

## Geodemografía: coberturas del suelo, sistemas de información geográfica y distribución de la población \*

Francisco J. Goerlich Gisbert \*\*, Isidro Cantarino Martí \*\*\*

**RESUMEN:** Este trabajo examina la reciente aplicación de las técnicas derivadas de los *Sistemas de Información Geográfica* (GIS) al análisis de la distribución de la población sobre el territorio. Se muestran los recientes esfuerzos de los Institutos Nacionales de Estadística Europeos en esta dirección con ocasión del último censo, así como los resultados esperables de dicho trabajo.

El punto de partida es una creciente necesidad de disponer de cifras de población en sistemas zonales no ligados a los caprichosos lindes administrativos de un país o región. Ello permite tanto homogeneizar el espacio físico sobre el que medir la distribución de la población, como evitar distorsiones ligadas al tamaño de los municipios o provincias. Pero el argumento de más enjundia deriva de la necesidad de globalizar las estadísticas; es decir de ser capaces de integrar información cuyo marco de recopilación natural no son las artificiales fronteras delineadas por el hombre, fundamentalmente estadísticas medioambientales y geográficas, con información demográfica y de carácter socio-económico, y poder de esta forma estudiar las relaciones entre economía y medio ambiente en un sistema unificado. Finalmente, el trabajo ilustra una aplicación concreta: la construcción de una *grid* de densidad poblacional para España utilizando como información auxiliar las recientes bases de datos sobre ocupación del suelo (SIOSE), todo ello en el marco de los sistemas de referencia espacial armonizados con la Unión Europea.

**Clasificación JEL:** J11.

**Palabras clave:** población; Sistemas de Información Geográfica; Corine Land Cover; SIOSE; desagregación espacial.

---

\* Los autores agradecen a Matilde Mas los comentarios realizados a una versión inicial de este trabajo. Al mismo tiempo se agradece la ayuda financiera del proyecto del Ministerio de Ciencia y Tecnología ECO2011-23248 y del programa de investigación Fundación BBVA-Ivie. Los gráficos y mapas de este trabajo se aprecian con más nitidez en la versión electrónica en color, que en su traslación al papel impreso en blanco y negro, además de ser más sostenible. <http://www.investigacionesregionales.org>.

\*\* Universidad de Valencia e Instituto Valenciano de Investigaciones Económicas (Ivie).

\*\*\* Universidad Politécnica de Valencia.

Correspondencia: F. J. Goerlich Gisbert, Universidad de Valencia, Departamento de Análisis Económico, Campus de Tarongers, Avda. de Tarongers s/n, 46022-Valencia. E-mail: [Francisco.J.Goerlich@uv.es](mailto:Francisco.J.Goerlich@uv.es).

Recibido: 27 de junio de 2012 / Aceptado: 24 de enero de 2013.

## Geodemography: Land cover, geographical information systems and population distribution

**ABSTRACT:** This paper examines the recent application of the *Geographical Information Systems* (GIS) to the analysis of population distribution. We mention the efforts of the National Statistical Institutes in this direction boosted by the last census 2011.

The starting point is a growing need to have available population figures for areas not related to administrative boundaries, either user defined zones or in grid format. This allows a convenient zonal system to combine demographic characteristics with environmental and pure geographic data, so the relation between the man and the environment can be analyzed in a unified way.

Eventually, we offer a practical illustration of the interactions between GIS techniques and administrative population data in the study of spatial population distribution: We build a density grid for Spain by dasymetric methods from census tracts population data and Land Cover and Use Information System of Spain (SIOSE). The analysis is done within the spatial reference framework of the European Union.

**JEL Classification:** J11.

**Keywords:** population; Geographical Information Systems; Corine Land Cover; SIOSE; spatial disaggregation.

### 1. Introducción

El interés por representar en un mapa la distribución de la población se remonta al propio desarrollo de la cartografía (Reher, 2006). El mapa de densidades más antiguo del que tenemos noticias en España se remonta a los datos demográficos del censo de 1860, y representa, en un mapa de coroplemas, la densidad de población por partidos judiciales. En él se observan ya los rasgos fundamentales de la actual distribución de la población, con su elevada concentración en el litoral y la «isla» de Madrid<sup>1</sup>.

En la actual organización administrativa del Estado español, los municipios constituyen las unidades administrativas menores en las que se divide el territorio nacional y que tienen asignados lindes precisos sobre los que se extienden sus competencias<sup>2</sup>. Por esta razón, y también por cuestiones de disponibilidad estadística, los trabajos que estudian la localización de la población para áreas geográficas amplias suelen hacerlo, en el mejor de los casos, descendiendo al nivel municipal (Reques y Rodrí-

---

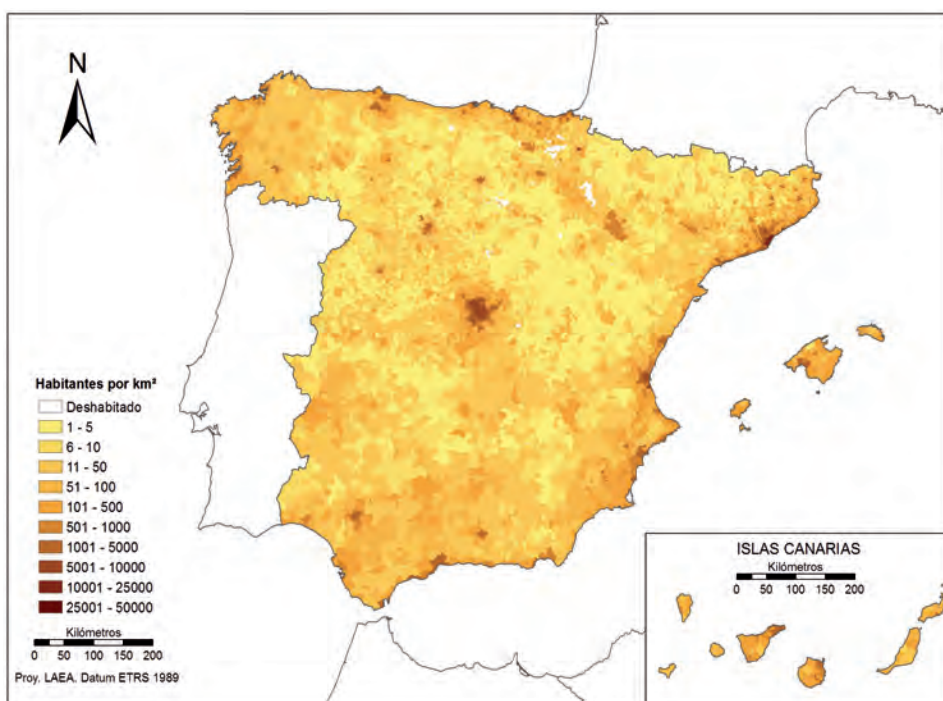
<sup>1</sup> Este mapa está accesible en los fondos cartográficos del Instituto Geográfico Nacional (IGN): <http://www2.ign.es/fondoscartograficos/>.

<sup>2</sup> Es cierto, sin embargo, que por debajo del municipio existe el concepto de Entidad Local de Ámbito Territorial Inferior al Municipio (entidades locales menores, ELM), definido por la Ley 7/1985, de 2 de abril, Reguladora de las Bases del Régimen Local como una unidad de gestión, administración descentralizada y representación política dentro del municipio.

guez, 1998; Zoido y Arroyo, 2004; De Cos y Reques, 2005; Goerlich, Mas, Azagra y Chorén, 2006; Goerlich y Mas, 2008a, 2008b, 2009).

La figura 1 muestra la densidad de población de los municipios españoles a partir del padrón de 2010. Aunque éste es un mapa típico sobre la distribución de la población resulta evidente que es poco realista; por una parte todo el espacio se encuentra aparentemente habitado, por otra los valores representados dependen críticamente de las superficies municipales, que nada tienen que ver con la distribución de la población. Como han reconocido numerosos autores (Reher, 1994), desde el punto de vista del estudio del asentamiento de la población sobre el territorio, la división municipal es claramente insuficiente y debemos aumentar la escala geográfica del análisis. La cuestión es, pues, si podemos mejorar la representación de la distribución espacial de la población con la información actualmente disponible.

**Figura 1.** Densidad de población municipal. 2010



Fuente: Padrón Municipal: Instituto Nacional de Estadística; Instituto Geográfico Nacional y elaboración propia.

Desde el punto de vista geográfico-estadístico, es bien conocido que la división del territorio es relevante para un tema particular si las áreas que constituyen la partición son homogéneas, o cumplen una determinada función, para el problema que deseamos investigar (Vidal, Gallego y Kayadjanian, 2001). No hay una partición geográfica uniformemente mejor para todas las situaciones, sin embargo la captación

de estadísticas demográficas y socioeconómicas parte de áreas administrativas predefinidas, organizadas según una estructura jerárquica, y no relacionadas directamente con los problemas que se pretenden analizar, sino más bien con la estructura de toma de decisiones política.

Este problema es más evidente cuando analizamos variables climáticas o medioambientales, en las que las zonas homogéneas o funcionales de análisis tienen que ver con características geográficas o relacionadas con el paisaje: cortes altimétricos, coberturas del suelo, zonas climáticas homogéneas o cuencas hidrográficas, por ejemplo; pero subsiste igualmente en el caso de variables demográficas o socioeconómicas. Así pues, si deseamos analizar la distribución de la población a partir de datos municipales deberíamos preguntarnos sobre la homogeneidad de los municipios para este fin. En este sentido es bien conocido que los municipios españoles presentan un grado considerable de heterogeneidad; su estructura ha sido determinada por razones históricas y sus delimitaciones no responden necesariamente a criterios relacionados con la distribución de la población sobre el territorio, sino más bien a criterios de conveniencia administrativa y voluntad política (Rodríguez, Martín-Asín y Astudillo, 1997; Capdevila i Subirana, 2009). Uno de los problemas más graves en este contexto es la presencia de algunos términos municipales muy grandes, que conviven con algunos de tamaño ridículo, como tendremos ocasión de comprobar más adelante.

De esta forma, un examen detallado de la distribución de la población sobre el territorio exige, por una parte, descender más allá del nivel de agregación municipal, y por otra, superar de alguna forma el rígido marco de los lindes administrativos (Muguruza y Santos, 1989). La superación de los lindes administrativos se hace todavía más necesaria cuando se desea integrar datos demográficos y socioeconómicos con datos geográficos, climáticos o medioambientales, recogidos en un sistema zonal diferente y que nada tiene que ver con estos lindes.

En este sentido, la mejor forma de generar mapas sobre la distribución de la población con arreglo a un determinado criterio, por ejemplo residencia, sería geo-referenciar todos los edificios de un país, determinar la población que reside habitualmente en ellos y contarla. La distribución vendría dada entonces por un conjunto de coordenadas con el volumen de población en cada punto como atributo. Resulta obvio que la impresión visual del mapa resultante de esta aproximación geográfica a la distribución de la población sería muy diferente de la visión que observamos cuando representamos las densidades municipales a partir de un mapa de coropletas, como el dibujado en la figura 1, ya que gran parte de la superficie parecería como vacía, lo que nunca sucede cuando esparcimos la población de forma uniforme sobre un territorio dado.

Esta tendencia a geo-codificar la población según su residencia ha ido ganando popularidad en los últimos años, hasta el punto de que algunos Institutos Nacionales de Estadística Europeos, fundamentalmente aquéllos donde la estadística oficial está fuertemente sustentada por registros administrativos, disponen de la totalidad de su población completamente geo-referenciada.

Las posibilidades de este tipo de información son enormes ya que una determinación precisa de la localización de la población sobre el territorio, no directamente

relacionada con los límites administrativos en los que se divide un país, es absolutamente esencial para numerosas cuestiones prácticas de organización social; y va más lejos de la simple representación cartográfica o del mero estudio descriptivo de la geografía humana. Quizá el ejemplo más directo es el de la determinación de la población en riesgo ante determinadas catástrofes naturales (Tralli *et al.*, 2005). En estos casos una localización exacta de la población es absolutamente esencial, ya que la evaluación de daños, el diseño y la efectividad de políticas de intervención dependen en gran medida de la localización de la población y otros activos físicos y económicos (Chen *et al.*, 2004). Otro ejemplo de interés radica en el análisis de la accesibilidad a determinadas infraestructuras y servicios, y que no puede basarse en «situar» la población de un municipio o región en un solo punto, ya sea éste el núcleo principal o el centro del polígono (Gutiérrez, Gómez y García, 2009). Esta relación entre accesibilidad y localización es fundamental en el desarrollo de los modelos dinámicos de movilidad *intra*-día, lo que a su vez tiene interés en la gestión y planificación de infraestructuras urbanas y en el análisis de la dinámica de las grandes urbes (Smith, 2011).

Es por esta razón por la que Eurostat ha impulsado la geo-codificación de las variables asociadas al Censo de 2011, promoviendo la cooperación de los Institutos Nacionales de Estadística, y atendiendo a las recomendaciones de la directiva comunitaria INSPIRE. El Instituto Nacional de Estadística (INE) hace especial énfasis en el proyecto censal (INE, 2011) en la geo-referenciación de todos los edificios que tengan algún inmueble que sea una vivienda. De esta forma se pretende generar un sistema de difusión de resultados censales por áreas definidas por el propio usuario, sin correspondencia directa con los ámbitos administrativos tradicionales.

La combinación de estadísticas demográficas geo-referenciadas con otro tipo de información, por ejemplo medioambiental, requiere una unidad de comparación, ya que manejar coordenadas, no coincidentes normalmente, resulta extremadamente complejo y poco operativo en la práctica. Dicho de otra forma necesitamos un sistema zonal que constituya la base de comparación. Descartados los lindes administrativos por su artificialidad, una alternativa natural en un contexto espacial son las **zonas geográficamente regulares o de rejilla cartográfica (*grid*)**, en el que todas las unidades tienen el mismo tamaño. La base de este sistema zonal está en la propia cartografía, se trata de efectuar una partición del espacio con una base geográfica, evitando cualquier hipótesis *a priori* sobre la distribución espacial de la información en la que estamos interesados y sin ninguna referencia ni a zonas funcionales ni a regiones administrativas. Las rejillas cartográficas existen desde la propia aparición de los mapas, ya que fueron utilizadas, en un principio, como sistema de referencia para la localización de accidentes geográficos.

Este sistema zonal ha ganado popularidad en los últimos años por varias razones: i) La observación de la Tierra mediante satélite facilita la recogida de la información en este sistema zonal. ¡Desde el espacio no se observan fronteras! De hecho la mayor parte de la información de tipo medioambiental se recoge actualmente en este formato, y posteriormente se procesa en otro formato si se considera conveniente. ii) El desarrollo de las técnicas de los Sistemas de Información Geográfica (GIS) ha hecho

tremendamente eficiente la manipulación y procesado de grandes volúmenes de información almacenada en este tipo de formato. En este caso la *grid* es básicamente un paso intermedio para el procesamiento posterior de la información. iii) Convenientemente definidas, las *grids* aumentan la comparabilidad entre las unidades de partida, al tener todas ellas el mismo tamaño. Ello es cierto independientemente de la información que contengan. iv) Los últimos desarrollos muestran que las *grids* juegan un papel primordial en el trasvase de información entre sistemas zonales, lo que es esencial en la integración entre estadísticas demográficas y socioeconómicas por una parte, y estadísticas medioambientales por otra. En este caso, una *grid* puede ser el sistema zonal último en el que almacenar información, que será rápidamente comparable, o simplemente un eslabón intermedio en el proceso de manipulación de la misma.

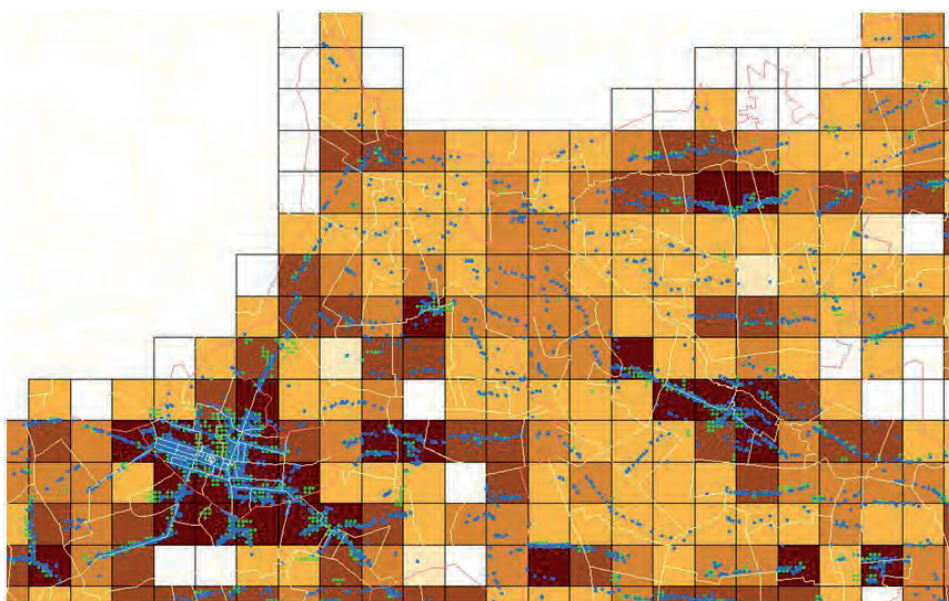
Una vez definida la rejilla de referencia, para lo cual existen ya estándares europeos (INSPIRE, 2010a; 2010b), la construcción de una *grid* de densidad de población es inmediata si disponemos de un fichero de población geo-referenciada: basta con agregar los puntos de población que caen dentro de cada celda de la *grid*. Esta forma de proceder, que construye el mapa desde su base, se conoce como la *bottom-up approach*.

Cuando la información de partida no es suficiente para lograr este resultado exacto, y partimos de datos demográficos por unidades administrativas (secciones censales, municipios o provincias), debemos utilizar métodos de desagregación espacial con información auxiliar. Esta forma de proceder se conoce como *downscaling* o *top-down approach*. Normalmente en estos casos se procede en dos etapas: i) Inicialmente se acota la población dentro de las unidades administrativas a partir de información sobre los edificios o los usos del suelo. Se genera de esta forma lo que se conoce como un mapa dasimétrico, es decir, un mapa cuyos lindes definen zonas homogéneas, con la intención de representar información cuantitativa de forma que represente lo más fielmente posible la realidad estadística subyacente. ii) La población de los polígonos resultantes se re-muestra en las celdas de la rejilla. La calidad del resultado final dependerá en gran medida de la resolución y calidad de la información de partida, y en menor medida del algoritmo concreto de desagregación espacial utilizado (Martin, Tate y Langford, 2000), pero al tratarse de un método estadístico es necesariamente menos perfecto que el anterior, y es inevitable etiquetar celdas como vacías, cuando en realidad contienen población.

La figura 2 contrapone ambos métodos. Los puntos azules representan la población geo-referenciada. Para cada habitante disponemos de sus coordenadas de acuerdo a un determinado criterio, residencia o lugar de trabajo, por ejemplo. La superposición sobre estos puntos de una rejilla cartográfica de la resolución deseada nos genera de forma automática la *grid* que estamos buscando. Esta forma de proceder genera resultados exactos, en el sentido de que si la base de datos es completa, la localización de la población es perfecta. Obsérvese que en la figura 2a, a todas las celdas que contienen puntos azules se les asigna densidad. Por el contrario la



**Figura 2a.** Población geo-referenciada: aproximación *Bottom-up*



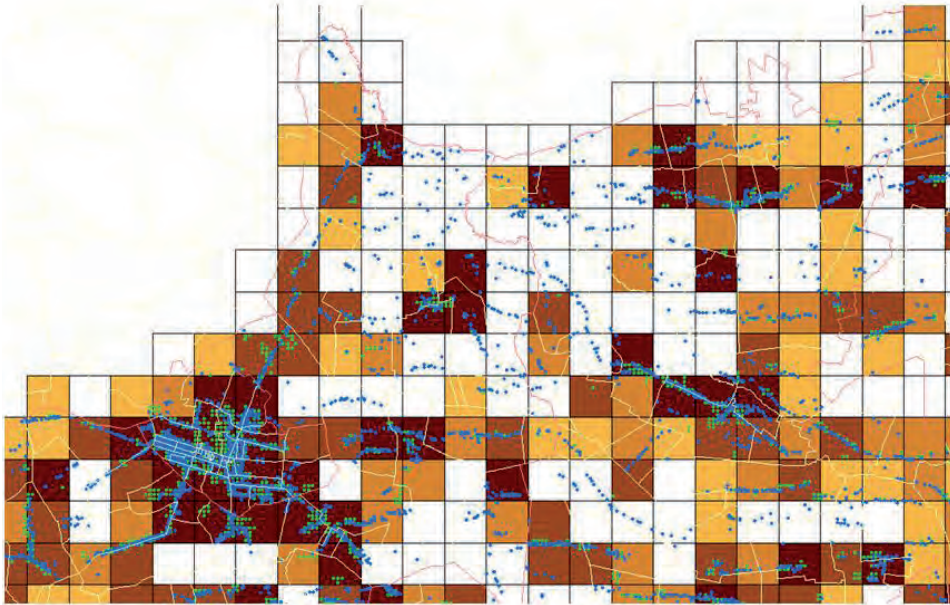
Para cada habitante disponemos sus coordenadas de acuerdo con el criterio de residencia. La superposición de la rejilla cartográfica deriva de forma inmediata la *grid* buscada.  
 Fuente: Jablonski (2011).

figura 2b representa el resultado de un ejercicio de desagregación espacial, y observamos celdas en blanco con puntos azules, es decir, celdas a las que el método de desagregación espacial no asigna población, aunque en realidad sí la tienen. Si mantenemos la restricción de volumen sobre la población representada, esto es, que la suma de la población de las celdas de una determinada área administrativa sume la población conocida de dicha área, ello implica que en otras celdas tendremos más población de la que deberíamos.

Por las razones que ya hemos mencionado, Eurostat, a través del Centro Común de Investigación (Joint Research Centre, JRC) de la Unión Europea (UE), impulsó la elaboración de una *grid* de población a nivel europeo que desagregara la población municipal mediante la utilización de Corine Land Cover (CLC) como información auxiliar (Gallego y Peedell, 2001), y de esta forma mostrara una distribución más realista de lo que se observa en un mapa de coropletas. Gracias a este esfuerzo existe una *grid* de población para España con resolución de una ha., fecha de referencia el censo de 2001, y que se puede obtener en formato *raster* del sitio web de la Agencia Europea del Medio Ambiente (AEMA; European Environment Agency, EEA)<sup>3</sup>. Esta *grid* ha sido ya actualizada a fecha de referencia 2006, si bien su resolución se ha

<sup>3</sup> <http://www.eea.europa.eu/data-and-maps/data/population-density-disaggregated-with-corine-land-cover-2000-2>.

**Figura 2b.** Mapa dasimétrico: aproximación *Top-down* o *downscaling*



Información auxiliar sobre usos del suelo, contornos urbanos o de edificios residenciales, es utilizada como información auxiliar en el algoritmo de desagregación espacial.

Fuente: Jablonski (2011).

fijado en un  $\text{km}^2$ <sup>4</sup>, y se espera que tras el censo de 2011 se actualice de nuevo, esta vez utilizando la población geo-referenciada para la mayoría de países.

El resto del trabajo se estructura de la siguiente forma. La sección siguiente describe, en primer lugar, los esfuerzos realizados en el contexto europeo por generar una *grid* de población utilizando CLC como información auxiliar. Se muestra cómo, en el caso de España, los resultados de dicho ejercicio ofrecen una calidad muy pobre, lo que deriva de la escasa resolución de CLC para este fin, de la heterogeneidad de los municipios españoles y del elevado grado de concentración de la población española. Dado que el fin último del trabajo es mejorar para España la *grid* construida a nivel europeo, en segundo lugar, se presenta una base de datos de usos del suelo alternativa, de reciente publicación por parte del Instituto Geográfico Nacional (IGN), así como el resto de información utilizada en el ejercicio de desagregación espacial. Los detalles más técnicos sobre el proceso de construcción de esta *grid* se ofrecen en la sección 3. Un epígrafe final ofrece unas breves conclusiones.

---

<sup>4</sup> [http://epp.eurostat.ec.europa.eu/portal/page/portal/gisco\\_Geographical\\_information\\_maps/pops/references/population\\_distribution\\_demography](http://epp.eurostat.ec.europa.eu/portal/page/portal/gisco_Geographical_information_maps/pops/references/population_distribution_demography).



## 2. Coberturas del suelo y distribución de la población

### 2.1. Corine Land Cover y la distribución de la población en España

En 1985, la Comisión Europea lanzó el proyecto Corine Land Cover (CLC) con el objetivo de producir bases de datos sobre coberturas y usos del suelo homogenizadas a nivel europeo. Existen en la actualidad tres versiones de CLC que corresponden a los años 1990, 2000 y 2006; estando prevista la siguiente actualización para 2012. La información es pública y está disponible, en formatos *raster* y vectorial, en el sitio web de la EEA. CLC es considerado como el estándar europeo en coberturas del suelo, puesto que ha sido producido de forma armonizada y bajo reglas comunes a partir de la foto-interpretación de imágenes de satélite (*Landsat ETM+*). La nomenclatura de CLC tiene una estructura jerárquica y consta de 44 clases agrupadas en tres niveles. Las superficies artificiales comprenden 11 clases, dos de las cuales son zonas urbanas: continua y discontinua, y donde esperamos que resida la mayor parte de la población.

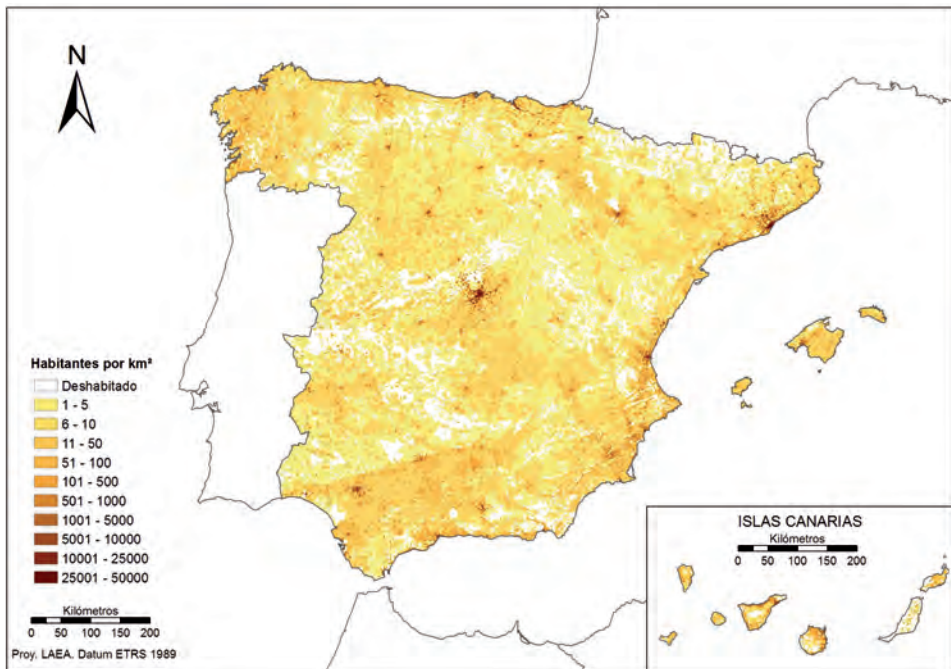
A partir de esta información y la población municipal del censo de 2001 el JRC generó una *grid* de población a nivel europeo, que para España se muestra en la figura 3, una vez agregada a una resolución de un km<sup>2</sup>.

Comparando las figuras 1 y 3 como representación visual de la distribución espacial de la población se observan diferencias claramente visibles. La representación en formato de *grid* de la figura 3 muestra algunos espacios vacíos que corresponden a coberturas del suelo que se suponen no albergan población residente. Sin embargo, es evidente que la figura 3 muestra una población excesivamente dispersa en muchas partes de nuestra geografía. De hecho, un 85% de las celdas de la *grid* contienen población, lo que no encaja con la observación de la geografía española, especialmente en el interior y el sur peninsular, caracterizado por pequeños núcleos de población y grandes espacios deshabitados. La principal razón de este resultado hay que buscarlo en que en muchos municipios CLC no reporta cobertura urbana, lo que unido a la restricción de volumen (la población de las celdas de un municipio deben sumar la población de dicho municipio), hace que el algoritmo de distribución disperse la población sobre otras coberturas (agrícolas y/o forestales).

Esta falta de precisión en la utilización de CLC como información auxiliar en la desagregación espacial de la población se debe básicamente a su baja resolución para un ejercicio de este estilo: en particular la unidad mínima de mapeo son 25 hectáreas. Polígonos con menor superficie se etiquetan con la cobertura dominante. Si no hay una cobertura dominante clara en el polígono, éste es clasificado como heterogéneo. En CLC 2006, esto ocurre en el 11% de la superficie europea, pero en el 18% de la superficie española.

Otros problemas que contribuyen a los pobres resultados mostrados en la figura 3 hacen referencia a la heterogeneidad en el tamaño y la densidad de los municipios es-

**Figura 3.** Grid de densidad de población de un km<sup>2</sup>. 2001



*Fuente:* Información original: Joint Research Center, Eurostat, European Environment Agency.  
Población: Censo 2001, Instituto Nacional de Estadística.  
Agregación a una resolución de un km<sup>2</sup>: European Forum for Geostatistics y elaboración propia.

pañoles. La tabla 1 ofrece una breve descripción de esta realidad, utilizando las cifras de población del Padrón de 2010. Mientras que en términos de superficie el tamaño medio son 62 km<sup>2</sup>, el tamaño mediano es casi la mitad, 35 km<sup>2</sup>. Es cierto que esta heterogeneidad en tamaños no es tremendamente diferente de la que encontramos en Europa, pero los resultados relativos a la concentración de la población sí lo son. Vale la pena observar que en los 60 municipios de mayor tamaño (0,7%), que cubren el 9% de la superficie nacional (menos de la mitad de lo que encontraríamos en Europa para el intervalo de más de 500 km<sup>2</sup>) vive el 14% de la población, más del doble del porcentaje europeo<sup>5</sup>. En el otro extremo de la distribución, el 5% de la población vive en el 10% de los municipios de superficie más reducida. Estas cifras muestran una clara concentración de la población en las grandes ciudades, con escasos habitantes en los municipios más pequeños, algunos de ellos de un tamaño realmente minúsculo: un 60% de los municipios tienen una población residente inferior a los 1.000 habitantes, y el 13% menos de 100. Esta tendencia hacia una concentración creciente de la población puede ser vista históricamente en España desde hace más

<sup>5</sup> Las comparaciones europeas se basan en Gallego (2010), que utiliza las cifras de población del censo de 2001 y CLC 2000, mientras que nosotros utilizamos el Padrón de 2010 y CLC 2006.

**Tabla 1.** Heterogeneidad de los tamaños municipales y concentración de la población en España

| Tamaño (km <sup>2</sup> )     | Municipios |        | Superficie      |        | Población 2010      |        |
|-------------------------------|------------|--------|-----------------|--------|---------------------|--------|
|                               | Número     | %      | km <sup>2</sup> | %      | Habitantes          | %      |
| (0, 10]                       | 771        | 9,5%   | 4.931           | 1,0%   | 2.562.385           | 5,4%   |
| (10, 100]                     | 6.054      | 74,6%  | 231.903         | 46,0%  | 23.119.824          | 49,2%  |
| (100, 500]                    | 1.229      | 15,1%  | 221.696         | 43,9%  | 14.799.103          | 31,5%  |
| > 500                         | 60         | 0,7%   | 46.105          | 9,1%   | 6.539.719           | 13,9%  |
| Total                         | 8.114      | 100,0% | 504.636         | 100,0% | 47.021.031          | 100,0% |
|                               | Mediana    |        | Media           |        | Desviación Estándar |        |
| Superficie (km <sup>2</sup> ) | 34,9       |        | 62,2            |        | 92,3                |        |
| Población 2010                | 582        |        | 5.795           |        | 47.527              |        |

Fuente: Elaboración propia a partir de INE (<http://www.ine.es>). Padrón Municipal y Sección de Territorio.

de un siglo, y todavía continúa: la población se desplaza de las zonas montañosas y las áreas rurales hacia la costa, los valles y las capitales provinciales (Goerlich y Mas, 2008a; 2008b; 2009).

El reducido peso poblacional de muchos municipios y la asimetría en la distribución de tamaños es claramente visible si comparamos el tamaño medio, 5.795 habitantes, con el tamaño mediano, 582 residentes. Además, la población no se encuentra dispersa sobre el territorio en la mayor parte de España, sino más bien al contrario, concentrada en pequeños núcleos urbanos dentro de cada municipio. Dadas las 25 ha. de unidad mínima de mapeo de CLC, esto implica que en muchos municipios CLC no reporta área urbana, simplemente porque no contienen ningún polígono urbano de más de 25 ha. dentro de su término municipal. Gallego (2010: 463) reporta que esto sucede en el 29% de los municipios europeos a partir de CLC 2000, pero en España la situación es mucho peor. A nivel nacional, CLC 2006 no reporta área urbana (clases 1.1.1 y 1.1.2) en más de la mitad de los municipios (57%), aunque estos municipios albergan sólo al 4% de la población, ocupan un 39% de la superficie nacional. No es necesario indicar que estos municipios sí disponen de áreas urbanas donde albergar su población, aunque sean de un tamaño tan reducido que no sean captadas por CLC. A nivel provincial la situación es muy heterogénea: en diez provincias CLC 2006 no reporta área urbana en más del 75% de los municipios, y en dos provincias esta cifra se eleva al 90%.

Estas cifras todavía son más llamativas si nos restringimos a las superficies artificiales. En la mitad de los municipios CLC 2006 no informa sobre ningún tipo de cobertura artificial. A nivel provincial, en seis provincias esto sucede en más del 75% de sus municipios. Como ya hemos indicado, dada la restricción de volumen a nivel municipal (Tobler, 1979), la ausencia de superficie urbana implica la distribución de la población sobre clases no urbanas; el resultado es que la población en superficies agrícolas y forestales es sobreestimada, incluso si *a priori* las densidades iniciales se

mantienen en valores muy reducidos. Además, este efecto es muy heterogéneo por provincias. De esta forma el mapa de la figura 3 muestra una distribución espacial de la población más dispersa de lo que encontramos en la realidad.

Precisamente la falta de resolución de CLC para este tipo de ejercicio, es lo que ha motivado la incorporación de información adicional en la actualización a 2006 de la *grid* mostrada en la figura 3. Para refinar CLC se utiliza básicamente la capa de sellado del suelo de la EEA, con resolución de una ha. (Kopecky y Kahabka, 2009), y otra información geográfica de mayor resolución que la de CLC (Batista e Silva, Lavalle y Koomen, 2012). El objetivo es proporcionar una distribución más realista de la población cuando no se dispone de su geo-referenciación. Como veremos a continuación, este tipo de complejidades son innecesarias en el caso español, ya que disponemos de información sobre coberturas del suelo de elevada resolución y cobertura la totalidad del territorio nacional. Esta información puede ser utilizada para una localización muy precisa de la población residente dentro de las áreas administrativas, y este mapa dasimétrico puede ser posteriormente re-muestreado en el formato de *grid* a la resolución deseada. A continuación describimos la información de base, para presentar el ejercicio de desagregación y sus resultados en la sección siguiente.

## **2.2. Sistema de Información sobre Ocupación del Suelo en España (SIOSE)**

A partir de la experiencia en la elaboración de CLC, el IGN desarrolló un nuevo modelo de Sistema de Información sobre la Ocupación del Suelo en España (SIOSE) con la misma fecha de referencia que la última versión de CLC, 2006. SIOSE trata de resolver muchos de los problemas que plantea CLC y ha sido generado a partir de la foto-interpretación de las imágenes de satélite SPOT5, fotografía aérea, mapas topográficos a escala 1:25.000 e información catastral. SIOSE es público, puede descargarse de forma gratuita en el centro de descargas del IGN (<http://centrodedescargas.cnig.es/CentroDescargas/index.jsp>), y dispone de una dirección web con abundante información sobre el proyecto (<http://www.siose.es/>).

Desde nuestro punto de vista, dos son las características básicas de SIOSE que merecen especial atención. La primera es su resolución, ofrece una escala de 1:25.000 frente a 1:100.000 en CLC, lo que se traslada en una unidad mínima de mapeo variable de acuerdo con el tipo de cobertura:

- Superficies artificiales y láminas de agua: una ha.
- Cultivos forzados, coberturas húmedas, playas, vegetación de ribera y acantilados marinos: 0,5 ha.
- Cultivos y resto de áreas de vegetación natural: 0,5 ha.

Por tanto, en términos de información auxiliar para la distribución de la población, donde las superficies urbanas son la principal cobertura a ser considerada, la unidad mínima de mapeo es de una ha. frente a las 25 de CLC. Con este nivel de resolución todos los municipios presentan algo de área urbana, y por tanto un lugar preciso donde localizar la población dentro de sus términos municipales.

La segunda es el novedoso modelo de datos de SIOSE. Como hemos indicado CLC es un modelo jerárquico: a cada polígono se le asigna una única cobertura de una nomenclatura predefinida que comprende 44 clases a tres niveles de desagregación. SIOSE es en realidad una base de datos orientada a objetos (Villa *et al.*, 2008): el polígono es la unidad básica de análisis, con una cobertura relativamente homogénea. Sin embargo, el objetivo de SIOSE no es clasificar el polígono a partir de una nomenclatura predefinida, sino describir sus contenidos a partir de una serie de reglas. El modelo de datos SIOSE consiste en objetos, atributos, relaciones y reglas de consistencia que se integran en una base de datos geográfica de carácter vectorial y que describe cada polígono en que se divide el territorio nacional. Al contrario que CLC, SIOSE sólo está disponible en formato vectorial.

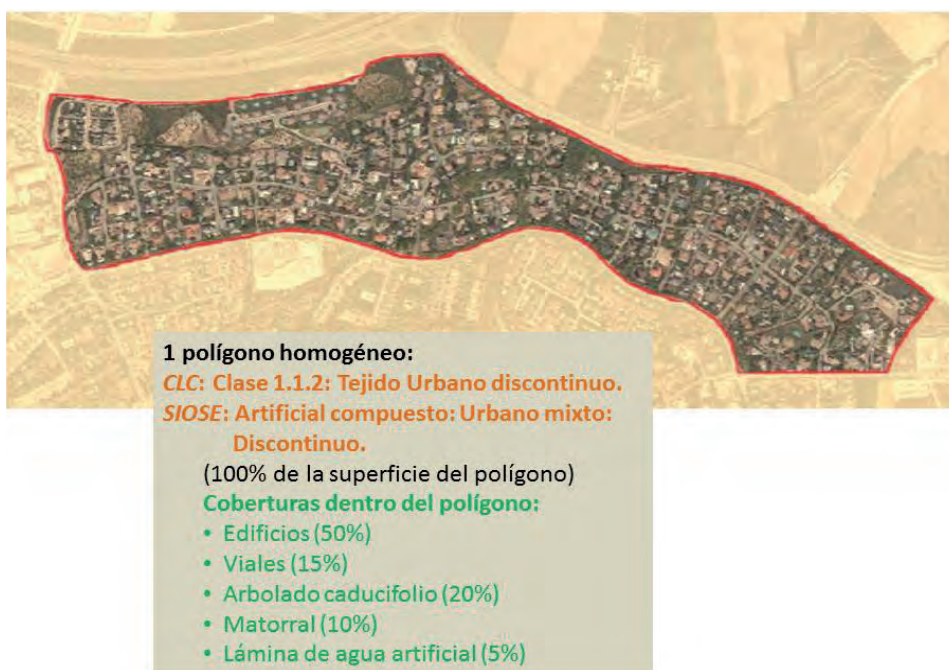
La compleja estructura de SIOSE parte de una nomenclatura de 40 coberturas simples, y la estructura de la base de datos hace posible atribuir una o más de estas coberturas simples a cada polígono, con ciertas restricciones, y sujeto todo ello a determinadas reglas de consistencia. Los polígonos pueden, por tanto, disponer de una «cobertura simple», cuando sólo una cobertura está presente en el polígono (*i. e.* edificios, embalses, prados, suelo desnudo,...); pero más generalmente, cada polígono presenta una «cobertura compuesta», esto es, una combinación de «coberturas simples» (*i. e.* asentamiento agrícola residencial, dehesa, polígono industrial, tejido urbano,...). Los elementos de esta combinación tienen asignados porcentajes de superficie, cuya suma debe ser igual al 100%. Adicionalmente las «coberturas simples» disponen de atributos que nos proporcionan información adicional sobre el contenido del polígono (*i. e.* los edificios pueden ser bloques aislados o casas pareadas). Las «coberturas compuestas» pueden anidarse, y un atributo adicional a éstas nos indica cómo los elementos se distribuyen dentro del polígono: en forma de «asociación» o «mosaico». Aunque no existe una nomenclatura cerrada de «coberturas compuestas», SIOSE incluye una lista de «coberturas compuestas predefinidas». Esta nomenclatura es característica del paisaje español (*i. e.* huerta familiar, olivar-viñedo, tejido urbano de tipo casco, ensanche o discontinuo,...), y ni es exhaustiva del territorio nacional, ni constituye una lista cerrada, pero facilita enormemente la extracción y manipulación de la base de datos.

En principio, SIOSE permite diferentes y potencialmente infinitas combinaciones de las «coberturas simples». La información es muy versátil y puede adaptarse a las necesidades del investigador para el estudio de un problema particular; como contrapartida la manipulación de la base de datos es mucho más compleja que los modelos de coberturas del suelo de tipo jerárquico como CLC.

Un ejemplo visual puede clarificar el complejo modelo de datos de SIOSE. La figura 4 representa un polígono de tejido urbano discontinuo (una urbanización) que sería clasificado en CLC como de la clase 1.1.2: *Tejido Urbano Discontinuo*. El mismo polígono tratado por SIOSE tendría asociada una larga etiqueta que nos describiría su contenido en términos de los porcentajes de superficie que ocupan cada cobertura simple presente dentro del mismo. En primer lugar se nos informaría de que se trata de una cobertura compuesta de *Tejido Urbano Discontinuo*: etiqueta UDS. A continuación, entre paréntesis encontraríamos la lista de etiquetas de las coberturas



**Figura 4.** Corine Land Cover *versus* SIOSE: Tejido urbano



Fuente: Villa (2009).

simples con porcentajes asociados, de esta forma podemos saber que el polígono está compuesto por: 50% de edificios, con un atributo que nos indica que son del tipo vivienda unifamiliar aislada, 15% de viales, 20% de arbolado, con un atributo que indica que son de tipo caducifolio, 10% de matorral, y 5% de lámina de agua artificial (piscinas o estanques). Dado que la etiqueta del polígono sigue estrictas reglas, es posible manipularla para determinar la superficie de cada cobertura simple dentro del polígono. De esta forma, si sabemos que la superficie del polígono es de 1.000 ha., la superficie neta construida es de sólo 500 ha.

Claramente la información proporcionada por SIOSE es mucho más detallada que la proporcionada, no sólo por CLC (dada su mayor resolución), sino por cualquier otro modelo jerárquico de coberturas del suelo. De hecho, CLC 2006 divide el territorio nacional en unos 150.000 polígonos, con un tamaño medio de 3,3 km<sup>2</sup>, perteneciendo a las 44 clases diferentes. En SIOSE disponemos de unos 2,5 millones de polígonos, con un tamaño medio de 0,2 km<sup>2</sup>, y alrededor de 820.000 categorías de coberturas diferentes, en el sentido de combinaciones diferentes de coberturas simples (polígonos diferentes).

Vale la pena observar que, dada la estructura de SIOSE, en relación con la superficie residencial y el cálculo de densidades de población, disponemos de dos superficies de cálculo. La superficie del polígono donde los edificios se localizan:

1.000 ha. en el ejemplo anterior; y la superficie neta residencial edificada dentro del polígono: 500 ha. en el ejemplo anterior. En el ejercicio de desagregación espacial de la población que describimos en la sección siguiente, utilizamos como superficies para el cálculo de densidades la superficie neta edificada. De esta forma esperamos eliminar al máximo los efectos distorsionadores derivados de la heterogeneidad en los tamaños de los polígonos donde asignamos la población. Obsérvese que incluso si la unidad mínima de mapeo para áreas urbanas es de una ha., las coberturas simples se representan dentro de un determinado polígono con tal de que ellas representen al menos un 5% de la superficie total del polígono. Por tanto, esperamos que cualquier área residencial de al menos 500 m<sup>2</sup> haya quedado representada en SIOSE.

### **2.3. Información adicional utilizada**

Además de SIOSE el ejercicio de desagregación espacial descrito en la sección siguiente utilizó la siguiente información de partida.

Para la población utilizamos las secciones censales en lugar de los municipios. Esta información es pública y dispone de una cartografía asociada. Existen en España algo más de 8.000 municipios, pero alrededor de 35.000 secciones censales. Muchos municipios, los más pequeños, sólo tienen una sección censal (el 72% englobando un 6% de la población); en consecuencia no esperamos mejorar mucho en la distribución de la población en las pequeñas áreas rurales. Sin embargo, la utilización de las secciones censales, en lugar de los municipios, permitirá grandes mejoras en ciudades de tamaño medio y grande, como veremos en la sección siguiente, y ello a pesar de que la heterogeneidad en tamaños de las secciones censales supera ampliamente a la de los términos municipales. De hecho, un 60% de las secciones censales tienen una superficie inferior a un km<sup>2</sup>; su tamaño medio es 14,2 km<sup>2</sup>, pero el tamaño mediano es de solamente 0,22 km<sup>2</sup>, dado el reducido tamaño de las mismas en las grandes ciudades. La desviación típica de la distribución de tamaños es casi tres veces la media, alcanzando un valor de 38,7 km<sup>2</sup>.

El año de referencia para la población es el 1 de enero de 2010, y la fuente de información es el Padrón municipal. Se trata pues de la población residente, por lo que estamos hablando de lo que en el argot se conoce como «población nocturna», frente a la «población diurna», cuyo criterio de localización sería el lugar de trabajo o actividad durante el día.

El origen de la cartografía asociada a las secciones censales es un fichero vectorial de todas ellas que proporciona el INE. Debido a que el fichero de 2006, que es la fecha más cercana a SIOSE, estaba lleno de errores topológicos que hacían prácticamente imposible su manipulación en los Sistemas de Información Geográfica, se decidió hacer el ejercicio con el fichero de 2010, que no presentaba errores topológicos, y ofrecía una representación continua y sin solapamientos del territorio.

Con la información demográfica anterior y SIOSE es posible elaborar un mapa dasimétrico que muestre la distribución de la población de forma bastante aproxima-

da a la realidad mediante operaciones estándar en el contexto del GIS y modelización estadística. Si finalmente se desea un formato de *grid* su construcción es directa a partir de la capa vectorial de la misma. El origen de la *grid* de un km<sup>2</sup> procede del sitio web del European Forum for Geostatistics (EFGS, <http://www.efgs.info/data/eurogrid>) y cumple con las directivas de INSPIRE (2010a; 2010b).

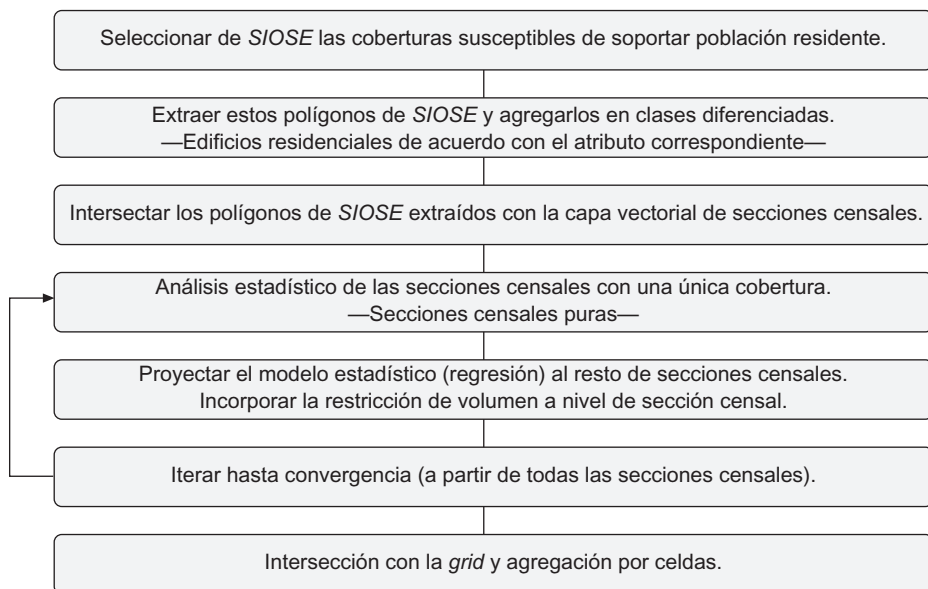
### 3. Desagregación espacial de la población a partir de las secciones censales y SIOSE

Presentamos a continuación un ejercicio de desagregación espacial con el objeto de mejorar la distribución de la población observada en la figura 3 y acercarse lo más posible a la realidad en este campo.

Los métodos finalmente utilizados son básicamente métodos de regresión (Yuan, Smith y Limp, 1997) apoyados por operaciones estándar en el contexto del GIS. Son similares al denominado método de la variable límite, ya utilizado por Gallego, Batista, Rocha y Mubareka (2011), y que se inspira en el trabajo de Eicher y Brewer (2001).

El proceso comienza con la propia definición de densidad, escrita como *Población = Densidad × Superficie*, y una modelización estadística de *Densidad*, dado que *Superficie* es conocida, tras una estratificación y dada la información disponible. El diagrama simplificado del flujo de trabajo se muestra en la figura 5. Finalmente, rea-

**Figura 5.** Diagrama simplificado del flujo de trabajo



lizamos un ejercicio de validación a partir de una *grid* construida a partir del Padrón geo-referenciado para la Comunidad de Madrid<sup>6</sup>.

Todo el trabajo en GIS es realizado a partir de información vectorial, lo que presenta la ventaja de evitar distorsiones derivadas del paso a formato *raster* de información originalmente vectorial, lo que dado el reducido tamaño de algunas secciones censales puede ser importante. El coste es una elevada carga computacional, puesto que la base de datos SIOSE tiene un tamaño de 13,7 Gb para el conjunto nacional.

A partir de la estructura de SIOSE, y tras un detallado análisis de coberturas, elegimos como soporte para la población residente todos aquellos polígonos que incluyan edificaciones de tipo no industrial. Dada la información sobre atributos de las edificaciones, éstas deben ser de los siguientes tipos: bloques de apartamentos, aislados o no, y viviendas unifamiliares, adosadas o aisladas. Obsérvese que los polígonos seleccionados pueden ser urbanos o rurales, puesto que la característica que los define es simplemente que tengan edificaciones residenciales de alguno de los cuatro tipos considerados. Por supuesto, podemos excluir *a priori* coberturas susceptibles de soportar población residente, por ejemplo gente que viva en polígonos industriales, pero incluir más coberturas con generalidad podría tener el efecto contrario, acabar dispersando población sobre coberturas que generalmente no albergan población residente. A partir del fichero de población geo-referenciado para la Comunidad de Madrid comprobamos que el 97,8% de la población cae dentro de los polígonos seleccionados, y que es poco claro que extendiendo la lista de coberturas mejore potencialmente los resultados.

Una vez extraídos los polígonos se agregan de acuerdo con el tipo de edificación, de esta forma finalmente tenemos cuatro tipos. El criterio para la agregación por atributo es simplemente que ésta es la característica que más afecta a la densidad de población, y esta variable es la que necesitamos modelar. Un análisis de varianza para aquellas secciones censales que sólo contenían un tipo de edificación confirmó claramente este hecho. La densidad de población es muy heterogénea de acuerdo no sólo con el tipo de edificio, sino también con arreglo a estratos por tamaño de población. Sin embargo, la mayor variabilidad se observa por tipo de edificio, de forma que finalmente mantenemos cuatro clases diferenciadas en el algoritmo de desagregación.

El siguiente paso consiste en una intersección geométrica de los polígonos extraídos con la capa vectorial de secciones censales. Tras esta intersección muchas secciones censales contienen en ellas un solo tipo de edificación. Éstas son denominadas secciones censales puras, y juegan un papel fundamental en el algoritmo de distribución. Esto no sucede en el caso de CLC, dada su resolución y la utilización de datos municipales, pero en nuestro caso tenemos un 56% de secciones censales con un solo tipo de edificación, muchas de ellas en las grandes ciudades, pero también en pequeños municipios en el caso de viviendas adosadas. En conjunto estas secciones censales engloban al 51% de la población. En estos casos no hay problema de redis-

---

<sup>6</sup> La información de base para la construcción de esta *bottom-up grid* fue amablemente cedida por el Instituto de Estadística de la Comunidad de Madrid (IEM).

tribución; SIOSE ayuda a localizar de forma precisa la población residente dentro de los lindes de la sección censal. La modelización estadística de estas secciones censales puras es esencial para redistribuir la población en el resto.

Dado un conjunto de coberturas  $c$ , con densidades diferenciadas por unidad administrativa (sección censal),  $m$ ; la población total puede ser escrita como

$$P^m = \sum_c P_c^m = \sum_c d_c^m \times S_c^m \quad (1)$$

donde  $d_c^m$  representa la densidad de la clase  $c$  en la sección censal  $m$  y  $S_c^m$  su superficie, que en nuestro caso representa la superficie edificada neta del tipo de edificación correspondiente, tal y como se ha indicado anteriormente.

La ecuación (1) es una identidad. Para secciones censales puras  $c = 1$ , y no hay problema de redistribución; para el resto  $c > 1$ . Conocemos  $S_c^m$ , a partir de la intersección geométrica entre SIOSE y la geometría de las secciones censales, y diferentes métodos de estimar  $d_c^m$  nos proporcionan diferentes algoritmos de reparto. Adicionalmente, el algoritmo debe incorporar la restricción de volumen (Tobler, 1979), de forma que  $P^m = \sum_c P_c^m$ . Esta población es conocida a partir del Padrón.

El caso más simple es suponer que  $d_c^m$  es constante para cada  $c$ , hasta un factor de proporcionalidad,

$$d_c^m = \theta_c \times \lambda^m \quad (2)$$

donde  $\theta_c$  depende sólo de la clase, por ejemplo del tipo de edificación, y puede ser estimado a partir de la información proporcionada por las secciones censales puras para cada  $c$ . Un candidato natural para  $\theta_c$  es la densidad por clase determinada a partir de las secciones censales puras. Estas densidades pueden aplicarse al resto de secciones censales para conseguir una estimación inicial de la población por sección censal y clase, dado  $S_c^m$ . El factor  $\lambda^m$  asegura que, al final del proceso, la restricción de volumen es satisfecha, de forma que finalmente  $P^m = \sum_c P_c^m$ . Sustituyendo (2) en (1)

$$P^m = \sum_c \theta_c \times \lambda^m \times S_c^m = \lambda^m \times \sum_c \theta_c \times S_c^m \Rightarrow \lambda^m = \frac{P^m}{\sum_c \theta_c \times S_c^m} \quad (3)$$

que representa una distribución proporcional de las discrepancias para un  $\theta_c$  dado.

De esta forma, iniciando el proceso con valores iniciales de las densidades por clase, que coinciden con las densidades derivadas de las secciones censales puras, acabamos con las densidades  $d_c^m = \theta_c \times \frac{P^m}{\sum_c \theta_c \times S_c^m}$ . Una vez disponemos de una cifra

de población por clase y sección censal, que satisface la restricción de volumen,  $P_c^m$ . Es posible iterar el proceso hasta la convergencia, con nuevas densidades iniciales calculadas esta vez a partir de todas las observaciones.



Un proceso idéntico puede seguirse si estratificamos las clases de edificios por tamaños municipales. De hecho un análisis de varianza corrobora esta estratificación. Cualquier otra partición sería posible, por ejemplo provincias. Estos métodos son comparables a los introducidos por Mennis (2003, 2009), y es esencialmente el utilizado por Gallego y Peedel (2001) en los primeros ejercicios de desagregación con CLC, convenientemente adaptado a la estructura de nuestra información.

Este tipo de métodos se los conoce en la literatura como métodos de ratio-fija, ya que el supuesto simplificador es que la ratio entre densidades de población de diferentes clases dentro un área administrativa es constante, e igual para todas las áreas administrativas en el mismo estrato. Escapar de esta restricción exige relajar (2), con lo que podemos suponer

$$d_c^m = \theta_c^m \times \lambda^m \quad (4)$$

donde  $\lambda^m$  juega el mismo papel que antes, y  $\theta_c^m$  es un umbral de densidad inicial en el algoritmo de reparto por área administrativa y clase, obtenida a partir de la modelización estadística de las densidades por clase en las secciones censales puras.

Dado que  $S_c^m$  son las superficies netas edificadas, debemos esperar una relación claramente positiva entre densidades a nivel de sección censal por clase y densidad a nivel de municipio, puesto que en este caso la heterogeneidad en tamaños no juega ningún papel. En logaritmos, las correlaciones simples varían entre 0,68 para bloques de apartamentos aislados y 0,91 para viviendas unifamiliares adosadas. Esto sugiere estimar una relación log-log para las secciones censales puras de la forma

$$\log d_c^m = \alpha + \beta \log d^n + u \quad (5)$$

donde  $d^n$  es la densidad a nivel municipal en términos de superficie edificada neta,  $d^n = \frac{P^n}{S^n}$ . La proyección de este modelo para el resto de secciones censales proporciona una estimación de  $\theta_c^m$  en (4).

Sin embargo, disponemos de más información demográfica a nivel de sección censal que puede ser potencialmente incorporada en la regresión (5), por lo que es natural aumentarla con regresores adicionales

$$\log d_c^m = \alpha + \beta \log d^n + \gamma' x^m + u \quad (6)$$

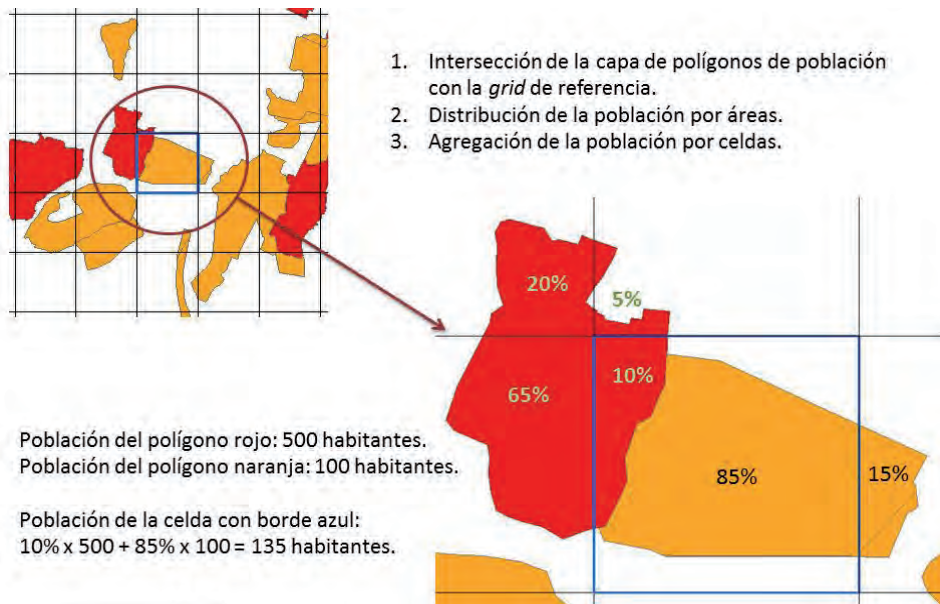
Las variables incluidas en  $x^m$  son el porcentaje de población mayor de sesenta y cinco años, el porcentaje de población extranjera, y el porcentaje de residentes en la sección censal que nacieron en el mismo municipio al que pertenece dicha sección. Estas variables fueron significativas en la regresión (6) y finalmente incluidas al objeto de permitir una mayor generalidad en los umbrales de densidad inicial en el algoritmo de distribución.

En resumen: i) Estimamos (6) para las secciones censales puras de forma separada para cada clase. ii) Proyectamos el modelo al resto de secciones censales para obtener umbrales de densidad inicial, diferentes para cada clase y sección censal,  $\theta_c^m$ , con los que iniciar la distribución. iii) Ajustamos las discrepancias con respecto a la cifra de población conocida de la sección censal correspondiente. iv) Finalmente, una vez hemos estimado la población para todos los polígonos de las diferentes clases dentro de cada sección censal, re-estimamos la ecuación (6), ahora con el total de secciones censales, y repetimos el proceso de redistribución hasta la convergencia de los coeficientes en la regresión. De esta forma construimos un mapa dasimétrico en el que dentro de cada sección censal tenemos uno o varios polígonos con su población residente.

El método es similar al de la variable límite modificado de Gallego, Batista, Rocha y Mubareka (2011), sin embargo nosotros modelizamos nuestros propios datos en lugar de utilizar información auxiliar a partir de *grids* de referencia *bottom-up* de otros países. Ello es posible gracias a la elevada resolución de nuestras bases de datos de partida.

Después de que este proceso ha sido completado, la elaboración de la *grid*, al nivel de resolución deseado, es directa. Efectuamos una intersección geométrica entre la capa vectorial de la población por sección censal y clase, con la *grid* también en formato vectorial, y agregamos la población mediante una distribución por áreas (*areal weighting*). La figura 6 ilustra el proceso de forma gráfica.

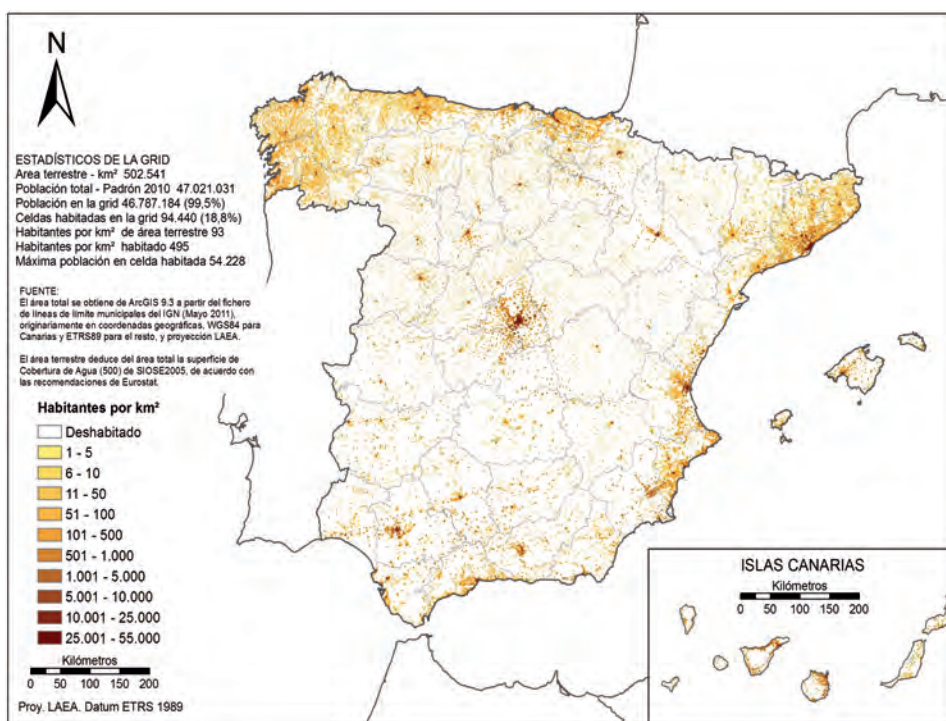
**Figura 6.** Generación de la *grid*: distribución proporcional por área



Fuente: IGN, SIOSE2005 y elaboración propia.

La *grid* resultante de un km<sup>2</sup> de resolución, con sus correspondientes estadísticos y distribución de frecuencias, se muestra en la figura 7<sup>7</sup>. De acuerdo con nuestros cálculos la superficie habitada apenas alcanza el 20% de la superficie terrestre<sup>8</sup>, y mientras que la densidad promedio para el conjunto nacional es de alrededor de 93 habitantes por km<sup>2</sup>, la densidad por kilómetro cuadrado habitado es notablemente superior, 495 habitantes. Estimamos 2.852 celdas con un solo habitante, y 2.007 celdas con una población de al menos 5.000 personas. La celda con mayor población, 54.228 residentes, se encuentra en el área metropolitana de Barcelona.

**Figura 7.** *Grid* de densidad de población, un km<sup>2</sup>. 2010



**Distribución de frecuencias. Celdas por intervalo**

| 1-4   | 5-19   | 20-199 | 200-499 | 500-4.999 | ≥ 5.000 |
|-------|--------|--------|---------|-----------|---------|
| 8,637 | 18,574 | 45,448 | 10,060  | 9,714     | 2,007   |

Fuente: Padrón Municipal: Instituto Nacional de Estadística; SIOSE: Instituto Geográfico Nacional y elaboración propia.

<sup>7</sup> Aunque en principio es posible obtener una *grid* con mayor resolución, utilizamos la de un km<sup>2</sup> de tamaño de celda, porque éste es el estándar europeo (Van Eupen *et al.*, 2012), y de esta forma podemos compararla con otras *grids*.

<sup>8</sup> Siguiendo las recomendaciones de Eurostat (1999), la superficie terrestre excluye láminas de agua de la superficie total. Estas superficies se estiman a partir de SIOSE.

De esta forma la figura 7 ofrece una visión mucho más realista de la distribución de la población que la observada en el mapa de coropletas de densidades municipales, o incluso a nivel de sección censal.

### 3.1. Validación

Saber si nuestras fuentes de información y métodos mejoran los utilizados en el contexto Europeo (Gallego, 2010; Gallego, Batista, Rocha y Mubareka, 2011), y cuyos resultados para España se muestran en la figura 3 requiere *grids* de población construidas por agregación. Gallego (2010) y Gallego, Batista, Rocha y Mubareka (2011) realizan este ejercicio de validación a partir de *bottom-up grids* para diversos países europeos en los que se dispone de información, utilizando el estadístico relativo de discrepancia

$$\delta = \frac{\sum_j |P_j - P_j^{ref}|}{2 \cdot \sum_j P_j} \quad (7)$$

donde  $j$  indexa las celdas y el superíndice *ref* hace referencia a la *grid* de referencia, la construida por agregación, y que se toma como la verdadera distribución de la población. El índice está normalizado en el intervalo [0, 1] y  $100 \times \delta$  puede ser interpretado como el porcentaje de población que es asignado a una celda incorrecta, dada la resolución de la *grid*. Aunque este índice es bastante rudimentario, porque no tiene en cuenta la distancia en el error cometido cuando asignamos población a una celda que no le corresponde, es decir la contribución al error es la misma si situamos a un habitante en la celda de al lado respecto a la que le correspondería o en la otra parte del país, lo utilizamos por motivos de comparación.

La tabla 2 muestra el estadístico relativo de discrepancia reportado por Gallego, Batista, Rocha y Mubareka (2011) para diversos países europeos y recogiendo el

**Tabla 2.** Estadístico relativo de discrepancia,  $100 \times \delta$ , frente a *bottom-up grids* en diversos países europeos. *Grid* JRC - Eurostat - EEA

|                                | <i>Austria</i> | <i>Dinamarca</i> | <i>Finlandia</i> | <i>Suecia</i> | <i>Países Bajos</i> | <i>Irlanda del Norte</i> | <i>Estonia</i> |
|--------------------------------|----------------|------------------|------------------|---------------|---------------------|--------------------------|----------------|
| <i>Población (millones)</i>    | 8,03           | 5,35             | 5,18             | 8,88          | 15,99               | 1,69                     | 1,31           |
| <i>Municipios (coropletas)</i> | 55,79          | 56,82            | 65,54            | 70,27         | 58,54               | 33,14                    | 58,02          |
| <i>Desagregación espacial</i>  | 27,09          | 20,47            | 30,89            | 35,70         | 19,20               | 17,75                    | 35,50          |

*Fuente:* Gallego, Batista, Rocha y Mubareka (2011). Los resultados se ofrecen, en cada caso, para el mejor método de los diferentes ensayados por estos autores.

menor error en cada caso para los diferentes métodos ensayados por estos autores. Claramente la información y los métodos utilizados por estos autores mejoran la distribución de la población, reduciendo el error a aproximadamente la mitad en comparación con la representación estándar de las densidades municipales. Visualmente éstas son las mejoras observadas al pasar de la figura 1 a la figura 3. En cualquier caso, las discrepancias difícilmente bajan del 20%.

En nuestro caso podemos validar el ejercicio frente a la Comunidad de Madrid. La tabla 3 presenta resultados comparables, añadiendo las discrepancias que obtendríamos a partir de un mapa de coropletas para las secciones censales.

**Tabla 3.** Estadístico relativo de discrepancia,  $100 \times \delta$ , frente la *bottom-up grid* de la Comunidad de Madrid

|                                      |       |
|--------------------------------------|-------|
| Población (millones)                 | 6,44  |
| Municipios (coropletas)              | 69,02 |
| Secciones Censales (coropletas)      | 12,03 |
| Desagregación espacial: ecuación (6) | 4,41  |

Fuente: Elaboración propia, INE, IEM, IGN y SIOSE2005.

Para el mapa de densidades municipales, el mapa de coropletas presenta una discrepancia comparable a la observada en el caso de los países europeos mostrados en la tabla 2. En este caso, las discrepancias alcanzan casi el 70%. Las discrepancias para el mapa de coropletas a nivel de sección censal son mucho menores, del orden del 12%. En este caso todavía no hemos utilizado información auxiliar procedente de las coberturas del suelo, y los resultados muestran de forma clara los beneficios de hacer pública información demográfica para áreas pequeñas. El error es menor al que se observa en la tabla 2 para cualquier país utilizando poblaciones municipales e información sobre coberturas del suelo de CLC. La última fila muestra los beneficios derivados de utilizar información sobre coberturas del suelo con elevada resolución. El error se reduce en un orden de magnitud de tres.

Como nota de precaución, vale la pena observar que la Comunidad de Madrid es, probablemente, una región poco representativa: altamente urbanizada, con una muy elevada densidad y concentración de la población. De hecho, un 67% de la población reside en secciones censales puras, de forma que el problema de redistribución afecta sólo a un 33% de la población.

#### 4. Comentarios finales

Una localización precisa de la población es clave para muchas cuestiones prácticas de la organización social. El trabajo ha presentado las tendencias recientes en este campo, desde los esfuerzos de los Institutos Nacionales de Estadística y Eurostat para mejorar la correspondencia entre la medición de la población y su localización, hasta



un ejercicio de desagregación espacial que construye una *grid* de población residente para España con un km<sup>2</sup> de resolución.

Las aplicaciones para este tipo de información son enormes, especialmente en el análisis de las relaciones entre la población y el medio ambiente (Verburg *et al.*, 2010), algunas de ellas objeto actual de investigación. Cuestiones como la accesibilidad de los servicios públicos (Verburg, Overmars y Witte, 2004; Uchida y Nelson, 2009), la definición de áreas urbanas y rurales (Eurostat, 2010, ch. 15; Van Eupen *et al.*, 2012), así como la inter-relación entre ellas (Dijkstra y Poelman, 2008), engarzan de forma natural en este contexto. Sin duda alguna, un mayor número de variables socio-demográficas, tales como edad, sexo o nacionalidad, serían de gran interés.

Incluso, cuando en el futuro, las variables demográficas estén geo-referenciadas y disponibles en formato de *grid*, los métodos presentados en este trabajo pueden ser re-orientados en dos direcciones: i) hacia la reformulación de modelos dinámicos de distribución de la población (Martin, 2010), o ii) hacia la desagregación de información económica en este formato (Sachs, Mellinger y Gallup, 2001; Nordhaus, 2002; 2006; 2008).

## Bibliografía

- Batista e Silva, F.; Lavalle, C., y Koomen, E. (2012): «A procedure to obtain a refined European land use/covermap», *Journal of Land Use Science*, en prensa. Disponible en línea, 29 de marzo de 2012 (doi:10.1080/1747423X.2012.667450).
- Capdevila i Subirana, J. (2009): *Historia del deslinde de la frontera hispano-francesa. Del tratado de los Pirineos (1659) a los tratados de Bayona (1856-1868)*, Centro Nacional de Información Geográfica, Instituto Geográfico Nacional, Ministerio de Fomento, Madrid. Disponible en internet: <http://www.ign.es/ign/layout/ libdigitalespublicaciones.do>.
- Chen, K.; McAneney, J.; Blong, R.; Leigh, R.; Hunter, L., y Magill, C. (2004): «Defining area at risk and its effect in catastrophe loss estimation: A dasymetric mapping approach», *Applied Geography*, 24, núm. 2 (abril), 97-117.
- De Cos Guerra, O., y Reques Velasco, P. (2005): «Los cambios en los patrones territoriales de la población española (1900-2001)», *Papeles de Economía Española*, 104, 167-192.
- Dijkstra, L., y Poelman, H. (2008): «Remote rural regions: How proximity to a city influences the performance of rural regions», *Regional Focus*, 1/2008, DG-Regio, Bruselas.
- Eicher, C., y Brewer, C. (2001): «Dasymetric mapping and areal interpolation: Implementation and evaluation», *Cartography and Geographic Information Science*, 28, 125-138.
- Eurostat (1999): *Recommendations for a harmonized definition of calculation of surface area of territorial units. Methods and Nomenclature. Theme 1*, European Commission, Luxembourg.
- (2010): *Eurostat regional yearbook 2010*, Publication Office of the European Union, Eurostat, Luxemburg.
- Gallego, F. J. (2010): «A population density grid of the European Union», *Population & Environment*, 31, 6, julio, 460-473. Disponible en internet: <http://www.springerlink.com/content/0199-0039/31/6/>.
- Gallego, F. J., y Peedell, S. (2001): «Using land cover to map population density», en *Towards agri-environmental indicators. Integrating statistical and administrative data*

- with land cover information, Topic report núm. 6 (pp. 94-105), Eurostat, DG Agriculture, DG Environment, Joint Research Center, European Environment Agency, Luxemburgo.
- Gallego, F. J., Batista, F.; Rocha, C., y Mubareka, S. (2011): «Disaggregating population density of the European Union with Corine land cover», *International Journal of Geographical Information Science*, 25, 12, diciembre, 2051-2069. Disponible en internet: <http://dx.doi.org/10.1080/13658816.2011.583653>.
- Goerlich, F. J., y Mas, M. (2008a): «Algunas pautas de localización de la población española a lo largo del siglo xx», *Investigaciones Regionales*, 12, primavera, 5-33.
- (2008b): «Empirical evidence of population concentration in Spain», *Population-E*, 63, 4 635-650.
- (2009): «Drivers of agglomeration: Geography versus History», *The Open Urban Studies Journal (TOUSJ)*, 2, 28-42. Disponible en internet: <http://www.bentham.org/open/tousj/openaccess2.htm>.
- Goerlich, F. J.; Mas, M.; Azagra, J., y Chorén, P. (2006): *La localización de la población española sobre el territorio. Un siglo de cambios. Un estudio basado en series homogéneas (1900-2001)*, Fundación BBVA, Bilbao.
- Gutiérrez Puebla, J.; Gómez Cerdá, G., y García Palomares, J. C. (2009): «Estudio de accesibilidad del PEIT (Plan Estratégico de Infraestructuras y Transportes del Ministerio de Fomento)», Departamento de Geografía Humana, Universidad Complutense de Madrid, Madrid.
- IGN (Instituto Geográfico Nacional) (2002): *Corine 2000. Descripción de la Nomenclatura del Corine Land Cover al nivel 5.º*, diciembre, Madrid.
- (2007a): *Necesidad de un nuevo «Datum»*. Versión 1.0, octubre, Grupo de trabajo para la transición a ETRS89, Consejo Superior Geográfico, Madrid.
- (2007b): *Análisis de la problemática del cambio de datum geodésico a ETRS89*. Versión 1.0, octubre, Grupo de trabajo para la transición a ETRS89, Consejo Superior Geográfico, Madrid.
- (2007c): *Análisis de los sistemas de cambio*. Versión 1.0, octubre, Grupo de trabajo para la transición a ETRS89, Consejo Superior Geográfico, Madrid.
- (2007d): *Herramientas para facilitar el cambio*. Versión 1.0, octubre, Grupo de trabajo para la transición a ETRS89, Consejo Superior Geográfico, Madrid.
- (2007e): *El sistema ETRS89 y la nueva cartografía (Relación entre las series cartográficas y «mapas continuos»)*. Versión 1.0, octubre, Grupo de trabajo para la transición a ETRS89, Consejo Superior Geográfico, Madrid.
- (2010a): *Metodología de Producción de la Base de Datos CLC-Change 2000-2006*, enero, Madrid. Disponible en internet: <http://centrodedescargas.cnig.es/CentroDescargas/>.
- (2010b): *BTN25 Base Topográfica Nacional de España 1:25.000. Anexo C Diccionario de Datos y Normas de Captura BTN25*. Versión 2.1, enero, Subdirección General de Producción Cartográfica, Madrid. Disponible en internet: <http://centrodedescargas.cnig.es/centrodedescargas/>.
- (2010c): Fichero vectorial de CORINE LandCover 2006 del Centro de Descargas del Centro Nacional de Información Geográfica (CNIG) (consulta: 13 de septiembre de 2010).
- (2011a): *Sistema de Información de Ocupación del Suelo en España —SIOSE2005—*. Documento Resumen, mayo, Madrid.
- (2011b): *Modelo de Datos de la Base Cartográfica Nacional 1:200.000 BCN200*. Versión 4.0, enero, Madrid. Disponible en internet: <http://centrodedescargas.cnig.es/centrodedescargas/>.
- (2012): *Fondos cartográficos del Instituto Geográfico Nacional. España siglos XVI y XIX*, Madrid. Disponible en internet: <http://www2.ign.es/fondoscartograficos/> (consulta: marzo del 2012).

- INE (Instituto Nacional de Estadística) (2011): *Proyecto de los Censos Demográficos 2011*, febrero, Subdirección General de Estadísticas de la Población. Madrid. Disponible en internet: [http://www.ine.es/censos2011/censos2011\\_proyecto.pdf](http://www.ine.es/censos2011/censos2011_proyecto.pdf).
- INSPIRE (2010a): *D2.8.I.1 INSPIRE Specification on Coordinate Reference Systems - Guidelines*. Version 3.1, 26 de abril de 2010, INSPIRE Thematic Working Group Coordinate Reference Systems and Geographical Grid Systems, Bruselas. Disponible en internet: <http://inspire.jrc.ec.europa.eu/index.cfm/pageid/2>.
- (2010b): *D2.8.I.2 INSPIRE Specification on Geographical Grid Systems - Guidelines*. Version 3.0.1, 26 de abril de 2010, INSPIRE Thematic Working Group Coordinate Reference Systems and Geographical Grid Systems, Bruselas. Disponible en internet: <http://inspire.jrc.ec.europa.eu/index.cfm/pageid/2>.
- Jablonski, R. (2011): «Lessons learned from disaggregating population data by using different data sources». Ponencia presentada en el *European Forum for Geostatistics 2011*, Lisboa, 12-14 de octubre. Disponible en internet: <http://www.efgs.info/workshops/efgs-2011-lisboa-portugal>.
- Kopecky, M., y Kahabka, H. (2009): *2006 GMES Fast track service precursor on land monitoring. Updated delivery report*. European Mosaic, European Environment Agency, Copenhagen. Disponible en internet: <http://www.eea.europa.eu/data-and-maps/data/eea-fast-track-service-precursor-on-land-monitoring-degree-of-soil-sealing-100m-1>.
- Martin, D. (2010): «Progress report: 24-hour gridded population models». Ponencia presentada en el *European Forum for Geostatistics 2010*, Tallin (Estonia), 5-7 de octubre. Disponible en internet: <http://www.efgs.info/workshops/efgs-2010-tallinn-estonia>.
- Martin, D.; Tate, N. J., y Langford, M. (2000): «Refining population surface models: Experiments with Northern Ireland Census data», *Transactions in GIS*, 3, 285-301.
- Mennis, J. (2003): «Generating surface models of population using dasymetric mapping», *Professional Geographer*, 55, 1, 31-42.
- (2009): «Dasymetric mapping for estimating population in small areas», *Geography Compass*, 3, 2, 727-745.
- Muguruza Cañas, C., y Santos Preciado, J. M. (1989): «La importancia de las unidades de análisis en el modelo de ecología factorial», *Espacio, tiempo y forma* (Revista de la Facultad de Geografía e Historia de la UNED), 2, 87-101.
- Nordhaus, W. D. (2002): *Alternative approaches to spatial rescaling*. Version 2.2.2. Mimeo, Yale University, New Haven (Connecticut), 28 de febrero. Disponible en internet: <http://gecon.yale.edu/research-papers>.
- (2006): «Geography and macroeconomics: New data and new findings». Ponencia en *National Academy of Sciences of the United States of America*, 103, 10, marzo, 3510-3517. Disponible en internet: <http://gecon.yale.edu/research-papers>.
- (2008): «New metrics for environmental economics: Gridded economic data», *The Integrated Assessment Journal, Bridging Sciences & Policy*, 8, 1, 73-84.
- Reher, D. S. (1994): «Ciudades, procesos de urbanización y sistemas urbanos en la Península Ibérica, 1550-1991», en M. Guardia, F. J. Monclús y J. Oyón (dirs.), *Atlas histórico de ciudades europeas*, Centre de Cultura Contemporània de Barcelona y Salvat, Barcelona, 1-29.
- (2006): «Perfiles regionales de la población española en tiempos históricos», en J. A. Fernández Cordón y J. Leal Maldonado (eds.), *Análisis territorial de la demografía española 2006*, Fundación Fernando Abril Martorell, Madrid, julio, 53-83.
- Reques Velasco, P., y Rodríguez Rodríguez, V. (1998): *Atlas de la Población Española. Análisis de base municipal*, Servicio de Publicaciones de la Universidad de Cantabria, CSIC y ESRI-España, Santander.
- Rodríguez Pascual, A. F.; Martín-Asín López, G., y Astudillo Muñoz, B. (1997): «La Base Cartográfica Numérica 1:25.000 (BCN25)», *Mapping* (Revista Internacional de Ciencias de la

- Tierra), abril-mayo. Disponible en internet: [http://www.mappinginteractivo.com/plantilla-ante.asp?Id\\_articulo=782](http://www.mappinginteractivo.com/plantilla-ante.asp?Id_articulo=782).
- Sachs, J. D.; Mellinger, A. D., y Gallup, J. L. (2001): «The geography of poverty and wealth», *Scientific American*, 284, marzo, 70-75.
- SIOSE (2010): *Land Cover and Use Information System (SIOSE). Technical Document Version 2.0*, National Geographic Institute (IGN), Madrid, 29 de enero. Disponible en internet: <http://inspire.jrc.ec.europa.eu/index.cfm/pageid/241/referenceid/32883>.
- Smith, D. A. (2011): «Polycentricity and sustainable urban form. An intra/urban study of accessibility, employment and travel sustainability for the strategic planning of the London Region». Thesis, Center for Advanced Spatial Analysis & Department of Geography, University College London, London, agosto. Disponible en internet: [http://casa.ucl.ac.uk/duncan/dasmith\\_phdthesis\\_2011.pdf](http://casa.ucl.ac.uk/duncan/dasmith_phdthesis_2011.pdf).
- Tobler, W. R. (1979): «Smooth pycnophylactic interpolation for geographical regions», *Journal of the American Statistical Association*, 74, 367, septiembre, 519-530.
- Tralli, D. M.; Blom, R. G.; Zlotniki, V.; Donnellan, A., y Evans, D. L. (2005): «Satellite remote sensing of earthquake, volcano, flood, landslide and coastal inundation hazards», *ISPRS Journal of Photogrammetry and Remote Sensing*, 59, 4, 185-198.
- Uchida, H., y Nelson, A. (2009): «Agglomeration index: Towards a new measure of urban concentration». Background paper para World Bank's World Development Report 2009: *Reshaping Economic Geography*, 16. Disponible en internet: <http://www.worldbank.org/>.
- Van Eupen, M.; Metzger, M. J.; Pérez-Soba, M.; Verburg, P. H.; van Doorn, A., y Bunce, R. G. H. (2012): «A rural typology for strategic European policies», *Land Use Policy*, 29, 3, julio, 473-482.
- Verburg, P. H.; Overmars, K. P., y Witte, N. (2004): «Accessibility and land use patterns at the forest fringe in the Northeastern part of the Philippines», *The Geographical Journal*, 170, 3, septiembre, 238-255.
- Verburg, P. H.; van Berkel, D. B.; van Doorn, A. M.; van Eupen, M., y van den Heiligenberg, H. A. R. M. (2010): «Trajectories of land use change in Europe: a model-based exploration of rural futures», *Landscape Ecology*, 25, 2, febrero, 217-232.
- Vidal, C.; Gallego, J., y Kayadjanian, M. (2001): «Geographical use of statistical data. Methodological overview», en *Towards agri-environmental indicators. Integrating statistical and administrative data with land cover information*, European Environment Agency, Copenhagen, topic report núm. 6, capítulo 1, 11-24.
- Villa, G. (2009): *CLC-OO. Data model. Application and examples*, 2<sup>nd</sup> EIONET OODM Working Group Meeting, National Geographic Institute (IGN), Madrid, 10-11 diciembre. Disponible en internet: <http://inspire.jrc.ec.europa.eu/index.cfm/pageid/241/referenceid/32916>.
- Villa, G.; Valcárcel, N.; Arozarena, A.; García-Asensio, L.; Caballero, M. E.; Porcuna, A.; Domenech, E., y Peces, J. J. (2008): «Land cover classifications: An obsolete paradigm», *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XXXVII, parte B4, 609-614, Beijing.
- Yuan, Y.; Smith, R. M., y Limp, W. F. (1997): «Remodeling census population with spatial information from landsat TM imagery», *Computers, Environment and Urban Systems*, 21, 3/4, mayo-junio, 245-258.
- Zoido Naranjo, F., y Arroyo Pérez, A. (2004): «La población de España», en A. Arroyo Pérez (coord.), *Tendencias demográficas durante el siglo xx en España*, Universidad de Sevilla e Instituto Nacional de Estadística, Madrid, 1 de febrero, 17-75. Disponible en internet: [http://www.ine.es/prodyser/pubweb/tend\\_demo\\_s20/tend\\_demo\\_s20.htm](http://www.ine.es/prodyser/pubweb/tend_demo_s20/tend_demo_s20.htm).