TECHNICAL REPORT

ESPON

UMZ: a data base now operational for urban studies (M4D improvements)

CONTENT

Updating and improving the methods for naming UMZ:

Generic methods have been elaborated for improving the computation of population starting from grids, for assigning names to UMZ and for checking the validity of the assigned names according to different urban databases.

Elaborating the UMZ-LAU2 dictionary of correspondence:

In order to populate UMZ database with LAU2 databases, a generic method for adjusting urban objects to local administrative units has been elaborated.

Populating UMZ with SIRE database:

A generic method has been constructed for allocating LAU2 indicators into UMZ objects. Different SIRE indicators have been expertised, regarding completeness and robustness criteria. Results are presented for age structure.

33 pages

ESPON M4D -
MULTI DIMENSIONAL DATABASE DESIGN & DEVELOPMENT

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

LIST OF AUTHORS

Anne Bretagnolle, University Paris 1, UMR 8504 Géographie-Cités

Marianne Guérois, University Paris 7, UMR 8504 Géographie-Cités

Hélène Mathian, C.N.R.S., UMR 8504 Géographie-Cités

Antonin Pavard, C.N.R.S., UMR 8504 Géographie-Cités

Contact

anne.bretagnolle@parisgeo.cnrs.fr

UMR 8504 Géographie-Cités
Tel. (+ 33) 1 40 46 40 00

TABLE OF CONTENT

Introduction

Urban Morphological Zones have been created in 2004 by the European Environment Agency. It has been constructed using highly automated methods and regularly updated (three dates are now available for UMZ perimeters, 1990, 2000 and 2006). However, different tasks had to be done before rendering UMZ operational for urban studies. In the version freely available on EEA website, the objects are simply spots or patches, without names, and hence without semantic links with the territory. They only constitute a set of geometrical objects, and not of geographical objects.

In the first phase of ESPON 2013 Database, a generic method has been elaborated for assigning names to UMZ (see technical report "Naming UMZ", March 2011). The Urban Morphological Zone database has been uploaded in the ESPON Database in may 2013. However, three additional tasks were necessary to carry out in order to ensure a full and friendly utilisation of the DB by ESPON users.

The first one concerns the improvements of the ESPON 2013 Database naming methods. The present report updates these methods in three different ways, by improving the parameters choice for selecting the final name(s) of the UMZ, by using a most appropriate version of LAU2 for selecting the names and by improving the automatic algorithm used for giving the names as well as the procedure for checking the results.

The second task concerns the construction of a dictionary of correspondence between UMZ and LAU2. The creation of this dictionary was a complex but necessary step for enriching UMZ database with other statistical indicators, for example socio-economic or demographic ones. The method that has been elaborated is a generic one. It can be applied for other urban databases.

The third task concerns the enrichment of the UMZ database with the SIRE database (European infra-regional information system[1]). The dictionary of correspondence between UMZ and LAU2 allows populating this data base with local data, as it gives the composition of the UMZ in terms of LAU2. We have constructed a generic method for characterizing the intensity of each link LAU2-UMZ and for allocating the indicator according to this intensity. We have tested different SIRE indicators, regarding the completeness of the geographical coverage and the coherence of the data. A very few indicators have passed successfully these different tests, among them the age structure. We present finally a typology of European cities according to their age structure.

[1] Data for census 2001 in SIRE Database 2008, Eurostat, BSI. We also used Hampson, P., Raxis, P., 2008. "Database documentation, Managment of SIRE Data Base", Eurostat, BSI, 145p.

1 Improving the methodology for naming UMZ

The total number of UMZ has been reduced, from the Database 2013 version to the current M4D version. Indeed, we have eliminated the UMZ of Balkans (except those of Croatia), because the population data did not come from the same source and could create inconsistencies. So we have now 4304 UMZ instead of 4437.

1.1 Parameters for choosing a simple or a multiple name

The different parameters used to calibrate the choice of a simple name or a multiple name have been improved since the first version of the naming UMZ. These parameters consist in population thresholds and are computed in areas that intersect UMZ and reference administrative units (LAU2, LAU1, etc.). Until now, the very simple rule based on the threshold of 50% of the population concentrated in the selected LAU2, had been used in the naming procedure, as in the original method used by the French census board INSEE.

We have tested different thresholds, from 40% to 70% and examined systematically the results for a set of well known conurbations, where multiple names are expected.

Table 1: Different population thresholds for naming

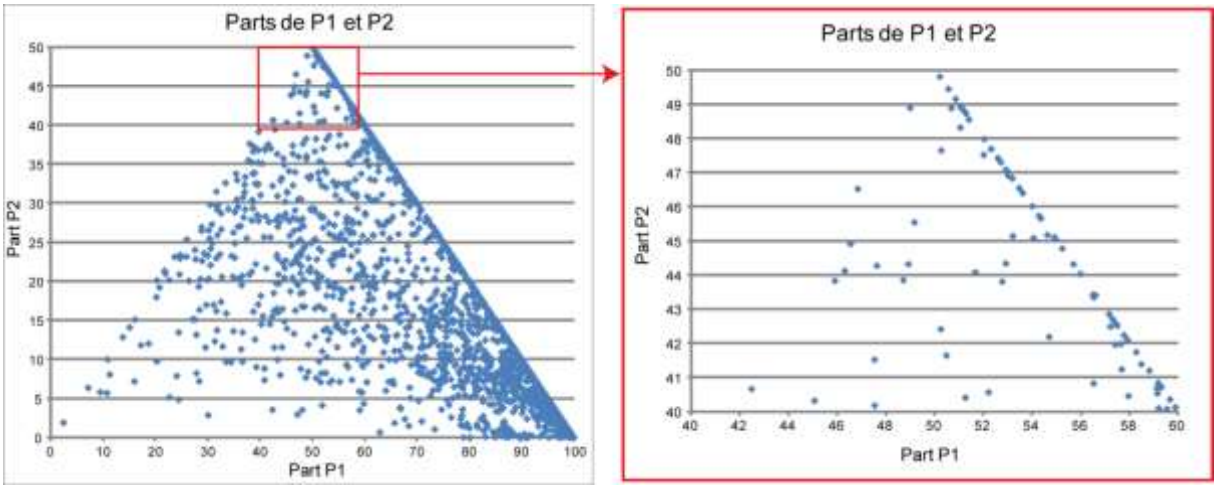
α level	β level	Status 1 Number of UMZ (One strong core)	Status 2 Number of UMZ (Several names)	Status 3 Number of UMZ (One weak core)	Total
50	50	4036	189	79	4304
100	50	1320	382	2602	4304
100	45	1320	439	2545	4304
100	43	1320	472	2512	4304
100	40	1320	501	2483	4304
70	45	3572	439	293	4304
70	43	3572	472	260	4304
70	40	3572	496	236	4304
60	45	3798	353	153	4304
60	43	3798	366	140	4304
60	40	3798	382	124	4304

Alpha level : Absolute dominance of the first administrative unit. If the population of this administrative unit is larger than xx% (for instance 40%) of the UMZ population, then the UMZ receives the name of this unit (Status 1).

Beta level : Relative dominance of the first administrative unit. If there is no administrative unit with a positive *alpha* test, then we check if the population of administrative units have significantly different contributions. More precisely, we measure the difference between the contribution of the first administrative unit and the contribution of all other administrative units. If this difference is larger than a given beta level (($PartP_k$ / $PartP_1$)> = β%), then the UMZ takes several names (the name of the dominant basic unit, plus the name of the administrative units being positive for the beta test).

The first line of Table 1 corresponds to the situation of the ESPON Database 2013 naming (threshold of 50% for α and β). We had 4036 UMZ with one strong core (one name – 1sC), 189 UMZ with several cores (several names - xC), and 79 UMZ with one weak core (one name – 1C). As the 50% threshold is fairly arbitrary (what to do with cases that fit the 49% threshold?), we have explored the range of situations where an administrative unit of rank 1 (P_1) and an administrative unit of rang 2 (P_2) contributes to 40% up to 60% of a UMZ population (Figure 1). There are around 90 UMZ for which these two administrative units have a balanced contribution.

Figure 1: Administrative unit of rank 1 and administrative unit of rank 2



In order to know if there is a less arbitrary threshold than the 50% one for α and β levels, we have realized some tests (see Table 2). The objective was to find a balance between α and β.

Table 2: Different naming results according to population thresholds: examples for 6 UMZ

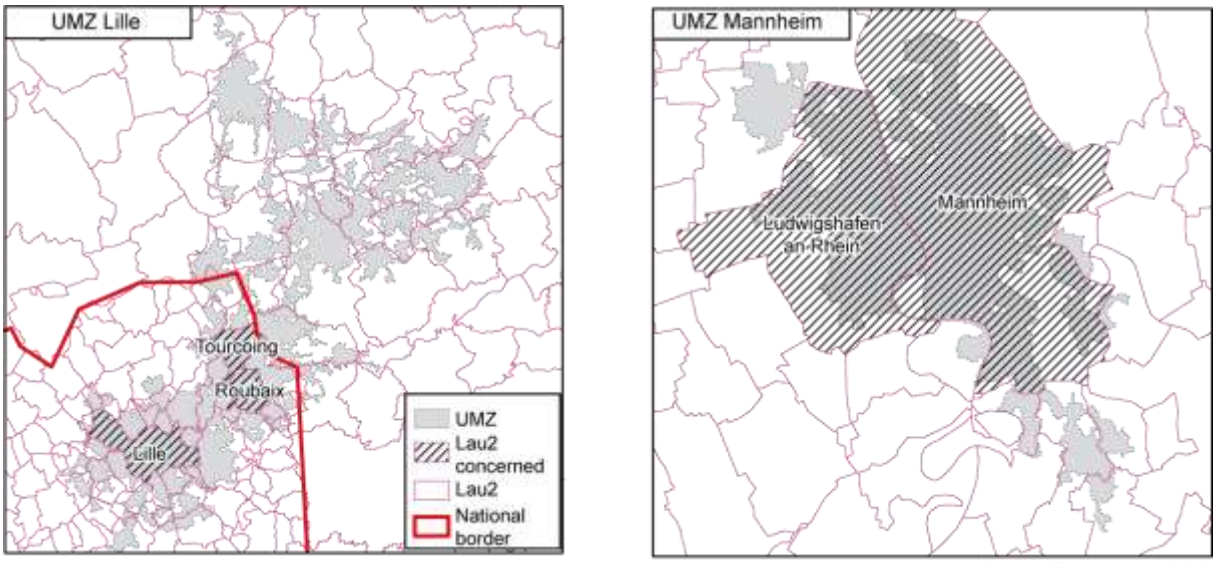
α level	β level	Lille	Mannheim	Hennigsdorf	Roche-la-Molière	Vichy	Cuijk
50	50	Lille	Mannheim	Berlin	Roche-la-Molière	Vichy	Cuijk
100	50	Lille	Mannheim	Berlin // Hennigsdorf	Roche-la-Molière	Vichy	Cuijk
100	45	Lille // Roubaix	Mannheim // Ludwigshafen am Rhein	Berlin // Hennigsdorf	Roche-la-Moliere // Saint-Genest-Lerpt	Vichy	Cuijk
100	43	Lille // Roubaix // Tourcoing	Mannheim // Ludwigshafen am Rhein	Berlin // Hennigsdorf	Roche-la-Moliere // Saint-Genest-Lerpt	Vichy	Cuijk
100	40	Lille // Roubaix // Tourcoing	Mannheim // Ludwigshafen am Rhein	Berlin // Hennigsdorf	Roche-la-Moliere // Saint-Genest-Lerpt	Vichy // Cusset	Cuijk // Mook en Middelaar
70	45	Lille // Roubaix	Mannheim // Ludwigshafen am Rhein	Berlin // Hennigsdorf	Roche-la-Moliere // Saint-Genest-Lerpt	Vichy	Cuijk
70	43	Lille // Roubaix // Tourcoing	Mannheim // Ludwigshafen am Rhein	Berlin // Hennigsdorf	Roche-la-Moliere // Saint-Genest-Lerpt	Vichy	Cuijk
70	40	Lille // Roubaix // Tourcoing	Mannheim // Ludwigshafen am Rhein	Berlin // Hennigsdorf	Roche-la-Moliere // Saint-Genest-Lerpt	Vichy // Cusset	Cuijk
60	45	Lille // Roubaix	Mannheim // Ludwigshafen am Rhein	Berlin // Hennigsdorf	Roche-la-Moliere	Vichy	Cuijk
60	43	Lille // Roubaix // Tourcoing	Mannheim // Ludwigshafen am Rhein	Berlin // Hennigsdorf	Roche-la-Moliere	Vichy	Cuijk
60	40	Lille // Roubaix // Tourcoing	Mannheim // Ludwigshafen am Rhein	Berlin // Hennigsdorf	Roche-la-Moliere	Vichy // Cusset	Cuijk

We have decomposed the different examples in order to better understand the spatial configurations (Table 3 and Figure 2).

Table 3: Importance of the administrative units according to their share of the population included within UMZ

Info	Lille	Mannheim	Hennigsdorf	Roche-la-Molière	Vichy	Cuijk
UMZ Population	1335026	557008	80304	13775	53315	21614
P1	212135 (Lille)	304734 (Mannheim)	41555 (Berlin)	9446 (Roche)	26373 (Vichy)	15244 (Cuijk)
P2	96183 (Roubaix)	141676 (Ludwigshafen)	25917 (Hennigsdorf)	4316 (St-Genest)	11333 (Cusset)	6243 (Mook)
P3	93530 (Tourcoing)					

Figure 2: Examples of spatial configurations



Antonin Pavard, June 2012

Starting from these 6 examples, and more generally from our knowledge on European conurbations (for instance, Manchester-Liverpool or Lille-Roubaix-Tourcoing), we have finally selected two different thresholds, **α = 70% and β = 40%.**

We have summarized the results using the new thresholds according to two indicators: the type of the UMZ (one strong core, several cores, one weak core), and the name of the UMZ. Table 4 indicates the number of UMZ that have been modified according to the type of the UMZ (Column A) and to the name (Column B). These new parameters have a clear impact on countries characterized by important conurbations (ex: Ruhr basin, main seashores).

Tableau 4: Modifications due to the change of thresholds by country

Country	Type of UMZ (Column A)	Name of UMZ (Column B)	Name of UMZ (in %)
AT	22	15	25
BE	12	6	12
BG	2	2	2,4
CY	3	1	20
CZ	3	1	0,9
DE	78	66	7,8
DK	2	2	3,9
EE	1	1	6,3
ES	39	34	9,8
FI	3	2	4,5
FR	100	58	14,8
GR	1	1	2
HR	3	2	5,6
HU	4	8	7,5
IE	2	1	3,8
IT	97	83	14,4
LU	1	1	100
NL	18	1	100
PL	19	13	6,5
PT	6	18	5,5
RO	15	7	10,0
SE	7	19	11,9
SI	2	6	6,8
SK	2	2	16,7
UK	52	2	3
Total	496	3	25

Table 5 presents the modifications for the 10 largest UMZ. For instance, Greater Manchester was classified as a "UMZ with one strong core", whereas this UMZ covers two important cities, Manchester and Liverpool. Manchester city is a potential core of the UMZ, but Liverpool too. In the new version of the database, the UMZ receives the name "Greater Manchester" but is qualified as a weak core.

Table 5: Top 10 modifications (10 UMZ most populous have undergone changes)

ID_Up	Pop_Up	NAME_Up	Country	Seuil50_50	Seuil70_40
13409	3546819	Greater Manchester	UK	1sC	1C
36402	1767659	Koeln	DE	1sC	1C
33959	1335026	Lille	FR	1C	xC
91932	1278016	Torino	IT	1sC	1C
4410	1233147	Stockholm	SE	1sC	1C
24115	1072014	Rotterdam	NL	1sC	1C
2956	917813	Helsinki#Helsingfors	FI	1sC	1C
30617	809770	Düsseldorf	DE	1sC	1C
57105	715660	Stuttgart	DE	1sC	1C
100258	703413	Firenze	IT	1C	xC

Table 6 illustrates the contribution of each administrative unit inside one UMZ in the Top 10 presented in the previous table. Table 7 gives the resulting names, depending on the selected thresholds.

Table 6: Contribution of administrative units by UMZ

ID_Up	Pop_Up	NAME_Up	P1	P2	P2 / P1
13409	3546819	Greater Manchester	(Manchester) 57%	(Liverpool)22%	38%
36402	1767659	Koeln	(Koeln) 50%	(Bonn) 17%	33%
33959	1335026	Lille	(Lille) 16%	(Roubaix / Tourcoing) 7%	44%
91932	1278016	Torino	(Torino) 68%	(Mocalieri) 4%	6%
4410	1233147	Stockholm	(Stockholm) 59,5%	(Huddinge) 6%	10%
24115	1072014	Rotterdam	(Rotterdam) 53%	(Schiedam) 7%	13%
2956	917813	Helsinki#Helsingfors	(Helsinki) 57,5%	(Espoo) 20%	33%
30617	809770	Duesseldorf	(Dusseldorf) 67%	(Rleuss) 15%	22%
57105	715660	Stuttgart	(Stuttgart) 67%	(Ludwigsburg) 11%	17%
100258	703413	Firenze	(Firenze) 49,6%	(Prato) 22%	44%

Table 7: Examples of UMZ with different names depending on different thresholds

ID_Up	Pop_Up	Ctr	Seuil50_50	Seuil70_40
33959	1335026	FR	Lille	Lille // Roubaix // Tourcoing
30174	1138180	DE	Wuppertal // Hagen	Wuppertal // Hagen // Solingen
100258	703413	IT	Firenze	Firenze // Prato
11471	699727	PL	M Gdansk	M Gdansk // M Gdynia
49012	557008	DE	Mannheim	Mannheim // Ludwigshafen am Rhein
32915	411544	BE	Hasselt // Genk // Beringen	Hasselt // Genk // Beringen // Sint-Truiden // Heusden-Zolder
20665	341191	NL	Haarlem	Haarlem // Velsen
125593	291210	ES	Santa Cruz de Tenerife	Santa Cruz de Tenerife // San Cristobal de la Laguna
32645	259447	NL	Heerlen // Kerkrade	Heerlen // Kerkrade // Landgraaf
115222	246440	IT	Caserta // Marcianise	Caserta // Marcianise // Santa Maria Capua Vetere
36650	203498	FR	Valenciennes	Valenciennes // Denain
91629	201106	BG	Ruse	Ruse // Giurgiu
38720	195674	PL	M Rybnik	M Rybnik // Wodzislaw Slaski
29724	166984	BE	Sint-Niklaas	Sint-Niklaas // Beveren
19168	160594	NL	Alkmaar	Alkmaar // Heerhugowaard
100609	128632	IT	Pisa	Pisa // Cascina
104736	124795	PT	Guimaraes	Guimaraes // Santo Tirso
21694	123319	UK	Rhondda	Rhondda // Pontypridd
61395	123095	DE	Ulm	Ulm // Neu-Ulm
70485	122512	DE	Loerrach // Weil am Rhein	Loerrach // Weil am Rhein // Saint-Louis

1.2 LAU2 reference year

In the ESPON Database 2013 version of the naming, we used the 2006 LAU2 database. During the M4D project, we obtained the 2001 LAU2 database, that fit better with the reference year of the UMEZ perimeters. The differences between the 2001 and the 2006 LAU2 may be important in some countries (Figure 3).

In France (case A on Figure 3), the 16 july 1971's law allows the association of municipalities, which can now combine to form a single one. It was the case of Lille and Lomme, which have joined in 2000.

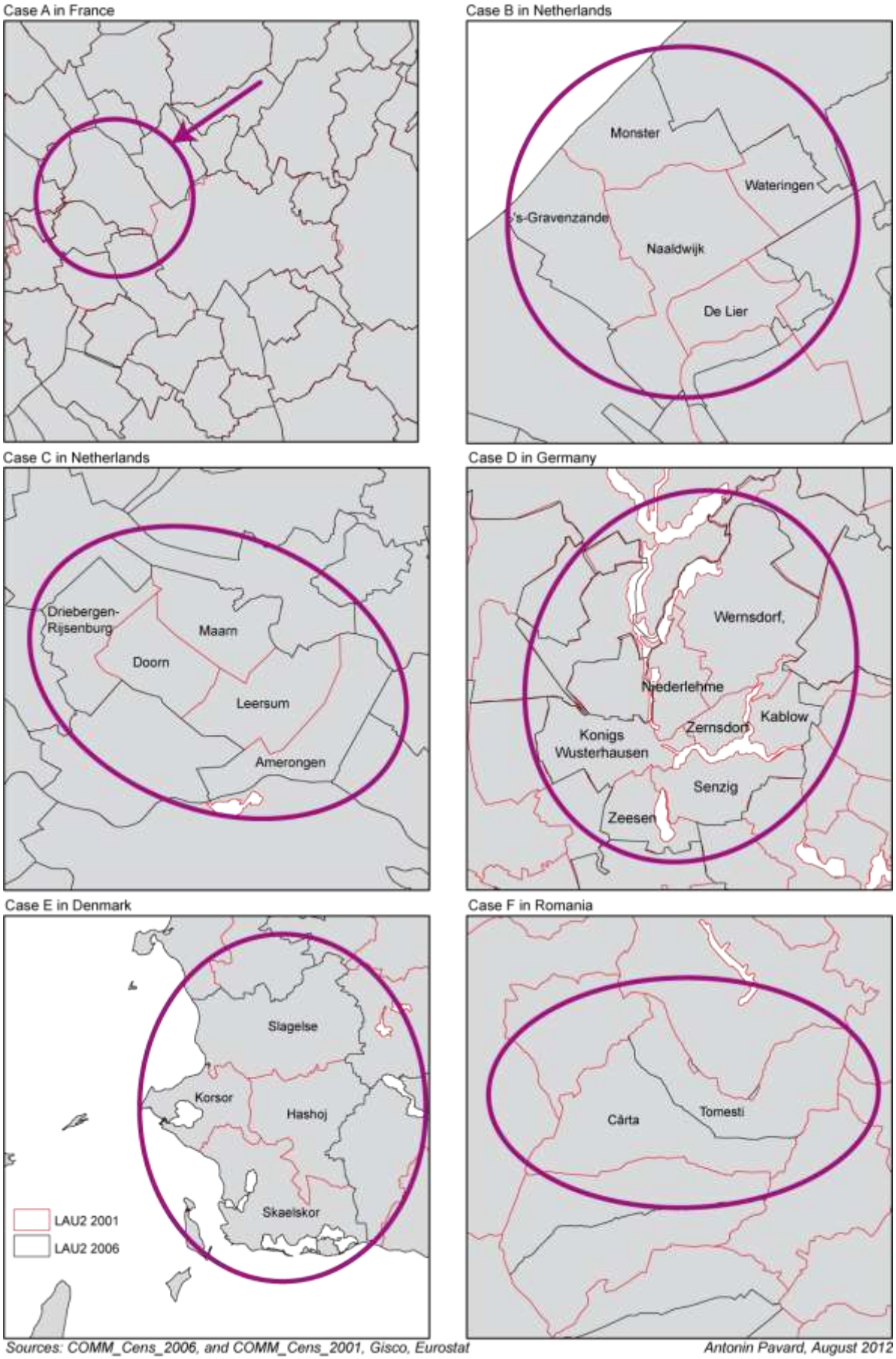
In Netherlands (case B and C), there was an administrative reform between 2004 and 2006. Several municipalities were merged to obtain larger municipalities. Case B illustrates the creation of the municipality of Westland from five municipalities (De Lier, Monster, Naaldwijk, Wateringen, and Gravenzande), in 2004. Case C can be illustrated by the creation of the municipality of Utrechtse from five municipalities (Amerongen, Doorn, Driebergen-Rijsemburg, Leersum, and Maarn), in 2006.

In Germany (case D), there was an administrative reform in 2003. Some municipalities were created by merging several municipalities. Case D can be illustrated by the creation of the municipality of Königs Wusterheursen from seven municipalities in 2003.

In Denmark (case E), there was a municipal reform between 2006 and 2007. The municipalities were merged for creating larger ones. For example, the municipality of Slagelse was created from four municipalities (Korsor, Hashoj, Skaelskor, and Slagelse).

In Romania (case F), there was a municipal reform in 2004. Few municipalities were divided into two units. For example, the municipality of Cârta was divided into Cârta and Tomesti.

Figure 3: Several examples of differences between 2001 and 2006 LAU2.



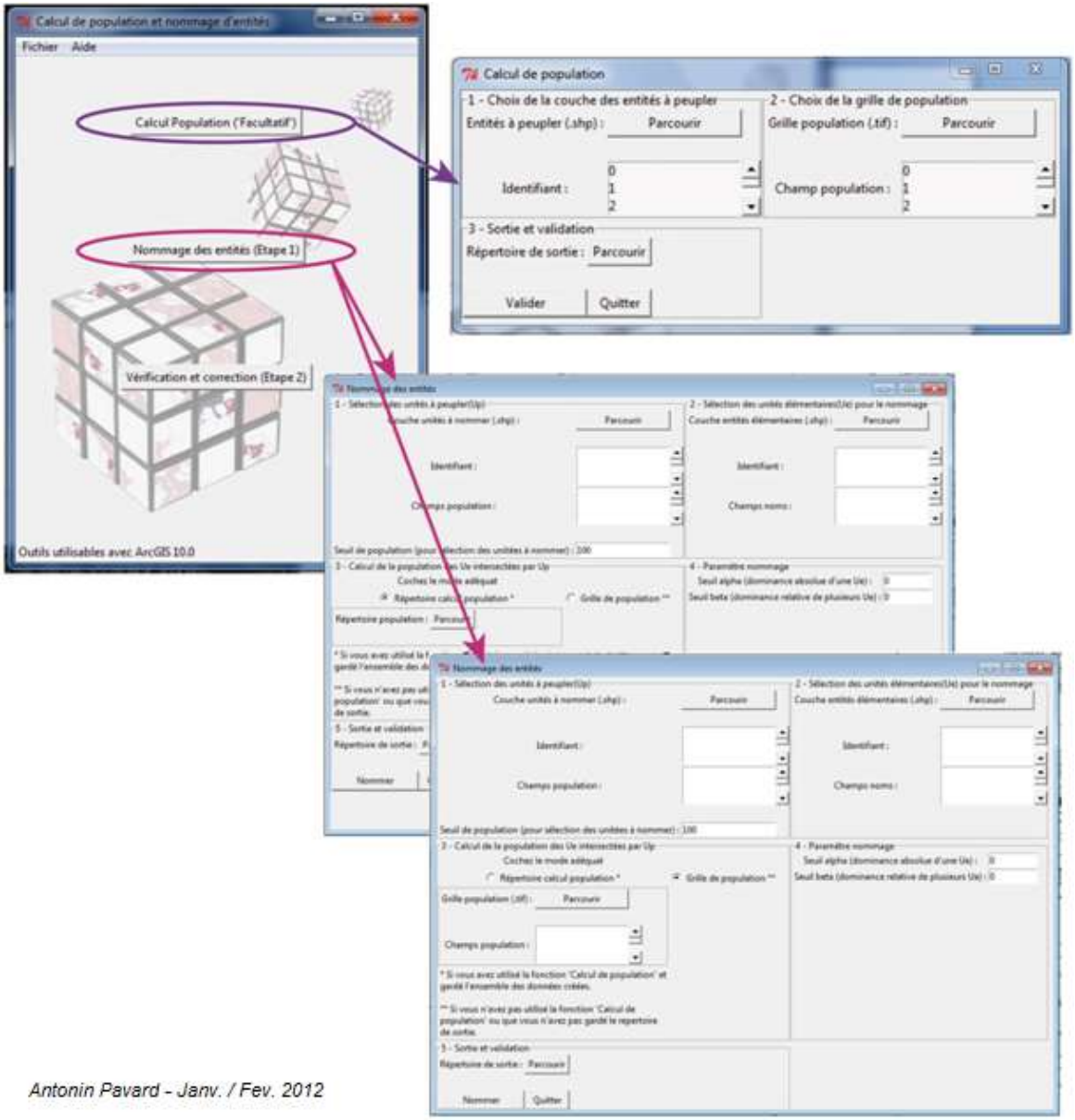
Sources: COMM_Cens_2006, and COMM_Cens_2001, Gisco, Eurostat

Antonin Pavard, August 2012

1.3 Improvement of the automatic algorithms

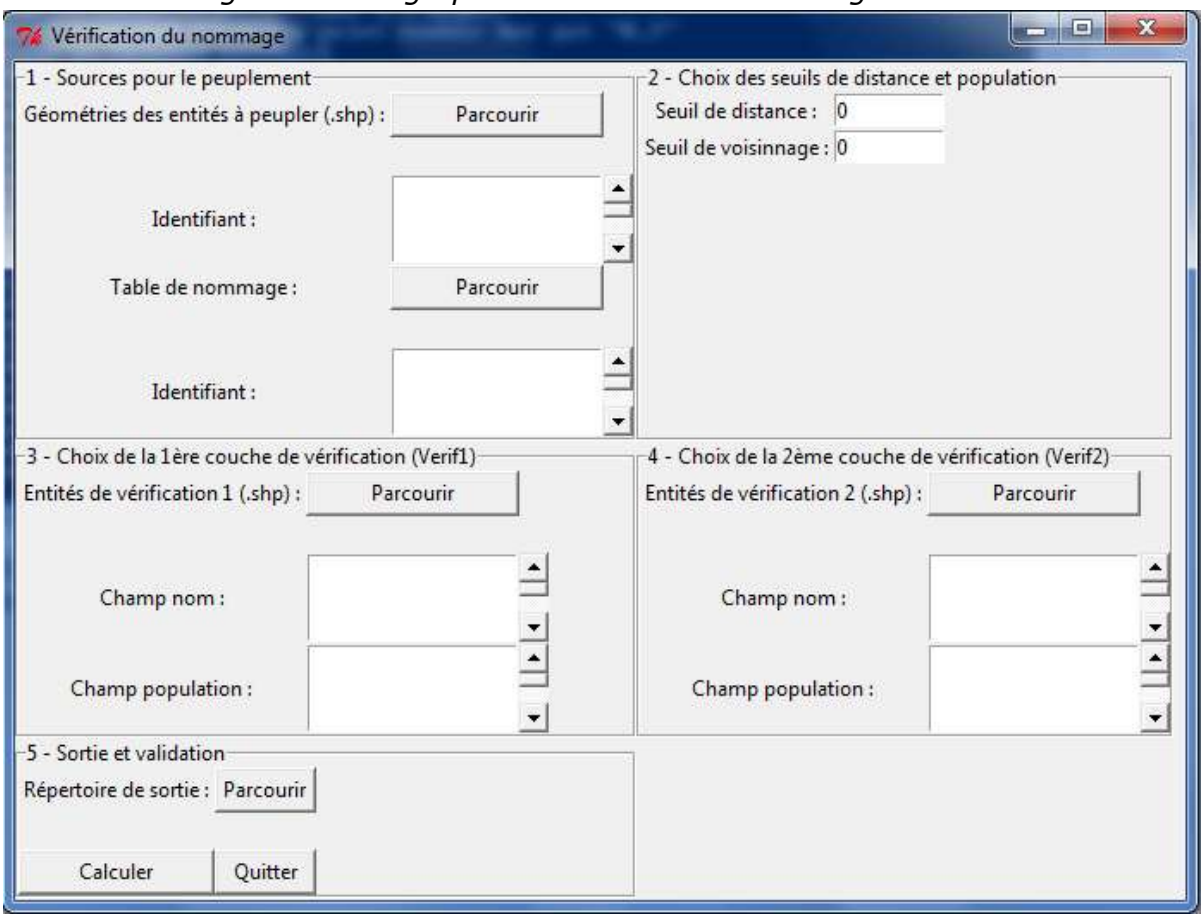
The automatic algorithm used for giving the names as well as the procedure for checking the results have been improved in four ways. Instead of using different computing languages and environments (like Python, Post Gray and Post Gis), the procedure now only uses the Python language and the Arc Gis tool. Furthermore, we have simplified this procedure by creating a graphical interface (Figure 4), that allows for choosing the geometric layers, the parameters, the directory files etc. In particular, the choice of a specific version of LAU (for example LAU2 2001 or LAU2 2006) may bring very different resulting names for countries that have dramatically reformed their administrative boundaries, such as Poland.

Figure 4: The graphical interface for the UMZ naming algorithm



Lastly, the procedure for checking results by comparing the names with Eurostat database (*Geographical names: Settlements*) and Geopolis database (F. Moriconi-Ebrard 1994) has also been automated and has been simplified by the construction of a graphical interface (Figure 5). Two principles are used for the data checking: the research of the closest cities around each named UMZ (the distance thresholds are chosen by the interface user) and the comparison between the closest cities names and the UMZ names (the Levenstein distance is used and consists in calculating the number of modifications – replacement, removal or displacing of letters – necessary to go from the UMZ names to the Eurostat or Geopolis ones).

Figure 5: The graphical interface for checking the UMZ names



Finally, the whole chain of procedures necessary to give a name and check the results only takes now a couple of hours, instead of two days. This huge reduction of time makes this tool very useful and efficient for the future versions of UMZ database (for example the 2006 UMZ) or for naming UMZ smaller than 10 000 inhabitants (which is the current minimal threshold of population in the ESPON Database). Let us also notice that these tools have been conceived as generic and could be used for naming other types of urban objects. The different menus can be translated in English easily.

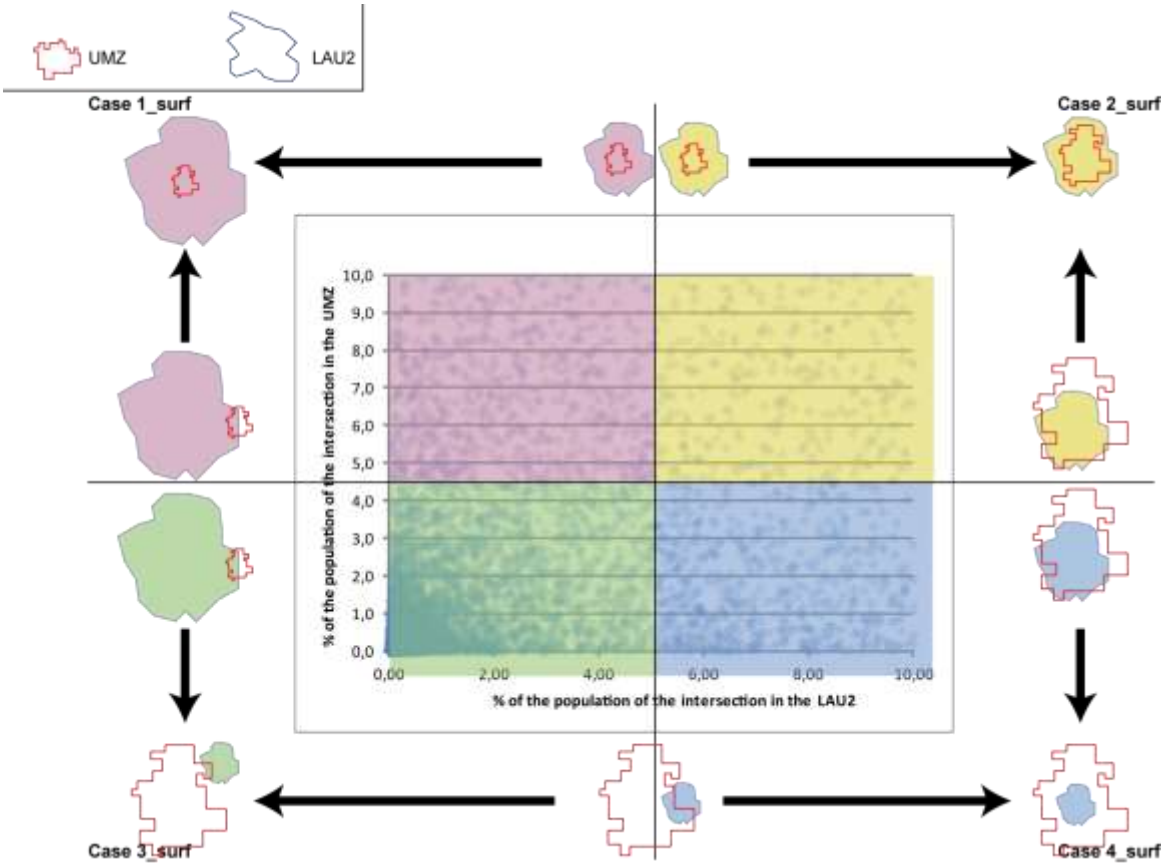
2 Creating a dictionary of correspondence with LAU2

2.1 A generic methodology

The UMZ delineations are not based on a collection of LAU2 but on a collection of pixels, as they are created from CORINE Land cover. In order to facilitate the use of socio-economic indicators based on LAU2 (for instance the SIRE database, Eurostat or national census), we have worked on a generic methodology to create a dictionary of LAU2. This dictionary allows creating a simplified version of UMZ database, called UMZ_LAU2, that gives, for each UMZ, the collection of LAU2 whose aggregation contains this UMZ: one single LAU2 if it is an "isolated city", several LAU2 if it is a "multi-LAU2 city". The LAU2 version that has been used here is "Comm_census_2001", from Gisco 2001. In order to simplify the procedure, we have aggregated all the UMZ inside one LAU2 when they do not cross other LAU2. The total number of UMZ is then reduced from 4185 to 4070.

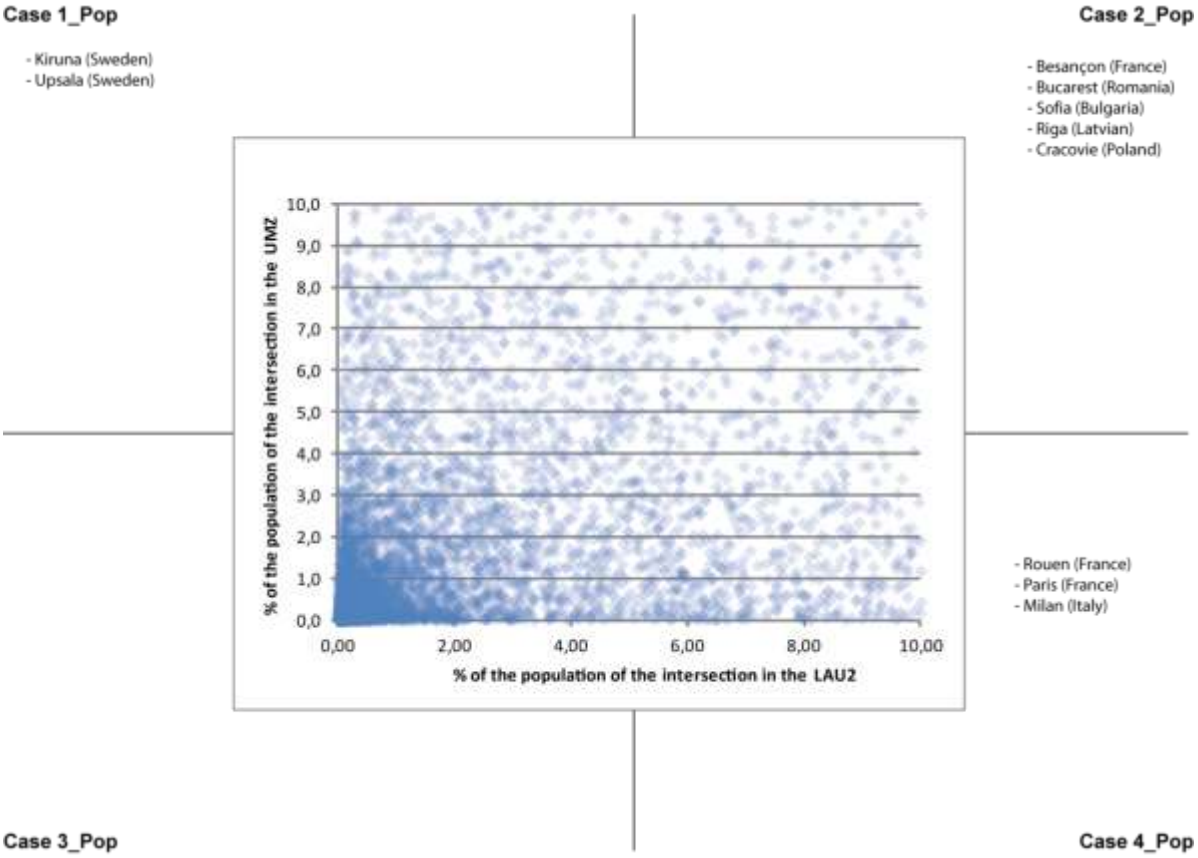
The first step consists in intersecting the geometries of UMZ and LAU2 and computing, for each resulting area, the population and surface. A total of 23257 intersections have been considered for the UMZ larger than 10 000 inhabitants. Figure 6 and Figure 7 enlighten the variability of the resulting configurations (share of UMZ surface versus share of LAU2 surface, in Figure 6, and share of UMZ population versus share of LAU2 population, in Figure 7). Let us precise that the surface approach is just given here as an illustration of the nature of intersections between UMZ and LAU2. The following dictionaries are all based on population qualification of intersection and not on surface ones.

Figure 6: Qualifying the intersections between UMZ and LAU2 in terms of surfaces



Legend:
Case 1: UMZ surface much smaller than LAU2 surface
Case 2: UMZ surface very close to LAU 2 surface
Case 3: The intersected surface between UMZ(s) and LAU2(s) is small
Case 4: The intersected surface between UMZ(s) and LAU2(s) is large
The interoperability is high for case 1 and 2, whereas it is more complex in case 3 and 4.

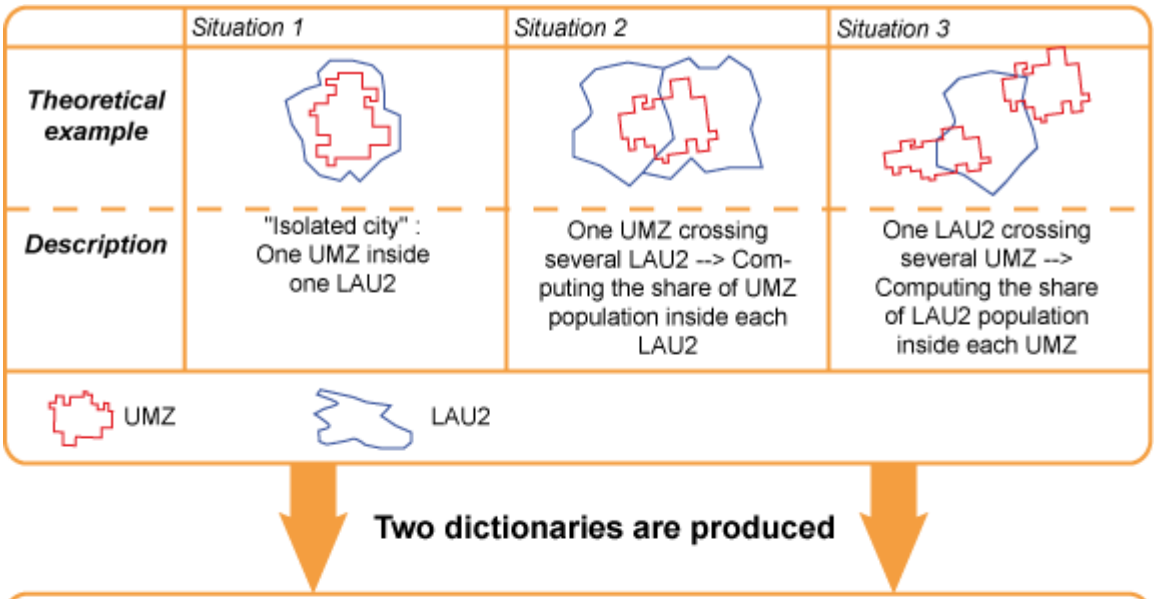
Figure 7: Qualifying the intersections between UMZ and LAU2 in terms of populations



Legend:
Case 1: UMZ population much smaller than LAU2 population
Case 2: UMZ population very close to LAU 2 population
Case 3: The intersected population between UMZ(s) and LAU2(s) is small
Case 4: The intersected population between UMZ(s) and LAU2(s) is large
The interoperability is high for case 1 and 2, whereas it is more complex in case 3 and 4

The complexity of the cases 3 and 4 has leaded us to propose two different final products. The first one, called **UMZ_LAU2_Restricted**, is constructed in a very simple way and its use is very easy, but it is relatively poor regarding the question of harmonization as it does not give the possibility for the user to change any threshold. The second one, called **UMZ_LAU2_Complete**, gives much more fields and possibilities to the user, who can, for example, vary himself the thresholds in order to adapt them to the average size of LAU2 per country, or in order to study the different configurations of intersections that characterize, for instance, polycentric areas or conurbations. The different variables of these two databases can be separated into three main fields (Figure 8), that are described in the following section.

Figure 8: Spatial configurations and main situations in the UMZ_LAU2 database



Two dictionaries are produced

Simple : UMZ_LAU2_Restricted *(selection of LAU2 for which at least 50% of the population is laying on the UMZ intersection + allocation of one LAU2 to the major UMZ when it crosses several UMZ)*

ID_UMZ_Agr	Country	NAME_Up	ID_LAU2	Type_UMZ	P_Pop_Int_Lau2 (>50%)	Nb_LAU	Flag_Inf_30
1	AT	Ansfelden	AT41002	0	95,25	1	0
2	AT	Bad Ischl	AT40703	0	92,39	2	0
2	AT	Bad Ischl	AT50336	0	56,05	2	0
3	AT	Bad Voslau // Kottingbrunn	AT30603	0	88,41	4	0
3	AT	Bad Voslau // Kottingbrunn	AT30618	0	95,40	4	0
3	AT	Bad Voslau // Kottingbrunn	AT30620	0	92,84	4	0
3	AT	Bad Voslau // Kottingbrunn	AT30635	0	92,89	4	0

ID_UMZ_Agr : *ID of the aggregated UMZ*
Country : *Country code of the UMZ*
NAME_Up : *Name of the most populated UMZ among the aggregated one*
ID_LAU2 : *ID of the LAU2 (COMM_Census2001)*
Type_UMZ : *Type of the UMZ (Isolated city or multi-LAU2)*

P_Pop_Int_Lau2 : *Share of the intersection population among the total LAU2 population*
Nb_LAU2 : *Number of LAU2 intersected by the UMZ*
Flag_Inf_30 : *Boolean information based on the field "P_Pop_Int_LAU2" (If value is < 30% : flag = 1 ; if value is >=30% : flag = 0). Concerns isolated cities.*

Elaborate : UMZ_LAU2_Complete

ID_UMZ_Agr	Country	NAME_Up	ID_LAU2	Type_UMZ	P_Pop_Int_Lau2	P_Pop_Int_UMZ	Nb_LAU2	Flag_Inf_30	Classif_P_Pop	RankInf50	Flag_Sup50	RankSup50	Rank_LAU2_Sim
1	AT	Ansfel	AT41002	0	95,25	97,74	3	0	1	1	1	1	1
1	AT	Ansfel	AT40101	0	0,18	2,27	3	1	10	2	0		2
1	AT	Ansfel	AT41019	0	0,00	0,00	3	1	10	3	0		1

P_Pop_Int_UMZ : *Share of the intersection population among the total UMZ population*
Classif_P_Pop : *Classification of the intersections according the field "P_Pop_Int_LAU2"*
RankInf50 : *Rank for each intersection in the UMZ according the field "P_Pop_Int_UMZ"*

Flag_Sup50 : *Boolean information based on the field " P_Pop_Int_LAU2" (If value is < 50% : flag = "" ; if value is >=50% : flag = 1).*
RankSup50 : *Rank for the intersections in the UMZ according the field "P_Pop_Int_UMZ" and if the Sup50 = 1*
RankLAU2_Sim : *Rank for UMZ crossing one LAU2 relative to the share of population*

2.2 Three different spatial configurations

2.2.1 The "isolated cities"

The first situation of Figure 8 corresponds to the cases 1 and 2 of Figure 7 (one UMZ is laying inside one LAU2). It concerns a total of 1091 UMZ that can be called "isolated cities". Of course, this result depends mainly on the mean size of LAU2, by country (see the difference between France and Great Britain and Sweden, Germany or Baltic countries on Table 8 or Figure 10) or even inside one country (for instance, the LAU2 are very different in size in Italy, very large along the coasts and smaller in the central regions). In the database UMZ_LAU2, there is a column indicating if the UMZ is an "isolated city" or a "multi-LAU2 city".

Figure 9: UMZ laying inside one LAU2 ("isolated cities")

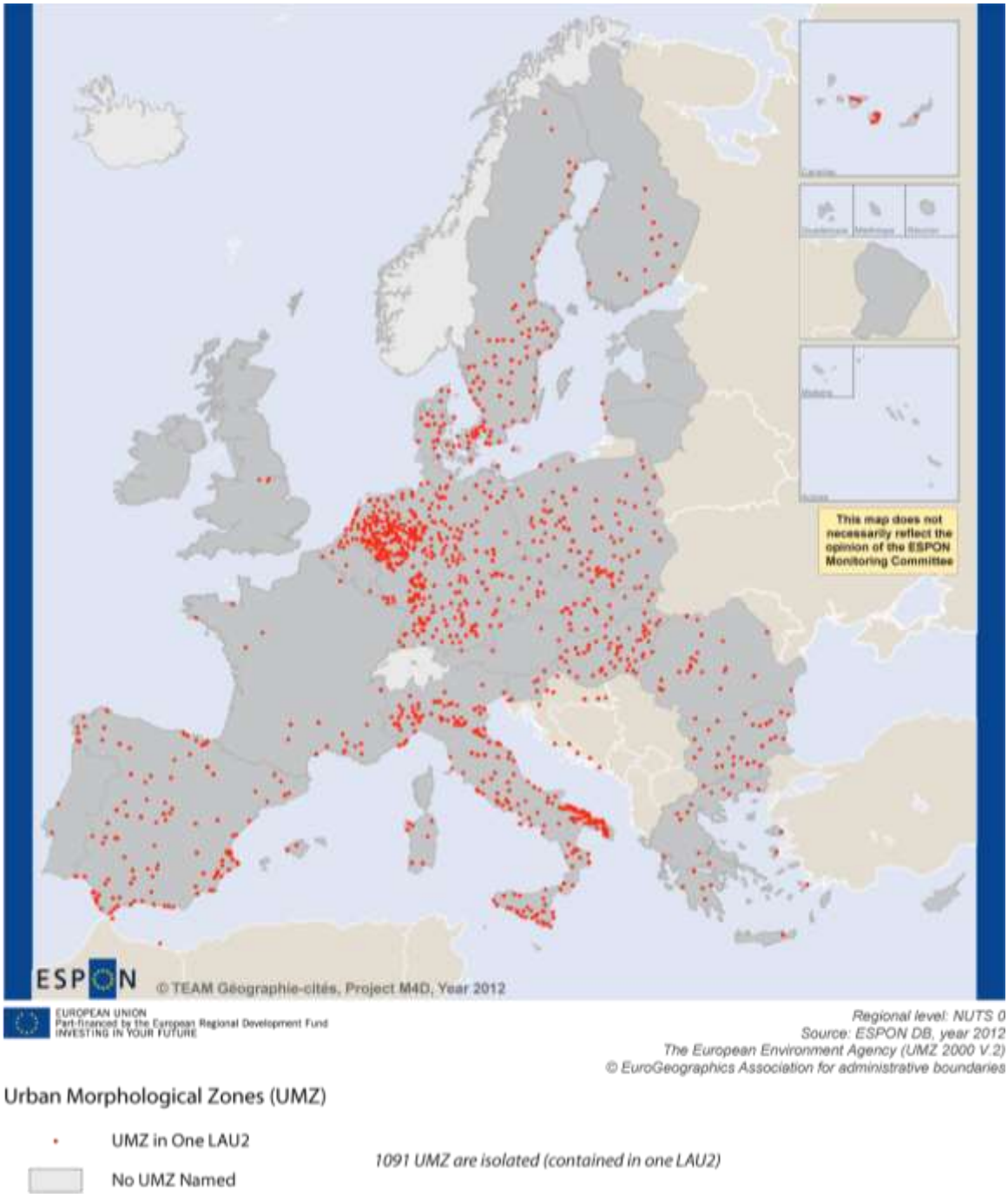


Figure 10: Number of "isolated cities" and average size of LAU2 per country

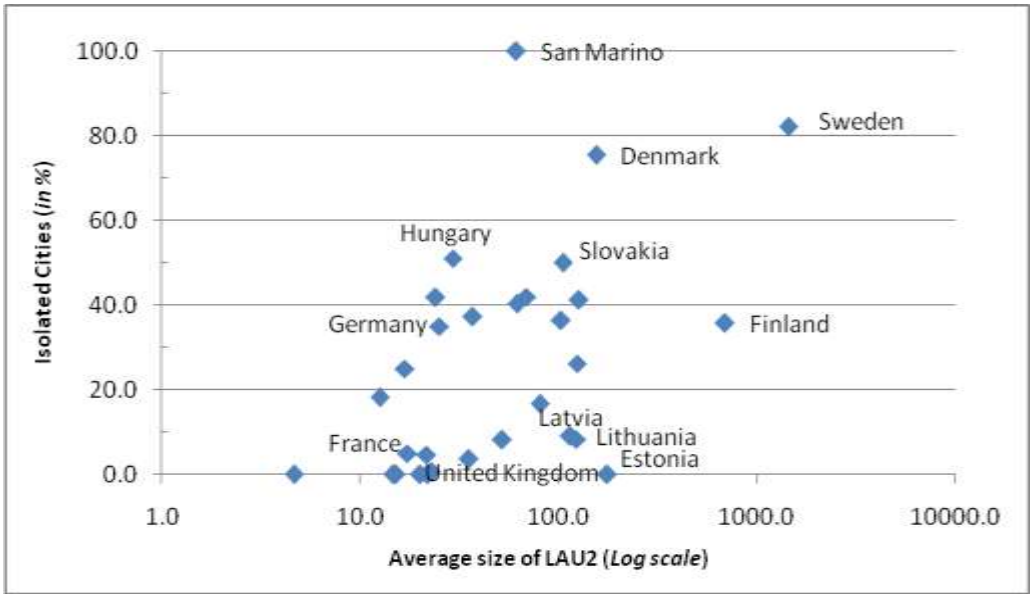


Table 8 : Number of « isolated cities » and average size of LAU2 per country

Ct	Av_Nb_LAU2_By_UMZ	Nb_Isolated_Cities	Ave_Size_LAU2	Nb_UMZ	P_Isolated_Cities
SE	2,9	69	**1444,1**	84	**82,1**
FI	3,4	15	693,2	42	35,7
EE	3,6	0	175,4	13	0,0
DK	7,8	37	154,8	49	75,5
GR	33,2	21	126,9	51	41,2
PL	5,3	81	124,2	310	26,1
LT	6,1	1	**123,3**	12	**8,3**
LV	4,2	2	**114,3**	22	**9,1**
SI	2,3	6	**105,5**	12	**50,0**
HR	3,3	12	103,1	33	36,4
RO	5,3	25	80,7	150	16,7
NL	6,0	78	69,1	186	41,9
ES	7,4	125	61,6	311	40,2
SM	1,0	1	61,0	1	100,0
BE	81,9	4	52,0	49	8,2
IT	49,8	199	37,0	535	37,2
AT	12,4	2	35,4	56	3,6
HU	7,6	49	**29,4**	96	**51,0**
DE	7,4	276	25,2	789	35,0
BG	4,1	28	23,9	67	41,8
UK	149,0	3	22,8	502	0,6
LU	10,0	0	22,0	2	0,0
PT	51,0	3	21,7	64	4,7
IE	182,3	0	20,1	24	0,0
FR	76,2	19	17,4	380	5,0
SK	5,4	16	16,7	64	25,0
CY	12,0	0	15,1	5	0,0
LI	7,0	0	14,6	1	0,0
CZ	5,7	19	12,6	105	18,1
MT	47,0	0	4,6	1	0,0

2.2.2 One UMZ intersects several LAU2

When one UMZ is crossing different LAU2s (Figure 8 situation 2), different choices are possible. The most simple consists in identifying only one LAU2, i.e. the one that contains the major part of the UMZ. We have not retained this method as it prevents the study of interesting urban forms, like conurbations for example (several urban cores historically well differentiated and laying in different LAU2 but latterly joined by urban sprawl into one single morphological patch). A more sophisticated solution consists in selecting several LAU2s until we exceed a certain share of the UMZ population, for example 50% or 75%. We have not retained this second method as it may lead to select some LAU2 that represent a very few part of the UMZ.
We have then adopted a third solution, which is currently used, for example in the cores of the new LUZ, or for the French urban areas ("unités urbaines"). It consists in selecting the different LAU2s for which more than a certain threshold of population (generally 50%) lay inside the UMZ. For the UMZ_LAU2_Restricted database, we have adopted the threshold of 50%. But for the UMZ_LAU2_Complete database, and considering the large heterogeneity of the average size of LAU2 per country, we have decided not to determine ourselves a certain threshold but to let the possibility to the user to choose this threshold. For instance, a high threshold such as 50% is more restrictive for the selection of LAU2 than a low threshold such as 40% (201 UMZ would be excluded with this high threshold, especially in Germany, Italy, Netherlands, Poland, Spain and Sweden). It would thus create more cities linked to only one LAU2 and, by this way, there would be more "rural" LAU2 in Sweden. The user can then choose to modify the threshold according to the average size of LAU2 in each country. Another possibility is to adopt the same threshold, but the user should then keep in mind this heterogeneity when interpreting, for instance, results coming from socio-economic indicators: if the UMZ only represents 40% of the LAU2 population, the use of indicators collected for this LAU2 should be done with caution. In the database, we have added a column with a flag in case the threshold in lower than 40%. We have also added a column with a discretisation of the shares in ten classes (from 0 to 10%, from 10 to 20% etc.).

2.2.3 One LAU2 intersects several UMZ

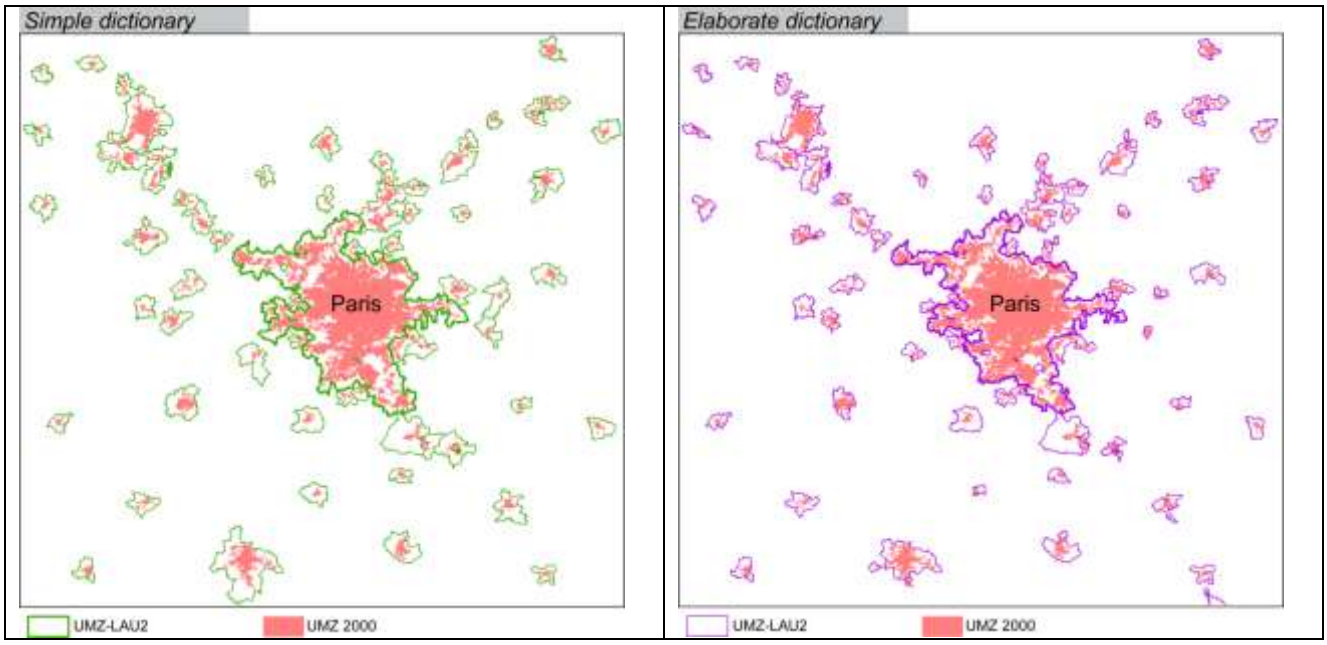
When one LAU2 is crossing different UMZ (Figure 8, situation 3), we have a particular case of the situation 2 that raises the question of "LAU2 multiple belonging" (one LAU2 that may belong to two or more different UMZ_LAU2). This case is not very frequent (it concerns less about 7% of the total number of LAU2 intersecting UMZ) but is very concentrated in certain areas in Europe, characterized by large LAU2 and/or dense urban settlement patterns, like Germany, Belgium, Netherlands, Italy or Spain. In each of these countries, some LAU2 may intersect 5, 6 or 7 different UMZ. Different choices are possible for managing this case. The most simple, selected for the UMZ_LAU2_Restricted database, consists in retaining only the UMZ that concentrates the largest share of LAU2 population, which relates directly to the situation 2 (if the share of LAU2 population inside this UMZ intersection is larger than 50%, the LAU2 is retained). For the UMZ_LAU2_Complete database, we have prepared three indicators that allow the user to make its own choices. First, the different UMZ that intersect the LAU2 are ranked according to the share of intersected population, so that the user may

choose only the the main UMZ (as in the UMZ_LAU2_Restricted version) or retain, for example, the two largest intersecting UMZ etc. Secondly, we make apparent the particular threshold of 50% as an under-category of the ranking (Rank1_>50% or Rank1_<50%) in order to allow the user selecting the largest intersecting UMZ whether its share is larger than 50% (Rank1_>50%) or not (Rank1_<50%).

2.3 An illustration: the case of Paris

In order to better illustrate the differences between the two dictionaries (the restricted and the complete one), we have mapped the resulting perimeters for the case of Paris (Figure 11). In the left part of the figure, the complete dictionary of correspondence has been used. The collection of LAU2s that are retained for Paris are not only included inside the UMZ but are also located around the UMZ, intersecting the limits, even if the intersected area is very small. The final perimeter is underlined in bold green in the figure. In the right part of the figure, the restricted dictionary of correspondence has been used. The collection of LAU2s that are retained for Paris is smaller than in the previous case. It includes not only the LAU2s located inside the UMZ but also the ones located at the limits, but only when 30% or more of their population is laying inside the UMZ. The final perimeter is underlined in bold violet in the figure.

Figure 11: Paris city, from the UMZ to a collection of LAU2 (left: using the complete dictionary; right: using the restricted dictionary)



3 Populating UMZ with local data

We present a generic methodology for populating urban databases with local data, applied to the case of Urban Morphological Zones and LAU2. The dictionary of correspondence between UMZ and LAU2 allows populating this data base with local data, as it gives the composition of the UMZ in terms of LAU2. The intensity of each link LAU2-UMZ may be characterized by two attributes, the share of the LAU2's area that is included in the UMZ (area's contribution), and the share of the LAU2's population that is included in the UMZ (calculated with the JRC population density grid). This will be called respectively LAU2 area contribution and LAU2 population contribution.
We present in a first part the methodology for populating the database as well as a validation procedure. In a second part we propose an illustration of its interest but also of potential problems of completeness, due to missing or inconsistent values in the SIRE database.

3.1 Methodology: illustration of the different steps based on the example of SIRE Database

3.1.1 Join between the correspondence table and attributes

We have first realized a join between the attributes of SIRE database and the correspondence table based on LAU2 ID.
We have then identified the unmatched records from one side and the other. Two methods have been used to reduce these cases:

- Codes: work on the correspondence between SIRE LAU2 ID and UMZ LAU2 ID (for instance, adding or retrieving a 0 in the code is sometimes enough to allow the matching)
- Names: using the names instead of the ID for improving the matching of SIRE and UMZ LAU2.

Finally, only one hundred LAU2 on a total of 23 021 did not match.

3.1.2 Allocation and aggregation of SIRE data in each UMZ

We have allocated SIRE data according to the intensity of each link LAU2-UMZ. Two different methods were tested, the allocation according to the LAU2 area contribution and the allocation according to the LAU2 population contribution. The first one gave too much incoherent results, possibly because of the heterogeneity of the LAU2 area sizes in Europe. We retained the LAU2 population contribution method.
In order to aggregate SIRE data into each UMZ, three different cases were considered.

- One UMZ is included inside one LAU2. Then, the previous step is enough for getting the data at the scale of the UMZ
- Different UMZ are included inside one LAU2. For instance, in the case of Roma, in Italy, 11 UMZ lay inside the LAU2 of Roma. Then, we have added

the total population of the included UMZs and the allocation/aggregation was done on the basis of this total population
- The other UMZ. In this case, the allocation/aggregation was done on the basis of each LAU2 population contribution to the UMZ and we have aggregated all these contributions.

We have then identified, for each UMZ, the number of unmatched LAU2.

3.1.3 Verification of the results

We have first chosen some indicators for the tests, and decided to select a very simple variable, the population. This indicator is present in the SIRE database and can be also obtained by the population density grid (JRC).
We have then compared the results obtained with the SIRE database by aggregation of the LAU2 share of population (previous procedure) and by aggregation of the cells of the JRC population grid. For this comparison, we have chosen a tolerance threshold, the one of 10%[2].
The results show that only 95 UMZ (out of a total of 4304 UMZ) present deviations exceeding this threshold. These results were expected for the 45 UMZ that are concerned by missing LAU2. For the 50 remaining ones, it is to be noted that most of the deviations are very concentrated near 10%. However, there are some exceptions that are, for most of them, due to some inconsistency in the density grid. For instance in Latvia (for unknown reason), or for UMZ located at the frontier with Switzerland where there is no data in the density grid. For these 95 UMZ, we have added an indicator of validity that indicates to which extent the data coming from SIRE database have to be considered with caution.

3.2 SIRE database: problems due to data completeness and consistency

3.2.1 Selection of different indicators in SIRE database

Different indicators have been selected for testing the populating methodology. We present here some conclusions regarding their level of quality, considering their potential use for populating UMZ database.
- **Total population**: the data seem to be coherent and have been used to populate the UMZ database (see above).
- **Commuting** data: we have presented these data and the problems we met with them in the June 2012 deliveries[3].

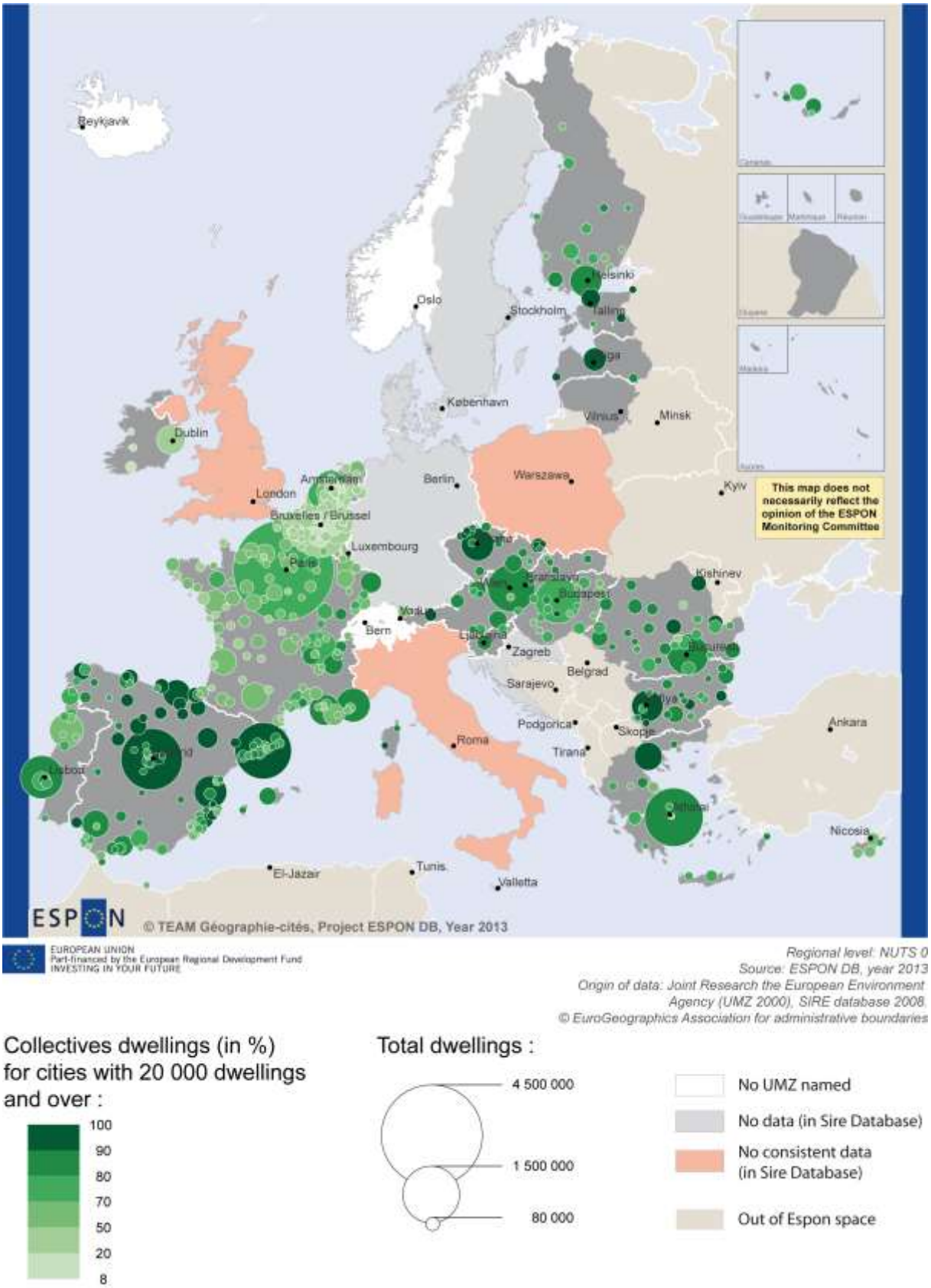
[2] It corresponds more or less to the range delimited by the quartiles of the statistical distribution of the deviation between the two measures. The deviation is computed as the ratio (in %) between the difference between JRC population and SIRE population obtained by allocation/aggregation, and the JRC population.
[3] The data about commuters are incomplete and not harmonised. A test for the Paris zone has shown that SIRE information did not completely match the INSEE national information about commuters. In particular, the lowest flows from each LAU2 are not documented and the threshold used to select data varies from one LAU2 to another.

- **Total population per age class**: the data seem to be coherent and have been used to populate the UMZ database (see below).
- **Level of education**: the data have been used but the results are not good. There are no data for Germany and Lithuania, and no consistent data for Sweden, Portugal, Italy, Greece, Austria, Bulgaria, Croatia, Ireland, Denmark, and Czech Republic (the total population is different from the addition of the population of the different classes)
- **Individual housing and collective dwellings**: the data have been used but the results seem to be incoherent for some countries (see below).

3.2.2 An example of problematic indicator: collective dwellings

Starting from the SIRE indicators related to housing, we have chosen to represent the share of collective dwellings in European cities (Figure 12). The map enlightens two types of difficulties. First, there is a problem of completeness in SIRE database. Two countries (Germany and Sweden) did not send any data for collective dwellings. Secondly, there is a problem of consistency, easy to detect here but sometimes much more difficult to identify. In three countries, like United Kingdom, Poland and Italy, the share of collective dwelling is abnormally low (less than 0,7% whereas the average in the other countries is 81%).

Figure 12: Share of collective dwellings in European cities (UMZ) in 2001



3.3 A coherent SIRE indicator: age structure

3.3.1 Selection of variables

Contrary to the problematic indicators described in the previous section, the total population per age class sounds to be coherent. In the June 2012 deliveries, we computed and mapped two different indicators:

- The **old-age dependency ratio**: it indicates the relationship between the working-age population and elderly persons (see for instance Eurostat Regional Yearbook 2011 or the ESPON project DEMIFER final report).
- The **demographic ageing**: this indicator is built as the ratio between the population older than 65 years and the population younger than 25 years. The map constructed with this latter indicator has been published in the ESPON M4D newsletter (June 2013), for the Dublin ESPON Seminar.

We have then chosen to improve the analysis of the European cities by age class, by taking into account not only two specific age classes but the whole structure of the age pyramid. The statistical method generally used for constructing this typology is the cluster analysis. Eight age classes have been used:

0-14 years
15-24
25-34
35-44
45-54
55-64
65-74
75 and more

3.3.2 A cluster analysis at the European scale

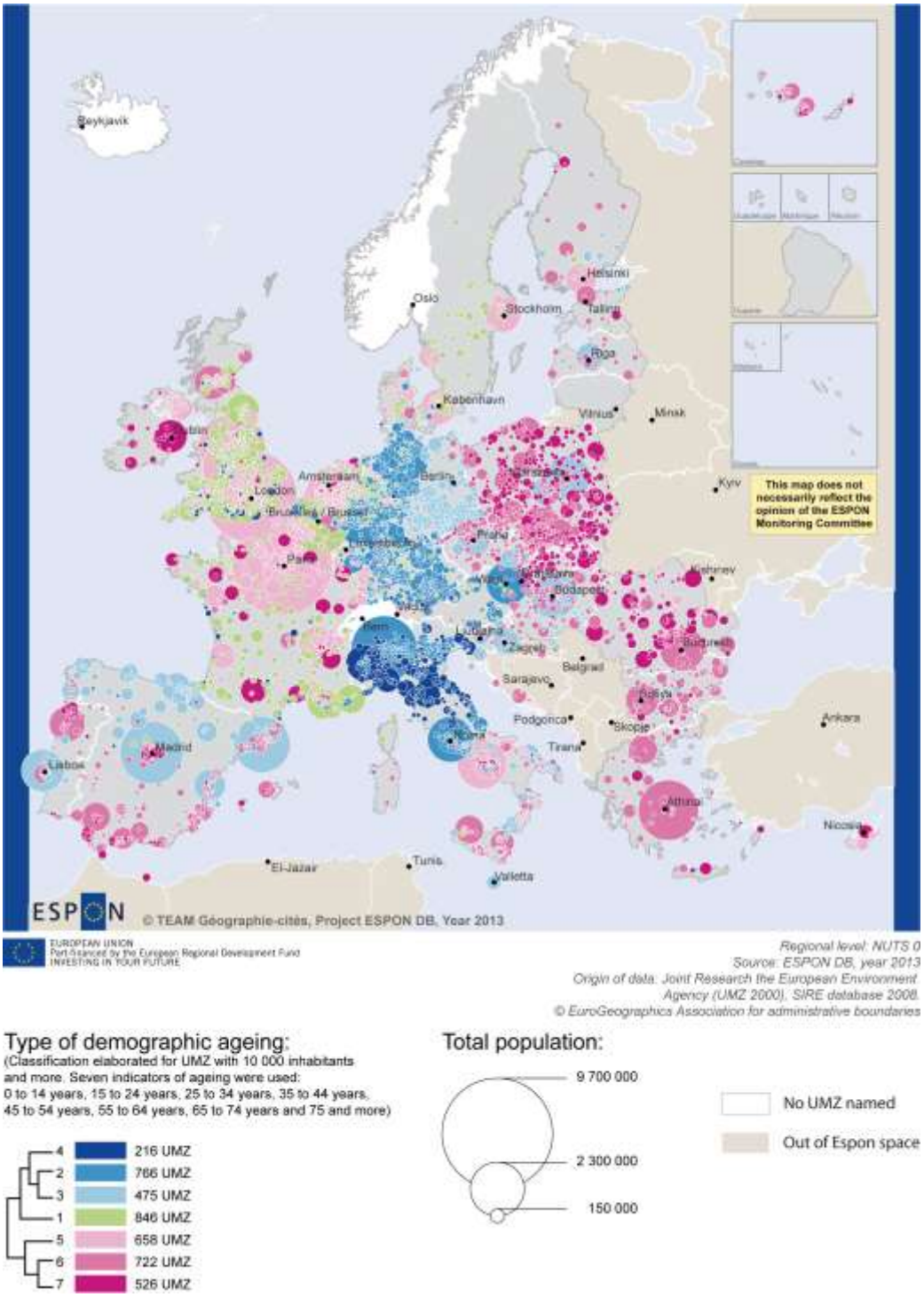
We have made a first typology of the European cities according to the age structure considering all the UMZ larger than 10 000 inhabitants. The cluster analysis has been built with 7 classes (Figure13). Cities with blue and green colours are characterised by ageing population, especially the ones with dark blue colour. At the opposite cities with rose colours are characterized by young populations, especially the ones with dark rose colour. The different profiles are represented by age pyramid schemas: for each pyramid, we read the European profile (average) at the left and the specific class profile at the right. In order to render easier the reading of the pyramids, the over-representations of some age classes in each pyramid are underlined in bold.

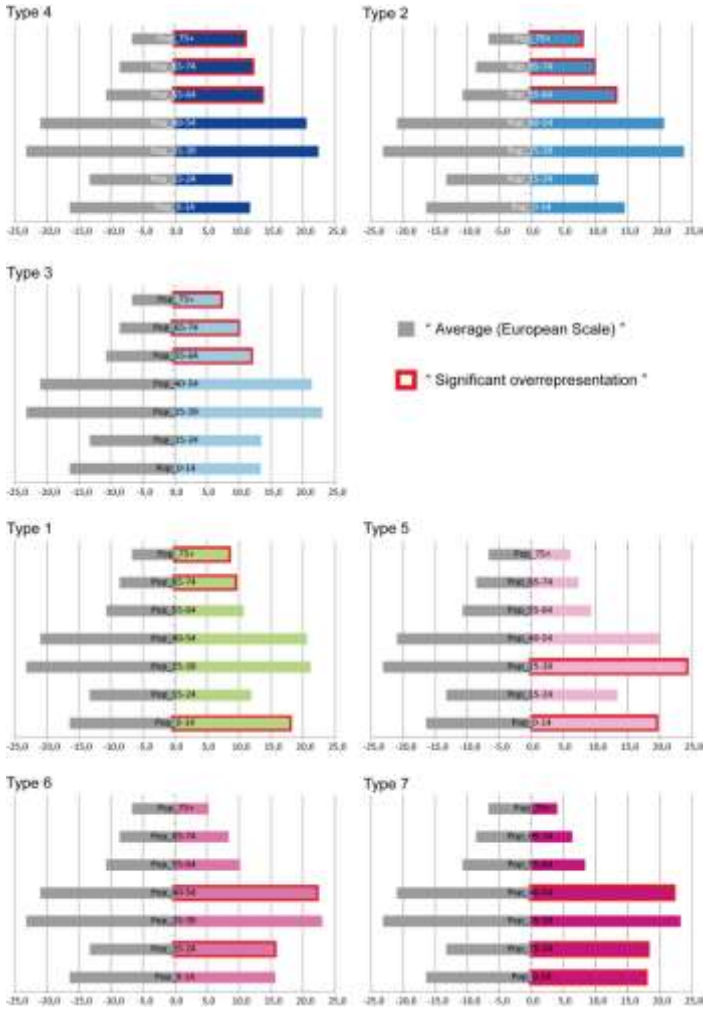
The map enlighten a very clear opposition between three types of regions:

- **Ageing** ones (Germany, Austria, northern Italy & Spain)
- **Intermediate** ones (UK, France, Belgium, Netherlands, northern Europe)
- **Young** ones (Central & Eastern Europe, southern Italy & Spain, Greece, Ireland)

Let us notice than the age class indicator is generally used at regional level and not at city level. This latter scale of analysis allows a very fine and detailed information, making clear some age contrasts between Rivieras and the rest of countries (see for instance in the east-south of France) or between some regions inside countries (see the north-south age contrasts in Portugal, Spain or Italy).

Figure 13: Typology of age structure by city at European level and class profiles

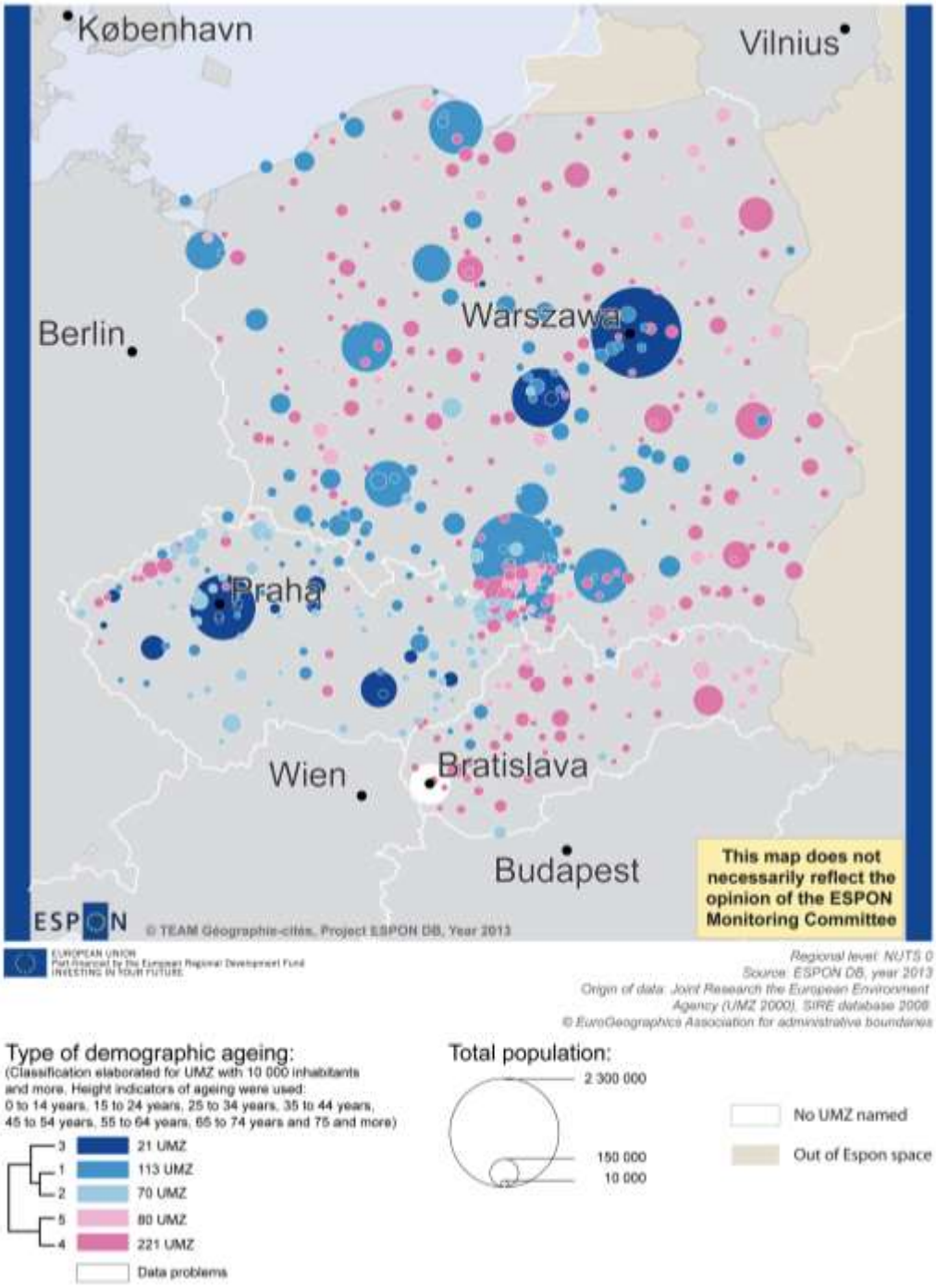


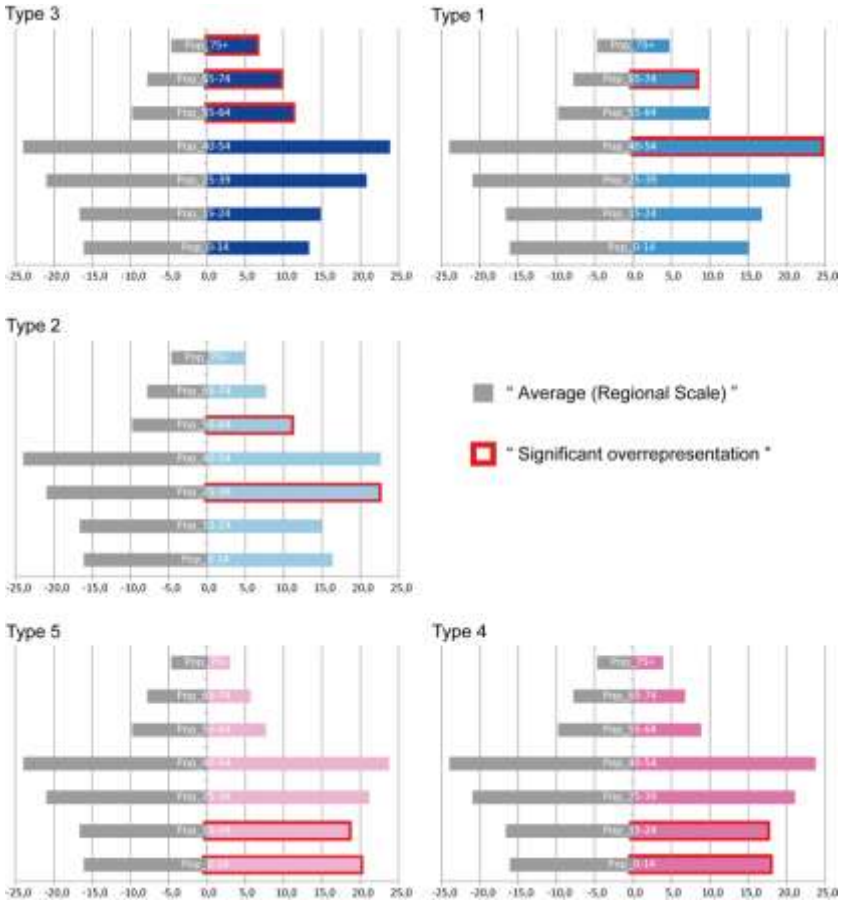


3.3.3 A cluster analysis at a regional scale

When we consider the typology at the European scale, some macro-regions seem to follow very similar profiles, such as the Central and Eastern countries, that appear as "young" comparatively to the rest of Europe. It may be interesting to deeper the analyse by changing the reference scale and by zooming on these countries. We have realized a second cluster analysis, with the same 7 age classes but only for three countries, Poland, Czech Republic and Slovakia (Figure 14). The results show that alongside regional contrasts between western and eastern parts of each country, we also read very clearly some city size effects. The largest cities (especially the capitals) tend to be populated by oldest inhabitants than medium cities and the smallest cities are the youngest. These spatial structures are very important to consider when analysing the current demographic trends (such as the questions of shrinking cities or shrinking regions, but also out-migrations of young people). It is not visible when the age classes are mapped by NUT2 or NUT3 nor when mapped at LAU2 level. It raises the importance of having spatial nomenclatures of urban objects and coherent LAU2 databases for populating them.

Figure 14: Typology of age structure by city at regional level (Central Europe) and class profiles




Type 3
Type 1
Type 2
" Average (Regional Scale) "
" Significant overrepresentation "
Type 5
Type 4

4 Conclusion

The approach that has been developed by ESPON Database in the two different phases of the work (ESPON DB 2013 and M4D) is based on a very coherent line that is to be recalled here.

We have insisted first on the importance of uploading in the Database several urban databases and not only one, which could be considered as "official". City is a very complex object that cannot be reduced to a sole approach and definition. One major stake for European statistics is to maintain two different types of urban databases, the ones defining cities as morphological objects (based on continuous built-up areas) and the ones defining cities as functional objects (based on commuters). We have also insisted on the fact that working at an international scale impose to define cities using harmonized criteria and not as a collection of national definitions, as it was done for example in Urban Audit 2004 for the Larger Urban Zones (see our March 2011 Technical Report on that topic).

Considering the functional urban areas (FUA), we have now the possibility for using the New Harmonized LUZ, provided by the consortium Eurostat/DG Regio/Urban Audit/OECD. However, the FUA generally relate only to the largest cities and not to the small and medium ones. For instance, there are 695 new LUZ, which means that the cities lower than 50 000 inhabitants are not represented by this database, whereas they raise crucial issues in terms of urban planning and knowledge on territorial dynamics. This is the reason why we have considered the Urban Morphological Zones with great interest. There are 4304 UMZ larger than 10 000 inhabitants, and some of them lay inside regions lacking in any new LUZ but needing the European support through attributions of structural funds.

UMZ is the only database freely available, defined at the European scale with harmonized criteria and including all the cities larger than 10 000 inhabitants. It is updated regularly by the EEA and the automation of the methods that have been constructed by ESPON Database for giving them a name render the name updates very easy and quick (see section 1 of this technical report).

The indicators available with the UMZ are interesting (population and surface) but too poor for a deep exploration of European urban structures and dynamics. This is the reason why we have investigated in methods for populating UMZ with LAU2 indicators. The methods which are described in this technical report (see section 2) are generic and can easily be transferred to other databases.

Populating UMZ with social, economic or demographic indicators suppose however to have robust and complete LAU2 databases. This is not the case with the current version of SIRE (year 2000). We have expertised the different indicators that are proposed in SIRE and retain only the age structure (see section 3 of this technical report). The results, described through a European and a regional cluster analysis, are very promising and show that there is an urgent need for improving the SIRE database in order to explore other dimensions of European cities.