

REVISTA DE ESTUDIOS REGIONALES

I.S.S.N.: 0213-7585

2ª EPOCA Enero-Abril 2020



117

SUMARIO

Mª Esther López Vizcaíno, Carlos Luis Iglesias Patiño y Mª Esther Calvo Ocampo. Propuesta metodológica para la georreferenciación de la población y primeras aplicaciones en Galicia

Laura Nieto Torrejón y Mª Concepción Pérez Cárcelos. Productividad y eficiencia de los sistemas universitarios regionales de España en el periodo 2009-2013

Ricardo Urrestarazu Capellán, José Luis Sánchez Ollero y Alejandro García Pozo. Una aproximación a la influencia del sector turístico sobre el derecho a la vivienda

Fernando Fernández-Llebrez González. Mecanismos ambientales, cognitivos y relacionales en el surgimiento del actual movimiento independentista Catalán

Gema Gómez-Casero, Miguel Jesús Medina-Viruel, Javier Jeménez Beltrán y Juan Choque Tito. Motivación y valoración de atributos en el gastroturismo de Oruro (Bolivia)

Luis Camarero, Renato Miguel Carmo y Sofía Santos. Mobility as a function of environmental conditions and sociodemographic differentiation: The case of gender inequality in the Lisbon Metropolitan Area

Recensiones y reseñas bibliográficas

I. Artículos

Propuesta metodológica para la georreferenciación de la población y primeras aplicaciones en Galicia

Methodological proposal for the georreferenciation of the population and first applications in Galicia

M^a Esther López Vizcaíno
Carlos Luis Iglesias Patiño
M^a Esther Calvo Ocampo
Instituto Galego de Estatística

Recibido, Febrero de 2018; Versión final aceptada, Diciembre de 2018.

PALABRAS CLAVES: Georreferenciación, Splines, Análisis de componentes principales, Población

KEYWORDS: Geocoded, Splines, Principal component analysis, Population

CLASIFICACIÓN JEL: C5, C63

RESUMEN

Las estadísticas se refieren a una base territorial y la información con una resolución geográfica mayor es cada vez más necesaria para llevar a cabo estudios sobre el asentamiento de la población. El objetivo de este artículo es presentar la metodología seguida en el Instituto Gallego de Estadística para geolocalizar la población de Galicia disponible en el Padrón Municipal de Habitantes, empleando para ello diversas técnicas estadísticas, como son la regresión con splines y el análisis de componentes principales. El proceso de georreferenciación se dividió en dos pasos, por un lado se georreferenció la parte urbana y por otro la parte rural. La metodología que se presenta permitió georreferenciar el 99% de la población de Galicia. Para complementar el objetivo general, se presentan ejemplos de uso de la población geolocalizada, que muestran la importancia y el potencial de la georreferenciación de los datos.

ABSTRACT

The statistics refer to a territorial base and the need for information with a greater geographical resolution is increasingly necessary to carry out studies on the settlement of the population. In the current administrative organization of the kingdom of Spain, municipalities constitute the smaller administrative units into which the territory is divided and which have precise boundaries assigned. For this reason, studies whose objective is the location of the population tend to descend to the municipal level. Anyway, several authors (Reher, 1994; Rúa and others, 2003) have recommended that, for the study of the settlement of the population on the territory, this division is clearly insufficient and it is necessary to increase the geographical resolution of analysis.

In this direction, according to Goerlich and Cantarino (2012), the best way to represent the population is to georeference all the residential buildings of a country, determine the population that resides in them, to count it and to assign it to the corresponding pixel. The determination of the location of the population on the territory, not directly related to the administrative boundaries, is absolutely essential for many practical issues of social organization (Goerlich and Cantarino, 2012). Having georeferenced statistical units it will allow social, economic or environmental planning easier to design and monitor, as well as a more appropriate territory management, less dependent on the administrative subdivision. It also make decisions easier, not only public, but also private, both for profit and non profit entities. This will reduce costs because, in the medium term, many of the ad hoc studies commissioned by stakeholders, and possibly also private ones, will be unnecessary to base their strategy. It is not intended to replace the decision maker but to help him.

Then, the aim of georeferencing is to identify by geographic coordinates where the statistical units are located in the territory. In the files that are used for georeferencing it is common to find a proportion of records that have incomplete information and that can not be assigned to a precise location in the space. This situation is particularly frequent in rural areas where addresses are not standardized and the scattering of the buildings leads to a wider positioning error. In this situation, instead of eliminating these records, what is usually done, in this work, missing coordinates will be imputed using indirect techniques, depending on the available information. The utility of the georeferenced information is multiple: in the resident population by territorial areas, in the transportation planning (Moreno y Prieto, 2003), in the planning of services or facilities, in the calculation of the population in influence zones of schools, senior residences, etc.

The objective of this article is to present the methodology followed in the Galician Institute of Statistics (IGE) to geolocalize the population available in the Municipal Population Register (PMH), using various statistical techniques, such as the regression with splines and principal component analysis. The process of georeferencing was divided into two steps, on the one hand the urban part was georeferenced and on the other the rural part. In the urban part

it was decided to apply a regression model with splines to the coordinates of the specified path. The advantages of using splines are that they adapt to non-rectilinear routes, which in Galicia abound due to their orography, and they can also be represented as mixed models, and therefore benefit from the theory and the existing software on this kind of models (Ruppert et al., 2003; Ugarte et al., 2009; Wood, 2006).

In the rural part, the georeferencing is done at the nucleus/scattering level and the number of the state. The corresponding coordinates are assigned to a weighted average point of the nucleus or the corresponding scattering, because the centroid of the points (vector of arithmetic means) was not satisfactory in some cases, due to the dispersion of the population in Galician territory

The methodology presented allowed us to georeference about 99% of the Galician population. In a later section and to complement this general objective, we present three examples of integration between both types of information, statistical and geographical, which show the importance and potential of georeferencing. The first example refers to the calculation of population in a geographical area that does not coincide with any of the administrative divisions of Galicia, the second and third examples are related to the calculation of statistical indicators of service coverage, allocating inhabitants to influence polygons; in the third, as a previous step to a spatial analysis.

The proposed methodology was implemented in the R software (R Core Team, 2016) and tested with the PMH data for Galicia and for 2016. The goal is to use this methodological proposal, or it

1. INTRODUCCIÓN

Doscientos años atrás, el matemático Domingo Fontán fijó la posición geográfica de la Torre del Reloj de la Catedral de Santiago de Compostela para poder levantar su Carta Geométrica de Galicia, esto es, puso a Galicia en el mapa. Ahora estamos intentando, en el Instituto Gallego de Estadística (IGE), colocar a sus habitantes sobre él.

La práctica totalidad de las estadísticas tienen una base territorial y, por tanto, geográfica y la integración entre la información estadística y geográfica es imparable, ya que permite sinergias capaces de responder a preguntas que aun recientemente no tenían respuesta.

En la actual organización administrativa del Estado español, los municipios constituyen las unidades administrativas menores en las que se divide el territorio y que tienen asignados lindes precisos. Por esta razón, los estudios cuyo objetivo es la localización de la población suelen descender hasta el nivel municipal (Goerlich y Mas, 2009). De todos modos, diversos autores (Reher, 1994, Rúa y otros, 2003) han recomendado que para el estudio del asentamiento de la población sobre el territorio esta división es claramente insuficiente y es necesario aumentar la resolución geográfica de análisis.

Además, el tratamiento histórico dado a las direcciones postales conducía a la asignación de divisiones territoriales, como es el caso de la población en municipios, entidades colectivas o secciones censales. El efecto final de esta asignación es que los resultados estadísticos sólo podían obtenerse respecto de estas zonas o sus agregados, produciéndose una pérdida de información no deseable.

Desde el punto de vista geográfico-estadístico, la división del territorio es relevante si los dominios de la partición son homogéneos, pero en el caso de los municipios y entidades colectivas, estamos ante una situación de falta de homogeneidad de superficie, población o concentración de unidades de producción. Teniendo en cuenta esta problemática, hace unos años, algunos institutos de estadística de la Unión Europea con el apoyo de Eurostat lanzaron el European Forum of Geography and Geostatistics (EFGS), con el objetivo de armonizar estadísticas europeas sobre la base de una cuadrícula de 1 km² y en un sistema geodésico de referencia común.

Además, la creciente demanda de estadísticas con un elevado detalle territorial para un análisis espacial más preciso se concreta normativamente en el Reglamento (UE) 2017/2391 del Parlamento Europeo y del Consejo de 12 de diciembre de 2017 por el que se modifica el Reglamento (CE) num. 1059/2003 en lo que respecta a las tipologías territoriales (Tercet), publicado en el Diario Oficial de la Unión Europea de 29 de diciembre de 2017. Este reglamento señala ya la malla estadística, las tipologías de mallas y considera que debe aplicarse un sistema de mallas estadísticas para calcular y atribuir los tipos territoriales a las regiones y zonas en cuestión,

ya que dichos tipos dependen de la distribución y la densidad de la población en celdas de malla de un kilómetro cuadrado. Este nuevo sistema zonal puede permitir combinar el máximo detalle territorial en la difusión con la preservación de la confidencialidad estadística.

En esta dirección, según Goerlich y Cantarino (2012), la mejor manera de representar la población es georreferenciar todos los edificios residenciales de un país, determinar la población que reside en ellos, contarla y asignarla al pixel correspondiente. Esta aproximación se conoce como *bottom-up approach* y su realización es posible a partir de :

- Enormes bases de datos geográficas sobre las construcciones residenciales del país
- Un censo o un registro administrativo de población como el padrón, convenientemente enlazado con la base de datos de los edificios.

En esta línea, y a nivel nacional, el Instituto de Estadística de Cataluña (Idescat) (Lagonigro et al, 2017) está usando técnicas espaciales de procesado de la información para obtener datos geográficos de todos los individuos en el registro de población de Cataluña del año 2014, con el objetivo de poner a disposición información relevante de alta calidad. Idescat también ha trabajado en la automatización de técnicas de visualización geoespacial lo más precisas posibles y que cumplan los requisitos de confidencialidad estadística.

Desde 2013, el Instituto de Estadística y Cartografía de Andalucía (IECA) difunde anualmente en la malla estadística de 250 metros de lado, información relativa a los residentes en Andalucía y sus características sociodemográficas, a partir de la información procedente de la Base de Datos Longitudinal de Población de Andalucía (IECA, 2017).

Por último, en el año 2011 Eurostat publicó una metodología para la clasificación de los territorios empleando un criterio a partir de las densidades de población de las celdas de 1 Km² que cubren todo el territorio (Eurostat, 2011). Para ello es necesario tener la población georreferenciada junto con la malla estadística. La muestra georreferenciada del Censo de población del 2011 (INE, 2013) permitió aplicar este nuevo enfoque también a Galicia y obtener una nueva clasificación y subclasificación de las parroquias y los municipios (IGE, 2016). En este sentido, en el IGE, y para la elaboración de esta clasificación se ha trabajado con la muestra de la población gallega georreferenciada en el Censo del año 2011. Si se quisiera actualizar esta clasificación sería necesario esperar a los resultados de un nuevo Censo de Población, o buscar alternativas a medio plazo, como la georreferenciación de registros de población.

Por tanto, el objetivo final de la georreferenciación es identificar mediante coordenadas geográficas dónde se sitúan las unidades estadísticas en el territo-

rio. En los ficheros que se usan para georreferenciar es frecuente encontrar una proporción de registros que tienen información incompleta y que no pueden ser asignados a una localización precisa en el espacio. Esta situación es particularmente frecuente en zonas rurales donde las direcciones no están estandarizadas y la dispersión de los edificios lleva a que el error de posicionamiento sea más amplio. En esta situación en lugar de eliminar estos registros lo que se suele hacer, y lo que se hará en este trabajo, es imputar las coordenadas utilizando técnicas indirectas, dependiendo de la información disponible (Walter et al. (2013), Dilekli et al. (2018) y Enrique (2016)).

Las utilidades de la información georreferenciada son múltiples: en la población residente por zonas territoriales, en la planificación del transporte (Moreno y Prieto, 2003), en la planificación de servicios o equipamientos, en el cálculo de la población en zonas de influencia de escuelas, residencias, etc. (Goerlich y Cantarino, 2011).

Habida cuenta de todo lo anterior, el objetivo de este artículo es presentar la metodología seguida en el IGE para geolocalizar la población disponible en el Padrón Municipal de Habitantes (PMH) (INE, 2016), empleando para ello diversas técnicas estadísticas. En una sección posterior y para complementar este objetivo general, se tratan ejemplos de integración entre ambos tipos de información, estadística y geográfica, que muestran la importancia y el potencial de la georreferenciación.

La metodología propuesta se implementó en el software R (R Core Team, 2016) y se probó con los datos del PMH del año 2016 para Galicia. La meta es emplear esta propuesta metodológica, o sus futuras mejoras, para que el IGE mantenga la población de Galicia georreferenciada a partir de ese año.

Teniendo en cuenta los precedentes anteriores el resto del artículo se organiza como sigue: en la siguiente sección se recorrerán las fuentes y los métodos empleados en este trabajo, para a continuación presentar los resultados obtenidos aplicando la metodología anterior a la base de datos del PMH del año 2016. Posteriormente, se describirán tres aplicaciones del uso de la información geolocalizada en tres aspectos que podrían interesar para la planificación en Galicia y, finalmente, se presentarán las conclusiones del trabajo.

2. FUENTES Y MÉTODOS

La información que se georreferenciará es la población de Galicia disponible en el PMH de Galicia a 1 de enero de cada año. En el PMH están disponibles el nombre, los apellidos y la dirección postal de todos los habitantes que tienen residencia en esta comunidad autónoma. El objetivo de este trabajo es asignar unas coordenadas

geográficas a las direcciones existentes en el PMH formadas por la vía y el número del portal. Es decir, las personas que viven en el mismo portal tendrán las mismas coordenadas geográficas.

Dado que aún no se dispone en el IGE de un callejero estándar, ni se conoce su existencia en la Comunidad Autónoma hubo que recurrir el Instituto Nacional de Estadística (INE) como fuente principal. Se empleará para georreferenciar la base de datos proporcionada por el INE denominada "Modelo de Direcciones de la Administración General del Estado" (MDAGE). En esta base:

- Están identificados los inmuebles de uso residencial.
- El inmueble está identificado siempre del mismo modo, a partir de una información estructurada y actualizada.
- Las direcciones están localizadas espacialmente (georreferenciadas), con sus coordenadas UTM X e Y en el huso 30
- Se presenta como modelo de referencia para la armonización, integración, intercambio y consulta de direcciones en toda España

El primer trabajo de preparación de la base de datos que se realizó fue transformar las coordenadas UTM del huso 30 al 29. El problema que se tendrá que resolver a partir de este punto es cómo asignar las personas del PMH a los inmuebles de la MDAGE y conseguir así obtener una coordenada geográfica para cada uno de los habitantes de Galicia.

Las variables que se emplean para georreferenciar son:

- Cvia: código de la vía de comunicación construida para la circulación.
- Tvia: tipo de vía
- Nvia: nombre de la vía
- Num: número de portal
- Cec: código de entidad colectiva
- Ces: código de entidad singular
- Cnuc: código de núcleo o diseminado
- Dices: código de distrito censal
- Secc: código de sección censal

En este punto hay que resaltar que la base de datos MDAGE no es completa, existen inmuebles que aún no tienen unas coordenadas UTM válidas o, simplemente, no están georreferenciados. De todas formas, esta será la fuente de información principal, aunque después se complete con las siguientes fuentes de información auxiliares:

- Cartociudad: proyecto colaborativo liderado por el Instituto Geográfico Nacional (IGN) de producción y publicación mediante servicios web de datos espaciales de cobertura nacional. Contiene información de la red viaria continua con las calles, con portales, y las carreteras, con puntos kilométricos. Dispone de un servicio de procesamiento web basado en cálculos programados que operan sobre la información georreferenciada. Este servicio se empleará para georreferenciar direcciones no disponibles en la MDAGE.
- Dirección General del Catastro: dispone de servicios web que permiten la descarga de la cartografía catastral en formato vectorial e información alfanumérica, así como servicios web para la descarga de direcciones y coordenadas. Este servicio se empleará para georreferenciar direcciones no disponibles en la MDAGE ni en Cartociudad.
- Google: para aquellas direcciones que no se puedan georreferenciar con la MDAGE, Cartociudad o la Dirección General de Catastro, se empleará este servicio para buscar coordenadas geográficas entregándole una dirección postal.

El procedimiento de georreferenciación se llevó a cabo en dos pasos, en el primer paso se georreferenció la población que vive en la zona urbana y en el segundo paso la zona rural.

Paso 1. Georreferenciación de la población que vive en la zona urbana

El procedimiento comienza por georreferenciar las personas que residen en viales que tienen disponible el código de vía, tipo de vía, nombre de vía y número, que corresponde principalmente a las zonas urbanas (Paso 1.1). En este caso se cruza la base de datos MDAGE con el PMH empleando las variables anteriores y se le asigna a cada persona las coordenadas correspondientes. En este proceso surge el problema de que existen viales que tienen código de vía pero el número del portal del PMH no está disponible en la MDAGE, en cuyo caso hay que imputar las coordenadas X e Y correspondientes a este nuevo número de portal (coordenadas indirectas).

Se optó por aplicar un modelo de regresión con splines a las coordenadas de la vía especificada, y se tendrá en cuenta si los números del portal son pares o impares. La ventaja de emplear los splines es que se adaptan a vías no rectilíneas, que en Galicia abundan debido a su orografía, y además pueden representarse como modelos mixtos, y por lo tanto se benefician de la teoría y del software existente sobre este tipo de modelos (Ruppert et al., 2003, Ugarte et al., 2009, Wood, 2006). El modelo que se aplicó, para cada vía, fue el siguiente:

$$Y_i = f(X_i) + aI_i + e_i \quad (1)$$

donde

- Y coordenada geográfica UTM Y en el Huso 29,
- X coordenada geográfica UTM X en el Huso 29,
- I indicadora de la paridad del número (1 si es par; 0, impar),
- e variable aleatoria $N(0, \sigma^2)$,
- f función suave,
- i=1, ..., n° total de portales disponibles de la vía.

Para su ajuste se emplean los splines penalizados.

Una vez establecido este primer modelo se necesita un segundo modelo que nos permita calcular la coordenada geográfica X de un nuevo portal que se va a imputar. El planteamiento para cada una de las vías en este segundo modelo es:

$$X_i = g(N_i) + bI_i + e_i' \quad (2)$$

donde

- X coordenada geográfica UTM X en el Huso 29,
- N n° del portal del inmueble,
- I indicadora de la paridad del número (1 si es par; 0, impar),
- g función suave,
- e' variable aleatoria $N(0, \sigma'^2)$,
- i=1, ..., n° total de portales disponibles de la vía.

Para el ajuste de este segundo modelo también se emplearán los splines penalizados. Por tanto, el procedimiento para obtener las coordenadas X e Y para un número de portal no disponible en la MDAGE en una vía concreta es el siguiente:

- Con las coordenadas geográficas X e Y que nos provee la MDAGE para una vía, ajustamos el modelo 1.
- Con las coordenadas geográficas X y los números de portal que nos provee MDAGE para una vía, ajustamos el modelo 2.
- Para predecir la coordenada geográfica X, empleamos el número que queremos predecir y el modelo 2.
- Una vez que se obtiene la coordenada geográfica X del número que vamos a predecir, aplicamos el modelo 1 para obtener la coordenada geográfica Y.

En caso de que no haya suficientes números de portales disponibles en la vía para aplicar la regresión con splines, se optó por asignarle la coordenada del número de portal más próximo. En este sentido se utilizó la distancia numérica entre un portal y otro, en futuras versiones de este estudio sería conveniente investigar la

distancia espacial. Para las personas que quedan sin georreferenciar se emplea la información que proporciona Cartociudad (Paso 1.2) y para las que aún permanecen sin georreferenciar, tras los dos pasos anteriores, se empleará la información que proporciona la Dirección General de Catastro. En concreto se utiliza la información de la cartografía urbana digital en formato shape, que ofrece las coordenadas de las parcelas donde se ubican los inmuebles. El problema que ofrecen estas bases de datos es que no existe un nexo de unión entre los ficheros del PMH y los del Catastro. Como en ambos casos se dispone de las direcciones de los inmuebles, se unirá por ellas tratando de vincular las vías que proporciona el PMH con las de los ficheros de la Dirección General de Catastro.

Para ello se empleará un procedimiento basado en unir tablas mediante variables que no tienen una coincidencia exacta con el paquete de R *fuzzyjoin* (Robinson, 2016). Este paquete emplea un procedimiento basado en distancias entre cadenas de caracteres, en concreto se utilizó la distancia de Levenshtein (Robinson, 2016). Esta viene medida por el número de caracteres que hay que eliminar, incluir o sustituir para transformar una en la otra. Cuanto menor es la distancia de Levenshtein entre dos cadenas de caracteres, más similares son. Este procedimiento comienza con la normalización de los nombres de las vías de las dos bases de datos: se eliminan espacios, caracteres raros, artículos, ... También se emplea el mapa de secciones censales, de tal manera que la vía del Catastro debe pertenecer a la misma sección censal que la disponible en el PMH (Paso 1.3).

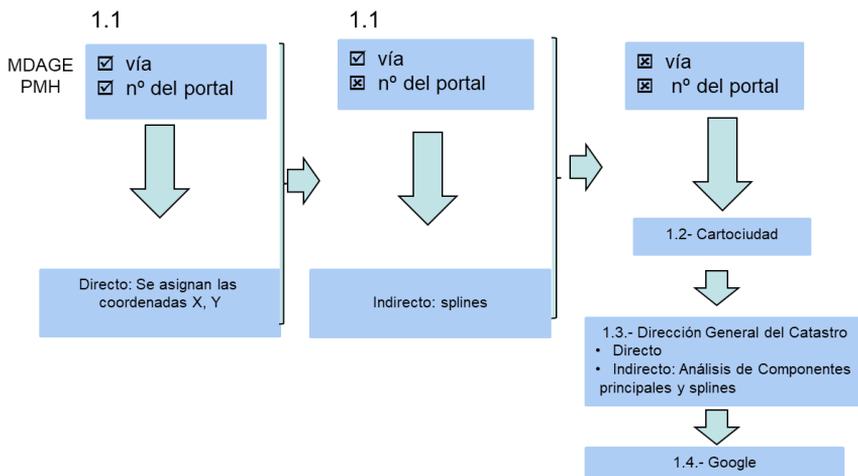
Una vez realizada la correspondencia entre los códigos de vía del PMH y Catastro, es posible que el Catastro no contenga todos los números de portal disponibles en el PMH. Esto ocurre, sobre todo, en las vías de reciente creación. En este caso pueden ocurrir dos situaciones diferentes:

1. Algunos números de portales que dispone Catastro son nulos: cuando estos contienen ceros, no se puede saber si se trata de portales con numeración par o impar. Para averiguar esto se aplicará el análisis de componentes principales (Uriel y Aldás, 2005) a las coordenadas X e Y que nos aporta el Catastro para la vía. Se puede comprobar que la primera componente principal se corresponde con la dirección de la vía, en el sentido geométrico, y la segunda, ortogonal a la primera, con la paridad. Una vez averiguado cuáles son los portales pares e impares, ya podemos calcular el centroide de las coordenadas de los pares y de los impares, respectivamente, y asignárselo al número de portal que se quiere georreferenciar, dependiendo de su paridad.
2. Ningún número de portal es nulo: en esta situación para determinar las coordenadas del nuevo número se aplica la regresión con splines, que se utilizó en el paso 1.1, a la vía especificada.

En la Figura 1 se presenta un resumen del proceso de georreferenciación seguido en las zonas urbanas, especificando si el proceso es directo o indirecto. Se considera un proceso directo cuando la relación entre las dos bases de datos que se utilizan es directa e indirecta cuando no hay una correspondencia entre la base de datos del PMH y la base de datos de las coordenadas y se emplean técnicas de imputación.

FIGURA 1
ESQUEMA DE LOS PASOS DE LA GEORREFERENCIACIÓN EN LAS ZONAS URBANAS

1. Georreferenciar a las personas que viven en las zonas urbanas



Fuente: Elaboración propia.

Paso 2. Georreferenciación de la población que vive en la zona rural

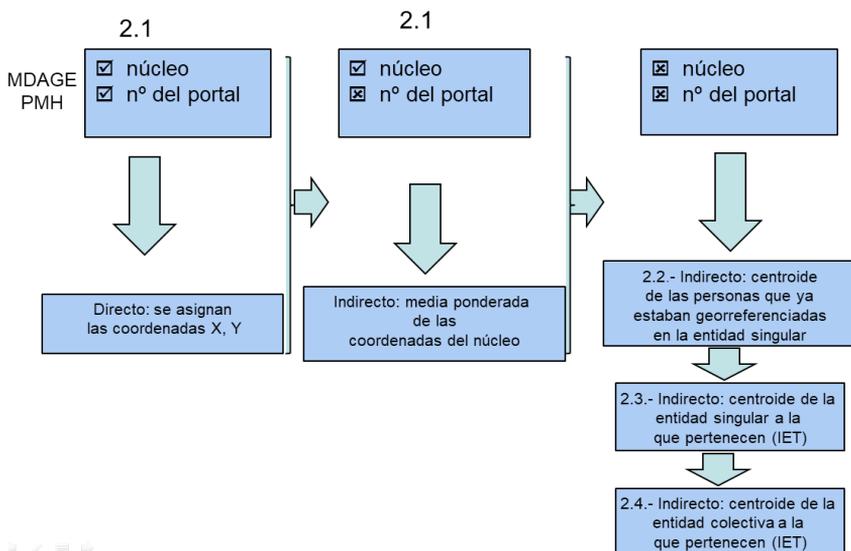
Por último, es necesario georreferenciar las personas que viven en entidades que no tienen disponibles, en el PMH, la clase de vía, código o nombre de vía, que se corresponde principalmente con la zona rural. En este caso la georreferenciación se realiza a nivel de núcleo/diseminado y número del inmueble. Para aquellas personas que tienen un número de portal en el PMH que está disponible en la MDAGE, se le asigna la coordenada disponible en la MDAGE. Para aquellas personas que viven en inmuebles que tienen un número de portal en el PMH y este número no está disponible en la MDAGE se le asigna el promedio ponderado de las coordenadas del núcleo/diseminado al que pertenecen según la siguiente metodología (Paso 2.1).

Se distingue el núcleo del diseminado de una entidad singular y se tratan por separado. Tanto para el núcleo como para el diseminado, se asigna las coordenadas correspondientes a un punto medio ponderado del núcleo o del diseminado correspondiente, porque el centroide de los puntos (vector de medias aritméticas) no resultaba satisfactorio en algunos casos, debido a la dispersión de la población en el territorio gallego.

El peso de un punto (x_p, y_p) es igual a d_i^{-1} siendo $d_i = d((x_p, y_p); ((x_j, y_j)))$ con $((x_j, y_j))$ el centroide de la nube de puntos original excepto el propio punto (x_p, y_p) . En el caso de que exista vía, asignamos el vecino más próximo si disponen de número o el centroide si no disponen de él. Esta metodología se comprobó mediante un caso real en un municipio tipo de Galicia como se puede consultar en la ponencia García, Iglesias y López (2017).

FIGURA 2
ESQUEMA DE LOS PASOS DE LA GEORREFERENCIACIÓN EN LAS ZONAS RURALES

2. Georreferenciar las zonas rurales (no tienen código de vía).



Fuente: Elaboración propia.

Para las personas que permanecen sin georreferenciar en el paso anterior, se les asignó el centroide de la entidad singular a la que pertenecen. Para ello se siguió el siguiente procedimiento:

- Para aquellas entidades singulares en las que ya se tenía personas georreferenciadas, se les asignó el punto medio de las personas que ya estaban georreferenciadas en estas entidades singulares (Paso 2.2).
- Para aquella entidad singular que carecía de personas georreferenciadas, se le asignó el centroide de la entidad singular del mapa de estas entidades que nos proporcionó el Instituto de Estudios do Territorio (IET), elaborado a partir de la información del Censo de Viviendas del año 2011 y de otros datos disponibles en ese instituto (Paso 2.3).
- Para las personas de las restantes entidades singulares (no georreferenciadas en el paso anterior), se le asignó el centroide de la entidad colectiva a la que pertenecen (Paso 2.4).

En la Figura 2 se presenta un resumen del proceso de georreferenciación seguido en las zonas rurales.

3. RESULTADOS

El procedimiento anterior se aplicó a todos los municipios de Galicia y al PMH del año 2016. De las 2.718.525 personas disponibles en el PMH, se georreferenciaron 2.690.299 personas que se corresponden con 550.532 portales (Cuadro 1). En el paso 1.1 y 2.1 no se pudieron geolocalizar 190.705 personas (7%). En el paso 1.1 quedaron sin georreferenciar 43.605 personas y en el paso 2.1 quedaron sin georreferenciar 147.100 personas.

En el paso 1.2, llamando a los servicios de Cartociudad se pudieron georreferenciar 12.793 personas. Para las personas que quedan sin georreferenciar, se acudió a la información georreferenciada que publica la Dirección General de Catastro completada con la información de Google. Se consiguió georreferenciar a 2.665 personas más.

En el caso de las 147.100 personas sin georreferenciar del paso 2.1, a 4.678 se le asignó el punto medio de las personas ya georreferenciadas en el paso 2.1, a 129.456 se le asignó el centroide de la entidad singular proporcionado por el IET, a 12.887 se le asignó el centroide de la entidad colectiva (parroquia) y 79 no se consiguieron clasificar.

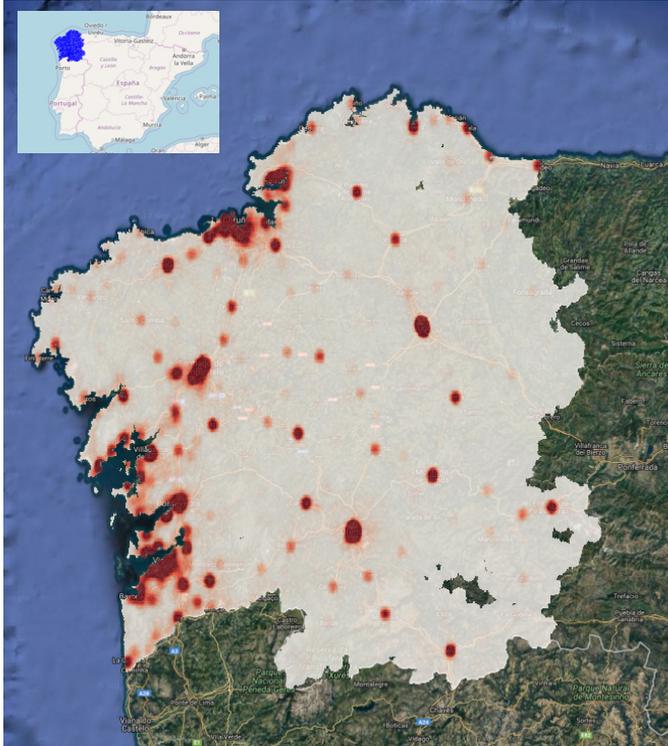
En la Figura 3, se presenta un mapa de calor de las personas georreferenciadas en Galicia siguiendo el procedimiento descrito.

CUADRO 1
PERSONAS Y PORTALES GEORREFERENCIADAS EN CADA UNO DE LOS PASOS DEL PROCEDIMIENTO

Paso	Fuente	Proceso	Personas	Personas georreferenciadas	Portales georreferenciados
Urbano (con vía)			1.977.813		
1.1	MDADE	Directo		1.851.933	267.841
	MDAGE	Indirecto		82.275	20.166
1.2	Cartociudad			12.793	3.704
1.3	Catastro	Directo		1.192	244
	Catastro	Indirecto		1.211	206
1.4	Google			262	59
Rural (sin vía)			740.712		
2.1	MDAGE	Directo		487.870	173.102
	MDAGE	Indirecto		105.742	35.068
2.2	MDAGE	Indirecto		4.678	1.521
2.3	Centroide entidad singular	Indirecto		129.456	44.424
2.4	Centroide entidad colectiva	Indirecto		12.887	4.197
Total			2.718.525	2.690.299	550.532

Fuente: Elaboración propia

FIGURA 3
MAPA DE CALOR DE LAS PERSONAS GEORREFERENCIADAS



Fuente: Elaboración propia

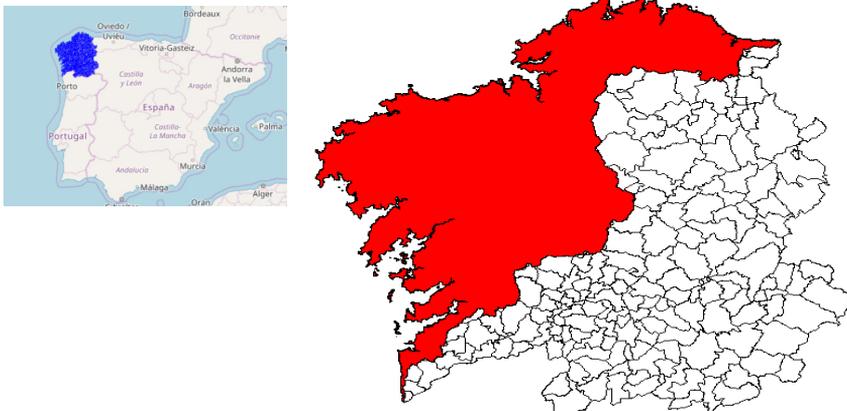
4. APLICACIONES

En esta sección se presentan tres aplicaciones de uso de la información georreferenciada de la sección anterior, que muestra la importancia y el potencial de tener una localización aproximada de la población. El primer ejemplo se refiere al cálculo de población en una zona geográfica que no coincide con ninguna de las divisiones administrativas de Galicia, el segundo y tercer ejemplo tienen relación con el cálculo de indicadores estadísticos de cobertura de servicios, contando con la población residente a la que van dirigidos según polígonos de influencia; en el tercero, como paso previo a un análisis espacial.

4.1 Cálculo de población en zonas que no coinciden con divisiones administrativas: la Demarcación Hidrográfica Galicia-Costa

En este caso el objetivo es calcular la población que reside en la Demarcación Hidrográfica Galicia-Costa, que comprende las cuencas de los ríos que vierten al mar Cantábrico (excepto las de los ríos Eo y Navia) o al océano Atlántico (con la exclusión del sistema Miño-Sil). El territorio de esta demarcación se presenta en rojo en la Figura 4. Esta división no se corresponde con ninguna administrativa disponible en la Comunidad Autónoma de Galicia, por lo que, para calcular la población que vive en esta Demarcación Hidrográfica es preciso situar las personas con sus coordenadas (X, Y) sobre el espacio geográfico y luego hacer el recuento de habitantes.

FIGURA 4
DEMARCACIÓN HIDROGRÁFICA GALICIA-COSTA (ROJO)



Demarcación Hidrográfica Galicia-Costa
Escala 1 cm: 3.030.303 cm

Fuente: Xunta de Galicia. Augas de Galicia

En el Cuadro 2 se muestran las poblaciones calculadas empleando las coordenadas (X, Y), la envolvente de entidades colectivas (parroquias) y la de municipios. Como era de esperar, se comete más error aproximando la población de la zona con la envolvente de municipios o con la envolvente de parroquias. En el caso de los municipios la Demarcación Hidrográfica contaría con 153.622 personas más que si el cálculo se realiza aproximando la población por sus coordenadas, lo que supone más de la mitad de los habitantes del municipio de Vigo.

CUADRO 2
**POBLACIÓN DE LA DEMARCACIÓN HIDROGRÁFICA GALICIA-COSTA
 CALCULADA CON DIVERSOS MÉTODOS**

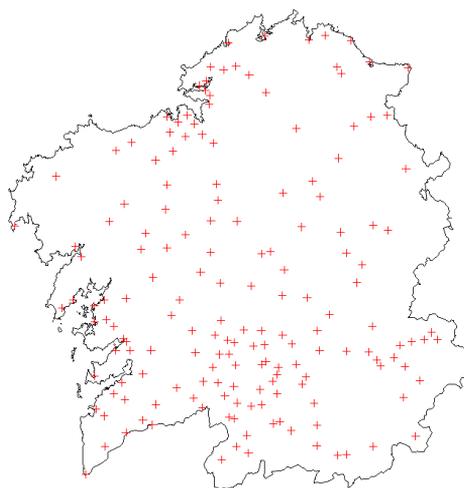
	Población
Cálculo con coordenadas (X,Y)	1.995.402
Cálculo de la envolvente por parroquias	2.070.645
Cálculo de la envolvente por municipios	2.149.024
Galicia	2.718.525

Fuente: IGE y Elaboración propia

4.2 Indicadores de cobertura: Las Residencias de Mayores.

El objetivo de esta aplicación es determinar un polígono de influencia de las residencias de mayores que nos pueda informar sobre la adecuada cobertura de este tipo de centros a nivel local. Galicia en el año 2016 contaba con 661.255 personas de 65 o más años y un total de 342 residencias de mayores. En la Figura 5 se puede observar la ubicación de este tipo de centros en Galicia.

FIGURA 5
UBICACIÓN DE LOS CENTROS RESIDENCIALES DE MAYORES EN GALICIA



+ Centro residencial
 Escala 1 cm: 2.500.000 cm

Fuente: Xunta de Galicia. Consellería de Política Social

Para determinar la cobertura de estos centros, en primer lugar se calcula el centro de masas¹ de las residencias por municipios, con lo cual contaremos con un punto central en cada municipio donde disponen de residencias de mayores. Este primer paso se hace con el objeto de reducir el número de zonas de influencia calculadas y para tratar de homogeneizar las grandes ciudades con el resto de los municipios. En las ciudades y sus coronas suele haber varias posibles candidatas dentro del municipio, por eso se calculó el centro de masas para representarlos a todos. El objetivo ahora es determinar cuál es la zona de atracción de cada residencia única o de cada conjunto municipal de residencias y comprobar la cobertura en la dotación de plazas.

Por tanto, se debe determinar una configuración de zonas que permita que el número de mayores atraídos por cada una ellas sea el más factible. Se supondrá que los mayores irían, en principio, al establecimiento más próximo a su domicilio y no a aquellos que estén más lejos (esta hipótesis debe matizarse en las ciudades debido a la existencia de varios centros de mayores representados por el centro de masas).

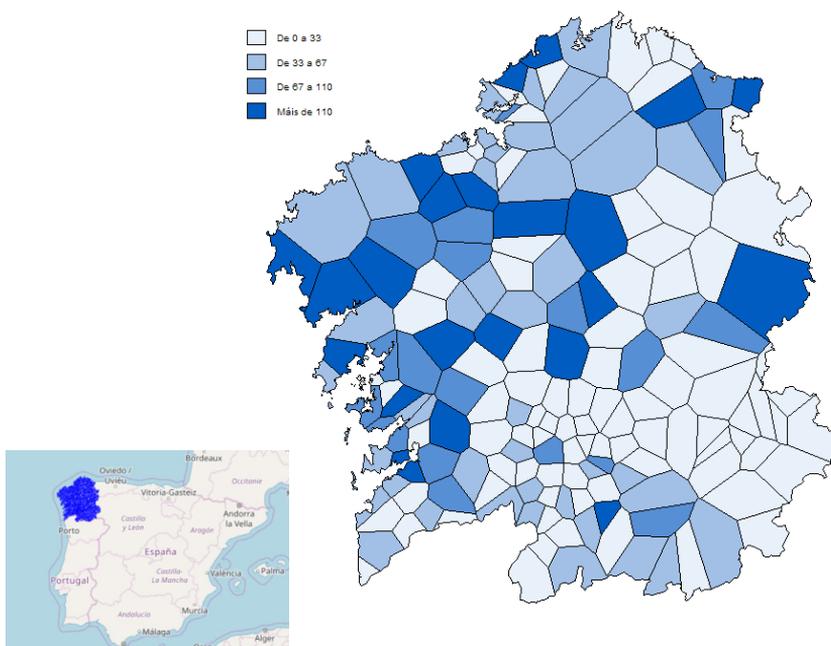
Considerando esto, se emplearán los diagramas de Voronoi que inducen una partición de la región geográfica en función de un conjunto de puntos dado, en este caso, cada uno es un centro de masas (recuérdese que si solo existe una residencia, el punto correspondiente es su propio centro de masas). Esta partición está constituida por polígonos. Cualquiera de ellos contiene un único centro de masas y engloba todos los puntos más próximos a este centro.

Por tanto, se calcula la partición de Voronoi de Galicia y se obtiene el mapa que se presenta en la Figura 6. Cada uno de los polígonos de este gráfico determina el subconjunto de puntos de Galicia más cercanos y por tanto la población que habita más cerca de la residencia representada por el centro de masas. Una vez que se dispone de la partición, se compara la población residente en cada polígono de influencia con las plazas de sus respectivas residencias.

1 Los centros de masas se calculan como $(\frac{\sum x_i p_i}{\sum p_i}, \frac{\sum y_i p_i}{\sum p_i})$ donde, p_i plazas de la residencia i , x_i 1ª coordenada de la residencia i e y_i 2ª coordenada de la residencia i .

FIGURA 6
POLÍGONOS DE VORONOI DE LAS RESIDENCIAS DE MAYORES EN GALICIA

Mapa del ratio de población de 65 y más años por plaza



Escala 1 cm: 2.000.000 cm

Fuente: Elaboración propia

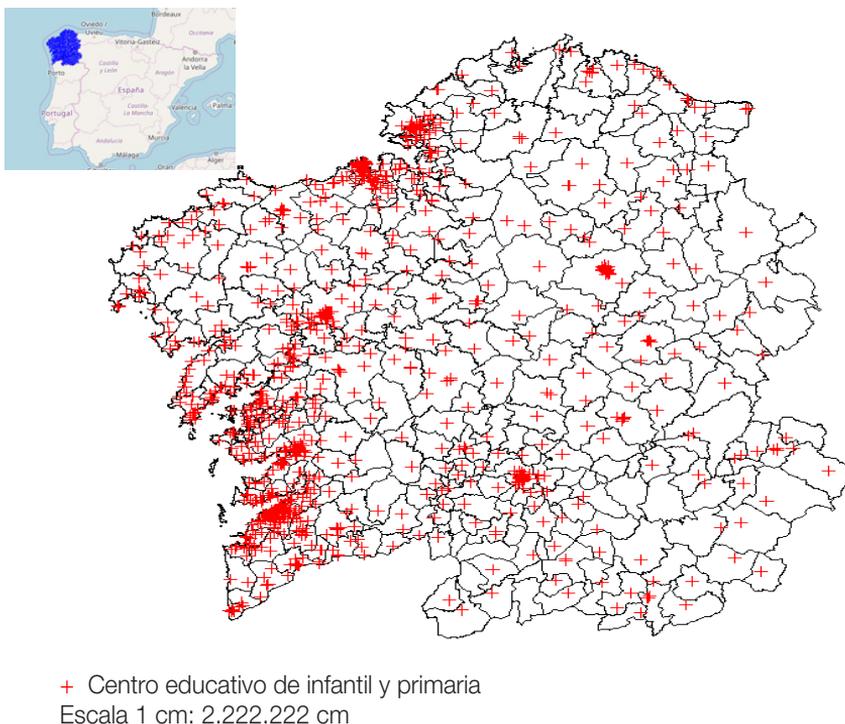
En la Figura 6 se presenta, para cada uno de los polígonos, el ratio de población de 65 y más años entre el número de plazas disponibles. Los polígonos con color azul oscuro tienen un ratio de más de 110 personas por plaza y conforman las zonas de influencia más deficitarias de Galicia, en cuanto a número de plazas en residencias de mayores. Se corresponde fundamentalmente con el interior de la provincia de A Coruña y con algunas zonas de la parte nororiental de la provincia de Pontevedra.

4.3 Aproximación a la distribución de los centros educativos de infantil y primaria

El objetivo de esta aplicación es aproximarnos a la distribución de los centros educativos de infantil y primaria, teniendo en cuenta la ubicación de los alumnos gallegos de 3 a 11 años. En el año 2016, Galicia dispone de 201.874 niños, en este tramo de edades y de 971 centros educativos con 10.099 unidades educativas con capacidad para 252.475 niños. A priori la oferta supera la demanda, con lo cual resulta interesante observar la distribución por zonas.

En la Figura 7 se muestra el mapa de los municipios de Galicia con los centros educativos de infantil y primaria disponibles y su ubicación.

FIGURA 7
**UBICACIÓN DE LOS CENTROS EDUCATIVOS DE INFANTIL Y PRIMARIA
SOBRE LOS MUNICIPIOS GALLEGOS**



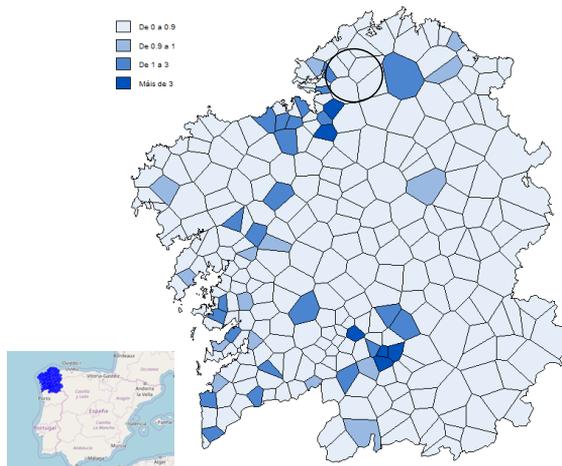
Fuente: Xunta de Galicia. Consellería de Cultura, Educación e Ordenación Universitaria

Para observar una primera distribución de los centros educativos, contando con la distribución de la población por el territorio, se calcula el centro de masas de los centros educativos por municipios, con lo cual se dispone de un punto central en cada municipio donde existen centros. Ahora determinaremos la partición de Voronoi para Galicia, considerando los centros de masas calculados en el punto anterior. Con esto se calcula un polígono de influencia de cada centro de masas, de tal manera que los niños que viven en ese polígono, tienen por centro educativo más próximo el que está representado por su centro de masas.

En la Figura 8 se muestra el ratio de población de 3 a 11 años entre las plazas disponibles en los centros educativos. Las zonas donde los centros educativos tienen mayor ocupación están en las coronas de las grandes ciudades. Sin embargo, se observa que la mayor parte de los polígonos de Galicia tienen un número de plazas superior al necesario para satisfacer al alumnado disponible en su entorno. En esta situación se podría plantear algún tipo de redistribución de los centros, con el fin de reducir costes sin que los alumnos se vean demasiado perjudicados.

FIGURA 8 MAPA CON LOS POLÍGONOS DE VORONOI DE LOS CENTROS EDUCATIVOS POR MUNICIPIOS

Mapa del ratio de población de 3 a 11 años por plazas



Escala 1 cm: 2.000.000 cm

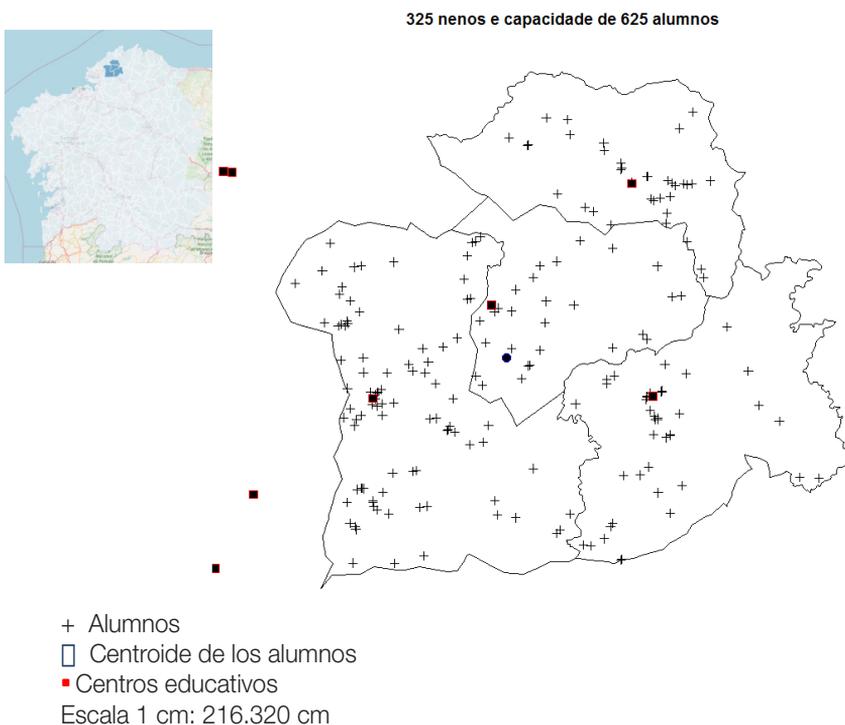


Zona formada por los municipios de Cerdido, Moeche, Somozas y San Sadurniño

Fuente: Elaboración propia

Para ejemplificar la situación anterior, consideramos una zona de Galicia donde la oferta supera a la demanda, como es la formada por los municipios de Cerdido, Moeche, Somozas y San Sadurniño, marcada en la Figura 8 con un círculo. En todos estos municipios el ratio de población por plaza es inferior a 0,9, y en As Somozas se sitúa en 0,63. Esta zona tiene 271 Km² y tiene 4 centros educativos con 25 unidades, con capacidad para 625 alumnos, y cuenta con una población susceptible de escolarizar en infantil y primaria de 325 niños. Estamos ante una situación donde la oferta supera a la demanda con claridad.

FIGURA 9
MAPA DE LOS MUNICIPIOS DE CERDIDO, MOECHE, AS SOMOZAS Y SAN SADURNIÑO CON LOS CENTROS EDUCATIVOS (ROJO) Y CON EL CENTROIDE DE LOS ALUMNOS (AZUL)



Fuente: Xunta de Galicia y Elaboración propia.

En primer lugar calculamos el centroide de los niños de estos cuatro municipios, y comprobamos que se encuentra en el concello de Moeche (punto azul de la Figura 9). Como el centro educativo de Moeche está muy próximo al centroide, consideraremos que el centro de referencia de estos cuatro municipios sería el centro educativo del municipio de Moeche. A continuación, se calcula las distancias por carretera y el tiempo de desplazamiento de todos los niños al colegio de Moeche. Para ello se emplea el software R con la función *mapdist* (Kahle y Wickham, 2013) que, a su vez, consiste en una llamada a la API que proporciona Google. Las distancias por carretera que tendrían que recorrer los alumnos de toda la zona para ir al centro educativo del municipio de Moeche serían de 21 km y emplearían 30 minutos como máximo. La distancia media sería de 9 km con un tiempo medio de 11,8 minutos (Cuadro 3).

CUADRO 3
DISTANCIA Y TIEMPO DE LOS NIÑOS DE SOMOZAS, CERDIDO, SAN SADURNIÑO Y MOECHE AL CENTRO EDUCATIVO DE MOECHE

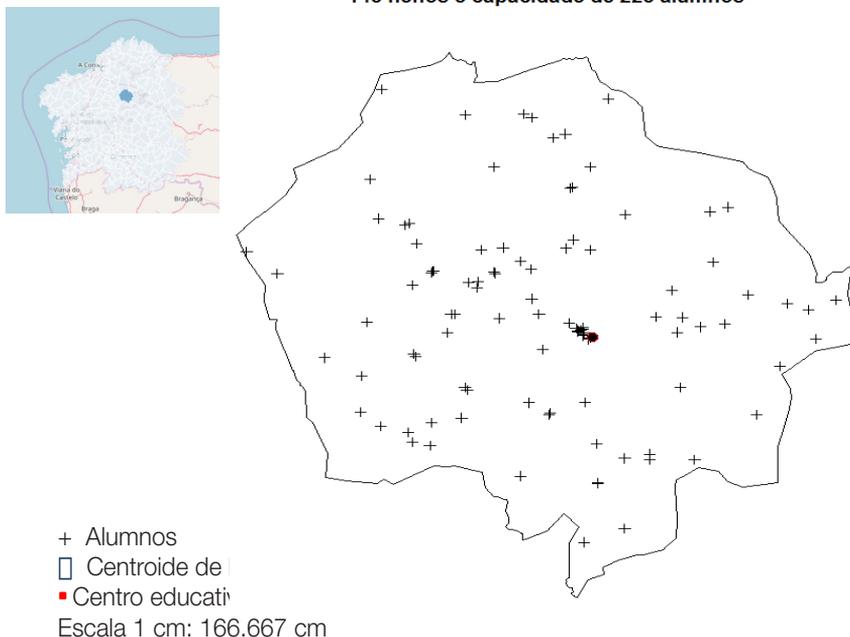
	Mínimo	1º cuartil	Mediana	Promedio	3º Cuartil	Máximo
Distancia (km)	0	5,83	8,74	8,95	11,67	21,47
Tiempo (minutos)	0	9,08	11,42	11,79	14,12	30,15

Fuente: Elaboración propia

A priori puede parecer mucha distancia, por lo que lo vamos a comparar con las distancias que tienen que recorrer los alumnos que viven en un municipio de Galicia con una superficie similar a la de estos cuatro, como puede ser Friol (293,3 km²). El municipio de Friol cuenta con un colegio situado en el centro, en el sentido geográfico, y con capacidad para 225 alumnos. En este municipio en el año 2016 había 145 niños de entre 3 y 11 años con escolarización en infantil y primaria (Figura 10). Las distancias que tienen que recorrer se presentan en el Cuadro 4. Se puede observar que la máxima distancia que recorren es de cerca de 20 km con un promedio de 6 km por niño. El tiempo máximo del recorrido es de 25 minutos y el tiempo promedio es de 7 minutos.

FIGURA 10
MAPA DEL MUNICIPIO DE FRIOL CON EL CENTRO EDUCATIVO (ROJO)
Y LA DISTRIBUCIÓN DE LOS ALUMNOS

145 niños e capacidade de 225 alumnos



Fuente: Xunta de Galicia y Elaboración propia

CUADRO 4
DISTANCIA Y TIEMPO DE RECORRIDO DE LOS NIÑOS DE FRIOL AL
CENTRO EDUCATIVO

	Mínimo	1º cuartil	Mediana	Promedio	3º Cuartil	Máximo
Distancia (km)	0	3,93	6,09	5,94	8,44	19,46
Tiempo (minutos)	0,017	4,44	6,95	7,23	10,37	25,02

Fuente: Elaboración propia

Podemos observar que los tiempos de desplazamiento y las distancias son similares a los del ejemplo anterior, con lo cual parece factible una reagrupación de alumnos sin que estos se vean demasiado perjudicados.

5. CONCLUSIONES

Este artículo presenta una aproximación a la geolocalización de la población gallega aprovechando la información existente en las bases de datos geográficas e integrándolas con bases de datos sociales.

La determinación de la localización de la población sobre el territorio, no directamente relacionada con los límites administrativos, es absolutamente esencial para numerosas cuestiones prácticas de organización social (Goerlich y Cantarino, 2012). Disponer de las unidades estadísticas georreferenciadas permitirá una planificación social, económica o ambiental más fácil de diseñar y de evaluar, así como, una ordenación territorial más adecuada, menos dependiente de la subdivisión administrativa. Procurando que uno de los problemas más acuciantes de Galicia, el territorio, se convierta en una oportunidad.

También facilita la toma de decisiones, no sólo públicas, sino también privadas, tanto con ánimo de lucro como sin él. Esto reducirá costes, porque a medio plazo serán innecesarios muchos de los estudios ad hoc encargados por los agentes públicos, y posiblemente también por los privados, para fundamentar su estrategia. No se pretende sustituir al decisor sino ayudarlo.

La base de datos MDAGE del INE (principalmente) junto con las técnicas de imputación descritas en este trabajo, permiten en este momento la obtención de esta valiosa información. En este punto conviene destacar que las técnicas de imputación utilizadas constituyen una de las aportaciones novedosas de este trabajo. De todos modos, para la obtención de nuevas localizaciones el Idescat (Suñé, 2014) también ha empleado técnicas de imputación, pero en este caso usan la cuadrícula de población combinada con métodos probabilísticos.

La disponibilidad de información georreferenciada de la población tiene diversas aplicaciones, algunas de las cuales se han presentado en este trabajo. Estas aplicaciones y el método que se utilizó para su resolución constituyen la segunda novedad que aporta este trabajo. Otros ejemplos también pueden ser cuestiones relacionadas con la evaluación de daños potenciales, como por ejemplo la determinación de la población en riesgo ante determinadas catástrofes naturales, o bien cuestiones relacionadas con la accesibilidad a determinadas infraestructuras o con el análisis de la efectividad de las políticas públicas.

En cuanto a los resultados que aportan las dos últimas aplicaciones prácticas, es necesario comentar que el criterio de centralidad, que puede ser óptimo o admisible para infraestructuras de nueva planta, no es el único, ya que existe patrimonio construido. Se deben tener en cuenta otros criterios como el económico (costes inducidos por los traslados, ahorro en las remuneraciones de equipos directivos, amortización del inmovilizado,...) o la presencia en el territorio de la administración de la Comunidad Autónoma. Tampoco debe olvidarse en Galicia, la posibilidad de

accidentes geográficos (gran río, ría, paso por zonas de altitud o de gran pendiente) que no aconsejen la circulación durante todo el curso o parte de él. Partir del inmovilizado existente ha sido otro criterio que se ha tenido en cuenta en la sección 4.3 cuando se ha elegido el colegio más próximo al centro de masas.

Por todo lo anterior, se podría destinar el parque de edificios de la consejería que queden libres, a causa de la agrupación de alumnos, a otros asuntos de distinta materia (conservatorios de música o danza, artes y oficios, todas las especialidades de Bachillerato o más especialidades de FP) o a otros de la misma materia (preescolar, educación permanente 'long life learning'...) que lo único que supondrían es un cambio organizativo (toda la enseñanza dependiente de una consejería, y no de varias como actualmente) o, simplemente, una coordinación administrativa de las consejerías que intervienen en la misma materia. O bien, destinar el inmovilizado a otra: residencia de ancianos, centros polivalentes, escuelas musicales... que podría aconsejar ceder el edificio a la administración local.

Si se opta por la solución más fácil de mantener los edificios destinados a la misma materia y dependientes de la misma consejería, podría comenzarse por la coordinación de los centros bajo un equipo único de dirección. Si se optara por la última, obsérvese que en nuestro caso particular no urge destinarlos a residencia de ancianos, como se puede comprobar en el mapa del Gráfico 4, donde se aprecia, incluso, un polígono con una baja ratio de ancianos por plaza.

Finalmente, la metodología presentada es fácilmente replicable, tanto con los datos disponibles en otra comunidad autónoma como con los datos disponibles a nivel nacional o incluso con otros países. Y en cuanto a las extensiones de este trabajo, Goerlich y Cantarino (2010) comentaron que la población residente debe ser el objetivo principal para geolocalizar, aunque no el único. Es necesario extender este estudio y, por ejemplo, cambiar el criterio de localización de la población del lugar de residencia al de trabajo (García-Palomares, 2017), estudio que, aunque complicado por la cantidad de fuentes de información que es necesario emplear, pensamos que se puede acometer en el futuro. Otro aspecto interesante es la incorporación de la perspectiva longitudinal y ofrecer información novedosa sobre las trayectorias vitales de la población.

BIBLIOGRAFÍA

- DILEKLI, N., JANITZ, A. E. AND CAMPBELL, J.E. AND BEURS K.M.. (2018). Evaluation of geoimputation strategies in a large case study. *International Journal of Health Geographics*. <https://doi.org/10.1186/s12942-018-0151-y>
- ENRIQUE REGUEIRA, I. (2016) Official grid-based statistics: a new approach to regional statistics in Andalusia. *European Forum for Geography and Statistics*, Paris
- EUROPEAN FORUM FOR GEOGRAPHY AND STATISTICS (EFGS) (2017). Grid. Recuperado de: <http://www.efgs.info/data/>
- EUROSTAT (2011). The degree of urbanization (DEGURBA). Recuperado de: http://ec.europa.eu/eurostat/ramon/miscellaneous/index.cfm?TargetUrl=DSP_DEGURBA
- GALLEGO, F.J. (2010): A population density grid of the European Union. *Population & Environment*, 31, 460-473
- GARCÍA, J.R., IGLESIAS, C.L e LÓPEZ, M.E. (2017) Coordenadas xeográficas de unidades estatísticas e a súa asignación: un caso típico. *Actas do XIII Congreso Galego de Estatística e Investigación de Operacións*. Ferrol
- GARCÍA-PALOMARES, J.C., SALAS-OLMEDO, M.H., MOYA-GÓMEZ, B., CONDEÇO-MELHORADO, A. M. y GUTIÉRREZ, J. (2017). City dynamics through Twitter: relationships between land use and spatiotemporal demographics. *Cities*
- GOERLICH, F.J. y CANTARINO, I. (2011). Cartografía y Demografía: una grid de población para la Comunitat Valenciana. *Estadística Española*. Vol. 53, Núm. 176, 127-161
- GOERLICH, F.J. y CANTARINO, I. (2012). *Una "grid" de densidad de población para España*. Bilbao, Fundación BBVA
- GOERLICH, F. J., MAS, M (2009). Drivers of agglomeration: Geography versus History. *The Open Urban Studies Journal* 2. Recuperado de: <http://www.bentham.org/open/tousj/openaccess2.htm>
- IGE (2016). Clasificación do grao de urbanización das parroquias e dos concellos galegos (GU 2016). Recuperado de: http://www.ige.eu/estatico/pdfs/s3/clasificacions/urbanizacion/MetodoloxiaGU2016eAnexo_Modificada.pdf
- IECA (2017). Distribución Espacial de la Población en Andalucía. Recuperado de: <http://www.juntadeandalucia.es/institutodeestadisticaycartografia/distribucionpob/index.htm>
- INE (2013). Censos de población y viviendas 2011. Recuperado de: http://www.ine.es/censos2011_datos/cien11_datos_inicio.htm
- INE (2016). Modelo de Direcciones de la Administración general del Estado.
- INE (2016). Padrón Municipal de Habitantes.
- KAHLE, D. and WICKHAM, H. (2013). ggmap: Spatial Visualization with ggplot2. *The R Journal*, 5(1), 144-161. Recuperado de: <http://journal.r-project.org/archive/2013-1/kahle-wickham.pdf>
- LAGONIGRO, R, OLLER, R., MARTORI, J.C. (2017). A quadtree approach based on European geographic grids: reconciling data privacy and accuracy. *SORT* 41 (1) January-June, 139-158.
- MORENO, A. y PRIETO, M.E. (2003). Evaluación de procedimientos para delimitar áreas de servicio de líneas de transporte urbano con sistemas de información geográfica. *Investigaciones Regionales*, 2, pp 85-102
- R CORE TEAM (2016). R: Language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Recuperado de: <https://www.r-project.org/>.
- ROBINSON, D. (2016). fuzzyjoin: Join Tables Together on Inexact Matching. R package version 0.1.2. Recuperado de: <https://CRAN.R-project.org/package=fuzzyjoin>
- REHER, D. (1994). Ciudades, procesos de urbanización y sistemas urbanos en la Península Ibérica 1550-1991. En M. Guardia, F.J. Monclús, J. Oyón (dirs.). *Atlas histórico de ciudades europeas*. Barcelona: Centre de Cultura Contemporànea de Barcelona y Salvat (1994): 1.29.
- RÚA, A., REDONDO, R., DEL CAMPO C. (2003). Distribución municipal de la realidad socioeconómica gallega. *Revista Galega de Economía*, vol 12, núm 2, pp 243-262
- RUPPERT, D., WAND, M.P., CARROL, R.J. (2003). *Semiparametric Regression*. Cambridge University Press, New York.

- SUNÉ, E. (2014). La georreferenciación de la población de Catalunya. VIII Jornadas de SIG Libre. Recuperado de: <http://www.sigte.udg.edu/jornadassiglibre2014/uploads/2014/articulos/art15.pdf>
- WALTER, S.R. and ROSE, N. (2013). Random property allocation: A novel geographic imputation procedure based on a complete geocoded address file. *Spatial an Spatio-temporal Epidemiology*, vol 6, pp. 7-16
- UGARTE, M., GOICOA T., MILITINO., A y DURBÁN, M. (2009). Spline smoothing in small area trend estimation and forecasting. *Computational Statistics and Data Analysis.*, 53: 3616-3629.
- URIEL, E. y ALDÁS, J. (2005). *Análisis multivariante aplicado. Aplicaciones al Marketing, Investigación de Mercados, Economía, Dirección de Empresas y Turismo*. Madrid: Thompson Editores.
- WOOD, S.N. (2006) *Generalized Additive Models: An Introduction with R*. Chapman and Hall/CRC.

