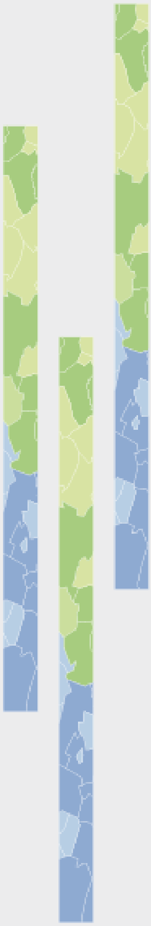
TECHNICAL
REPORT



Local and regional data

MAIN RESULTS

CLC 2006 data integration on LAU2 geometry for selected indicators

Road network density at LAU2 scale`

Theoretical issues related to the local data integration

ESPON 2013 DATABASE

EUROPEAN UNION
Part-financed by the European Regional Development Fund
INVESTING IN YOUR FUTURE

28 PAGES

LIST OF AUTHORS

Octavian Groza, UAIC, CUGUAT-TIGRIS, Iasi, Romania
Alexandru Rusu, UAIC, CUGUAT-TIGRIS, Iasi, Romania

Contact
octaviangroza@yahoo.com

tel. + 40 0232 20 14 88

TABLE OF CONTENT

Introduction

The problems explored in this technical report can be organized by three topics of interest:

- the issues related to the CLC 2006 data integration in the LAU2

- the integration of the road density at local scale of analysis

- problems of data integration from other databases and theoretical models applied at local scale

The construction of indicators derived from the intersection of CLC 2006 datasets with the LAU2 geometry of 2006 is a matter of technical management of the spatial units and of methodology. Not all the CLC 2006 categories were included in our approach, a selection was made for artificial surfaces, arable land, permanent crops, heterogeneous agricultural areas and forests. Excepting three countries where data is not available, all other ESPON states are present in the dataset resulted from the analysis.

In the case of the road density indicator construction, we have applied a simple but efficient methodology that enabled the calculus of road network length for every LAU2 present in the geometry. The technical difficulties forced us to split the road network geometry in hierarchical segments (local roads, regional roads, long distance roads and motorways), a good thing after all because this split allows us building a qualitative classification of the LAU2, a classification based on the intersection between administrative geometry and the hierarchical segments. One country is absent from the analysis - Bulgaria (no proper road network available).

The last part of this report describes other issues linked to the exploration of local data integration. There are three points discussed in this final part - the issue of commuters flow management between the LAU2, the application of theoretical models at local scale (for the moment, only models of spatial interaction) and the issue of the centroids of the LAU2. Two methods were tested in relation to the last issue, methods that allowed us to create alternative centers to the LAU2, using population disaggregated in a grid of 1km.

1 The production of basic indicators using the LAU2 geometries and the CLC 2006 data

The problem of data integration from the CLC 2006 in the LAU2 frame was already explored in the previous ESPON DB 2013 project, the data being collected for selected countries from the Eastern Europe (Czech Republic, Slovakia, Hungary, Romania and Bulgaria). The methodology involved for analyzing the land use in these countries offered new indicators, such as:

- the surface of any CLC 2006 category at LAU2 level
- the share of the categories in the total surface of the LAU2

The basic steps in order to obtain these values are simple operations involving logical tools of spatial analysis (intersections, surface calculations, summarization by LAU2 code and mapping exercises for verification). As the CLC 2006 layers (categories) are amorphous and administratively independent, we need to intersect theme with the LAU2 frame for obtaining two different codes: a code describing the CLC 2006 polygon of origin and a code for the LAU2 geometry. Obviously, the LAU2 geometry and the CLC 2006 layers need to be properly projected so that the calculated surfaces are as accurate as possible. The surfaces of the intersected objects are summarized using the LAU2 code and attached as new fields in the LAU2 geometry. The only method of validation is to make the sum of the CLC 2006 integrated categories and to verify that this sum is equal to the LAU2 surface. With 44 layers in the CLC 2006 dataset and with few options for automatization, obtaining the indicators implies a large quantity of time. By experience, using a model builder for repetitive steps (intersection and projection) can be used with reasonable geometries that contain less than 20 000 objects. Anyway, even with a functional model builder, the final calculation steps still involve a layer by layer approach. In this case is wiser to execute all the processes for a single country or for a limited number of countries.

To resume, these steps can be synthesized as follows:

1) chose of a LAU2 geometry. As we use CLC 2006 data, we also used a 2006 polygon layer for the LAU2 (extracted from the GISCO database - COMM_RG_2006 with attributes)
2) project the COMM_RG_2006 in a projection appropriate for surface calculation
3) in the same logic, project all the CLC 2006 layers
4) intersect the LAU2 geometry with a layer from CLC 2006 (e.g. Artificial surfaces, "Artificial, non-agricultural vegetated areas", Green urban areas -141 category)
5) calculate the surfaces of the intersected objects. We have made an option for square meters as unit of measure.
6) summarize the surfaces of the intersected objects by LAU2 code and export the output
7) join the output to the COMM_RG_2006 (or to the group of selected and extracted LAU2 objects, when working for a country or a group of countries)
8) map the indicator and verify that its share in the LAU2 is not bigger than 1.
9) verify that the results are reliable (if all the steps were properly implemented, you should not find olive groves in Finland).

Before integration the CLC 2006 for all the LAU2 in the ESPON space, we have tested the methodology on three countries from Western and Southern Europe: Spain, Portugal and France. There are two lessons to be learned from this experience. Some of the indicators are so spatially concentrated (probably due to ecological conditions and agricultural structures) that they cannot be fully exploited. It should be the case, for the CLC 222 category in Spain and Portugal. Aggregating this indicator in a superior class will provide more information about the land use territorial patterns. On the other hand, the share itself (ratio between CLC 222 category and the LAU2 surface) is a limited indicator that might not always suggest the concentration or the localization trends. Considering a new indicator that will take into account the problem of geographical could be more useful, in this context.

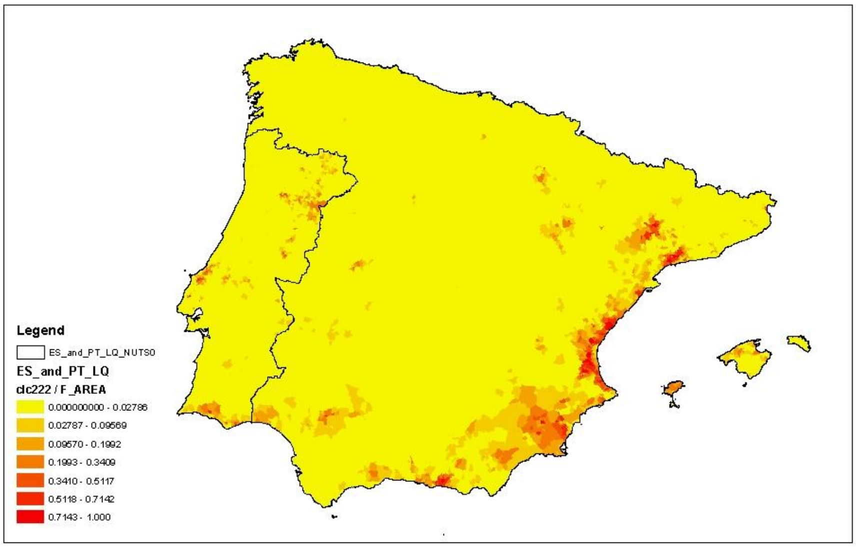


Fig 1 Draft/working map for Spain and Portugal CLC 2006 data integration at LAU2 scale - Permanent crops, Fruit trees and berry plantations.

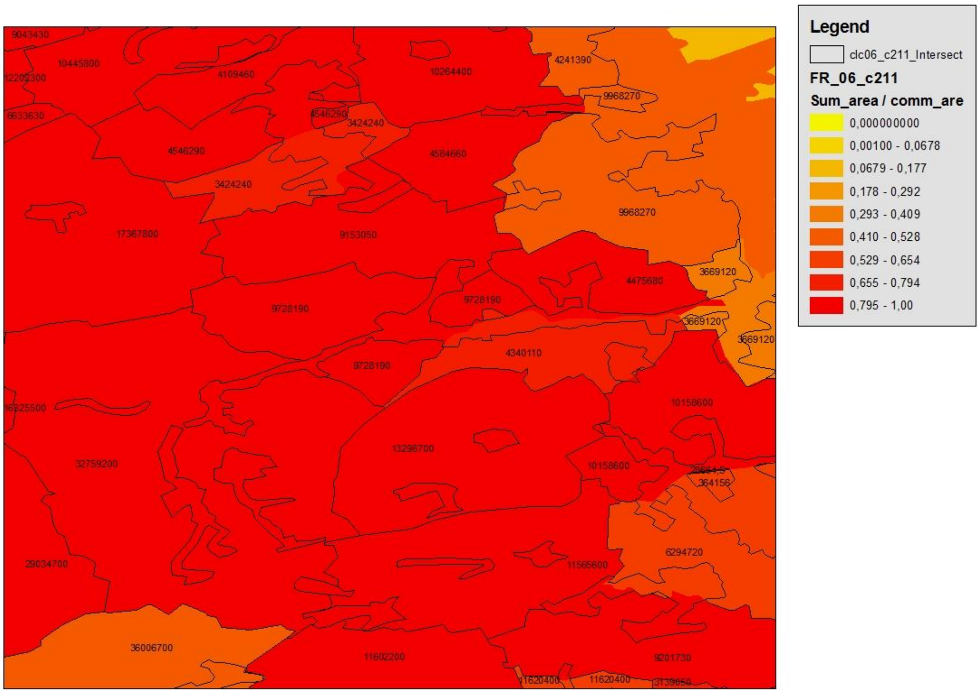


Fig 2 Draft/working map for France CLC 2006 data integration at LAU2 scale - Arable land, Non-irrigated arable land. The map shows the overlay between the LAU2 frame and the intersected CLC 2006 category.

In the methdological and conceptual arsenal of geography there are a lot of tools able to describe the concentration of spatial phenomena or distributions, some of them being inspired by other disciplines (more often economy). There are also some theoretical debates regarding the proper use of these tools, especially when one makes confusion between a concentration index with an indicators of equirepartition (Hoover or Gini). As the ratios are somehow tricky (percentage only refers to the local context), an indicator that will put in relation the relative share of one CLC 2006 category with a macro-spatial context (ESPON space, NUTS0, NUTS2 or NUTS3) will be more efficient.

In the frame of the economic base theory, planners, geographers and practionners developed several methods in order to evaluate the concentration (or the lack of concentration) of employement by economic branch, assuming that this trend will have an impact on the economic perfromance of regions. One of these methods is to calculate the location quotient of employees in the regional economy. The formalization is simple and can be easily implemented in any geographical software or calculus table :

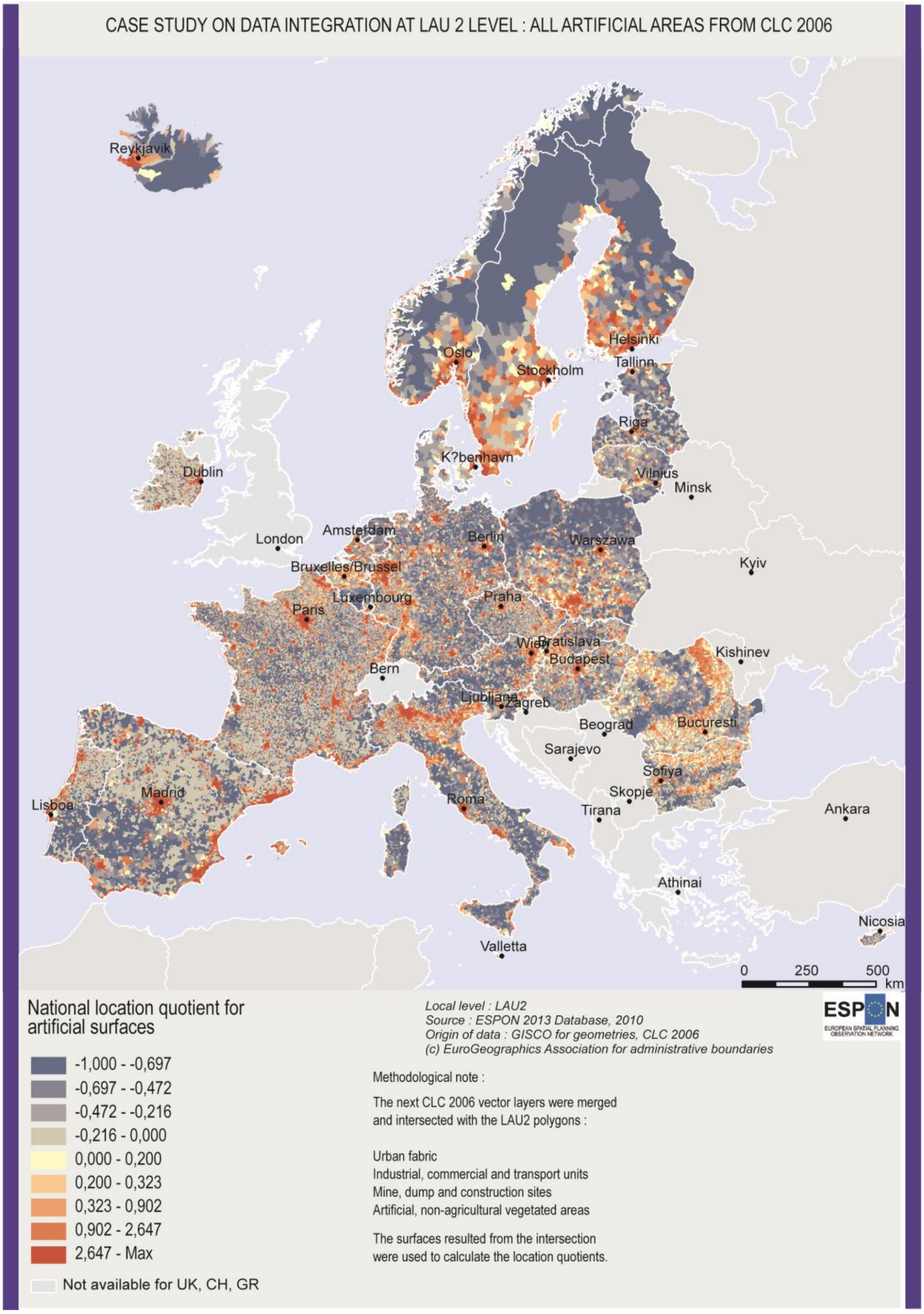
LQ = (R.E.i /R.T.E.) / (N.E.i / N.T.E.)

R.E.i = regional employment in economic branch i (manufacturing, for example)
R.T.E. = regional total employment
N.E.i = national employment in branch i
N.T.E = national employment

The reference value of this ration will be 1. When the value 1 (or very close to 1) appears, we can assume that there is no local pattern of concentration. If the value is larger than 1, we deal with local concentration because the local share is larger than the national one. When the value is less than 1 (0.33 for example) we have a ratio inferior to the national one and we deal with a relative abscence/lack of concentration.

In this logic, we can use the CLC 2006 categories integrated in the LAU2 frame to measure concentration trends for land use. Replacing the employment with different CLC 2006 categories, the national employement with the national surfaces give us a more accurate measure of spatial patterns of land use. Substracting 1 from the result replaces the limits of the indicator to -1 for complete abscence, 0 for no concentration trends (similar to the national share) and positive values - concentration of one CLC 2006 category.

Two suplimentary steps are compulsory in order to calculate the location quotient: summarizing the national and the ESPON space CLC 2006 surfaces and the effective implementation of the formula. A double join of tables to the initial files is needed. All these steps and methodological points create huge table files, difficult to manage, if we take into account the large number of spatial units involved. Different tests on different hardware platforms were performed, hoping that an optimization of the calculus will be found. For the moment, the results of these tests are not very encouraging.



Local level : LAU2
Source : ESPON 2013 Database, 2010
Origin of data : GISCO for geometries, CLC 2006
(c) EuroGeographics Association for administrative boundaries

Methodological note :

The next CLC 2006 vector layers were merged and intersected with the LAU2 polygons :

Urban fabric
Industrial, commercial and transport units
Mine, dump and construction sites
Artificial, non-agricultural vegetated areas

The surfaces resulted from the intersection were used to calculate the location quotients.

Fig 3 Location quotient of the artificial surfaces in 2006 - national reference

The first test and the first integration of data concern the artificial surfaces. Excepting three countries (UK, CH, GR), all the other ESPON states were included in the analysis. The next layers were merged in a single spatial reference and intersected with the LAU2 frame :

111, Artificial surfaces, Urban fabric, Continuous urban fabric
112, Artificial surfaces, Urban fabric, Discontinuous urban fabric
121, Artificial surfaces, "Industrial, commercial and transport units", Industrial or commercial units
122, Artificial surfaces, "Industrial, commercial and transport units", Road and rail networks and associated land

123, Artificial surfaces, "Industrial, commercial and transport units", Port areas
124, Artificial surfaces, "Industrial, commercial and transport units", Airports
131, Artificial surfaces, "Mine, dump and construction sites", Mineral extraction sites
132, Artificial surfaces, "Mine, dump and construction sites", Dump sites
133, Artificial surfaces, "Mine, dump and construction sites", Construction sites
141, Artificial surfaces, "Artificial, non-agricultural vegetated areas", Green urban areas
142, Artificial surfaces, "Artificial, non-agricultural vegetated areas", Sport and leisure facilities

Two indicators were created for this category : a location quotient at national scale and a location quotient for the ESPON space. The two of them are positevely strongly correlated (r = 0.83 and R2 = 0.69). The map shows a classic pattern of the European space with the metropolized core highly artificialized spaces (the Pentagon) and with the regions where the relative abscence of artificial surfaces is very prononced (Northern Poland, Central and Northern parts of the Scandinavian countries or areas in Spain, Portugal and Italy). This spatial repartition is partially (and arguably) explained by the natural and ecological features of the European territory. Internal (national) logics of planning would explain to a certain extent the distribution of the values : rural migration, diffusion of economic practices, volontarist interventions in some states - Hungary or Romania.

The second map is based on the integration of agricultural data - the arable areas. Much more dependent on the natural and ecological constraints, the arable land concentration is a sensitive subject for agricultural policies. It also indicates how rural territories functions or if they are monospecialized. In this case, the map has a double interest - allowing comparations at national scale and between states. Three layers from CLC 2006 were merged and integrated in order to produce this map:

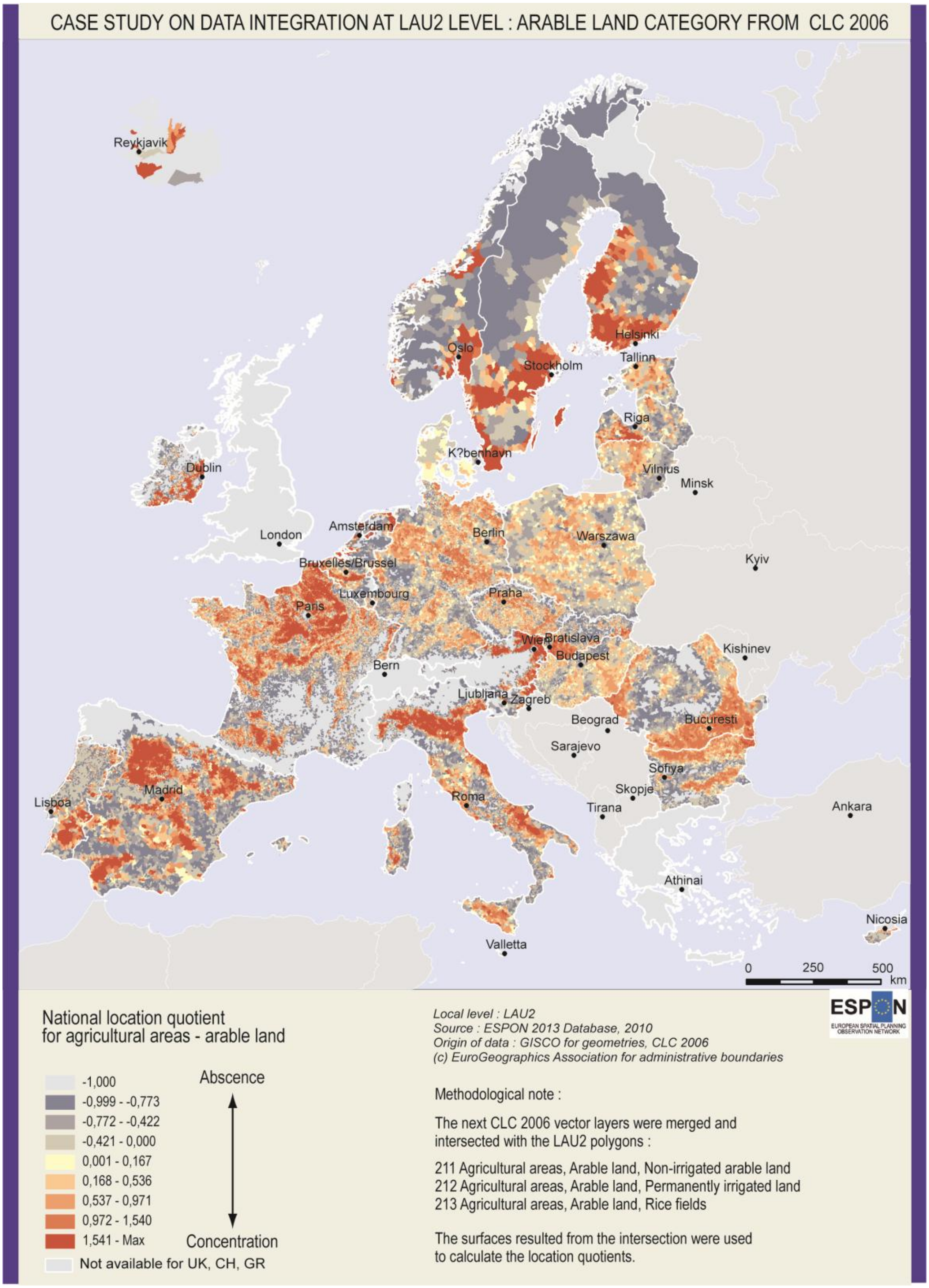
211, Agricultural areas, Arable land, Non-irrigated arable land
212, Agricultural areas, Arable land, Permanently irrigated land
213, Agricultural areas, Arable land, Rice fields

The third map deals with the problem of the agricultural heterogeneous areas, emphasiying European regions with possibly fragmented landscapes (Western France, Northern Spain or Central Transylvania in Romania). The next layers were merged in a single spatial reference and intersected with the LAU2 frame:

241, Agricultural areas, Heterogeneous agricultural areas, Annual crops associated with permanent crops
242, Agricultural areas, Heterogeneous agricultural areas, Complex cultivation patterns
243, Agricultural areas, Heterogeneous agricultural areas, "Land principally occupied by agriculture, with significant areas of natural vegetation
244, Agricultural areas, Heterogeneous agricultural areas, Agro-forestry areas

The two maps are partially completing each other in certain regions like the North-West of Spain or in Latvia and Lithuania, where the border regions present a completely different pattern of spatial organization regarding the agricultural areas. One notable association appears between the distribution of the heterogeneous agricultural land and the low-mountain areas - the

Subcarpathian regions in Poland and Romania, the Central Massif in France present positive values of the indicator. One solution to better observe these relations is to cross the values of the location quotient with different regional frames/geometries, using the OLAP cube. The results could be used in order to refine typologies of land occupation at different scales of analysis. It could also be usefull in order to better seize the geographical specificities of regions.


CASE STUDY ON DATA INTEGRATION AT LAU2 LEVEL : ARABLE LAND CATEGORY FROM CLC 2006

Local level : LAU2
Source : ESPON 2013 Database, 2010
Origin of data : GISCO for geometries, CLC 2006
(c) EuroGeographics Association for administrative boundaries

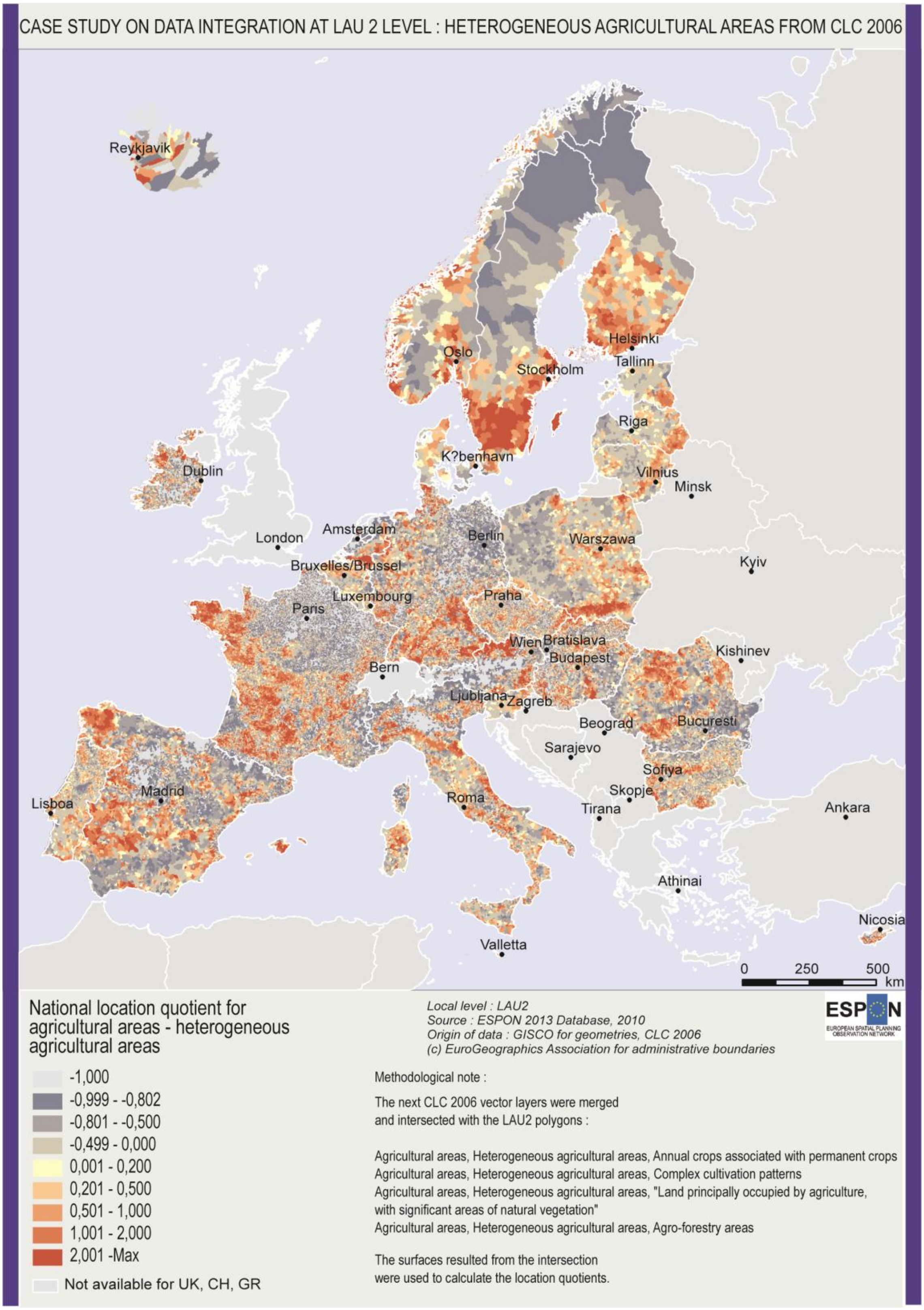
Methodological note :

The next CLC 2006 vector layers were merged and intersected with the LAU2 polygons :

211 Agricultural areas, Arable land, Non-irrigated arable land
212 Agricultural areas, Arable land, Permanently irrigated land
213 Agricultural areas, Arable land, Rice fields

The surfaces resulted from the intersection were used to calculate the location quotients.

Fig 4 Location quotient of the arable land in 2006 - national reference



CASE STUDY ON DATA INTEGRATION AT LAU 2 LEVEL : HETEROGENEOUS AGRICULTURAL AREAS FROM CLC 2006

Local level : LAU2
Source : ESPON 2013 Database, 2010
Origin of data : GISCO for geometries, CLC 2006
(c) EuroGeographics Association for administrative boundaries

Methodological note :

The next CLC 2006 vector layers were merged and intersected with the LAU2 polygons :

Agricultural areas, Heterogeneous agricultural areas, Annual crops associated with permanent crops
Agricultural areas, Heterogeneous agricultural areas, Complex cultivation patterns
Agricultural areas, Heterogeneous agricultural areas, "Land principally occupied by agriculture, with significant areas of natural vegetation"
Agricultural areas, Heterogeneous agricultural areas, Agro-forestry areas

The surfaces resulted from the intersection were used to calculate the location quotients.

Fig 5 Location quotient of the heterogeneous agricultural areas in 2006 - national reference

The issue of the permanent cultures has multiple stakes because their spatial repartition is not only a matter of natural conditions. Territories specialized in permanent cultures are socially structured by traditions in this agricultural practice, at least in theory. This specialization is definitely market oriented, creating economical linkages between territories and functioning as

]

an engine of extrovertion. Some of the regions involved in this analysis are easy to locate and associate with their specialization : Bordeaux, Côte de Rhone, Vallée de la Loire in vineyards, some other in Spain or Italy are more difficult to label. These generally rural spaces should be regarded as regions where the territorial competitiveness functions as a pre-condition for the economic performance, some of the well-known European brands being produced or located (Champagne). The next layers were merged in a single spatial reference and intersected with the LAU2 frame:

221, Agricultural areas, Permanent crops, Vineyards
222, Agricultural areas, Permanent crops, Fruit trees and berry plantations
223, Agricultural areas, Permanent crops, Olive groves

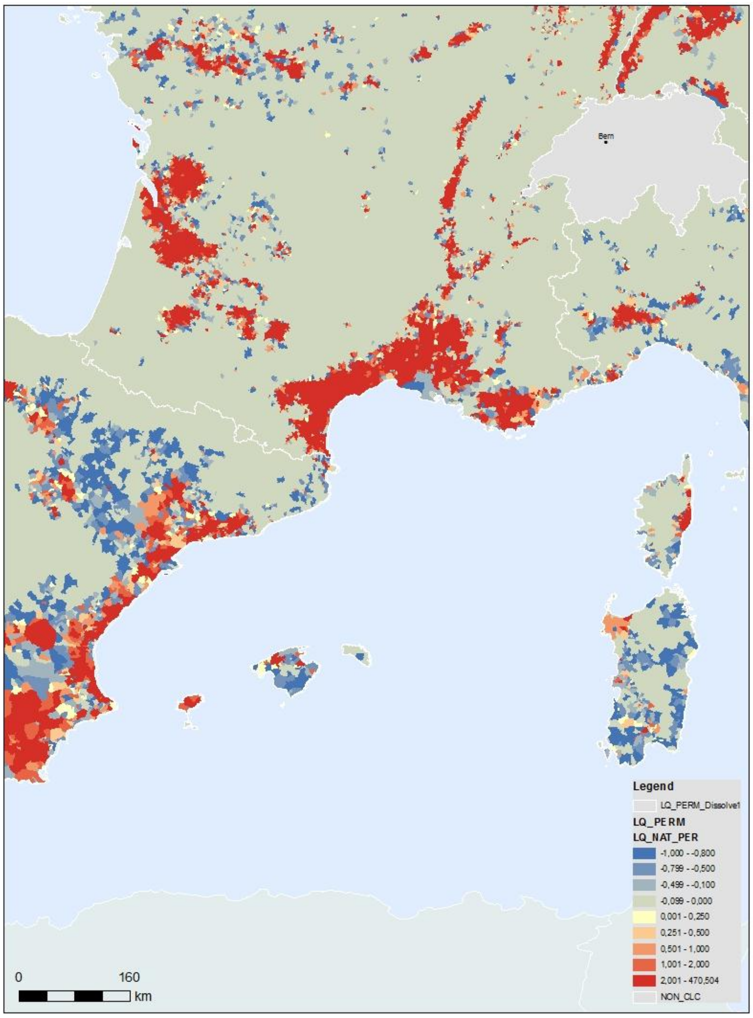
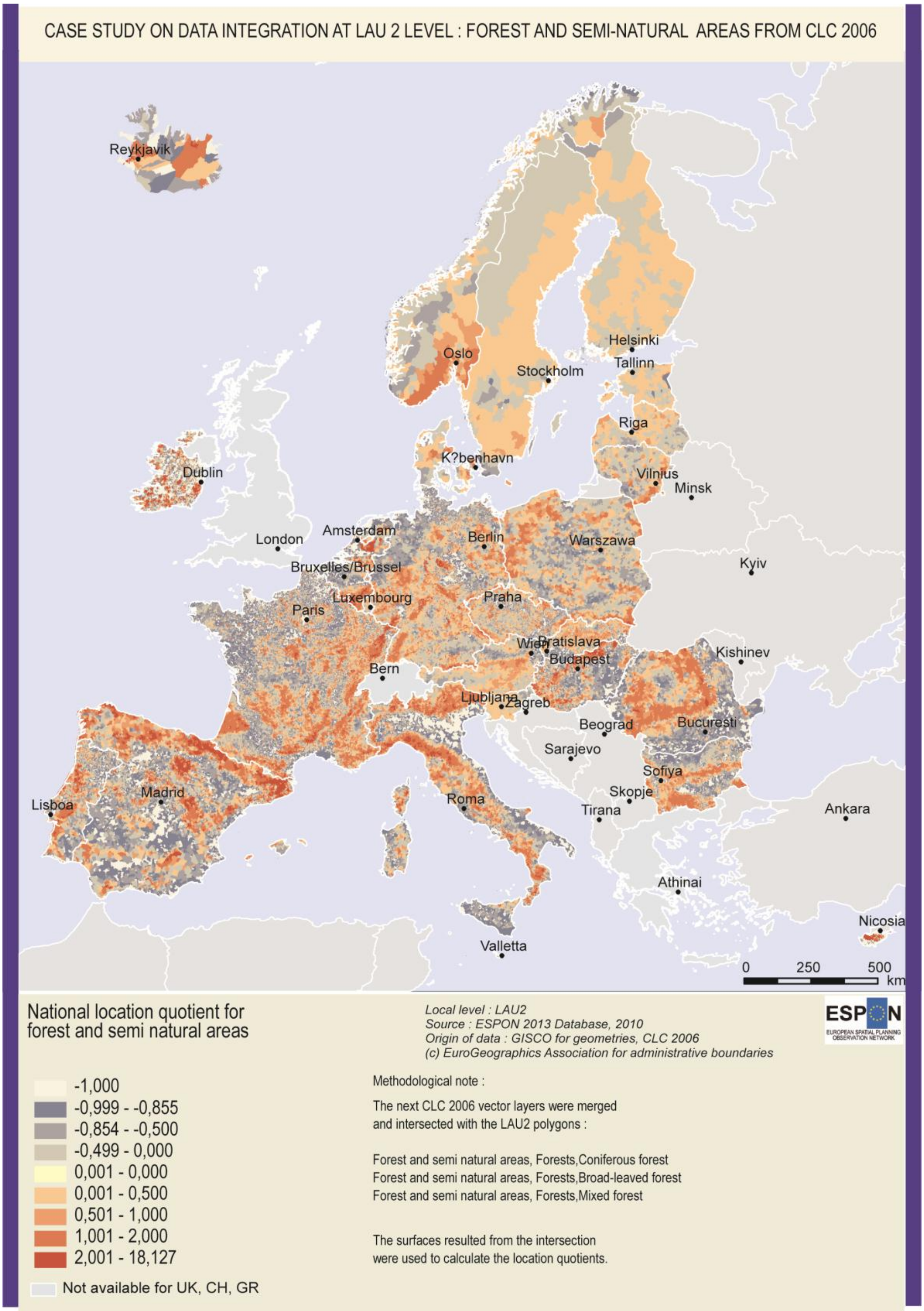


Fig 6 Location quotient of the permanent crops in 2006 - national reference. Draft/working map.

The last map we present as an illustrative output of our methodological approach integrated the forests and the semi-natural areas. Theoretically, this map should closely follow the natural limits or the ecological constraints in the development of these natural areas and be less spectacular. Consequentely, it is no surprise that the map emphasize mountain areas or regions with genuine natural potential. What is more interesting on the map is the repartition of areas that are practically deforested, despite their potential. Some LAU2 situated in Central Germany, in France, in Italy, in Hungary or in Romania are the signs of the deep impact of economic activities in the territory. Better understanding their spatial distribution should be a matter of interest, relevant for policy decision and scientific analysis.



CASE STUDY ON DATA INTEGRATION AT LAU 2 LEVEL : FOREST AND SEMI-NATURAL AREAS FROM CLC 2006

Local level : LAU2
Source : ESPON 2013 Database, 2010
Origin of data : GISCO for geometries, CLC 2006
(c) EuroGeographics Association for administrative boundaries

Methodological note :

The next CLC 2006 vector layers were merged and intersected with the LAU2 polygons :

Forest and semi natural areas, Forests,Coniferous forest
Forest and semi natural areas, Forests,Broad-leaved forest
Forest and semi natural areas, Forests,Mixed forest

The surfaces resulted from the intersection were used to calculate the location quotients.

Fig 7 Location quotient of the forest and semi-natural areas in 2006 - national reference

The next layers were merged in a single spatial reference and intersected with the LAU2 frame:

311, Forest and semi natural areas, Forests, Broad-leaved forest
312, Forest and semi natural areas, Forests, Coniferous forest
313, Forest and semi natural areas, Forests, Mixed forest

The CLC 2006 data integration in the LAU2 frame produced 20 indicators grouped by three major categories : artificial surfaces, agricultural land and forest. The coverage of the ESPON space is almost complete, excepting three countries where data is not available.

Indicators	Total surface	Share of in %	LQ_NAT	LQ_EUR
Artificial surfaces	1	1	2	2
Arable land	1	1	2	2
Permanent crops	1	1	3	3
Heterogeneous agricultural areas	3	1	2	2
Forest and semi natural areas	1	1	2	2

Legend	
1	Easy to calculate
2	Complex to calculate
3	Reserved to interpretation

Tab. 1 Synthetic table - degree of difficulty in the indicators construction

As the construction of the CLC 2006 vector layers is a matter of photo-interpretation, some of the information should be precautiously regarded, especially the category heterogeneous agricultural areas. In the case of the permanent crops, another reference for the location quotient could be considerent as relevant (NUTS2 scale), taking into account the very limited presence of this category in the North of the ESPON Space. The size difference between the LAU2 included in the data integration process is also a problem. The geometry of France is difficult to compare with the LAU2 frame of Poland and may influence the results, especially for the calculated ratios. However, this problem of scale is partially solved by the superior reference layer (national or ESPON Space) in the construction of the location quotients. One question that also deserves to be explored is how mobile in time is the land use and what transfers are realized between these categories. For the moment what we have is almost a complete picture of the ESPON Space land occupation in 2006.

2 The production of basic indicators using the LAU2 geometries and the road network

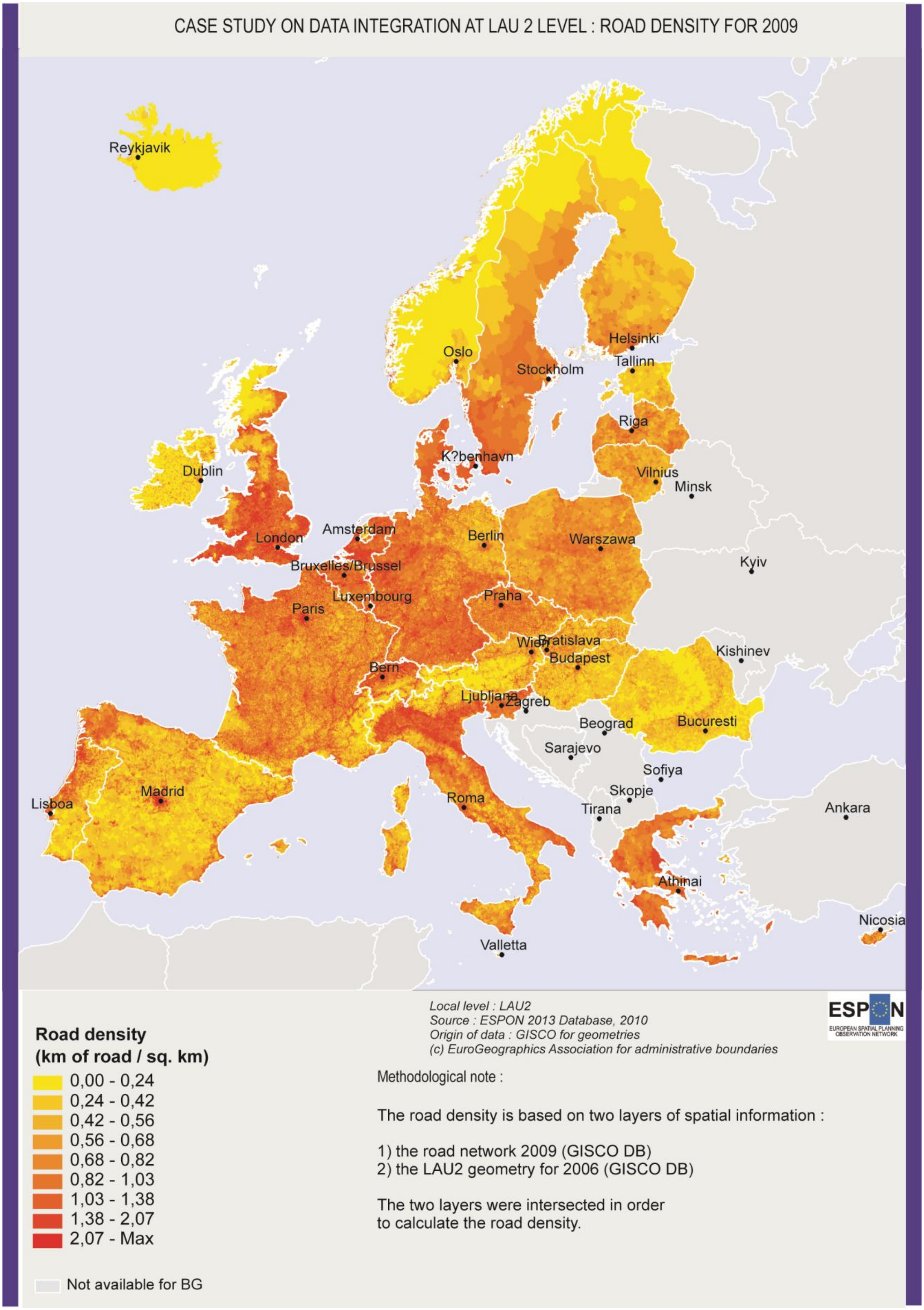
While in the case of the CLC 2006 data integration we dealed with spatial structures of the same type (polygons), obtaining the road density involves an intersection between a network and a geometry which has an effect on the calculation time and the calculation steps. The road network for 2009 was extracted from the GISCO DB and was dissagregated in four different networks, using the hierarchical code of the segments :

Code 14 : Primary route (long-distance road)
Code 15 : Secondary route (regional road)
Code 16 : Limited access route (motorway)
Code 984 : Local road

These four networks were intersected with the LAU2 frame and lenghth for the new objects were calculated. As in the previous case, the intersected objects keep a double attribute (road and LAU2) that can allow us to use a summurize function in order to obtain the network coverage for every LAU2. The four files were joined to the LAU2 geometry of 2006 (COMM_RG_06 - GISCO DB) and the total length of the road network was calculated. Obtaining the road density by LAU2 was just a matter of a new field calculation.

The result was mapped for all the ESPON Space (excepting Bulgaria) and the cartographic illustration puts some problems of interpretation. The pattern is again very clear, maybe more clear at LAU2 level than at superior scales (NUTS3) and it shows how the ESPON Space is structured by a core vs. peripheral regions model, complicated by macro-regional gradients. Beeing a matter of territorial endowment, the road density could present interest for policy makers because this indicator can be crossed with other variables (local accesibility to cities or services of general interest, for example). Appearentely, it could be considered an indicator that is very related to the scale of intersection (LAU2). Evidences from the map rather suggest the contrary. Large LAU2 from Netherlands or Finland have similar values with LAU2 from France or Italy, making it not so easy to explain the results using the size of the local administrative units. Another aspect that might interfere with the mapping process is the classification of roads, the regions where the local roads are numerous or densely distributed having an influence on the final map. We have also taken into account the hypothesis that the distribution of the road density is influenced by the population density repartition in space. Using an OLS regression we have obtained an unconclusive result, the R2 being very small (0,067).

Maybe more interesting than the simple road map density, having separated the network by the hierarchy allowed us to provide a qualitative classification of the LAU2, based on the road type that intersects the spatial unit. The interstitial areas that result from this analysis can be compared at regional, national and European level. The mosaic of situations that results from this map might be relevant for policy decision.



Local level : LAU2
Source : ESPON 2013 Database, 2010
Origin of data : GISCO for geometries
(c) EuroGeographics Association for administrative boundaries

Methodological note :

The road density is based on two layers of spatial information :

1) the road network 2009 (GISCO DB)
2) the LAU2 geometry for 2006 (GISCO DB)

The two layers were intersected in order to calculate the road density.

Fig 8 Road density at local scale (LAU2) in 2009

The map may look interesting but it is also tricky, being this time highly dependent on the roads classification that also relates on national standards. Can we compare a primary road from Poland with a primary road from UK ? Maybe yes, but we don't have enough information to verify it. That's why this classification is based on what we called pseudo-connectivity.

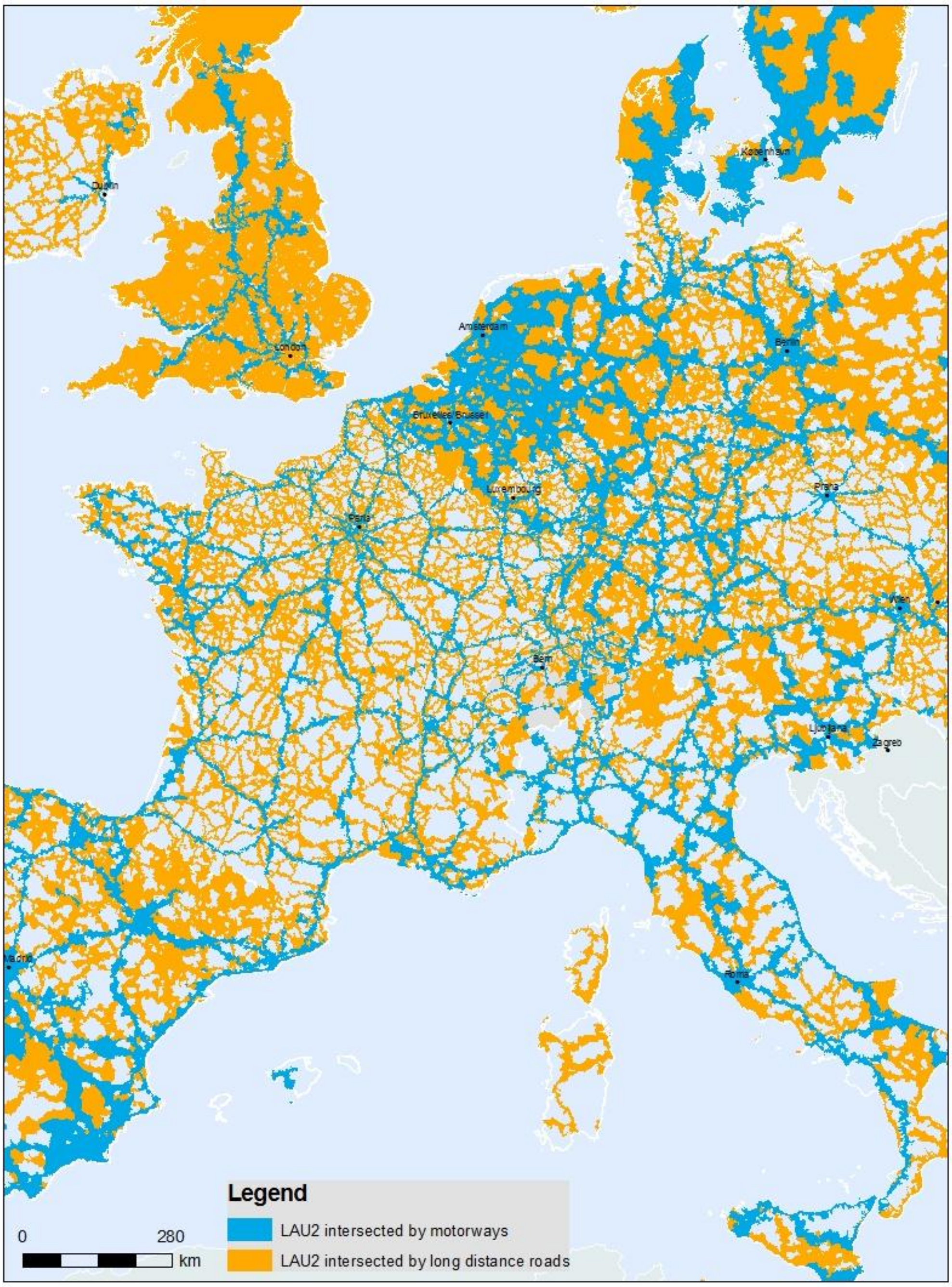


Fig 9 Qualitative classification of the LAU2 based on their pseudo-connectivity to superior road segments. Draft/working map for output verification.

As the road map refers to the situation of 2009 and despite the precautios reading of this cartographic illustration, we consider that this classification could qualify for the core database strategy, eventually allowing the construction of time series at LAU2 scale. If the territorial endowment with road networks is enriching, having an image of the evolution might present some interest.

The final comments on this part will present the ingredients and the methods we used in order to create this indicators:

1) as usual, a proper projection of the files is needed, both for the road network and the LAU2 geometry (COMM_RG_06)
2) split the network by the hierarchical code in 4 different parts.
3) intersect each part of the network with the LAU2 geometry. A model can be used in order to accelerate this process, but it will rely on the hardware endowment.
4) calculate the length of each segement resulted from the intersection
5) summarize the length by LAU2 code (COMM_ID in the COMM_RG_06 file)
6) join the four created tables to the COMM_RG_06 and export it to a new product (output)
7) map the lenghth in order to superficially verify for errors
8) calculate the road density or other indicators

3 Other issues related to the data integration at local scale

This short part will deal with the concept of spatial interaction and its exploration at local scale. The spatial interaction is usually associated with the exchanges of goods, energy, information or people from one place to another. The formalizations and the theoretical background of geography, regional science and economy allowed even more developments of this intuitive definition, linking the spatial interaction to more elaborate concepts such as area of influence for services or potential of interaction. Due to the difficulty to manage many geographical objects (LAU2), researchers sometimes avoid attacking this subject at local scale. Excepting the case studies or the grey literature, is quite difficult to find examples of studies centered on the spatial interaction at local scale.

The technical difficulties are not easy to overcome, but even if one will manage to create the technical frame to study this topic, he will still face the problem of data collection. Limited as they are, the data exist and can be used in order to better describe the structures of territorial organization at local scale. The SIRE DB contains information about one aspect of the spatial interaction (the commuters' flows between the LAU2), but the data quality is questionable. However, it can be linked to a geometry and at least explore these flows to satisfy a genuine scientific curiosity. Linking the data involves the existence of a common key to make the join possible and to allow the cartographic illustration.

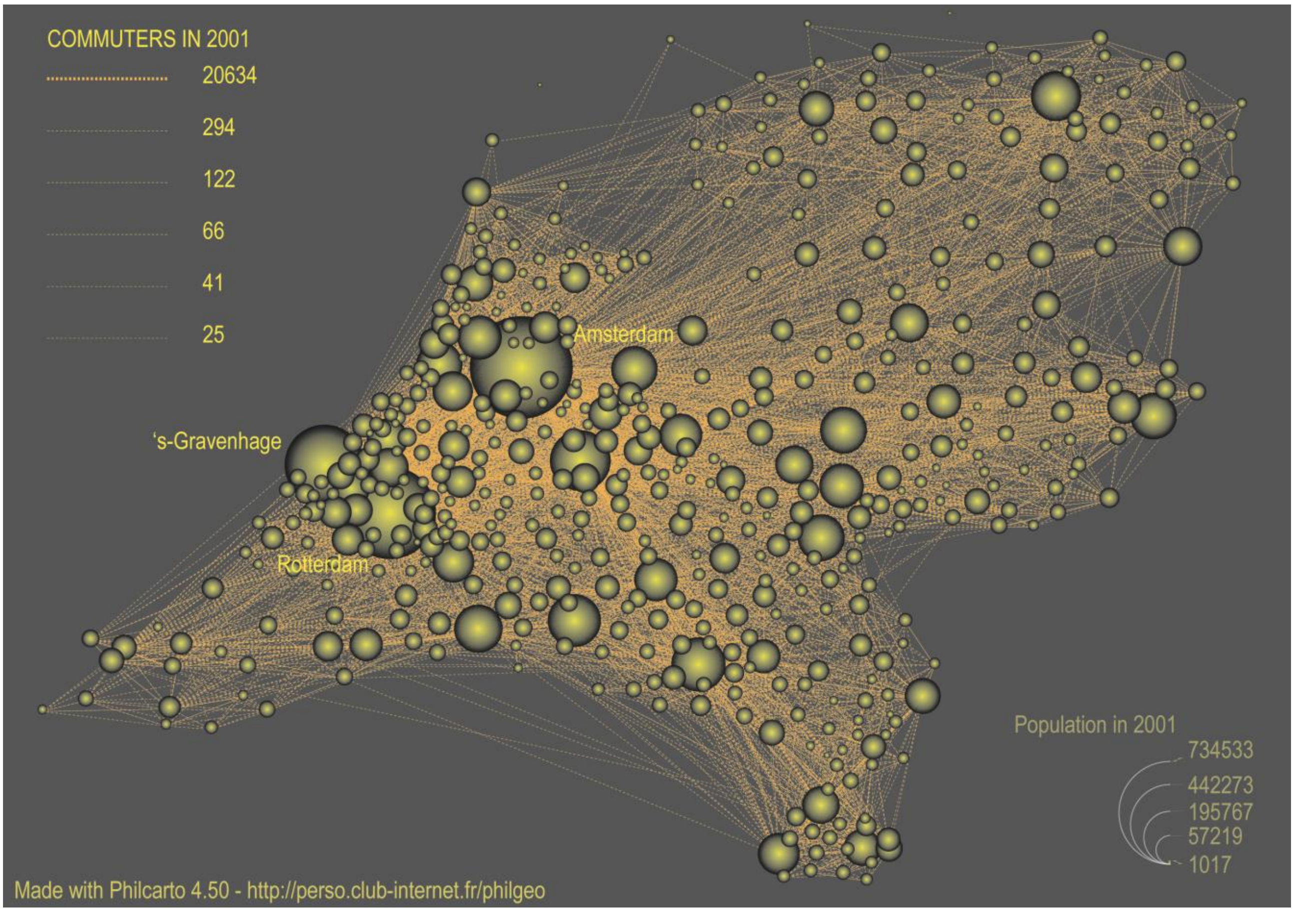


Fig 10 Flows of commuters between the LAU2 in Netherlands. Draft/working map for data visualisation.

Both the cases of Netherlands and Spain show how much diversity in the organization of the flows can exist, in the same time, in the ESPON Space. If we use flows only to provide descriptions about how cities act, a large part of the territory will be ignored.

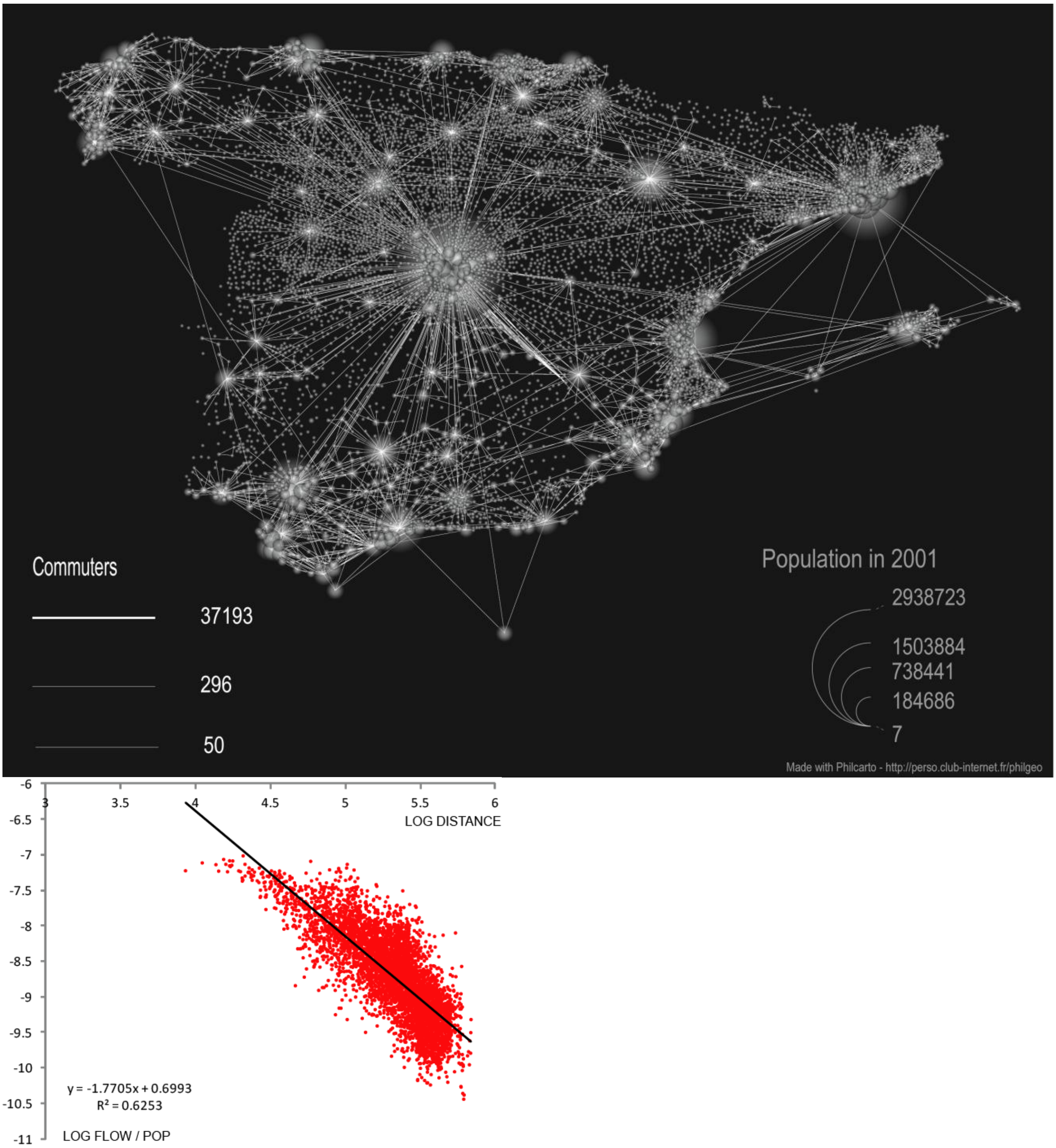


Fig 11 Flows of commuters between the LAU2 in Spain (mainland) and OLS regression between distance to Madrid and flows'intensity. Draft/working map for data visualisation.

The opposition between two models of urban organization (highly polycentric in Netherlands vs. moderate to high monocentric in Spain). This opposition could be a source for the construction of new indicators : emmisivity and attractivity of the LAU2, slope of the regression line between the distance to the poles of attraction (urban and rural) and the flows intensity or probabilities of spatial interaction based on more accurate parameters. The slope of the regression line between distance and the flows intensity (ratio between flows and the product of populations at origin and destination) is also a descriptor of the networks efficiency. In the case of Madrid, we used straight-line distances between LAU2 centroids and the result is questionnable. Using road distances or multimodal distances will eventually make the result more accurate.

For the moment, we have managed to integrate flow data for selected test countries :

Austria - monocentric urban system dominated by Wien, complicated situation of the LAU2 spatial configuration due to natural constraints
Belgium - low to moderate monocentric, christallerian urban dispositif, highly urbanized and artificialized
Germany - polycentric urban system, internal historical barrier induced by the discontinuities between Eastern and Western Germany. The attempt to integrate the data on this country was possible only for a limited number of spatial units because a lack of proper matching between the key codes.
Italy - relatively polycentric, regional gradients of development
Netherlands - polycentric, densely populated and artificialized.
Spain - moderate to high monocentric urban system, population concentration on the coasts

Data on Czech Republic, Slovakia and Bulgaria is also available, for the last country the scale of referernce being the LAU1 frame and not the LAU2 geometry. An attempt was made to integrate data on Hungary too. Mismatch between the coding system of the spatial units was again a barrier in accomplishing the task. A new system of matching for Hungary is in preparation and it is based on the concatenation of different spatial references (NUTS3, LAU1 and toponym) in order to identify the spatial units. The new system will propose a match based on fragments of toponyms that have an unique probability of matching, eliminating possible errors.

Mapping these flows is a matter of compromise between cartographic visualisation and exhaustiveness of the data reprsentation. In the case of Netherlands the flow limit was set to 25 commuters. For Spain a choice was made for 50 commuters, while for Italy the needed limit was 150 commuters. The representation of data in draft/working maps was made using Philcarto rather than a GIS software for a rapid visualization. It is a fast solution that sacrifice georeference in the name of fast result verification.

Concerning the data quality issues, we don't have a clear answer but a sincere one. The map shows a relative good matching between the spatial units, at origin and at destination. The intensity and the volume of flows look also sufficientely satisfying, according to the basic gravity modeles. The coefficient of determination (R2) between distance and flows' intensity is rather appropiate for a real flow matrix than for an estimation (we should have expected a larger value, in a different case). From the point of view of an XVIII century expert in litterature, the data on the commuters respect all the rules of the *vraisemblance*. Without proper metadata about this topic, our approach is marked by the limits of the fuzzy probabilities.

Exploring the issues of spatial interaction at local scale is not only linked with the difficulties of dealing with commuters extracted from the SIRE DB. The theoretical models, if properly adusted and calibtrated can also offer an inisght on the territorial organization at LAU2 scale. Two models were tested but only one was illustrated on the map - the Huff model and the model of potential of interaction, the last using a gaussian function with a spawn of 25 or 50 km at 50 % probability. In order to better calibrate the model, we put into relation all the LAU2 in some selected countries of Eastern and Central Europe with all the LAU2 with more than 10 000 inhab, in a search radius of 100 km around each "destination". This radius will eliminate the theoretical probability, even if reduced, to put into interaction spatial units that are unlikely to interact, focusing thus on the "origins" rather than the "destinations". The canonical formalization use several parameters that are based on the assumptions that one made about the spatial interaction potential.

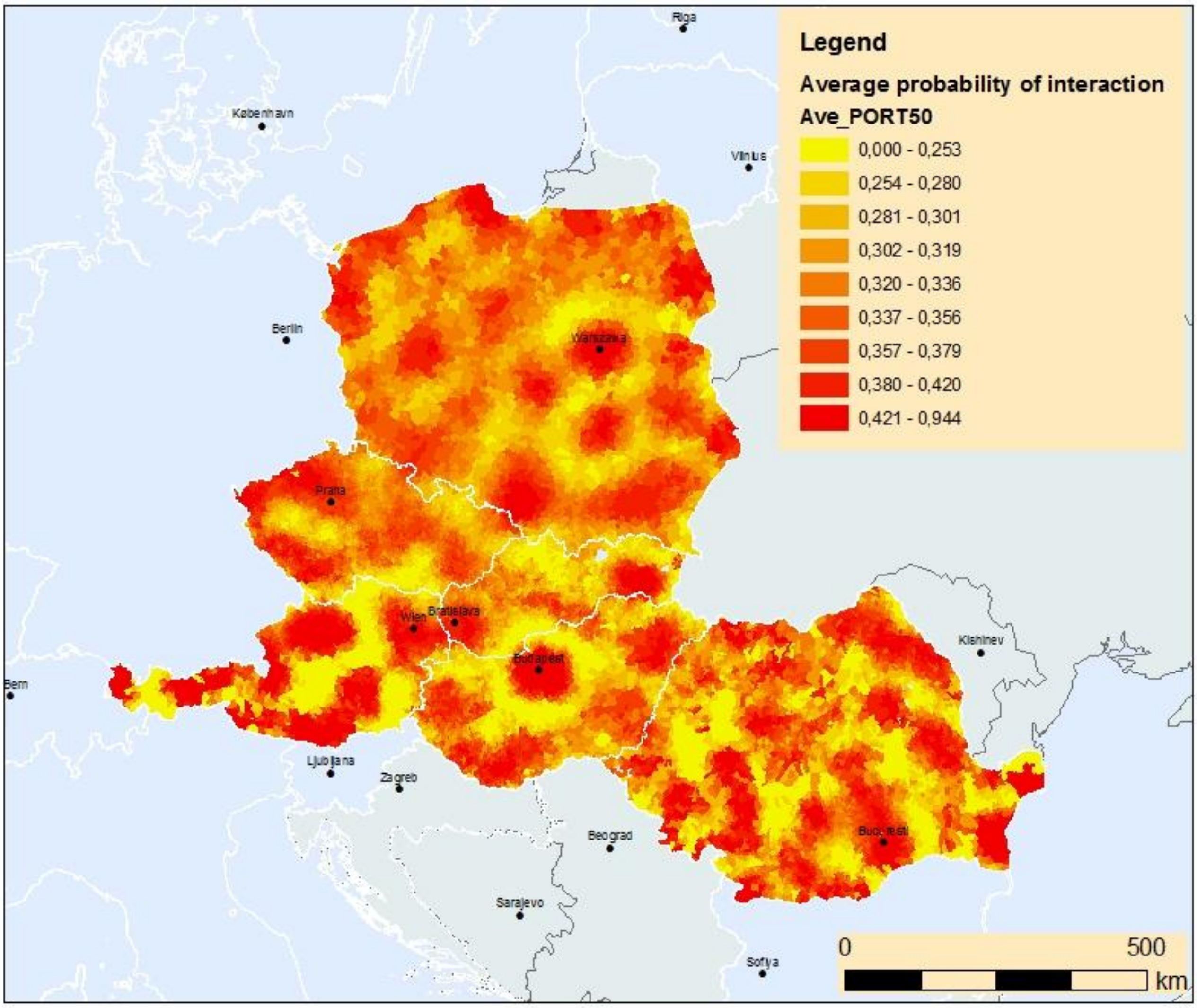


Fig 12 Average probability of interaction calculated using a gaussian function with a spawn of 50 km for LAU2 smaller than 10 000 inhab. in 2006 (the function was applied on a road distance limit of 100 km around each LAU2). Draft/working map for data visualisation.

It is not only a problem of spawn in the initial model, this aspect also involves a good approximation of the distance interference with the spatial interaction, else being just another elegant mathematical function. In the case of countries from Eastern and Central Europe, an exponent of distance of 2 looks still apropriate because of severe limits in the road density distribution. Obtaining this indicator involved the next steps in the calculation:

1) double projection of geometries (COMM_LB_06 centroids of LAU2 - GISCO DB and the road network for selected countries - PL, CZ, SK, AT, HU, RO)
2) elaboration of a network dataset based on road distances
3) extraction of the origin centroids (LAU2 < 10 000 inhab.) and the destination centroids (LAU2 > 10 000 inhab.)
4) construction of the origin-destination road distance matrix, limited at 100 km radius around the origins.
5) estimation of the probability index based on a Gaussian negative function with different spawns (25 and 50 km) for each LAU2 at origin
6) estimation of the central values (mean of the probability index) and dispersion values (standard deviation, maximum and minimum)
7) if a complete model is needed, multiply the population at origin with the probabilities elaborated in the model. Summarize the data by LAU2 code and map the result. Else, map central or dispersion values calculated at step no. 6

A plan B network dataset, based on time distances, was also used to estimate the potential of interaction for the LAU2 of the selected countries. The time distances were associated with the average speed for every road segment in the studied region. The speed was allocated in the following way:
1) 25 km/h for the local roads
2) 45 km/h for the regional roads
3) 60 km/h for the long distance roads
4) 90 km/h for the motorways
These speeds might look underestimated, but this should be considered as a testing scenario with pessimistic pre-conditions, especially for one having visited the Casimcea Plateau in Romania. As the calculation time with minutes as impedance was discouraging for the whole region, we have tested the model only for one country (Romania - 3175 LAU2). Working and fitting quite well the reality, our approach is fragile when extrapolating to other territories. It was, however, a good opportunity to make us reflect more on the model, especially on the definition of centroids. The mean center of the LAU2 polygons involved in the model depends on the administrative fragmentation of the geometry we use (COMM_RG_06 from GISCO DB). In this case, the centroid will always reflect the spatial situation of the polygon limits and not the distribution of population. That's why the model always suffers from an approximation of distances, instead of reflecting more accurate demographic repartitions in space. The question we had in mind was: is there any method to better approximate this "centroid"?

Theoretically, using the settlements' layer from the GISCO DB would help, if only it would be provided with some demographic information, which is not always the case. In the layer we used (2008 as time reference) only 127 501

populated places had a demographic mass attached as attribute (from more than 285 000 spatial units). Here is the critical point where new spatial products interfere with our intentions, the very recent population grid from Eurostat (officially named GEOSTAT_Grid_POP_2006_1K). In this grid, the population was disaggregated using an 1 km spatial step, from Iceland to Bulgaria and from Italy to Finland. As any grid, the information is submitted to a surface spatial structure. Converting this polygon geometry to a point structure is an easy task. Associating the population to the settlements is also an easy task, if one will use classical tools such as the Thiessen polygons in order to allocate inhabitants to the closest settlement. Theoretically again, there are three different candidates for the function of centroid:
- the GISCO center (the mean center of the X,Y coordinates describing the LAU2 perimeter, if our hypothesis is good)
- the official center as described by national standards
- the calculated center derived from several intersection between population grid, Thiessen polygons based on settlements and LAU2 geometry.

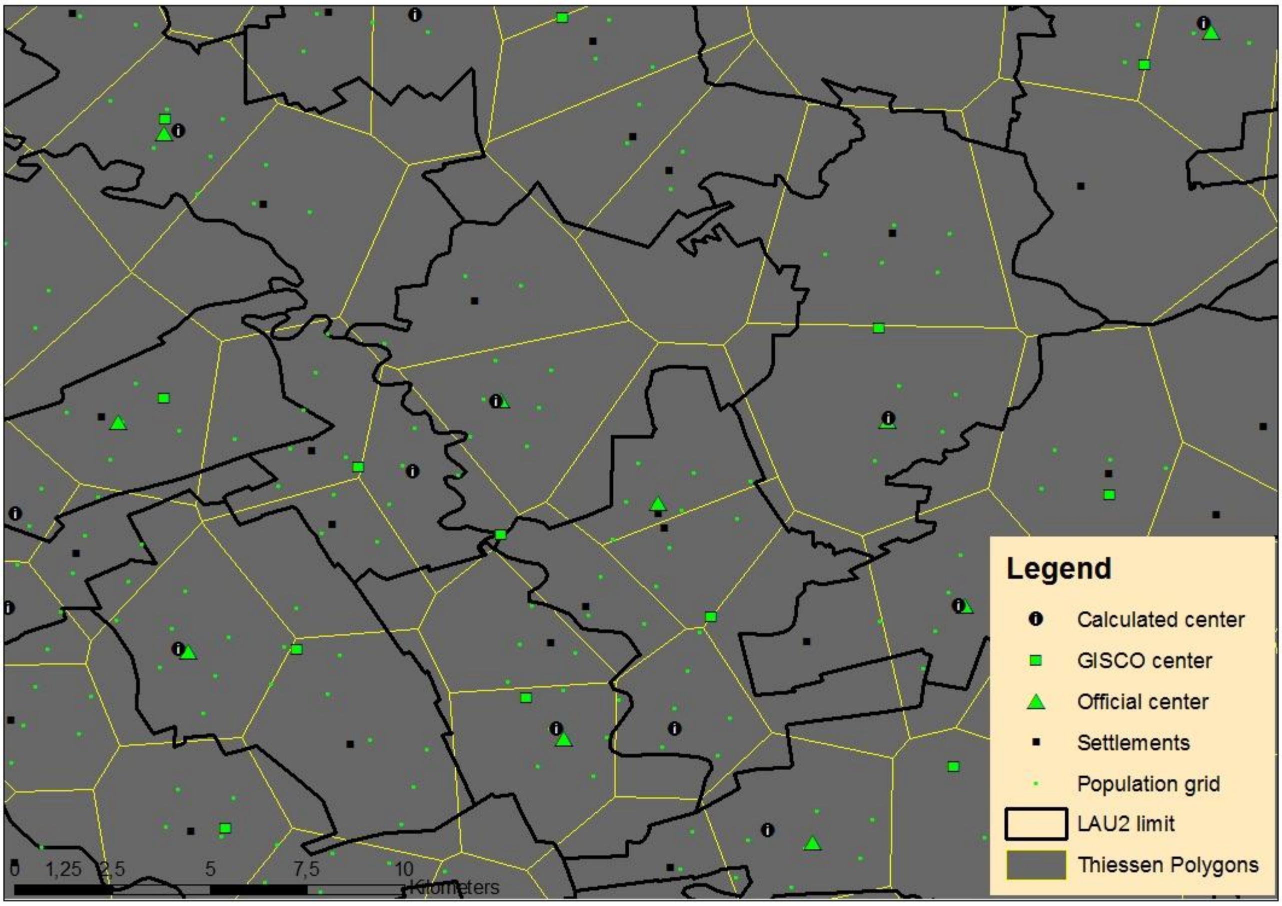


Fig 13Layers needed to calculate an alternative centroid to the LAU2 mean center

A solution was found in a case study on Romania (playground for testing models and hypothesis due to the spatial or territorial variety of geographical features). The steps needed to create a layer with more apropriate centroids are:
1) project the geometries (COMM_RG_06, SETTL_08 and the grid of population)
2) transform the grid from polygon to points
3) create Thiessen polygons for each settlement

4) intersect de Thiessesn polygons with the LAU2 geometry and obtain 3 codes - Thiessen code, settlement code and LAU2 code
5) intersect the new objects from point 4 with the grid of population (point structure)
6) summarize the file from point 5 in order to obtain the population for each Thiessen polygon and settlement. Use the code of the settlement to calculate the summurized result - SETTL_ID
7) join the summarized table to the settlement file intersected with the LAU2
8) summarize again the table from point 7, using the LAU2 code in order to detect the maximum of population for each settlement situated in a LAU2 polygon. The demographic centroid was obtained (called calculated center on the map on fig 13)

As the image show, in some cases the coincidence between the official centroids and the calculated centers (demographic centroids) is strong. In other cases, there is a lack of spatial matching between the two, possibly caused by differences between the settlement layer and the official layer of LAU2 centers. In any of the cases, the calculated center is more appropriate than the GISCO centroid, especially when used as an entry layer in distance calculation.

An option to this method was also considered, calculating the weighted mean center for every LAU2. The weight is given by the population grid. For each cell in the grid we attach the X, Y couple of coordinates of the centroid. The grid is intersected with the LAU2 frame and a double code is obtained in the new file : cell code and LAU2 code. The X and Y coordinates of each cell are weighted with the population of the cell. The file is summarize by LAU2 code and three new fields are obtained : the sum of weighted X coordinate, the sum of weighted Y coordinate, the population of each LAU2 as sum of cells intersected by LAU2. The next step is to calculate the weighted center of the LAU2 with two coordinates X and Y.

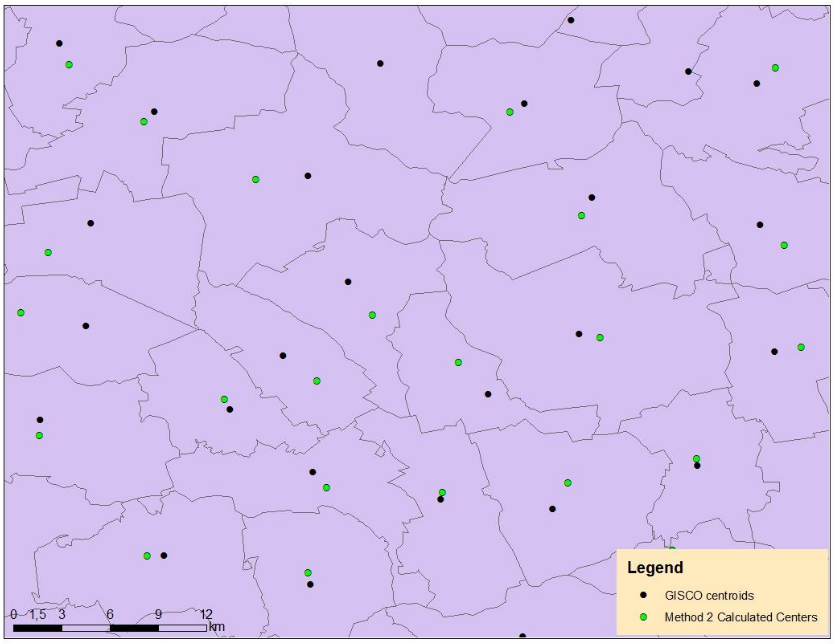


Fig. 14 Alternative method to calculate the centers of the LAU2 - weighted centers

Three topics of interest were explored in this part :

- how to integrate data from SIRE DB, especially flow data related to the commuter population. The methdological exploration showed that the integration is possible and that the data can be mapped, if one will use a LAU2 geometry for 2001. There are still some issues to solve in the case of some countries (Germany or Hungary), where administrative modifications of the geometry complicate the data integration process.

- applying theoretical models of spatial interaction to selected countries from Eastern Europe. The calibration of the models and the use of a network geometry for distance calculation provide some encouraging results. However, problems still exist and this time are related to the point layers introduced in the modelling process.

- defining new centroids for the LAU2 datasets. The question of the centroids is one of major interest because their spatial position will influence models based on distance (from interpolation to spatial interaction models). There are two solutions to better aproximate these centroids, using a population grid and the settlements layer.

Conclusions

The integration of data from CLC 2006 in the LAU2 geometry show that the construction of indicators at local scale is possible, for all the countries in the ESPON Space. Despite the technical difficulties involved by the management of more than 120 000 spatial units, we have integrated data related to the land use and elaborated simple tools of data transformation (adaptation of the location quotient to surface analysis). The main CLC 2006 categories integrated describe the artificial areas, the arable land, the heterogeneous agricultural surfaces, the permanent crops and the forest and natural areas. For case studies (Spain and Portugal), all the 44 CLC 2006 categories were integrated in the LAU2 frame.

A second topic of interest was the construction of road density indicators. Using a road network from 2009 and the LAU2 geometry of 2006, we have applied methods of spatial analysis (intersection and measurements of network segments) in order to derive the density. The output of this methodology also allowed us to build a LAU2 classification based on the connectivity to different hierarchical network segments (motorways, regional roads, long distance roads and local roads). Mapping the outputs illustrates the diversity of territorial endowment at local scale and how different countries or regions are situated in relation with the road newtork.

The last part was dedicated to different issues related to the local data integration : commuters flows between the LAU2, spatial interaction models applied at local scale (using distances calculated on the road network) and alternative methods to derive centroids for the LAU2. The experimental work allocated to these explorations can be summarized in a list of recommendations related to the working flow involved in these topics.