasing amount of data around the total mean. All in all, it is complicated to determine how well a final map mirrors the target phenomenon.

Northern Sweden is one of the most sparsely populated areas in Europe. Consequently, the sizes of NUTS 2 and NUTS 3 territorial units are very large. This produces a problem of unit heterogeneity compared not only to the rest of Europe but also within Sweden. The aim of this study is to investigate the modifiable areal unit problem in the Swedish context. Specifically, the aim is to explore the impact of spatial scale as well as grid contra administrative region effects on map representation and interpretation. Beyond the problem of cheating with maps lurks an eventual bias in results of a causal analysis. Similar analyses might end up with different (biased) results and answers only because of the choice of high or low resolution in used spatial data.

The study is carried out by examining and comparing the visual appearance of three indicators in different map types with varying levels of aggregation. The indicators are 1) population density (2002), 2) population change 1990–2002, and 3) mean disposable income (2002). The distribution of the indicators – as well as the number of people – in the map classes are also summarized separately. In addition, descriptive statistics of the underlying distribution of values are presented. The variable disposable income is treated a bit more in-depth. First, a close-up look at the data beyond what is visible on the maps is presented. Also, an attempt is made to provide an estimation of the error that is introduced when full information is not available, by means of a simple experiment based on OLS regression.

Spatial Modelling Centre (SMC) at Dept. of Social and Economic Geography, Umeå University is in possession of a database – ASTRID – that contains Swedish individual register data for the years 1985–2002. The database can be used for specific research purposes only. It contains constant personal data such as birth year, gender, and country of origin as well as time-dependent data such as civil status, education level, and incomes. In addition, the database contains the geographical coordinates of all individuals' place-of-residence with a resolution of 100 meters. Thus, the database provides a good platform for trying out different mapping solutions (grids, administrative territories, smoothing methods, etc.) of the selected indicators.

In the study, the chosen indicators are presented in grid maps and territorial maps – each at four different levels of aggregation. The *grid maps* consist of squares covering the whole country. However, for obvious reasons squares intersecting the national border will not have the predetermined area. This is most evident at the lowest resolution (100 km): only around one third of the squares are even close to the stipulated area of 10,000 km^2. Uninhabited squares, which are most common at the highest resolution (10 km), are not displayed. The grid maps have the following resolutions:

- 100 km squares (N=68 or 69).
- 50 km squares (N=226 or 228).
- 30 km squares (N=562 or 565).
- 10 km squares (N=3,876, 3,954, or 4,040).

The *territorial maps* are based on existing administrative or statistical units. The highest resolution is found in the maps based on the 284 municipalities that existed in Sweden in 1990.[1] The next level is one commonly used definition of local labor market regions – aggregations of municipalities into functional

1 Subsequently, six additional municipalities have been created from parts of existing ones. However, since one variable involves data for the years 1990 and 2002, the lowest resolution – i.e. the smallest number of municipalities – was used in the study.

regions based on commuting patterns. The territorial maps that will be shown are:

- National areas, i.e. NUTS 2 (N=8).
- Counties, i.e. NUTS 3 (N=21).
- Local labor market regions (N=81).
- Municipalities, i.e. LAU level 2 (N=284).

Since the 10 km square maps have the highest resolution (and therefore the most extreme minimum and maximum values), the distribution of the variables at this level formed the basis for the classification of all maps. The attributes of the 10 km squares were divided into six groups with an approximately equal number of squares in each. The same class breaks were then been applied to the other seven maps of the concerned phenomenon. As will be revealed later this might be one reason for the emergence of different interpretation problems.

Population Density

In Figure 1 and 2, population density (inhabitants/ km^2) for the year 2002 is displayed in territorial and grid maps, respectively.[2] The territorial map with the highest resolution – i.e. the Swedish municipalities – exhibits higher population densities along the coasts and in southern Sweden compared to the inland of northern Sweden. A similar pattern can be recognized for the labor market regions, but at this level there is a contiguous densely populated area between Stockholm and Gothenburg. At the municipality level, this high-density area only appears as a number of scattered heavily populated municipalities.

This high-density belt remains in almost unchanged form at the NUTS 2 and NUTS 3 levels. With increased aggregation, the heterogeneity of the map decreases substantially. In the NUTS 2 map, there are only three classes. The main cities Stockholm, Gothenburg, and Malmö and their surroundings are densely populated, while northern Sweden exhibits the lower values. The comparatively densly populated coastal areas have incorporated the areas with very low population density in the inland of northern Sweden and the final result is a map where more than half of Sweden has a medium dense population.

A somewhat similar picture is discernable in the grid maps. The 10 km square map is irregular and fuzzy, but some low-density areas are seen in southern Sweden and minor high-density areas can be identified in the north. The higher the aggregation of the squares, the more the characteristics from the thematic maps can be recognized. It should be noted, however, that the lowest density class (<0.19 inhabitants/km^2) – which was not present at all in the territorial maps – is represented in all grid maps. Comparing the 50 km square map (N= 226) and the municipality map (N=284), a much more diversified picture is apparent in the grid map. The area corresponding to the low-density municipalities in the inland of northern Sweden is represented as a mosaic of the three lowest density classes.

2 Water surfaces are not included in the area used to calculate the indicator. The actual land area for each municipality has been collected from Statistics Sweden (<http://www.scb.se/statistik/MI/MI0802/2005A01A/mi0802tab3.xls>). The other territorial groups are all aggregations of municipalities and their land area were acquired by adding the area of the comprising municipalities. The land area of each square was calculated by addition of the comprising 100 m squares from the Corine Land Cover 2000 data [Corine data: ©EEA, Copenhagen, 2005]. A few the 10 km squares have a land area of 1,000 km^2, but none of the larger squares reach this level.

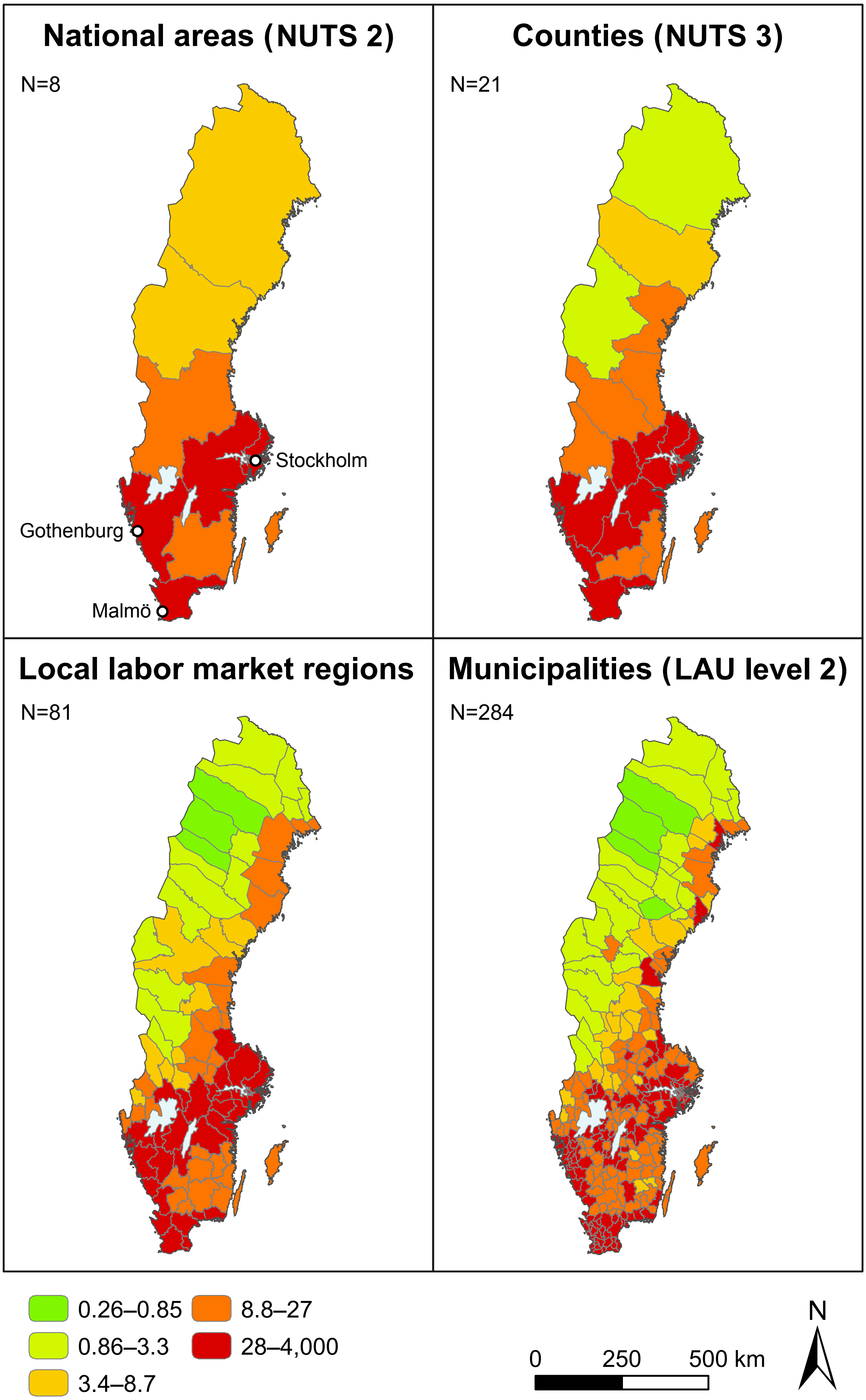


Figure 1 Population density (inhabitants/km² 2002) – regions.

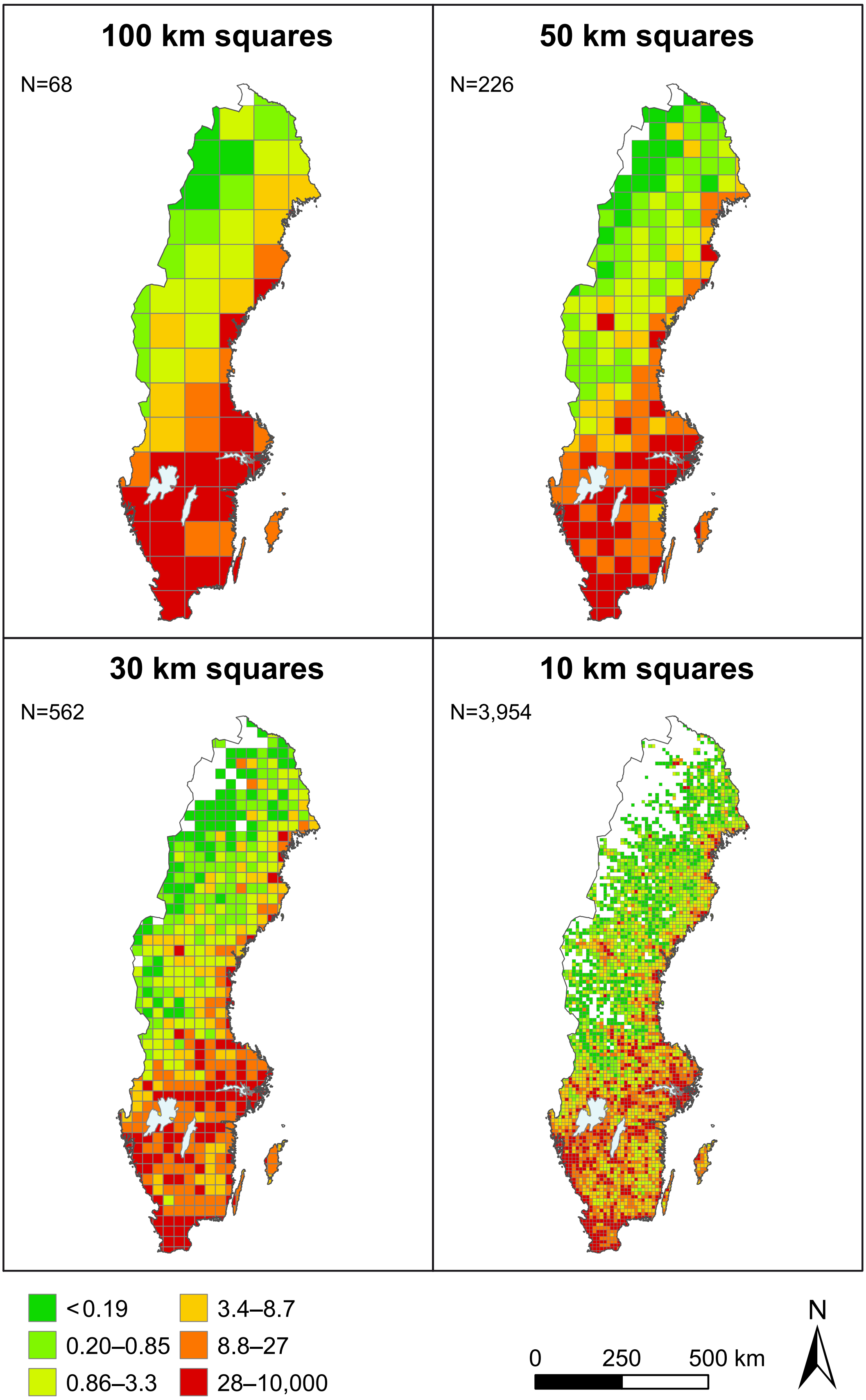


Figure 2 Population density (inhabitants/km² 2002) – squares.

Figure 3 summarizes the proportion of units in the different population density classes. The number of units in the high-density classes increases with aggregation. Moreover, these classes are generally more frequently represented in the territorial maps compared to the grid maps. Since larger urban localities are present in more or less all regions in the territorial maps, but not all of the squares, this is hardly surprising.

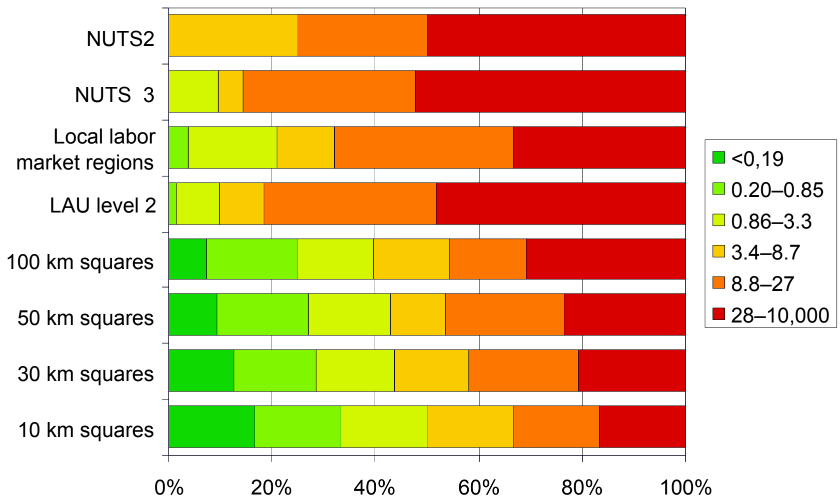


Figure 3 Population density (inhabitants/km² 2002): proportion of units in the different classes. Note that the 10 km square class has an equal amount of squares in each population density class.

Figure 4 displays the percentage of the population living in different unit classes of population density. The population distribution is more or less the same independently of aggregation level and map type. About 80% of the population lives in the most densely populated class. Even though half of the 10 km squares are found in the three classes with lowest population densities only approximately 1.5% of the population lives there.

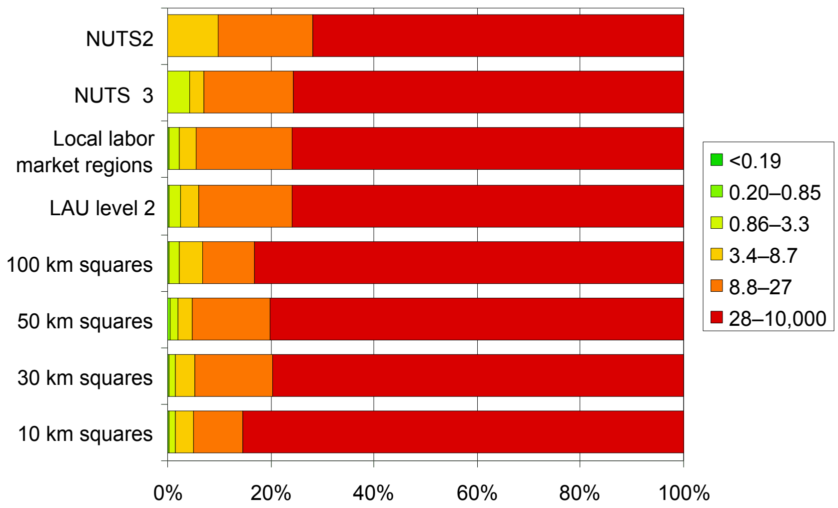


Figure 4 Population density (inhabitants/km² 2002): share of population in the different unit classes.

Table 1 contains descriptive statistics such as mean and standard deviation for the indicator population density. In general, range (maximum value-minimum value) decreases with increased aggregation. For instance, the 10 km squares have by far the highest range. Notably, the municipality level (N=284) exhibits the second highest range and the highest standard deviation, while the smallest range and the lowest standard deviation appear at the local labor market level (N=81). The municipalities also have the highest mean and local labor market regions the lowest. The main explanation for this is that the local labor market subdivision represents an attempt to depict functional regions; many densely populated municipalities with a net influx of commuters are therefore aggregated with sparsely populated municipalities in the vicinity. The distributions are consistently skewed to the right, i.e. there are comparatively few regions and squares with very high-density values.

Table 1 Descriptive statistics, population density (inhabitants/km^2) (2002).

	N	Min.	Max.	Range	Mean	Median	Std. dev.	Skew-ness	Kurtosis
NUTS 2	8	3.3	283.7	280.4	65.2	31.4	93.4	2.3	5.6
NUTS 3	21	2.6	283.7	281.1	42.9	31.2	60.2	3.5	13.9
Local labor market regions	81	0.3	234.8	234.6	28.0	17.1	38.6	3.3	13.2
LAU level 2	284	0.3	4,036.2	4,036.0	126.7	26.3	423.0	6.9	54.6
100 km squares	68	0.01	691.4	691.4	36.7	7.5	95.4	5.4	34.0
50 km squares	226	0.001	801.5	801.5	31.3	7.2	86.2	5.9	41.0
30 km squares	562	0.001	2,666.3	2,666.2	35.2	5.7	149.7	12.3	188.1
10 km squares	3,954	0.01	10,344.3	10,344.3	36.4	3.3	250.0	26.1	901.4

Population Change

Figure 5 and 6 show population change in Sweden between 1990 and 2002 in territorial and grid maps, respectively. As with population density, the reduction of heterogeneity is considerable when changing the aggregation from the 10 km squares (N=4,040) to the 100 km (N=69) squares. The reduction is even more substantial when the 284 municipalities are

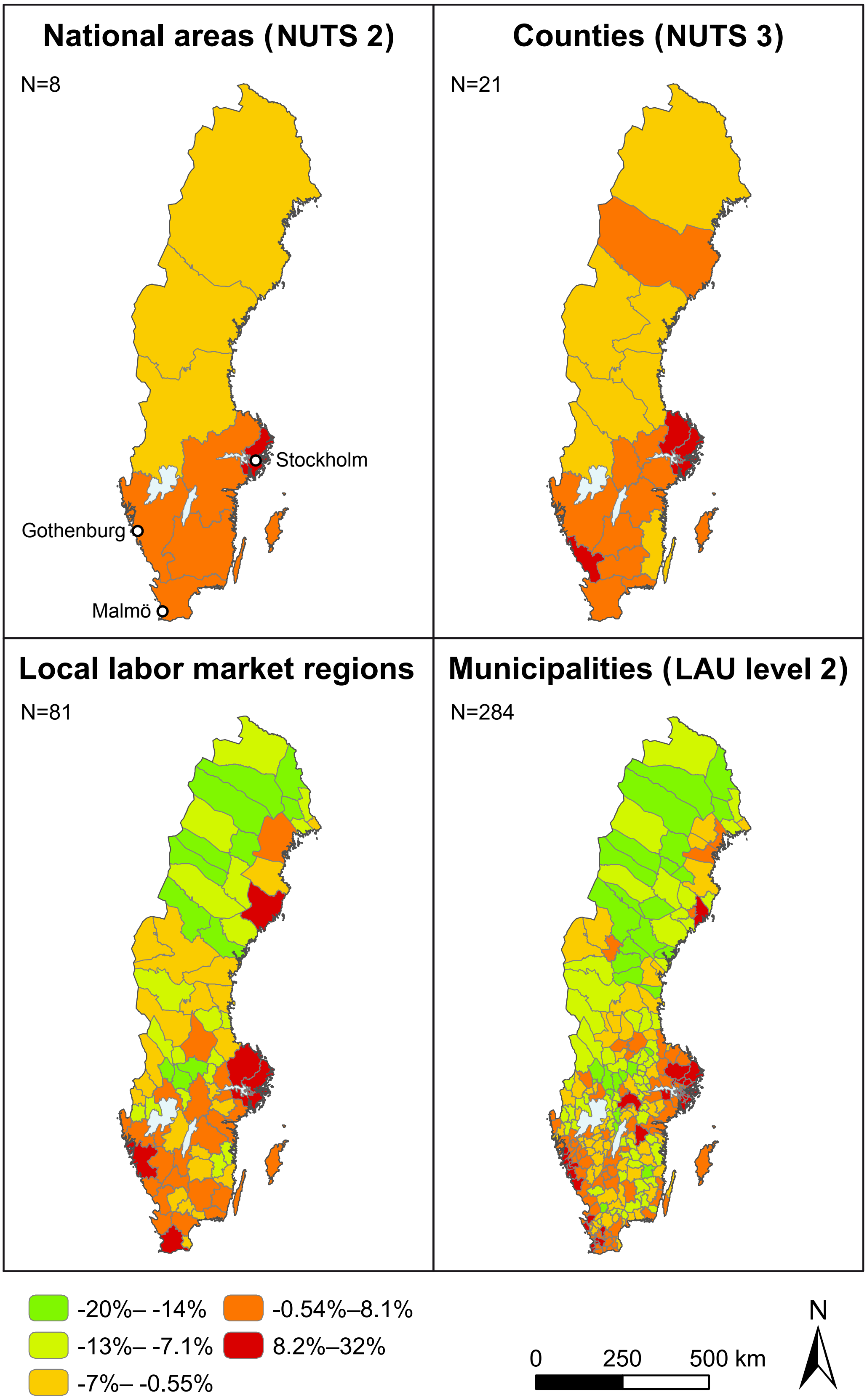


Figure 5 Population change 1990–2002 – regions.[2]

compared to the eight NUTS 2 regions. Looking at the NUTS 2 map, the reader gets the impression that the population redistribution has been rather limited during the period. However, the grid maps – especially the 10 km square map – show that numerous areas have suffered from a reduction of inhabitants, while a few areas have gained and often gained a lot.

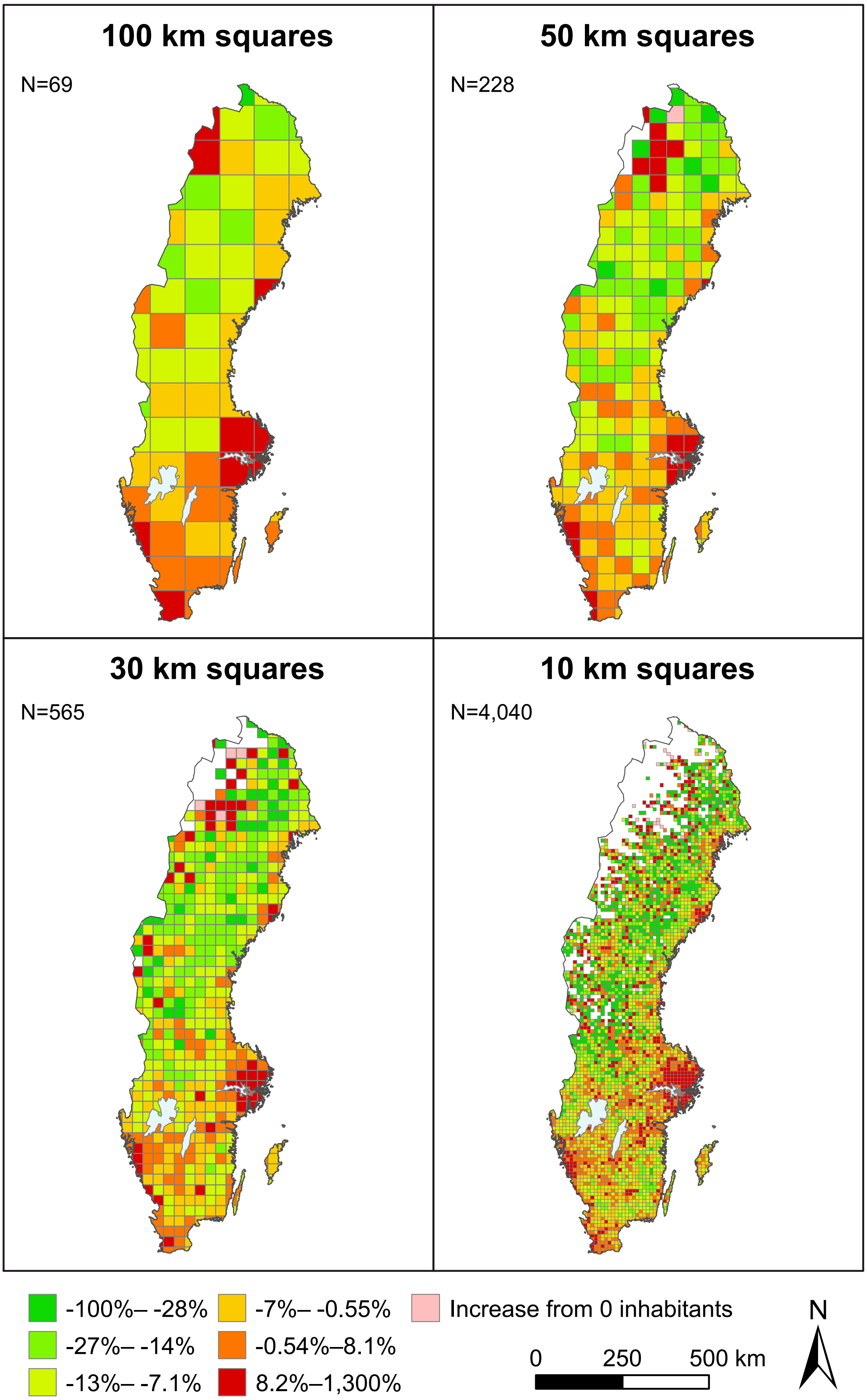


Figure 6 Population change 1990–2002 – squares.[2]

In the sparsely populated inland of northern Sweden, every birth, death, and move has a big influence on the demographic situation. Consequently, all degrees of change, from 100% to more than +8.2%, are found here. A small number of previously uninhabited squares in northern Sweden have been inhabited between 1990 and 2002. However, only a few people live in these places. (In total, there are 181 inhabitants in the 51 newly inhabited 10 km squares.)

Figure 7 summarizes the proportion of units in the different classes of population change. The majority of the NUTS 2 and NUTS 3 regions are found in the two classes with the least change (-7%–8.1%), while many local labor market regions and municipalities exhibit larger population losses. Of course, this means that the most populated – and therefore most influential – parts of these NUTS 2 and NUTS 3 regions have a close to constant population or exhibit a population increase that outweighs population losses in other parts of the territories (cf. Figure 5 and 6).

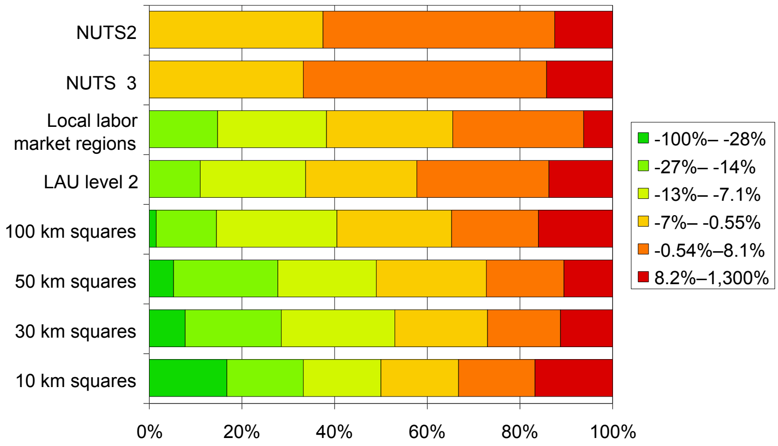


Figure 7 Population change 1990–2002: proportion of units in the different classes.[3]

Figure 8 displays the percentage of the population living in different unit classes of population change. Compared to the population density variable where approximately 80% of the population lived in the areas with high population density, the population is less concentrated to the class with the highest increase of inhabitants (cf. Figure 4). Only around 40% live in these areas. Contemporary place attraction (change 1990 – 2002) is obviously more selective compared to the result of long-term attraction (cross-section 2002). The number of inhabitants that reside in the 10 km squares with the highest population loss is negligible (0.22%) and the squares with the next highest loss also have a tiny population (2.7%).

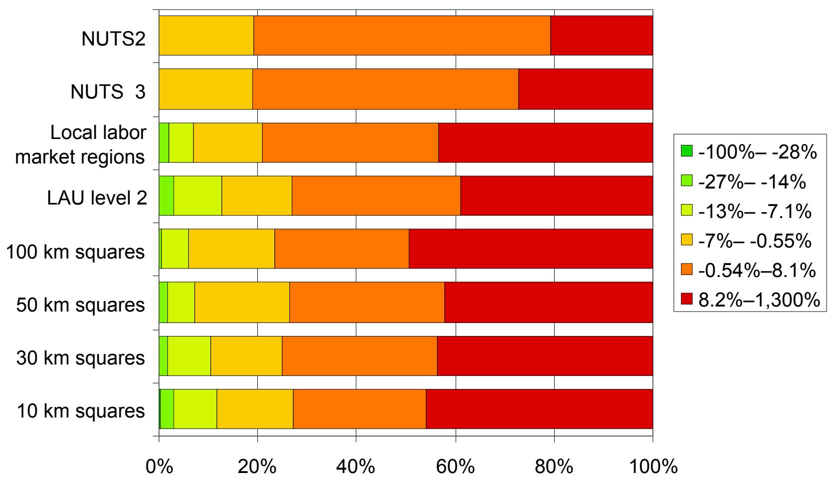


Figure 8 Population change 1990–2002: share of population (2002) in the different unit classes.[3]

3 Excluding squares with an increase from zero inhabitants.

Table 2 contains descriptive statistics for the indicator population change. The mean and median values are negative for all map types and aggregation levels, with the exception of NUTS 2 and NUTS 3. In terms of range, there is a substantial reduction with increased aggregation – even more dramatic than the corresponding change for the population density indicator.

Table 2 Descriptive statistics, population change 1990–2002.[3]

	N	Min.	Max.	Range	Mean	Median	Std. dev.	Skew-ness	Kurtosis
NUTS 2	8	-6.6	11.4	17.9	1.6	1.3	5.9	0.3	-0.5
NUTS 3	21	-6.9	11.4	18.2	0.7	0.4	5.1	0.6	-0.3
Local labor market regions	81	-20.3	11.4	31.7	-4.9	-4.8	7.7	-0.002	-0.7
LAU level 2	284	-20.3	32.5	52.8	-2.3	-2.9	9.2	0.4	0.1
100 km squares	69	-100	26.5	126.5	-5.2	-4.5	15.6	-3.2	19.6
50 km squares	227	-100	400	500	-4.9	-6.7	34.2	8.0	90.8
30 km squares	561	-100	406.3	506.3	-5.1	-7.9	39.2	8.1	81.6
10 km squares	3,989	-100	1,250	1,350	-6.0	-7.1	48.3	8.7	159.1

Disposable Income

Figure 9 and 10 show mean disposable income in Sweden for the year 2002 in territorial and grid maps, respectively. Disposable income is total earnings when taxes are deducted. The mean is calculated for inhabitants between 20 and 64 years of age and displayed in the Swedish currency (SEK). In the municipality map, the wealthiest areas can be found in southern Sweden, particularly in the vicinity of Stockholm. Most inland municipalities in northern Sweden are comparatively poor. However, the class with the lowest mean income is not present in the territorial maps, and even the second poorest income class is rarely represented. Interestingly, in the far

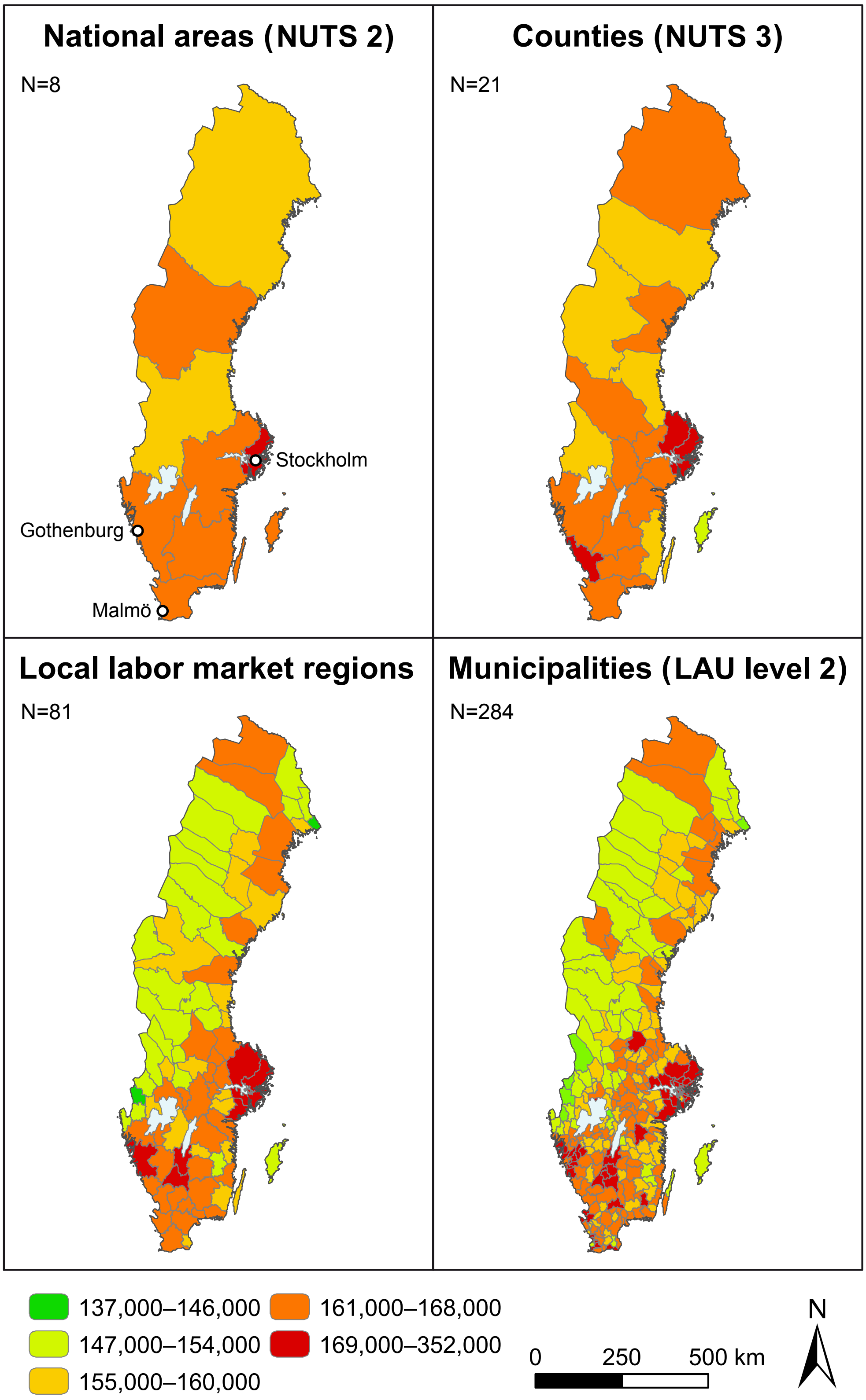


Figure 9 Disposable income (SEK)/inhabitant (20–64 years; 2002) – regions.

north of the country there is a ”wedge” of comparatively wealthy municipalities that stretches from the coastal municipalities towards northwest into the inland. Just like with the other two indicators, heterogeneity decreases with increased aggregation. At the NUTS 2 level, only the three highest income classes are still visible. At this aggregation level, the high-income belt in northern Sweden has disappeared. Due to the impact of the poorer districts in the south, this NUTS 2 region ends up as a territory with quite moderate mean disposable incomes.

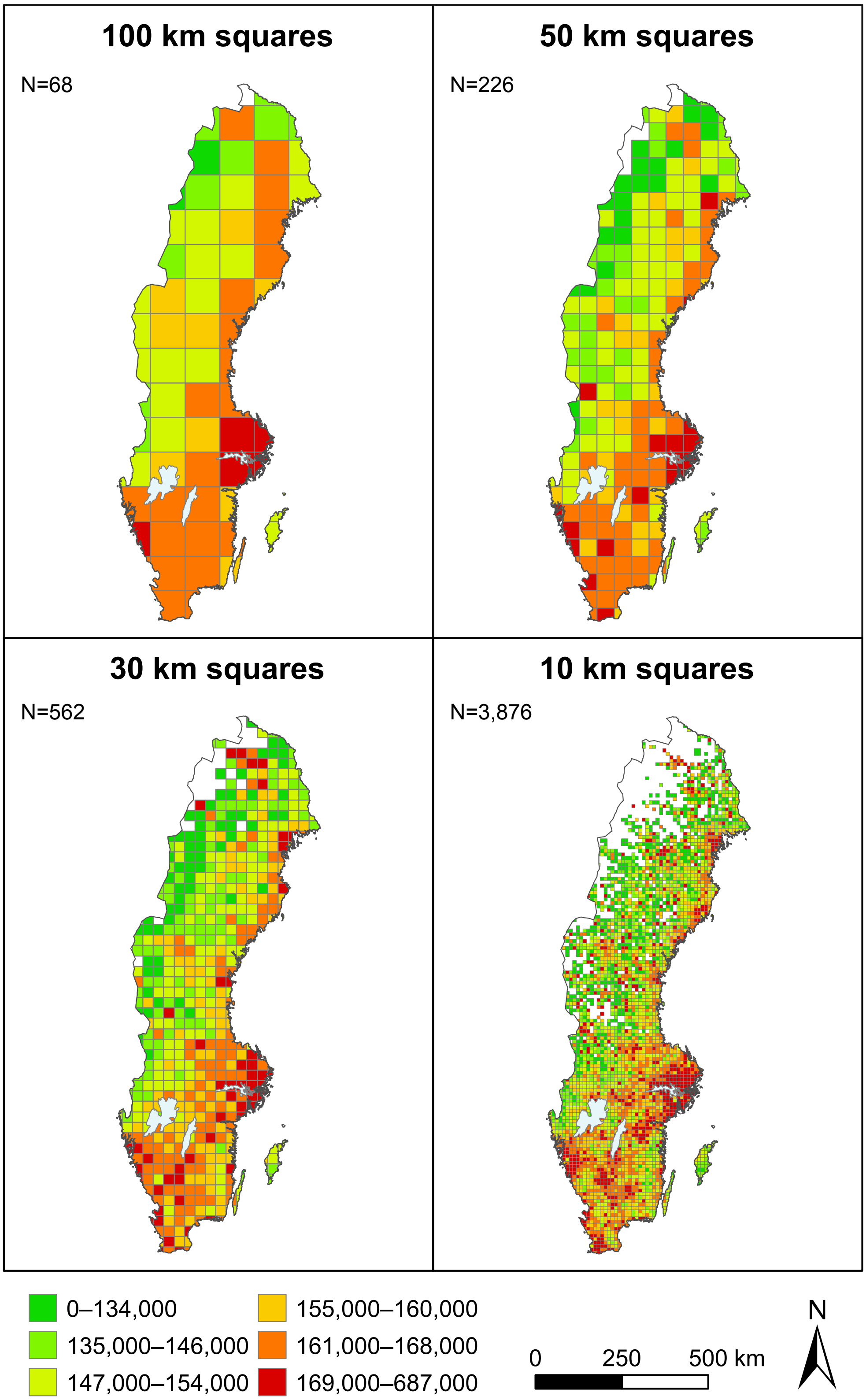


Figure 10 Disposable income (SEK)/inhabitant (20–64 years; 2002) – squares.

The grid maps reveal that the aforementioned "wedge" can be attributed to a limited number of wealthy and densely populated 10 km squares. The squares in question are mainly located in areas with large deposits of iron ore and associated mining activity, the "iron belt." In addition, a number of areas in the inland of northern Sweden that have very low mean incomes are clearly visible up to the 50 km level.

Figure 11 summarizes the proportion of units in the different income classes, while Figure 12 shows the percentage of the population living in different unit classes of mean income. As is the case with the population change indicator, the proportion of inhabitants living in the squares with the lowest incomes is very small.

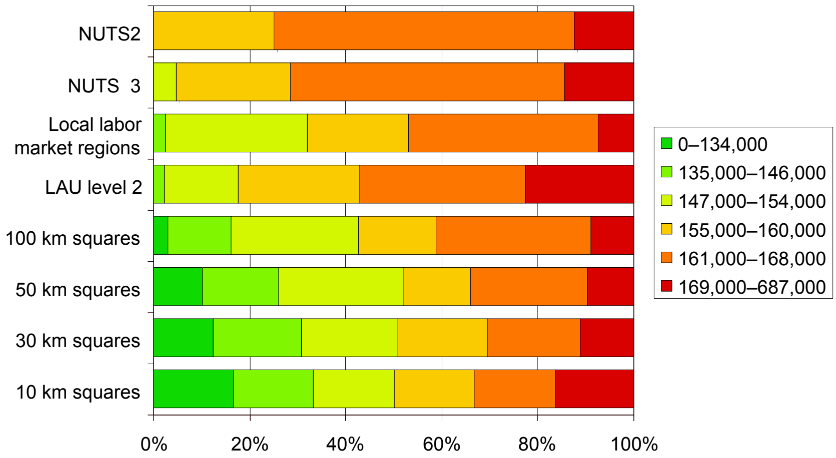


Figure 11 Disposable income (SEK)/inhabitant (20–64 years; 2002): proportion of units in the different classes.

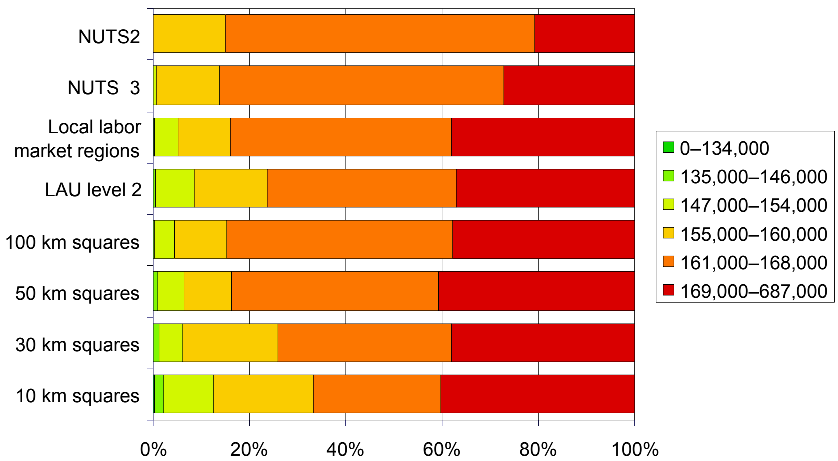


Figure 12 Disposable income (SEK)/inhabitant (20–64 years; 2002): share of population (all ages) in the different unit classes.

Table 3 contains descriptive statistics for the disposable income indicator. The reduction in range is still considerable, but the least conspicuous compared to the other variables. The minimum and maximum values converge with higher aggregation, while the means – and also the medians – are quite similar for all groups. The grids have slightly lower means/medians compared to the territorial maps. This is mainly due to the fact that administrative areas tend to contain both richer urban and poorer rural areas, while many grid squares are wholly rural in character.

Table 3 Descriptive statistics, disposable income (SEK)/inhabitant (20–64 years; 2002).

	N	Min.	Max.	Range	Mean	Median	Std. dev.	Skew-ness	Kurtosis
NUTS 2	8	158,400	195,400	37,000	166,700	163,700	12,000	2.5	6.5
NUTS 3	21	150,000	195,400	45,400	163,900	163,000	8,900	2.2	7.6
Local labor market regions	81	137,200	195,200	58,000	158,400	159,000	8,600	0.7	3.4
LAU level 2	284	137,200	352,400	215,200	164,500	160,600	18,600	4.7	38.7
100 km squares	68	124,200	20,7400	83,200	156,465	158,500	13,300	0.8	3.2
50 km squares	226	800	251,100	250,300	152,366	153,100	19,300	-2.0	20.4
30 km squares	562	800	476,500	475,700	152,120	153,400	23,700	3.9	67.1
10 km squares	3,876	0	686,800	686,800	151,387	153,600	27,900	2.3	48.5

In the Hypercarte Project, Grasland showed the discontinuity that appears along the French-Belgian border when mapping population change between the years 1980 and 1990. For the reader it is impossible to compare the differences in population change between the two sides of the border, since the NUTS regions are so differently sized. In similar maps of population density, the heterogeneity of the NUTS regions becomes less troublesome, due to higher spatial autocorrelation – greater similarity between neighboring areas – at the concerned level of resolution [Grasland, personal communication]. Tests for spatial autocorrelation (Moran's *I*)[4], show that – in the Swedish context, too – the process of population change in many cases tends to exhibit lower spatial autocorrelation compared to population density (Table 4). At the 10, 30, and 50 km square levels, as well as in the municipality maps, population change exhibits the lowest spatial autocorrelation of the three indicators. Naturally, the wide variety of degrees of population change in northern Sweden is an important factor contributing towards this result. Interestingly, comparing the grid maps at the different levels, mean disposable income is consistently the indicator that has the highest value for Moran's *I*.

Table 4 Tests for spatial autocorrelation (Moran's *I*).[4]

	Population density	Population change	Disposable income
NUTS 2	Not significant	Not significant	Not significant
NUTS 3	Not significant	Not significant	Not significant
Local labor market regions	0.098	0.145	0.126
LAU level 2	0.389	0.230	0.332
100 km squares	0.040	0.073	0.123
50 km squares	0.064	0.021	0.117
30 km squares	0.037	0.016	0.077
10 km squares	0.022	0.014	0.065

4 Moran's *I* is the perhaps most commonly used test statistic for spatial autocorrelation – degree of similarity between adjacent regions. The possible values of Moran's *I* range between -1 and 1. Positive values indicate clustering, while negative values indicate dispersion. In these calculations, an inverse distance function was used to conceptualize the spatial relationship between the map units.

A Close-Up Look

When maps are interpreted, it is very important be aware of the weaknesses – that they may present a distorted picture, or at least conceal some aspects. The maps that have been shown until now display a sparsely populated and quite poor area in northern Sweden that contrasts with a more well-off southern part of the country. However, a closer look at the data behind the squares reveals that there is more to the picture than meets the eye.

There are, in total, twenty-three 50 km-squares in the poorest income class (mean disposable income ≤134,000 SEK) (cf. Figure 10). In these squares, totaling a land area of over 40,000 km^2, only 3,124 inhabitants in the age 20–64 years reside (6,185 in total). By contrast, in Malmö municipality there are twelve contiguous 1 km squares with a similarly low mean disposable income (Figure 13). In this small area, 42,129 individuals in the age range are found (50,807 in total)! The area is centered on "Rosengård," a council estate with a high proportion of foreign-born and second-generation immigrants. Similar pockets of poverty – albeit perhaps not quite as extreme – can be found in Stockholm and Gothenburg as well as several other large Swedish cities. Obviously, such low-income enclaves disappear in the crowd of more wealthy people in the surroundings when data is displayed with squares of higher resolution.

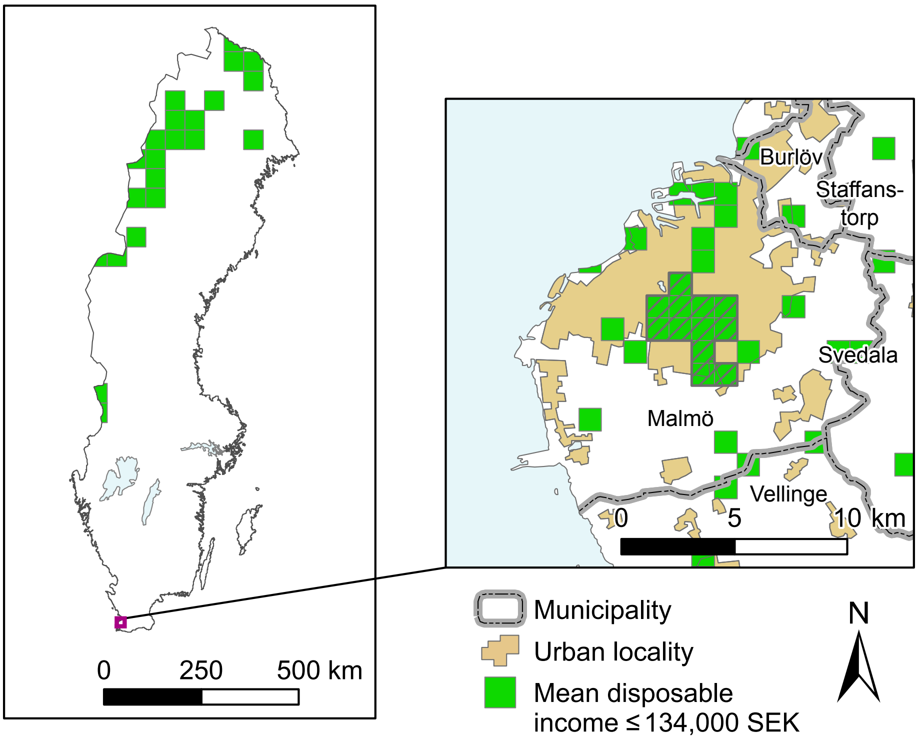


Figure 13 The twenty-three poorest 50 km squares (≤134,000 SEK; left) and twelve contiguous 1 km squares in Malmö with similarly low mean disposable income (right).

As stated above, the twenty-three poorest 50 km squares – all with a mean disposable income ≤134,000 SEK – contain 3,124 individuals between 20 and 64 years. The mean income of these few thousand individuals is about 128,500 SEK. In the richest 50 km square – which is centered on Stockholm – there are about 60 times as many inhabitants with similar low disposable incomes. If the richer part of the 20–64 years old in this wealthiest 50 km square in Sweden decided to move away, the population in the concerned age group would drop from 611,595 to 182,889, and the mean income from 205,003 SEK to 77,898 SEK. Thus, especially low incomes might not be the biggest "problem" in the poorest 50 km squares, but rather a severe deficit of rich people, which contributes to keeping average incomes at low levels.

A Small Experiment

In Table 5, the mean disposable income for the different map types and levels of resolution is displayed. In addition, the mean value is displayed for squares with a resolution of 1 km and at the individual level (population 20–64 years). In a sense, the empirical means at the different levels constitute overall measures of the errors that are introduced in the different maps. When aggregating income levels from very small grids, the mean value decreases at first until somewhere between 10 and 30 km whereupon it increases. In the beginning of the aggregation process, there are lots of uninhabited squares outside urban areas. Therefore, the comparatively small number of rich squares concentrated in urban and densely populated areas decrease at the fastest rate. When the squares reach a certain size, however, the process is reversed, since densely populated richer squares will increasingly incorporate sparsely populated poorer ones.

Table 5 Mean disposable income at different levels.

Level	Mean disposable income
NUTS 2	166,700
NUTS 3	163,900
Local labor market regions	158,400
LAU level 2	164,500
100 km squares	156,465
50 km squares	152,366
30 km squares	152,120
10 km squares	151,387
1 km squares	160,100
Individuals	170,800

In another approach to estimate the error that is introduced when full information is not available, an, as simple as possible, OLS regression was employed. It serves the purpose of representing any member of the set of possible, more realistic causal models. Population density was used to predict disposable income at the different levels represented in the maps (Table 6). The density values used varied depending on map type and level of aggregation. Not surprisingly, R^2 is higher but significance lower at more aggregated levels. The intercept and – in particular – the slope coefficients exhibit different values in the various regressions.

Table 6 Population density as a predictor of disposable income at different levels.

Level	Density value	N	R^2	Parameters		Sig.
NUTS 2	National area	8	0.95	Const.: B:	148,073 112.87	3.15E-05
NUTS 3)	County	21	0.76	Const.: B:	151,013 31.43	2.61E-07
Local labor market regions	Local labor market region	81	0.45	Const.: B:	149,880 79.46	5.58E-12
LAU level 2	Local labor market region	284	0.44	Const.: B:	153,497 80.83	1.68E-37
100 km squares	Square	69	0.34	Const.: B:	153,462 169.03	2.13E-07
50 km squares	Square	227	0.13	Const.: B:	154,202 149.42	4.36E-08
30 km squares	Square	561	0.04	Const.: B:	158,390 129.09	2.03E-06
10 km squares	30 km radius[5]	3,989	0.09	Const.: B:	158,577 125.10	2.61E-82

5 The calculation of area in this context was carried out in a somewhat simplified way, compared to the procedure described in footnote 2.

The regression equation was then applied to all individuals (20–64 years); each individual was assigned the density value of the 10 km square they belong to. The empirical individual mean (170,800 SEK) was subtracted from the estimated means, producing a form of residual values (Table 7). Expressed in this way, the ”error” is the smallest at the 100 km square level, followed by the 10 km square level.

Density impact on income is obviously scale dependent and seems to increase with square size. Neighborhood density is something different from density in larger regions. The construction of municipalities and local labor market regions seems to internalize a large part of the variation. On regions as large as NUTS 2 and 3, the meaning of density becomes blurred and whatever results arise, might be the effect of pure chance.

Regression	Estimated mean	Residual
National area regression	185,683	14,883
County regression	186,361	15,561
Local labor market region regression	186,578	15,778
Municipality regression	190,088	19,288
100 km square regression	171,010	210
50 km square regression	167,097	-3,703
30 km square regression	157,822	-12,978
10 km square regression	172,529	1,729

Table 7 Regression equations applied on individuals (mean: 170,800 SEK).

Discussion

For the studied indicators, an increase in unit aggregation is associated with decreased heterogeneity on the maps, as well as smaller ranges and lower standard deviations. This finding is hardly surprising. The general effect of aggregation on mean and median values is less straightforward. In the twenty-four maps that have been presented, increased aggregation is associated with higher means and medians. In the case of these statistics, however, other indicators, map types, and aggregation levels may yield other outcomes. For instance, when aggregating income levels from very small grids, the mean and median values decrease at first.

Both territorial and grid maps have advantages as well as disadvantages. When comparing administrative systems and the consequences of different political decisions, territorial maps should be used. Preferably, the maps should be constructed with comparable areas for the whole area of interest. Even when this is feasible, however, the areas will still be difficult to compare. Different population densities and different population distributions within the areas complicate all socio-economic comparisons. When unique and atypical areas in Europe should be identified, it might be necessary to use an equal area method, for instance a grid map. Most localities that exhibit rare or exceptional values will appear in such a map if only the resolution is high enough. A high-resolution map can, however, easily become difficult to interpret due to a cluttered appearance.

A comparison between the territorial and the grid maps can be done by looking at the 50 km square maps (N=226 or 228) and the municipality maps that are divided into approximately the same number of units (N=284). Regional dissimilarities are quite obvious in the grids, but harder to discern in the municipality maps. All municipalities have a principal city or town where population densities and incomes are higher and population change often positive compared to their rural areas. When average values are displayed at the municipality level, the extreme values that actually are found in some places in most municipalities will disappear. Only a few squares will contain the – for the municipalities – dominating principal city, which leads to bigger differences in the grid maps. None of the territorial maps have units with the lowest class values, while in the grid maps such extreme values are quite common – especially in remote and sparsely populated areas.

Even with a resolution of 10 km squares, there is a scale problem. This is illustrated when the proportion of squares in the different classes are compared to share of the population in the different unit classes. The squares have been homogenized when it comes to area, but the population distribution is extremely irregular. With increased aggregation, the proportion of squares in the extreme ends tends to decrease while the "mainstream" squares increase. A similar but even more conspicuous change is seen concerning the percentage of individuals living in the different kinds of squares. At the highest map resolution (10 km squares), each classification group contains the same number of squares; the classification was defined this way. Already on this level, the population distribution is irregular due to varying population densities, but aggregation causes additional population redistribution. For instance, at the 10 km square level, about 0.3% of the whole population lives in the poorest 1/6 of the squares. When aggregated to 50 km^2, a mere 0.00083% lives in the about 3% squares of this type. This phenomenon explains a great deal of the difficulties in illustrating socio-economic data on maps in a satisfactory manner.

What this study has shown is that even with a high geographical resolution and equally sized units, the information displayed on the maps to some extent can be inadequate or even misleading. Moreover, high-resolution maps are unpractical to use even at the national level: the grid maps with the resolution of 10 km squares are fuzzy and quite hard to interpret. Yet, the advantage of micro data is that the user is free to aggregate it as he/she pleases and that it is possible to estimate the errors of the chosen aggregation.

The target for most social, economic, and geographic studies is human individuals – not pieces of land. That is just one out of several attributes of the core entity at study, the individual in his/her spatial and socioeconomic context. While an equal area-based classification, like the one used here, is well suited for visualization of e.g. the morphology of natural resources it might not be the best choice for attributes of individuals. A classification based on equal number of people per class should better represent the distribution of individuals. However, that alone doesn't help much in avoiding the problem of hiding large minorities with extreme characteristics in dense regions and exaggerating similar extreme attributes for small minorities in sparsely populated regions. When the distance to the socially constructed border of the administrative region is less important compared to i.e. the cognitive distance to friends, relatives, and activities, it might be more informative to use transformed maps based on equal density per region for visualizing attributes of individuals.

ESPON Action 3.4.3

"THE MODIFIABLE AREAS UNIT PROBLEM"

An Example of Statistical Analysis using Spatial Units from Ireland and Northern Ireland

Submitted by:

Professor A. Stewart Fotheringham
Martin Charlton
Dr. Jianquan Cheng

National Centre for Geocomputation
John Hume Building
National University of Ireland, Maynooth
Maynooth
Co. Kildare
IRELAND

Web: **www.ncg.nuim.ie**
Tel: **+353 1 708 6455**
Fax: **+353 1 708 6456**
Email: **ncg@nuim.ie**

An Example of Statistical Analysis using Spatial Units from Ireland and Northern Ireland

Abstract

The Modifiable Areal Unit Problem (MAUP) is a generalised term for several related problems in which the results of some type of analysis vary substantially when the spatial units used in the analysis vary. For instance, statistical information may be available for a series of different scales of spatial units such as NUTS level 2 or 3 zones or there might be a variety of ways of aggregating smaller units into larger units so that the results of the analysis when using the larger zones depend on how those zones have been aggregated from the set of smaller zones. A third type of modifiable areal problem occurs when the results of an analysis are sensitive to the definition of the spatial extent of the study region.

This report examines the effect of the MAUP in a particularly sensitive area of analysis – that of the border region between Ireland and Northern Ireland. Different sets of spatial units are defined on either side of the border, yet there is a growing need to understand the social processes on the region which straddles the border. Here, we examine the effects of the border on the analysis of the determinants of educational attainment by constructing a statistical model in which the proportion of people with a university degree in a small area is related to a series of socio-economic characteristics of that area. We utilise both a traditional 'global' regression model and the recently developed local technique of Geographically Weighted Regression (GWR). The results highlight the effects of the MAUP and the need for a consistently defined set of spatial units for statistical analysis throughout Europe.

1. Background

Within Europe most countries share a land border with at least one other country. Such borders are interesting for understanding the processes operating within such countries because they highlight potential differences in socio-economic processes caused by the different institutional, political, social and economic frameworks within different countries. An important research question is: "Are such differences diminishing under the umbrella of the European Union and the general movement towards globalisation?" The border regions are also interesting, however, for statistical analysis because often the reporting units for data will be different on either side of the border. A research question here is then: "To what extent do observed differences in statistical analyses either side of a national border result from data on either side being drawn from incompatible spatial units?" To some extent, the EU is tackling this problem by developing a consistently defined set of spatial units (NUTS units) but much work remains to be done particularly in terms of obtaining pan-European data sets for small spatial units.(NUTS levels IV and V).

The border between Ireland and Northern Ireland (the latter being part of the United Kingdom) provides perhaps the best study region in Europe for examining the effect of a national border on socio-economic processes and for examining what statistical problems are created by using different spatial units on either side of the border to examine these processes. Both countries were part of the United Kingdom until 1926 and both have much in common despite the presence of a national border for the past 80 years. Indeed, the border, once quite tightly controlled, is now virtually invisible.

As well as having an ideal set of data for which to examine the effect of different spatial units on statistical analysis, this project takes advantage of recent developments in Geographic Information Systems (GIS) and spatial statistics to improve our understanding of the modifiable areal unit problem (MAUP).

2. The Study Area

The study area consists of spatial units within two countries: Northern Ireland and the Republic of Ireland (Ireland). Ireland is an island on the western fringe of Europe. Its greatest length, from Malin Head in the north to Mizen Head in the south, is 486 km and its greatest width from east to west is approximately 275 km. Since 1921 the island has been divided politically into two parts: Northern Ireland in the North and the Republic of Ireland in the south with a land border (figure 1), the former is under the governing of United Kingdom.

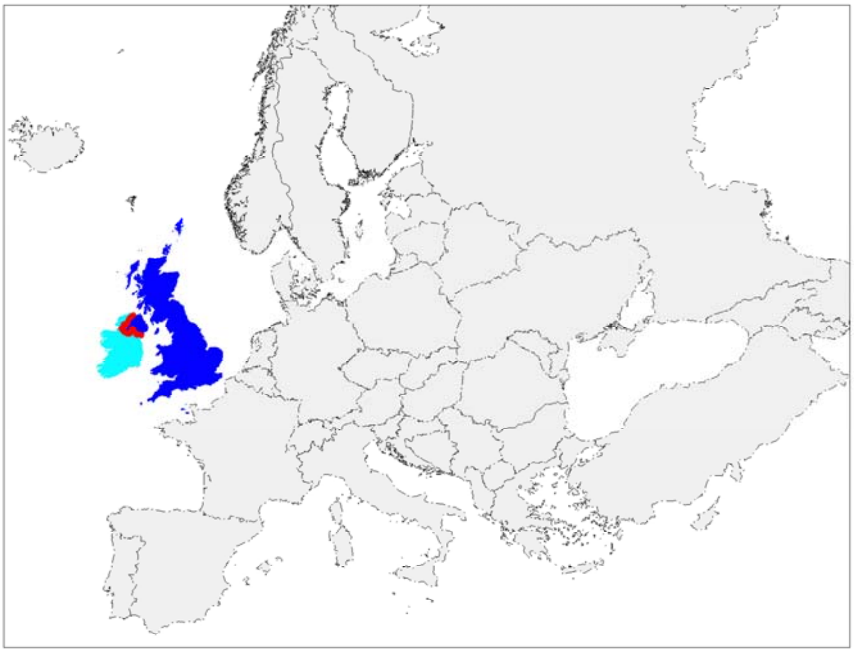


Fig 1. Ireland and N. Ireland in Europe

The Republic of Ireland has a different hierarchy of spatial statistical units for reporting population census data than does Northern Ireland; the latter follows the system of United Kingdom. Northern Irish data can be downloaded from the Northern Ireland Statistics and Research Agency (http://www.nisra.gov.uk/census/) whereas the Irish population census data are available through the National University of Ireland, Maynooth (http://www.nuim.ie/staff/dpringle/saps/index.htm).

The hierarchy of spatial units in the Republic of Ireland is as follows (from largest to smallest) (Figure 2):

County (and county Borough)
UD (Urban District) and RD (Rural District)
DED (District Electoral Division)

The DED is the smallest unit for which population census data are generally available, with 3440 covering the whole country in 2002.

The hierarchy of spatial units in Northern Ireland is as follows (from largest to smallest) (Figure 2):

County
District
Ward
Output Area (OA)

The OA is the smallest unit for which population census data are generally available with 5023 covering the whole country in 2001 (Northern Ireland is much smaller than the Republic of Ireland so the data reported for OAs in N. Ireland convey much more detail than do the DEDs of the Republic. The average size of an OA is 2.8km^2 whereas the average size of a DED is 20.4km^2

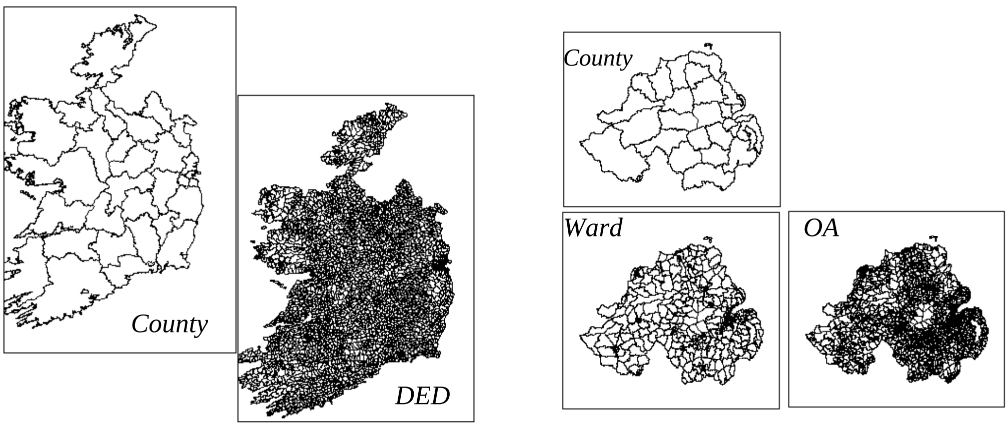


Fig 2. Spatial units in the Republic and the Northern respectively

The population census survey is held every 5 years in the Republic of Ireland, and every 10 years since in Northern Ireland. A population census was held in Ireland 1981, 1986, 1991, 1996 and 2002; the next one is scheduled to be held 2006. The reason for the delay in holding the census in 2002 was because of Foot and Mouth disease which prevented the census from being taken in 2001. A population census was held in Northern Ireland 1951, 1961, 1971, 1981, 1991 and 2001. As a result, the latest census survey from the two countries (2001 from Northern Ireland and 2002 from Republic) are reasonably temporally comparable.

Consequently, there are two reasonable choices that could be made regarding the spatial units either side of the border:

1.Use DEDS in the Republic of Ireland and OAs in Northern Ireland

2. Use DEDs in the Republic of Ireland and Wards in Northern Ireland

Finally, as well as the susceptibility of analytical results to the definition of the spatial units either side of the border, there is also a decision to be made in examining any cross-border issues of how far from the border should one define the spatial units which will provide the data for the analysis. To examine this effect, we chose three definitions:

1. All of the Republic of Ireland and all of Northern Ireland
2. Those spatial units within 100kms of the border (this includes both capital cities of Dublin and Belfast)
3. Those spatial units within 50kms of the border (this definition excludes both Dublin and Belfast)

Table 1 Variables used in the analysis

Variable	Meaning
BScPer Educational Attainment	Percentage of people over 17 yrs old with post secondary school education
Cars2H	Percentage of households with 2 and more cars;
Socc1P	Percentage of people in social class 1 over the population (between 16-74 for Northern and over 15 for Republic)
Erate	Percentage of people in employment over the population in the same age group (16-74 or >15)
PopDen	Total population divided by the total area
CathoPer	Percentage of population with Catholic religion

3. Modelling Educational Attainment Levels

3.1 Data

In order to demonstrate the sensitivity of analytical results to the definition of spatial units, we first construct a plausible model in which the proportion of population aged 18 years or over having a university degree is regressed on a set of socio-economic indicators which are given in the following table 1 and their spatial patterns are mapped in figure 3.

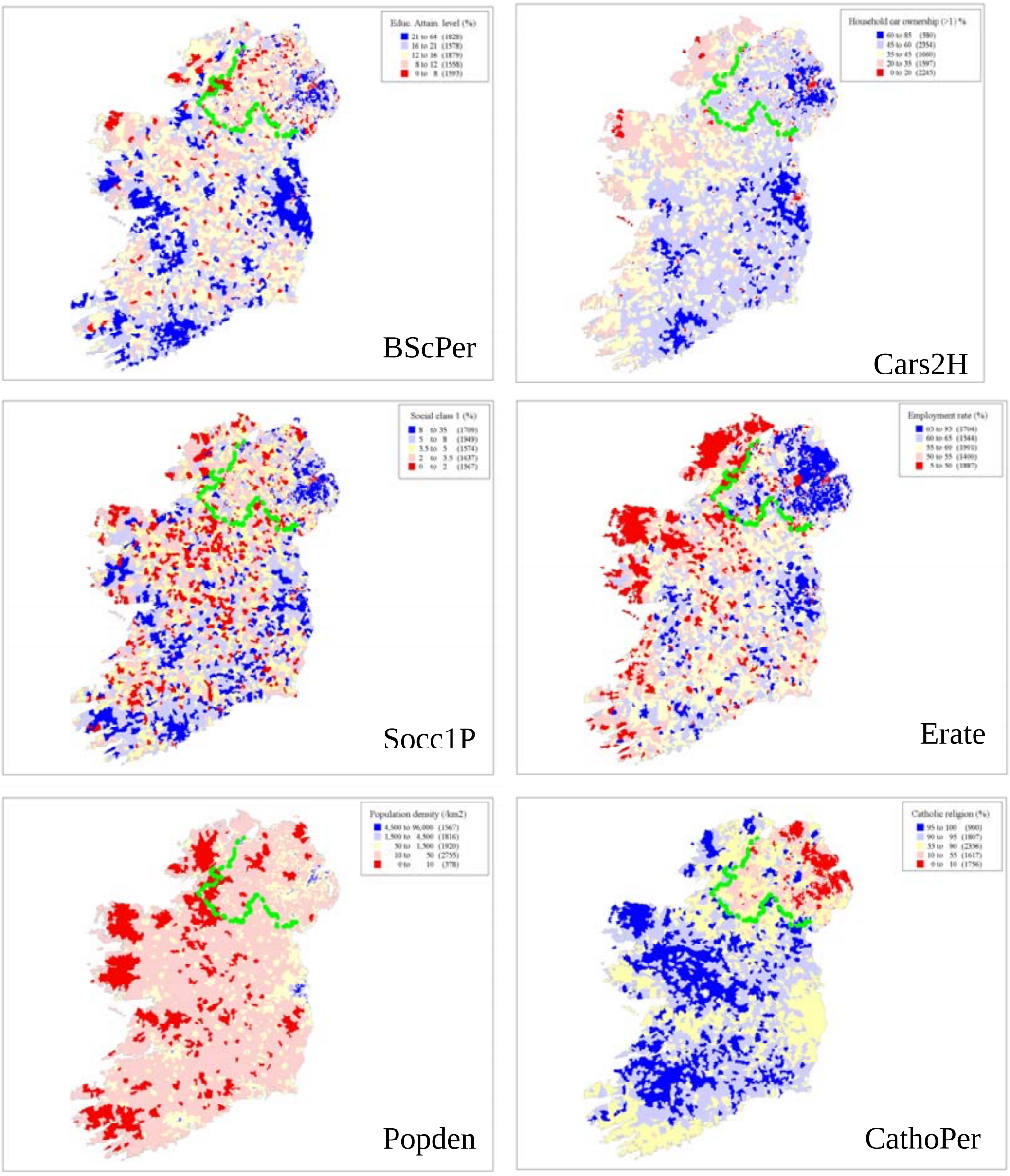


Fig 3. Spatial patterns of the defined variables

3.2 Methodology

A model was constructed in which educational attainment was related to the variables listed above in the following way:

$$BscPer = \alpha + \beta_1\ Cars2H + \beta_2\ Socc1P + \beta_3\ Erate + \beta_4\ Popden + \beta_5\ CathoPer \quad (1)$$

and the parameters of the model are estimated with data reported from the following sets of spatial units:

1. All OAs in Northern Ireland and all DEDs in the Republic of Ireland

2. Only those OAs in Northern Ireland

3. Only those DEDs in the Republic of Ireland

4. Only those wards in Northern Ireland

5. All wards in Northern Ireland and all DEDs in the Republic of Ireland

6. Only those OAs in Northern Ireland and those DEDs in the Republic of Ireland within 100kms of the border

7. Only those OAs in Northern Ireland and those DEDs in the Republic of Ireland within 50kms of the border

8. Only those OAs in Northern Ireland within 100kms of the border

9. Only those OAs in Northern Ireland within 50kms of the border

10 Only those DEDs in the Republic of Ireland within 100kms of the border

11. Only those DEDs in the Republic of Ireland within 50kms of the border

A comparison of the results from the model calibration using the units defined in 1 and 5 indicate the effect of using wards instead of Output Areas (smaller units) in Northern Ireland in an all-Ireland analysis.

A comparison of the results from the model calibration using the units 1, 2 and 3 and also the units in 3, 4 and 5 indicate the effect of the border on the determinants of educational attainment and whether the measurement of this effect is influenced by the choice of spatial units in Northern Ireland.

A comparison of the results from the model calibration using the units 1,6 and 7 and those in 2, 8 and 9 and those in 3, 10 and 11 indicate the effect of the spatial extent of the study region on the analytical results.

In addition to these different sets of spatial units, the model in equation (1) was calibrated both by traditional global regression and also by

Geographically Weighted Regression (selected cases) and the results of both types of model calibration compared to further examine the effects of the MAUP on statistical analysis of spatial data.

4 Global Regression Results

4.1 Using Data from all OAs in Northern Ireland and all DEDs in the Republic of Ireland

BscPer = -8.64 *(0.36)* + 0.01 Cars2H *(0.005)* +1.55 Socc1P *(0.014)* + 0.2 Erate (0.007)+0.0003 Popden (0.00002)+0.06 CathoPer *(0.0016)*

(2)

R^2 = 0.75

The figures in parentheses are the standard errors of the parameter estimates.

From these results we would infer the following conclusions about the determinants of educational attainment:

a. All five explanatory variables are significant in determining educational attainment levels.
b. Educational attainment is positively related to the percentage of people in high socio-economic groups and to the percentage of people with 2 or more cars – both measures of 'wealth'.
c. Educational attainment is positively related to the rate of employment in an area – measure of national economic situation.
d. Educational attainment is related to the proportion of Catholics in an area.
e. Educational attainment is slightly related to the population density, which differs urban from rural areas.

Table 2. Global regression models

Case	1	2	3
Unit	DED/OA	OA	DED
Extent	Ireland	N. Ireland	The Republic
Sample	3414 /5022	5022	3414
R^2	0.75	0.83	0.62
Regression	-8.6 (0.36, -24.3)	-7 (0.35, -20)	12.5 (1.3, 9.7)
Co-efficient	0.01 (0.005, 2.4)	0.03 (0.005, 5.7)	-0.003 (0.009, -0.31)
(α, β_1, β_2,	1.55 (0.014, 115)	1.63 (0.015, 109)	1.17 (0.025, 47)
β_3, β_4, β_5)	0.2 (0.007, 30)	0.16 (0.007, 23)	0.26 (0.016, 16.6)
(Sta. err,	0.0003 (0.00002, 13)	0.0002 (0.00002, 12)	0.0008 (0.00006, 13.8)
t-value)	0.06 (0.0016, 37)	0.05 (0.002, 27)	-0.18 (0.011, -16.3)

4.2 Using Other Data Sets

The same model was calibrated with 10 other data sets described above and the results are presented in Table 2. These show the sensitivity of the results of the model calibration on the definition of the spatial units for which data are reported and the spatial extent of the area being examined.

Table 2. Global regression models (continued-1)

Case	4	5	6
Unit	WARD	DED/WARD	DED/OA
Extent	N. Ireland	Ireland	100 km
Sample	582	3414/582	1465/5022
R^2	0.89	0.6	0.81
Regression	-4.7 (0.96, -4.9)	-9.5 (0.8, -11.7)	-8.3 (0.35, -24)
Co-efficient	-	0.003 (0.009, 0.38)	0.02 (0.005, 3.7)
(α, β_1, β_2,	1.86 (0.035, 53)	1.35(0.023, 57.6)	1.7 (0.014, 120)
β_3, β_4, β_5)	(0.02, 5.4)	0.25 (0.015, 16.9)	0.18 (0.007, 27)
(Sta. err,	0.0003 (0.00007, 3.8)	0.0009(0.00006, 16.4)	0.00023 (0.00002, 12.6)
t-value)	0.05 (0.0036, 14)	0.054(0.004, 14)	0.06 (0.0016, 38)

Table 2. Global regression models (continued-2)

Case	7	8	9
Unit	DED/ OA	OA	OA
Extent	50 km	100 km	50 km
Sample	662/2551		2551
R^2	0.70		0.75
Regression	-6 (0.5, -12)	As 2	-6 (0.53, -11.3)
Co-efficient	0.066 (0.007, 9)		0.07(0.008, 8.7)
(α, β_1, β_2, β_3,	1.48 (0.03, 56)		1.57 (0.03, 56)
β_4, β_5)	0.13 (0.01, 13.6)		0.12 (0.01, 12)
(Sta. err.,	0.00015 (0.00004, 3.7)		0.0002 (0.00004, 5.2)
t-value)	0.05 (0.002, 22)		0.05 (0.003, 19)

Table 2. Global regression models (continued-3)

Case	10	11
Unit	DED	DED
Extent	100 km	50 km
Sample	1465	662
R^2	0.79	0.36
Regression	7.9 (1.66, 4.8)	2.77 (2.03, 1.36)
Co-efficient	-0.026 (0.013, -2.1)	0.007 (0.022, 0.32)
(α, β_1, β_2, β_3,	1.62 (0.033, 49)	0.9 (0.07, 12.6)
β_4, β_5)	0.24 (0.02, 12)	0.2 (0.032, 6.3)
(Sta. err.,	0.0005 (0.00007, 7.7)	0.0023 (0.0006, 4.15)
t-value)	-0.13 (0.014, -9)	-0.03 (0.017, -1.94)

For instance, it is clear that the effects of Catholic religion on educational attainment are remarkbly different between Northern Ireland (positive) and the Republic (negative) (compare the parameter estimates for data sets 2 and 4 with those for data set 3). This table also accounts for the weaker coefficient when a combination of DEDs and wards rather than DEDS and OAs is used because when wards are used the number of spatial units in the analysis from the North is far fewer than when OAs are used and hence the influence of the processes in Northern Ireland are greatly diminished in the analysis (compare the parameter estimates between data sets 1 with 5). Equally, it is clear that the effects of social class on education attainment are greater in N. Ireland than in the Republic.

The results (data sets 1, 9, 10, 11) also show the effect of changing the spatial extent of the study region on the analyses.

4.3 Calibrated by Local GWR Modelling

GWR is an innovative local spatial exploratory analysis method, which incorporates spatial weight, based on distance decay function, into global regression equation. It can optimally search for an adaptive bandwidth based on a cross-validation procedure. Previous studies have proven its effectiveness in exploring spatial non-stationarity and also the possibility in searching local determinants. In this project, GWR modelling is implemented based on GWR 3.0 (http://ncg.nuim.ie/ncg/GWR/), with a loose coupling with GIS (MapInfo). To test the spatial non-stationarity of the determinants affacting educational attainment, we make four combinations of units with varied resolutions or extents, corresponding to the units 1,3,5,7 defined above. Major outcomes from the GWR modelling are listed in table 3.

Table 3. Outcomes from GWR modelling

Case	1	5	3	7
Unit	DED/OA	DED/WARD	DED	DED/ OA
Extent	Ireland	Ireland	Republic	50km
Sample	3414 /5022	3414/582	3414	662/2551
R^2	0.75	0.6	0.62	0.7
Gwr-R^2	0.85	0.81	0.79	0.79
Local esti. (0.1%)	α, β_1, β_2, β_3, β_4, β_5	α, -, β_2, -, β_4, β_5	α, -, β_2, -, β_4, β_5	α, β_1, β_2, β_3 (5%), β_4, β_5
Bandwidth	433	205	205	185
Residual. Impro.	41%	56%	49.5%	34%

Note: '-' is not locally significant

Case 1 is the most plausiable local GWR model using data from all smallest units in both sides. The results (from table 3) showed a much improved accuracy of modelling and especially all five variables are locally significant as a determinant. The spatial patterns of the major outputs from case 1 are displayed in figures 4-8 (left parts). However, with the changes of resolutions and extents, the spatial variability of determinants is sufferring from a great diversity. The comparisons between them are able to test the spatial effects of varied resolutions and extents on the variability of local determinants. For example, a comporison of the results from local model calibration using the units defined in 1 and 5 can disclose the effects of using wards instead of OA in Northern Ireland in an all-Ireland analysis. In table 3, a major change is detected as the decline of the significance of local variability of two variables (Cars2H and Erate) despite of the fact that the residual improvement of the later is much greater than the former (56>41). This result has showed the higher dicriminability of local GWR modelling than global one as the table 3 implied much difference existing between the results of case 1 with case 5 if compared with those in table 2.

A comparison of the results from the local model calibration using the units 1 and 3 indicate the effect of the border on the determinants of educational attainment and whether the measurement of this effect is influenced by the addition of the units in Northern Ireland. The results in table 3 shows the remarkable role of the units in Northern Ireland, which help the significance of local variability of the two variables.

A comparison of the results from the model calibration using the units 1 and 7 indicate the effect of the spatial extent of the study region on local analytical results. Table 3 shows a slight change indicated by the decline of local variability significance of the variable Erate.

The effects of the MAUP on spatial analysis described above are mainly focused on the perpective of amount. Table 3 implies that the determinants (social class 1, populdation density and Catholics) are scale independent across the defined scales (the four combinations).

In some cases, the spatial patterns of the effects will help find out the spatially ‘detailed’ hotspots. To be comparable and focused, figures 4-19 only illsutrate the areas within the buffering distance 50 km to the border. The comparisons are made in the parameters of standard residual, adjusted R^2, local parameter estimation and their corresponding t-values. Figures 4-8 aim to show the effects of changing units from OA to Ward in Northern Ireland in the model of taking the whole Ireland into account. Figures 9-13 attempt to illustrate the impacts of changing boundary with and without Northern Ireland being included in the model. Figures 14-19 can explain the effects of changing extent when the spatial coverage is shrank from all Ireland to the areas within 50 km buffering distance to the border. The comparisons between the spatial patterns in these figures can help us understand where and how much the areas will be impacted when the resolutions, boundary and extent are changed.

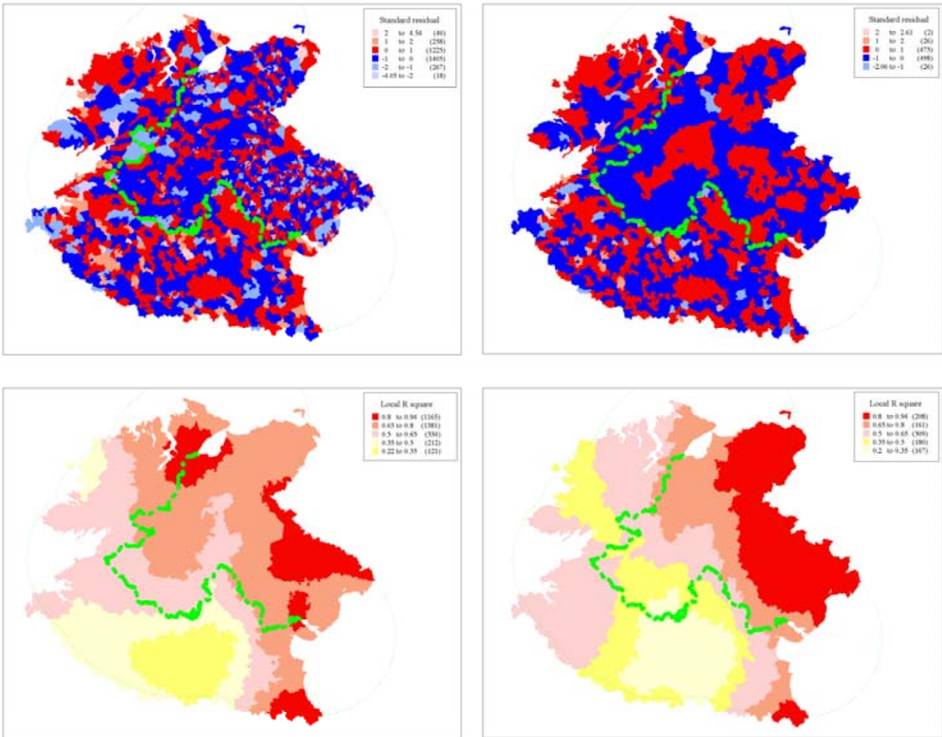


Fig 4. Local standard residual and adjusted R^2 (case 1 and case 5)

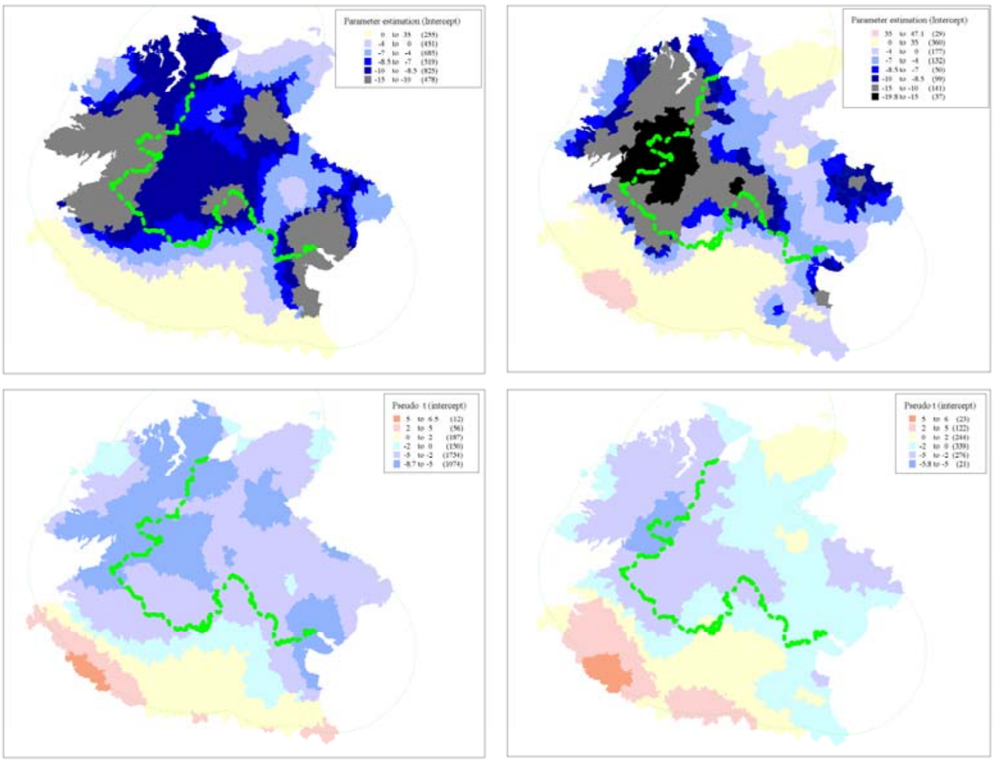


Fig 5. Local parameter estimation and Pseudo t (Intercept) (Case 1 and 5)

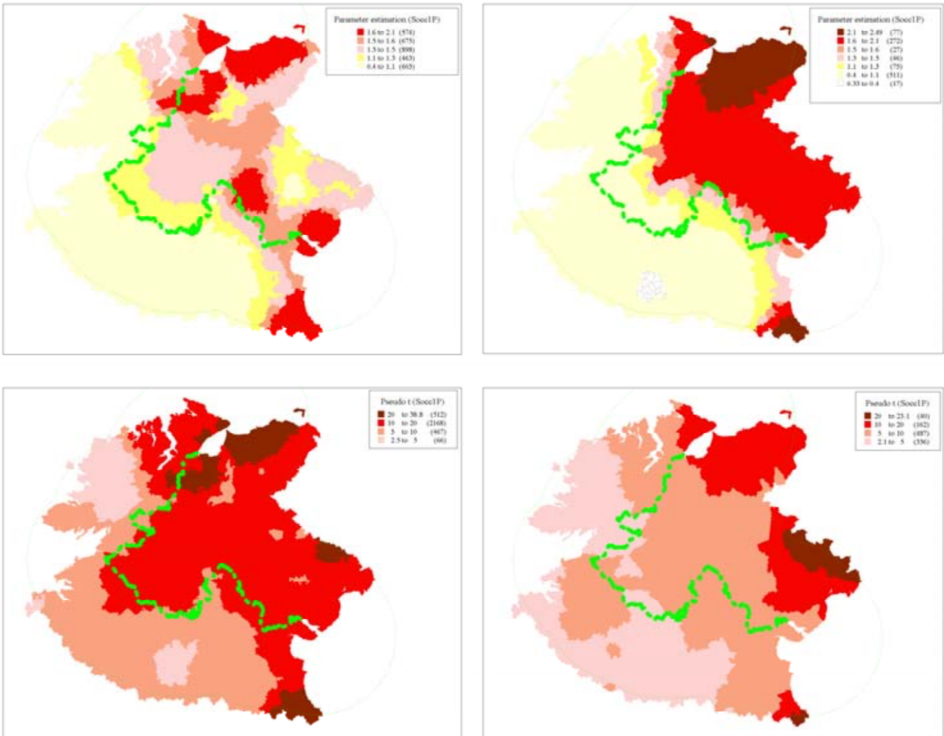


Fig 6. Local parameter estimation and Pseudo t (Socc1P) (Case 1 and 5)

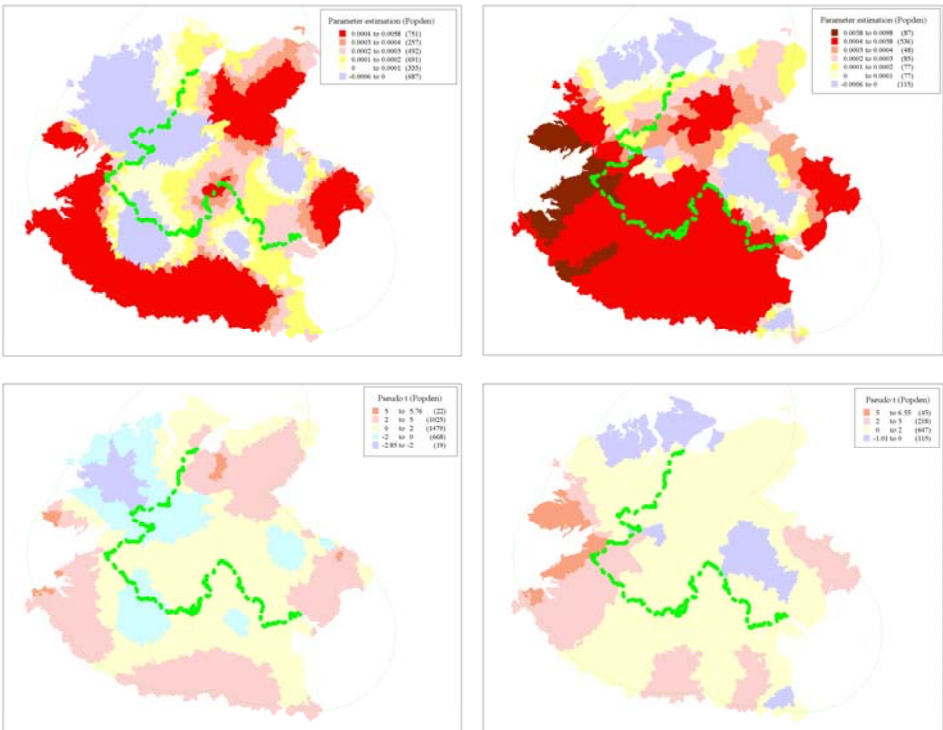


Fig 7. Local parameter estimation and Pseudo t (Popden) (Case 1 and 5)

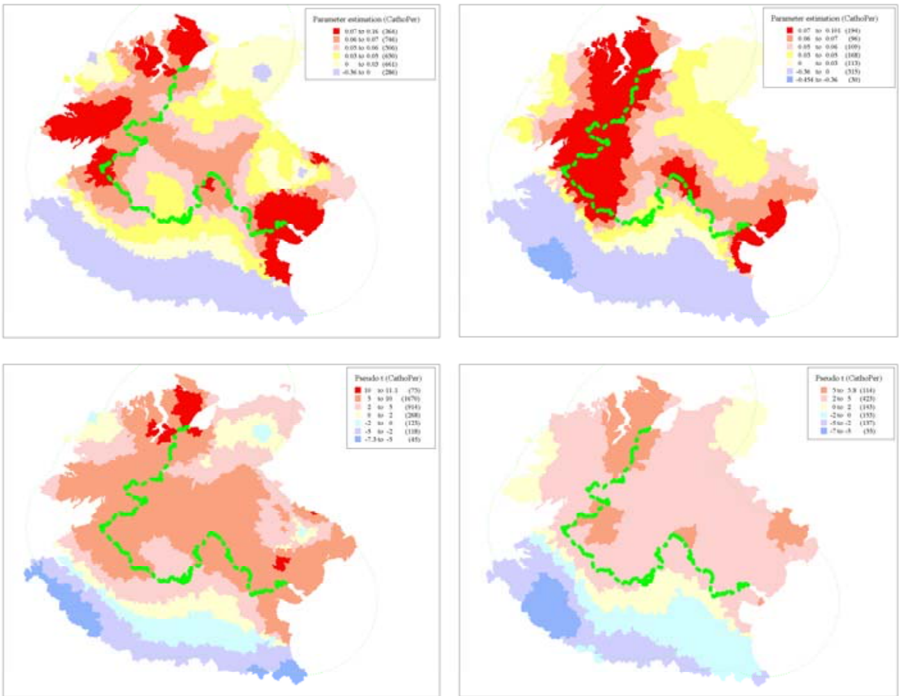


Fig 8. Local parameter estimation and Pseudo t (CathoPer) (Case 1 and 5)

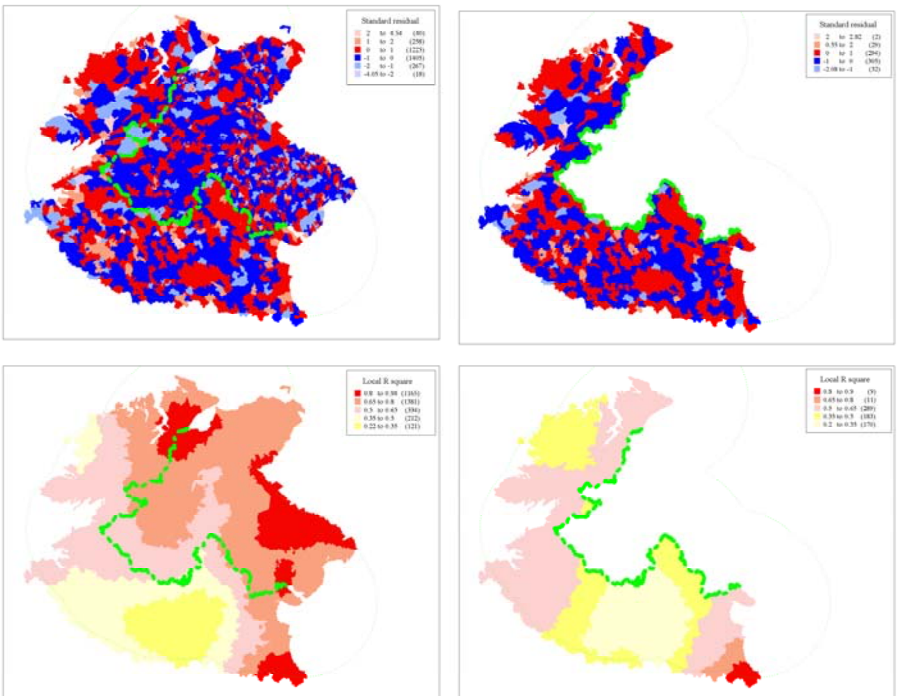


Fig 9. Local standard residual and adjusted R^2 (case 1 and case 3)

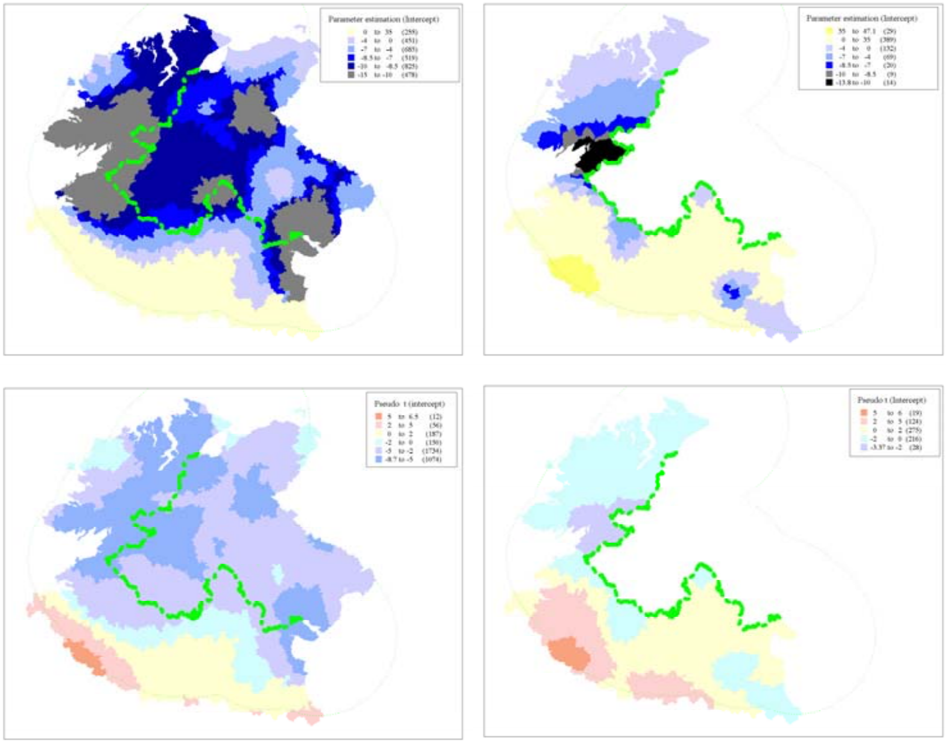


Fig 10. Local parameter estimation and Pseudo t (Intercept) (Case 1 and 3)

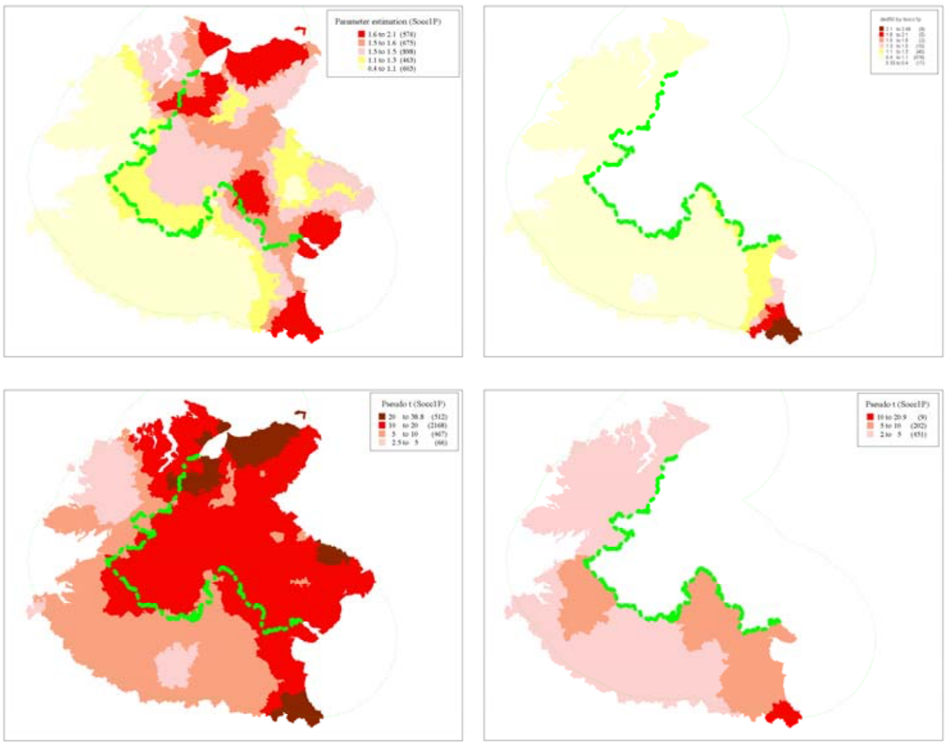


Fig 11. Local parameter estimation and Pseudo t (Socc1P) (Case 1 and 3)

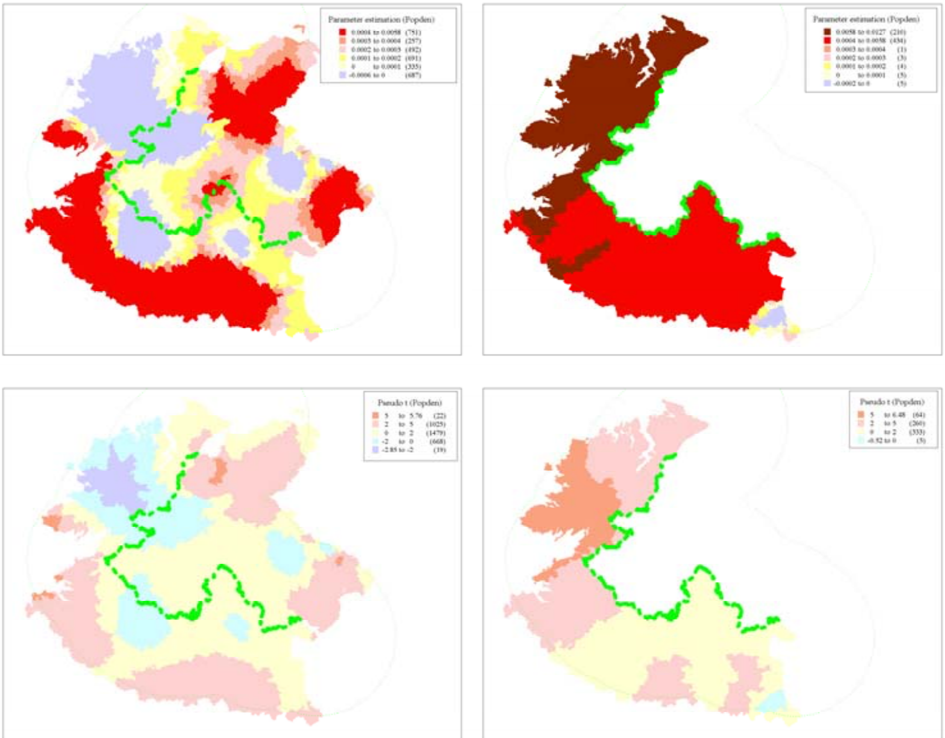


Fig 12. Local parameter estimation and Pseudo t (Popden) (Case 1 and 3)

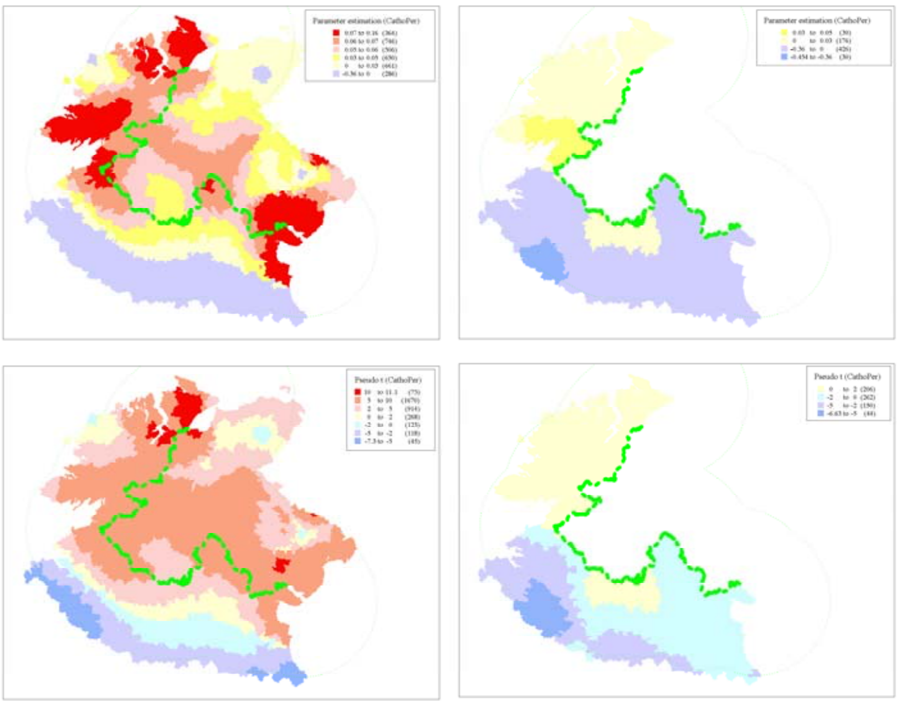


Fig 13. Local parameter estimation and Pseudo t (CathoPer) (Case 1 and 3)

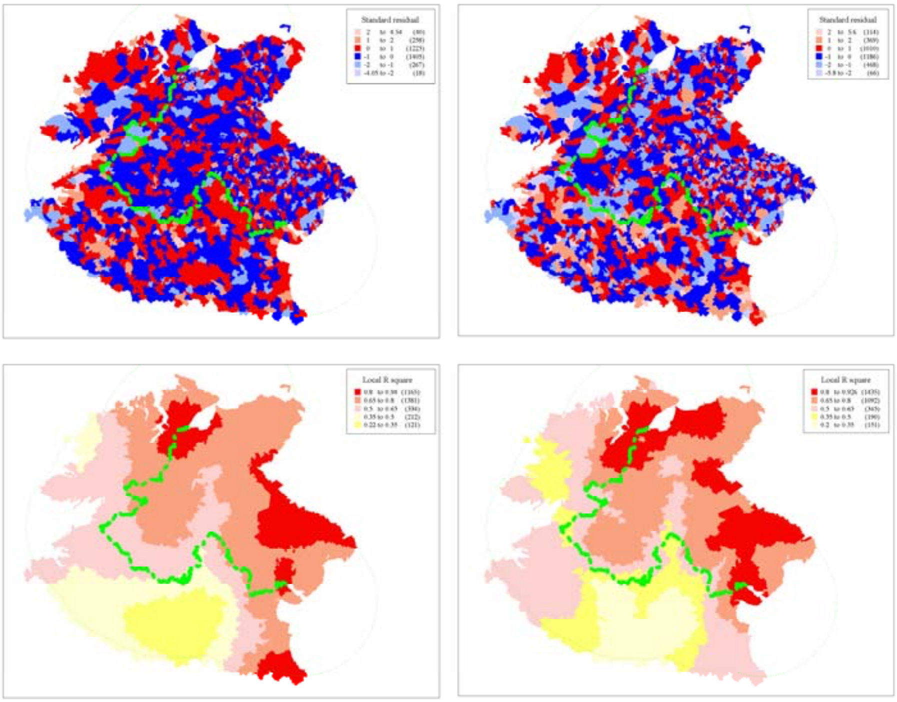


Fig 14. Local standard residual and adjusted R^2 (case 1 and case 7)

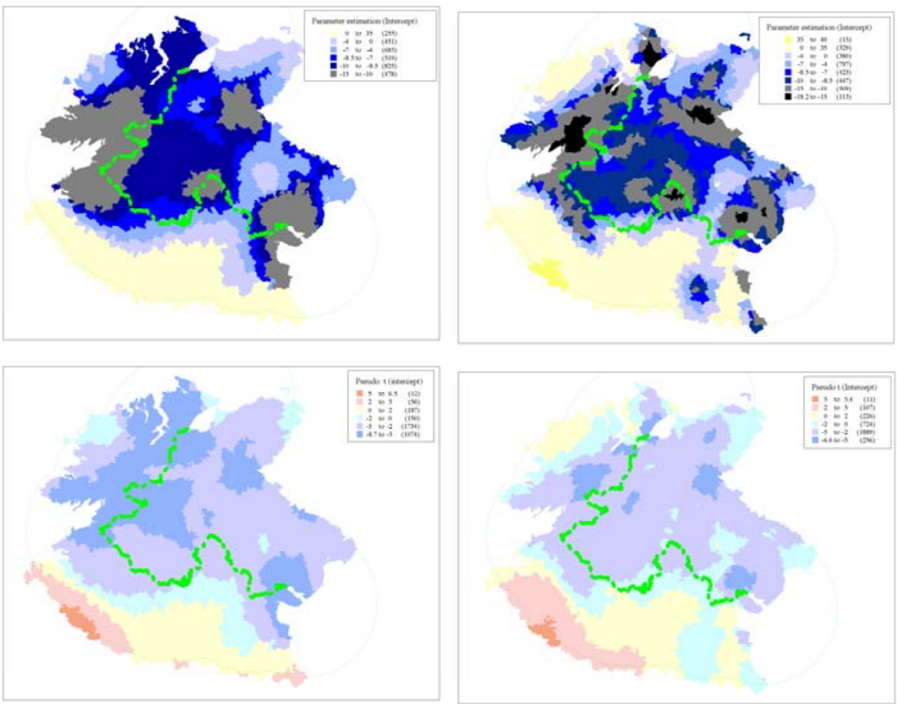


Fig 15. Local parameter estimation and Pseudo t (Intercept) (Case 1 and 7)

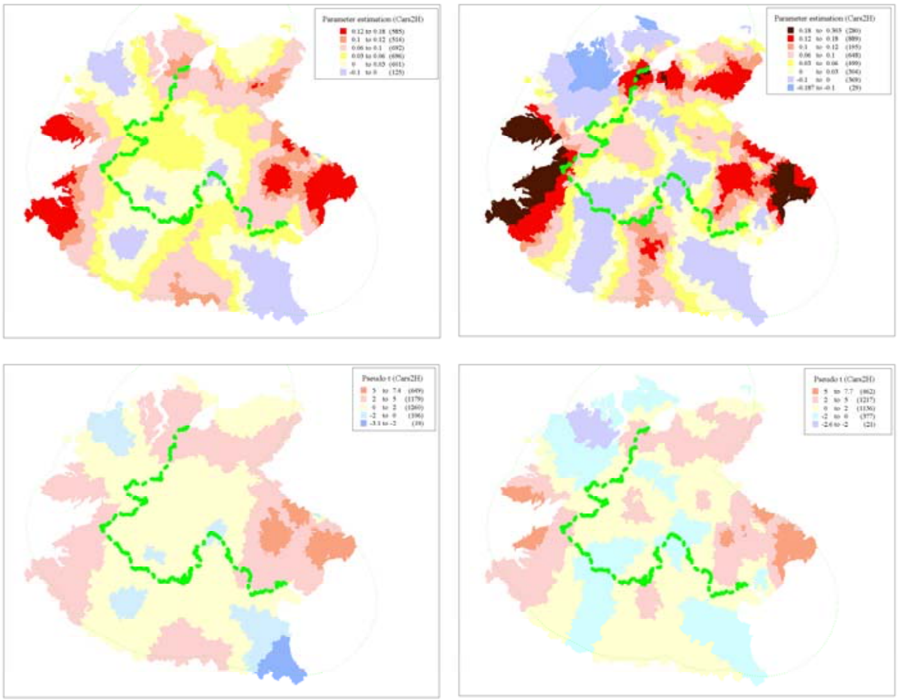


Fig 16. Local parameter estimation and Pseudo t (Cars2H) (Case 1 and 7)

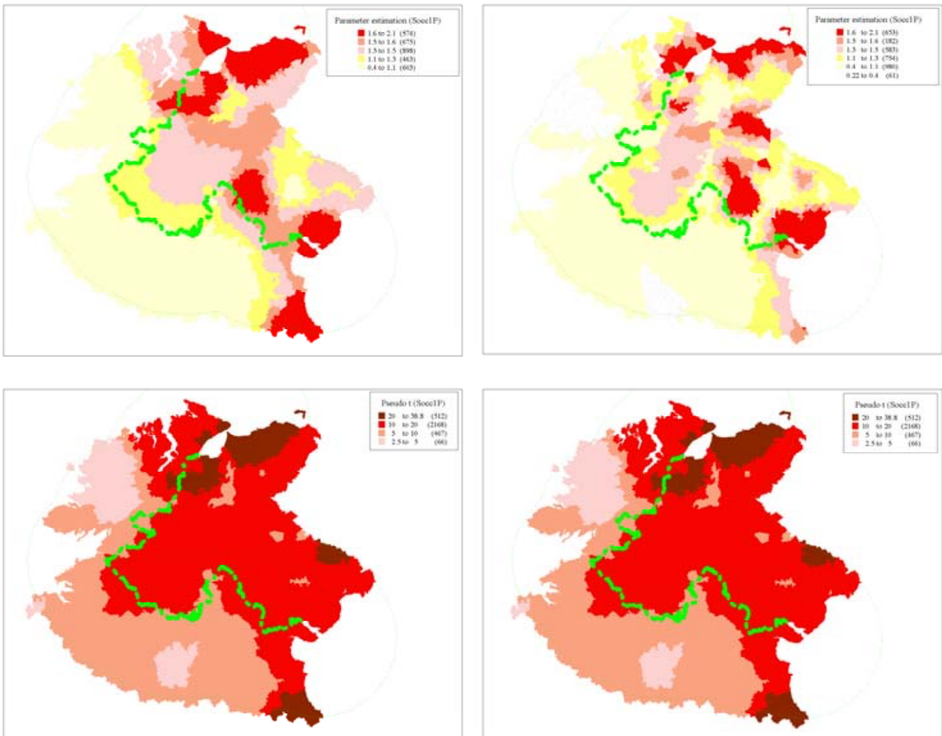


Fig 17. Local parameter estimation and Pseudo t (Socc1P) (Case 1 and 7)

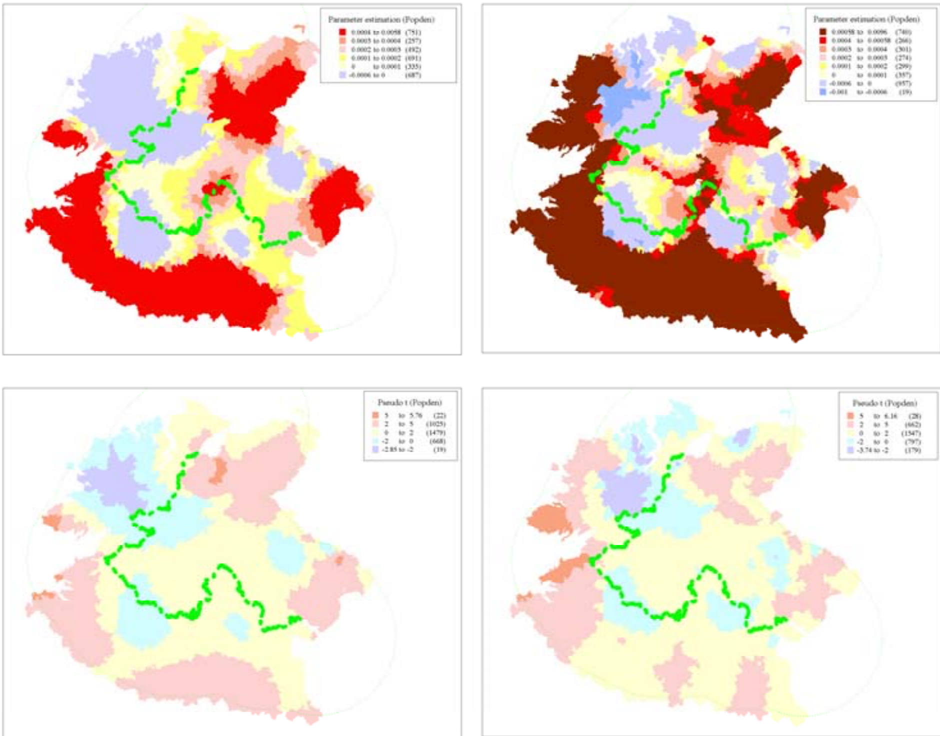


Fig 18. Local parameter estimation and Pseudo t (Popden) (Case 1 and 7)

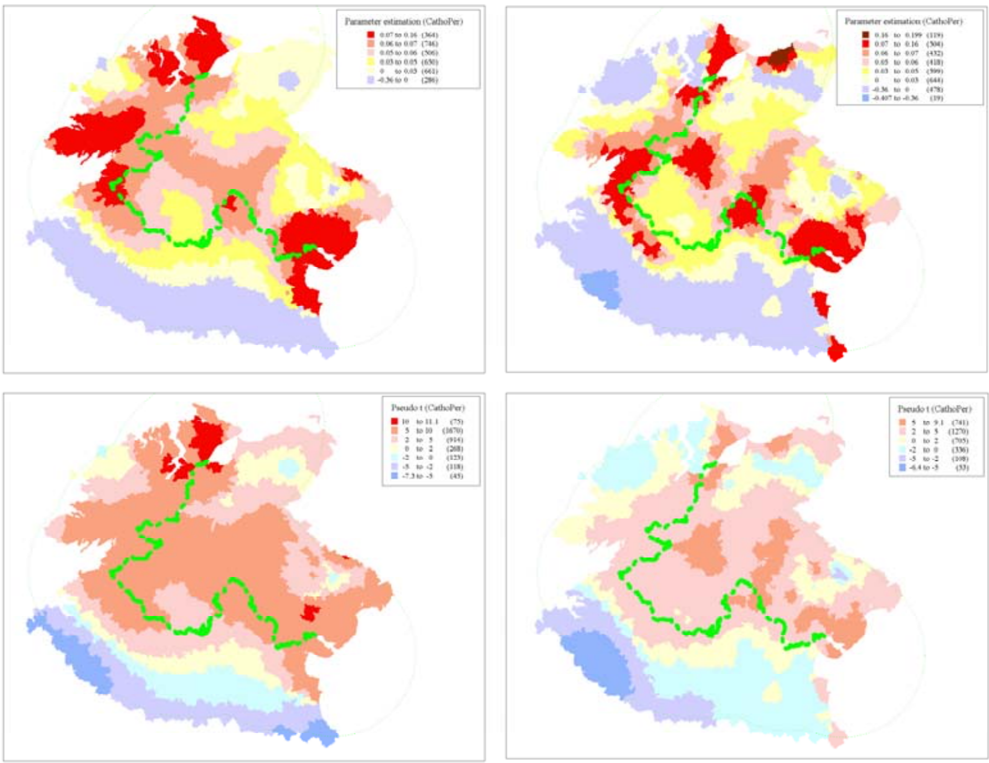


Fig 19. Local parameter estimation and Pseudo t (CathoPer) (Case 1 and 7)

5. Conclusions

Cross-border studies are receiving many attentions recently as demanded by various socio-economic and political cooperations across nations and regions.

This report has sufficiently showed the MAUP effects on the determinants of educational attainment level, caused by the changes of the resolution of units and the extents of analysis based on a specific field of cross-border case study. ‘Global’ and local regression (GWR) modelling are all able to detect these effects, however, GWR has higher discriminability and exploratory capability in detecting hotspots. These effects further strenghten the requirements of developping consistent spatial units across countries and regions.

Annexes

Annex 1: ESPON and the MAUP : the example of GDP/inh. PPS in 1999

Table 1 : Number and population of ESPON territorial units according to NUTS99 delimitations

State		Population	Number of territorial units			Mean population (1999)		
Code	**Name**	**1999**	**Nuts1**	**Nuts2**	**Nuts3**	**Nuts1**	**Nuts2**	**Nuts3**
Old Member States		***373784***	***74***	***204***	***1085***	***5051***	***1832***	***345***
AT	Austria	8092	3	9	35	2697	899	231
BE	Belgium	10214	3	11	43	3405	929	238
DE	Germany	82189	16	40	441	5137	2055	186
DK	Denmark	5319	1	1	15	5319	5319	355
ES	Spain*	37981	6	17	50	6330	2234	760
FI	Finland	5165	2	6	20	2583	861	258
FR	France*	58624	8	22	96	7328	2665	611
GR	Greece	10538	4	13	51	2634	811	207
IE	Ireland	3745	1	2	8	3745	1872	468
IT	Italia	57647	11	20	103	5241	2882	560
LU	Luxembourg	433	1	1	1	433	433	433
NL	Netherlands	15812	4	12	40	3953	1318	395
PT	Portugal*	9666	1	5	28	9666	1933	345
SE	Sweden	8858	1	8	21	8858	1107	422
UK	United Kingdom	59501	12	37	133	4958	1608	447
New Member States		***75099***	***10***	***41***	***121***	***7510***	***1832***	***621***
CY	Cyprus	763	1	1	1	763	763	763
CZ	Czech Rep.	10283	1	8	14	10283	1285	735
EE	Estonia	1441	1	1	5	1441	1441	288
HU	Hungary	10066	1	7	20	10066	1438	503
LT	Latvia	3699	1	1	10	3699	3699	370
LV	Latvia	2432	1	1	5	2432	2432	486
MT	Malta	380	1	1	2	380	380	190
PL	Poland	38655	1	16	44	38655	2416	879
SI	Slovenia	1985	1	1	12	1985	1985	165
SK	Slovakia	5395	1	4	8	5395	1349	674
Candidate countries		***30666***	***2***	***14***	***70***	***15333***	***2190***	***438***
BG	Bulgaria	8209	1	6	28	8209	1368	293
RO	Romania	22457	1	8	42	22457	2807	535
Other Espon		***11589***	***2***	***14***	***45***	***5795***	***828***	***258***
CH	Switzerland	7144	1	7	26	7144	1021	275
NO	Norway	4445	1	7	19	4445	635	234
UE25		***448883***	***84***	***245***	***1206***	***5344***	***1832***	***372***
UE25+2		***479549***	***86***	***259***	***1276***	***5576***	***1852***	***376***
ESPON 29		***491138***	***88***	***273***	***1321***	***5581***	***1799***	***372***

* without ultra-peripheral regions

Table 2 : Number of territorial units with GDP/inh. PPS lower than 75% of EU25 average in 1999

State Code	Name	Units with GDP/inh < 75			% of units eligible to Obj. 1		
		Nuts1	Nuts2	Nuts3	Nuts1	Nuts2	Nuts3
Old Member States		**7**	**25**	**242**	**9%**	**12%**	**22%**
AT	Austria	0	0	5	0%	0%	14%
BE	Belgium	0	0	9	0%	0%	21%
DE	Germany	0	2	101	0%	5%	23%
DK	Denmark	0	0	0	0%	0%	0%
ES	Spain*	2	6	21	33%	35%	42%
FI	Finland	0	0	0	0%	0%	0%
FR	France*	0	0	1	0%	0%	1%
GR	Greece	2	8	42	50%	62%	82%
IE	Ireland	0	0	0	0%	0%	0%
IT	Italia	3	4	17	27%	20%	17%
LU	Luxembourg	0	0	0	0%	0%	0%
NL	Netherlands	0	0	0	0%	0%	0%
PT	Portugal*	0	4	24	0%	80%	86%
SE	Sweden	0	0	0	0%	0%	0%
UK	United Kingdom	0	1	22	0%	3%	17%
New Member States		**8**	**36**	**112**	**80%**	**88%**	**93%**
CY	Cyprus	0	0	0	0%	0%	0%
CZ	Czech Rep.	1	7	13	100%	88%	93%
EE	Estonia	1	1	5	100%	100%	100%
HU	Hungary	1	6	19	100%	86%	95%
LT	Latvia	1	1	10	100%	100%	100%
LV	Latvia	1	1	5	100%	100%	100%
MT	Malta	1	1	2	100%	100%	100%
PL	Poland	1	16	42	100%	100%	95%
SI	Slovenia	0	0	9	0%	0%	75%
SK	Slovakia	1	3	7	100%	75%	88%
Candidate countries		**2**	**14**	**70**	**100%**	**100%**	**100%**
BG	Bulgaria	1	6	28	100%	100%	100%
RO	Romania	1	8	42	100%	100%	100%
Other Espon		**0**	**0**	**0**	**0%**	**0%**	**0%**
CH	Switzerland	0	0	0	0%	0%	0%
NO	Norway	0	0	0	0%	0%	0%
UE25		**15**	**61**	**354**	**18%**	**25%**	**29%**
UE25+2		**17**	**75**	**424**	**20%**	**29%**	**33%**
ESPON 29		**17**	**75**	**424**	**19%**	**27%**	**32%**

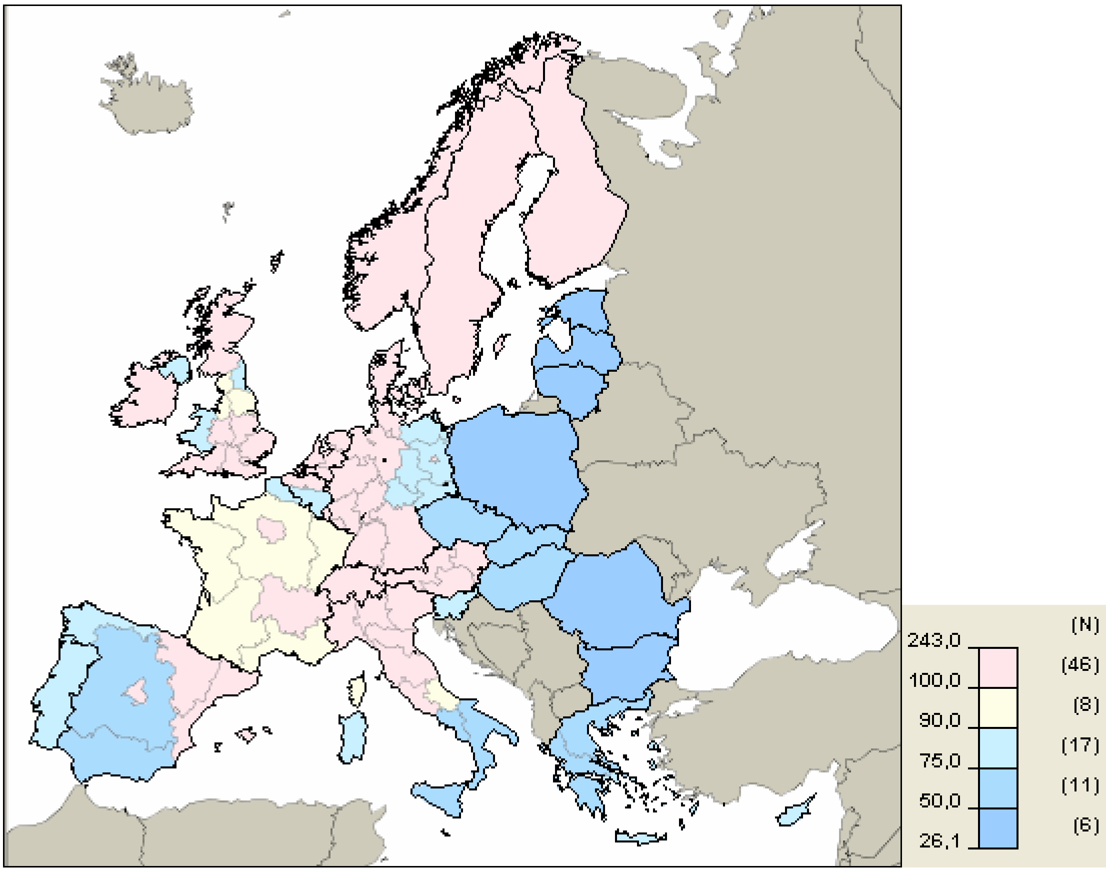
* without peripheral regions

Table 3 : Population located in territorial units with GDP/inh. PPS lower than 75% of EU25 average in 1999

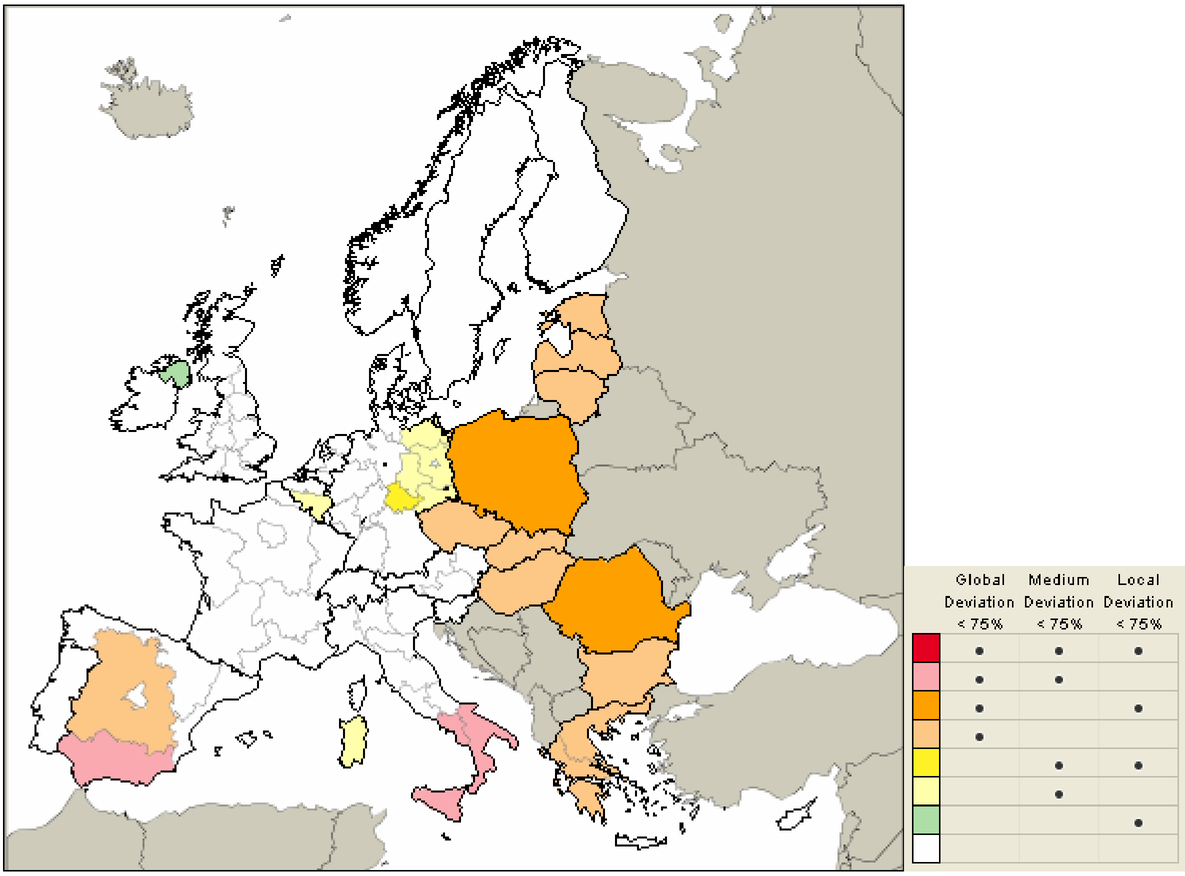
State Code	Name	Units with GDP/inh < 75			% of units eligible to Obj. 1		
		Nuts1	Nuts2	Nuts3	Nuts1	Nuts2	Nuts3
Old Member States		**37388**	**43947**	**59459**	**10%**	**12%**	**16%**
AT	Austria	0	0	732	0%	0%	9%
BE	Belgium	0	0	1105	0%	0%	11%
DE	Germany	0	2203	12933	0%	3%	16%
DK	Denmark	0	0	0	0%	0%	0%
ES	Spain*	13690	13921	12721	36%	37%	33%
FI	Finland	0	0	0	0%	0%	0%
FR	France*	0	0	124	0%	0%	0%
GR	Greece	6067	4043	5173	58%	38%	49%
IE	Ireland	0	0	0	0%	0%	0%
IT	Italia	17630	17023	12120	31%	30%	21%
LU	Luxembourg	0	0	0	0%	0%	0%
NL	Netherlands	0	0	0	0%	0%	0%
PT	Portugal*	0	6262	6230	0%	65%	64%
SE	Sweden	0	0	0	0%	0%	0%
UK	United Kingdom	0	495	8321	0%	1%	14%
New Member States		**72351**	**67694**	**67802**	**96%**	**90%**	**90%**
CY	Cyprus	0	0	0	0%	0%	0%
CZ	Czech Rep.	10283	9093	9093	100%	88%	88%
EE	Estonia	1441	1441	1441	100%	100%	100%
HU	Hungary	10066	7216	8241	100%	72%	82%
LT	Latvia	3699	3699	3699	100%	100%	100%
LV	Latvia	2432	2432	2432	100%	100%	100%
MT	Malta	380	380	380	100%	100%	100%
PL	Poland	38655	38655	36462	100%	100%	94%
SI	Slovenia	0	0	1276	0%	0%	64%
SK	Slovakia	5395	4778	4778	100%	89%	89%
Candidate countries		**30666**	**30666**	**30666**	**100%**	**100%**	**100%**
BG	Bulgaria	8209	8209	8209	100%	100%	100%
RO	Romania	22457	22457	22457	100%	100%	100%
Other Espon		**0**	**0**	**0**	**0%**	**0%**	**0%**
CH	Switzerland	0	0	0	0%	0%	0%
NO	Norway	0	0	0	0%	0%	0%
UE25		**109739**	**111642**	**127261**	**24%**	**25%**	**28%**
UE25+2		**140405**	**142308**	**157927**	**29%**	**30%**	**33%**
ESPON 29		**140405**	**142308**	**157927**	**29%**	**29%**	**32%**

* without peripheral regions

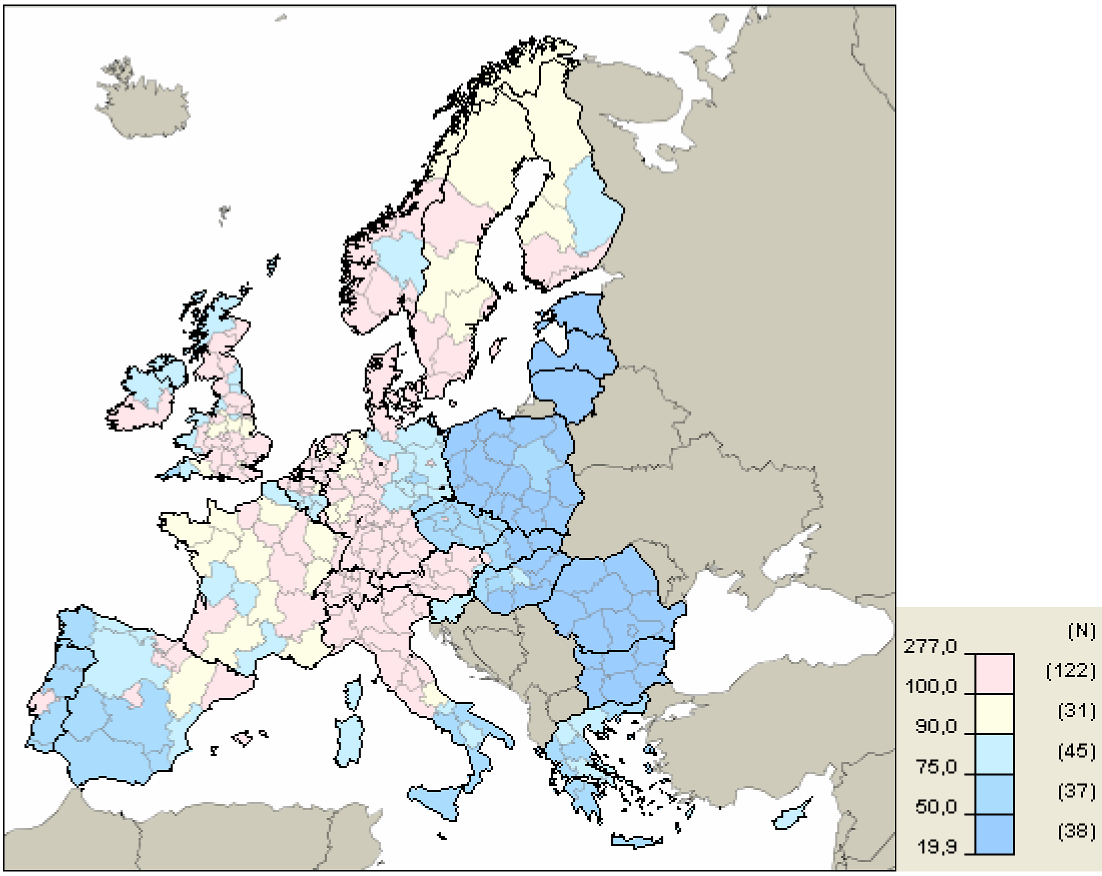
Map 1 : GDP/inh. PPS in 1999 at NUTS 1 level (index 100 = UE25)



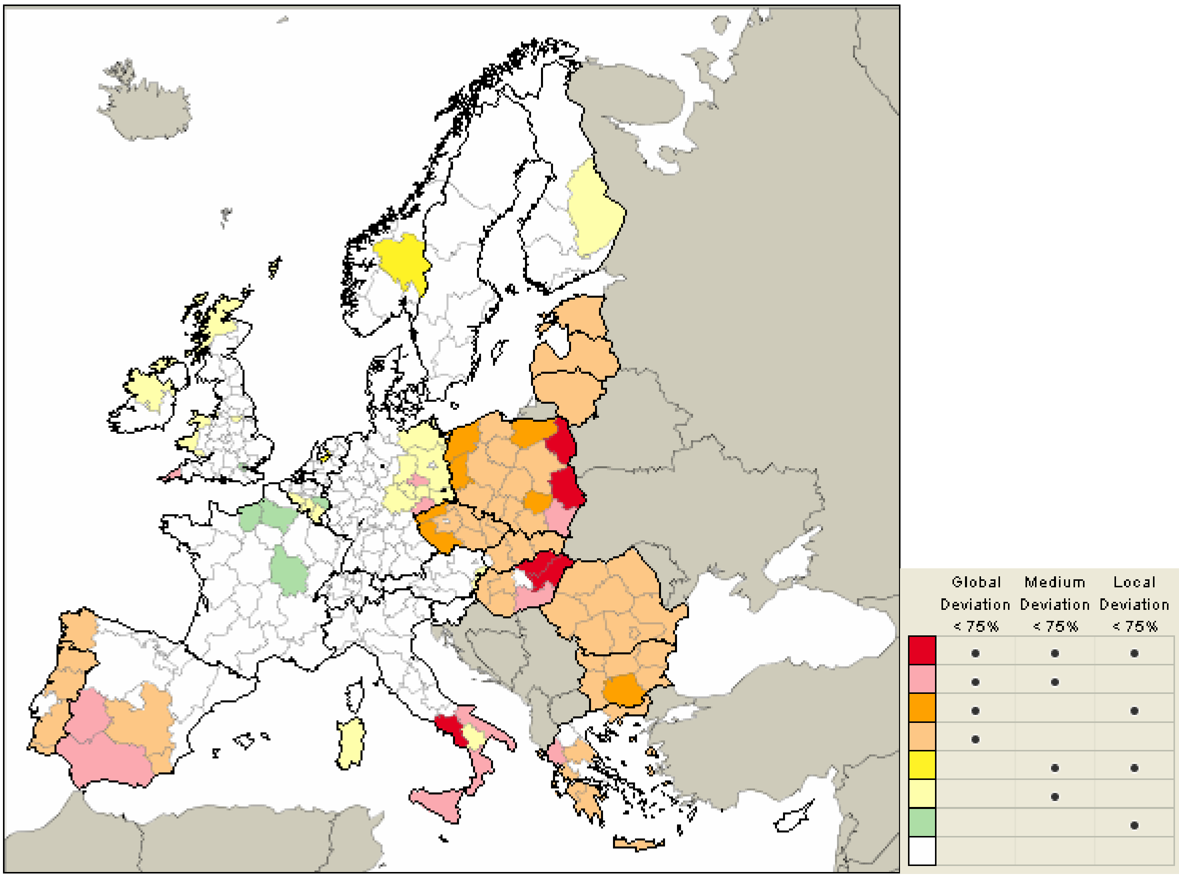
Map 2 : Multiscalar territorial analysis of GDP/inh. PPS 1999 at NUTS 1 level



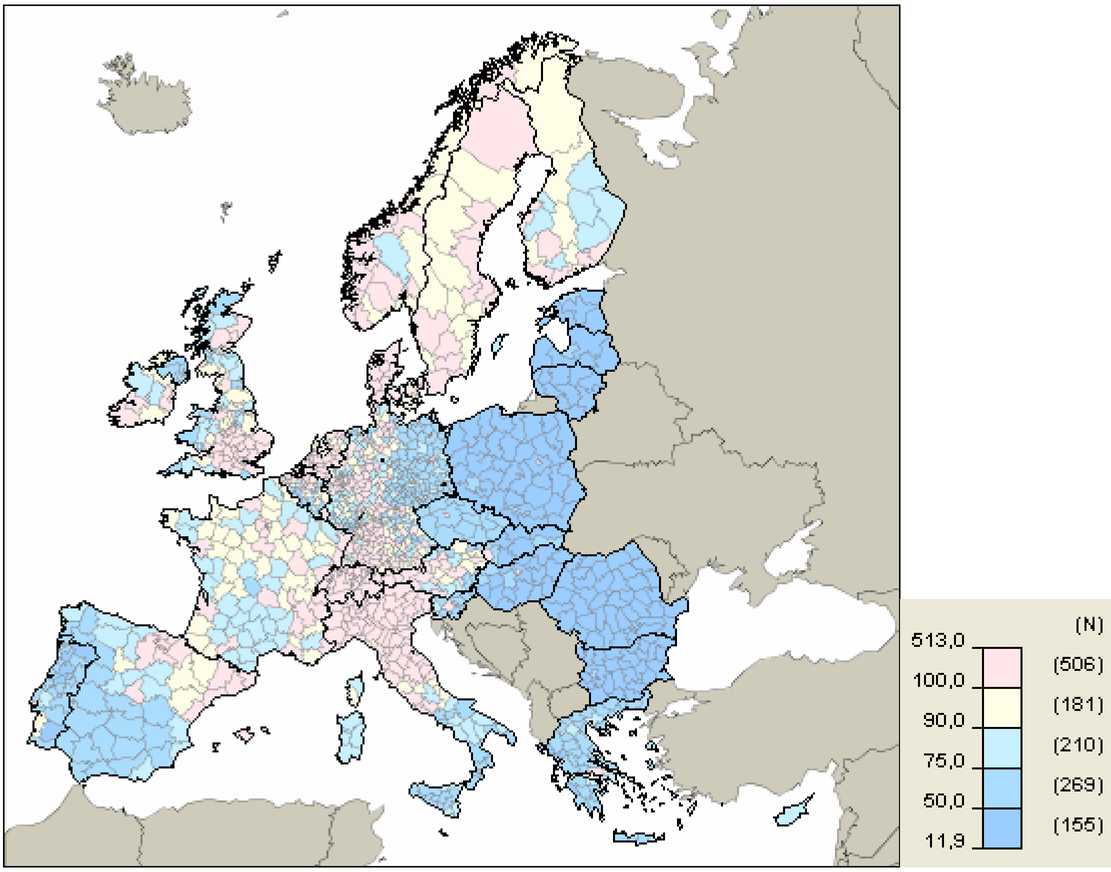
Map 3 : GDP/inh. PPS in 1999 at NUTS 2 level (index 100 = UE25)



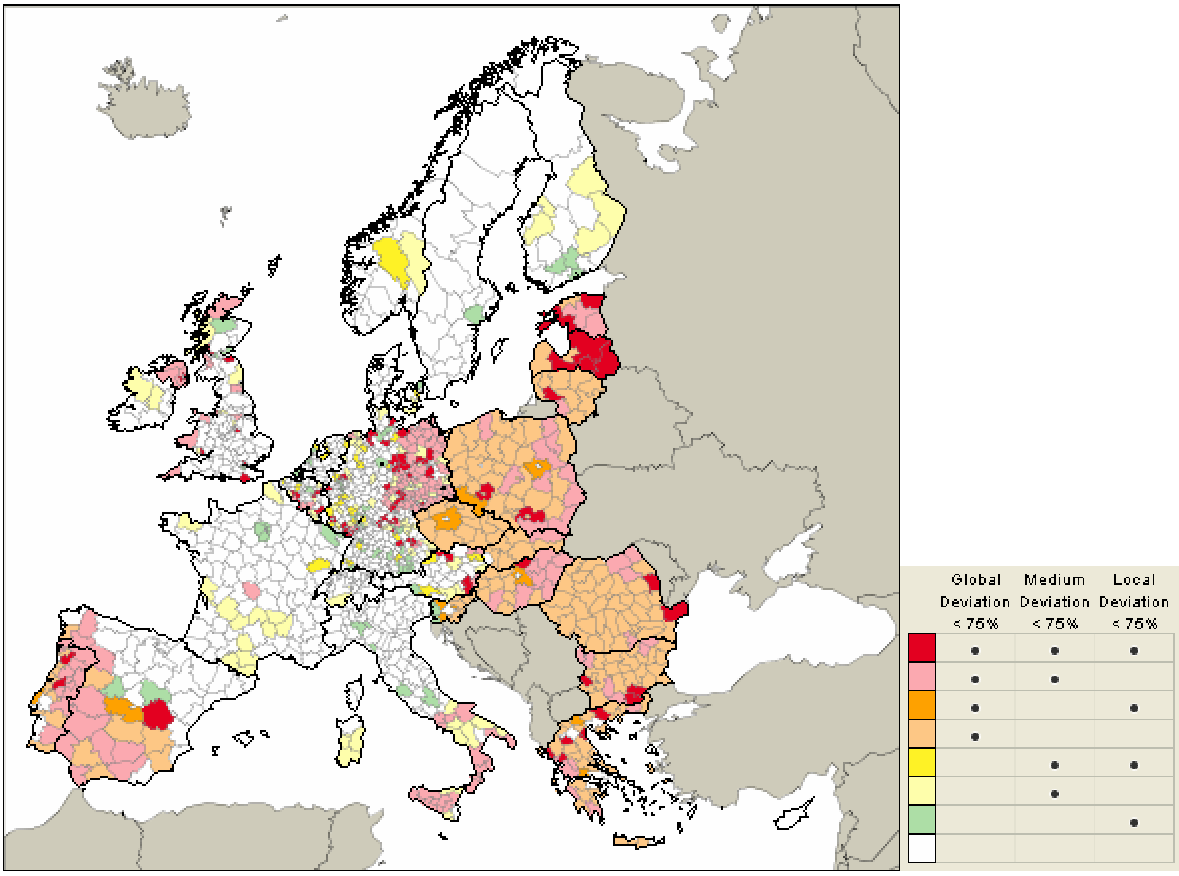
Map 4 : Multiscalar territorial analysis of GDP/inh. PPS 1999 at NUTS 2 level



Map 5 : GDP/inh. PPS in 1999 at NUTS 3 level (index 100 =UE25)

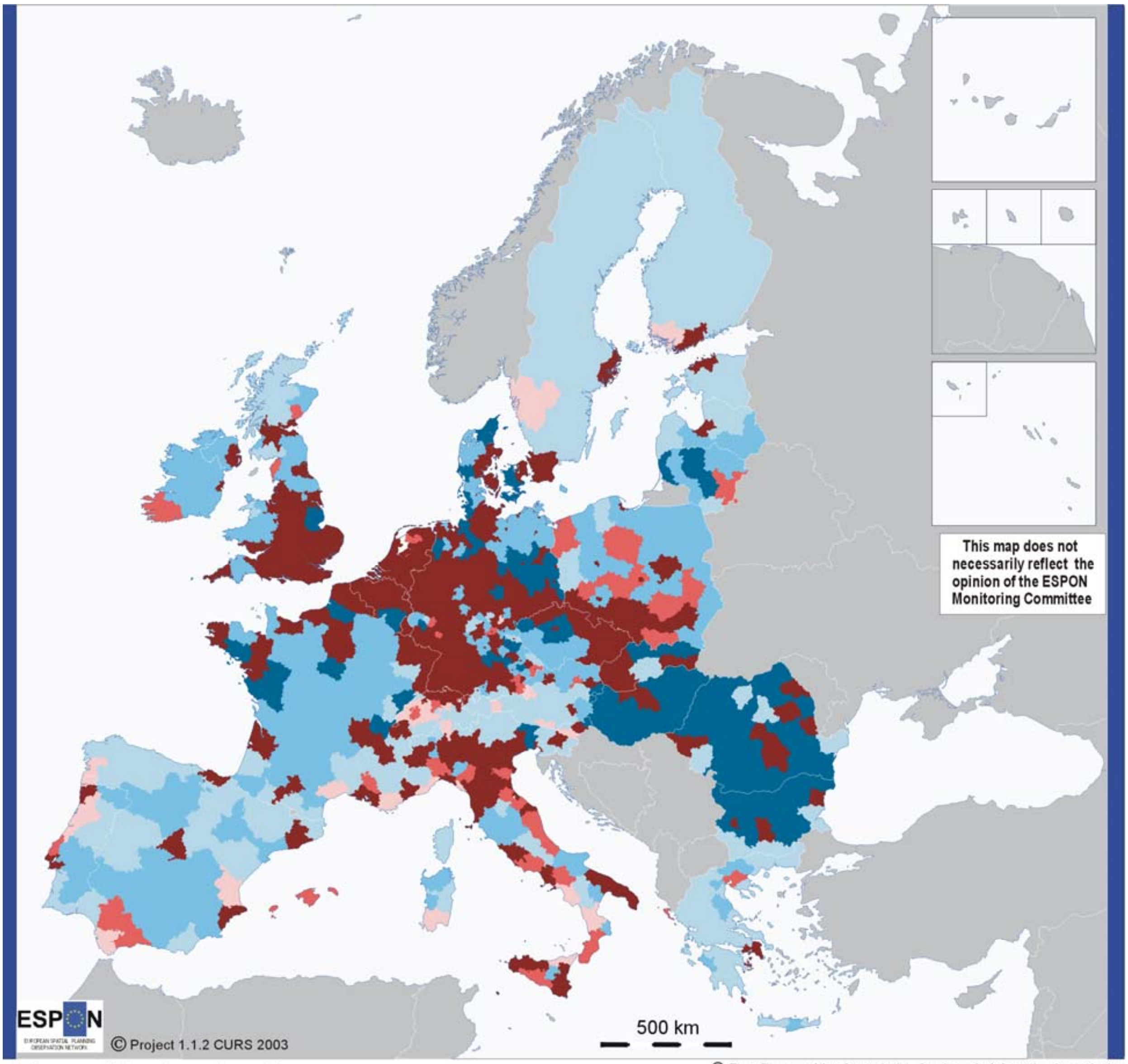


Map 6 : Multiscalar territorial analysis of GDP/inh. PPS 1999 at NUTS 3 level



Annex 2: the urban-rural typology extracted from the ESPON project 1.1.2

From *fr-1.1.2_part1_revised_31-03-05.pdf*



Urban-rural typology, based on population density, FUA ranking and land cover

- High urban influence, high human intervention
- High urban influence, medium human intervention
- High urban influence, low human intervention
- Low urban influence, high human intervention
- Low urban influence, medium human intervention
- Low urban influence, low human intervention

Ranking of Functional Urban Areas (FUAs):
Origin of data: EUROSTAT, National Statistical Offices, National experts
Source: Nordregio, ESPON Data Base

Population density:
Origin of data: EU15 and CC's: Eurostat
Norway and Switzerland: National Statistical Offices
Time reference:1999

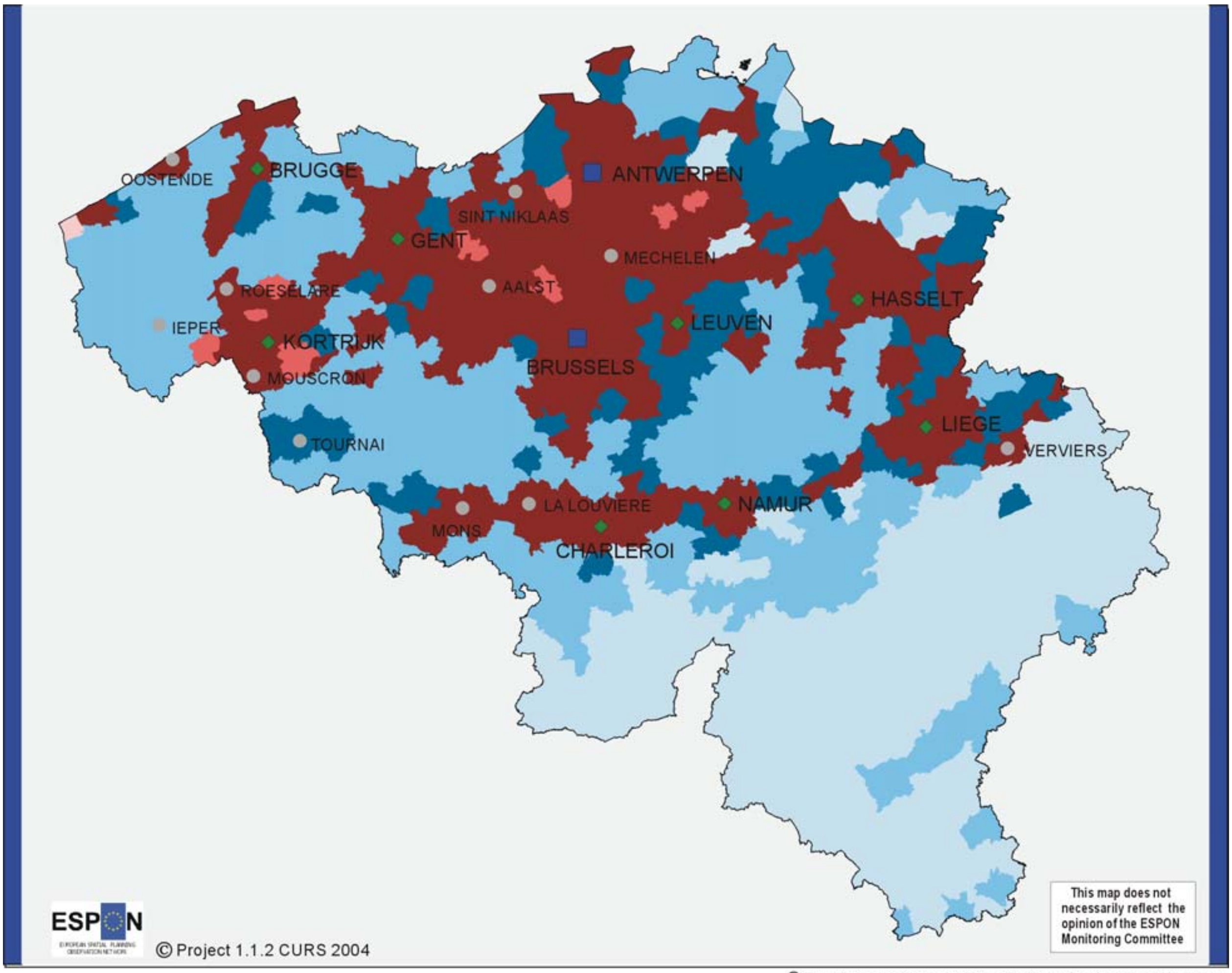
Land cover types:
Origin of data: EEA, Corine Land Cover 90

Source: ESPON Data Base

The criteria for urban influence:
- Population density above the average (107 inhabitants/km^2 in EU25+4)
- And/or at least a European level functional urban area (based on typology made by ESPON Action 1.1.1)

Degree of human intervention is estimated through the average shares of land covers (in EU23+3, no data on Cyprus, Malta and Norway):
High human intervention: at least the share of artificial surfaces above average (3,48%)
Medium human intervention: at least the share of agricultural land above average (50,36%)
Low human intervention: only the share of residual land use above average (46,16%)

Map 1. Urban–rural typology



© EuroGeographics Association for the administrative boundaries

Urban-rural typology in Belgium at NUTS5-level based on national averages

- High urban influence, high human intervention (239)
- High urban influence, medium human intervention (9)
- High urban influence, low human intervention (1)
- Low urban influence, high human intervention (77)
- Low urban influence, medium human intervention (179)
- Low urban influence, low human intervention (84)

Land cover:
Origin of data: Corine Land Cover 90

Population:
Origin of data: National Statistical Office
Time reference:2000

Source: CURS

Ranking of Functional Urban Areas (FUAs):
Origin of data: EUROSTAT, National Statistical Offices, National experts
Source: Nordregio, ESPON Data Base

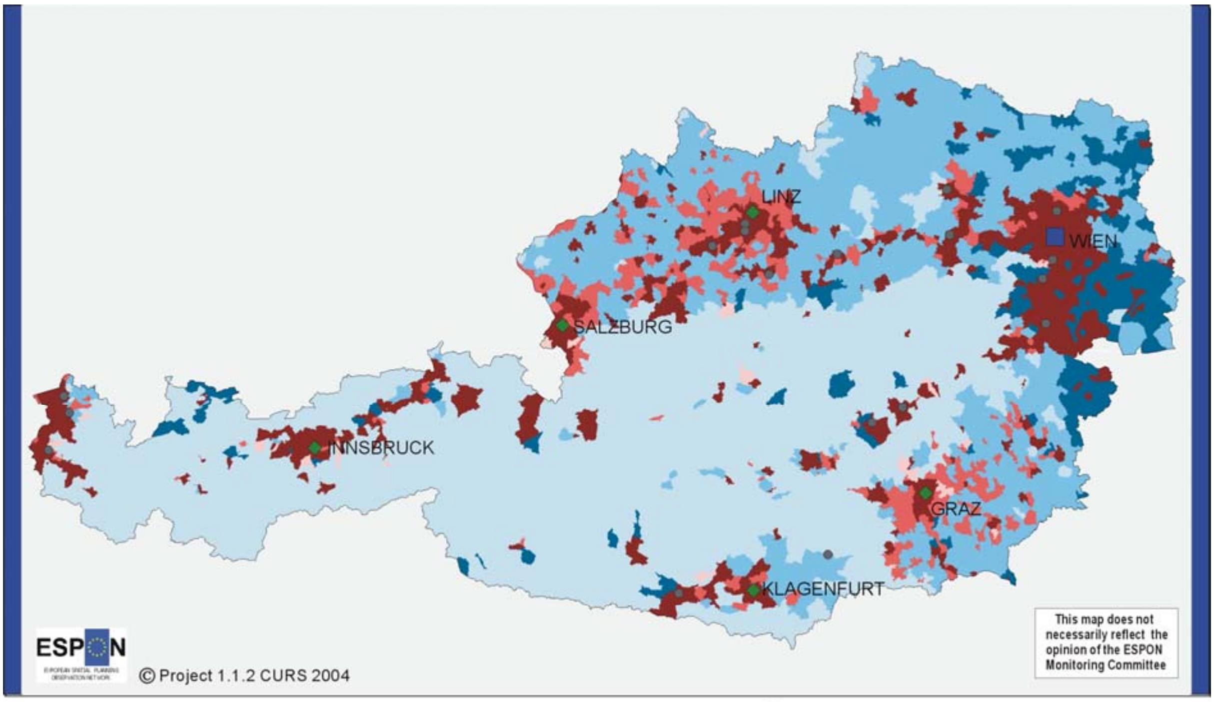
Typology of Funtional urban areas (from ESPON Action 1.1.1)

- European/Global
- National/Transnational
- Local/Regional

The criteria for urban influence:
- Population density above the average (336 inhabitants/km^2 in Belgium)
- And/or at least a European level functional urban area (based on typology made by ESPON Action 1.1.1)

Degree of human intervention is estimated through the average shares of land covers (in Belgium):
High human intervention: at least the share of artificial surfaces above average (19,21%)
Medium human intervention: at least the share of agricultural land above average (58,64%)
Low human intervention: only the share of residual land use above average (28,19%)

Map 2. Urban–rural typology in Belgium at NUTS 5 level, based on national averages



Urban-rural typology in Austria at NUTS5 level, based on national averages

© EuroGeographics Association for the administrative boundaries

High urban influence, high human intervention	(438)
High urban influence, medium human intervention	(294)
High urban influence, low human intervention	(28)
Low urban influence, high human intervention	(163)
Low urban influence, medium human intervention	(769)
Low urban influence, low human intervention	(666)

Typology of Functional urban areas (from ESPON Action 1.1.1):

- European/Global
- National/Transnational
- Local/Regional

Land Cover:
Origin of data: Corine Land Cover 90
Source: CURS

Population density:
Origin of data: National Statistical Office
Time reference:1999
Source: ÖIR

Ranking of Functional Urban Areas (FUAs):
Origin of data: EUROSTAT, National Statistical Offices, National experts
Source: Nordregio, ESPON Data Base

The criteria for urban influence:
- Population density above the average (96,8 inhabitants/km^2)
- And/or at least a European level functional urban area (based on typology made by ESPON Action 1.1.1)

Degree of human intervention is estimated through the average shares of land covers:
High human intervention: at least the share of artificial surfaces above average (1,76%)
Medium human intervention: at least the share of agricultural land above average (36,34%)
Low human intervention: only the share of residual land use above average (61,9%)

Map 3. Urban–rural typology in Austria at NUTS 5 level, based on national averages

From *fr-1.1.2_part2_revised_31-03-05.pdf*

"... The total territory under scrutiny concerns 29 countries: EU15+2 (Norway, Switzerland), and EU10+2 (Bulgaria, Romania). The total area of the territory under consideration is 4.69 million sq km, and it covers 1293 NUTS3 areas, which is the basic spatial unit of the study. The total population of the 29 countries is 495 million people. The relative of the EU10+2 countries is 21 percent of population and 23 percent of the total territory. The contribution of these countries to the total GDP is 4.8 percent (Table 3.1). In terms of land cover, which to some extent indicates the functional land use as well, the territory is composed of 50.36 percent of agricultural land, 3.48 percent of artificial surfaces, and 46.16 percent of residual land cover.

The task was to elaborate a typology that would define the character of regions on NUTS3 level in Europe according to their urban respective rural characteristics. Such a typology would not as such indicate urban-rural relations in a specified manner, but it would rather mirror the relative degree of urban respective rural features on a successive scale according to which each region (NUTS3) would be classified. In doing so, the typology would, however, indicate structural and functional relations as well, but on a very general level of abstraction.

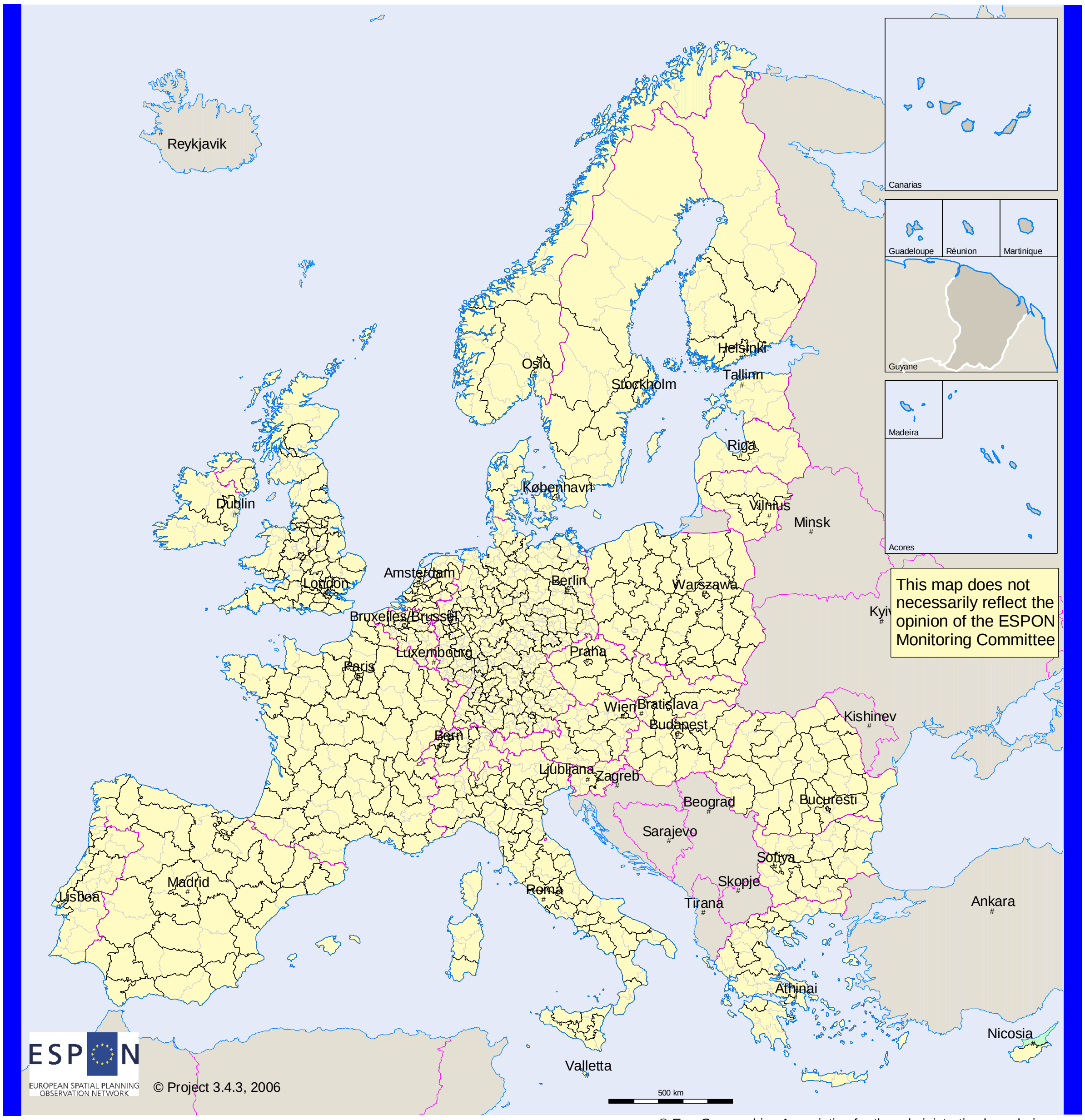
The first task was to identify criteria, which could be applied in order to define the degree of urbanity respective rurality. In order to get an idea of current practices in defining urban and rural, a survey of national definitions was carried out. The national definitions were tested in terms of correspondence with a set of chosen criteria based on national averages. Based on national definitions, a typology of urban and rural population across Europe was elaborated, and a harmonising element of population density was introduced.

The interrelations among a set of indicators were identified based on a factor analysis, and the criteria for defining the urban versus rural character of the NUTS3 regions were chosen. The inter-correlations of the chosen criteria were studied, and each criterion was related to a set of other criteria indicating economic performance, population change, etc. as well. The purpose was to analyse interrelations in a systematic way in order to provide knowledge for a discussion on policy implications.

A typology of 6 classes was elaborated, and scrutinised in terms of population, economic performance, prospects and accessibility to infrastructure and knowledge based on information acquired from other relevant ESPON projects. The typology based on national definitions was compared to the harmonised typology..."

Annex 3: Preliminary attempt of delimitation of new NUTS regions (with same population)

Recomposition of the NUTS division into units of same population (1.000.000 inhab.)



Origin of data: EU 15 and CC's : Eurostat,
Norway and Switzerland : National Statistical Offices.

Base of recompostion : NUTS 3 units

- coast
- National boundary
- # Capital cities
- NUTS 1.000.000
- NUTS 3

Recomposition of the NUTS division into units of same population (2000000 inhab.)

Origin of data: EU 15 and CC's : Eurostat,
Norway and Switzerland : National Statistical Offices.

Base of recompostion : NUTS 3 units

- coast
- National boundary
- # Capital cities
- NUTS 2.000.000
- NUTS 3

Recomposition of the NUTS division into units of same population (5.000.000 inhab.)

© EuroGeographics Association for the administrative boundaries

Origin of data: EU 15 and CC's : Eurostat,
Norway and Switzerland : National Statistical Offices.

Base of recompostion : NUTS 3 units

coast
National boundary
Capital cities
NUTS 2.000.000
NUTS 3