Downscaling Population with a High Resolution Land Cover Data Set for Spain.

(In progress…)

Francisco J. Goerlich, University of Valencia and Ivie

Isidro Cantarino, Polytechnic University of Valencia

EFGS Conference - Lisbon, October 13, 2011

Downscaling…

- Our starting point is a **well known fact**: The quality, and resolution, of the land cover layer is much more important than the particular choice of the downscaling algorithm (Martin et al 2000).

- So there is no point in trying to improve over what has already been done with Corine (Gallego 2010), but would be much more profitable looking for **alternative auxiliary information**.

- Assuming bottom-up approaches are not available…

…and this is what we have done.

- We have applied similar methods to the ones employed by Gallego (2010), whose population grid is available at the EEA *web* site, but using a high resolution land cover data set with a complex structure:

  Sistema de Información de Ocupación del Suelo en España –SIOSE– (Information System on Land Cover in Spain)

- The SIOSE data base has been developed by our *National Geographical Institute*, which is a different body than the *National Statistical Institute*, which produce the population data.

CORINE Land Cover *versus* SIOSE

	CLC2006	SIOSE2005
Escala Cartográfica	1:100.0000	1:25.000
Unidad mínima cartografiable	25 ha	- 0.5 ha: cultivos forzados, coberturas húmedas, playas, vegetación de ribera y acantilados marinos. - 1 ha: Zonas urbanas y láminas de agua. - 2 ha: Zonas agrícolas, forestales y naturales.
Anchura mínima de elementos lineales	100 m	15 m
Modelo de Datos	Jerárquico (44 clases)	Orientado a Objetos (40 clases simples y 46 compuestas predefinidas)
Sistema geodésico de referencia	ETRS89	ETRS89 excepto Canarias que utiliza WGS84
Sistema cartográfico de proyección	UTM, huso 30 extendido	UTM, huso correspondiente a cada Comunidad Autónoma

CORINE Land Cover *versus* SIOSE

- **Cartographic scale**:

  1:25.000 SIOSE *versus* 1:100.000 CLC

- **Minimum Mapping Unit**:

  1ha. Urban zones, SIOSE *versus* 25ha. CLC

- **Minimum width lineal elements:**

  15 m. SIOSE *versus* 100 m. CLC

- **Data Model**:

  Object oriented, SIOSE *versus* hierarchical (44 classes) CLC

Improved resolution…

- **Spatial resolution**:
  CLC2006   155.801 polygons
  SIOSE2005   2.477.593 polygons

- **Average polygon size**:
  CLC2006   3,24 Km^2
  SIOSE2005   0,20 Km^2

- **Thematic resolution**:
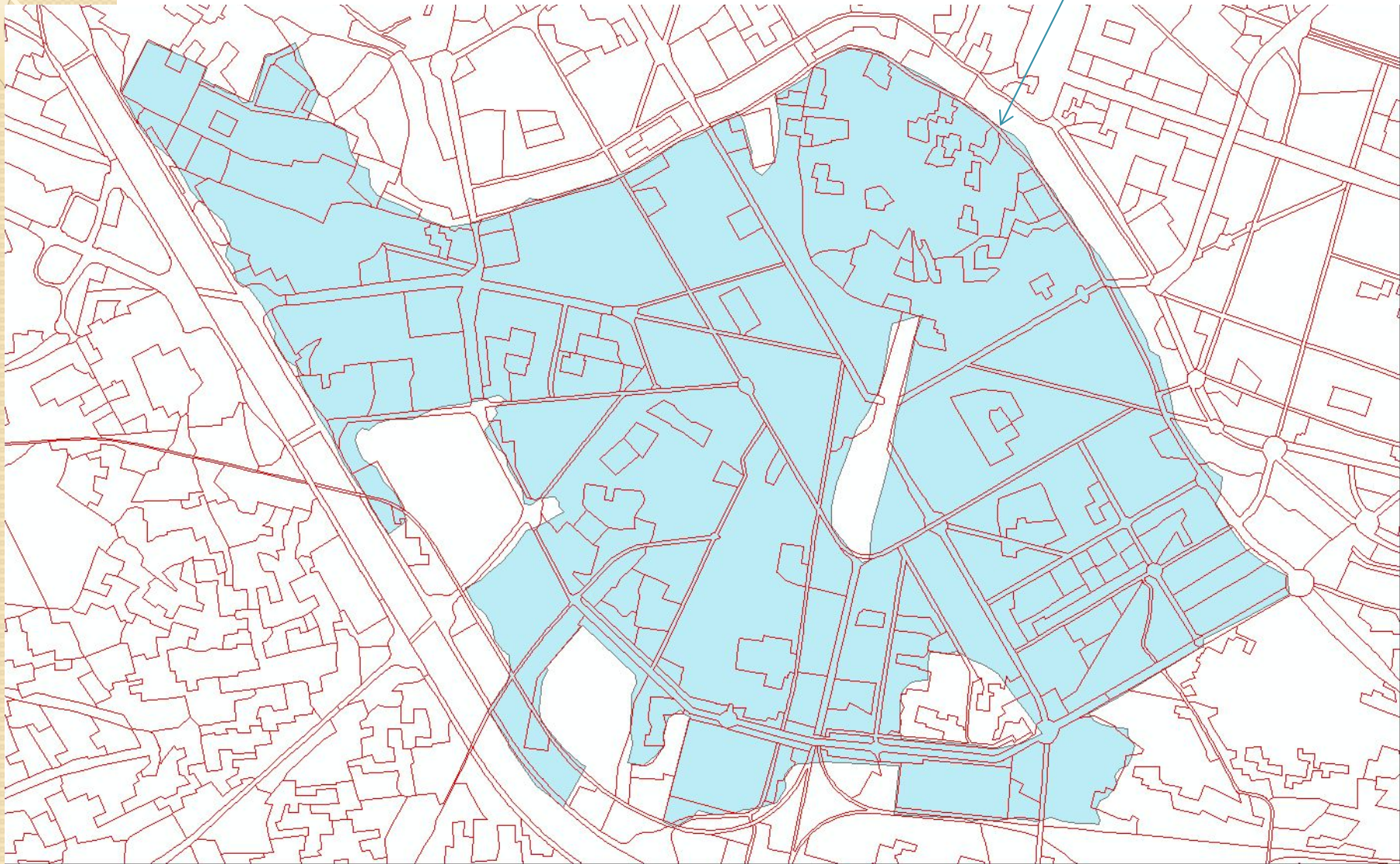 CLC2006:   44 covers (= 44 classes)
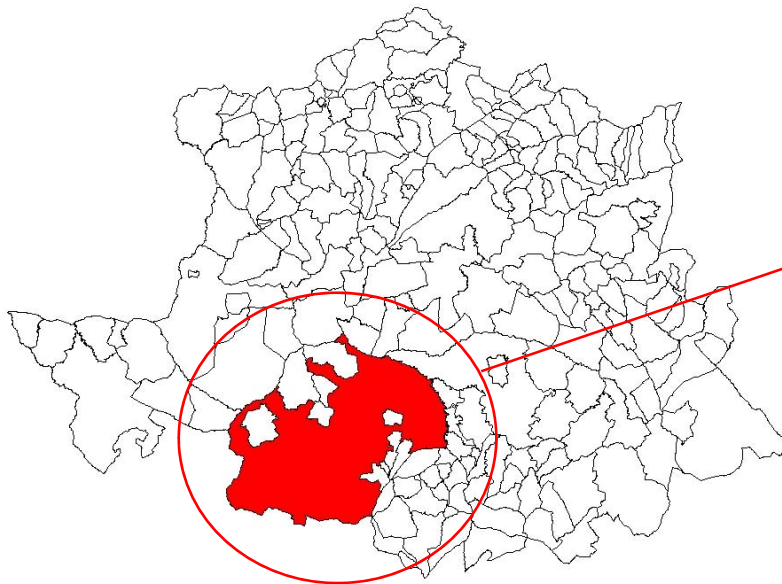 SIOSE2005: 820.632 covers (different combination of classes)

CLC *versus* SIOSE

SIOSE polygons

I CLC polygon

The Data Model: CLC
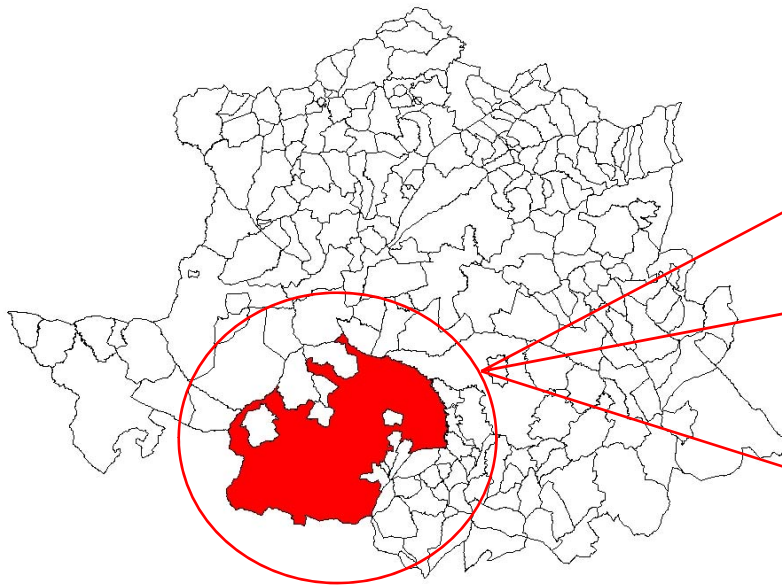
- In CLC only one cover is assigned to each polygon:

Table 3.1 -CORINE land cover nomenclature

Level 1	Level 2	Level 3
1. Artificial surfaces	1.1. Urban fabric	1.1.1. Continuous urban fabric
		1.1.2. Discontinuous urban fabric
	1.2. Industrial, commercial and transport units	1.2.1. Industrial or commercial units
		1.2.2. Road and rail networks and associated land
		1.2.3. Port areas
		1.2.4. Airports
	1.3. Mine, dump and construction sites	1.3.1. Mineral extraction sites
		1.3.2. Dump sites
		1.3.3. Construction sites
	1.4. Artificial non-agricultural vegetated areas	1.4.1. Green urban areas
		1.4.2. Sport and leisures facilities
2. Agricultural areas	2.1. Arable land	2.1.1. Non-irrigated arable land
		2.1.2. Permanently irrigated land
		2.1.3. Rice fields
	2.2. Permanent crops	2.2.1. Vineyards
		2.2.2. Fruit trees and berry plantations
		2.2.3. Olive groves
	2.3. Pastures	2.3.1. Pastures
	2.4. Heterogeneous agricultural areas	2.4.1. Annual crops associated with permanent crops
		2.4.2. Complex cultivation patterns
		2.4.3. Land principally occupied by agriculture, with significant areas of natural vegetation
		2.4.4. Agro-forestry areas
3. Forests and semi-natural areas	3.1. Forests	3.1.1. Broad-leaved forest
		3.1.2. Coniferous forest
		3.1.3. Mixed forest
	3.2. Shrub and/or herbaceous vegetation associations	3.2.1. Natural grassland
		3.2.2. Moors and heathland
		3.2.3. Sclerophyllous vegetation
		3.2.4. Transitional woodland shrub
	3.3. Open spaces with little or no vegetation	3.3.1. Beaches, dunes and sand plains
		3.3.2. Bare rock
		3.3.3. Sparsely vegetated areas
		3.3.4. Burnt areas
		3.3.5. Glaciers and perpetual snow
4. Wetlands	4.1. Inland wetlands	4.1.1. Inland marshes
		4.1.2. Peatbogs
	4.2. Coastal wetlands	4.2.1. Salt marshes
		4.2.2. Salines
		4.2.3. Intertidal flats
5. Water bodies	5.1. Inland waters	5.1.1. Water courses
		5.1.2. Water bodies
	5.2. Marine waters	5.2.1. Coastal lagoons
		5.2.2. Estuaries
		5.2.3. Sea and oceans

8

The Data Model: SIOSE

- In SIOSE a polygon is characterized by many classes, with different spatial structures:

NOMBRE	ETIQUETA	ID
COBERTURA ARTIFICIAL		100
Edificación	EDF	101
Zona verde artificial y arbolado urbano	ZAU	102
Lamina de agua artificial	LAA	103
Vial, aparcamiento o zona peatonal sin vegetación	VAP	104
Otras construcciones	OCT	111
Suelo no edificado	SNE	121
Zonas de extracción o vertido	ZEV	131
CULTIVOS		200
Cultivos Herbáceos		210
Arroz	CHA	211
Cultivos Herbáceos distintos de Arroz	CHL	212
Cultivos Leñosos		220
Frutales		221
Frutales Cítricos	LFC	222
Frutales no Cítricos	LFN	223
Viñedo	LVI	231
Olivar	LOL	232
Otros cultivos leñosos	LOC	241
Prados	PRD	290
PASTIZAL	PST	300
ARBOLADO FORESTAL		310
Frondosas		311
Frondosas Caducifolias	FDC	312
Frondosas Perennifolias	FDP	313
Coníferas	CNF	316
MATORRAL	MTR	320
TERRENOS SIN VEGETACIÓN		330
Playas, dunas y arenales	PDA	331
Suelo desnudo	SDN	333
Zonas quemadas	ZQM	334
Glaciares y nieves permanentes	GNP	335
Ramblas	RMB	336
Roquedo		350
Acantilados marinos	ACM	351
Afloramientos rocosos y roquedos	ARR	352
Canchales	CCH	353
Coladas lavicas cuaternarias	CLC	354

The Data Model: SIOSE

- SIOSE does not classify a given **polygon** within a fixed hierarchical nomenclature, but it allows to assign **one or more covers** to the same polygon, using shares of occupation, that eventually add to 100%.

- In addition, (simple) covers have **attributes**, providing additional information on the spatial distribution of covers, or signaling its potential use.

- This is a much more complex information than the usual thematic maps, but it is also much more versatile and adaptable to the researcher needs.

SIOSE: Simple covers

NOMBRE				ETIQUETA	ID
COBERTURA ARTIFICIAL					100
	Edificación			EDF	101
	Zona verde artificial y arbolado urbano			ZAU	102
	Lamina de agua artificial			LAA	103
	Vial, aparcamiento o zona peatonal sin vegetación			VAP	104
	Otras construcciones			OCT	111
	Suelo no edificado			SNE	121
	Zonas de extracción o vertido			ZEV	131
CULTIVOS					200
	Cultivos Herbáceos				210
		Arroz		CHA	211
		Cultivos Herbáceos distintos de Arroz		CHL	212
	Cultivos Leñosos				220
		Frutales			221
			Frutales Cítricos	LFC	222
			Frutales no Cítricos	LFN	223
		Viñedo		LVI	231
		Olivar		LOL	232
		Otros cultivos leñosos		LOC	241
	Prados			PRD	290
PASTIZAL				PST	300
ARBOLADO FORESTAL					310
	Frondosas				311
		Frondosas Caducifolias		FDC	312
		Frondosas Perennifolias		FDP	313
	Coníferas			CNF	316
MATORRAL				MTR	320
TERRENOS SIN VEGETACIÓN					330
	Playas, dunas y arenales			PDA	331
	Suelo desnudo			SDN	333
	Zonas quemadas			ZQM	334
	Glaciares y nieves permanentes			GNP	335
	Ramblas			RMB	336
	Roquedo				350
		Acantilados marinos		ACM	351
		Afloramientos rocosos y roquedos		ARR	352
		Canchales		CCH	353
		Coladas lavicas cuaternarias		CLC	354

- The data model begins with a basic list of 40 **simple covers**

COBERTURAS HÚMEDAS				400
	Humedales continentales			410
		Zonas pantanosas	HPA	411
		Turberas	HTU	412
		Salinas continentales	HSA	413
	Humedales marinos			420
		Marismas	HMA	421
		Salinas marinas	HSM	422
COBERTURA DE AGUA				500
	Aguas continentales			510
		Cursos de agua	ACU	511
		Láminas de agua		512
			Lagos y lagunas	ALG	513
			Embalses	AEM	514
	Aguas marinas			520
		Lagunas costeras	ALC	521
		Estuarios	AES	522
		Mares y océanos	AMO	523

SIOSE: Complex covers

- Simple covers are aggregated into complex covers to describe fully the composition and structure of a polygon.
- Each simple cover is assigned its share of the surface of the polygon.
- Complex covers can be nested.

SIOSE: Predefined Complex covers

- Some complex covers are predefined, and have a particular structure.

NOMBRE	ETIQUETA	ID
DEHESA	DHS	701
OLIVAR VIÑEDO	OVD	702
ASENTAMIENTO AGRÍCOLA RESIDENCIAL	AAR	703
HUERTAS FAMILIAR	UER	704
ARTIFICIAL COMPUESTO		800
Urbano mixto		810
Casco	UCS	811
Ensanche	UEN	812
Discontinuo	UDS	813
Industrial		820
Polígono industrial ordenado	IPO	821
Polígono industrial sin ordenar	IPS	822
Industria aislada	IAS	823
Primario		830
Agrícola/Ganadero	PAG	831
Forestal	PFT	832
Minero extractivo	PMX	833
Piscifactoría	PPS	834
Terciario		840
Comercial y oficinas	TCO	841
Complejo hotelero	TCH	842
Parque recreativo	TPR	843
Camping	TCG	844

NOMBRE	ETIQUETA	ID
Equipamiento/Dotacional		850
Administrativo institucional	EAI	851
Sanitario	ESN	852
Cementerio	ECM	853
Educación	EDU	854
Penitenciario	EPN	855
Religioso	ERG	856
Cultural	ECL	857
Deportivo	EDP	858
Campo de golf	ECG	859
Parque urbano	EPU	860
Infraestructuras		870
Transporte		880
Red viaria	NRV	881
Red ferroviaria	NRF	882
Portuario	NPO	883
Aeroportuario	NAP	884
Energía		890
Eólica	NEO	891
Solar	NSL	892
Nuclear	NCL	893
Eléctrica	NEL	894
Térmica	NTM	895
Hidroeléctrica	NHD	896
Gaseoducto/Oleoducto	NGO	897
Telecomunicaciones	NTC	900
Suministro de agua		910
Depuradoras y potabilizadoras	NDP	911
Desalinizadoras	NDS	913
Conducciones y canales	NCC	912
Residuos		920
Vertederos y escombreras	NVE	921
Plantas de tratamiento	NPT	922

SIOSE: Attributes

- In addition, covers can be assigned an attribute, that provides additional information on the particular cover.

NOMBRE		ETIQUETA	ID
distribucionEspacial			10
	ASOCIACION	A	11
	MOSAICO REGULAR	M	12
	MOSAICO REGULAR	I	13
tipoEdificacion			20
	EDIFICIO AISLADO	ea	21
	EDIFICIO ENTRE MEDIANERAS	em	22
	VIVIENDA UNIFAMILIAR. AISLADA	va	23
	VIVIENDA UNIFAMILIAR. ADOSADA	vd	24
	NAVE	nv	25
enConstruccion	EN CONSTRUCCIÓN	ec	28
irrigacion			30
	SECANO	sc	31
	REGADÍO REGADO	rr	32
	REGADÍO NO REGADO	rn	33
abancalado	ABANCALADO	ab	35
esForzado	FORZADO	fz	36
plantacion	PLANTACIÓN	pl	40
formacionDeRibera	FORMACIÓN DE RIBERA	fr	41
funcionDeCortafuegos	FUNCIÓN DE CORTAFUEGOS	fc	44
cortas	CORTAS	ct	45
procedenciaDeCultivo	PROCEDENCIA DE CULTIVOS	pc	46
altaMontaña	ALTA MONTAÑA	am	47
esRoturadoNoAgricola	ROTURADO NO AGRÍCOLA	ra	48
esZonaErosionada	ZONAS EROSIONADAS	ze	49
esCuaternaria	CUATERNARIAS	cu	50

SIOSE: Polygon labels

- Each polygon has a code label.

1. Polygon with a simple cover:

$$100\text{FDCfr} = \textbf{FDCfr}$$

2. Polygon with a complex cover:

$$\textbf{R(50LFNfzrr\_40CNFpl\_10SDNfc)}$$

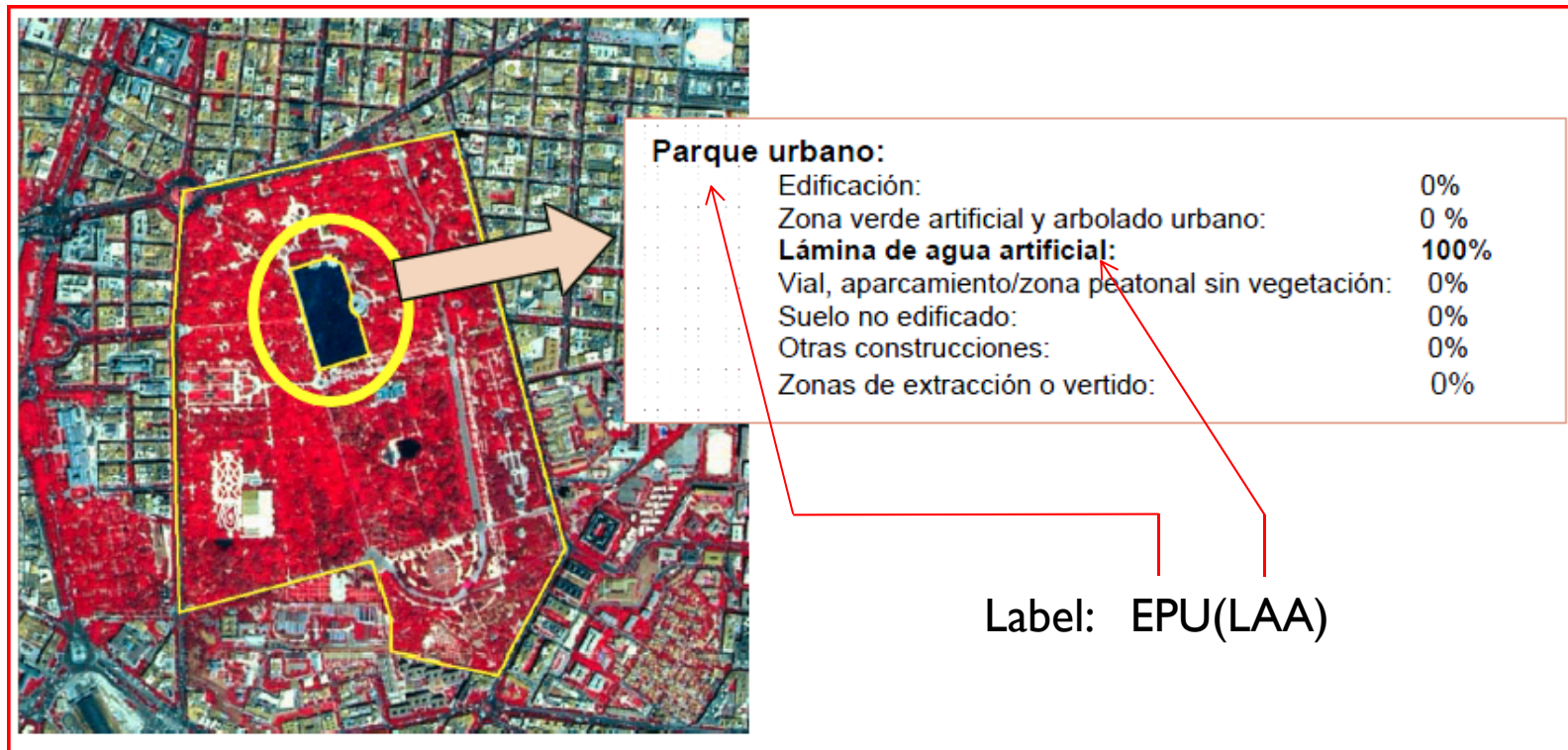3. Polygon with a complex predefined cover:

$$\textbf{UER(30LFCfzsc\_25EDFva\_20CHLfzrr\_20FDPpl\_5LAA)}$$

4. Polygon with a nested complex cover:

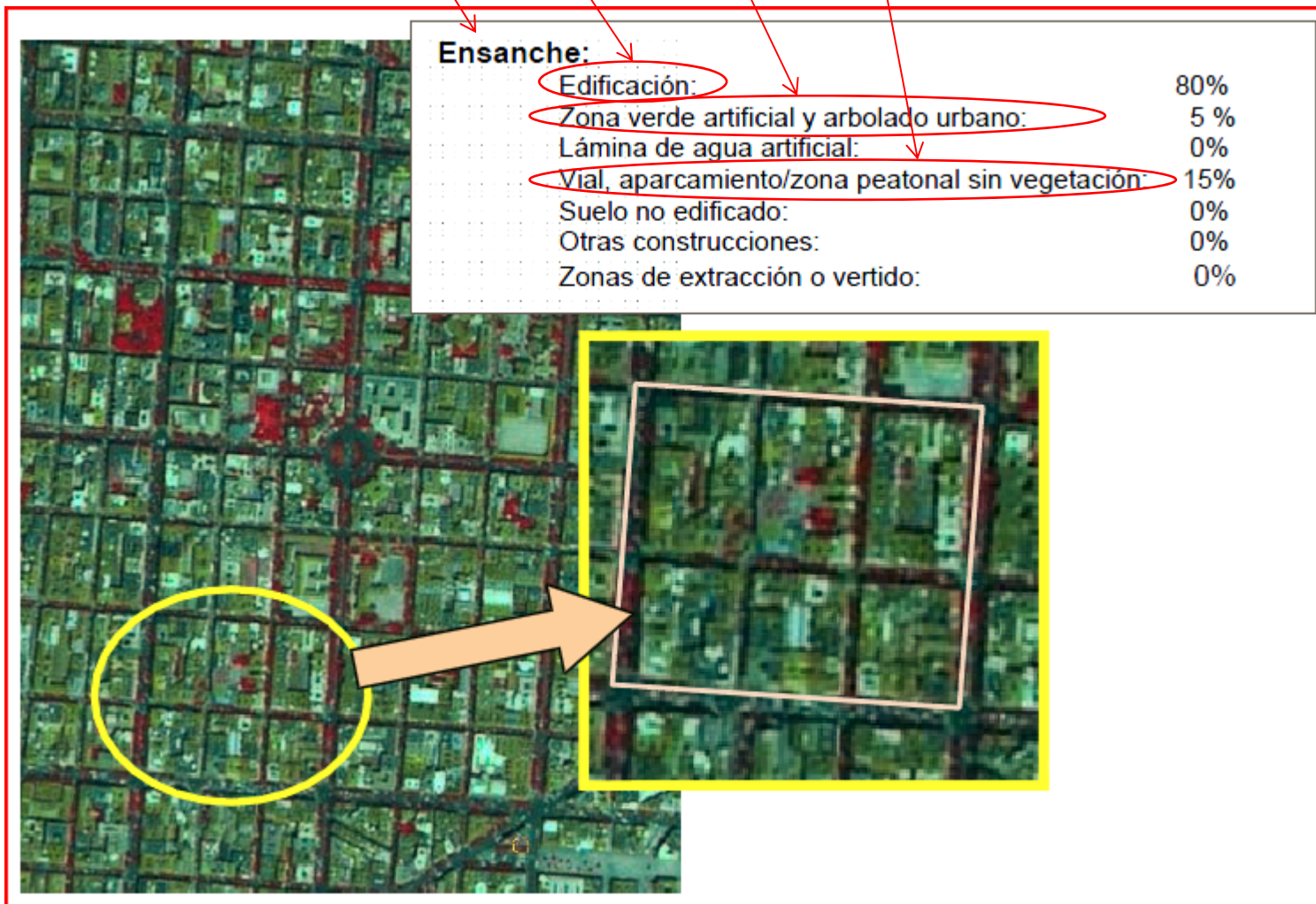$$\textbf{R(80A(70MTRfr\_30ZQM)\_20OVD(90LVlfzsc\_10LOLfzsc))}$$

In this way we get a lot of information

- Artificial water body within a green urban area. In this way we do not lose the information that the lake is inside a city.


Parque urbano:

Edificación:	0%
Zona verde artificial y arbolado urbano:	0 %
Lámina de agua artificial:	**100%**
Vial, aparcamiento/zona peatonal sin vegetación:	0%
Suelo no edificado:	0%
Otras construcciones:	0%
Zonas de extracción o vertido:	0%

Label: EPU(LAA)

Which is specially useful in urban areas

Label: UEN(80EDFem_5ZAU_15VAP)

Ensanche:	
Edificación:	80%
Zona verde artificial y arbolado urbano:	5 %
Lámina de agua artificial:	0%
Vial, aparcamiento/zona peatonal sin vegetación:	15%
Suelo no edificado:	0%
Otras construcciones:	0%
Zonas de extracción o vertido:	0%

Which is specially useful in urban areas

- Because for a given urban polygon we can know the surface occupied by building and by other elements of the urban structure.

1 Vial, aparcamiento o zona peatonal sin vegetación
2 Zona verde artificial y arbolado urbano
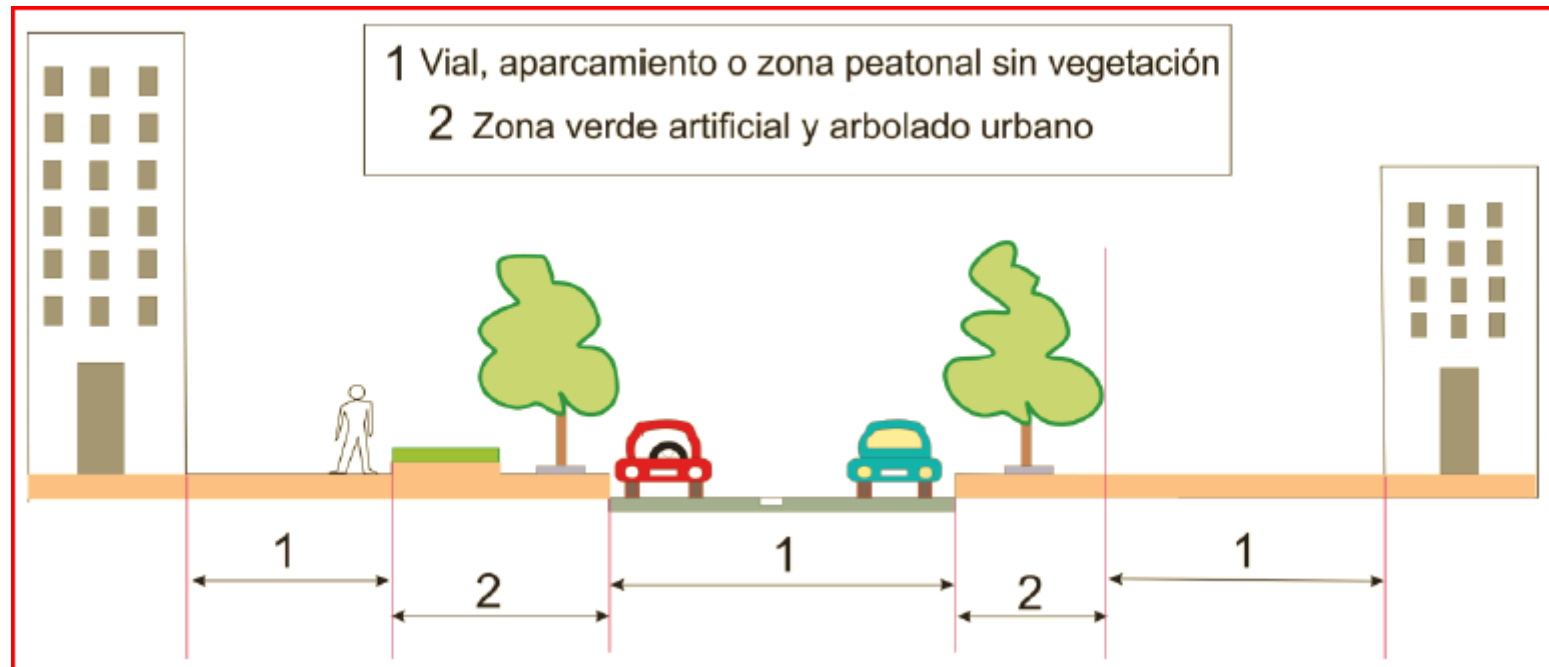
CLC *versus* SIOSE…

- A comparison between both data sets is not straight forward, but at Corine Land Cover level 1, we observe a clear underestimation of the artificial surfaces.

	CLC Level 1 Classification		CLC2006		SIOSE2005	
Level 1			Sup (Km^2)	%	Sup (Km^2)	%
1	Artificial surfaces		10,174	2.01%	18,701	3.70%
2	Agricultural areas		253,642	50.05%	280,339	55.40%
3	Forest and semi-natural areas		238,521	47.07%	202,536	40.02%
4	Wetlands		1,111	0.22%	994	0.20%
5	Water bodies		3,282	0.65%	3,470	0.69%
	Total		**506,730**	**100.00%**	**506,040**	**100.00%**

Downscaling residential population...

- Using SIOSE as land cover information, and census tracts as population data (about 35.000) we build a 1 Km^2 grid for Spain following similar methods to the ones applied by Gallego (2010).

- We proceed as follow:

1. Determine the polygons susceptible to hold residential population: All polygons which include the simple cover "buildings" of non-industrial type.

Residential buildings…

- We don´t have population outside these polygons.
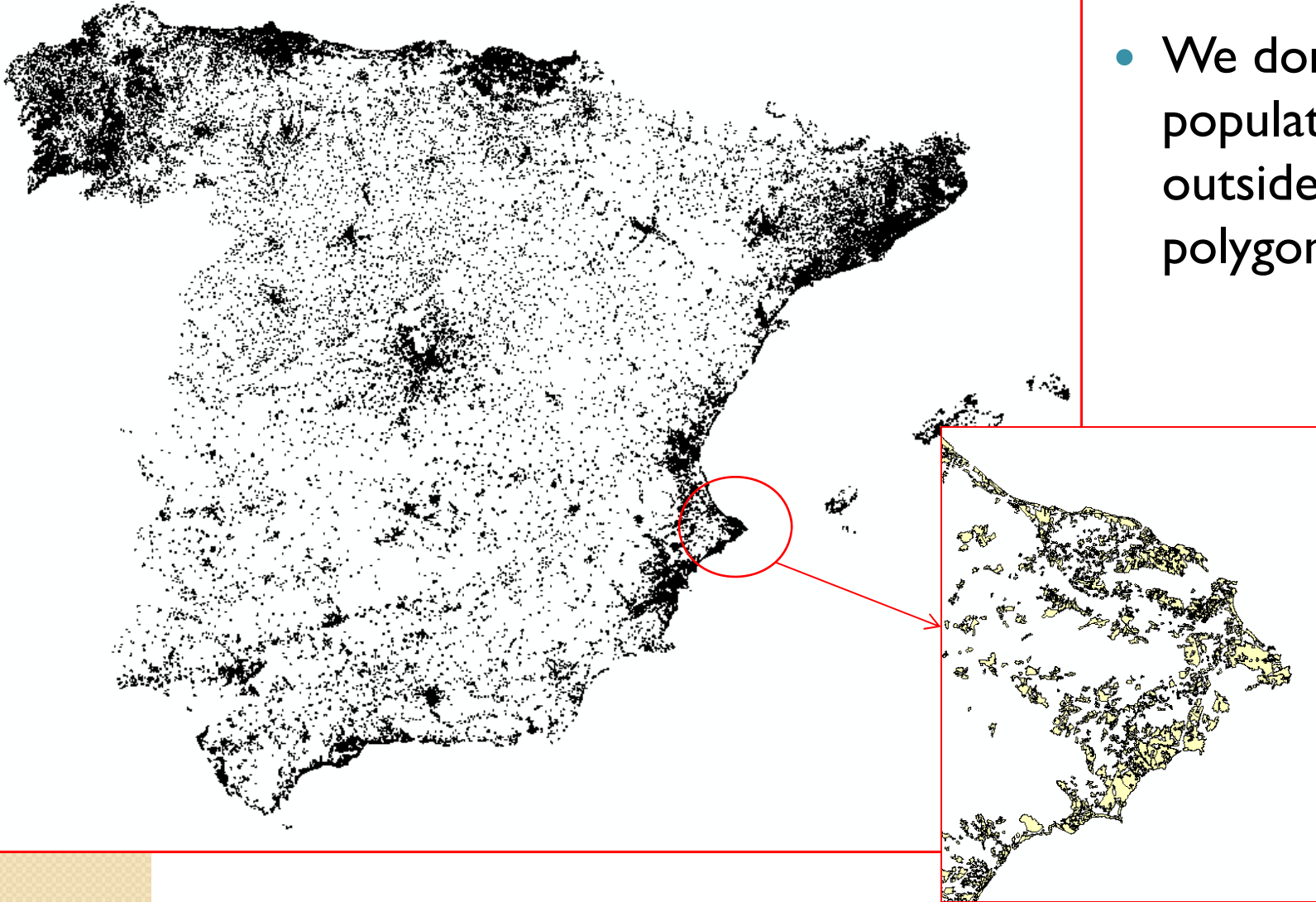
Algorithm…

2. Aggregate residential buildings by type:
4 classes with different threshold densities:

➤ Isolated buildings

➤ Apartment block

➤ Detached house

➤ Semi-detached house

Algorithm…

3. Intersect these polygons with census tracts, so for each census tract we have, at most, 4 classes of residential buildings.

4. But for more than half of the census tracts (55%) we get pure classes (only one type of building).
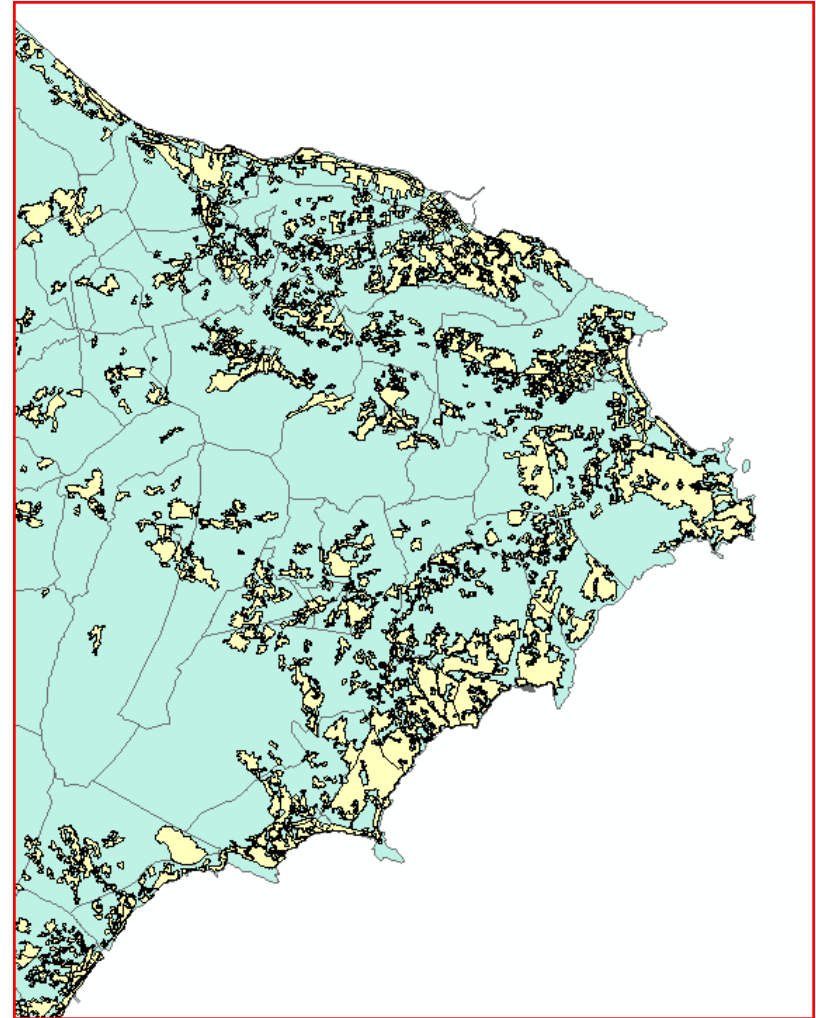
Algorithm…

5. Pure classes cover 50% of the population (and about 33% of the residential surface), and from them we determine average national density by class: $d_c^0 = \dfrac{P_c}{S_c}$.

6. Apply these densities to the rest of census tracts, m, by class, to get an initial population estimate for class c: $P_{c,m}^1 = W_m \times d_c^0 \times S_{c,m}$, where W_m is an scale factor that ensures that the total population of the census tract sums its known value (pycnophylactic constraint).

7. Determine new average national density by class (all census tracts combined), and repeat the process until convergence: $P_{c,m}^i = W_m \times d_c^{i-1} \times S_{c,m}$.
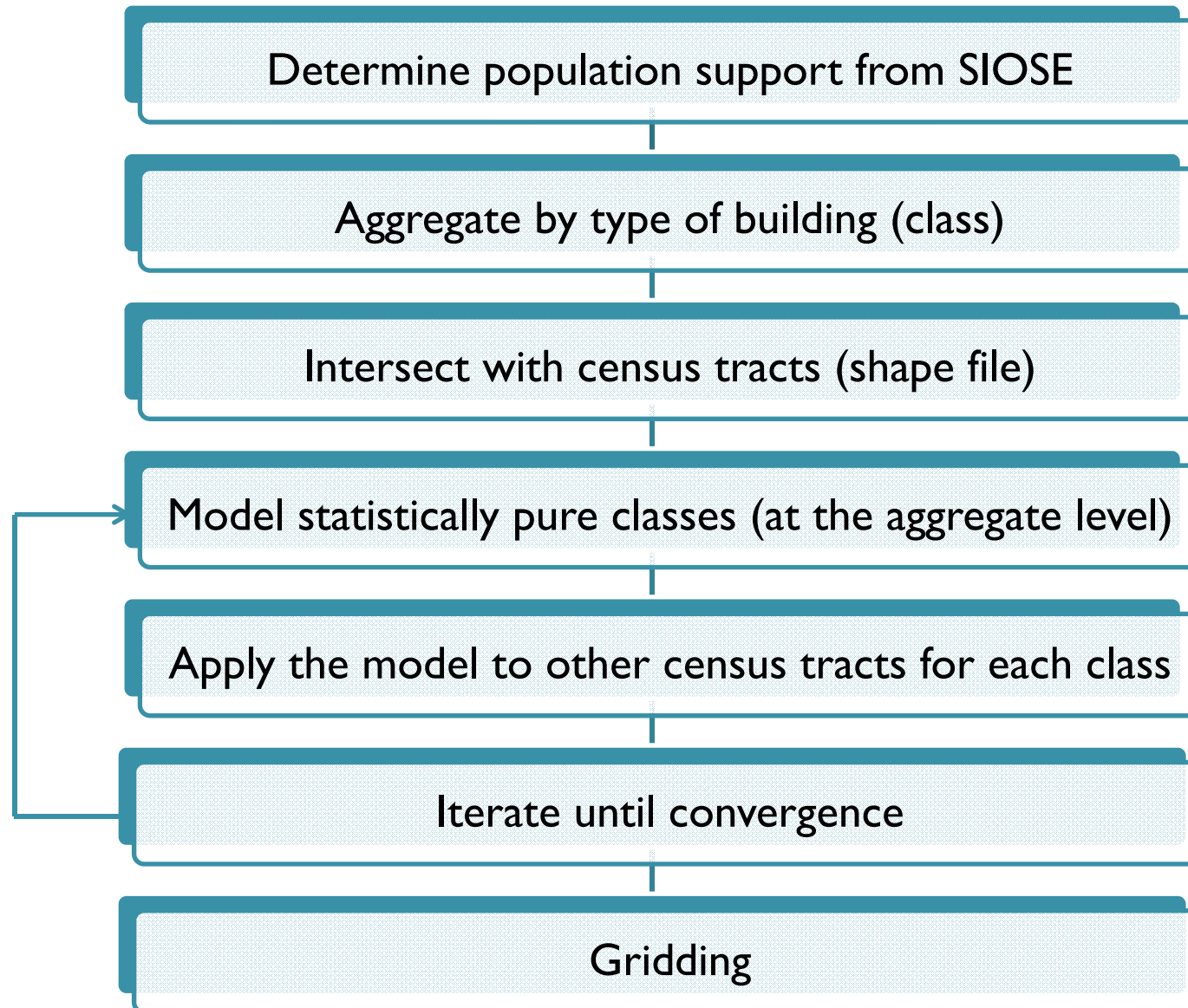
Algorithm…

8. This process redistribute the population of each census tract into its populated polygons by class.
    This vector layer of polygons is useful by its own sake.
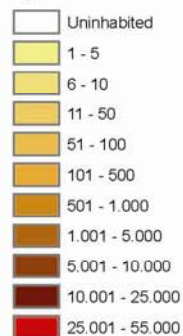
9. Gridding from this layer is straightforward.

Work flow…

Determine population support from SIOSE

Aggregate by type of building (class)

Intersect with census tracts (shape file)

Model statistically pure classes (at the aggregate level)

Apply the model to other census tracts for each class

Iterate until convergence

Gridding
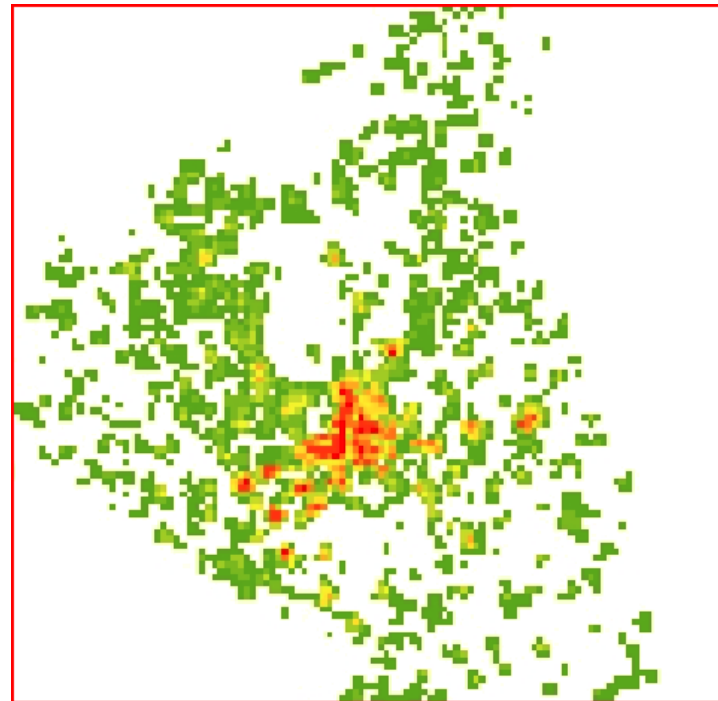
SPAIN POPULATION GRID
Peninsula & Balearic Islands

Number of inhabitants
per individual
square kilometer

- Uninhabited
- 1 - 5
- 6 - 10
- 11 - 50
- 51 - 100
- 101 - 500
- 501 - 1.000
- 1.001 - 5.000
- 5.001 - 10.000
- 10.001 - 25.000
- 25.001 - 55.000

Total area - km² 506.010
Land area - km² 502.540
Total population 47.021.031
Population - grid dataset 46.785.548
Population in grid data set - per cent 99,5
Inhabited grid cells 93.467
Inhabited grid cells - per cent of land areas 18,6
Inhabitants per km² land area 93,6
Inhabitants per inhabited km² 500,6
Max population in one grid cell 54.170
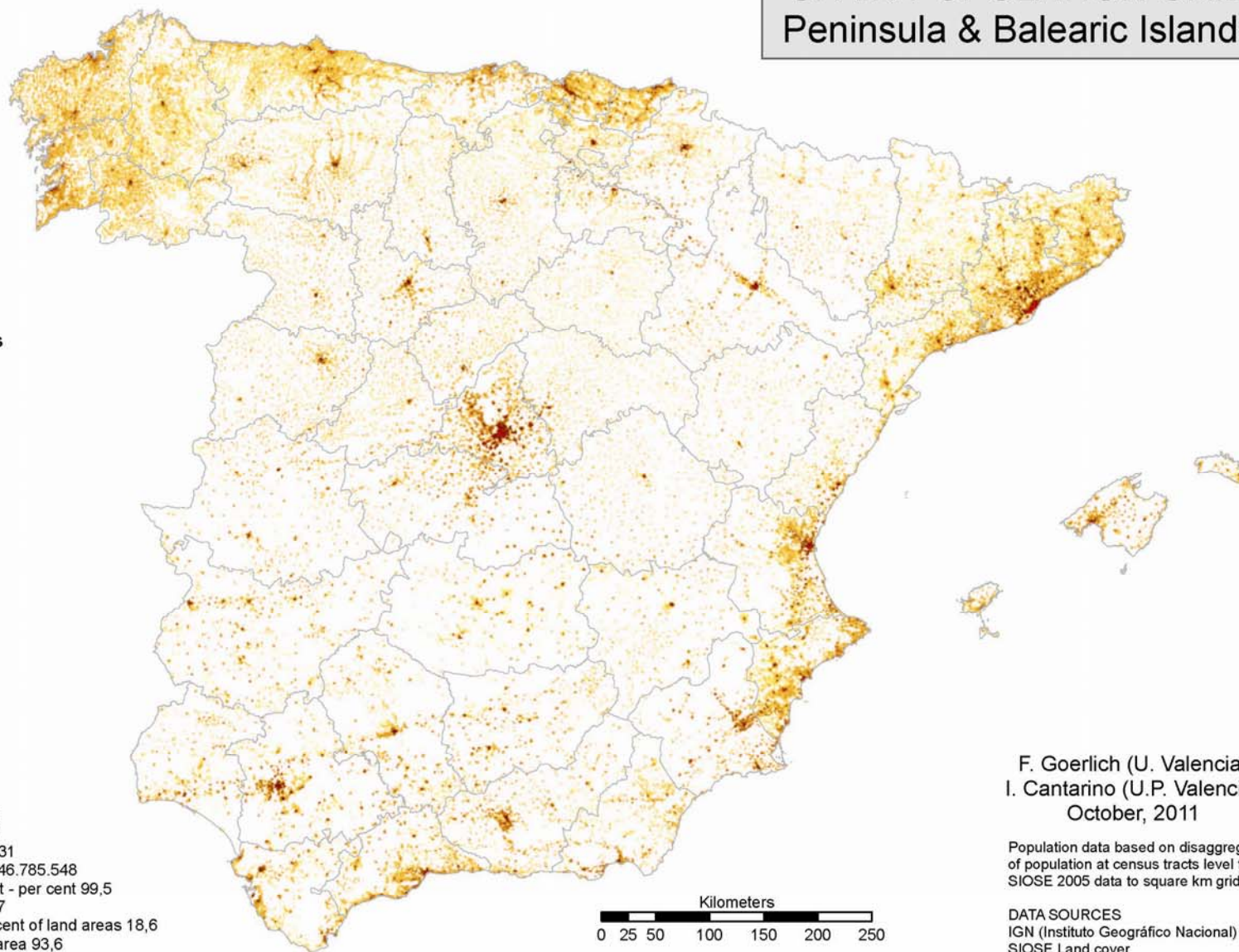
Kilometers
0 25 50 100 150 200 250

1:3.500.000
Proj. LAEA. Datum ETRS 1989

F. Goerlich (U. Valencia)
I. Cantarino (U.P. Valencia)
October, 2011

Population data based on disaggregation
of population at census tracts level through
SIOSE 2005 data to square km grids.

DATA SOURCES
IGN (Instituto Geográfico Nacional) for
SIOSE Land cover
INE (Instituto Nacional de Estadística) for
population and census tracts limits

Validation…

- Validation of our results is a problem, but fortunately we have got a data set with for a NUTS 3 with coordinates for 96% of total population, so it can be very useful for validation.

- We are currently undertaking such evaluation exercise and statistically modeling the densities by class of the pure classes at the census tract level.

- Some work still to be done…

Many thanks for your attention.