

5 Documento de Trabajo Ivie

WP-Ivie 2024-05

GRID STATISTICS -ESTADÍSTICAS EN FORMATO GRID-

Francisco Goerlich Gisbert

Los documentos de trabajo del Ivie ofrecen un avance de los resultados de las investigaciones económicas en curso o análisis específicos sobre debates de actualidad, con objeto de divulgar el conocimiento generado por diferentes investigadores.

Ivie working papers offer a preview of the results of economic research under way, as well as an analysis on current debate topics, with the aim of disseminating the knowledge generated by different researchers.

La edición y difusión de los documentos de trabajo del Ivie es una actividad subvencionada por la Generalitat Valenciana, Conselleria de Hacienda y Modelo Económico, en el marco del convenio de colaboración para la promoción y consolidación de las actividades de investigación económica básica y aplicada del Ivie.

The editing and dissemination process of Ivie working papers is funded by the Valencian Regional Government's Ministry for Finance and the Economic Model, through the cooperation agreement signed between both institutions to promote and consolidate the Ivie's basic and applied economic research activities.

Todos los documentos de trabajo están disponibles de forma gratuita en la web del Ivie <http://www.ivie.es>. Al publicar este documento de trabajo, el Ivie no asume responsabilidad sobre su contenido.

Working papers can be downloaded free of charge from the Ivie website <http://www.ivie.es>. Ivie's decision to publish this working paper does not imply any responsibility for its content.

Cómo citar/How to cite:

Goerlich Gisbert, F. « Grid Statistics –Estadísticas en formato grid–». Working Papers Ivie n.º 2024-5. València: Ivie.
http://doi.org/10.12842/WPIVIE_0524

Versión: Julio 2024 / Version: July 2024

Edita / Published by:

Instituto Valenciano de Investigaciones Económicas, S.A.

C/ Guardia Civil, 22 esc. 2 1º - 46020 València (Spain)

DOI: http://doi.org/10.12842/WPIVIE_0524

WP-Ivie 2024-5

Grid Statistics

–Estadísticas en formato grid–

Francisco Goerlich¹

Resumen

Con la popularización de las *grids* como unidad básica de distribución de estadísticas territoriales, impulsada por Eurostat, se hace necesario la disponibilidad de estadísticas en este formato.

Este trabajo representa un primer intento en esta dirección para facilitar la manipulación de estadísticas en este sistema de distribución y, al mismo tiempo poder agregar los resultados por entidades administrativas: municipios, provincias o comunidad autónoma.

Se trata de un trabajo y una base de datos abierta que irá creciendo en el futuro conforme nuevas variables se vayan incorporando a este formato. La versión más reciente del trabajo puede encontrarse en <https://go.uv.es/goerlich/GridStatistics> y las variables descritas en el mismo pueden descargarse de <https://zenodo.org/records/12954007>.

Palabras clave: Estadísticas en formato *grid*, sistemas zonales.

Clasificación JEL: C80

Abstract

With the popularization of grids as the basic unit of distribution of territorial statistics, promoted by Eurostat, the availability of statistics in this format has become necessary.

This work represents a first attempt in this direction to facilitate the manipulation of statistics in this distribution system and, at the same time, to be able to aggregate the results by administrative entities: municipalities, provinces or autonomous communities.

It is an open work and an open database that will grow in the future as new variables are incorporated into this format. The most recent version of the work can be found at <https://go.uv.es/goerlich/GridStatistics> and the variables described in it can be downloaded from <https://zenodo.org/records/12954007>.

Keywords: Grid statistics, zonal systems.

JEL classification: C80

¹ Francisco Goerlich, Universidad de Valencia e Ivie.

1.

Introducción

Este trabajo es una base de datos abierta², con la popularización de las *grids*, impulsada por Eurostat, se hace necesario la disponibilidad de estadísticas en este formato. Muchas de ellas de carácter meramente geográfico, resultado de la combinación con otras capas geográficas y derivadas mediante la aplicación de operaciones de los Sistemas de Información Geográfica (GIS). Las razones son múltiples y sencillas de entender.

Por una parte, esta información tiene utilidad en sí misma –al fin y al cabo, las *grids* son un sistema zonal al igual que los municipios o las provincias, con la particularidad de que sus unidades elementales, las celdas o píxeles, son homogéneas en tamaño e invariantes en el tiempo–.

Por otra parte, las *grids* se utilizan muchas veces como geografía intermedia para transferir información de unos sistemas zonales a otros, y esta información auxiliar es esencial para este propósito (Goerlich 2016).

Finalmente, muchas estadísticas son recogidas a partir de unidades administrativas en las que se divide el territorio –Secciones Censales (SSCC) o Comunidades Autónomas (CCAA)–, pero se desea su desagregación o transferencia a un formato de *grid*. Esta información vuelve a ser esencial para este tipo de tarea (Goerlich y Cantarino 2012, 2013a, 2015a)

El sistema de *grids* europeo está normalizado por INSPIRE (2014a), por lo que las

estadísticas generadas lo serán en este formato normalizado. En particular, la *grid* con 1km de resolución –celdas de 1km × 1km– se ha convertido en el estándar para distribución de cierta información a nivel europeo –las *grids* de población de GEOSTAT se distribuyen con esta resolución³–. Así pues, la información generada lo será a esta resolución, a menos que se indique lo contrario. La agregación a una menor resolución –5km o 10km– es, normalmente, inmediata. Sin embargo, el proceso contrario, la desagregación a una mayor resolución –250m o 100m– no lo es en absoluto.

La estructura del trabajo es la siguiente. A continuación, exponemos la generación de las *grids*, sus características y el contorno que cubre. Una vez disponemos de la *grid* de referencia los siguientes apartados describen, por bloques, la información generada indicando, brevemente, el proceso de elaboración y las fuentes primarias de datos utilizadas.

A menos que se indique lo contrario los procesos fueron implementados en software libre basado en el sistema de cálculo estadístico R (R Core Team 2022), utilizando las librerías de *tidyverse* (Wickham et al 2019) para data *wrangling*, la librería *sf* (Pebesma 2018) para el manejo de información vectorial y la librería *terra* (Hijmans 2022) para el manejo de información *raster*.

La información de la base de datos está disponible si se solicita al autor.

² La versión más reciente de este documento puede ser encontrada en <https://go.uv.es/goerlich/GridStatistics>. La base de

datos a la que hace referencia este documento puede descargarse de <https://zenodo.org/records/12954007>

³ El censo 2021 distribuirá información demográfica, más allá de los totales de población, con esta resolución.

2.

Grids generadas y contorno utilizado

La Comunidad Valenciana sabemos lo que es. Tenemos en mente su localización geográfica, su forma aproximada y algunas de sus características geográficas básicas –por ejemplo, sabemos que linda con el mar mediterráneo ¡pero que no es una isla! y que no tiene fronteras con otros países–. Disponemos, por tanto, de una imagen visual de la misma, aunque para trabajar con ella desde el punto de vista geográfico en un *GIS* necesitaremos resolver algunas cuestiones técnicas referidas al Sistema de Referencia de Coordenadas (*CRS*).

Esta intuición visual desaparece cuando hablamos de *grids*. La primera pregunta que surge a la hora de trabajar con una *grid* es: **¿de que contorno estamos hablando?** Necesitamos un espacio de referencia en el que situar la *grid*. Luego vendrán las mismas cuestiones técnicas que son necesarias para manipular la información geográfica en un *SIG*, pero primero necesitamos saber de qué estamos hablando.

Aunque el GISCO de Eurostat ofrece las *grids* normalizadas que cubren todo el territorio Europeo con diferentes resoluciones, y de donde podíamos haber extraído las celdas correspondientes a nuestro país⁴, nosotros preferimos generar las *grids* para el contorno de España de acuerdo con la información de la **Base de Datos de Líneas Límite** de Instituto Geográfico Nacional (IGN). **Nuestro contorno de referencia es toda España**, hasta la resolución y el detalle ofrecido por el IGN. Esta información se describe en el

epígrafe 2.2. El proceso de generación de las *grids* y los detalles técnicos que se mencionan al principio se ofrecen a continuación.

2.1 Grids: Características básicas

Dado el contorno de referencia, las *grid* se generaron con *GridMaker*, versión 1.3. Una utilidad *opensource* producida por Eurostat (2021) para tal fin. **Se generaron sin buffer**, lo que implica que **todas** las celdas de las *grids* generadas tienen intersección no nula con alguna parte del territorio, por pequeña que sea, y no hay ninguna parte del territorio que no sea cubierta por la *grid* generada. *GridMaker* genera las *grids* en vectorial.

Siguiendo las recomendaciones de INSPIRE (2014a), todas las *grid* se generaron con *CRS* ETRS89-LAEA⁵. ETRS89 es el Sistema Geodésico de referencia Europeo y LAEA –*Lambert Azimutal Equal Area*– es una proyección que mantiene áreas en todas las regiones de la esfera, y es la recomendada por INSPIRE (2014b) para cartografía y análisis espacial.

El **centro de la proyección** ETRS89-LAEA se fija en la coordenada 52°N y 10°E y **falso origen** en $x_0 = x_0 = 4321000\text{m}$ –falso este– y $y_0 = y_0 = 3210000\text{m}$ –falso norte–. El origen de la *grid*, $x = x = 0$ y $y = y = 0$, coincide con el falso origen del *CRS* ETRS89-LAEA, por lo que esta *grid* se la conoce como **Grid_ETRS89-LAEA5210**, y en la que, para la identificación de la resolución, se añade al final el tamaño de la celda en metros –_10K, para identificar una resolución de 10km =

⁴ El Instituto Nacional de Estadística (INE) también ofrece dicha *grid*, con resolución de 1km, en la información del Censo 2011, aunque curiosamente dicha *grid* no cubre el total de la

población del censo, sino que solo incluye las celdas que contienen al menos una vivienda –principal o no principal–

⁵ Código EPSG:3035

10,000m–. La *grid* se define de forma jerárquica en coordenadas métricas y potencias de 10. De esta forma la resolución de las celdas es 1m, 10m, 100m, 1000m, 10k, 100k y 1000k. Existe, por tanto, la posibilidad de que algunas *grids* basadas en ETRS89-LAEA no sean acordes con la **Grid_ETRS89-LAEA5210** –por ejemplo, si tienen una resolución de 5km–. A efectos de maximizar la compatibilidad, el origen de todas las *grids* basadas en ETRS89-LAEA coincidirá con el origen de la **Grid_ETRS89-LAEA5210**. De esta forma los puntos de la *grid* basados en ETRS89-LAEA coincidirán con puntos de la **Grid_ETRS89-LAEA5210**. La orientación de la *grid* es sur-norte y oeste-este.

INSPIRE (2014a) establece una **identificación inequívoca** para las **celdas** de la *grid* basada en el **tamaño** de la celda, en metros, y las **coordenadas** de un **punto de referencia**. Como punto de referencia de cada celda de toma la coordenada de la **esquina inferior izquierda** de la celda.

Aunque la *grid* de referencia, y aquella para la que se producen estadísticas, es la de 1km de resolución –celdas de 1km × 1km– se elaboraron muchas otras. Todas ellas se generaron en formato poligonal y puntual, y se almacenaron en ficheros independientes en formato *open source Geopackage*, *.gpkg, con arreglo a la nomenclatura:

- Spain2019_grid_<res>km_<geom>_ETRS89_LAEA

Donde <res> es la resolución, que como se indica a continuación es en kilómetros, y <geom> es la geometría –point para geometría puntual y surf para geometría poligonal–. En total 20 *grids*.

Se generaron *grids* con las siguientes resoluciones:

Resolución (km)	Celdas
0.1	50,669,552
0.2	12,682,116
0.5	2,035,567
1	511,294
2	128,950
5	21,126
10	5,474
20	1,467
50	283
100	92
200	34
500	10
1000	6

Fuente:

GridMaker genera la *grid* con 3 campos básicos:

- **GRD_ID: Identificador único de celda** con formato CRS3035RES<res>mN<Y_LLC>E<X_LLC>, donde <res> es la resolución en metros.
- **X_LLC:** Coordenada X –longitud– de la esquina inferior izquierda en metros.
- **Y_LLC:** Coordeanda Y –latitud– de la esquina inferior izquierda en metros.

Naturalmente las coordenadas son ETRS89-LAEA –métricas–.

Los campos son idénticos en las *grids* poligonales que en las puntuales, de forma que las coordenadas en la *grid* puntual hacen referencia a la identificación de la celda, mientras que la geometría es la del centroide de la celda.⁶

Nota técnica sobre la generación de las *grids* Por debajo de los 500m de resolución *GridMaker* fue incapaz de generar la *grid* por problemas de memoria. El resultado pasó por construir una función específica en *R* que generara la *grid*. Dicha función incorpora la posibilidad de teselación, lo que reduce enormemente la memoria utilizada, además de aumentar notablemente la velocidad de cómputo, y paralelización.

La *grid* de 200m de resolución, con algo más de 12 millones de celdas, fue obtenida en un

ordenador personal con teselación y paralelización –14 *cores*–.

La *grid* de 100m de resolución, con algo más de 50 millones de celdas, requirió de computación paralela –36 *cores*– en un *cluster* de computación científica de la Universidad de Valencia –Lluis Vives–, y la versión polygonal ocupó 13Gb.

2.2 Contornos administrativos

El contorno de España, utilizado para generar las *grids* descritas en el epígrafe anterior, procede de la Base de Datos de Líneas Límite disponible en el Centro de Descargas del Centro Nacional de Información Geográfica (CNIG), dependiente del IGN.⁷ Según información en la *web* de descarga la “...*geometría responde a la interpretación de los títulos jurídicos inscritos en el Registro Central de Cartografía...*”. Y de acuerdo con los metadatos, en el caso general –un 90% aproximadamente en 2017–, las líneas límite tienen una incertidumbre geométrica correspondiente a un rango de escalas entre 1:25,000 y 1:100,000. Esta información vectorial se distribuye en formato *shape* y coordenadas geográficas *ETRS89* para la península y balears y *WGS84* para Canarias, en dos capas diferentes. Después de transformar la capa para Canarias a *ETRS89* ⁸se fusionaron ambas capas en una sola. Dicha información se re-proyectó a LAEA, que es nuestro *CRS* de referencia.

La capa está ajustada a los municipios a 1 de enero de 2019 –desde esa fecha no se han producido alteraciones municipales⁹ por lo que a 1 de enero de 2022 disponemos de los

⁶ Existen algunas diferencias entre el identificador de celda generado por *GridMaker* y el propuesto por INSPIRE (2014a), si bien en ambos casos es posible identificar la resolución de la celda y las coordenadas de la esquina inferior izquierda.

⁷ La información se descargó el 04/02/2020.

⁸ Aunque en la práctica ambos sistemas de referencia son equivalentes a nuestros efectos.↵

⁹ Lo que no significa que no haya habido alteraciones de linderos.

mismos municipios–, y tiene un registro por municipio –no por polígono–. En total disponemos de 8,212 registros –8,131 corresponden a municipios y los 83 restantes a territorios mancomunados–. La codificación en la descarga se ajusta a la normativa INSPIRE para unidades administrativas (INSPIRE 2014c), por lo que de esta se extrajo la codificación municipal de 5 dígitos, añadiéndose los campos correspondientes a los códigos y nombres de las provincias y Comunidad Autónoma (CCAA). Esta capa forma parte de la librería de Goerlich y Pérez (2021).

La capa de contornos administrativos incluye recintos para el Peñón de Vélez de la Gomera, el Peñón de Alhucemas, las Islas Alhucemas y las Islas Chafarinas, posesiones españolas en el norte de África. Dicha capa atribuye estos polígonos a la Ciudad Autónoma de Melilla, aunque versiones más recientes de la capa de contornos administrativos incluye 5 recintos que el IGN no atribuye a ninguna Comunidad Autónoma (CCAA), con código de CCAA 20 y código de provincia 54¹⁰. Además de los 4 recintos mencionados, también se incluye, entre estos territorios no asignados a ninguna CCAA,

la Isla de Perejil, que en nuestra capa de contornos está asignada a la Ciudad Autónoma de Ceuta. Finalmente, la isla de Alborán está asignada al municipio de Almería, tanto en nuestra capa como en las más recientes. Desde el punto de vista de la generación de la *grid* para el conjunto nacional esta atribución es irrelevante, puesto que solo el contorno exterior de España importa, pero para hacer una asignación de celdas a unidades administrativas –municipios, provincias y CCAA– no lo es.

El contorno de España se obtuvo por disolución, lo que generó un una capa multipolígono con registro único en la tabla de atributos y una superficie –calculada en ETRS89-LAEA– de 506,027.25 km² (Goerlich 2022). Esta capa es la que se envió a *GridMaker* para la generación de las *grids* que acabamos de describir en el epígrafe anterior.

También se obtuvieron por disolución capas a nivel de provincia y de CCAA. Utilizadas en la identificación de las celdas con las correspondientes unidades administrativas.

¹⁰ La razón para utilizar el código de provincia 54 y no el 53 es que, aunque los territorios mancomunados sí están asignados a la provincia a la que pertenecen –a pesar de que algunos de ellos están administrados por municipios limítrofes de provincias diferentes– lógicamente no disponen de código municipal y el IGN les atribuye un código de 5 dígitos en el que los 2

primeros se corresponden con el 53, y los 3 restantes con un número correlativo. Así pues, el código de provincia 53 no existe, ¡aunque sí el 54!

3.

Características básicas de las celdas

Una vez generada la *grid* se derivó la siguiente información básica –para la *grid* de 1km²–:

- **Observaciones:** Observaciones de las celdas particulares. Normalmente este es un campo vacío, que simplemente identifica ciertas características que pueden ser de interés. Por el momento simplemente identifica las celdas correspondientes al Islote Perejil, Peñón Velez de la Gomera, Peñón de Alhucemas, Islas Chafarinas, Isla de Alborán, Islas Columbretes, Illot de Benidorm y Roque del Este. Además de una celda perteneciente a Canarias, CRS3035RES1000mN1001000E1965000, que parece situarse en el mar según identificación visual con las ortofotos del Plan Nacional de Ortofotografía Aérea (PNOA).

Identificación de las celdas que perteneces a islas:

- **CodIsla:** Código de Isla de acuerdo con la [codificación del Instituto Nacional de Estadística \(INE\)](#).
- **Isla:** Nombre de las Isla.
- **AREA:** Area, en km², de la celda una vez la *grid* ha sido recortada por el contorno administrativo. El área es 1km² para las celdas interiores y menor de 1km² para las celdas exteriores. Aunque la unidad son km² el valor puede ser interpretado en términos porcentuales.

Nota: Existen 2 celdas con AREA inferior al m²: CRS3035RES1000mN1881000E2814000

–0.220m², por lo que redondeada al m² daría una superficie de 0m²– y

CRS3035RES1000mN1438000E3181000 – 0.763m²–. Y otras 2 celdas con AREA ligeramente superior al m²:

CRS3035RES1000mN1891000E3737000 – 1.07m²– y

CRS3035RES1000mN2185000E3663000 – 1.18–. Por esta razón, no se efectuó redondeo alguno sobre los cálculos almacenados.

- **TIPO:** Tipo de celda Interior/Exterior. Existen 10,056 celdas exteriores y 501,238 celdas interiores.
- **CodCCAA:** Código de CCAA con el que intersecta la celda: 01 a 19.
- **CCAA:** Nombre de CCAA con el que intersecta la celda.
- **nCCAA:** Número de intersecciones a nivel de CCAA.
- **CodProv:** Código de Provincia con el que intersecta la celda: 01 a 52.
- **Provincia:** Nombre de Provincia con el que intersecta la celda.
- **nProv:** Número de intersecciones a nivel de Provincia.
- **CodMuni:** Código de Municipio con el que intersecta la celda: Código [INE](#) de 5 dígitos y código IGN de 5 dígitos para los condominios –cuyos 2 primeros dígitos son 53–.
- **Municipio:** Nombre de Municipio –o condominio– con el que intersecta la celda.
- **nMuni:** Número de intersecciones a nivel de Municipio.

El propósito de las variables relacionadas con la división administrativa – CCAA, Provincias y Municipios– es facilitar la transferencia de información entre la *grid* y dichas divisiones –o viceversa–.

Las celdas borde entre 2 o más unidades administrativas contienen los códigos de todas aquellas unidades con las que tienen intersección no nula, separados por |. Por tanto, para extraer las celdas de la provincia 15 –por ejemplo– debemos filtrar **CodProv** por los valores que contengan "15", **no** por los valores que sean iguales a "15" –filter [grid, stringr: str_detect (CodProv, "15")], **no** filter (grid, CodProv == "15"), esta última expresión solo extraerá las celdas que son interiores a la provincia 15–.

El número de intersecciones es útil para distinguir entre las celdas interiores a un contorno dado y las que están entre dos o más unidades administrativas. La Tabla 1 muestra el número de celdas según el número de intersecciones a nivel de CCAA, Provincia y Municipio. Aunque a nivel de CCAA y Provincia solo encontramos celdas que están repartidas entre 3 unidades administrativas – 27 celdas en el caso de CCAA y 110 celdas en el caso de Provincias–, a nivel de Municipios encontramos 2 celdas que están repartidas entre 7 municipios, en ambos casos en la provincia de Valencia, y 9 celdas se encuentran repartidas entre entre 6 municipios.

Cuadro 1.

Intersecciones	Comunidad autónoma	Provincias	Municipios
1	503,407	494,514	348,401
2	7,860	16,670	139,297
3	27	110	21,711
4			1,769
5			105
6			9
7			2
Celdas	511,294	511,294	511,294

Fuente: CNIG del IGN y elaboración propia

- **DIST_COSTA:** Distancia a la costa más cercana en metros, m. Para las celdas exteriores que lindan con el mar, dicha distancia es 0.
- **COSTA:** Costa más cercana. Valores: Mar Cantábrico, Mar Mediterráneo y Océano Atlántico.
- **DIST_FRONTERA:** Distancia a la frontera más cercana en metros, m. Para las celdas frontera, dicha distancia es 0.
- **FRONTERA:** Frontera más cercana. Valores: Andorra, Francia, Gibraltar, Marruecos y Portugal.
- **DIST_EXTERIOR:** Distancia al exterior –costa o frontera– más cercana en metros, m. Para las celdas TIPO == "Exterior", dicha distancia es 0.
- **EXTERIOR:** Costa o frontera más cercana. Valores: la unión de COSTA y FRONTERA.

DIST_EXTERIOR es simplemente la distancia mínima de **DIST_COSTA** y **DIST_FRONTERA**, y **EXTERIOR** es el valor correspondiente de **COSTA** o **FRONTERA**. Existen algunos pocos casos en los que la celda intersecta con la costa y con una frontera, en estos casos todas las distancias son 0, y en **EXTERIOR** puede aparecer la etiqueta de la costa o de la frontera, según lo que el algoritmo encuentre antes. Estos casos son fácilmente identificables porque **DIST_COSTA == DIST_FRONTERA == 0**.

Naturalmente, todas las **distancias** son en **línea recta** y calculadas en la proyección de la *grid*, [ETRS89-LAEA](#).

4

Altitud media, superficie 3D y rugosidad

La información para estas variables procede del **Modelo Digital del Terreno con paso de malla de 5m, MDT05**, disponible en el Centro de Descargas del Centro Nacional de Información Geográfica (CNIG), dependiente del IGN.

El análisis de dicha información de partida se ofrece en Goerlich (2022a), y una aplicación a los municipios en Goerlich y Cantarino (2022). Las variables ofrecidas son las mismas que las de Goerlich y Cantarino (2022) a nivel de municipio, pero ahora nivel de celda de la *grid* de 1 km². Los estadísticos se corresponden a las celdas una vez han sido recortadas por el contorno administrativo, de forma que tienen como soporte el **AREA** del apartado anterior. En este contexto **AREA** es la superficie 2D sobre la que se sustentan el resto de los estadísticos.

El método de obtención de las variables sigue las recomendaciones de la literatura de que, en la mayoría de los casos, cuando debemos combinar datos *raster* y vectoriales es mejor evitar la reproyección de los *raster* y reproyectar la información vectorial (Love-lace, Nowosad y Muenchow 2022). Así se hizo en Goerlich (2022a) y en Goerlich y Cantarino (2022), y así procedemos ahora, de forma que el **MDT05** generado en Goerlich (2022a) se mantuvo en su proyección original. De esta forma la *grid* fue reproyectada a ETRS89-UTM en el Huso original de la información –Huso 30 extendido para Península y Baleares y Huso 28 para Canarias– antes de contar los *pixeles* que caen en cada celda y calcular los estadísticos correspondientes.

Las variables generadas son las siguientes:

- **Altitud:** Altitud media de la celda – en realidad del **AREA** cubierta por la celda–.
- **AREA3D:** Superficie 3D entendida como que tenga en cuenta la orografía y compatible con la superficie 2D, **AREA**.
- **TRI_1: Índice de rugosidad** obtenido como la ratio entre **AREA3D** y **AREA**. Rango de variación: $[1, +\infty+\infty]$
- **Crecimiento:** Crecimiento porcentual de la superficie 3D respecto a la superficie 2D: $100 * (\text{TRI}_1 - 1)$
- **Pendiente:** Pendiente porcentual derivada de la relación entre **AREA3D** y **AREA**.
- **Grados:** Pendiente en grados sexagesimales derivada de la relación entre **AREA3D** y **AREA**.
- **TRI_2: Índice de rugosidad** acotado entre 0 –terreno plano– y 1 –máxima rugosidad–. Ordinalmente equivalente a **TRI_1**: $1 - (1/\text{TRI}_1)$.

Todas **las variables toman como referencia la celda recortada por el contorno administrativo**. Además, hay 9 celdas que no disponen de información, por no existir datos en el **MDT05**, y que corresponden al Peñón Velez de la Gomera, Peñón de Alhucemas, Islas de Alhucemas e Islas Chafarinas.

Nota técnica sobre los cálculos y el ajuste de superficies:

Naturalmente las diferentes proyecciones tienen un error en la determinación de las superficies que fue preciso evaluar. Tomando como referencia la superficie de la *grid*, que utiliza una proyección que mantiene deliberadamente áreas –*Lambert*

Azimuthal Equal Area–, y tomando estas como referencia, no cometeríamos ningún error si en cada celda interior cayeran 40.000 *pixeles* del **MDT05**. En la práctica, el error – medido sobre las celdas interiores– no alcanzó nunca el 1%, el **error máximo** a nivel de celda fue de 0.89% y el **mínimo** de – 0.12%, pero en la inmensa mayoría de los casos fue mucho menor. El **error medio** fue solo del **0.4 por mil**, lo que se consideró más que aceptable.¹¹

Como altitud representativa de la celda se tomó la media de los valores del **MDT05** que caían en la misma¹². Las superficies 2D y 3D se obtuvieron como en Goerlich y Cantarino (2022), lo que generó unas superficies agregadas consistentes con las ofrecidas en dicho trabajo, y en la proyección original de los datos. A partir de estas superficies se obtuvo el **índice de rugosidad** a nivel de celda

como **conciente entre la superficie 3D y la superficie 2D, TRI_1**. Puesto que la superficie 2D es conocida en este caso, **AREA**, y es exactamente 1 para las celdas interiores, la superficie 3D a nivel de celda –es decir, en proyección ETRS89-LAEA– fue derivada del **TRI_1** obtenido en las superficies originales con un ajuste proporcional final que garantizara que el valor de **TRI_1** agregado a nivel nacional –descontando las 9 celdas para las que no disponemos de información– coincidiera exactamente con el obtenido en las superficies originales y manteniendo la restricción de que **AREA3D** \geq **AREA**. Los ajustes fueron mínimos. La correlación entre las superficies 3D en ambas proyecciones –original y ETRS89-LAEA– fue de 1 con una precisión de 4 decimales y entre los índices de rugosidad de 1 con una precisión de 8 decimales.

¹¹ Se ensayó también con el procedimiento inverso, reprojectar el *raster* a ETRS89-LAEA antes de calcular los estadísticos, pero el error fue inasumiblemente elevado, el error medio fue del –1.5%, lo que distorsionaba de forma importante el soporte que da justificación a los cálculos. Un método alternativo, consistente en pasar el *raster* a vectorial de puntos –centroide de la celda– y reprojectar esta información puntual a ETRS89-LAEA, fue numéricamente equivalente al método

expuesto en el texto, pero resultó considerablemente más lento desde el punto de vista de cálculo

¹² Donde el criterio de pertenencia es que caiga el centroide del *pixel* del *raster* en la celda de la *grid*, lo que evita el problema de la doble contabilidad de *pixeles*.

5.

Población

Recopilación de las diversas series de población para las que se dispone de información en este formato. En todos los casos es **población residente fechada a 1 de enero** del año correspondiente, salvo para 2011, ya que el Censo 2011 tiene fecha de referencia 1 de noviembre de dicho año *–night population–*.

- **POB2006:** Población residente en 2006. Básicamente **GEOSTAT2006** para España. Dicha *grid* fue producida por Goerlich y Cantarino (2013a), con los métodos descritos en dicho trabajo, pero utilizando 2006 como fecha de referencia. Celdas habitadas: 94,915. Población en la *grid*: 44,708,964 *–coincidente con la de Padrón–*.
- **POB2010:** Población residente en 2010. *Grid* derivada del trabajo de Goerlich y Cantarino (2012). Celdas habitadas: 94,439. Población en la *grid*: 46,787,184.
- **POB2011:** Población residente en 2011. Básicamente **GEOSTAT2011** para España. Dicha *grid* fue producida por el INE, y *–con algunas modificaciones–* es la utilizada en el trabajo de Reig, Goerlich y Cantarino (2016). Celdas habitadas: 63,528. Población en la *grid*: 46,815,916 *–coincidente con la del Censo 2011 y por tanto se trata de una población en números reales–*.
- **POB2018:** Población residente en 2018. **GEOSTAT2018** para España. Dicha *grid* fue producida por el *Joint*

Research Center (Batista e Silva, Poelman y Dijkstra 2021) como actualización de GEOSTAT2011. Celdas habitadas: 74,175. Población en la *grid*: 46,744,251.

- **POB2021:** Población residente en 2021 procedente del Censo 2021 INE sin redondeo. Celdas habitadas: 115,410. Población en la *grid*: 47,400,798.
- **POB2021_V1_0:** Población residente en 2021 provisional (V1-0). A finales de marzo de 2023 *–version de 13/03/2023–* Eurostat publicó una versión provisional de **GEOSTAT2021** *–renombrada ahora como Census Grid 2021–*. Celdas habitadas: 95,734. Población en la *grid*: 47,428,082.¹³

Las celdas sin población aparecen sin valor *–missing o na–* en la base de datos, **no** con 0. Lo que es consistente con el resto de las variables incluidas en la misma.

Todas las variables de población se pasaron, cada una de ellas, a una capa *raster* con la misma resolución *–1km ×× 1km–*, extensión *–extend–* y CRS *–ETRS89-LAEA–* que la *grid* de referencia. En este caso las celdas sin valor de población, pero dentro de la *grid* de referencia, tienen valor 0, lo que constituye una diferencia respecto a las capas vectoriales para estas variables. Los valores *missing* o *na* en las capas *raster* se reservan para píxeles del *extend* que caen fuera de la *grid* de referencia.

¹³ Esta *grid* se hizo pública por parte de Eurostat el 13/03/2023, mientras que la *grid* del INE se hizo pública el 15/01/2024.

A efectos de su utilización es necesario destacar que **ninguna** de estas *grids* anteriores a la de 2021 es realmente *bottom-up* y que las diferentes metodologías de elaboración afectan –en mi opinión de forma importante en algunos casos– a la comparabilidad temporal (Goerlich y Cantarino 2014, 2015b, 2017).

Notas técnicas sobre las series de población:

Las *grids* poblacionales fueron ajustadas a nuestra *grid* de referencia –apartado 2– que contiene 511,294 celdas por lo que algunos comentarios técnicos son pertinentes.

- **POB2006:** La comparación de esta variable con **GEOSTAT2006** puede arrojar algunas diferencias menores que es necesario precisar. Por una parte, la *grid* enviada a Eurostat no contenía toda la población del Padrón por problemas derivados de la información de usos del suelo utilizada –SIOSE 2005–, aunque esta población se recuperó posteriormente (Goerlich y Cantarino 2014). Por otra parte, la *grid* generada en aquel momento utilizaba como marco de referencia unos contornos administrativos –apartado 2.2– descargados con anterioridad. Al comparar las celdas habitadas en 2006 con las celdas disponibles en la *grid* utilizada ahora perdíamos una celda que contenía solo 2 habitantes. Investigado el origen del problema se identificó que era debido a una ligera alteración en la línea de costa en el municipio de Fasnia (38012), Tenerife. Dicha población fue añadida a la celda inmediatamente superior, y perteneciente al mismo municipio. La *grid* original de 2006 contenía, pues, 94,916 celdas habitadas y la misma población.
- **POB2010:** Esta *grid* no contiene la totalidad de la población del Padrón de 2010 –47,021,031–, solo el 99.5% de la misma, ya que la información de base para su elaboración fue la misma que la que se usó para **POB2006** y teníamos el mismo problema con la información de usos del suelo –SIOSE 2005–. En este caso no se recuperó la población faltante, ya que era un trabajo muy manual y esta *grid* no se utilizó posteriormente una vez el INE hizo pública la *grid* derivada del censo 2011. También se observó el mismo problema con la celda del municipio de Fasnia (38012), Tenerife. En este caso solo afectaba a 1 habitante. Se le dió la misma solución. La *grid* original de 2010 contenía, pues, 94,440 celdas habitadas y la misma población.
- **POB2011:** Para 2011 se dispone de 2 *grids* derivadas del censo 2011 que no son enteramente consistentes. Por una parte, la publicada por el INE en su *web*, redondeada al 5 y que excluye la población en establecimientos colectivos. Por otra parte, la disponible en el GISCO de Eurostat, **GEOSTAT2011**, que incluye toda la población del censo –redondeada a enteros–.¹⁴ El punto de partida de Reig, Goerlich y Cantarino (2016) para la elaboración de **POB2011** no fue ninguna de estas dos *grids*, sino la original que el INE envió a Eurostat –en números reales y ajustada exactamente a la población del censo 2011–.

¹⁴ El proceso de redondeo llevado a cabo por Eurostat suponía la pérdida de algunos efectivos demográficos y 5 celdas habitadas respecto a la *grid* original que le remitió el INE.

Un análisis detallado de esta información reveló numerosas inconsistencias cuando se hacía cuadrar con la estructura municipal del censo, por lo que se le efectuaron algunos ajustes sin alterar la población censal –en reales–. Dichos ajustes están descritos en el Apéndice A.1 de Reig, Goerlich y Cantarino (2016) y básicamente consistieron en hacer que dicha *grid* fuera consistente con las poblaciones municipales cuando se cruzaba con los lindes administrativos. Todas las celdas de esta *grid* tienen su correspondencia en la *grid* de este trabajo y, por tanto, no se requirieron ajustes en este sentido.

- **POB2018:** Esta *grid* procede directamente de la descarga de **GEOS-TAT2018** y la extracción de las celdas para España, al tratarse de una *grid* Europea. Las celdas son únicas, lo que significa que en las celdas frontera no es posible identificar a que país concreto pertenece la población o si hay población –y cuanta– de cada país. La extracción de las celdas de España, cuando ES aparece en el identificador de país, CNTR_ID, arrojó 74,698 celdas habitadas con una población de 46,804,347 habitantes. Existen celdas frontera con población en Andorra, 10, Francia, 321, Gibraltar, 6, y Portugal, 824.¹⁵ Cuando cruzamos estas celdas con la *grid* de este trabajo perdemos 523 celdas habitadas. Las 74,175 celdas que tienen correspondencia con nuestra *grid* arrojan una población de 46,744,251 habitantes. Algo superior a la del Padrón de

2018 –46,722,980–. Por lo que nos quedamos con dichas celdas como **POB2018**. Las celdas que perdemos son todas frontera con Francia, 164, y Portugal, 359.¹⁶

- **POB2021:** Esta *grid* procede directamente de la descarga del INE. El fichero original dispone de 115,410 registros, que se corresponden con el número de celdas habitadas. La población coincide exactamente con la del censo – 47,400,798 personas–.
- **POB2021_V1_0:** Esta *grid* procede directamente de la descarga de **GEOS-TAT2021** y la extracción de las celdas para España, al tratarse de una *grid* Europea. La versión provisional V1-0 publicada por Eurostat a finales de marzo 2023 solo incluye dos variables: (i) el identificador de celda, GRD_ID, y (ii) el total de población, OBS_VALUE_T. No hay identificador geográfico, por lo que no es posible saber a qué país o región pertenece la celda. Las celdas para España se obtuvieron por intersección con la *grid* generada –descrita en el apartado 2–, lo que arrojó 95,734 celdas habitadas con una población de 47,428,082 habitantes, ligeramente superior a la ofrecida por el Censo 2021 – 47,400,798 personas–. Esta variable se actualizará conforme Eurostat o el INE vayan actualizando la información en este formato.

¹⁵ Es posible encontrar **GEOSTAT2018** en dos sitios en la *web* de Eurostat, en el GISCO y en Grids. ¡La *grid* es la misma. Sin embargo, en la disponible en el GISCO, además de CNTR_ID existe un identificador de país, Country, no descrito en la información de descarga y que contiene identificadores de 1 solo país –sin identificar celdas frontera–. En este caso, en las celdas

anteriores encontramos, además de celdas de España, celdas de Francia, 219, y de Portugal, 489.

¹⁶ Si miramos al identificador de país, Country, las celdas que perdemos corresponden a Francia, 164, y Portugal, 359.

6.

Población 2021 próxima dentro de un radio de 10km

Para **POB2021** se calculó la **población próxima dentro de una vecindad circular con un radio de 10km** para cada una de las celdas con población.

El objeto es calcular medidas de densidad del tipo propuesto por [De La Roca y Puga \(2017\)](#) y [Duranton y Puga \(2020\)](#), aunque naturalmente otros radios de proximidad y otras opciones de ponderación por distancia entre las celdas implicadas –que están ausentes en este cálculo– serían posibles ([Henderson, Nigmatulina y Kriticos 2021](#)), lo que requeriría cálculos específicos.

El procedimiento es básicamente el siguiente:

1. Para una celda dada se calcula el centroide de esta¹⁷ y se traza un círculo de 10km de radio –lo que representa una superficie de 314km²–.
2. A partir de ahí, se determinan las celdas que caen dentro de dicho círculo –que constituye la vecindad que determina la población próxima–.
3. Una vez conocidas dichas celdas, se suma la población de estas. Esta es considerada la **población próxima** a la celda en cuestión.

El procedimiento se itera para todas las celdas con población –115.410–.

Obsérvese que la población de la propia celda está incluida en la población próxima, pero como esta es conocida puede deducirse con facilidad si se desea, y que el

cálculo no incluye ponderaciones por distancia entre las celdas, de forma que la población de todas las celdas implicadas pesa lo mismo, independientemente de lo cerca o lejos que estén de la celda de interés.

Al objeto de determinar las **celdas que caen dentro de la vecindad circular** se consideran **2 criterios alternativos** que afectan a las celdas en el límite de la vecindad. Por una parte, se consideran aquellas celdas cuyo centroide cae dentro de la vecindad circular. Por otra parte, se consideran aquellas celdas que tienen intersección no nula con la vecindad circular, lo que naturalmente arroja un mayor número de celdas, y de población próxima, que en el caso anterior. Se almacenan los resultados para ambos criterios, lo que permite examinar la sensibilidad de los resultados en función de uno u otro supuesto.

Las variables generadas son las siguientes:

- **POB2021in10km_np**: Número de celdas cuyo centroide cae dentro de la vecindad circular con un radio de 10km alrededor de la celda en cuestión.
- **POB2021in10km_p**: Población de las celdas cuyo centroide cae dentro de la vecindad circular con un radio de 10km alrededor de la celda en cuestión.
- **POB2021in10km_ns**: Número de celdas con intersección no nula con una vecindad circular con un radio de 10km alrededor de la celda en cuestión.

¹⁷ En la práctica utilizamos la *grid* puntual.

- **POB2021in10km_s**: Población de las celdas con intersección no nula con una vecindad circular con un radio de 10km alrededor de la celda en cuestión.

Por construcción, **POB2021in10km_ns** \geq **POB2021in10km_np** y **POB2021in10km_s** \geq **POB2021in10km_p**.

Notas técnicas sobre el cálculo de la población próxima:

Los cálculos se realizan enteramente en vectorial, a pesar de que en *raster*, a partir de estadísticos focales, serían mucho más eficientes.

En la práctica, los cálculos para los centroides de las celdas se realizan enteramente a partir de la *grid* puntual, mientras que los cálculos que consideran una intersección no nula de las celdas con la vecindad circular correspondiente utilizan la *grid* poligonal, si bien la vecindad circular se genera mediante un *buffer* de 10km del centroide de la celda de interés –utilizando en la práctica la *grid* puntual–.

Aunque la población próxima se ha calculado para un radio único de 10km, el *script* de cálculo permite cambiar este parámetro y calcular la población próxima con otro radio de vecindad. El tiempo de ejecución es de aproximadamente 1 día. La introducción de ponderaciones por distancia entre las celdas de la vecindad respecto a la celda de interés requeriría, sin embargo, de modificaciones en el *script*.

7. Coberturas y usos del suelo: SIOSE 2014

La información para estas variables procede del **Sistema de Información de Ocupación del Suelo de España (SIOSE) 2014**,¹⁸ disponible en el Centro de Descargas del Centro Nacional de Información Geográfica (CNIG), dependiente del IGN.

SIOSE es una base de datos de ocupación de suelo orientada a objetos, lo que implica que cada polígono tiene un descriptor que debe guardar determinadas reglas a partir de un conjunto de coberturas simples. Su estructura está descrita con detalle en la información técnica de SIOSE (Equipo Técnico Nacional SIOSE 2015, 2018a, 2018b), o en trabajos previos donde ha sido utilizada de forma intensiva (Goerlich y Cantarino 2012, 2013b; Reig, Goerlich y Cantarino 2016). La información contenida en **SIOSE** es de extraordinaria riqueza y puede combinarse de múltiples formas, pero para transformar dicha información a las celdas de nuestra *grid* es mucho más útil un modelo jerárquico en el que cada polígono tiene asignada una cobertura, aunque naturalmente ello suponga una cierta pérdida de información¹⁹. Esta es la estructura de *Corine Land Cover*, la base de datos de ocupación de suelo de referencia europea.

Afortunadamente **todos** los polígonos de **SIOSE 2014** tiene una doble clasificación jerárquica que deriva de las recomendaciones de Inspire sobre la **cubierta terrestre** –

Anexo II, tema 2, *Land Cover*– y los **usos del suelo** –Anexo III, tema 4, *Land Use*–. Son estas clasificaciones las que se ofrecen en el visor iberpix del IGN.

Para el caso de las **coberturas del suelo** el campo se denomina **CODIIGE** y contempla la clasificación en las 46 coberturas, a 3 niveles, que se ofrecen en el cuadro 2.

Para el caso de los **usos del suelo** el campo se denomina **HILUCS** y contempla la clasificación en los 16 usos, a 3 niveles, que se ofrecen en el cuadro 3.

La información sobre coberturas –**CODIIGE**– es mucho más rica e interpretable que la información sobre usos –**HILUCS**–. No obstante, casi por el mismo precio, se decidió transformar a la *grid* ambas clasificaciones jerárquicas de **SIOSE**. El proceso fue relativamente simple, pero computacionalmente muy intensivo. **SIOSE** es una base de datos vectorial, cuya unidad geométrica es el polígono, y que se distribuye por CCAA –con fichero único para Ceuta y Melilla–. Consta de unos 2.5 millones de polígonos, cada uno de ellos con su rótulo **SIOSE** y los campos **CODIIGE** e **HILUCS** –además de otra información de interés que no utilizamos en esta aplicación (Equipo Técnico Nacional SIOSE 2018b)²⁰–.

¹⁸ En el momento de acometer la elaboración de estas variables –diciembre 2022– esta es la última fecha para la que está disponible **SIOSE** completo para el conjunto nacional, estando parcialmente disponible **SIOSE** de **Alta Resolución (SIOSE-AR)** para 2017. **SIOSE-AR** supone, además de un aumento considerable en la resolución, un cambio en la estructura de la información disponible.

¹⁹ **SIOSE 2014** tiene unos 2.5 millones de polígonos, mientras que nuestra *grid* algo más de medio millón, por lo que

transformar **SIOSE 2014** a la *grid* de 1km × 1km supone, en conjunto, además de una cierta pérdida de información, una disminución de la resolución.

²⁰ Por ejemplo, para cada polígono disponemos de su superficie en hectáreas –calculada en UTM en el Huso correspondiente–, el porcentaje de arbolado forestal o de sellado del suelo, en el caso de que el polígono en cuestión tenga presencia de arbolado forestal o coberturas artificiales.

Cuadro 2. Clasificación de Coberturas del Suelo - CODIIGE

CODIIGE	Descripción
111	Casco
112	Ensanche
113	Discontinuo
114	Zona verde urbana
121	Instalación agrícola y/o ganadera
122	Instalación forestal
123	Extracción minera
130	Industrial
140	Servicio dotacional
150	Asentamiento agrícola y huerta
161	Red viaria o ferroviaria
162	Puerto
163	Aeropuerto
171	Infraestructura de suministro
172	Infraestructura de residuos
210	Cultivo herbáceo
220	Invernadero
231	Frutal cítrico
232	Frutal no cítrico
233	Viñedo
234	Olivar
235	Otros cultivos leñosos
236	Combinación de cultivos leñosos
240	Prado
250	Combinación de cultivos
260	Combinación de cultivos con vegetación
311	Bosque de frondosas
312	Bosque de coníferas
313	Bosque mixto
320	Pastizal o herbazal
330	Matorral
340	Combinación de vegetación
351	Playa, duna o arenal
352	Roquedo
353	Temporalmente desarbolado por incendios
354	Suelo desnudo
411	Zona húmeda y pantanosa
412	Turbera
413	Marisma
414	Salina
511	Curso de agua
512	Lago o laguna
513	Embalse
514	Lámina de agua artificial
515	Mar
516	Glaciar y/o nieve perpetua

Fuente: CNIG del IGN y elaboración propia

Cuadro 3. Clasificación de Usos del Suelo - HILUCS

HILUCS	Descripción
110	1_1_Agriculture
120	1_2_Forestry
130	1_3_MiningAndQuarrying
140	1_4_AquacultureAndFishing
200	2_SecondaryProduction
310	3_1_CommercialServices
330	3_3_CommunityServices
340	3_4_CulturalEntertainmentAndRecreationalServices
410	4_1_TransportNetworks
430	4_3_Uilities
500	5_ResidentialUse
610	6_1_TransitionalAreas
620	6_2_AbandonedAreas
631	6_3_1_LandAreasNotInOtherEconomicUse
632	6_3_2_WaterAreasNotInOtherEconomicUse
660	6_6_NotKnownUse

Fuente: SIOSE del IGN

El proceso de transferencia de **SIOSE** a la *grid* procedió secuencialmente como sigue:

1. Reproyección de **SIOSE** a ETRS89-LAEA²¹.
2. Disolución de los polígonos por el campo **CODIIGE** e **HILUCS** para cada CCAA²².
3. Juntamos todas las CCAA y disolvemos de nuevo por el campo **CODIIGE** e **HILUCS** para disponer de una capa nacional con 46 o 16 registros respectivamente *-features-*.
4. Intersectamos la capa anterior con la *grid*, de forma que disponemos de las coberturas **CODIIGE** e **HILUCS** que hay en cada celda de 1km².

²¹ La proyección original es ETRS89-UTM, Canarias en el Huso 28, Galicia en el Huso 29, Illes Balears y Cataluña en el Huso 31 y el resto de CCAA en el Huso 30.

²² Hay 9 polígonos en el fichero de Ceuta y Melilla que no tienen CCAA (**CODBLQ**) asignado, es *NA* o *missing*. Lo más razonable es que se trate del código "20" que pertenece a las Islas

Alhucemas, Chafarinas y Peñón Vélez de la Gomera, ya que dicho código no aparece en la base de datos, pero sí en la documentación *-mail Julián Delgado, IGN, 12/12/2022-*. Naturalmente estos polígonos no se perdieron ya que finalmente se obtuvo una capa nacional por disolución.

5. Calculamos el área de la intersección anterior. Los valores obtenidos representan la superficie de cada cobertura **CODIIGE** e **HILUCS** en cada celda de la *grid*. Como estas superficies, en principio, suman 1, pueden interpretarse como la distribución porcentual de coberturas **CODIIGE** e **HILUCS** de cada celda.

Algunos comentarios son de interés.

1. El **formato** de la información finalmente almacenada no son **porcentajes** que suman 1, sino que **suman 100** –con 2 decimales– de forma exacta para el conjunto de coberturas **CODIIGE** e **HILUCS**, y que deben interpretarse como el **porcentaje de la cobertura o uso correspondiente de la totalidad de cada celda de 1km²**.
2. Las superficies hacen referencia a las de la *grid* original, **no** a las recortadas por los lindes administrativos. Por tanto, **todas** las superficies suman 100 para cada celda. **SIOSE** contiene, implícitamente, lindes administrativos a nivel de CCAA –que es la unidad de distribución–, al igual que una línea de costa, sin embargo, estas líneas no tienen por qué coincidir con las de la Base de Datos de Líneas Límite (BDLL) del **IGN**, de hecho, no lo hacen. Por tanto, el proceso de intersección geométrica presenta ciertas discrepancias entre superficies, sobre todo en las celdas exteriores, que es necesario acomodar.
3. **SIOSE 2014** se distribuye con un *buffer* de 2.5km respecto a la línea de costa. Se trata de un polígono con cobertura **CODIIGE** de *Mar*, código 515, y uso **HILUCS** de *Water Areas Not In Other Economic Use*, código 632, por lo que las

celdas exteriores que **lindan con la costa** tienen intersección completa no nula con **SIOSE**. Esto es así para todas las CCAA con salida al mar excepto Ceuta y Melilla, en las que **SIOSE 2014** está ajustado a la línea de costa y carece de dicho *buffer*²³. Para estas celdas la intersección con **SIOSE 2014** genera una superficie menor que 1.

4. Las **celdas exteriores** que **lindan con países extranjeros** no disponen de ningún *buffer*, por lo que su intersección con **SIOSE** también genera una superficie menor que 1. Además, los límites de **SIOSE 2014** no coinciden con los límites exteriores de la BDLL del **IGN** utilizada.
5. Para algunas **celdas interiores** el proceso de intersección geométrica generó una superficie ligeramente menor que la unidad. Probablemente por cuestiones de precisión geométrica, y porque algunas celdas son interiores según la BDLL del **IGN**, pero no según **SIOSE**.
6. Finalmente, **SIOSE** carece de información –intersección geométrica nula– para 25 celdas de nuestra *grid*. Todas estas celdas son exteriores y con un **AREA** muy pequeña –la mayor presenta una superficie de 3.9Ha y la menor de 1m²–, 3 de ellas son de línea de costa y los 22 restantes de línea de frontera.

Para generar la información de coberturas y usos del suelo a nivel de celda –para todas ellas– de forma exacta se procedió de la siguiente forma:

1. Se acepta una precisión del 0.01% –1 por 10,000–, de forma que para aquellas celdas en las que la intersección con **SIOSE 2014** genera una superficie agregada de al menos 0.9999km² –lo que

²³ Ello se debió a un error por parte del IGN –mail Julián Delgado, **IGN**, 12/12/2022–.

equivale a un error máximo admisible de 100m²– las áreas de dicha intersección simplemente se ajustan de forma exacta a enteros después de haber sido multiplicadas por 10,000.

2. Tanto para la clasificación **CODIIGE** como para la clasificación **HILUCS**, las celdas para las que la intersección con **SIOSE 2014** generan una superficie agregada inferior a 0.9999km² resultaron ser 2,435. Solo 12 de ellas son celdas interiores –por problemas de consistencia entre los lindes de **SIOSE** y los de la BDLL del **IGN**–²⁴. Los 2,423 restantes son celdas exteriores. Para las celdas interiores simplemente se ajustaron las superficies como en el punto anterior. Para las exteriores se procedió de la siguiente forma:

- (i) Si son exteriores por la línea de costa se completó la superficie hasta 1km² con la cobertura **CODIIGE** de *Mar*, código 515, o uso **HILUCS** de *Water Areas Not In Other Economic Use*, código 632, y a continuación se ajustaron las superficies como en el punto anterior.
- (ii) Si son exteriores por lindar con un país extranjero se completó la superficie hasta 1km² con una cobertura o uso adicional a la clasificación **CODIIGE** o **HILUCS** de *Exterior*, código 999 en ambos casos, y a continuación se ajustaron las superficies como en el punto anterior. En estos casos no conocemos la cobertura o uso de la superficie faltante, que por otra parte no es de interés.

3. Para las 25 celdas para las que no tenemos información simplemente se siguió la regla del punto anterior.

- (i) Si son exteriores por la línea de costa se les asignó el 100% de superficie con la cobertura **CODIIGE** de *Mar*, código 515, o uso **HILUCS** de *Water Areas Not In Other Economic Use*, código 632.
- (ii) Si son exteriores por lindar con un país extranjero se les asignó el 100% de la superficie con la cobertura o uso adicional a la clasificación **CODIIGE** o **HILUCS** de *Exterior*, código 999 en ambos casos.

De esta forma disponemos de todas las celdas con coberturas y usos del suelo. En el caso de coberturas disponemos de las 46 de la Tabla 2 más la de *Exterior*, código 999, y en el caso de usos disponemos de las 16 de la Tabla 3 más la de *Exterior*, código 999. La información generada no se transformó en variables de la base de datos, sino que se grabó en formato *long* en *csv*²⁵, con 3 campos:

- **GRD_ID**: Identificador de celda.
- **CODIIGE** o **HILUCS**: Según sea el caso, con el código de cobertura –Tabla 2– o uso –Tabla 3–.
- **AREAPC**: Porcentaje, con 2 decimales, de la superficie de la celda y la cobertura o uso correspondiente.

Nota técnica sobre **SIOSE 2014**:

El análisis de la información de **SIOSE** reveló, además de 2,491 polígonos con geometrías

²⁴ De estas 12 solo 1 celda, CRS3035RES1000mN2278000E2828000, presenta discrepancias apreciables –del 15%– entre las áreas de la *grid* –1km²– y las derivadas de **SIOSE 2014** –0.8451km²–. Se trata de una celda del municipio de Entrimo (32030), en la provincia de Ourense, y cerca de la frontera con Portugal –a 35.7m de distancia de dicha frontera según nuestros cálculos–.

²⁵ Naturalmente se dispone de los ficheros originales de la intersección sin ningún tipo de ajuste, y también de un fichero intermedio que incluye, por celda y cobertura o uso, los porcentajes de superficies sin ajustar, ajustados y el número de intersecciones de los polígonos de **SIOSE** con la *grid*, **nCODIIGE** o **nHILUCS** según se trate de coberturas o usos.

inválidas, la existencia de 2 polígonos con geometría CURVEPOLYGON en la CCAA de Aragón. Se trata de dos polígonos contiguos, uno de ellos en Huesca y otro en Zaragoza. Las geometrías inválidas fueron transformadas a válidas y los dos polígonos curvos fueron transformados a geometría MULTIPOLYGON para poder operar con ellos sin problemas.

En relación con las celdas exteriores que se completaron a la unidad, en el caso de Ceuta y Melilla –y alguna que otra esquina– existen unas pocas celdas que lindan con el mar y con Marruecos –u otro país–. En estos casos se otorgó prioridad a la cobertura de *Mar*, código 515 de **CODIIGE**, o uso *Water Areas Not In Other Economic Use*, código 632 de **HILUCS**.

8.

Capas tierra –Land– y agua –Water–: SIOSE 2014

A partir de la información de **SIOSE 2014** del apartado anterior se generó una capa de **tierra**, *land*, y otra complementaria de **agua**, *water*, que sea útil para determinar densidades por celda. Específicamente, el objeto es generar una capa *raster* que pueda servir como *input* en la herramienta *Degree of Urbanisation Grid (GHS-DUG)* del sistema de información desarrollado en el marco de la *Global Human Settlement Layer (GHSL)* (Maffeni, Schiavina, Melchiori, Pesaresi y Kemper 2023).

Se consideraron como **tierra** las coberturas 111 a 354 de CODIIGE en la tabla 2, y como **agua** el resto, es decir, las coberturas 411 a 516, más la 999 que representa las celdas de *Exterior*.

A partir de los resultados del apartado anterior, que obtienen la distribución porcentual de coberturas CODIIGE por celda, la agregación es inmediata. La distribución porcentual **tierra/agua** de cada celda se almacenó en dos variables:

- **TIERRA:** Porcentaje de cobertura terrestre en la celda.
- **AGUA:** Porcentaje de cobertura de agua en la celda.

Si la ocupación es nula se consigna el valor 0.

Al igual que sucede con la información generada en el apartado anterior, los valores almacenados no suman 1, sino 100 –con 2 decimales– de forma exacta para las dos coberturas consideradas, **tierra/agua**, y que deben interpretarse como el **porcentaje de**

tierra/agua correspondiente de la totalidad de cada celda de 1km².

Estas dos variables se pasaron, cada una de ellas, a una capa *raster* con la misma resolución –1km ×× 1km–, extensión –*extend*– y CRS –ETRS89-LAEA– que la *grid* de referencia.

9. **Built-up, altura y volumen edificado**

A partir de Goerlich (2023) simplemente se pasaron a la *grid* en vectorial –tabla– valores del porcentaje edificado por celda, altura media de los edificios (m) y volumen edificado por celda (m³). Esta información existe en *raster* derivada de Goerlich (2023), por lo que el proceso simplemente extrae los valores del *raster* correspondiente y lo pasa al vectorial. La información se corresponde, pues, con la primera cobertura LiDAR y no está disponible para Ceuta, Melilla, Isla de Alborán y posesiones africanas –Isla Perejil, Peñón Velez de la Gomera, Peñón de Alucemas, Islas de Alucemas e Islas Chafarinas–.

Si no existen valores se consigna el valor 0, excepto en las celdas donde no existe información que se consigna *missing, NA*.

Las variables generadas son:

- **BUILT_UP**: Porcentaje de ocupación edificado de la celda sin redondeo respecto a la información original. Capa *raster* de origen: Built-up_epsg3035_1km.tif
- **HEIGHT**: Altura media (m) de los edificios en la celda redondeado a dos decimales (cm). Capa *raster* de origen: Height_epsg3035_1km.tif
- **VOLUME**: Volumen edificado (m³) en la celda redondeado a enteros (m³). Capa *raster* de origen: Volume_epsg3035_1km_FLT8S_2.tif

Hay que tener en cuenta que de los 365,201 píxeles originales con valor positivo solo 362,118 caen dentro de nuestra *grid*. Ya se observó en Goerlich (2023) que esta información no está recortada por lindes

administrativos y existen valores fuera de nuestras fronteras. Tampoco hemos aplicado ahora una máscara por lindes administrativos para la obtención de los valores por celda, limitándonos simplemente a transferir los valores del *raster* de 1km × × 1km a las celdas de la *grid* vectorial. Esto implica que las celdas frontera contienen valores de edificaciones fuera de nuestras fronteras.

A partir de la información de **BUILT_UP** se generó una capa *raster* de **BUILT_UP** que sea útil para reducir la fragmentación de los centros urbanos en la determinación del grado de urbanización a nivel de *grid*. Específicamente, el objeto es generar una capa *raster* que pueda servir como *input* en la herramienta *Degree of Urbanisation Grid (GHS-DUG)* del sistema de información desarrollado en el marco de la *Global Human Settlement Layer (GHSL)* (Maffenini, Schiavina, Melchiori, Pesaresi y Kemper 2023). Esta capa es similar a la original, Built-up_epsg3035_1km.tif, excepto que solo contiene valores entre 0 y 100 para las celdas de la *grid*, y *missing, NA*, para valores no cubiertos por los contornos administrativos que han generado la *grid*. Por otra parte, para los territorios para los que no disponemos de información –Ceuta, Melilla, Isla de Alborán y posesiones africanas– se consigna 0, y no *missing, NA*, lo que constituye una diferencia respecto a los valores consignados en el fichero vectorial.

10. Servicios Públicos

La información para estas variables procede de la Monografía del Ivие para la Fundación Ramón Areces sobre *Accesibilidad de la Población y Acceso a los Servicios Públicos* (Goerlich, Maudos y Mollá 2021) y comprende servicios sanitarios y educativos.

10.1 Servicios de Salud

A partir de la coordenada puntual de los centros de servicio de salud se pasan a la *grid* el número de (i) **Hospitales**, (ii) **Centros de Salud** y (iii) **Consultorios Locales**. Las variables generadas son:

- **HOSP**: Número de Hospitales (466).
- **CS**: Número de Centros de Salud (3,051).
- **CL**: Número de Consultorios Locales (10,104).

Si se desea el número de **Centros de Atención Primaria** basta con sumar CS y CL.

No se almacena ninguna información relacionada con los centros de servicio más allá de su número por celda. Esto se podría modificar a partir del *script* que hace el trasvase de información de las coordenadas puntuales a la *grid*, por ejemplo, si se quisiera distinguir entre los hospitales públicos y los privados con concierto.

Si el valor de la celda es *missing*, *NA*, es que el número de centros de servicio es 0.

10.2 Servicios Educativos

A partir de la coordenada puntual de los centros de servicio educativos se pasan a la *grid* el número de (i) Centros de **Infantil de segunda etapa**, (ii) Centros de **Primaria**, (iii)

Centros de **Secundaria** y (iv) Centros de **Bachillerato**. Las variables generadas son:

- **INFANTIL2**: Número de Centros de Infantil de segunda etapa (13,832).
- **PRIMARIA**: Número de Centros de Primaria (13,598).
- **SECUNDARIA**: Número de Centros de Secundaria (7,013).
- **BACHILLERATO**: Número de Centros de Bachiller (3,539).

Si se desea la agrupación de determinados centros basta con sumarlos, aunque en muchos casos el mismo centro imparte diferentes niveles educativos.

No se almacena ninguna información relacionada con los centros de servicio más allá de su número por celda. Esto se podría modificar a partir del *script* que hace el trasvase de información de las coordenadas puntuales a la *grid*, por ejemplo, si se quisiera distinguir entre los centros públicos y los privados con concierto.

Si el valor de la celda es *missing*, *NA*, es que el número de centros de servicio es 0.

11. Servicios Privados

11.1 Oficinas Bancarias

La información para estas variables procede de la Monografía del Ivie para la Fundación Ramón Areces sobre *Accesibilidad de la Población y Acceso a los Servicios Públicos* en lo que respecta a los años 2008 y 2019 (Goerlich, Maudos y Mollá 2021). El año 2020 procede de una actualización que se hizo para los seminarios sobre geocodificación en el Ivie, que se tomó dicho año como ejemplo.

A partir de la coordenada puntual de las Oficinas Bancarias se pasan a la *grid* el número de oficinas de cada año. Las variables generadas son:

- **Oficinas2008:** Número de Oficinas en 2008 (45,146).
- **Oficinas2019:** Número de Oficinas en 2019 (24,238).
- **Oficinas2020:** Número de Oficinas en 2020 (22,558).

No se almacena ninguna información relacionada con las oficinas más allá de su número por celda. Esto se podría modificar a partir del *script* que hace el trasvase de información de las coordenadas puntuales a la *grid*, por ejemplo, si se quisiera distinguir entre Bancos y Cajas, o se desearan las estadísticas en *grid* por entidades.

Si el valor de la celda es *missing*, *NA*, es que el número de centros de servicio es 0.

11.2 Gasolineras

La información para estas variables procede de la base de datos de la Monografía del Ivie para la Fundación Ramón Areces sobre

Accesibilidad de la Población y Acceso a los Servicios Públicos (Goerlich, Maudos y Mollá 2021), aunque esta variable no fué finalmente incluida en el trabajo. La fuente original es el Geoportal de Gasolineras, aunque los datos fueron minimamente arreglados, básicamente se eliminaron algunos duplicados. La descarga de la información se realizó en 21/12/2019.

A partir de la coordenada puntual de las gasolineras se pasan a la *grid* el número de gasolineras por celda. La variable generada es:

- **Gasolineras:** Número de Gasolineras en 2019 (10,319).

Si el valor de la celda es *missing*, *NA*, es que el número de gasolineras es 0.

Nota técnica sobre las gasolineras:

En el trasvase de la información de las gasolineras a partir de su coordenada puntual a la *grid* se encontró dos gasolineras que caen fuera de la *grid*. Una de ellas, asignada al municipio de Badajoz (06015), se sitúa en territorio portugués. La otra, asignada al municipio de Camargo (39016) se sitúa en el mar. En ambos casos la coordenada se asignó a la celda más cercana,

CRS3035RES1000mN1921000E2847000
en el caso de Badajoz (06015) y
CRS3035RES1000mN2360000E3207000 en
el caso de Camargo (39016), al objeto de no perder gasolineras en la *grid*.

12. *Direcciones y puntos kilométricos*

La información para estas variables procede de la capa *PORTAL_PK* de [Cartociudad](#) y se corresponde con la utilizada en [Goerlich \(2022b\)](#).

La fuente original dispone de 12,636,163 puntos que engloban tanto direcciones postales como puntos kilométricos de carreteras de la red viaria. La separación entre lo que se consideran direcciones y puntos kilométricos se realizó a partir de la variable *TIPO_PORPK* que toma el valor 1 si se trata de un portal –*TIPOPORPKD* = Portal– y valor 2 si se trata de un punto kilométrico –*TIPOPORPKD* = PK–. No es la única opción posible, pero es la más natural de acuerdo con [Cartociudad \(2021\)](#). Esta variable está completa para todos los registros. La descarga de la información se realizó en febrero de 2022, aunque la fecha de referencia de la información que constaba en el centro de descargas era noviembre de 2021.

A partir de la coordenada puntual de las direcciones y los puntos kilométricos se pasan a la *grid* el número de direcciones y puntos kilométricos por celda, conjuntamente y por separado. Las variables generadas son:

- **PORTAL_PK:** Número de direcciones y puntos kilométricos (12,636,163).
- **PORTAL:** Número de direcciones (12,444,757).
- **PK:** Número de puntos kilométricos (191,406).

Si el valor de la celda es *missing*, *NA*, es que la variable correspondiente toma el valor 0.

13. DEGURBA

Clasificación **DEGURBA** de las celdas hecha por la herramienta *Degree of Urbanisation Grid (GHS-DUG)* del sistema de información desarrollado en el marco de la *Global Human Settlement Layer (GHSL)* (Maffenini, Schiavina, Melchiori, Pesaresi y Kemper 2023) del *Joint Research Center (JRC)* a partir de la *grid* censal de 2021 –la población de referencia es pues **POB2021** de esta misma base de datos–.

La herramienta *Degree of Urbanisation Grid (GHS-DUG)* genera dos capas *raster* correspondientes a dos niveles jerárquicos en la clasificación **DEGURBA**. Las capas *raster* generadas tienen la misma resolución, *extend* y *CRS* que la capa *raster* de población de partida.

El nivel 1 de la clasificación **DEGURBA** tiene 3 clases:

- **Class 3:** "Urban Centre grid cell"
- **Class 2:** "Urban Cluster grid cell"
- **Class 1:** "Rural grid cell"

El nivel 2 de la clasificación **DEGURBA**, anidado en el anterior, tiene 8 clases:

- **Class 30:** "Urban Centre grid cell"
- **Class 23:** "Dense Urban Cluster grid cell"
- **Class 22:** "Semi-dense Urban Cluster grid cell"
- **Class 21:** "Suburban or peri-urban grid cell"
- **Class 13:** "Rural cluster grid cell"
- **Class 12:** "Low Density Rural grid cell"
- **Class 11:** "Very low density rural grid cell"
- **Class 10:** "Water grid cell"

Todas las celdas son clasificadas en ambos niveles, tengan o no población, y las variables generadas simplemente son la transposición de los valores de las capas *raster* a su contrapartida vectorial.

A efectos de contabilidad de las celdas habitadas en cada una de las clases tendría sentido eliminar las celdas que no tienen población. Esto afecta, sobre todo, a las celdas rurales en el nivel 1 y las celdas rurales de muy baja densidad en el nivel 2, que son identificadas de forma residual, tengan o no población. Algunas de las celdas de los centros urbanos o de los clusters urbanos de alta densidad también carecen de población, como consecuencia de la aplicación del filtro de suavizado de los bordes *-3-by-3 conditional majority filtering-* y el rellenado de huecos (Maffenini, Schiavina, Melchiori, Pesaresi y Kemper 2023).

Las variables generadas son:

- **DEGURBA_L1:** Clasificación DEGURBA de nivel 1: 3 clases.
- **DEGURBA_L2:** Clasificación DEGURBA de nivel 2: 8 clases.

14. Referencias

BATISTA E SILVA, F.; POELMAN, H. Y DIJKSTRA, L. (2021) *The JRC-GEOSTAT 2018 a novel small-area population estimation for Europe*. Presentado en el 14th European Forum for Geography and Statistics (EFGS) Conference - 7-8 September 2021.

CARTOCIUDAD (2021) *Especificaciones del Producto Cartociudad* version 13. 11-05-2021. Centro Nacional de Información Geográfica. Instituto Geográfico Nacional.

DE LA ROCA, J. Y PUGA, D. (2017) "Learning by Working in Big Cities". *The Review of Economic Studies*. 84, 1, (January), 106–142.

DURANTON, G. Y PUGA, D. (2020) "The Economics of Urban Density". *Journal of Economic Perspectives*. 34, 3, (Summer), 3-26.

EQUIPO TÉCNICO NACIONAL SIOSE (2015) *Descripción del Modelo de Datos SIOSE* Versión 2. Instituto Geográfico Nacional. Ministerio de Fomento. Madrid.

EQUIPO TÉCNICO NACIONAL SIOSE (2018A) *Documento Técnico SIOSE 2014* Versión 1. Instituto Geográfico Nacional. Ministerio de Fomento. Madrid.

EQUIPO TÉCNICO NACIONAL SIOSE (2018B) *Estructura y Consulta de la base de datos SIOSE* Versión 3. Instituto Geográfico Nacional. Ministerio de Fomento. Madrid.

EUROSTAT (2021) *GridMaker*. Version 1.3.

GOERLICH, F. J. (2016) "Una aproximación volumétrica a la desagregación espacial de la población combinando cartografía temática y datos LIDAR". *Revista de Teledetección* 46, (June), 147-163. ISSN: 1133-0953. EISSN: 1988-8740.

GOERLICH, F. J. (2022A) "Superficie planimétrica versus superficie del paisaje en España. –superficie 2D versus superficie 3D–" Documento del Trabajo del Ivie. WP-2022-07.

GOERLICH, F. J. (2022B) "Elaboración de un mapa de Códigos Postales de España con recursos libres. –como evitar pagar a  6000€ por información de referencia–" Documento del Trabajo del Ivie. WP-2022-03.

GOERLICH, F. J. (2023) "A System of Built-up, Building Height and Built-up Volume layers for Spain". 23/10/2023.

GOERLICH, F. J. Y CANTARINO, M. (2012) *Una grid de densidad de población para España*. Informes. Economía y Sociedad. Fundación BBVA. (Noviembre).

GOERLICH, F. J. Y CANTARINO, M. (2013A) "A population density grid for Spain". *International Journal of Geographical Information Science*. 27, 12, 2247-2263. doi: 10.1080/13658816.2013.799283.

GOERLICH, F. J. Y CANTARINO, M. (2013B) *Zonas de morfología urbana: Coberturas del suelo y demografía*. Informes. Economía y Sociedad. Fundación BBVA. (Octubre).

GOERLICH, F. J. Y CANTARINO, M. (2014) *Comparing bottom-up and top-down population density grids: The Spanish Census 2011*. Presentado en el 7th European Forum for Geography and Statistics (EFGS) Conference - 22-24 October 2014.

GOERLICH, F. J. Y CANTARINO, M. (2015A) "Estimaciones de la población rural y urbana a nivel municipal". *Estadística Española*. 57, 186, (1^{er} Cuatrimestre), 5-28.

GOERLICH, F. J. Y CANTARINO, M. (2015B) "Census Grid 2011. Una evaluación metodológica". *Working Paper WP-EC 2015-02*. Abril. Ivie.

GOERLICH, F. J. Y CANTARINO, M. (2017) "Grid poblacional 2011 para España. Evaluación metodológica de diversas posibilidades de elaboración". *Estudios Geográficos*. 78, 282, 135-163.

GOERLICH, F. J. Y CANTARINO, I. (2022) "Un índice de rugosidad del terreno a escala municipal – updated– Superficie 2D versus superficie 3D y rugosidad." Documento del Trabajo del Ivie. WP-2022-08.

GOERLICH, F. J.; MAUDOS, J. Y MOLLÁ, S. (2021) *Distribución de la población y acceso a los servicios públicos* Fundación Ramón Areces. Madrid.

GOERLICH, F. J. Y PÉREZ, P. (2021) *LAU2boundaries4spain: R package providing LAU2 (municipalities) data geometries for Spain for 2002-2021* ROpenSpain. Repositorio en GitHub.

HENDERSON, J. V.; NIGMATULINA, D. Y KRITICOS, S. (2021) "Measuring urban economic density". *Journal of Urban Economics*. 125, (September), 103188.

HIJMANS R. (2022) *terra: Spatial Data Analysis*. R package version 1.6-17.

INSPIRE (2014A) *D2.8.I.2 Data Specification on Geographical Grid Systems – Technical Guidelines v3.1* INSPIRE Infrastructure for Spatial Information in Europe. European Commission.

INSPIRE (2014B) *D2.8.I.1 Data Specification on Coordinate Reference Systems – Technical Guidelines v3.2* INSPIRE Infrastructure for Spatial Information in Europe. European Commission.

INSPIRE (2014C) *D2.8.I.2 Data Specification on Administrative Units – Technical Guidelines v3.1* INSPIRE Infrastructure for Spatial Information in Europe. European Commission.

LOVELACE, NOWOSAD Y MUENCHOW (2022) *Geocomputation with R* Second Edition *on line*. First Edition (2021) *R Series* - CRC Press. Routledge Taylor & Francis Group.

MAFFENINI, L.; SCHIAVINA, M.; MELCHIORI, M.; PESARESI, M. Y KEMPER, T. (2023) *GHS-DUG User Guide. Degree of Urbanisation Grid User Guide*. Version 6. JRC Scientific Information Systems and Databases Report. Joint Research Centre. Luxembourg. Publication Office of the European Union.

PEBESMA, E. (2018) "Simple Features for R: Standardized Support for Spatial Vector Data". *The R Journal*. 10, 1, 439-446. doi: 10.32614/RJ-2018-009.

R CORE TEAM (2022) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria.

REIG, E.; GOERLICH, F. J. Y CANTARINO, M. (2016) *Delimitación de áreas rurales y urbanas a nivel local. Demografía, coberturas del suelo y accesibilidad*. Informes. Economía y Sociedad. Fundación BBVA. (Enero).

WICKHAM ET AL., (2019) "Welcome to the tidyverse" *Journal of Open Source Software*. 4, 43, 1686. doi: 10.21105/joss.01686.



Ivie