

IMPORTANCIA DE LOS ESPACIOS COMUNES: UNA ADAPTACIÓN DE LA TÉCNICA DE INTERPOLACIÓN ESPACIAL INVERSE DISTANCE WEIGHTED (IDW) EN LA PREDICCIÓN DE DATOS SOCIOECONÓMICOS AUSENTES

Mónica Navarrete Álvarez¹
Ana María López²

RESUMEN

Disponer de estadísticas socioeconómicas microterritoriales no es frecuente a pesar de los esfuerzos de instituciones oficiales de generarlas, particularmente en el caso de lugares de difícil acceso y/o despoblados. En este documento, las autoras recurren al uso de técnicas de interpolación espacial para predecir microdatos de variables socioeconómicas ausentes. En particular, se utiliza la técnica del Inverso de la Distancia (IDW), conocida por su facilidad de implementación e interpretación. El IDW estima el valor de una variable desconocida a través de la ponderación inversa de la distancia a valores cercanos conocidos. Tal asunción se basa en la ley geográfica de Tobler de que “todo está relacionado con todo, pero las cosas cercanas están más relacionadas entre sí que las distantes”. Para mejorar la estimación de la predicción, se propone una nueva especificación de IDW compuesta por una función gravitatoria, en que la distancia inversa (variable de fricción) se combina específicamente con la longitud relativa de la frontera común entre dos unidades espaciales (variable de atracción). Probamos el funcionamiento de esta función para predecir los ingresos disponibles en algunos municipios metropolitanos en Madrid (España). La eficacia de las predicciones espaciales se evalúa mediante el error medio cuadrático (EMC) por el método de la validación cruzada, el error producto de las diferencias entre los valores reales e interpolados y el coeficiente de correlación entre los valores reales y estimados.

Palabras clave: interpolación espacial, Inverse Distance Weighted, renta de los hogares

1. INTRODUCCIÓN

La ausencia de variables socioeconómicas suele ser un problema frecuente que la econometría aborda mediante técnicas de predicción. Tal problema se agudiza aun más cuando los datos corresponden a unidades microterritoriales, pese a los esfuerzos que realizan las instituciones oficiales de recogida de datos para ofrecer, bases de datos completas y actualizadas. Puede ocurrir entonces, que a nivel provincial algunos municipios dispongan de datos en la escala y nivel territorial adecuado a los propósitos del analista, mientras que en otros, lo anterior se plantea como un problema. Para Bennett et al (1984), la ausencia de datos puede ocurrir bajo dos modalidades: la *falta de información* y la *inexistencia de datos*. La primera ocurre producto del truncamiento o agregación de los datos recogidos a una escala diferente a la requerida por el analista mientras que en la segunda los datos están disponibles en la forma correcta para los propósitos de un análisis, pero algunos de ellos no están

¹ Doctora en Modelización Económica Aplicada. Universidad de Tarapacá-Chile, e-mail: mnavarre@uta.cl

² Doctora en Ciencias Económicas y Empresariales Universidad Autónoma de Madrid, e-mail: ana.lopez@uam.es

disponibles o simplemente no han sido obtenidos. Desde la óptica temporal³, la estimación de datos ausentes se puede suplir mediante una estimación simple de interpolación lineal. En el caso de desagregación a una frecuencia distinta, las propuestas metodológicas Boot, Feibes y Lisman, las variables proxy de Chow-Lin y la construcción de indicadores sintéticos, son herramientas recurrentemente utilizadas en econometría. Sin embargo, desde la óptica espacial⁴, la estimación de datos ausentes debe considerar explícitamente el carácter georreferenciado y multidireccional que distingue a los datos espaciales⁵. Es en este punto, donde la presente investigación propone la técnica de interpolación inversa de la distancia (IDW) proveniente de otras áreas del conocimiento para predecir microdatos de variables socioeconómicas.

Si la predicción es entendida como “una afirmación (probabilística) acerca de un punto muestral no observado” ([Haavelmo 1950](#)), las técnicas de interpolación espacial, utilizadas para estimar valores ausentes por la geografía para el cartografiado, en estimación de variables climatológicas o en estudios geológicos, por ejemplo; pueden ser consideradas como métodos de predicción espacial, según lo planteado en [Chasco \(2003\)](#). En esta obra, la autora propone la predicción de datos de ámbito municipal a partir de datos procedentes de un nivel agregado provincial, mediante un ejercicio de extrapolación espacial. En el mismo, la obtención de los valores ausentes se lleva a cabo mediante una regresión espacial, en la que los datos provinciales (variable dependiente) se explican a partir de unas variables exógenas cuyos valores estén también disponibles para las localizaciones carentes de dato. Así, de la misma forma que la econometría clásica se apoya en indicadores sintéticos como ejercicio que ponderan la correlación entre variables disponibles en una frecuencia (anual, semestral, trimestral) para estimar el valor de la variable en otra frecuencia (no disponible), en este artículo se propone una alternativa a la predicción de datos espaciales inexistentes a partir de valores conocidos de esa variable en lugares cercanos (predicción muestral-no muestral), mediante la técnica de interpolación espacial. En concreto, nos centramos en el método *Inverse Distance Weighted* (IDW), que estima el valor de una variable desconocida a través de la ponderación inversa de la distancia a los valores cercanos conocidos, incorporando en su proceso de cálculo, la sugerencia de Cliff y Ord consistente en la utilización de una combinación de medidas de distancia y longitud relativa de la frontera que tienen en común las unidades espaciales. A modo de ilustración, se aplica la técnica IDW para estimar la renta per cápita de dos municipios del Área Metropolitana de la Comunidad de Madrid mediante la técnica de interpolación IDW y se compara los resultados con los obtenidos mediante la sugerencia de Cliff y Ord.

El artículo presenta inicialmente la técnica de interpolación espacial IDW. A continuación, se presenta un ejercicio de predicción espacial de la variable renta personal disponible para 41 municipios que no disponen de renta per cápita y por último, se presenta un apartado de conclusiones y las referencias.

2. TÉCNICA DE INTERPOLACIÓN IDW

La cartografía proporciona uno de los métodos más directos de estimación de valores ausentes mediante la interpolación espacial de datos (Bennett, et al 1984). El relieve y la topografía del terreno se han representado en la cartografía mediante algoritmos matemáticos que permiten la representación de superficies continuas a partir de datos puntuales, donde aquellos datos faltantes son interpolados

³ Los datos se requieren para una serie de tiempo T pero están disponibles para solamente un subconjunto $t < T$ de instantes de muestreo.

⁴ Los datos se requieren para N sitios pero están disponibles solamente para un subconjunto $n < N$ de sitios.

⁵ Los datos espaciales se caracterizan por su naturaleza georreferenciada, es decir, que su posición relativa o absoluta sobre el espacio contiene información valiosa para interpretar las relaciones consideradas, naturaleza multidireccional en las relaciones que sobre él se establecen (Chasco, C. 2003).

espacialmente. La interpolación espacial es un procedimiento que, “*dado una serie de datos espaciales, sea en forma de puntos o sub-áreas, busca la función que mejor representará la superficie total y que predecirá los valores en otras unidades espaciales*” (Lam, N. 1983). Bosque, J. (1997), le atribuye a la interpolación espacial, la función de “rellenar la información que falta en aquellas áreas en las que no están disponibles las observaciones del mundo real”.

Existe una gran variedad de algoritmos de interpolación y la elección de uno u otro depende de las características de los datos, el grado de precisión y del proceso de cálculo requerido. Se pueden encontrar revisiones muy completas sobre las técnicas de interpolación y su clasificación en Franke, (1982), Lam, N. (1983), Wingle, W. (1992), Burrough y Mc Donnell (1998), Nalder y Ross (1998) Echegoyen, T. (2002), Coley y Clabburn (2005). Detrás de cada técnica de interpolación está implícita la primera ley geográfica de [Tobler \(1979\)](#), según la cual “todo está relacionado con todo, pero las cosas cercanas están más relacionadas entre sí que las más lejanas”.

Dentro de las técnicas de interpolación, una de las más utilizadas por su sencillez, es la de “medias móviles con ponderación por la inversa de la distancia”, técnica exacta⁶, local⁷ y determinística⁸ propuesta en 1968 por Donal Shepard para interpolar un modelo de estimación de superficies con datos geológicos (Shepard, 1968). La técnica se basa en una función gravitacional de distancia (Gordon y Wixon, 1978), donde las observaciones más cercanas tienen más peso en la determinación del valor interpolado al punto desconocido. Este interpolador es conocido como Inverse Distance Weighted (IDW) o también método de Shepard (Franke, 1982; Coman, 1998; Shiou y Lin 2002), aunque “*Pelto, et al, dan el crédito del método cuadrático de la distancia inversa a McLain*” (Franke, 1982, pp. 186)

El IDW supone que las diferencias entre los datos espaciales dependen solo de la distancia entre dos localizaciones. Así, los valores de la variable que estén más cerca de la ubicación sujeta a predicción, tendrán mayor ponderación o influencia en el valor de predicción que aquéllos que están más lejos. En la Figura 1, se plantea la posibilidad de estimar el valor desconocido de una variable $Z(j)$ a partir del valor de la misma variable en localizaciones cercanas, separadas a una distancia d .

Para estimar el valor de la variable en un punto no muestral $Z(j)$, el IDW utiliza la siguiente ecuación:

$$Z(j) = \sum_{i=1}^N \lambda_i \cdot Z(i) \quad (1)$$

donde λ_i (con valores entre 0 y 1) es la función de ponderación de los N puntos muestrales i que intervienen en el cálculo de la distancia al punto no muestral $Z(j)$ a partir de los puntos conocidos $Z(i)$ que van a ser ponderados según la distancia (d_{ij}) que los separa (ecuación Nº 2).

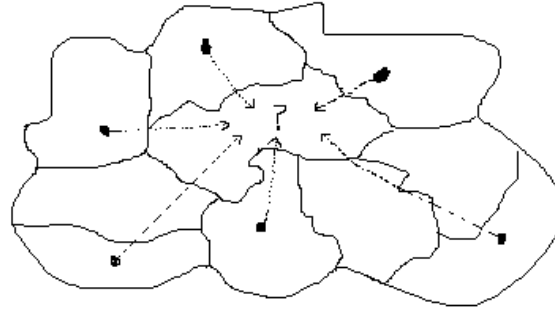
$$\lambda_i = (d_{ij}^{-p}) / \sum_{i=1}^N d_{ij}^{-p} \quad (2)$$

⁶ Los interpoladores exactos son aquellos en que tienen en cuenta todos los puntos para los cuales existen observaciones disponibles. Son técnicas especialmente adecuadas para aquellos casos en los que hay un alto grado de certeza en las mediciones.

⁷ Los interpoladores locales aplican el algoritmo a una pequeña parte del total de puntos muestreados.

⁸ Los interpoladores determinísticos realizan una suposición general, generalmente sin contrastar, acerca del carácter de la superficie a interpolar.

Interpolación de un dato espacial a partir de datos de unidades vecinas



Fuente: Elaboración propia.

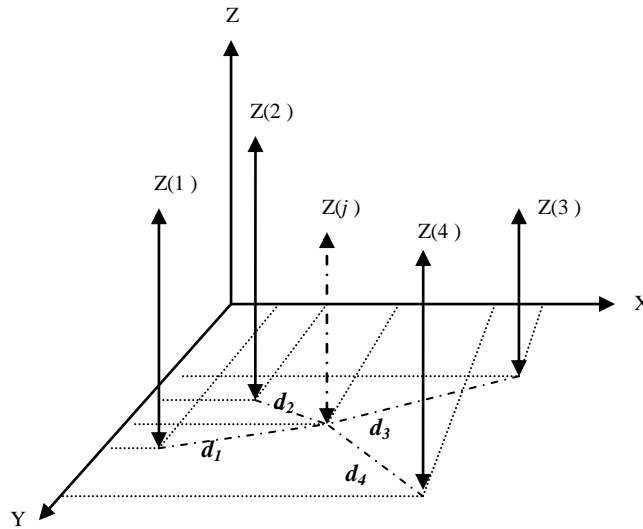
El uso de la función inversa de la distancia implica dar menos peso relativo a los puntos más alejados de j , y viceversa. El peso o ponderación del valor del atributo en los puntos vecinos λ_i , se atenúa en función de la distancia según un exponente p que indica la rapidez con que decrece el peso de un punto dado. Algunos autores establecen que el exponente 2 suele ofrecer unos resultados empíricos más satisfactorios, además de requerir un menor esfuerzo computacional (Declercq, 1996). El uso de exponentes elevados implica dar menos peso relativo a puntos más alejados del valor a interpolar así cuanto mas alta es esta potencia mas influyen los puntos próximos y menos los alejados. Un valor bajo del exponente limita el predominio de cada punto muestral a un corto radio de influencia, dando lugar a curvas concéntricas alrededor de los puntos a interpolar parecidas a afilados embudos y agujas conocidas como efecto ojos de buey o “bulls eyes” (Grinstead et al 2006).

La distancia de cada punto muestral en la interpolación al punto no muestral, se obtiene mediante la distancia euclídea a partir de las coordenadas geográficas (x,y) de longitud y latitud, respectivamente (véase ecuación nº3).

$$d(ij) = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \quad (3)$$

Los datos espaciales requeridos en el proceso interpolador, tienen la forma $(x_i, y_i, Z(i))$, siendo Z la variable observada en los puntos i con localización espacial (x_i, y_i) cercana al punto no muestral $Z(j)$. En la Figura 2, se presenta esta situación en un espacio tridimensional. En los ejes X, Y , se representan las coordenadas geográficas, mientras que el eje Z representa los valores de la variable. En concreto, se desea interpolar el valor de la variable Z en el punto j a partir de 4 valores muestrales de Z disponibles (1, 2, 3, 4).

Figura 1. Interpolación de un dato espacial a partir de datos de unidades vecinas



Fuente: Elaboración propia.

Ahora bien, la elección de los valores de p y N durante el proceso de interpolación se obtiene mediante el método de validación cruzada, en función del error cuadrático medio (ECM), recomendado por varios autores como la mejor medida de la bondad del ajuste (Yang y Doler, 2000). El ECM es la raíz cuadrada de la diferencia media entre el valor estimado del atributo medido y su valor real al cuadrado (ecuación N° 4).

$$ECM = \sqrt{\frac{\sum_{i=1}^N (\hat{Z}(i) - Z(i))^2}{N}} \quad (4)$$

Siendo $\hat{Z}(i)$ el valor esperado o predicho por el interpolador para el punto i , y $Z(i)$ el valor real de la variable medida. El resultado de la interpolación será más satisfactorio mientras menor ECM se obtenga.

Alternativamente al error cuadrático medio (ECM) utilizado para comparar los valores estimados y reales (en este caso la renta per cápita para cada municipio), otra medida para examinar en qué medida el resultado de la interpolación arroja cifras discrepantes de los valores reales, se obtiene el error medio absoluto mediante sustracción simple entre estos dos valores, utilizando la siguiente ecuación N° 5.

$$EM = \frac{\sum_{i=1}^N (Re_i - Rr_i)}{N} \quad (5)$$

Siendo EM = error medio absoluto, Re_i = renta estimada por interpolación para i por el método de validación cruzada para el municipio i , y Rr_i = renta real del municipio i .

Respecto a la elección del número de los valores vecinos (N) una alternativa sería incluir todos los datos muestrales disponibles. Sin embargo, el incremento en costo y tiempo del manejo de un gran volumen

de datos resulta innecesario cuando la distancia se corrige con un exponente (p). Así, de incluir demasiados puntos; la superficie creada tendería a ser muy alisada, asignando finalmente valor medio para todo el área. Si el número de puntos es muy reducido, el valor a interpolar estará muy influido por los valores extremos en la superficie interpolada. Burrough (1986) sugiere que el número debe fluctuar entre 4 y 12 puntos de información, siendo lo normal utilizar entre 6 y 8. El resultado de la comparación de técnicas interpolantes por Dowdall y O’Dea (1999) sugieren un número de vecinos entre 6 y 12 e indican que esta recomendación es consistente con otros investigadores. Declercq (1996) establece que el número de vecinos depende de la característica de la superficie. Esto es, $4 \leq N \leq 8$ para superficies alisadas y $16 \leq N \leq 24$ para superficies muy abruptas. Por otro lado, el concepto de cercanía de los datos espaciales, se puede abordar mediante la construcción de una matriz de contigüidad (ecuación Nº 6) o de pesos espaciales en la que se establecen las relaciones espaciales entre cada par de localizaciones. Cada elemento de la matriz W refleja la intensidad de la interdependencia existente entre cada par de localizaciones i y j como consecuencia de su posición en el mapa (Larraz, B. 2004). Moran, P. (1948) y Geary, G. (1954), utilizan un criterio de contigüidad física de primer orden en que w_{ij} será unitario si las localizaciones i y j son físicamente adyacentes y cero en caso contrario.

$$\mathbf{W} = \begin{pmatrix} 0 & w_{12} & K & w_{1n} \\ w_{21} & 0 & K & w_{2n} \\ M & M & O & M \\ w_{n1} & w_{n2} & K & 0 \end{pmatrix} \quad (6)$$

El IDW es una técnica de interpolación especialmente adecuada para el caso de fenómenos que se propagan en el espacio a partir de un punto, siendo apropiada a efectos de esta investigación, en la estimación de la Renta per cápita, “que se construye bajo la hipótesis de efectos desbordamiento; esto es, lo que ocurre en una región, aún suponiendo que se corresponda con la pertinente, no solo tiene consecuencias en su interior, sino que genera influencias sobre las restantes regiones” (Mur y Angulo, 2006). Las debilidades adjudicadas a esta técnica son asociadas a la sensibilidad al tamaño del vecindario, sensibilidad a las distribuciones anisotrópicas⁹ de los puntos muestrales (Watson y Phillip 1985), dependencia del valor de p y la sensibilidad a la aparición de clusters.

Pese a sus debilidades, el IDW goza de una utilización bastante generalizada (Gemmer et al 2004, Isaaks y Sirastava 1989, Morillo et al, 2002, Mulugeta 1996, Aguilar et al 2001, Lozano 2003, Wolf-Dieter Rase, 2001). En Hutchinson et al (2000) así como en Nalder y Wein (1998) al interpolador IDW se le agrega una variante que asume autocorrelación espacial y en Tomczak (1998), se agrega es el valor un parámetro de alisado δ e indican que cuando los datos son isotrópicos, d_{ij} se calcula como la distancia euclidiana. Técnicas similares al IDW, derivadas de la geoestadística conocidas como Kriging, dan solución a todos esos problemas, debido a que reconocen que la variación espacial de cualquier atributo, sin embargo, en contraste con esta técnicas de interpolación, el IDW requiere solo la estimación de la superficie a interpolar¹⁰ lo cual influye en la sencillez de cálculo, la simplicidad del proceso, la velocidad de la ejecución y los resultados obtenidos.

⁹Característica de un proceso aleatorio que presenta altas autocorrelaciones en una dirección frente a las demás.

¹⁰ El interpolador “kriging” o krigeage (Chica 1994, Montero 2004, Larraz 2004, entre otros) considera en su estimación la variabilidad espacial de los datos siendo necesario la obtención del cálculo del variograma, ajuste del modelo y de la superficie a interpolar.

3. ESTIMACIÓN DE LA RENTA DISPONIBLE EN MUNICIPIOS DEL ÁREA METROPOLITANA DE MADRID

En algunas ocasiones, los datos municipales sólo se encuentran disponibles para un determinado colectivo de municipios, según su dimensión poblacional. Así, por ejemplo, la renta bruta disponible de los hogares es calculada por algunos institutos de estadística regionales para un cierto grupo de localidades: el Instituto de Estadística de Cataluña ([Idescat, 2006](#)) elabora los datos para los municipios catalanes de más de 5.000 habitantes y capitales comarcales, el Instituto Aragón de Estadística ([laest, 2006](#)), lo hace para los municipios aragoneses de más de 3.000 habitantes y capitales de comarca.

3.1. Planteamiento:

En esta línea, planteamos utilizar el método IDW para estimar los datos de renta bruta disponible de los hogares por habitante de algunos municipios del Área Metropolitana de Madrid. Se trata del grupo de 41 municipios que forman parte de las siguientes NUTS4 o comarcas de la Comunidad de Madrid¹¹: Municipio de Madrid, Este Metropolitano, Norte Metropolitano, Oeste Metropolitano y Sur Metropolitano. El Instituto de Estadística de la Comunidad de Madrid ha publicado la serie de renta bruta disponible municipal para 2004, como dato de avance. Aunque esta serie se encuentra completa para todos los municipios de esta comunidad autónoma, aquí supondremos, con fines ilustrativos; que únicamente se encuentran disponibles los datos correspondientes al colectivo municipal de 10.000 y más habitantes (39 municipios, según el Padrón de 1 de enero de 2004).

En el grupo de municipios metropolitanos, 6 de ellos tienen una población inferior a 10.000 habitantes, por lo que careceríamos del dato de renta para ellos. Todos se encuentran algo más alejados del centro de la Comunidad y muy próximos, en unos casos, a municipios con valores más elevados de renta per cápita y, en otros, a municipios con un nivel de renta similar. Se trata de los siguientes: Ajalvir, Brunete, Cobeña, Hoyo de Manzanares, Paracuellos del Jarama y Velilla de San Antonio (véase Figura Nº 3), de los cuales expondremos la técnica en dos de ellos: Hoyo de Manzanares y Velilla de San Antonio.

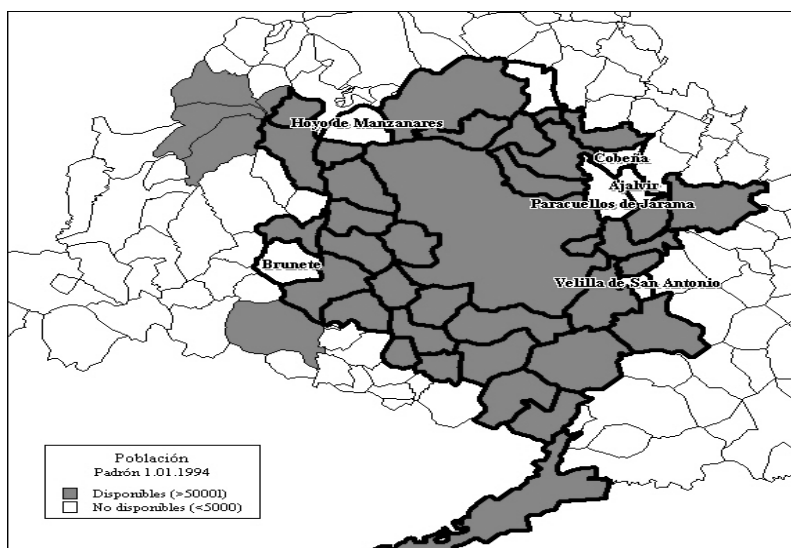
3.2. Objetivo:

Predicción de la renta bruta disponible de los hogares para los municipios Hoyo de Manzanares y Velilla de San Antonio con población en el padrón inferior a 10.000 habitantes (datos extra-muestrales), a partir de la información muestral disponible (39 municipios con población igual o superior a 10.000 habitantes).

La cercanía entre los municipios considera el criterio de contigüidad física de primer y segundo orden dado el número reducido de municipios en primer orden.

¹¹ El Instituto de Estadística de la Comunidad de Madrid ha elaborado una zonificación estadística NUTS4, como desagregación territorial superior a la municipal e inferior a la provincia. La información completa puede localizarse en <http://www.madrid.org/iestadis>.

Figura 2. Mapa de los municipios a interpolar



Fuente: Elaboración propia.

3.3. Datos municipales:

En la Tabla N° 1 se dispone de los datos de la renta per cápita para los municipios vecinos en primer y segundo orden, y las coordenadas geográficas (x,y) del centroide en cada municipio para la distancia¹² a los puntos interpolados.

Tabla N°1: Municipios considerados

Municipio	Coordenadas X	Coordenadas Y	Renta bruta disponible municipal año 2004*
Hoyo de Manzanares	425,004443	4496,64704	16156
Torrelones	421,368801	4491,24103	19.753
Colmenar viejo	437,959433	4499,12325	10.520
Galapagar	416,694101	4491,72365	15.070
Madrid	441,796657	4481,80583	14.671
Rozas de Madrid (Las)	423,769212	4487,18415	20.631
Collado Villalba	416,109172	4500,14371	12.443
Escorial (El)	407,153656	4492,0583	14.124
Tres cantos	437,299319	4497,47626	18.176
Velilla de San Antonio	459,27469	4468,85709	11922
Rivas-Vaciamadrid	455,874729	4466,61531	14.242
Arganda	462,296163	4461,15953	10.520
San Martín de la Vega	452,816047	4456,00344	9.964
Mejorada del Campo	459,445773	4472,0621	10.224

¹² La medición espacial de la distancia puede considerar varias alternativas, como la distancias en línea recta, distancias por vías de acceso, tiempo de trayecto, coste económico, entre otros (Moreno, A y Prieto, M^a, 2002)

Madrid	441,796657	4481,80583	14.671
Getafe	443,881604	4461,30985	11.311
San Fernando	459,46311	4476,71894	11.144

*Fuente: Instituto de Estadística de la Comunidad de Madrid

3.4. Interpolación por IDW:

Para cada uno de estos municipios, se calcula la distancia (en coordenadas x,y) a uno de los dos municipios a interpolar siguiendo la ecuación dada en (3). Consideraremos para el cálculo del valor a predecir, todos los municipios disponibles en I y II orden de vecindad, así; el IDW estima el valor desconocido de la renta per cápita otorgando un mayor peso a los municipios más cercanos. El IDW establece como restricción que los ponderadores deben ser tal que:

$$\sum_{i=1}^n \lambda_i = 1 \quad (7)$$

Los

valores de la tabla Nº 2 y tabla Nº3 se obtienen con un valor de p=2.

Tabla Nº 2: Estimación de ponderador, renta y medidas de errores

Municipio	d_{ij}	$\lambda_i = (d_{ij}^{-p})$	$\lambda_i = (d_{ij}^{-p}) / \sum_{n=1}^N d_{ij}^{-p}$	$Z(s_j) = \sum_{i=1}^{10} \lambda_i * Z(s_i)$
Torrelones	6,515	0,023561	0,32067	6334,250
Colmenar viejo	13,190	0,005748	0,07824	823,004
Galapagar	9,659	0,010718	0,14587	2198,289
Madrid	22,411	0,001991	0,02710	397,574
Rozas de Madrid (Las)	9,543	0,010980	0,14944	3083,174
Collado Villalba	9,558	0,010947	0,14899	1853,826
Escorial (El)	18,431	0,002944	0,04006	565,853
Tres cantos	12,323	0,006585	0,08963	1629,076
suma		0,0734	1	16885,05
Valor estimado renta municipal Hoyo de Manzanares				16885,05
ECM				4437,12
EM				590

De los municipios considerados en la predicción de Hoyo de Manzanares, un 32% es aportado por el valor de la renta del municipio de Torrelodones mientras que los municipios de Las Rozas de Madrid, Collado Villalba y Galapagar aportan cada uno un 14%, dejando en último lugar de aporte; al municipio de Madrid (2,7%).

La tabla Nº2 muestra por su parte que el municipio de Mejorada del campo y Rivas-Vaciamadrid explican casi el 80% del valor de renta que se estima para Velilla de San Antonio.

Tabla Nº 2: Estimación de ponderador, renta y medidas de errores

Municipio	d_{ij}	$\lambda_i = (d_{ij}^{-p})$	$\lambda_i = (d_{ij}^{-p}) / \sum_{n=1}^N d_{ij}^{-p}$	$Z(s_j) = \sum_{i=1}^{10} \lambda_i * Z(s_i)$
Rivas-Vaciamadrid	16,585	0,0603	0,3037	4325,6735
Arganda	68,382	0,0146	0,0737	774,9467
San Martín de la Vega	206,931	0,0048	0,0243	242,5588
Mejorada del Campo	10,301	0,0971	0,4890	4999,7899
Madrid	473,151	0,0021	0,0106	156,2003
Getafe	293,908	0,0034	0,0171	193,8759
San Fernando	61,844	0,0162	0,0815	907,7073
suma		0,1985	1	11600,75
Valor estimado renta municipal Velilla de San Antonio				11600,75
ECM				2001,27
EM				-198

3.5. Fronteras comunes: cálculo del IDW con ponderador de longitud de perímetro.

El interpolador IDW supone que las diferencias entre los datos espaciales dependen solo de las distancia entre dos localizaciones, situación que conlleva un sesgo de estimación de la variable renta per cápita municipal al valor medio ponderado que tiene la variable en los puntos muestrales. En lo que sigue del documento incorporaremos al cálculo del IDW, los denominados modelos gravitacionales usados en el comercio internacional. Estos modelos aplican un concepto análogo a la ley de Newton que relaciona la atracción o gravedad entre dos objetos al tamaño de su masa y a la distancia entre ellos. En este sentido, se suele afirmar que el flujo de comercio bilateral entre dos países está relacionado con el tamaño de sus economías (aproximado por su nivel de ingreso o su población), la distancia entre ambos, el lenguaje, la cultura, la existencia de una frontera común, entre otros factores (Cafiero, 2005). Es precisamente la existencia de la frontera común que tomaremos en cuenta para incorporar un segundo ponderador al IDW. La relación contigüidad y distancia se traduce a una expresión matemática dada por la ecuación Nº 8:

$$Z(j) = \sum_{i=1}^N \omega_i \cdot Z(i) \quad (8)$$

$$\text{Con } \omega_i = \frac{\lambda_i + \beta_i}{\sum_{i=1}^N (\lambda_i + \beta_i)}$$

Donde ω representa la participación de dos eventos: una ponderación de la longitud de perímetro compartido de frontera o ponderación de primer orden¹³ entre dos unidades espaciales (β_i) sumado a la ponderación inversa de la distancia entre ellos (λ_i). En este sentido, hacemos referencia a la matriz de contigüidad que Cliff y Ord utilizan para medir la autocorrelación o dependencia espacial donde establecen una combinación de medidas de distancia (inversa de la distancia) y la longitud relativa de la frontera que tienen en común dos unidades espaciales en relación a la longitud de frontera que tiene una unidad espacial con el resto de unidades de su alrededor. Los elementos resultantes de la matriz

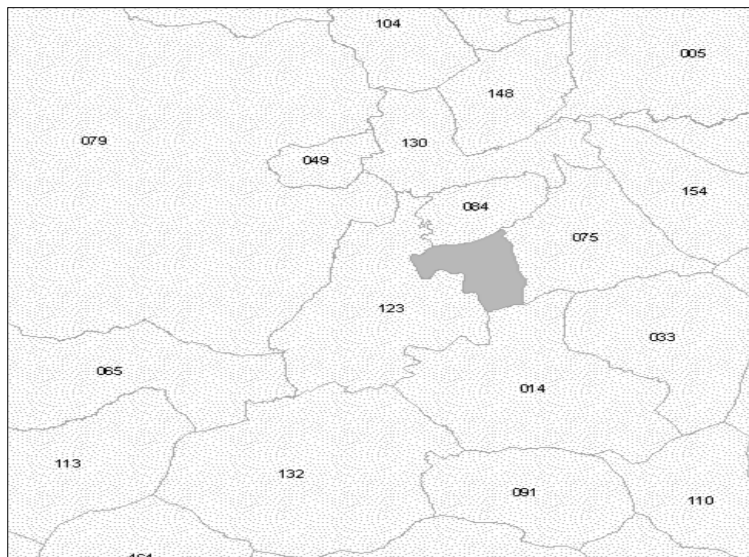
¹³El orden de contigüidad permite diferenciar a las unidades espaciales distribuidos en I y II orden con igualdad de distancias, las primeras poseen fronteras en común .

resultan asimétricos, a menos que ambas unidades espaciales tengan la misma longitud de frontera. Por otro lado, este ponderador tiene aplicación para unidades poligonales y estará muy relacionado con las características físicas de las unidades espaciales.

El ponderador inverso de la distancia (λ_i) se obtiene de la misma forma que en la ecuación (2). El ponderador β_i corresponde a la longitud relativa de la frontera entre la unidades espaciales j e i en relación a la longitud de frontera que tiene la unidad espacial j con el resto de unidades de su alrededor. De esta forma, el ponderador es capaz de diferenciar aquellas unidades que tienen una frontera en común muy amplia de las que apenas tienen un borde en común.

Para ilustrar el cálculo del ponderador de longitud, disponemos en la figura Nº 4 de los municipios vecinos a Velilla (destacado en gris). Los municipios vecinos con población sobre los 10.000 habitantes de primer y segundo orden son: Rivas-Vaciamadrid (123), Arganda (14), San Martín de la Vega (132), Mejorada del campo (84), Madrid (079), Getafe (065) y San Fernando (130). Los datos del vector anotado a un costado de la figura 4, representan el código identificador del municipio y la longitud relativa de vecindad conjunta de Velilla respecto de los vecinos de primer orden.

Figura 3. Mapa de los municipios cercanos a Velilla



$$\begin{bmatrix} 123 \\ 014 \\ 132 \\ 084 \\ 079 \\ 065 \\ 130 \end{bmatrix} = \begin{bmatrix} 0,5 \\ 0,2 \\ 0,0 \\ 0,3 \\ 0,0 \\ 0,0 \\ 0,0 \end{bmatrix}$$

La longitud relativa de la frontera entre la unidad espacial j con el resto de unidades de su alrededor indica que son los vecinos con quienes comparte frontera, los que influyen de una forma más significativa, es decir; el municipio San Martín de la Vega (132) tiene un peso del 0,5 o 50% en la estimación del dato. Mediante el ponderador ω_i es posible discriminar el peso que tendrían municipios que estando a una misma distancia del valor a interpolar, comparten fronteras de aquellas que están en un orden inferior de contigüidad¹⁴.

La tabla Nº 4 muestra la renta estimada mediante interpolación de longitud y distancia. Véase el caso del Municipio de Colmenar Viejo: dada la mayor distancia hacia el municipio de Hoyo de Manzanares, el ponderador de distancia λ_i le asignaba un quinto lugar a su contribución ($\lambda_i=0,07824$) en renta para estimar el valor de la variable en el municipio interpolado. Con la incorporación de un ponderador de

¹⁴ Vecinos de II, III orden por ejemplo.

longitud de perímetro β_i , su contribución en peso e importancia ha aumentado dado la existencia de una frontera en común ($\lambda_i + \beta_i = 0,37824$). Por otro lado, el aporte de los municipios Las Rozas de Madrid y Collado Villalba siendo equidistantes al municipio Hoyo de Manzanares, depende exclusivamente de la renta de cada uno de ellos, mientras que la adición del ponderador de longitud; provoca una diferenciación en la contribución al valor a estimar, dado que uno solo de ellos tiene una frontera en común con Hoyo de Manzanares.

Tabla Nº 4: Predicción renta Hoyo de Manzanares. Técnica de Interpolación IDW modificado

Interpolación Nº 1	Municipio	β_i	$\lambda_i = \frac{d_p^{-p}}{\sum_{i=1}^N d_p^{-p}}$	$\lambda_i + \beta_i$	$\omega_i = \frac{\lambda_i + \beta_i}{\sum_{i=1}^N (\lambda_i + \beta_i)}$	$Z(s_j) = \sum_{i=1}^{10} \omega_i * Z(s_i)$
	Torrelones	0,40000	0,32067	0,72067	0,36034	16914,5167
	Colmenar viejo	0,30000	0,07824	0,37824	0,18912	17352,4934
	Galapagar	0,00000	0,14587	0,14587	0,07294	16903,4362
	Madrid	0,20000	0,02710	0,22710	0,11355	16939,8325
	Rozas de Madrid (Las)	0,00000	0,14944	0,14944	0,07472	16839,9643
	Collado Villalba	0,10000	0,14899	0,24899	0,12449	15564,0647
	Escorial (El)	0,00000	0,04006	0,04006	0,02003	14936,028
	Tres cantos	0,00000	0,08963	0,08963	0,04481	11965,8103
	suma	1	1	2,00000	1,00000	16060,35
Valor estimado renta municipal Hoyo de Manzanares						16060,35

En la interpolación para Velilla de San Antonio, véase el caso de los municipios de Rivas-Vaciamadrid y Mejorada del Campo, que responden a los siguientes casos: municipios muy cercanos en distancia (a centroide) y con frontera común a Velilla de San Antonio, con lo cual su contribución es significativamente alta comparada con los otros municipios participantes del proceso. Tales municipios de acuerdo al ponderador de distancia son fácilmente discriminados respecto de por ejemplo, el municipio de Arganda, ubicado a una distancia mayor. Sin embargo, tal significatividad disminuye si se considera la condición de longitud de perímetro compartido (véase Tabla Nº 5).

Tabla Nº 5: Predicción renta Velilla de san Antonio. Técnica de Interpolación IDW modificado

Municipio	β_i	$\lambda_i = \frac{d_p^{-p}}{\sum_{i=1}^N d_p^{-p}}$	$\lambda_i + \beta_i$	$\omega_i = \frac{\lambda_i + \beta_i}{\sum_{i=1}^N (\lambda_i + \beta_i)}$	$Z(s_j) = \sum_{i=1}^{10} \omega_i * Z(s_i)$
Rivas-Vaciamadrid	0,50	0,3037	0,804	0,4019	5723,26
Arganda	0,25	0,0737	0,324	0,1618	1702,41
San Martín de la Vega	0,00	0,0243	0,024	0,0122	121,28
Mejorada del Campo	0,25	0,4890	0,739	0,3695	3777,92
Madrid	0,00	0,0106	0,011	0,0053	78,10
Getafe	0,00	0,0171	0,017	0,0086	96,94
San Fernando	0,00	0,0815	0,081	0,0407	453,85
suma	1	1	2,00	1	5723,26
Valor estimado renta municipal Velilla de San Antonio					11953,77

La Tabla Nº 6 muestra un resumen de los valores de la predicción de la renta para los municipios de Hoyo de Manzanares y Velilla de San Antonio mediante la técnica del IDW. En ambos casos, la incorporación de la longitud de perímetro compartido y su contigüidad mejora la precisión en el valor de estimación medido en el valor de ECM y EM.

Tabla Nº 6: Resultados de la interpolación

	Renta bruta municipal 2004	Renta interpolada con IDW	Renta interpolada con IDW modificado
Hoyo de Manzanares			
Renta	16156	16885,05	16060,35
ECM	-	4437,12	3977,97
EM	-	590	254
Velilla de San Antonio			
Renta	11922	11600,75	11953,76
ECM	-	2001,27	1921,08
EM	-	-198	249

4. CONCLUSIONES

Aunque la mejor solución al problema de la ausencia de datos sería la obtención de la información adicional necesaria, esto resulta inviable en muchas ocasiones. Y, de llevarse a cabo, tampoco garantizaría la inexistencia de valores ausentes, con lo cual hemos abordado esta primera aproximación a las técnicas de interpolación espacial, como solución a la estimación de valores ausentes de una variable. La técnica de interpolación IDW pondera la contribución de cada unidad espacial con valor conocido, exclusivamente en función de la distancia que los separa. Sin embargo, la técnica no diferencia entre variables que representan a datos puntuales de aquellos que corresponden a polígonos. Para provocar esta diferencia, se ha incluido en el IDW la valoración de la existencia de frontera compartida entre unidades espaciales. Lo anterior tiene raíces en los modelos gravitacionales de comercio internacional que establecen un flujo positivo y mayor cuando las unidades espaciales comparten frontera. El IDW propuesto considera que la longitud y la distancia tienen la misma importancia en la interpolación del valor ausente. Esto es, el ponderador por distancia asigna una contribución mayor a los puntos más cercanos, mientras que el ponderador por longitud, lo hace a los puntos que comparte frontera. La ponderación entre distancia y longitud, se convierte en un parámetro a decidir por el investigador, de la misma forma que elige p y n en el ponderador de la distancia.

El rango de valores en que podría fluctuar el valor interpolado se obtienen suponiendo toda la contribución a los vecinos que comparten frontera y toda la contribución en función inversa de la distancia. En el primer caso, el valor dependerá del valor que adopte la variable en los vecinos de 1º orden, con lo cual, la presencia de autocorrelación espacial positiva y no estacionariedad de segundo grado es altamente aconsejable.

REFERENCIAS

- Aguilar, M. et al.** (2001): Evaluación de diferentes técnicas de interpolación espacial para la generación de modelos digitales de elevación del terreno agrícola. *Mapping*, nº 74, pp. 72-88.
- Bennett; Haining; and Griffith** (1984): The Problem of Missing Data on Spatial Surfaces *Annals of the Association of American Geographers*, Vol. 74, No. 1, pp. 138-156.
- Bosque, J.** (1997): *Sistemas de información geográfica* Madrid, Rialp, 2ª edición corregida.
- Burrough, P.** (1986): *Principles of geographical information systems for Land Resources Assessment*. Oxford: Clarendon
- Burrough, P. y McDonnell, R.** (1998) *Principles of geographical information systems*. New York: Oxford University Press.
- Chasco, C.** (2003): *Econometría espacial aplicada a la predicción-extrapolación de datos microterritoriales*, Comunidad de Madrid. Madrid.
- Cafiero, J.** (2005): Modelos Gravitacionales para el Análisis del Comercio exterior. Revista del CEI Nº 4 Centro de Economía Internacional. Ministerio de Relaciones Exteriores, Comercio Internacional y Culto. Argentina.
- Coman, G.** (1998): Shepard operators of Birkhoff-type, *Calcolo* 35, pp. 197-203.
- Coley, A. y Clabburn, P.** (2005): GIS Visualisation and Analysis of Mobile Hydroacoustic Fisheries Data: a practical example. *Fisheries Management and Ecology*, Vol. 12, Nº6, pp 361–367.
- Declercq, F.** (1996): Interpolation methods for scattered sample data: accuracy, spatial patterns, processing time. *Cartography and Geographic Information Systems* Vol. 23, Nº 3, pp. 128-144.
- Dowdall, M. y O'Dea, J.** (1999): Comparison of Point Estimation Techniques in the Spatial Analysis of Radium-226, Radium-228 and Potassium-40 in Soil. *Environmental Monitoring and Assessment*. Vol. 59, Nº 2, pp. 123-247.
- Echegoyen, T.** (2002): Desarrollo de un lenguaje original de proceso de mallas y aplicaciones a la ingeniería civil. Tesis Doctoral, Universidad Politécnica de Madrid.
- Franke, R.** (1982): Scattered data interpolation: tests of some methods. *Mathematics of computation*, Vol. 38, Nº 157, pp. 181-200.
- Geary, R.** (1954):The contiguity ratio and statistical mapping. *The Incorporated Statistician*, Vol. 5, Nº 3 (Nov., 1954), pp. 115-127+129-146.
- Gemmer, M. et al** (2004):. Observed monthly precipitation trends in China 1951–2002. *Theor. Appl. Climatol.* Vol. 77, pp. 39–45.
- Grinstead, B. et al** (2006): Model Building for Simulation and Testing under Uncertain Conditions. *SPIE Modeling and Simulation for Military Applications*, Vol. 6228, pp. 98-109.
- Gordon, W. y Wixom, J.** (1978): Shepard's Method of "Metric Interpolation" to Bivariate and Multivariate Interpolation *Mathematics of Computation*, Vol. 32, Nº 141, pp. 253-264.

- Haavelmo, T.**(1950): The notion of involuntary economic decisions. *Econometrica* 18, pp.1-8.
- Hutchinson et al** (2000): A comparison of two statistical methods for spatial interpolation of Canadian monthly mean climate data. *Agricultural and Forest Meteorology*, Vol. 101, Nº 2, pp. 81-94.
- Institut d'Estadística de Catalunya, Idescat** (2006): *Renta familiar disponible bruta (RFDB) y componentes estimados. Base 2000. Años 2000-2002*. Disponible en <http://www.idescat.net>
- Instituto Aragonés de Estadística, Iae** (2006): *Renta disponible bruta. Aragón. Serie 1999-2002*. Disponible en <http://portal.aragob.es>
- Instituto de Estadística de la Comunidad de Madrid** (2006): Indicador de renta disponible bruta municipal 1999-2002, 2003 (provisional) y 2004 (avance). Disponible en <http://www.madrid.org/iestadis>
- Isaaks, E. y Srivastava, M.** (1989): *An Introduction to Applied Geostatistics*, Oxford University Press, New York.
- Lam, Nina S.** (1983): Spatial interpolation methods: A review. *The American Cartographer*, Vol. 10, Nº 2, pp. 129-149.
- Larraz, B.** (2004): Tesis Doctoral “Técnicas de Cokrigado para el análisis económico. Estimación de precios de bienes inmuebles en el casco histórico de la ciudad de Toledo”. Universidad Castilla la Mancha.
- Lozano, A.** (2003). Tesis Doctoral “Optimización del diseño de redes de vigilancia y control de la calidad del aire y su aplicación en Andalucía”. Universidad de Sevilla.
- Moran, P.** (1948): The interpretation of statistical maps. *Journal of the Royal Statistical Society, Series B* Vol.37, Nº 2, pp. 243–251.
- Moreno A. y Prieto M^a** (2002): “Estimación con sistemas de información geográfica del área de servicio y la demanda potencial para el transporte urbano: Una comparación de métodos” Departamento de Geografía. Universidad Autónoma de Madrid.
- Morillo, J. et al** (2002): Análisis de Calidad de un Modelo Digital de Elevaciones Generado con Distintas Técnicas de Interpolación. XIV Congreso Internacional de Ingeniería Gráfica.
- Mulugeta G.** (1996): Manual and Automated Interpolation of Climatic and Geomorphic Statistical Surfaces: An Evaluation. *Annals of the Association of American Geographers*. Vol. 86, Nº 2, pp. 324-342.
- Mur, J. y Angulo, A.** (2006): PROCESOS ESTOCÁSTICOS ESPACIALES, Documentos de clase. depto. ANÁLISIS ECONÓMICO Facultad de Ciencias Económicas y Empresariales, Universidad de Zaragoza.
- Nalder y Wein** (1998): Spatial interpolation of climatic Normals: test of a new method in the Canadian boreal forest. *Agricultural and Forest Meteorology*, Vol. 92, Nº 4, pp. 211-225.
- Shiou, F. y Lin, Y.** (2002): Calculation of the Normal Vector using the 3X3 Moving Mask Method for Freeform Surface Measurement and its Application. *Int J Adv Manuf Technol*, Vol. 19, pp. 516–524.
- Shepard, D.** (1968): A Two-Dimensional Interpolation Function for Irregularly Spaced Data, Proc. 23rd Nat. Conf. ACM, 1968, pp. 517-523.

Tobler, W. (1979): «Cellular Geography», en Gale, S. y Olson G. (dir.): *Philosophy in Geography*. Ed. Dordrecht. pp 279-386.

Tomczak, M. (1998): Spatial interpolation and its uncertainty using automated anisotropic Inverse Distance Weighting (IDW)- cross-validation/jackknife approach. *Journal of Geographic Information and Decision Analysis*. Vol. 2, Nº2, pp. 18-30.

Watson, D. y Phillip, G. (1985): A Refinement of Inverse Distance Weighted Interpolation. *Geo-Processing*, Elsevier Science Publishers, Nº 2, pp. 315-327.

Wingle, W.L. (1992): Examining Common Problems Associated with Various contouring Methods, Particularly Inverse-Distance Methods, Using Shaded Relief Surfaces, *Geotech '92 Conference Proceedings*, Lakewood, Colorado, pp. 362-376.

Wolf-Dieter, R (2001): Volume-preserving interpolation of a smooth surface from polygon-related data. *Journal of Geographical Systems*, Vol. 3, Nº. 2, pp. 199-213.

Yang, X. y Hodler, T. (2000): Visual and statistical comparisons of surface modeling techniques for point-based environmental data. *Cartography and Geographic Information Science*, Vol. 27, Nº 2, pp. 165-175.