

ANÁLISIS DE REDES Y SENSIBILIDAD A LA UNIDAD MÍNIMA DE INFORMACIÓN POBLACIONAL: SANLÚCAR DE BARRAMEDA (CÁDIZ)

JOAQUÍN MÁRQUEZ PÉREZ¹, JOSÉ OJEDA ZÚJAR² y JOSÉ I. ÁLVAREZ FRANCO³
Departamento de Geografía Física y Análisis Geográfico Regional. Universidad de Sevilla
C/ María de Padilla, s/n. 41004, Sevilla. España
jmarper@us.es, zujar@us.es, jalvarez2@us.es

RESUMEN

Actualmente existen fuentes de información que permiten disponer de datos de población desagregada al máximo nivel espacial. Sin embargo, debido al carácter secreto de los datos estadísticos, su grado de detalle debe evitar que éste sea vulnerado, y para ello es necesario emplear unidades mínimas de información poblacional que lo garanticen. Mediante procesos de agregación espacial, es posible encontrar esas unidades mínimas y realizar análisis complejos, respetando la confidencialidad de los datos.

El objetivo esencial de este trabajo reside en evaluar la sensibilidad que muestran los resultados obtenidos por distintos tipos de agregación espacial de la población a la asignación de la demanda a centros sanitarios, a partir de métodos derivados del análisis de redes. Para ello se ha empleado como fuentes de datos de población el Registro Longitudinal de la Población de Andalucía en 2011, y el Callejero Digital de Andalucía para la construcción de la red topológica, ambas fuentes referidas al núcleo urbano de Sanlúcar de Barrameda (Cádiz).

Palabras clave: unidad mínima poblacional, agregación espacial, análisis de redes, geocodificación, SIG.

NETWORK ANALYSIS AND SENSITIVITY TO MINIMAL POPULATION INFORMATION UNIT: SANLÚCAR DE BARRAMEDA (CÁDIZ)

ABSTRACT

Geocoding processes of postal addresses and, therefore, their linked data (e.g. population) allows the usage of spatially disaggregated population data at maximum level. However, the level of data-disaggregation should not violate statistical confidentiality. Raster and vector-based analysis can be useful to find intermediate representations of spatial population units that avoid disclosure of statistical confidentiality.

The main objective of this paper is to evaluate the sensitivity of network analysis results to the use of different minimum spatial information units, provided by different aggregation methods. Sanlúcar de Barrameda municipality has been chosen as study area. The *Registro Longitudinal de la Población de Andalucía (2011)* is used as population data source and the *Callejero Digital de Andalucía* have been selected to build up the topologic network.

Keywords: minimum population unit, spatial aggregation, network analysis, geocoding, GIS.

1. Introducción

Los procesos de geocodificación de las direcciones postales y, por tanto, de los datos asociados a ellas (la población en este caso) permiten disponer de información poblacional desagregada espacialmente al máximo nivel tomando como fuente de datos más actualizados el Padrón Municipal o registros similares. Desde este punto de vista, el diseño de una red topológica de calles y carreteras precisa y completa, con consistencia geométrica y topológica respecto a los datos georreferenciados, permitiría modelizar la ubicación óptima de servicios públicos (educación o salud) y constituiría un eficiente instrumento de análisis para el diseño, planificación y gestión de los mismos.

Los avances tecnológicos que tuvieron lugar durante las últimas décadas del siglo XX permitieron disponer de datos poblacionales de gran calidad y detalle, pero al mismo tiempo fomentaron el debate sobre el correcto uso de los mismos en cuanto a la preservación de la privacidad de los individuos a los que hacen referencia (Duncan y Pearson, 1991; Cox, 1996; Armstrong *et al.*, 1999).

Desde el punto de vista normativo, es clave el Reglamento del Parlamento Europeo y del Consejo de 11 de marzo de 2009 (DOUE de 31/03/2009) relativo a la estadística europea, dentro del cual pueden destacarse algunas consideraciones: "26) *Los investigadores deben poder disfrutar de un amplio acceso a los datos confidenciales utilizados para el desarrollo, elaboración y difusión de las estadísticas europeas, con el fin de analizarlos en pro del progreso científico europeo. También debe mejorarse el acceso con fines científicos a los datos confidenciales, sin menoscabo del alto nivel de protección que tales datos requieren*".

En el artículo 3 del mismo reglamento se recogen algunas definiciones, entre las que nos interesan especialmente dos: "7) *dato confidencial: dato que permite identificar, directa o indirectamente, a las unidades estadísticas y divulgar, por tanto, información sobre particulares; para determinar si una unidad estadística es identificable, deberán tenerse en cuenta todos los medios pertinentes que razonablemente podría utilizar un tercero para identificar a la unidad estadística, y 9) identificación directa: identificación de una unidad estadística por su nombre o apellidos, su domicilio o un número de identificación públicamente accesible*".

El análisis espacial de la información sobre los individuos, a escalas de detalle, se encuentra así condicionado por la necesidad de respetar la confidencialidad de los datos personales. Esta cuestión ha sido tratada por diferentes autores, principalmente desde la óptica de disciplinas como la

estadística (Duncan y Pearson, 1991), la demografía (Martin, 2002) o la salud pública (Cox, 1996; Armstrong *et al.*, 1996; Curtis *et al.*, 2006; Malin, 2007; Leitner y Curtis, 2007; Kamel Boulos *et al.*, 2009). Cuando los datos son agregados a escalas inferiores, las posibilidades del investigador para analizar y revelar las relaciones espaciales entre los datos se reducen considerablemente. Resulta obvio, como expone Armstrong (1999, p. 498), que "...el análisis geográfico debe investigar las relaciones en áreas que necesitan ser definidas con mucha más flexibilidad de la que permiten los datos agregados al nivel de las actuales entidades político-administrativas". En este sentido, el nivel de detalle del que se dispone actualmente, con agregaciones por secciones censales, resulta insuficiente para un gran número de análisis.

En la revisión bibliográfica pueden encontrarse diversos métodos propuestos para proteger el secreto estadístico, conservando en la medida de lo posible la información original (Cox, 1996; Armstrong *et al.*, 1999). Armstrong propone algunas técnicas de enmascaramiento (*geographical masks*), desde la transformación de las coordenadas del dato (por métodos deterministas o aleatorios) o las técnicas de agregación por área, hasta la sustitución de la información de posición por información contextual de interés para el investigador, recordando que la mejor solución depende tanto del propósito del investigador como del grado de confidencialidad de la información. Otros autores refieren soluciones más sencillas, basadas en la eliminación del dato (Duncan y Pearson, 1991), o en la agregación de la información al punto más cercano (Martin, 2002).

Deben emplearse, por tanto, unidades mínimas de información que permitan espacializar detalladamente las características poblacionales sin vulnerar el secreto estadístico. En el caso que nos ocupa, dado que la vulneración de la confidencialidad se produciría en el caso de revelar la ubicación exacta de una sola persona, la determinación de la unidad mínima de información espacial es clave, y debería oscilar entre aquella que ya está publicada (las secciones censales) y la que permite vulnerar el secreto estadístico (todos los portales). Entre ambos límites se encuentran, básicamente, los edificios, manzanas y parcelas, aunque también es posible la creación de nuevas unidades geométricas como la agrupación de portales cercanos en un mismo polígono (representado por su centroide), y las unidades elementales de igual tamaño y forma (celdillas raster), de resolución espacial acorde con el propósito buscado.

2. Objetivos

El objetivo básico de este trabajo, a diferencia de otros centrados en la desagregación espacial de la población (Moreno Jiménez, 1991; Suárez Vega *et al.*, 2008; Santos Preciado *et al.*, 2011), consiste en comparar distintos modos de agregar espacialmente la información poblacional del núcleo urbano de Sanlúcar de Barrameda con el objetivo no vulnerar el secreto estadístico y, posteriormente, evaluar la sensibilidad de estas agregaciones al análisis de redes orientado a la asignación de la demanda poblacional a determinados servicios generales básicos (salud en este caso), a través de la red vial. Asimismo, como un objetivo complementario, se busca comparar los resultados con los derivados de utilizar la información poblacional mas desagregada espacialmente disponible públicamente en la actualidad, es decir, las secciones censales.

3. Área de estudio

El estudio se centra en el núcleo urbano de Sanlúcar de Barrameda (figura 1), situado en el municipio homónimo de la provincia de Cádiz. Ubicado al noroeste de la provincia, en la costa atlántica y en el margen izquierdo de la desembocadura del río Guadalquivir. El término municipal tiene una extensión superficial de 170,9 Km². El núcleo de Sanlúcar de Barrameda que define el área de estudio proporciona, según el Registro Longitudinal de la Población de Andalucía de 2011, una población total de 66853 habitantes.

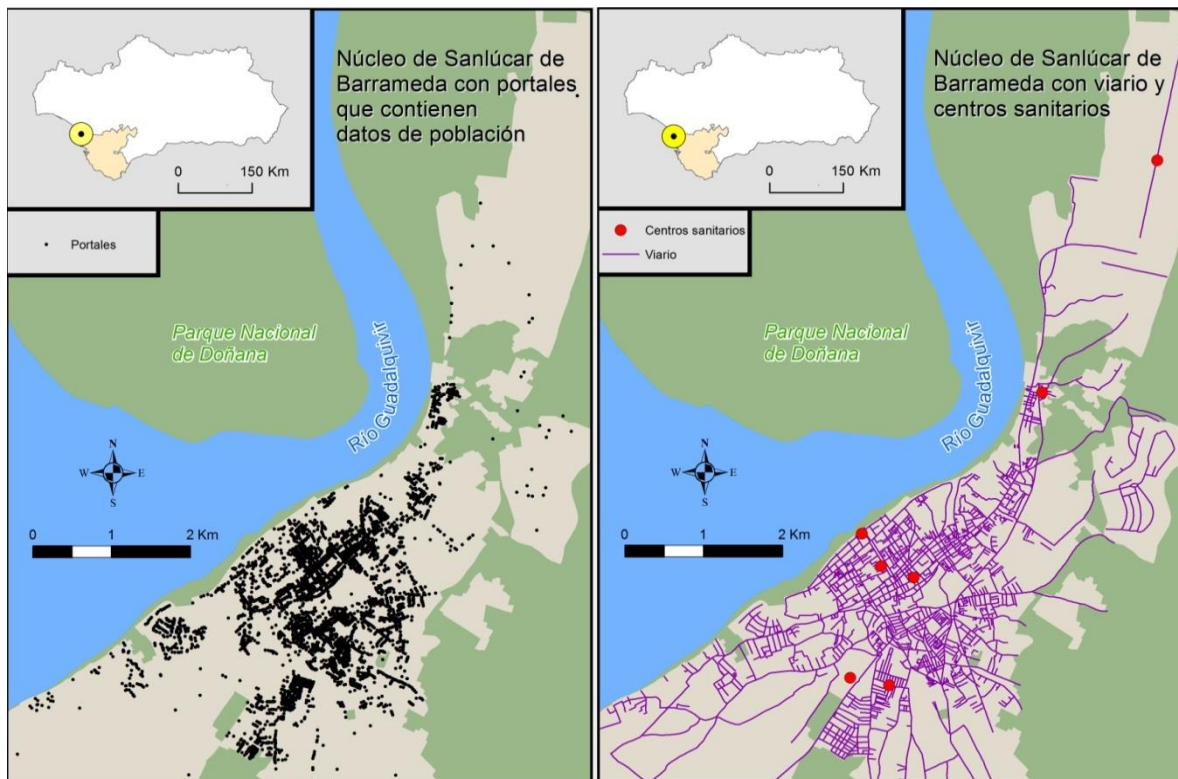


Figura 1. Zona de estudio: núcleo urbano de Sanlúcar de Barrameda. A la izquierda, la fuente de datos de población, los portales. A la derecha, el viario (Callejero Digital de Andalucía) y los centros sanitarios.

4. Las fuentes de información

El método elegido exige la disponibilidad de cuatro categorías de datos georreferenciados en un mismo Sistema de Referencia de Coordenadas (European Datum 1950, proyección UTM zona 30 Norte): (i) los destinados a generar la red de transporte, (ii) aquellos considerados como servicios, (iii) los que constituyen el objeto principal de este artículo, los puntos de demanda, y (iv) los elementos poligonales necesarios para agregar la información procedente de los puntos de demanda: secciones censales y edificios.

4.1. La red de transporte

Los datos con los que es construida la estructura de red topológica sobre la que realizar los análisis proceden del Callejero Digital de Andalucía (figura 1, derecha). Las características de la red son muy simples, pues solo permiten un modo de desplazamiento (pedestre): se trata de un tipo de análisis unimodal, donde el único atributo de coste (impedancia) es la longitud del desplazamiento.

4.2. Los servicios

Los datos utilizados como fuente de información para los servicios de oferta (centros de salud, figura 1, derecha) han sido extraídos de la cartografía vectorial DEA100, cuya localización geométrica se deriva del Mapa Topográfico de Andalucía (MTA10); su error posicional, en función de la escala de representación (1:10.000) y el error cartográfico (0.2 mm), es de 2 metros.

Dado que los datos geocodificados de población tan solo informan acerca del total de habitantes asociado a cada portal, interesa que el tipo de servicio elegido pueda ser empleado indistintamente por todo tipo de personas y, en ese sentido, los centros de salud presentan las características idóneas para la asignación de la demanda. Teniendo en cuenta que el objetivo principal del trabajo consiste en evaluar la sensibilidad del análisis de redes a los procesos de agregación poblacional, a todos los centros les ha sido asignado el mismo peso gravitacional. Los siete centros de salud son los siguientes: Barrio Alto, Barrio Bajo, Algaida, Bonanza, la Dehesilla, Cruz Roja-Protección Civil y Virgen del Camino.

4.3. Los puntos de demanda

Los datos originales utilizados, que constituyen la información acerca de los puntos de demanda (figura 1, izquierda), proceden de la geocodificación de una muestra del Registro Longitudinal de la Población de Andalucía (Viciano *et al.*, 2010) correspondiente a la provincia de Cádiz. Fueron elaborados conjuntamente por investigadores de la Universidad de Sevilla y técnicos del IECA (Díaz Cuevas, P. *et al.*, 2012), y han sido cedidos por este organismo. Para dicho trabajo se han utilizado las herramientas de geocodificación pertenecientes al SIG-Corporativo de la Junta de Andalucía (Zabala *et al.*, 2010).

4.4. Secciones censales y edificios

Las secciones censales proceden de la Cartografía Censal de Andalucía (periodo 2001-2009), información publicada por el Instituto de Estadística y Cartografía de Andalucía (2011). Los edificios, elementos poligonales extraídos de la capa homologa del Mapa Topográfico de Andalucía, escala 1:10.000 (MTA10), han sido elegidos debido a su consistencia geométrica con la red topológica utilizada (en un futuro se utilizarán parcelas del Catastro, pero debido a algunos errores de consistencia geométrica y topológica se ha pospuesto su utilización para futuros análisis). Por otra parte, su amplia distribución espacial los hace muy adecuados para representar los valores

poblacionales asociados a los portales geocodificados, y permite evitar, parcialmente, la vulneración del secreto estadístico.

5. Metodología

El proceso metodológico consta de dos fases: por una parte, la agregación de los datos de demanda, y por otra, la aplicación a esos datos agregados de distintos algoritmos basados en el análisis de redes, con la idea de evaluar su sensibilidad a los procesos de agregación empleados.

5.1. La agregación de los datos de demanda

Para la generación de los datos que contienen la información acerca de la demanda asignada a los servicios sanitarios, y que deben ser de tipo puntual, en este trabajo se han empleado cuatro aproximaciones:

A) Información asociada a portales

En primer lugar, se cuenta con los datos originales del proceso de geocodificación, es decir, la población total asociada a cada portal (figura 2, A). Posteriormente, estos datos son reestructurados para evitar la vulneración del secreto estadístico. Esta reestructuración ha sido llevada a cabo de dos formas diferentes:

La primera (figura 2, A1) consiste en emplear tan solo portales cuya población sea superior a un habitante. Para ello, todos los portales con un solo habitante son asociados al portal más próximo cuyo valor poblacional sea superior a 1. Como consecuencia, no son considerados una parte de los portales iniciales (los que correspondían a un solo habitante), aunque sí es tenida en cuenta su población.

La segunda posibilidad (figura 2, A2), algo más compleja que la anterior, tiene la ventaja de que aumenta el número de portales al no desechar aquellos correspondientes a un solo individuo. La idea es asociar a cada portal caracterizado por un solo habitante el identificador del portal más cercano, tanto si tiene varios habitantes como si tiene solo uno. En el caso de que dos portales con un solo habitante sean los más próximos el uno del otro, uno de ellos permanece como portal en la selección final, asignándosele un valor poblacional de 2.

Márquez Pérez J., Ojeda Zújar, J. y Álvarez, J. I. (2014): "Análisis de redes y sensibilidad a la unidad mínima de información poblacional: Sanlúcar de Barrameda (Cádiz)", *GeoFocus (Artículos)*, nº 14, p. 85-104. ISSN: 1578-5157

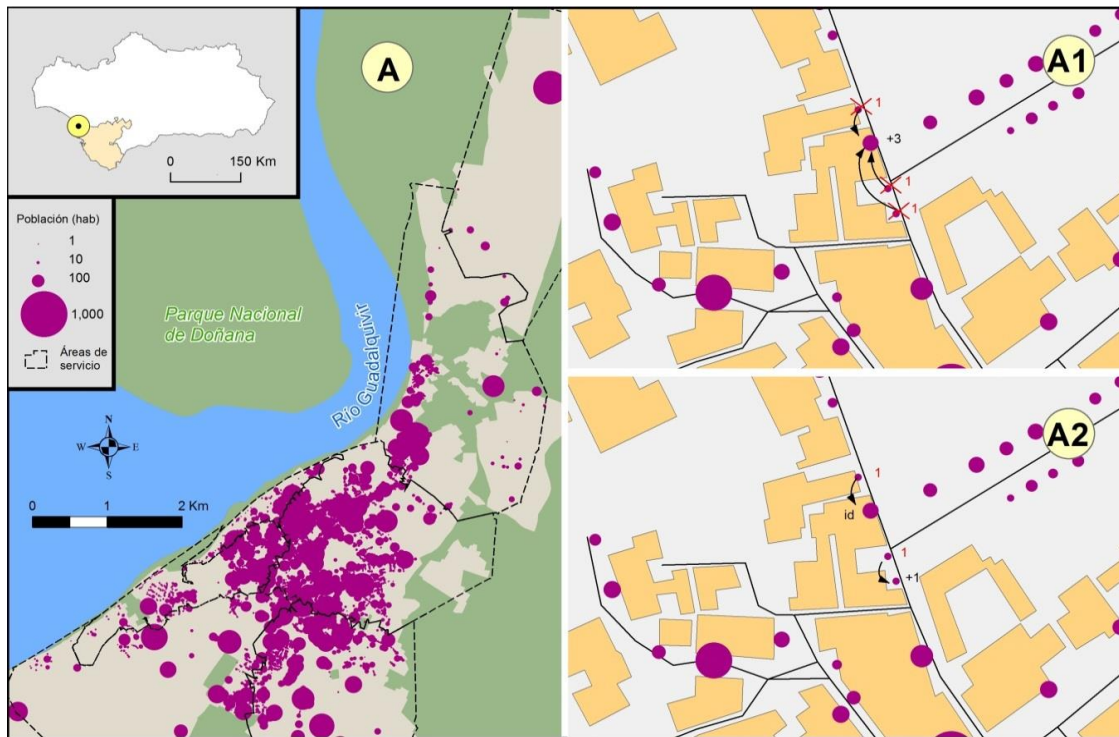


Figura 2. Asociación de la información a portales. A la izquierda, representación proporcional de la población asociada a los portales; A la derecha, reestructuración de los datos para evitar la vulneración del secreto estadístico.

B) Información asociada a edificios

Los datos procedentes de los portales geocodificados se agregan espacialmente a elementos poligonales, los edificios. La asociación de los valores poblacionales de los portales geocodificados a los elementos poligonales requiere un proceso que consta de cuatro fases: i) caracterización de cada portal por el identificador del elemento poligonal más cercano, ii) creación de una tabla resumen, que contenga la suma de los valores de cada registro en función del identificador del elemento, iii) unión de los valores poblacionales obtenidos en cada elemento, a los elementos de la capa poligonal, y iv) creación de centroides que representen el total de los valores poblacionales del elemento poligonal.

Dado que puede ser vulnerado el secreto estadístico en aquellos casos en que un polígono corresponda a un solo habitante (figura 3, B1), la solución pasa por agregar este valor a aquellos polígonos más próximos que posean un valor poblacional superior a 1 (figura 3, B2).

Márquez Pérez J., Ojeda Zújar, J. y Álvarez, J. I. (2014): "Análisis de redes y sensibilidad a la unidad mínima de información poblacional: Sanlúcar de Barrameda (Cádiz)", *GeoFocus (Artículos)*, nº 14, p. 85-104. ISSN: 1578-5157



Figura 3. Asociación de la información a edificios. A la izquierda, representación proporcional de la población a los edificios. A la derecha, reestructuración de los datos para evitar la vulneración del secreto estadístico.

C) Información asociada a celdillas raster

Los datos procedentes de los portales geocodificados son asociados al punto central de celdillas raster. Mediante la generación de una estructura donde las celdillas pasen a contener el valor poblacional de la suma de la población asociada a los portales ubicados en su interior, es posible obtener una distribución espacial de la población. Inicialmente, mediante un análisis iterativo (usando la aplicación *Modelbuilder*, incluida en *ArcGIS 10*) se busca la resolución espacial que evite que en una celdilla coincida un solo punto geocodificado que tenga un valor igual a 1. Posteriormente, la conversión de raster a vector devuelve los datos poblacionales al formato puntual para el análisis de redes.

Realizado el proceso, se comprueba que el tamaño mínimo de celdilla necesario para que el valor mínimo poblacional sea de 2 habitantes es de 450 metros (figura 4, C1). Igualmente, para la evaluación se han utilizado los resultados de celdillas de 250 y 100 metros de resolución espacial (figura 4, C2 y C3). En estos casos, cuando una celdilla arroja un valor poblacional de 1, ésta pasa a tener valor 0 y su valor se asocia al de la celdilla más próxima con valor superior a 1.

Márquez Pérez J., Ojeda Zújar, J. y Álvarez, J. I. (2014): "Análisis de redes y sensibilidad a la unidad mínima de información poblacional: Sanlúcar de Barrameda (Cádiz)", *GeoFocus (Artículos)*, nº 14, p. 85-104. ISSN: 1578-5157

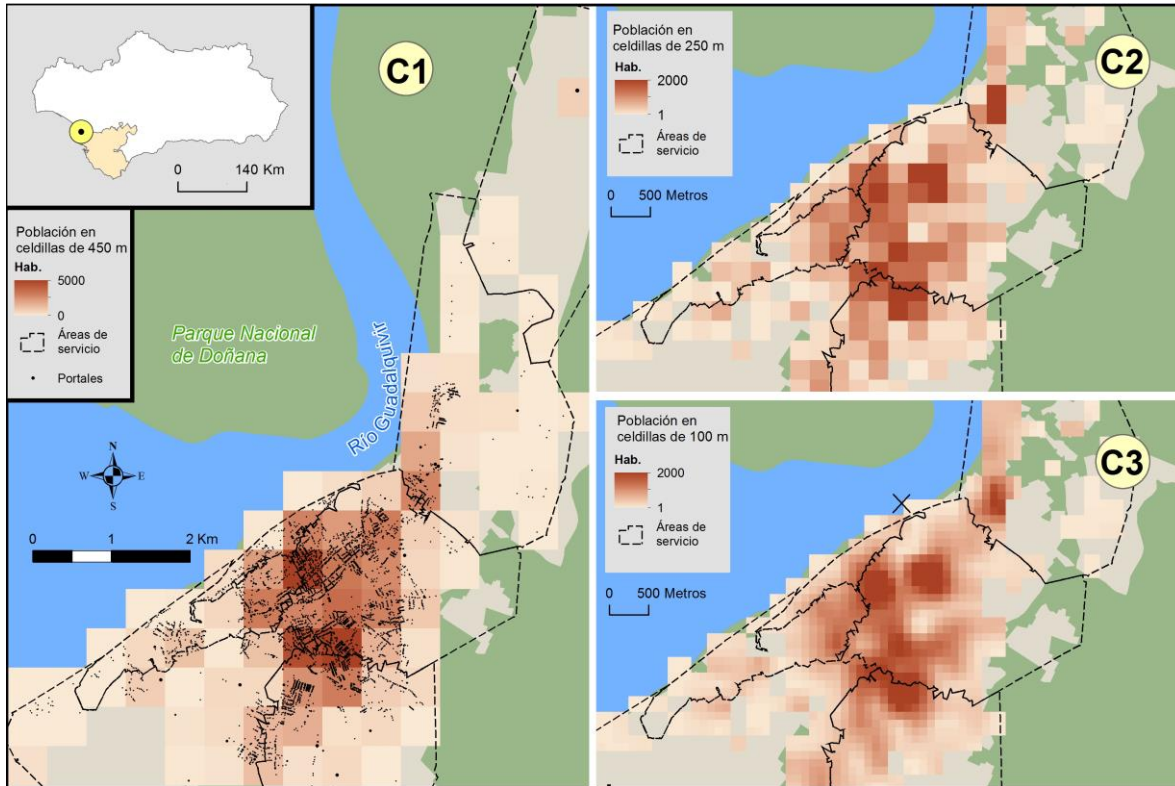


Figura 4. Asociación de la información a celdillas raster de 450 (C1), 250 (C2) y 100 metros (C3) de resolución espacial.

D) Información asociada a secciones censales

Por último, como análisis de contraste, se utilizan los datos pertenecientes a las secciones censales representadas por polígonos, tanto asociados en su totalidad a sus centroides (figura 5, D1), como distribuidos proporcionalmente (figura 5, D2) entre aquellos portales donde el proceso de geocodificación asigna elementos poblacionales. La asignación de los valores censales a los centroides es un método rápido de caracterizar la distribución poblacional de las secciones censales, aunque su escaso número dificulta la obtención de resultados adecuados.

Márquez Pérez J., Ojeda Zújar, J. y Álvarez, J. I. (2014): "Análisis de redes y sensibilidad a la unidad mínima de información poblacional: Sanlúcar de Barrameda (Cádiz)", *GeoFocus (Artículos)*, nº 14, p. 85-104. ISSN: 1578-5157

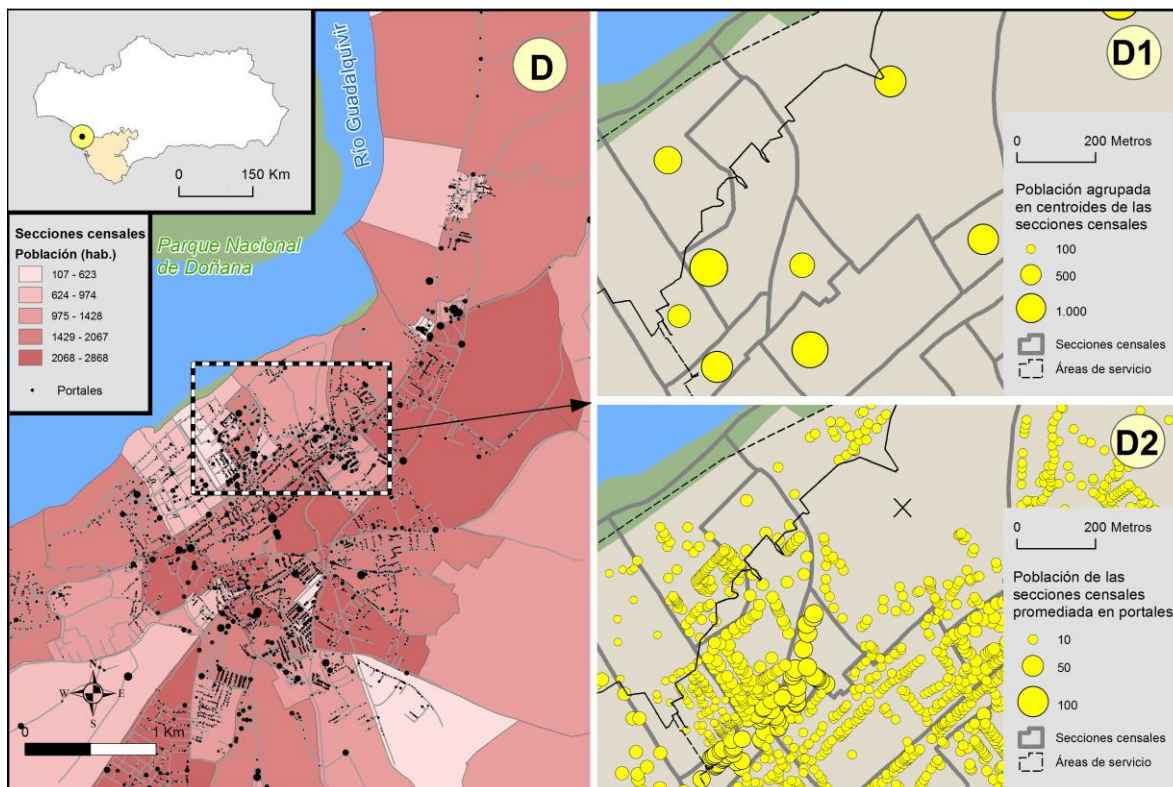


Figura 5. Asociación de la información a secciones censales. A la izquierda, representación proporcional en los portales y representación por coropletas de la población de las secciones censales; A la derecha, representación proporcional de la agrupación de la población en los centroides (D1) de las secciones y promediada en los portales (D2).

La asignación de los datos promediados a cada portal geocodificado, incluido en el polígono censal que lo contenga, permite trabajar con datos mejor distribuidos y mucho mejor localizados. El procedimiento de asignación requiere la caracterización de cada polígono por el número de portales positivos que contenga, la división del valor poblacional por este número, y la posterior reasignación del valor resultante a cada portal. La distribución de la información poblacional asociada a las secciones censales se ha abordado, usando métodos más o menos sofisticados, en diferentes publicaciones (Tobler, 1979, Santos Preciado y García Lázaro, 2003, Suárez Vega *et al.*, 2008, Santos Preciado *et al.*, 2011).

5.2. La aplicación del análisis de redes

El método de análisis supone el empleo del programa *ArcGIS Desktop*, cuya versión 10 incluye, dentro de la extensión destinada al análisis de redes (*Network Analyst*), un módulo que permite investigar la localización óptima de servicios, y la asignación de la demanda a esos servicios a través de una red (*Location-Allocation*).

Dentro de la extensión *Network Analyst* se ha utilizado la herramienta "Service Area", que permite generar polígonos (generalizados) que muestran el área de influencia de cada servicio utilizando la estructura topológica de la red (véase figura 6). Delimitando las zonas de influencia de cada centro sanitario, puede observarse en qué zona se ubica cada unidad mínima de información de la información poblacional, lo cual resulta especialmente útil para la planificación y gestión administrativa de estos servicios públicos (por ejemplo, la planificación de servicios sanitarios).

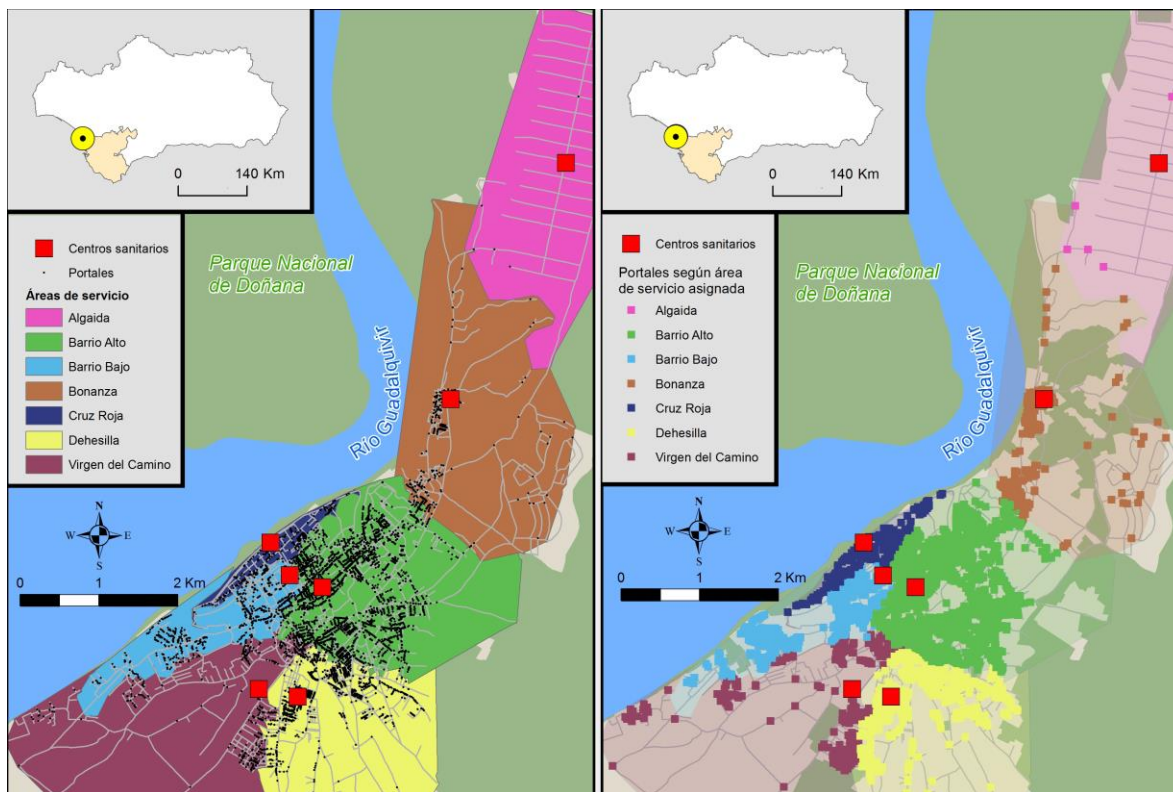


Figura 6. Áreas de servicio. A la izquierda, los polígonos que definen las áreas de servicio; A la derecha, los portales caracterizados según el área de servicio asignada a cada uno.

Para resolver el problema de la asignación de la demanda a los puntos de servicio se ha empleado el módulo "Location-Allocation" que, al incluir el algoritmo "Minimize Impedance", facilita la asignación de la demanda a esa tipología de servicios (educación u hospitales) que requieren que lo sea del modo más eficiente: cada punto de demanda es asociado al servicio más cercano, de modo que la suma de las impedancias sea mínima, e incluso puede no ser seleccionado alguno de estos servicios si su utilización supone un valor de impedancia superior. Si no se establece un tiempo máximo de desplazamiento, toda la demanda es satisfecha, y se considera un método de distribución equitativo para la demanda.

Los valores así obtenidos a partir de la población original geocodificada, son indicadores del número de habitantes que se desplazan al centro de salud más cercano, y constituyen la referencia necesaria para comparar los datos originales desagregados al máximo nivel, con las distintas opciones de datos agregados. Esta comparación permite evaluar la sensibilidad de los

métodos de agregación espacial aplicados, en los cuales se evita en cualquier caso la vulneración del secreto estadístico.

Los resultados de la asignación de la demanda a cada servicio son aquí exclusivamente utilizados como estimadores de la sensibilidad de los procesos de agregación espacial, a diferencia de otros estudios centrados en la localización óptima de servicios (Densham y Rushton, 1992, Bosque Sendra *et al.*, 2000, Moreno Jiménez y Buzai, 2008, Fuenzalida Díaz y Moreno Jiménez, 2010).

6. Resultados

Los resultados pueden desglosarse en función de la metodología empleada para la agregación espacial de los datos de población geocodificados, una vez corregidos para garantizar el secreto estadístico: los portales originales, los edificios, las celdillas raster, y las secciones censales.

6.1. Datos asociados a los portales originales

La incorporación de los datos de demanda, vinculados a los datos originales asociados a cada portal procedente de la geocodificación, ofrece los siguientes resultados en relación a la asignación de población a cada centro de salud (tabla 1).

Tabla 1. Resultados de la agregación en portales (A)

Centro Salud	Población adscrita (A)	Población adscrita (A1)	Variación (%)	Población adscrita (A2)	Variación (%)
Barrio Alto	29253	29254	0,00	29255	0,01
Barrio Bajo	7102	7102	0,00	7101	-0,01
Algaida	707	707	0,00	707	0,00
Bonanza	6475	6474	-0,02	6477	0,03
La Dehesilla	14064	14062	-0,01	14062	-0,01
Cruz Roja y P. Civil	2862	2863	0,03	2861	-0,03
Virgen del Camino	6390	6391	0,02	6390	0,00
Promedio			0.0114		0.0128

El análisis realizado a partir de todos los portales geocodificados con población adscrita (A) es el considerado óptimo al tener la máxima desagregación espacial, aunque viola el secreto estadístico al existir al menos 522 portales que solo registran un habitante. Sin embargo, sirven de referencia cuantitativa para evaluar la sensibilidad del resto de los métodos de agregación espacial de la población.

Las correcciones realizadas en las opciones A1 y A2 presentan resultados excelentes, muy escasa diferencia entre ellos y, en general, mínima sensibilidad cuantitativa y espacial al proceso de agregación.

6.2. Datos asociados a los edificios

Con los datos originales, agregados a los polígonos de los edificios, se viola el secreto estadístico en 71 de ellos, al ser asignado un solo habitante. La tabla 2 proporciona los resultados originales (todos los centroides de edificios, con al menos un habitante (B1)) y los corregidos (solo centroides de edificios con más de un habitante (B2)), ambos con porcentajes de variación muy aceptables.

Tabla 2. Resultados de la agregación en edificios (B)

Centro Salud	Población adscrita (A)	Población adscrita (B1)	Variación (%)	Población adscrita (B2)	Variación (%)
Barrio Alto	29253	29181	-0,25	29.180	-0,25
Barrio Bajo	7102	7400	4,20	7.403	4,24
Algaida	707	707	0,00	707	0,00
Bonanza	6475	6325	-2,32	6.325	-2,32
La Dehesilla	14064	13880	-1,31	13.881	-1,30
Cruz Roja y P. Civil	2862	2668	-6,78	2.666	-6,85
Virgen del Camino	6390	6692	4,73	6.691	4,71

6.3. Datos asociados a celdillas raster

La utilización de una estructura ráster con una resolución espacial que evite el secreto estadístico comporta un tamaño mínimo de celdilla relativamente amplio, de 450 metros (C1) y presenta resultados (véase tabla 3) con infravaloraciones de la demanda superiores al 10% y sobrevaloraciones que superan el 30%. Al aumentar la resolución espacial a 250 metros (C2) se encuentran solo 3 celdillas que violan el secreto estadístico; tras el proceso de agregación, los resultados son más aceptables, y solo aparece alguna sobrevaloración superior al 20%, por otra parte asociada a uno de los centros con menor demanda. Finalmente, con una resolución espacial de 100 metros (C3), se incrementa en 9 el número de celdillas que violan el secreto pero, tras el proceso de agregación, se consigue mejorar sustancialmente los resultados, con solo algunos porcentajes de variación significativos, y siempre por debajo del 10%.

Tabla 3. Resultados de la agregación en celdillas ráster (C)

Centro Salud	Población adscrita (A)	Población adscrita (C1)	Variación (%)	Población adscrita (C2)	Variación (%)	Población adscrita (C3)	Variación (%)
Barrio Alto	29253	29554	1,03	28782	-1,61	28693	-1,91
Barrio Bajo	7102	5705	-19,67	6711	-5,51	7420	4,48
Algaida	707	732	3,54	687	-2,83	707	0,00
Bonanza	6475	6032	-6,84	6335	-2,16	6522	0,73
Dehesilla	14064	12248	-12,91	14620	3,95	14004	-0,43
Cruz Roja	2862	4319	50,91	3532	23,41	2671	-6,67
V. Camino	6390	8263	29,31	6186	-3,19	6836	6,98

6.4. Datos asociados a secciones censales

La utilización de la población asociada a las secciones censales (las únicas actualmente accesibles) presenta modestos resultados por diferentes razones (localización geométrica de los centroides, adecuada cobertura espacial de la red, etc.) lo que da lugar a importantes valores de porcentaje de variación respecto a la demanda asignada (D1, tabla 4). Esta es la razón por la que se han propuesto diferentes formas de desagregación espacial sobre este tipo de estructura espacial de la población (Santos Preciado *et al.*, 2003). Una propuesta alternativa, en el caso de disponer de un callejero con portales, es utilizar éstos para la desagregación en el interior de la sección (D2), máxime si, como en este caso, se conocen los realmente habitados. Como puede observarse en la tabla 4, su utilización reduce sustancialmente los porcentajes de variación que, aunque presentan algunos valores altos de infravaloración, se asocian a los centros con menor demanda, por lo que su impacto en valores absolutos decrece.

Tabla 4. Resultados de la agregación en secciones censales (D)

Centro Salud	Población adscrita (A)	Población adscrita (D1)	Variación (%)	Población adscrita (D2)	Variación (%)
Barrio Alto	29253	5087	-82,61	31359	7,20
Barrio Bajo	7102	28051	294,97	6159	-13,28
Algaida	707	0	-100,00	192	-72,89
Bonanza	6475	6514	0,60	6548	1,13
La Dehesilla	14064	12013	-14,58	13514	-3,91
Cruz Roja y P. Civil	2862	3752	31,10	2990	4,49
Virgen del Camino	6390	7412	15,99	6091	-4,67

6.5. Síntesis de resultados

Los resultados (tabla 5) reflejan que la información poblacional asociada a portales no altera prácticamente la asignación poblacional al aplicar las correcciones del secreto estadístico en las dos opciones propuestas. Los resultados del uso de los edificios poligonales son también aceptables, al igual que las agregaciones raster (excepto con celdillas de 450 metros). Obviamente, en este último caso, su calidad aumenta al incrementarse la resolución espacial, lo que exige correcciones posteriores para eliminar el secreto estadístico. A diferencia de las propuestas anteriores, el uso de secciones censales introduce grandes alteraciones en los resultados aun cuando, si se dispone de la geometría de los portales geocodificados, su sola desagregación promediando el valor poblacional de la sección incrementa notablemente su calidad.

Tabla 5. Resumen de la población adscrita a cada área de servicio según el método de agregación. En rojo, las infravaloraciones superiores al 10 %; en azul, las sobrevaloraciones superiores al 10 %; a – variación absoluta; r – variación relativa.

	B. Alto	B. Bajo	Algaida	Bonanza	Dehesilla	Cruz roja	V. Camino
A	29253	7102	707	6475	14064	2862	6390
A1	29254 a: 1 r: 0%	7102 a: 0 r: 0%	707 a: 0 r: 0%	6474 a: -1 r: -0,02%	14062 a: -2 r: -0,01%	2863 a: 1 r: 0,03%	6391 a: 1 r: 0,02%
A2	29255 a: 2 r: 0,01%	7101 a: -1 r: -0,01%	707 a: 0 r: 0%	6477 a: 2 r: 0,03%	14062 a: -2 r: -0,01%	2861 a: -1 r: -0,03%	6390 a: 0 r: 0%
B1	29181 a: -72 r: -0,25%	7400 a: 298 r: 4,2%	707 a: 0 r: 0%	6325 a: -150 r: -2,32%	13880 a: -184 r: -1,31%	2668 a: -194 r: -6,78%	6692 a: 302 r: 4,73%
B2	29180 a: -73 r: -0,25%	7403 a: 301 r: 4,24%	707 a: 0 r: 0%	6325 a: -150 r: -2,32%	13881 a: -183 r: -1,30%	2666 a: -196 r: -6,85%	6691 a: 301 r: 4,71%
C1	29554 a: 301 r: 1,03%	5705 a: -1397 r: -19,67%	732 a: 25 r: 3,54%	6032 a: -443 r: -6,84%	12248 a: -1816 r: -12,91%	4319 a: 1457 r: 50,91%	8263 a: 1873 r: 29,31%
C2	28782 a: -471 r: -1,61%	6711 a: -391 r: -5,51%	687 a: -20 r: -2,83%	6335 a: -140 r: -2,16%	14620 a: 556 r: 3,95%	3532 a: 670 r: 23,41%	6186 a: -204 r: -3,19%
C3	28693 a: -560 r: -1,91%	7420 a: 318 r: 4,48%	707 a: 0 r: 0%	6522 a: 47 r: 0,73%	14004 a: -60 r: -0,43%	2671 a: -191 r: -6,67%	6836 a: 446 r: 6,98%
D1	28051 a: -1202 r: -4,11%	6087 a: -1015 r: -14,29%	0 a: -707 r: -100%	6514 a: 39 r: 0,6%	12013 a: -2051 r: -14,58%	3752 a: 890 r: 31,1%	7412 a: 1022 r: 15,99%
D2	31359 a: 2106 r: 7,2%	6159 a: -943 r: -13,28%	192 a: -515 r: -72,89%	6548 a: 73 r: 1,13%	13514 a: -550 r: -3,91%	2990 a: 128 r: 4,49%	6091 a: -299 r: -4,67%

Un resumen estadístico de los resultados de la tabla 5 permite contemplar con mayor claridad la bondad de los distintos métodos de agregación (tabla 6). El que aporta mejores resultados respecto al promedio y desviación estándar de las diferencias absolutas y porcentuales es el A1 y, sin embargo, aquellos que obtienen un menor coeficiente de variación (CV) son: D1 (diferencias absolutas) y B1-B2 (diferencias porcentuales), pues ambos poseen una desviación relativamente baja en relación con su media.

Tabla 6. Estadísticos de la variación absoluta y porcentual alcanzada por cada método de agregación. a – variación absoluta; r – variación relativa. En verde, valores óptimos.

	A1	A2	B1	B2	C1	C2	C3	D1	D2
Promedio a	0,86	1,14	171,43	172,00	1044,57	350,29	231,71	989,43	756,83
Desv. Std. a	0,64	0,83	102,44	102,71	710,13	218,91	200,03	556,36	653,35
C.V. a	0,74	0,72	0,59	0,59	0,67	0,62	0,86	0,56	0,86
Promedio r	0,01	0,03	2,80	2,81	17,74	6,09	3,03	25,81	17,74
Desv. Std. r	0,01	0,03	2,33	2,35	16,30	7,16	2,76	31,61	24,87
C.V. r	0,98	1,26	0,84	0,84	0,91	1,17	0,91	1,22	1,40

Por otra parte, los métodos de agregación, en conjunto, afectan de forma desigual a las áreas de influencia de los centros sanitarios: B. Alto y Bonanza son los que tienen un menor coeficiente de variación (inferior al 0.03) mientras que Cruz Roja y, especialmente, La Algaída, alcanzan valores superiores a 0.1 (tabla 7).

Tabla 7. Estadísticos poblacionales resultantes de los métodos de agregación, agrupados en función del área de influencia de los servicios. En verde, valores óptimos.

	B. Alto	B. Bajo	Algaída	Bonanza	Dehesilla	Cruz Roja	V. Camino
Media	29256,56	6787,56	571,78	6394,67	13587,11	3146,89	6772,44
Des. Std.	850,91	615,96	258,53	153,42	825,98	554,01	645,32
C.V.	0,029	0,091	0,452	0,024	0,061	0,176	0,095

Considerando que la asignación se lleva a cabo a partir de elementos puntuales, independientemente de sus valores poblacionales (pero siempre mayores a 1), resultan de interés las siguientes tablas, que muestran tanto el número de puntos empleados por cada método (tabla 8) como sus estadísticos (tabla 9). Sus resultados indican que el método A2 es aquel que proporciona un número de elementos puntuales asignado a cada servicio, más próximo a aquel derivado de la utilización de los datos reales; esto permite explicar por qué este método (A2), más preciso que el A1, da lugar a información algo menos relevante: el número de puntos seleccionados por A2 es igual o superior al empleado por A1 y, sin embargo, el promedio de su variación absoluta y porcentual respecto a los datos poblacionales originales es levemente superior, y ello se debe a la falta de control sobre los valores poblacionales asignados a los elementos puntuales. No debe considerarse, en este sentido, los valores alcanzados por D2, pues son idénticos a los originales.

Tabla 8. Porcentaje de puntos empleados por cada método y zona de influencia, respecto al número de portales originales asignados a cada Servicio. En verde, valores óptimos.

	B. Alto	B. Bajo	Algaida	Bonanza	Dehesilla	Cruz roja	V. Camino
A	9,39	7,79	0,71	5,13	7,07	8,98	6,96
A1	8,48	6,66	0,71	4,59	6,57	7,76	6,42
A2	8,57	6,76	0,71	4,65	6,59	7,86	6,45
B1	2,66	3,24	0,71	2,69	2,62	3,63	2,97
B2	2,60	3,01	0,71	2,46	2,57	3,35	2,85
C1	0,06	0,11	0,71	0,32	0,15	0,21	0,31
C2	0,19	0,28	0,57	0,59	0,34	0,52	0,59
C3	0,90	1,37	0,71	1,14	1,01	1,78	1,50
D1	0,06	0,04	0,00	0,08	0,06	0,10	0,08
D2	9,39	7,79	0,71	5,13	7,07	8,98	6,96

Tabla 9: Porcentaje de puntos empleado por cada método respecto al nº de portales iniciales. En verde, valores óptimos.

	A1	A2	B1	B2	C1	C2	C3	D1	D2
% de puntos	90,21	91,04	34,66	33,33	1,87	4,12	13,67	0,81	100,00

7. Discusión y conclusiones

Cualquier proceso de agregación de datos originales conlleva una disminución en la calidad de la información que aportan. Sin embargo, esta pérdida puede llegar a ser muy leve o casi inexistente dependiendo tanto de la distribución espacial y los valores poblacionales de los mismos datos, como del sistema de agregación empleado. Es el caso de los datos poblacionales que nos ocupan, asignados por geocodificación a portales y que, como resultado de distintos procesos de agregación con objeto de evitar vulnerar el secreto estadístico, permiten obtener resultados que, en muchos casos, denotan una mínima sensibilidad a la transformación efectuada.

El proceso de asignación de la población a centros de salud, utilizando dos algoritmos complementarios basados en la utilización de una red de comunicaciones (asignación de población a centros basada en la minimización de la impedancia, y generación del polígono que delimita el "área de servicio" de esos centros), proporciona las variables complementarias necesarias (demanda asignada y área de servicio, respectivamente) para la estimación de la sensibilidad de cada método de agregación propuesto, a partir de la comparación de los resultados obtenidos con los derivados del empleo de los datos geocodificados originales.

De los métodos de agregación planteados, los basados en la utilización de la localización original de los portales dan mejor resultado que aquellos asociados a la generación de puntos equiespaciados procedentes de la utilización de celdillas raster, si bien éstos proporcionan una ventaja adicional al posibilitar la difuminación posicional del dato, en función de la resolución espacial empleada, contribuyendo con ello a garantizar la confidencialidad del mismo.

Sin embargo, debe considerarse que el análisis se realiza a partir de elementos puntuales, caracterizados por información poblacional; a pesar de que la bondad de cada método se mida a partir de valores poblacionales asignados a servicios, son esos elementos puntuales los que portan esa información, por lo que si el volumen de población asignada a cada uno es desigual, el resultado final tendrá un margen de incertidumbre que dependerá tanto de la distribución espacial de los elementos puntuales, como de las diferencias entre sus valores poblacionales.

En cualquier caso, en este trabajo se presentan diferentes posibilidades de publicar información poblacional, con el secreto estadístico garantizado, derivadas de las diferentes opciones de agregación planteadas, con un nivel de desagregación espacial muy superior al actualmente disponible (secciones censales).

La escala de análisis podría influir parcialmente en los resultados obtenidos por los métodos de agregación propuestos, pues especialmente aquellos basados en celdillas tendrían más dificultad en representar fidedignamente población espacialmente más dispersa, como puede comprobarse si se observan los resultados asociados a las áreas de influencia de los distintos servicios; sin embargo, los basados en portales o edificios no tendrían por qué verse afectados, pues tantos unos como otros presentarían una distribución coherente respecto a los datos poblacionales originales.

Los resultados de este trabajo muestran que la disponibilidad de información poblacional georreferenciada a nivel del portal, junto a información de espacios construidos a escalas de detalle, consistentes tanto geométrica como topológicamente (básicamente, edificios), posibilita la utilización de diferentes métodos de agregación de la información poblacional original que, a su vez, permitiría su difusión por la administración sin vulnerar el secreto estadístico. En el caso de que no estuviese disponible la información asociada a los espacios, siempre sería posible el empleo del modelo de datos raster. Sin duda, la difusión de esta información proporcionaría a los investigadores y planificadores una variable espacializada del máximo interés tanto para la gestión y planificación de servicios públicos, como para numerosas actividades económicas.

Referencias bibliográficas

Armstrong, M.P., Rushton, G. y Zimmerman, D.L. (1999): "Geographically masking health data to preserve confidentiality", *Statistics in Medicine*, vol. 18, pp. 497-525.

Biang, L. y Butler, R. (1999): "Comparing effects of aggregation methods on statistical and spatial properties of simulated spatial data", *Photogrammetric Engineering & Remote Sensing*, 65, vol.1, pp. 73-84.

Bosque Sendra, J., Gómez Delgado, M., Moreno Jiménez, A. y Dal Pozzo, F. (2000): "Hacia un sistema de ayuda la decisión espacial para la localización de equipamientos", *Estudios Geográficos*, vol. 61, 241, pp. 567-598.

Cox, L.H. (1996): "Protecting confidentiality in small population health and environmental statistic", *Statistics in Medicine*, vol.15, pp. 1895-1905.

Márquez Pérez J., Ojeda Zújar, J. y Álvarez, J. I. (2014): "Análisis de redes y sensibilidad a la unidad mínima de información poblacional: Sanlúcar de Barrameda (Cádiz)", *GeoFocus (Artículos)*, nº 14, p. 85-104. ISSN: 1578-5157

Curtis, A.J., Mills, J.W., Leitner, M. (2006): "Spatial confidentiality and GIS: re-engineering mortality locations from published maps about Hurricane Katrina", *International Journal of Health Geographics*, vol. 5, pp. 44.

Densham, P.J. y Rusthon, G. (1992): "Strategies for Solving Large Location-Allocation Problems by Heuristic Methods", *Environment and Planning A*, vol. 24, pp. 289-304.

Díaz Cuevas, P., Vallejo Villalta, I. y Ojeda Zújar, J. (2012): "Espacialización de datos poblacionales de la provincia de Cádiz a escala de detalle", en Martínez Vega, J. y Martín Isabel, P. (Ed.): *Proceedings del XV Congreso Nacional de Tecnologías de la Información Geográfica*. Madrid, AGE-CSIC, pp. 279-287.

DOUE, Diario Oficial de la Unión Europea (2009): *Reglamento (CE) nº 223/2009 del Parlamento Europeo y del Consejo de 11 de marzo de 2009, relativo a la estadística europea y por el que se deroga el Reglamento (CE, EURATOM,) nº 1101/2008, relativo a la transmisión a la Oficina Estadística de las Comunidades Europeas de las informaciones amparadas por el secreto estadístico, el Reglamento (CE), nº 322/97 del Consejo sobre la estadística comunitaria y la Decisión 89/382/CEE, Euratom del Consejo por la que se crea un Comité del programa estadístico de las Comunidades Europeas*. Oficina de Publicaciones Oficiales de las Comunidades Europeas. © Unión Europea, <http://eur-lex.europa.eu/>.

Duncan, G.T. y Pearson, R.W. (1991): "Enhancing access to microdata while protecting confidentiality: Prospects for the future", *Statistical Science*, vol. 6, 3, pp. 219-232.

Fuenzalida Díaz, M. y Moreno Jiménez, A. (2010): "Diseño con SIG de la localización óptima de centros de atención primaria de salud, discriminando según estatus socioeconómico", en Ojeda, J., Pita, M.F. y Vallejo, I., (Ed.): *Tecnologías de la Información Geográfica: La Información Geográfica al servicio de los ciudadanos*. Sevilla, Secretariado de Publicaciones de la Universidad de Sevilla, pp. 453-465.

Kamel Boulos, M. N., Cai, Q., Padget, J.A. y Rushton, G. (2006): "Using software agents to preserve individual health data confidentiality in micro-scale geographical analyses", *Journal of Biomedical Informatics*, vol. 39, pp. 160-170.

Kamel Boulos, M.N., Curtis, A.J. y Abdelmalik, P. (2009): "Musings on privacy issues in health research involving disaggregate geographic data about individuals", *International Journal of Health Geographics*, vol. 8, pp. 46.

Leitner, M. y Curtis, A. (2006): "A first step towards a framework for presenting the location of confidential point data on maps: results of an empirical perceptual study", *International Journal of Geographical Information Science*, vol. 20, 7, 06, pp. 813-822.

Malin, B. (2007): "A computational model to protect patient data from location-based re-identification", *Artificial Intelligence in Medicine*, vol. 40, pp. 223-239.

Martin, D. (2002): "Geography for the 2001 Census in England and Wales", *Population Trends*, vol. 108, pp. 7-15.

Moreno Jiménez, A. (1991): "Modelización cartográfica de densidades mediante estimadores Kernel", *Treballs de la Societat Catalana de Geografia*, vol. 6, 30, pp. 155-170.

Márquez Pérez J., Ojeda Zújar, J. y Álvarez, J. I. (2014): "Análisis de redes y sensibilidad a la unidad mínima de información poblacional: Sanlúcar de Barrameda (Cádiz)", *GeoFocus (Artículos)*, nº 14, p. 85-104. ISSN: 1578-5157

Moreno Jiménez, A. y Buzai, G.D. (Coord.) (2008): *Análisis y planificación de servicios colectivos con sistemas de información geográfica*. Madrid, Universidad Autónoma de Madrid y Universidad Nacional de Luján, 1ª edición, 158 p.

Moreno Muñoz, J.A. (2011): "La base de referencia para la geocodificación: el Callejero Digital de Andalucía". *Mapping*, vol. 148, pp. 40-43.

Santos Preciado, J. M. y García Lázaro, F. J. (2003): "El método dasimétrico-picnofiláctico: un procedimiento para la desagregación de datos censales", *IX Conferencia Iberoamericana de SIG*, Universidad de Extremadura, CD-ROM, pp. 1-16.

Santos Preciado, J. M., Azcárate Luxán, M. V., Cocero Matesanz, D., García Lázaro, F. J. y Muguruza Cañas, C. (2011): "Los procedimientos de desagregación espacial de la población y su aplicación al análisis del modelo de la ciudad dispersa. El caso de las aglomeraciones urbanas de Madrid y Granada", *GeoFocus*, vol. 11, pp. 91-117.

Suárez Vega, R., Santos Peñate, D. R., y Dorta González, P. (2008): "Generación de un modelo superficial de la población de Gran Canaria", en Hernández Calvento, L. y Parreño Castellano, J. M. (Ed.): *Tecnologías de la información geográfica para el desarrollo territorial*. Las Palmas de Gran Canaria, Servicio de Publicaciones y Difusión Científica de la ULPGC, pp. 183-193.

Tobler, W. R. (1979): "Smooth pycnophylactic interpolation for geographical regions", *Journal of the American Statistical Association*, vol. 74, pp. 519-530.

Viciano Fernández, F., Montañés Cobo, V., Cánovas Balboa, M. R. y Poza Cruz, E. (2010): "Base de Datos Longitudinal de Población de Andalucía (BDLPA): Modelo de datos y sistema de gestión", *Jornadas de Estadísticas de las CCAA*, Cáceres. <http://www.jecas.org>.

Zabala Ordóñez, A., Guerrero de Mier, C., y Mañas Álvarez, B. (2010): "SIGC: Hacia una arquitectura orientada a servicio basada en software libre para los SIG de la Junta de Andalucía", en Ojeda, J. Pita, M.F. y Vallejo, I. (Ed.): *Tecnologías de la Información Geográfica: La Información Geográfica al servicio de los ciudadanos*. Sevilla, Universidad de Sevilla, pp. 615-623.

Agradecimientos

El presente trabajo se enmarca en el proyecto de excelencia de la Junta de Andalucía (RNM-6207) y en el correspondiente al Plan Nacional (CSO2010-15807): "Espacialización y difusión web de variables demográficas, turísticas y ambientales para la evaluación de la vulnerabilidad asociada a la erosión de playas en la costa andaluza (incidencia de la potencial subida del nivel del mar asociada al cambio climático)".

Los autores agradecen al Instituto de Estadística y Cartografía de Andalucía la colaboración prestada.