EUROPEAN FORUM
FOR GEOSTATISTICS 2012

Prague Conference

24-26 October 2012
Prague, Czech Republic

Detailed population distribution maps for Europe's cities

Filipe Batista e Silva
EC, Joint Research Centre, Ispra, Italy
Hugo Poelman
EC, DG Regio, Brussels, Belgium

CZECH STATISTICAL OFFICE

EUROPEAN FORUM FOR GEOSTATISTICS

CAGI

Structure of the presentation

- **Objective and scope of the work**
- **Data used**
- **Methodology**
- **Results and validation**
- **Summing up and conclusions**

Objective and scope of the work

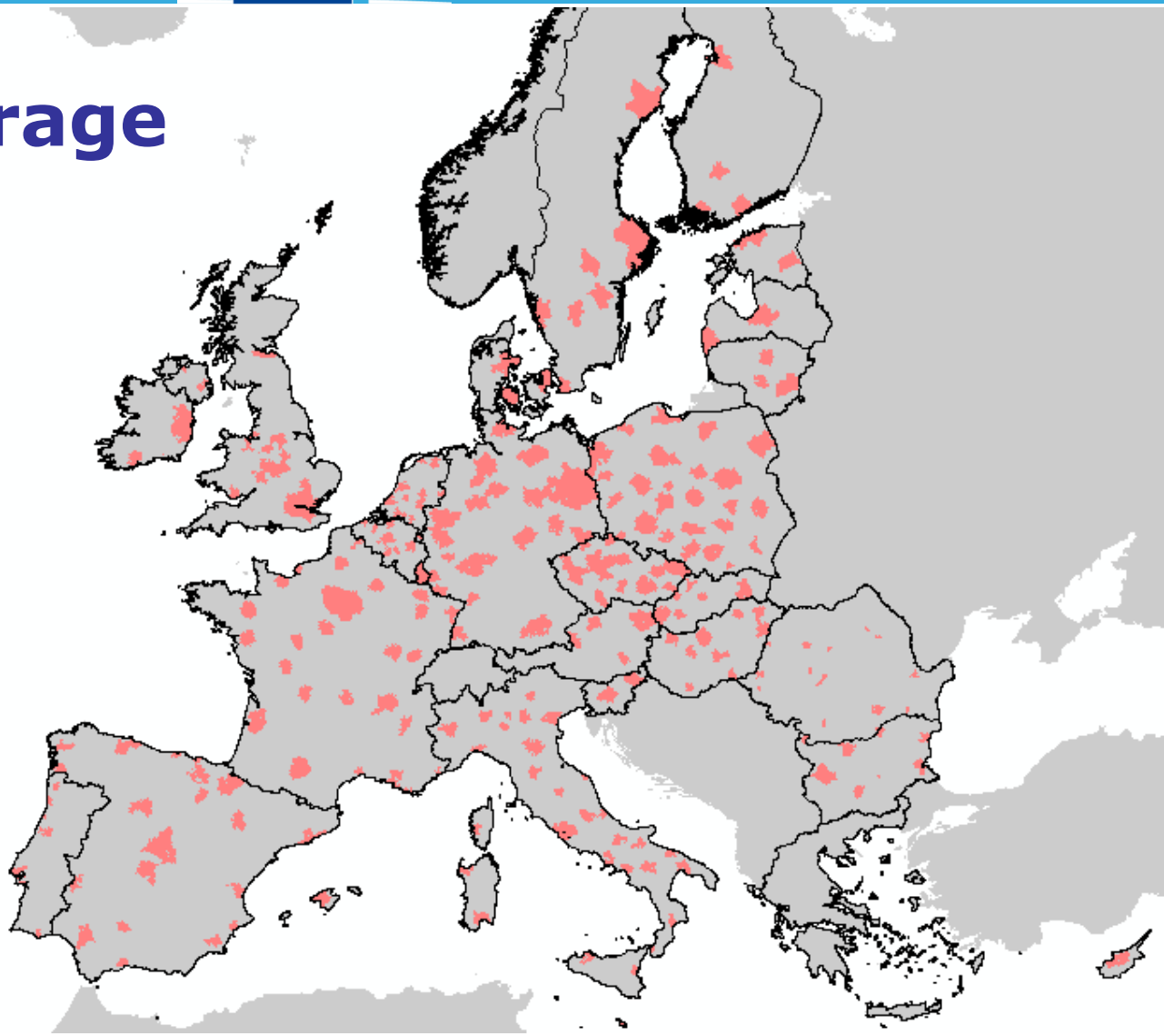
The main objective of the work was to

Leverage existing European spatial and statistical datasets in order to obtain detailed depictions and quantifications of the population distribution for major European cities.

→ Joint effort of *DG Regio* and *DG Joint Research Centre* (European Commission) to support urban and regional policy and analysis.

Spatial coverage

305 cities and towns in EU27



Data used

In areal interpolation problems, there are usually three types of data involved:

1. Source data
2. Target zones
3. Ancillary data

Data used

Source data

Definition:

Population counts at a given enumeration zoning system.

Data used in this project:

Population reported at bottom-up grids (1 Km or finer), census tracts or communes for *circa* 2006.

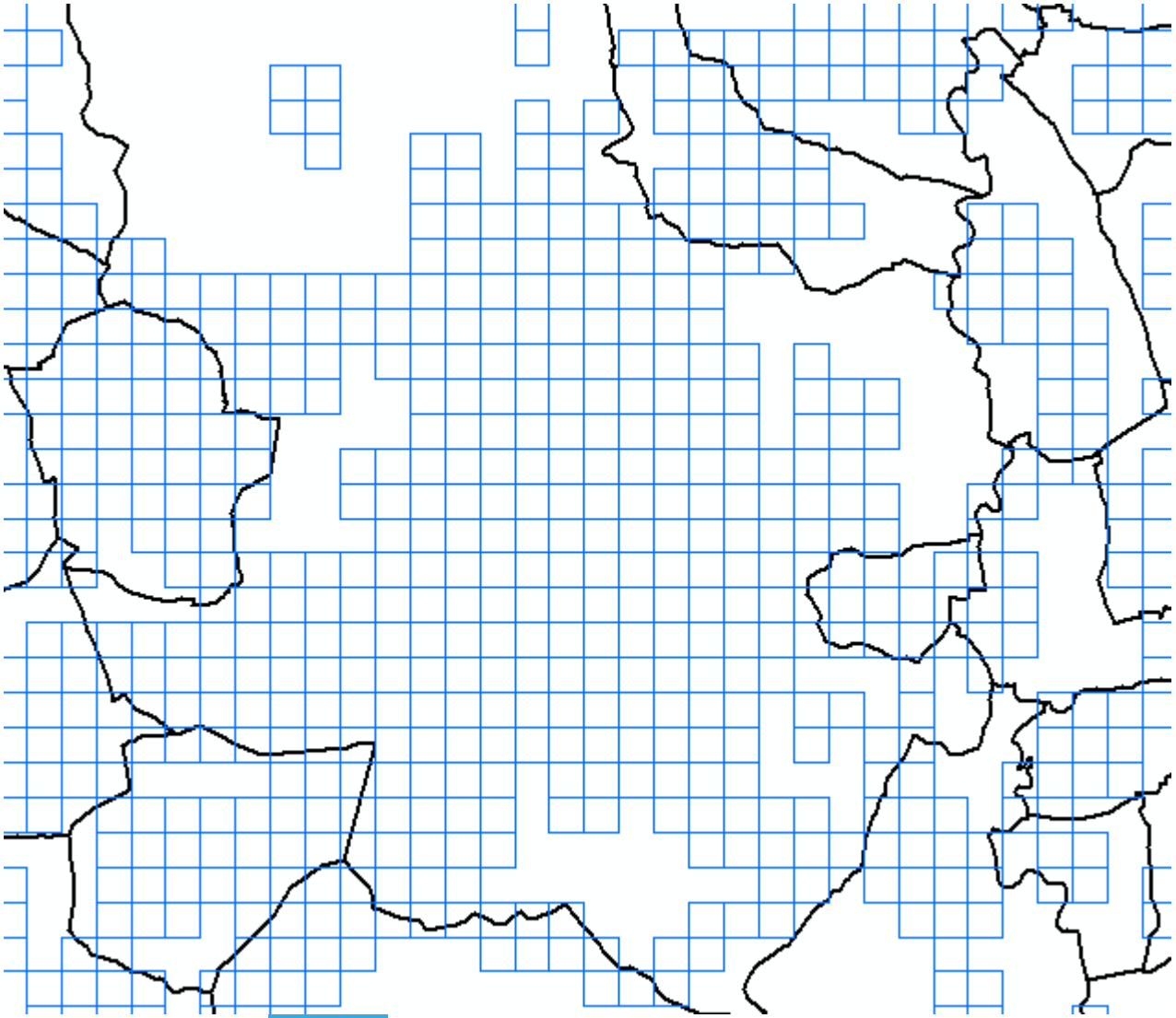
For each town/city, the most detailed available source was used.

Madrid, Spain

Communes

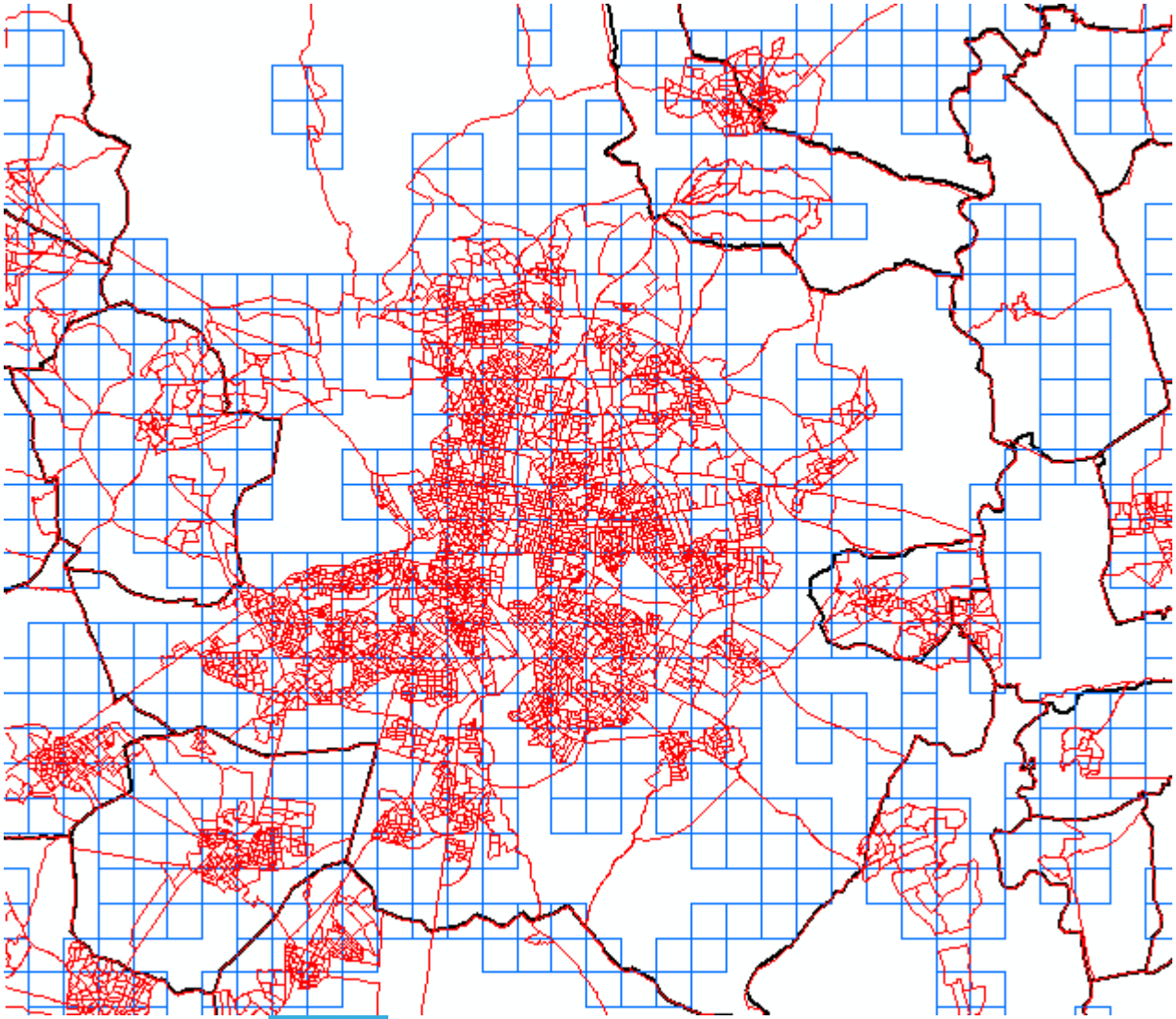
Madrid, Spain

1 km grid



Madrid, Spain

Census tracts



Data used

Target zones

Definition:

The set of geographical entities for which population needs to be estimated.

Data used in this project:

Land use/cover polygons of the Urban Atlas dataset.

Data used

Urban Atlas (main facts)

- Pan-European detailed and comparable land use/cover maps;
- Produced in the context of the European Programme for Global Monitoring for Environment and Security (GMES);
- Freely available (through EEA);
- Spatial coverage: 305 cities and towns across EU27;
- Temporal coverage: 2006 +/- 1 year;

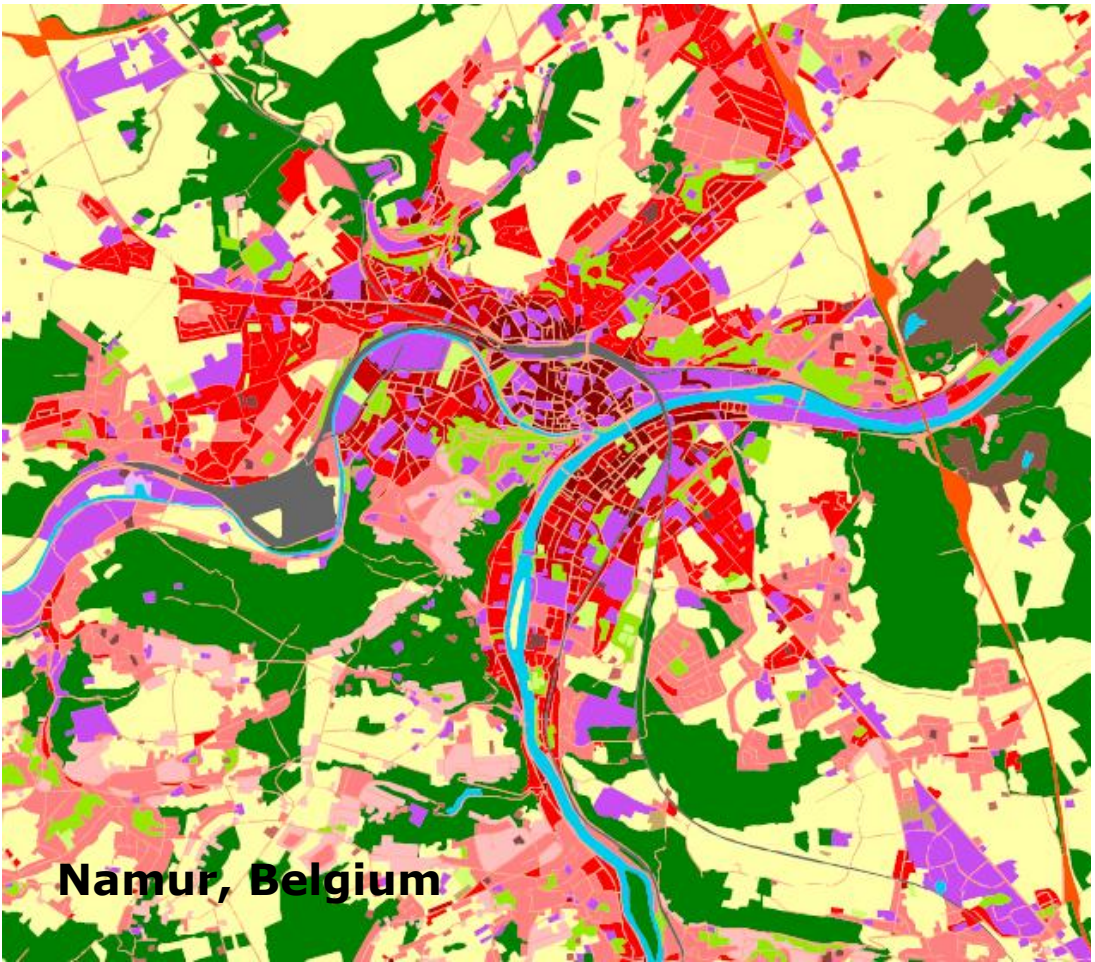
Data used

Urban Atlas (main facts)

- Spatial resolution (minimum mapping unit):
  - 0,25 hectares for artificial surfaces
  - 1 hectare for non artificial surfaces
- Thematic resolution: 20 land use/cover classes, with focus on artificial surfaces;
- Main data sources: Earth Observation (satellite imagery), road network, local topographic maps, local expertise.

Data used

Urban Atlas



- 11100: Continuous Urban fabric (S.L. > 80%)
- 11210: Discontinuous Dense Urban Fabric (S.L.: 50% - 80%)
- 11220: Discontinuous Medium Density Urban Fabric (S.L.: 30% - 50%)
- 11230: Discontinuous Low Density Urban Fabric (S.L.: 10% - 30%)
- 11240: Discontinuous very low density urban fabric (S.L. < 10%)
- 11300: Isolated Structures
- 12100: Industrial, commercial, public, military and private units
- 12210: Fast transit roads and associated land
- 12220: Other roads and associated land
- 12230: Railways and associated land
- 12300: Port areas
- 12400: Airports
- 13100: Mineral extraction and dump sites
- 13300: Construction sites
- 13400: Land without current use
- 14100: Green urban areas
- 14200: Sports and leisure facilities
- 20000: Agricultural, semi-natural and wetland areas
- 30000: Forest
- 50000: Water

Data used

Ancillary data

Definition:

Any kind of spatially explicit data that informs the population disaggregation process. It should be a proxy for population distribution.

Data used in this project:

Degree of soil sealing (wall to wall European dataset) (corrected version of 20 meter resolution).

Data used (summary)

Data category	Description		Reference year	Coverage
Source	Residential population	Type 1: High resolution bottom-up grids (<1km)	2006 +/- 1	Denmark; Finland; Sweden; Slovenia.
		Type 2: Census tracts		Belgium; England and Wales; Netherlands, Spain*.
		Type 3: Medium resolution bottom-up or hybrid grids (1km)		Austria; France; Portugal.
		Type 4: Commune boundaries		Remaining EU-27 countries.
Target	Urban Atlas polygons (only the polygons presumed to be populated)		2006 +/- 1	All EU-27
Ancillary	Soil Sealing Layer (adjusted version used for the production of the Urban Atlas)		2006	All EU-27

* For the Larger Urban Zone of Madrid, bottom-up population data (residential registry points) were aggregated to the Urban Atlas polygons. For the Larger Urban Zone of Seville, a hybrid 1 km grid was used as source data for the disaggregation.

Methodology

Areal interpolation with densities proportional to the soil sealing degree.

Preparatory steps:

1. Classify the Urban Atlas land use/cover classes into:
   a) Inhabited;
   b) Not inhabited.

Methodology

Preparatory steps (cont.):

2. Develop a weighting system representative of the population density for the inhabited classes:

   a) Weights directly derived from the soil sealing degree: for land use/cover classes where the soil sealing degree is assumed to be *correlated* with population density.

   b) Ad-hoc weights for the land use/cover classes where the soil sealing degree and population density are assumed to be *uncorrelated* with population density.

Methodology

Weighting scheme

Urban Atlas class	Description (from the Urban Atlas Mapping Guide)	Weight	
		Method	Value*
11100	Continuous urban fabric (S.L. > 80%)	Directly derived from the average soil sealing degree of polygons	80-100
11210	Discontinuous dense urban fabric (S.L. 50% - 80%)		50-80
11220	Discontinuous medium density urban fabric (S.L. 30 - 50%)		30-50
11230	Discontinuous low density urban fabric (S.L. 10% - 30%)		10-30
11240	Discontinuous very low density urban fabric (S.L. < 10%)		4-9**
11300	Isolated structures		4-9***

Methodology

Weighting scheme

Urban Atlas class	Description (from the Urban Atlas Mapping Guide)	Weight	
		Method	Value*
12100	Industrial, commercial, public military and private units	Attributed arbitrarily	1
12210	Fast transit roads and associated land		0
12220	Other roads and associated land		0
12230	Railways and associated land		0
12300	Port areas		0.1
12400	Airports		0

Methodology

Weighting scheme

Urban Atlas class	Description (from the Urban Atlas Mapping Guide)	Weight	
		Method	Value*
13100	Mineral extraction and dump sites	Attributed arbitrarily	0
13300	Construction sites		0
13400	Land without current use		0
14100	Green urban areas		0
14200	Sports and leisure facilities		1
20000	Agricultural areas, semi-natural areas and wetlands		0.1
30000	Forests		0
50000	Water		0

Notes:

* Indicative thresholds. Miner differences in the actual weights attributed may vary in individual polygons. See footnote 3.

** The lower threshold for the class 11240 was obtained through empirical analysis of observed sealing values.

*** For the class 11300 it was assumed a population weight equal to the one used for the class 11240. The only difference is that the polygons of this class are not contiguous to other urban fabric polygons.

Methodology

Areal interpolation with densities proportional to the soil sealing degree.

Main steps:

1. Source and target zones (vector) are geometrically intersected through a GIS operation.

   → A 'transitional' geometry is obtained.

Methodology

Main steps (cont.):

2. Estimate the population for each polygon of the 'transitional' geometry.

$$P_i' = P_s . \left(\frac{A_i . W_i}{\sum_i^n A_i . W_i} \right)$$

P_i' corresponds to estimated population of a given polygon i of the transitional geometry;

P_s is the known population in the source zone s;

A_i is the area of polygon i;

W_i is the weight assigned to polygon i, corresponding to the average soil sealing value;

n corresponds to the number of transitional polygons within each source polygon.

Methodology

Main steps (cont.):

3. Aggregate the estimates at the level of the target geometry (the Urban Atlas polygons).

$$P_t' = \sum_i^j P_i'$$

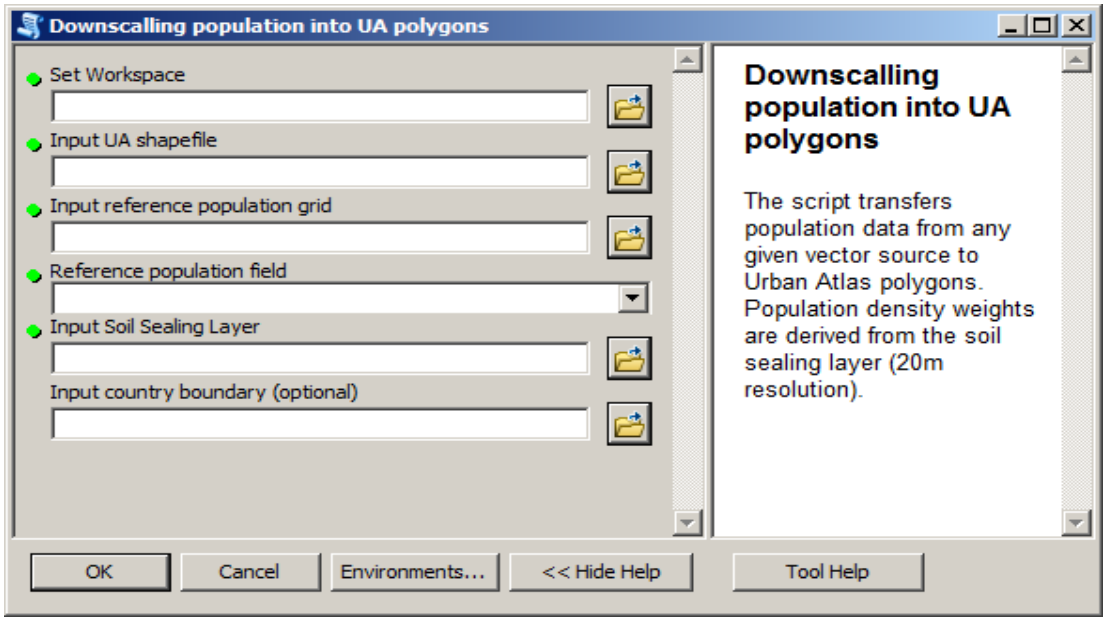
t denotes the target geometry (each Urban Atlas polygon)

j corresponds to the number of transitional polygons within each target polygon

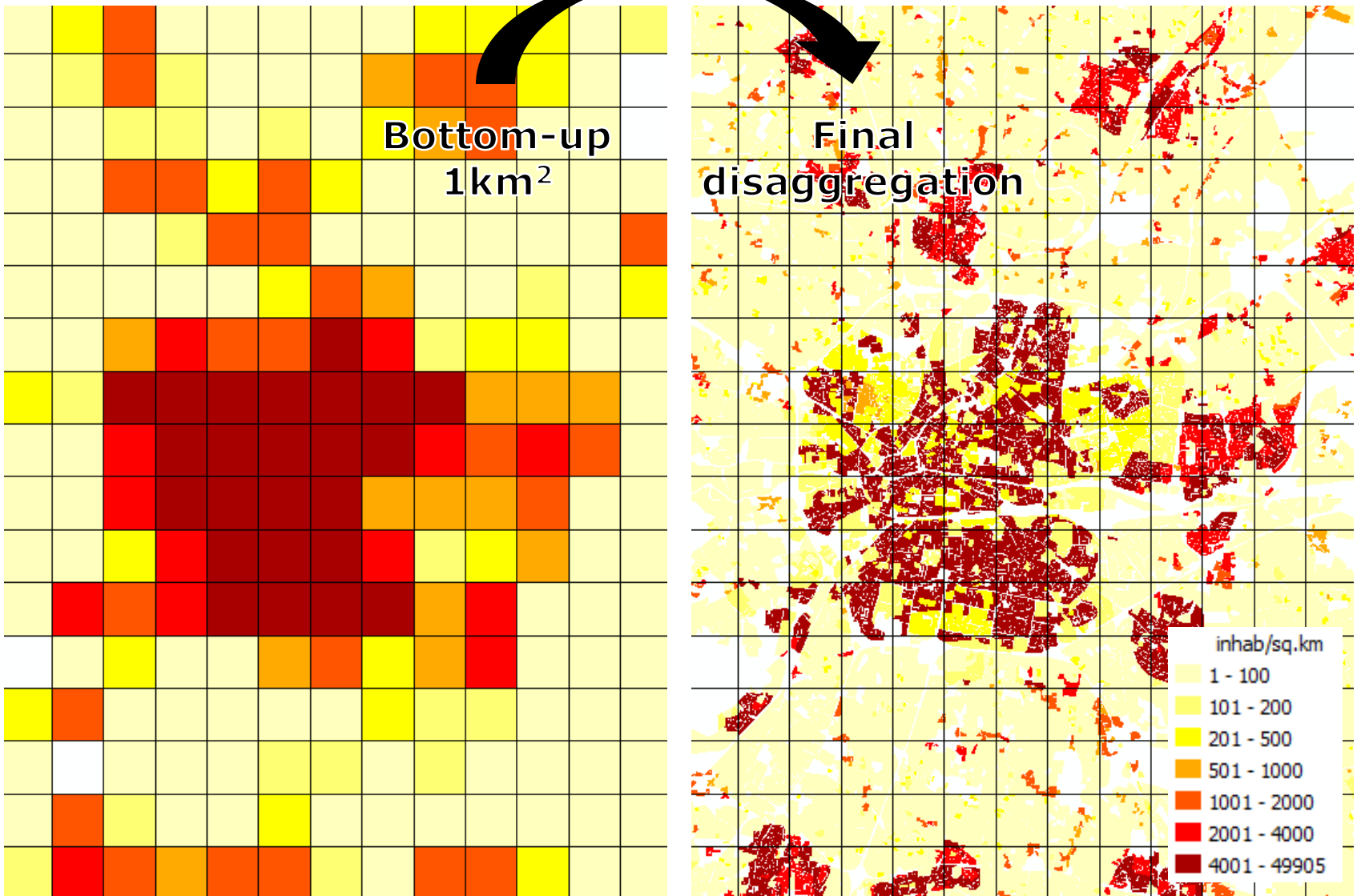
Methodology

Automation of the process

A script written in Python programming language, and accessible as a tool within the ArcGIS environment was created.



Results


Bottom-up
1km²
Final
disaggregation
inhab/sq.km
1 - 100
101 - 200
201 - 500
501 - 1000
1001 - 2000
2001 - 4000
4001 - 49905

Validation

Applied to a limited number of countries / regions:

- Austria, Finland, Portugal, Madrid.

The validation consisted of comparing the **estimated** against **known** number of residents for each Urban Atlas polygon.

The **known** number of residents for each polygon was obtained by aggregating point-data counts → ***ground truth data!***

(work done in collaboration with the Austrian, Finish and Portuguese Statistical Offices and with the University of Valencia, Spain)

Validation

Accuracy indicator: Total Absolute Error (TAE)

Measures the overall disagreement observed in a given study area by summing the absolute deviations between known and estimated values for all target zones:

$$TAE = \Sigma_t \, |P'_t - P_t \,|$$

By definition, TAE varies within the range $[0 \, , \, 2{*}P_{study\_area}]$.

$RTAE = TAE \, / \, P_{study\_area} \, , \in [0 \, , \, 2]$ (RTAE is easier to read)

Validation

if RTAE = 0

Perfect disaggregation. The population estimated for all target zones (Urban Atlas polygons) matches perfectly with ground truth data.

if RTAE = 2

Completely wrong disaggregation.

Validation (results)

UATL City			Source data		RTAE [0-2]	
Country	Code	Name	Type	Median unit size (sq. Km)		
Finland	FI001	Helsinki	Bottom-up	0,25	0,230	12 - 15%
	FI004	Oulu			0,260	
	FI002	Tampere			0,290	
	FI003	Turku			0,270	
Austria	AT001	Wien	Bottom-up	1,00	0,442	22 - 28%
	AT002	Graz			0,530	
	AT003	Linz			0,554	
	AT004	Salzburg			0,516	
	AT005	Innsbruck			0,437	
Portugal	PT001	Lisboa	Bottom-up	1,00	0,501	23 - 29%
	PT002	Porto			0,486	
	PT003	Braga			0,506	
	PT004	Funchal			0,522	
	PT005	Coimbra			0,524	
	PT006	Setubal			0,466	
	PT007	Ponta Delgada			0,470	
	PT008	Aveiro			0,469	
	PT009	Faro			0,579	

Range of total possible error

Validation (results)

UATL City			Source data		RTAE [0-2]	
Country	Code	Name	Type	Median unit size (sq. Km)		
Spain	ES001	Madrid	Communes	33,49	0,703	35%
			Bottom-up	1,00	0,516	26%
			Census tracts	0,05	0,390	20%

Possible logarithimic relationship between detail of input source zones and accuracy.

Validation (results)

Land Use Class		Austria		Finland		Portugal		
Code	Label	Absolute Error	% Error	Absolute Error	% Error	Absolute Error	% Error	
11100	Continuous Urban Fabric	47.484	4,7%	7.597	2,9%	-111.635	-4,2%	Weights derived from SSL
11210	Discont. Dense Urban Fabric	5.177	0,4%	18.043	4,6%	237.074	18,4%	
11220	Discont. Medium Density Urban Fabric	116.785	14,4%	21.012	4,4%	88.644	26,1%	
11230	Discont. Low Density Urban Fabric	19.354	7,5%	8.369	1,7%	9.376	7,6%	
11240	Discont. Very Low Density Urban Fabric	- 6.023	-39,6%	- 27.126	-11,0%	- 6.188	-39,1%	
11300	Isolated Structures	- 27.676	-48,3%	4.303	7,6%	- 19.398	-53,8%	
12100	Industrial, commercial, public…	-141.065	-83,4%	- 35.612	-49,8%	-121.755	-81,6%	Ad-hoc weights
12300	Port areas	- 135	-61,0%	- 104	-51,8%	- 53	-32,9%	
14200	Sports and leisure facilities	- 8.766	-53,7%	2.887	229,3%	- 1.815	-29,5%	
20000	Agric. + Semi-nat. + Wetlands	154	0,6%	3.020	33,6%	- 42.627	-67,3%	
12400	Airports	- 124	-100,0%	- 37	-100,0%	- 46	-100,0%	No pop. assigned
13300	Construction sites	- 4.215	-100,0%	- 1.741	-100,0%	- 26.343	-100,0%	
13400	Land without current use	- 951	-100,0%	- 257	-100,0%	- 5.234	-100,0%	

Summing up and conclusions

1. Collaborative work between institutions of the European Commission, including the active involvement of a number of National Statistical Offices;
2. Final outcome: detailed population estimates at the level of the Urban Atlas land use/cover polygons. Available for any analysis that might benefit from these small-area estimates.

Summing up and conclusions

3. Regarding the validation:

- Validation done by comparing estimates with ground-truth data. Uncertainties of the estimates are reasonably known;
- Reliability of the product varies across countries, mainly due to input source data of different resolutions: from high resolution bottom-up grids (< 1 km^2) to large commune boundaries.
- Some frailties are a consequence of the proxy used for population density (soil sealing degree) and from the weighting scheme.

Summing up and conclusions

4. Future developments:

   - Some of the detected weaknesses of the product can be addressed mainly by fine-tuning the weighting scheme;
   - The workflow is fully operational to produce new estimates as soon as new, updated or more detailed data are available;
   - The soil sealing degree *alone* is not the optimal proxy for population density. It does not capture the vertical dimension of the population distribution. Alternative and complementary proxies should be sought.

Acknowledgments

For providing detailed bottom-up population data (grid cells < 1 km²)

- Francisco Goerlich (University of Valencia)
- Madrid Regional Statistical Institute
- Statistics Denmark
- Statistics Slovenia
- Statistics Finland
- Statistics Sweden

Acknowledgments

For collaborating in the validation of the results by comparing our estimates with population point-data:

- Ana M. Santos (Statistics Portugal)
- Rina Tammisto and Markku Koivula(Statistics Finland)
- Ingrid Kaminger (Statistics Austria)