The effect of ancillary data in population dasymetric mapping

Filipe Batista e Silva
filipe.batista@jrc.ec.europa.eu

Javier Gallego
javier.gallego@jrc.ec.europa.eu

Carlo Lavalle
carlo.lavalle@jrc.ec.europa.eu

JOINT RESEARCH CENTRE
Institute for Environment and Sustainability
Land Management and Natural Hazards Unit
Ispra, Italy

JRC EUROPEAN COMMISSION
ies Institute for Environment and Sustainability
EFGS – Lisboa, Portugal 12-14 October 2011

STRUCTURE OF THE PRESENTATION

1. **Dasymetric mapping** (brief intro)
2. **Scope & objectives**
3. **Methodology**
    1. Overall workflow
    2. Refinement of CLC
    3. Disaggregation tecnhiques tested
4. **Results & validation**

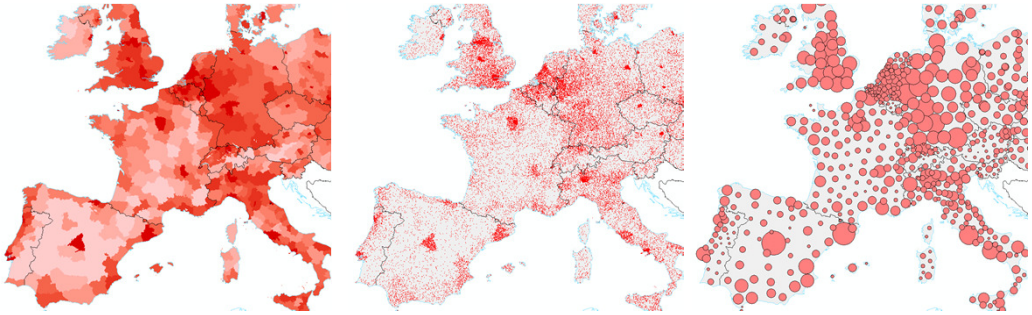
1. **Dasymetric mapping**
2. Scope & objectives
3. Methodology
    1. Overall workflow
    2. Refinement of CLC
    3. Disaggregation tecnhiques tested
4. Results & validation

JRC EUROPEAN COMMISSION
ies Institute for Environment and Sustainability
EFGS – Lisboa, Portugal 12-14 October 2011

CONVENTIONAL POPULATION MAPPING

Conventional population mapping uses traditional cartographic techniques such as:

choropleth mapping **dot mapping** **proportional symbols**



CONVENTIONAL POPULATION MAPPING

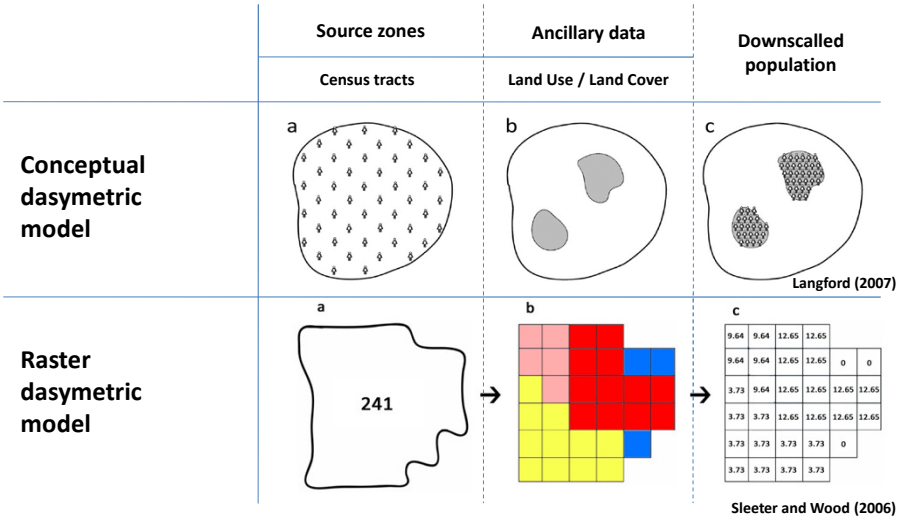
- These techniques **rely solely on the zoning layout**, usually administrative units of **heterogeneous size.**
- The resulting maps are subjected to the Modifiable Areal Unit Problem (**MAUP**).
- There is an **assumption of homogeneous distribution** within the spatial unit used (spatial average that masks the real distribution).
- Relevant **physiographic features are not considered**, such as water bodies, rocky and mountainous surfaces, natural areas - usually unpopulated.

JRC EUROPEAN COMMISSION
ies Institute for Environment and Sustainability
EFGS – Lisboa, Portugal 12-14 October 2011

DASYMETRIC MAPPING

- To address these problems, **dasymetric mapping techniques** have been developed in order to produce better and more realistic maps (population and not only).
- Dasymetric mapping can be described as a **cartographic technique whereby ancillary thematic data is used to configure the spatial distribution of a quantitative variable reported at coarse regional aggregation**.
- Dasymetric mapping is a type of areal interpolation since it involves the transformation of data from one set of spatial units (source units) to another (target units – generally of regular shape and size).
  - **Downscaling**: a type of areal interpolation that transforms data from one coarse set of spatial units to a finer set of units.

DASYMETRIC MAPPING



JRC EUROPEAN COMMISSION
ies Institute for Environment and Sustainability
EFGS – Lisboa, Portugal 12-14 October 2011

DASYMETRIC MAPPING - ANCILLARY DATA

For population dasymetric mapping, various ancillary spatial data may be used as predictors of **where population lives** (or doesn't live).

Ancillary data extracted from

Remote sensing imagery	Topographic/base maps
Land use / land cover	Road network
Night time lights	Buildings
Soil sealing degree	Physiographic features

LAND USE / LAND COVER AS ANCILLARY DATA

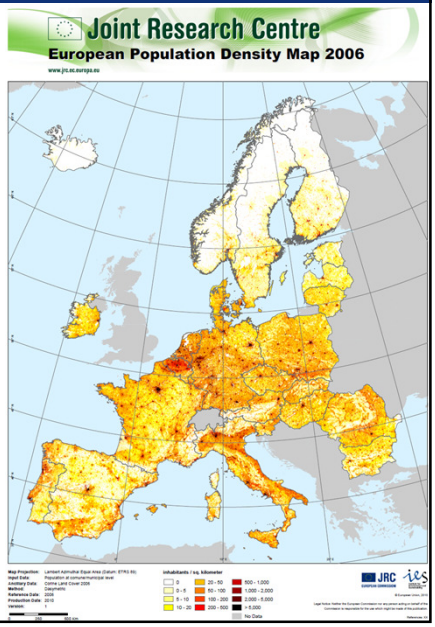
- The **choice** on the ancillary data is **often based on data availability**, spatial extent and scale/resolution of the final output map.
- **Land use / land cover (LULC) is usually a good predictor of where people live**.
- LULC is usually available at different scales/resolutions, for wide spatial extents and with a medium temporal frequency.
- Specific Land Cover classes can also be derived automatically from satellite imagery.
- **LULC are thus commonly used** as ancillary data for dasymetric mapping of the population distribution/density.

JRC EUROPEAN COMMISSION
ies Institute for Environment and Sustainability
EFGS – Lisboa, Portugal 12-14 October 2011

For European population mapping, downscaling procedure has been **relying on ancillary LULC datasets** (CORINE Land Cover).

European Population Density Map 2006 v.1

- Digital raster grid describing population density as inhabitants per square kilometer (inhab./km2);
- Spatial resolution: 100 meter;
- Input Data: Population at commune level and Corine Land Cover 2006 (v.13);
- Disaggregation method: Dasymetric areal interpolation / limiting variable algorithm



1. Dasymetric mapping
2. **Scope & objectives**
3. Methodology
    1. Overall workflow
    2. Refinement of CLC
    3. Disaggregation tecnhiques tested
4. Results & validation

JRC EUROPEAN COMMISSION
ies Institute for Environment and Sustainability
EFGS – Lisboa, Portugal 12-14 October 2011

SCOPE AND OBJECTIVES

PRIMARY OBJECTIVE

Test the effect ancillary data in population dasymetric mapping.

How relevant is ancillary data in the accuracy of a population distribution map?

To answer this question, ancillary data of different characteristics and resolutions will be used to disaggregate population.

FINAL OUTCOME

A refined, updated and comprehensive **European Population Grid** for the year **2006** at **1 hectare** resolution (100 x 100 meters).

SCOPE AND OBJECTIVES

SPATIAL SCOPE

EU 27*

+

Norway

+

Switzerland

+

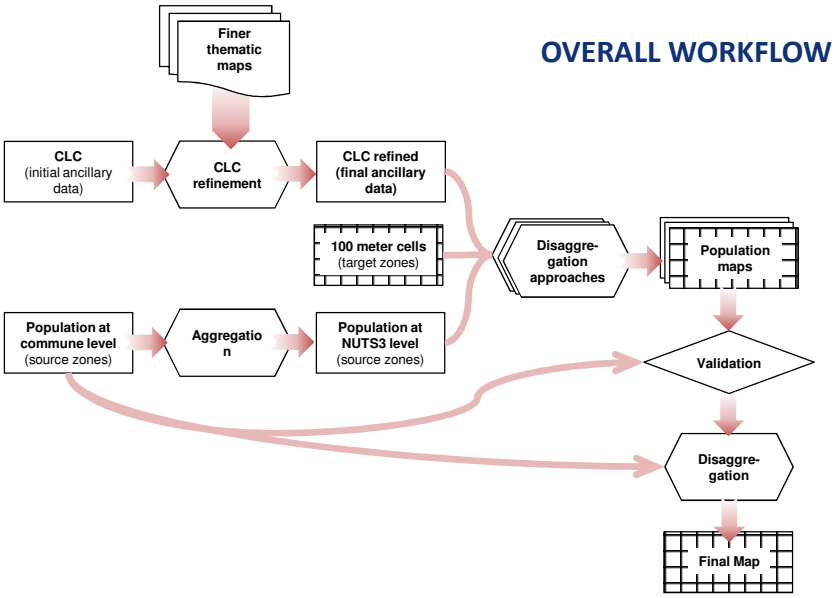
Micro States

* Greece missing due to lack of CORINE Land Cover 2006

JRC EUROPEAN COMMISSION
ies Institute for Environment and Sustainability
EFGS – Lisboa, Portugal 12-14 October 2011

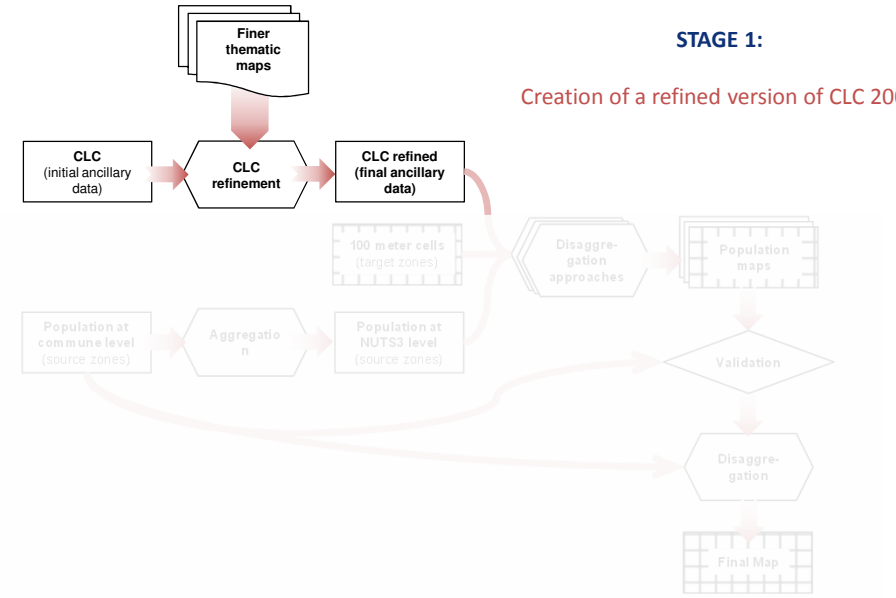
1. **Dasymetric mapping** (brief intro)
2. **Scope & objectives**
3. **Methodology**
   1. Overall workflow
   2. Datasets used
   3. Disaggregation tecnhiques tested
4. **Results & validation**

OVERALL WORKFLOW



JRC EUROPEAN COMMISSION
ies Institute for Environment and Sustainability
EFGS – Lisboa, Portugal 12-14 October 2011

1. Dasymetric mapping (brief intro)
2. Scope & objectives
3. **Methodology**
    1. Overall workflow
    2. Refinement of CLC
    3. Disaggregation tecnhiques tested
4. Results & validation



STAGE 1:

Creation of a refined version of CLC 2006

JRC EUROPEAN COMMISSION
ies Institute for Environment and Sustainability
EFGS – Lisboa, Portugal 12-14 October 2011

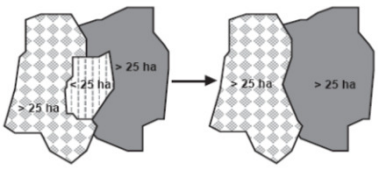
REFINEMENT OF CORINE LAND COVER 2006

Main advantages of using CLC:

- Temporal frequency (1990, 2000, 2006, and on…).
- European spatial extent.
- Methodological consistency leads to *reasonable* degree of comparison across countries.

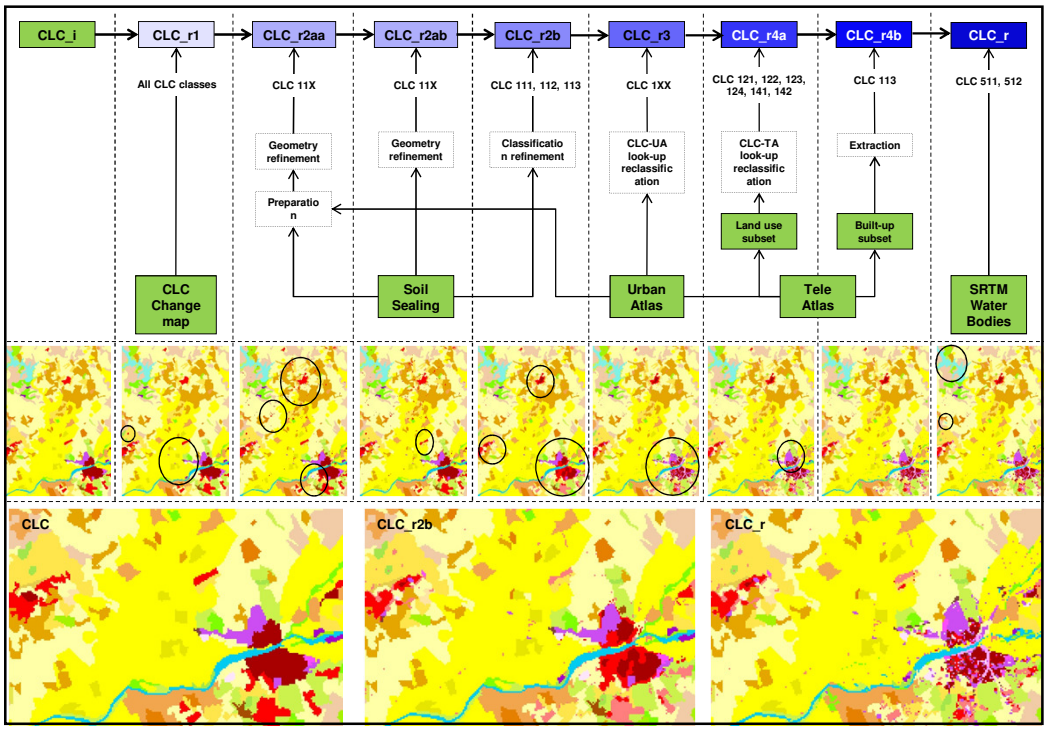
Main disadvantages of using CLC:

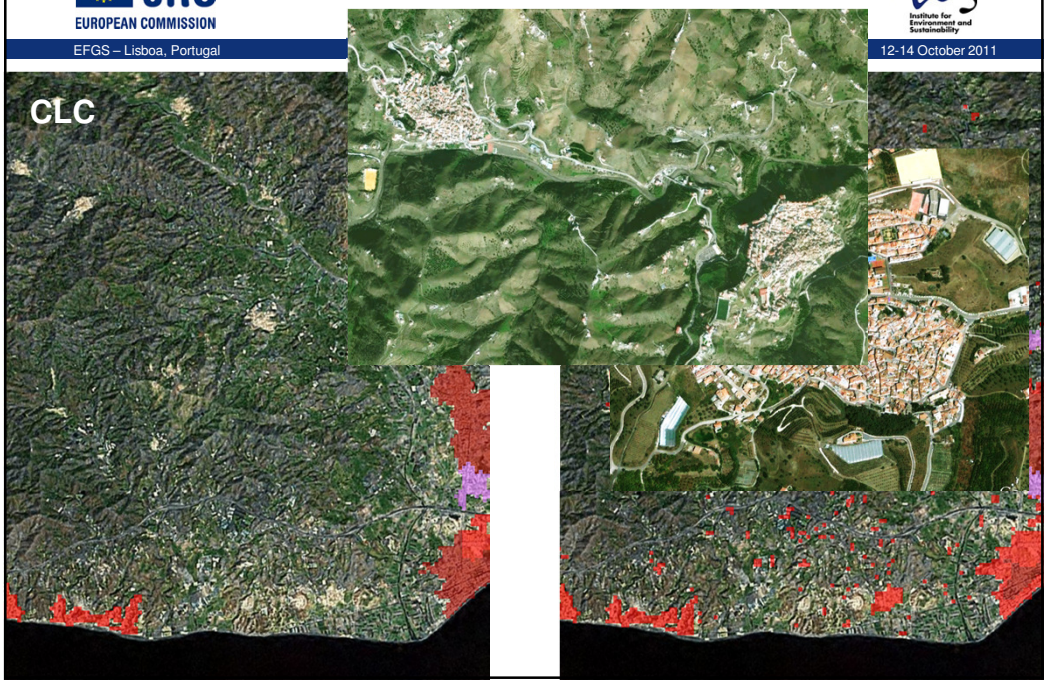
- Low spatial detail (MMU: 25 hectares)
- Low thematic resolution for artificial covers.



REFINEMENT OF CORINE LAND COVER 2006

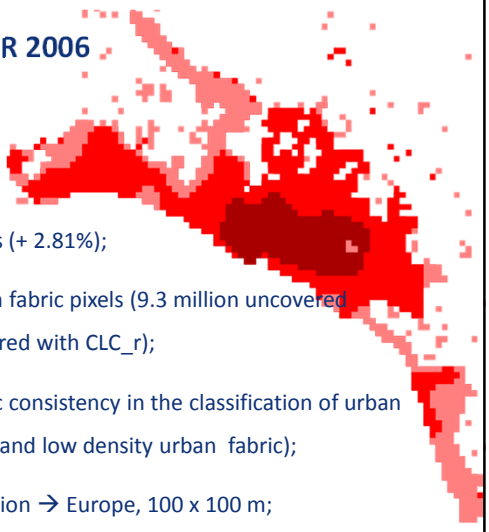
- Target mainly the artificial land cover classes, increasing the minimum mapping unit to 1 hectare, thus allowing a more complete representation of urban patterns;
- Was operated by incorporating land use/cover information present in various higher resolution thematic maps available for Europe such as:
  - **CLC change map;**
  - **Soil sealing layer;**
  - **Tele Atlas® Spatial Database;**
  - **Urban Atlas;**
  - **SRTM Water Bodies Data.**





JRC EUROPEAN COMMISSION
EFGS – Lisboa, Portugal
ies Institute for Environment and Sustainability
12-14 October 2011

REFINEMENT OF CORINE LAND COVER 2006



Relevant results:

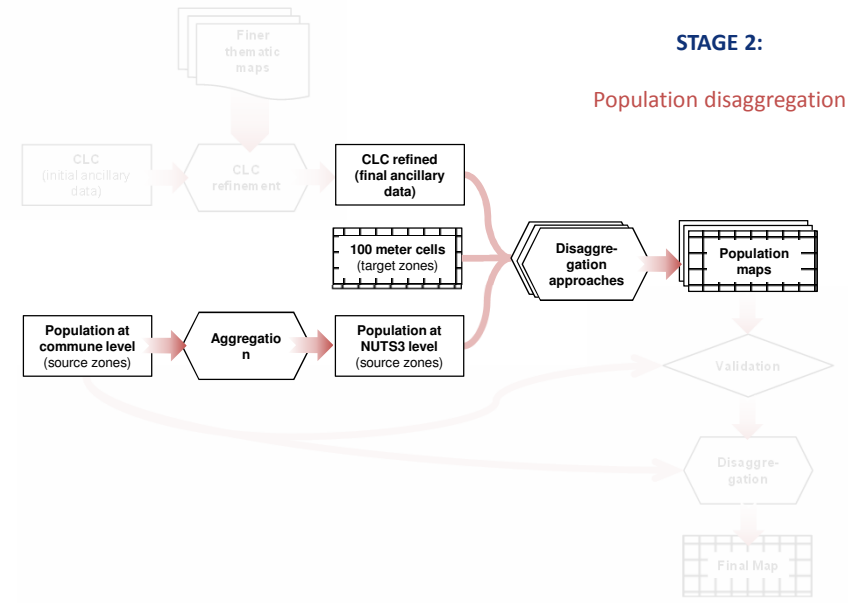
- Increased number of settlements;
- Increased reported area of urban areas (+ 2.81%);
- Increased population covered by urban fabric pixels (9.3 million uncovered inhabitants in CLC; 0.25 million uncovered with CLC_r);
- Increased thematic detail and thematic consistency in the classification of urban fabric classes (3 classes: high, medium and low density urban fabric);
- Same spatial extent, same pixel resolution → Europe, 100 x 100 m;
- **BUT…** mix of MMU's / loss of cartographic consistency…

1. Dasymetric mapping (brief intro)
2. Scope & objectives
3. **Methodology**
    1. Overall workflow
    2. Refinement of CLC
    3. Disaggregation tecnhiques tested
4. Results & validation

JRC EUROPEAN COMMISSION
ies Institute for Environment and Sustainability
EFGS – Lisboa, Portugal 12-14 October 2011

STAGE 2:

Population disaggregation



DISAGGREGATION APPROACHES

	Approach M1	Approach M2	Approach M3
Source pop. data	Population reported at NUTS3. Reference year 2006. Eurostat / NSOs.		
Ancillary data	CLC 2006 Classes used: all	CLC 2006 refined Classes used: urban fabric, industry & commercial	
Disaggregation rules/algorithm	'Limiting variable'*	Densities defined by regression**	Densities proportional to sealing degree***

* Gallego, F.J. *et al.* (in press)

** Adapted from Briggs D.J *et al.* (2007)

*** Adapted from Steinnocher K. *et al.*

JRC EUROPEAN COMMISSION
EFGS – Lisboa, Portugal 12-14 October 2011

DISAGGREGATION APPROACHES – M2

1. A multiple regression model was defined:

$$P_s = \beta_0 + \beta_{111}UA_{111} + \beta_{112}UA_{112} + \beta_{113}UA_{113} + \varepsilon_s$$

P_s = Total population at NUTS3 level (dependent variable)

β_0 = intercept

β_c = coefficient of CLC_r class c

Coefficients are used as weights to redistribute NUTS3 population within the CLC_r classes.

UA_c = Urban Area of CLC_r class c (explanatory variable)

ε_s = Residual

DISAGGREGATION APPROACHES – M2

2. The regression was applied in three strata:

Strata (NUTS3)

	PopDensity	Adj-R2
Stratum 1	< 32 inhab/ha	0.922
Stratum 2	32 - 65 inhab/ha	0.951
Stratum 3	> 65 inhab/ha	0.904

3. The obtained coefficients were transformed in weights

Coefficients

	Stratum 1	Stratum 2	Stratum 3
CLCr_111	29.041	88.255	167.893
CLCr_112	27.680	42.631	82.772
CLCr_113	19.787	30.296	66.872
Σ	76.508	161.182	317.537

Weights

	Stratum 1	Stratum 2	Stratum 3
CLCr_111	0.380	0.548	0.529
CLCr_112	0.362	0.264	0.261
CLCr_113	0.259	0.188	0.211
Σ	1.000	1.000	1.000

JRC EUROPEAN COMMISSION
ies Institute for Environment and Sustainability
EFGS – Lisboa, Portugal 12-14 October 2011

DISAGGREGATION APPROACHES – M2

4. Disaggregation based on the weights

Example for a NUTS3, Stratum 2, total population = **320,000**

Calculation example

	nr. Pixels (P)	weights (W)	P * W	share_pop	total_pop	Pop. Density
CLCr_111	350	0.548	191.64	22.02%	70,464	201.33
CLCr_112	1500	0.264	396.73	45.58%	145,871	97.25
CLCr_113	1500	0.188	281.94	32.40%	103,665	69.11
Σ	3350	1	870.32	100.00%	320,000	

DISAGGREGATION APPROACHES – M3

1. Find the 'relation' between Population *P* and Mean soil sealing degree *S*, through a constant *k* at NUTS3 level *s*, for each land cover class *c*:

$$k_s = P_s / \Sigma_c (UA_{cs} * S_c)$$

k = constant in NUTS3 s

P_s = Total population of NUTS3 s

UA_c = Urban Area of CLC_r class c in NUTS3 s

S_c = Mean soil sealing degree (1 - 100) of CLC_r class c in NUTS3 s

JRC EUROPEAN COMMISSION
EFGS – Lisboa, Portugal 12-14 October 2011

DISAGGREGATION APPROACHES – M3

2. Disaggregate:

$$D_{cs} = k_s * S_c$$

D_{cs} = Population density of CLC_r class c in NUTS3 s

Calculation example

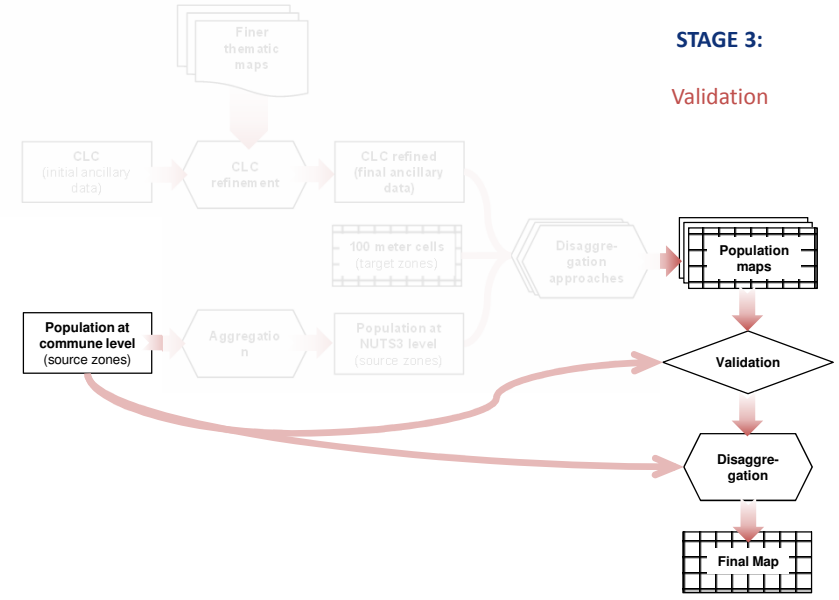
NUTS3	Population	Nr. Pixels			Mean soil sealing			k	Pop. density			Population			Total
		CLCr_111	CLCr_112	CLCr_113	CLCr_111	CLCr_112	CLCr_113		CLCr_111	CLCr_112	CLCr_113	CLCr_111	CLCr_112	CLCr_113	
AT122	251,280	134	4,166	10,815	86	51	23	0.528	45.6	26.8	12.3	6,108	111,776	133,396	251,280
AT123	146,412	111	2,422	5,907	86	52	28	0.487	42.0	25.4	13.6	4,664	61,625	80,122	146,412
AT124	222,205	62	3,312	11,905	88	54	32	0.390	34.3	21.1	12.6	2,129	69,795	150,281	222,205
AT125	123,653	0	3,692	8,954	83	52	28	0.280	23.3	14.6	7.8	0	54,025	69,628	123,653

1. **Dasymetric mapping** (brief intro)
2. **Scope & objectives**
3. **Methodology**
    1. Overall workflow
    2. Refinement of CLC
    3. Disaggregation tecnhiques tested
4. **Results & validation**

JRC EUROPEAN COMMISSION
ies Institute for Environment and Sustainability
EFGS – Lisboa, Portugal 12-14 October 2011

STAGE 3:

Validation

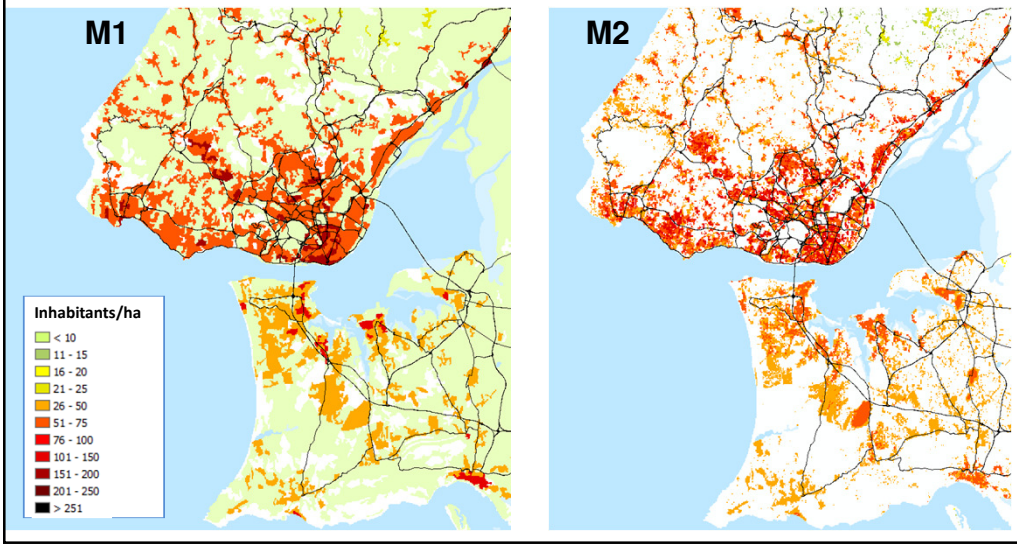


JRC EUROPEAN COMMISSION

Institute for Environment and Sustainability

EFGS – Lisboa, Portugal 12-14 October 2011

DISAGGREGATION APPROACHES – RESULTS

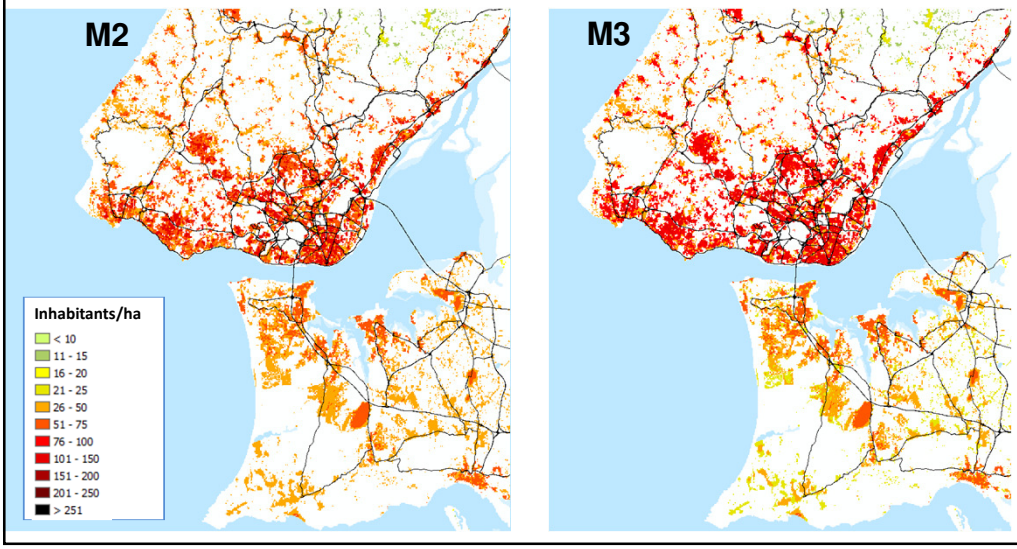


JRC EUROPEAN COMMISSION

Institute for Environment and Sustainability

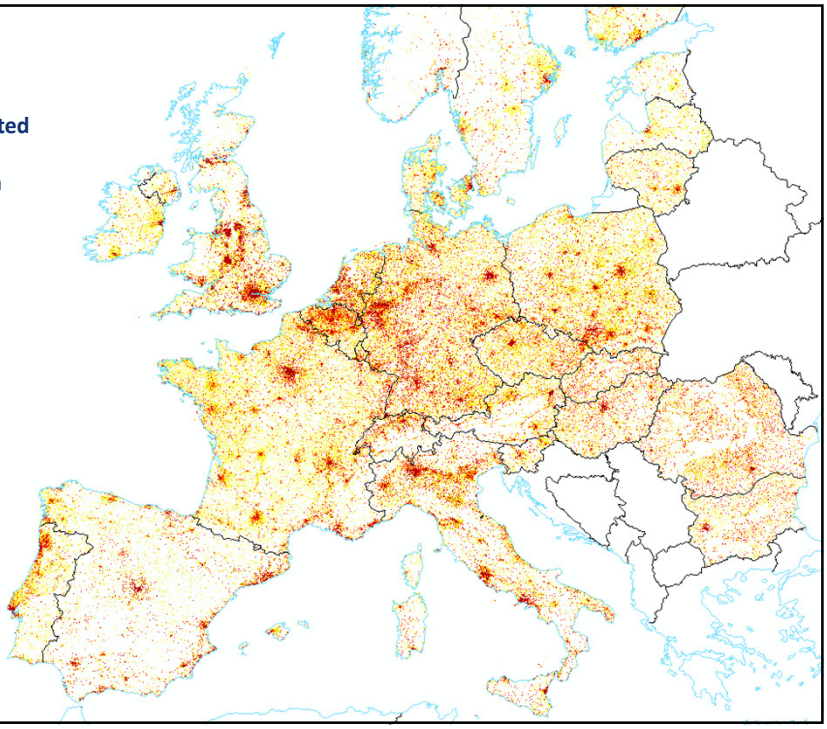
EFGS – Lisboa, Portugal 12-14 October 2011

DISAGGREGATION APPROACHES – RESULTS



M3

Results aggregated at 1km cells
(for visualization purposes)



EUROPEAN COMMISSION JRC

Institute for Environment and Sustainability

EFGS – Lisboa, Portugal 12-14 October 2011

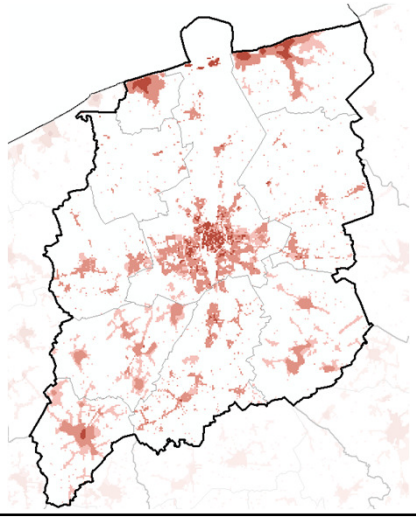
VALIDATION

Main indicator used: **Total Absolute Error (TAE)**

$$R_s = \widehat{P}_s - P_s \qquad TAE = \sum_{s=1}^{n} |R_s|$$

Calculation example

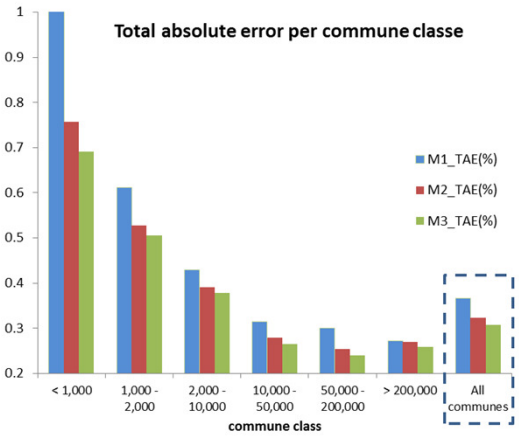
NUTS3	COMM	Real Pop	Estimated Pop	residual	|residual|	Relative Error
BE251	BE231003	14,642	18,462	3,820	3,820	26.1%
BE251	BE231004	18,175	16,479	-1,697	1,697	9.3%
BE251	BE231005	117,224	93,640	-23,584	23,584	20.1%
BE251	BE231006	10,899	14,263	3,364	3,364	30.9%
BE251	BE231012	13,572	16,369	2,797	2,797	20.6%
BE251	BE231022	21,796	26,791	4,995	4,995	22.9%
BE251	BE231033	19,453	24,414	4,961	4,961	25.5%
BE251	BE231040	21,835	25,093	3,258	3,258	14.9%
BE251	BE231042	2,776	3,007	231	231	8.3%
BE251	BE231043	34,063	35,918	1,855	1,855	5.4%
	Total population	274,435	Total Abs. Error	Σ	50,561	
				%	18.42%	



VALIDATION – RESULTS

Some conclusions:

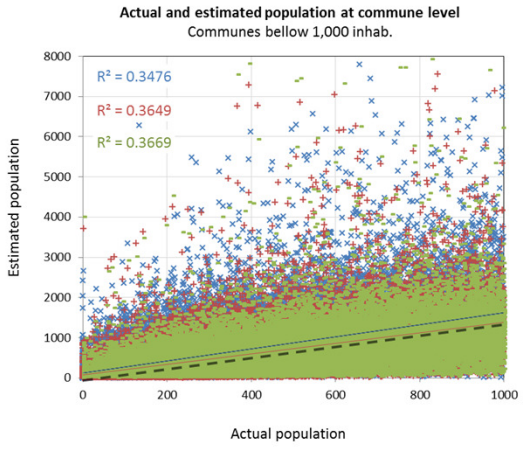
- Models M2 and M3 perform consistently better than M1:
  - ✓ The refinement of CLC seems to have a relevant and positive impact in population mapping.
- M3 performs better than M2:
  - ✓ Information on the soil sealing degree further improves the population mapping.
- Higher accuracies in communes with more inhabitants.

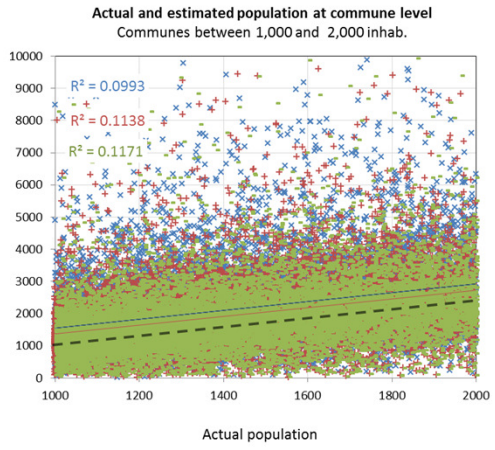


JRC EUROPEAN COMMISSION
ies Institute for Environment and Sustainability
EFGS – Lisboa, Portugal 12-14 October 2011

VALIDATION

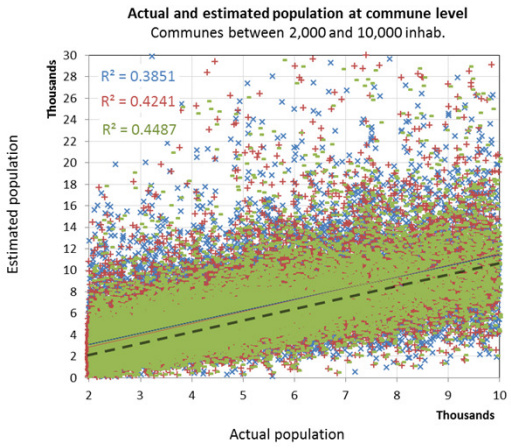
Scatter plots: Actual vs. Estimated population at commune level

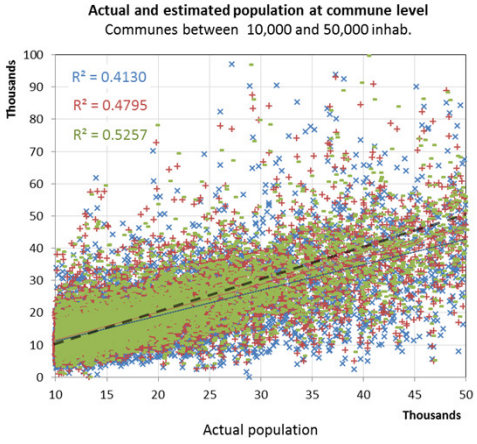




VALIDATION

Scatter plots: Actual vs. Estimated population at commune level

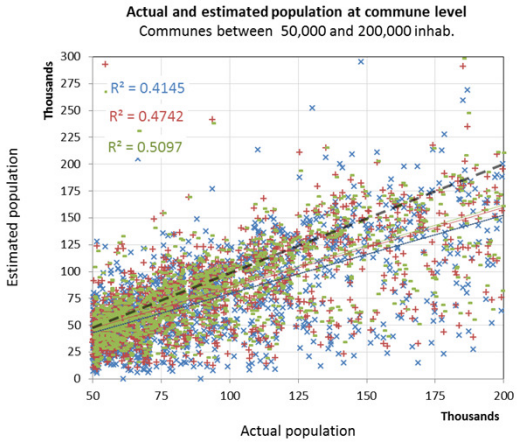


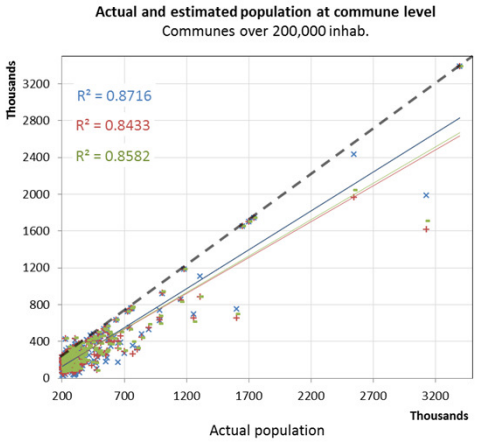


JRC EUROPEAN COMMISSION
Institute for Environment and Sustainability
EFGS – Lisboa, Portugal 12-14 October 2011

VALIDATION

Scatter plots: Actual vs. Estimated population at commune level





NEXT STEPS

- Receive your feedback…
- Produce a final population grid based on the lessons learnt in this study.