

# Detección de *clusters* y otras estructuras regionales y urbanas con técnicas de econometría espacial

C. CHASCO YRIGOYEN

Dpto. Economía Aplicada. Universidad Autónoma de Madrid

**RESUMEN:** A lo largo del siglo XX, y de forma general en el mundo, las ciudades han sufrido unos cambios tan espectaculares que han dado lugar a una transformación total del escenario urbano. Por este motivo, en los últimos años gran parte de la literatura geográfica se ha dedicado a analizar el grado de concentración espacial que tanto población como empleo/empresas tienen en las zonas urbanas. El interés por estas “formas” y estructuras ha puesto de manifiesto la importancia que tanto la localización como la distancia tienen en el análisis de datos espaciales. Este interés ha generado también una demanda creciente de métodos y técnicas capaces de realizar un tratamiento adecuado de los datos geográficos en trabajos empíricos. Un ejemplo de ello lo constituyen el Análisis Exploratorio de Datos Espaciales (AEDE) y los Sistemas de Información Geográfica (SIG), que son poderosos instrumentos al servicio del descubrimiento y visualización de todo tipo de estructuras espaciales. En este artículo, se presentan los métodos del AEDE propios de la econometría espacial que se han utilizado para la exploración de clusters geográficos, de carácter regional y urbano, en la población y la actividad económica. En concreto, se presenta una selección de herramientas interesantes para la detección de tendencias espaciales, agrupamientos, puntos atípicos y fenómenos de difusión, con objeto de disponer de un mejor conocimiento y/o predicción de los comportamientos humanos.

**DESCRIPTORES:** Análisis Exploratorio de Datos Espaciales (AEDE). Análisis regional y urbano. *Clusters* espaciales. Geomarketing. Econometría espacial.

## 1. Introducción

El análisis del espacio y la localización ha constituido uno de los pilares de la investigación en las ciencias sociales en los últimos 50 años. A ello han contribuido, sin duda, los profundos cambios geopolíticos acaecidos en las últimas décadas, entre los que cabe señalar la descomposición de la ex-Unión

Soviética (con la creación de los nuevos estados de la República Rusa y la Europa del Este), el proceso de unificación de la Unión Europea, que afianza la realidad de una Europa de las Regiones y, en España, la consolidación de un proceso de descentralización territorial, que concede una importancia creciente los ámbitos regional e infrarregional. Además, en un mundo cada vez más urbanizado, se han producido

Recibido: 16.09.2010 ; Revisado: 08.10.2010  
e-mail: [coro.chasco@uam.es](mailto:coro.chasco@uam.es)  
La autora agradece la financiación recibida por el Proyecto ECO2009-10534 del Ministerio de Ciencia e Inno-

vación y 2009/AX702 de la Xunta de Galicia. Y, a los evaluadores anónimos sus comentarios para la mejora del presente trabajo.

problemas nuevos de tipo medioambiental, climático, de recursos humanos y de salud pública que están íntimamente unidos a fenómenos como la descentralización de las ciudades, la congestión, el acceso a la vivienda, la pérdida de espacios abiertos, etc. Estos acontecimientos y las consecuencias derivadas de los mismos, necesitan de una atención que implica la consideración de diversos niveles de planificación espacial y toma de decisiones.

Por este motivo, en los últimos años gran parte de la literatura geográfica se ha dedicado a analizar cuestiones como el crecimiento económico y convergencia regional, el grado de concentración espacial que tanto población como empleo tienen en las zonas urbanas, las relaciones y el equilibrio entre los usos del suelo, los medios de transporte y el medio ambiente, etc. Todos estos problemas de tipo regional y urbano han potenciado el desarrollo del análisis espacial, tanto en sus aspectos teóricos, como en sus métodos y aplicaciones. Las primeras colecciones de libros colectivos dedicados al estudio monográfico del análisis de datos espaciales y la econometría espacial datan de los años noventa del siglo pasado (ANSELIN & *al.*, 1994; FISCHER & GETIS, 1997). Desde entonces hasta la publicación de obras similares más recientes (ANSELIN & *al.*, 2004; FISCHER, 2006; SMITH & *al.*, 2007; ANSELIN & REY, 2010; FISCHER & GETIS, 2010; JIANG & YAO, 2010; PÁEZ & *al.*, 2010), mucho se ha avanzado en este campo, que engloba a investigadores procedentes de ramas de saber tan diversas como la geografía cuantitativa, la economía y ciencia regional, la planificación regional y urbana, el sector inmobiliario, la ingeniería civil, la epidemiología y la criminología, entre otros.

Este interés por los fenómenos regionales y urbanos ha generado también una demanda creciente de métodos y técnicas capaces de realizar un tratamiento adecuado de los datos geográficos en trabajos empíricos. Un ejemplo de ello son los Sistemas de Información Geográfica (SIG) y el Análisis Exploratorio de Datos Espaciales (AEDE), capaces de visualizar y analizar todo tipo de estructuras espaciales (ANSELIN, 2006). Hay que tener en cuenta que los efectos espaciales de dependencia (autocorrelación espacial) y heterogeneidad (estructura espacial) son de tal calibre que no pueden ser ignorados sin caer en importantes sesgos. Es por eso, que requieren de métodos y modelos propios, así como de un “software” que, a la par que especializado, sea cercano también a los no especialistas en la materia (*cfr.* CHASCO & FERNÁNDEZ-AVILÉS, 2009: 21).

El AEDE podría ser también definido como un grupo de técnicas capaces de identificar valores espaciales atípicos, esquemas de asociación espacial o agrupamientos (*clusters*), puntos calientes/fríos (*hot spots*) u otras formas de heterogeneidad espacial (ANSELIN, 1999); es decir, tiene un carácter descriptivo (estadístico) más que confirmatorio (econométrico). Pero, tras la detección de estructuras espaciales en las variables geográficas, el AEDE hará también posible la formulación de hipótesis previas para la modelización econométrica y, en su caso, la posible predicción espacial de nuevos datos. Además, el AEDE combina el análisis estadístico con el gráfico, dando lugar a lo que podría denominarse una “visualización científica” (HAINING & *al.*, 2000) que, a los contrastes estadísticos sobre los efectos espaciales de dependencia y heterogeneidad, une un amplio marco de gráficos o “vistas” múltiples y dinámicas sobre la información geográfica (UNWIN, 2000). Por eso, estos métodos de visualización científica son más que simples mapas o gráficos estáticos de representación, siendo fundamental el papel jugado por los paquetes informáticos especializados.

En ciertas ocasiones, al analista le resultará suficiente con la utilización de un AEDE elemental o básico, que haga posible la exploración de aquellas características más sobresalientes de los datos. Sin embargo, en otras muchas circunstancias, el AEDE básico no permitirá detectar algunas estructuras y agrupamientos espaciales más complejos, por lo que será necesario recurrir a técnicas avanzadas del AEDE (ANSELIN, 2005; CHASCO & FERNÁNDEZ-AVILÉS, 2009). En este artículo, se presentan aquellos métodos del AEDE propios de la econometría espacial más utilizados para la detección y exploración de clusters geográficos en ámbitos regionales y urbanos. En concreto, se presentan algunas herramientas interesantes que permiten la afloración de tendencias espaciales, agrupamientos, puntos atípicos y fenómenos de difusión con objeto de disponer de un mejor conocimiento y/o predicción de los comportamientos humanos.

## 2. Técnicas básicas de análisis exploratorio espacial

El primer paso en todo análisis de datos espacial es la representación gráfica, a través de un mapa, del fenómeno objeto de estudio. La posibilidad de representar elementos cartográficos y distribuciones espaciales en un mapa digital, como la localización de los clientes de una empresa o la distribución de la renta regio-

nal ha revolucionado campos como el geomarketing o el análisis regional y urbano, respectivamente. Como es sabido, la presentación de datos en una tabla no permite visualizar las relaciones espaciales existentes entre las unidades, cosa que sí es posible cuando se utiliza un mapa u otros métodos gráficos, siempre en combinación dinámica con un mapa. En este apartado, presentamos algunos métodos básicos de AEDE, destinados no sólo a la mera representación de fenómenos espaciales, sino también a una primera detección de las características de su distribución (tendencia central, dispersión, concentración) y discontinuidades (puntos atípicos).

### 2.1. Análisis de la distribución y puntos atípicos en las variables geográficas

La tendencia central de una variable geográfica puede detectarse con ayuda de los llamados mapas temáticos, así como de gráficos estadísticos más o menos estándar como el histograma de frecuencias, función de densidad, el mapa de caja y el gráfico condicional. Los SIG permiten elaborar todo tipo de mapas temáticos, así como ciertos gráficos estadísticos básicos. En concreto, un *mapa temático* consiste en la representación cartográfica de una variable de datos espaciales utilizando símbolos y/o colores para poner de manifiesto las diferencias entre distintos intervalos de valores. En los mapas de cuantiles, los datos se dividen y agrupan en una serie de categorías (o cuantiles) con igual número de observaciones. Los mapas temáticos que mejor representan la tendencia espacial global de una variable son los mapas de cuantiles, el mapa de la desviación típica o el mapa de cortes naturales (JENKS & CASPALL, 1971). Este último es muy utilizado por ser útil para representar variables con una distribución muy desigual. Se trata de una técnica que, para determinar los intervalos, identifica los puntos de corte de forma que se haga mínima la varianza total intra-grupos. De este modo, detecta agrupamientos y estructuras subyacentes en los datos, que muchas veces no se vislumbran a simple vista. En la FIG. 1(a), hemos representado el mapa de cortes naturales de la densidad del empleo en los municipios de la Eurorregión “Galicia-Norte de Portugal” que, como puede observar-

se, destaca como muy densas las áreas metropolitanas de Porto, A Coruña y Vigo.

En ciertas ocasiones, el mapa puede no ser un instrumento suficiente para identificar la forma de una distribución espacial. Esto es especialmente cierto cuando en la misma, hay valores excepcionalmente bajos/altos que pueden no ser representativos de la distribución general. Por su parte, los valores atípicos pueden tener importancia en cuanto que se trata de datos interesantes (acontecimientos extraños, localizaciones calientes o frías para un negocio) o simplemente errores en la entrada de datos. Entre las técnicas más utilizadas para la detección de valores extremos en un mapa está el *cartograma*, que es un gráfico en el que las unidades espaciales originales son reemplazadas por círculos cuya área es proporcional al valor de la variable que se quiere representar. Gracias a un algoritmo de optimización no lineal es posible situar los círculos en el lugar más cercano posible a su ubicación espacial original. Los valores atípicos (tanto los inferiores como los superiores) se destacan con un color diferente al del resto de elementos. Así, en la FIG. 1(b) representa el cartograma de la variable de densidad de empleo municipal en la Eurorregión. A diferencia del mapa de cortes naturales, este gráfico permite detectar —y establecer una jerarquía entre— los polos y subpolos de empleo en esta zona.

Otro modo de visualizar valores atípicos en el territorio es el proporcionado por el *gráfico condicional*. El principio que subyace este tipo de gráficos es la utilización de las coordenadas geográficas X-Y como variables condicionales que dividen a la muestra total de datos en diferentes grupos. Si, por ejemplo, se diferencian tres categorías para cada coordenada, la superficie total del territorio habrá quedado dividida en nueve sub-espacios. Este gráfico representa, en cada uno de ellos, los datos sub-muestrales correspondientes a través de un gráfico o mapa, según se indique. Esto es lo que se ha hecho en la FIG. 2(a), donde se ha representado la variable del porcentaje de hogares con problemas de ruido en 5.080 localizaciones dentro de la “Almendra Central” de Madrid (los distritos rodeados por la carretera de circunvalación M-30). Una vez divide esta zona de la ciudad en nueve sub-espacios, se ha utilizado el *diagrama de caja*<sup>1</sup> como medio de represen-

<sup>1</sup> El diagrama de caja es un método de representación basado en el cálculo de los cuantiles y la mediana de una variable, así como en la obtención de las llamadas cotas o valores adyacentes superior e inferior. La “caja” es un rectángulo que se construye de forma que el valor infe-

rior de la misma es el primer cuartil (que contiene el 25% de las observaciones) y el valor superior, el tercer cuartil (que contiene el 75% de las observaciones). La mediana queda destacada en mitad de la caja con un círculo y una línea horizontal que la atraviesa. Las cotas se repre-

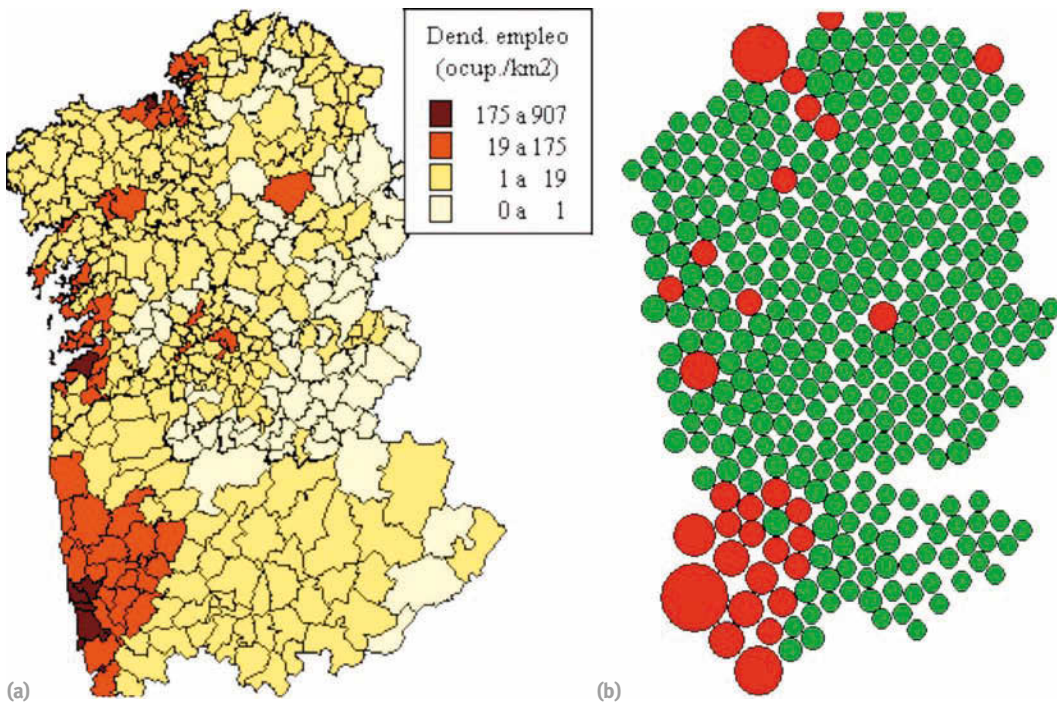


FIG. 1/ (a) Mapa de cortes naturales y (b) cartograma de la variable de densidad de empleo en los municipios de la Euroregión “Galicia-Norte de Portugal”

Fuente: Elaboración propia a partir de la base de datos SABI e Inebase (INE)

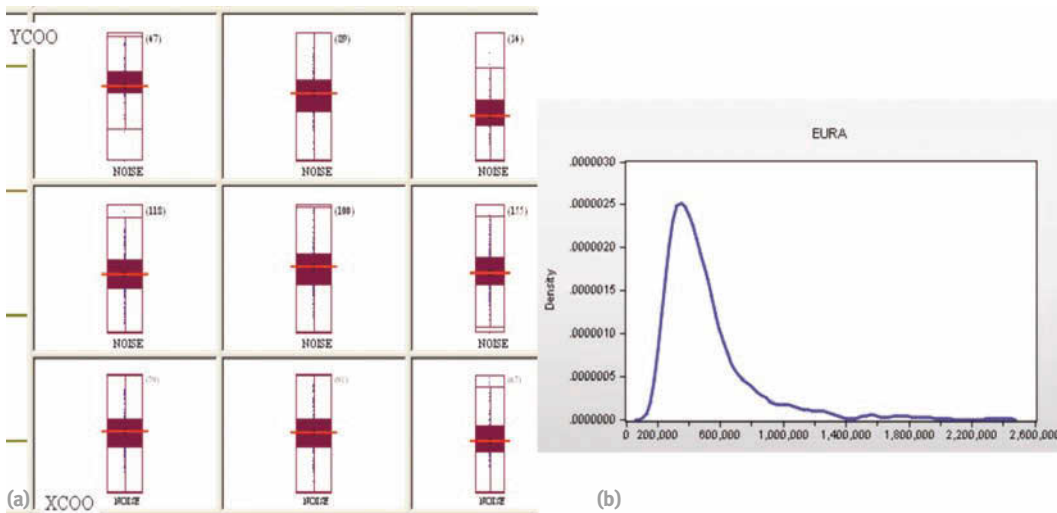


FIG. 2/ (a) Función de densidad del precio de la vivienda en localizaciones del centro de Madrid. (b) Gráfico condicional-diagrama de caja, del porcentaje de hogares del centro de Madrid con problemas de ruido

Fuente: Elaboración propia a partir de la base de datos de Idealista.com

sentan, en la parte superior e inferior de este gráfico en forma de T, como una línea horizontal gruesa. Estas cotas, a su vez, se obtienen sumando/restando a la mediana el producto de los valores del tercer (primer) cuartil por 1,5 veces el recorrido intercuartílico. De esta forma, se consideran como valores atípicos aquéllos que están

situados por encima (o por debajo) de dichas cotas, pudiendo no existir dichos valores (cuando la variable tenga valores muy concentrados en torno a la media). Si las cotas estuvieran bajo el valor inferior de la variable o sobre el valor superior, los bordes del gráfico serían dichas cotas (ANSELIN, 2005).



tación de las submuestras correspondientes a los nueve sub-espacios. De este ejemplo cabe destacar la existencia de valores atípicamente elevados en las distribuciones de los tres sub-espacios del este y la zona oeste-centro de la "Almendra Central". Es decir, en estos sub-espacios, se producen mayores diferencias en el porcentaje de hogares con problemas de ruido debido a la existencia de ciertas localizaciones donde este problema es especialmente importante. Lo contrario sucede en el sub-espacio noroeste, en el que encontramos valores atípicos bajos, es decir, con un bajo porcentaje de hogares con problemas de ruido.

Asimismo, la distribución de una variable puede visualizarse con un *gráfico de densidad*, que se trata de una aproximación a su función de densidad a través de algún algoritmo como un estimador kernel no paramétrico. En general, este gráfico puede ser interpretado como un histograma continuo en el que se lleva al infinito el número de intervalos. Un ejemplo de este tipo de gráfico se encuentra en la FIG. 2(b) donde se representa el precio de la vivienda en las citadas 5.080 localizaciones del centro de Madrid. Este gráfico deja al descubierto la gran asimetría a derechas de esta variable, con valores atípicamente elevados en el precio de algunas viviendas. Sin embargo, en esta zona no hay puntos extraordinariamente bajos en el precio de la vivienda, dado que la mayor parte de los datos se encuentra concentrada en torno al valor medio.

### 3. El efecto de la dependencia espacial

Este apartado está reservado al análisis espacial avanzado, que está basado en el concepto de dependencia espacial. Tras dos apartados dedicados al concepto y formulación matemática de este fenómeno, se analizan algunos métodos de contraste de autocorrelación espacial, así como ciertos instrumentos de detección de agrupaciones y estructuras espaciales.

#### 3.1. Concepto de dependencia o autocorrelación espacial

La dependencia espacial es, junto con la tendencia y los valores atípicos, un elemento caracterizador de las distribuciones espaciales. Se trata de un efecto que consiste en la coincidencia de valores (altos/bajos) de una variable en determinados lugares del espacio geográfico. Esta propiedad es también conocida con el nombre técnico de autocorrelación

espacial. El fenómeno de dependencia espacial en una variable puede expresarse matemáticamente como una relación funcional entre lo que ocurre en un punto determinado del espacio y lo que sucede en lugares cercanos —vecinos— al mismo (ANSELIN, 2001a). Es decir, en una variable se produce el fenómeno de la dependencia o autocorrelación espacial cuando los valores observados en un área (ciudad, distrito, etc.) dependen de los valores observados en áreas vecinas.

Este tema ha suscitado un interés creciente desde los primeros trabajos de CLIFF & ORD (1973 y 1981), donde los autores estudian este fenómeno para conocer las causas y formas de propagación de epidemias y enfermedades, aunque otros autores, como COX (1969), ya habían constatado previamente la existencia de continuidad geográfica en fenómenos sociales como la intención de voto. Posteriormente, este efecto espacial fue objeto de aplicaciones interesantes en otros campos, como los recursos naturales y medio ambiente (ANSELIN, 2001b), criminología (BALLER & *al.*, 2001; CECCATO & *al.*, 2002), actividad de I+D (ANSELIN & *al.*, 2000; FISCHER & VARGA, 2003), etc. En el terreno específicamente económico, se ha demostrado como muy común la existencia de autocorrelación espacial en fenómenos como la producción, paro, renta disponible, procesos de convergencia regional, demanda de transporte, impuestos, precio de la vivienda, población, etc. Algunos ejemplos en este campo serían REY & MONTOURI (1999), ARBIA (2006), ANSELIN & LOZANO-GRACIA (2008) o LE GALLO & CHASCO (2008), entre otros muchos.

El efecto de autocorrelación espacial puede ser de signo positivo o negativo, así como nulo. Se entiende por *autocorrelación espacial positiva* el fenómeno de asociación entre valores similares de una variable y localizaciones cercanas. Es decir, cuando, en el espacio geográfico, los valores altos de una variable están rodeados por valores altos, y viceversa. Éste sería el caso, del llamado efecto contagio o desbordamiento (*spillover*) que se produce en muchos fenómenos socioeconómicos de renta y desarrollo humano, en los que su presencia en una región es causa de una extensión de dicho fenómeno a regiones vecinas, favoreciendo la concentración del mismo en una zona geográfica determinada. Por el contrario, existe *autocorrelación espacial negativa* en un espacio cuando los valores altos de una variable se encuentran rodeados por valores bajos de la misma, y viceversa. Esto es lo que ocurre con la disposición de las casillas blancas y negras en un tablero de ajedrez. Según se ex-

pone en ANSELIN (1988), este efecto espacial es más difícil de encontrar en variables de tipo socioeconómico (si se compara con las ciencias naturales, por ejemplo) y está muy condicionado por la escala geográfica en la que se lleven a cabo los análisis<sup>2</sup>. Por último, se produce *ausencia de autocorrelación espacial* en una variable geográfica cuando ésta se distribuye de manera aleatoria sobre el espacio.

### 3.2. Formulación matemática del efecto de autocorrelación espacial

En un espacio compuesto por  $n$  observaciones, la existencia de autocorrelación espacial en una variable  $Y$  supondrá la existencia de una relación funcional entre el valor de  $Y$  en la localización  $i$  ( $y_i$ ) y los valores de dicha variable en todas, y cada una, de las  $n$  localizaciones situadas en el territorio en estudio:

$$y_i = f(y_1, y_2, \dots, y_n)$$

Esta expresión no resulta operativa en la práctica porque da lugar a un sistema no identificable, con muchos más parámetros a estimar, hasta un total de  $(n^2 - n)$ , que observaciones disponibles ( $n$ ). Por este motivo, habrá que suponer una estructura común y, por tanto, más general, en fenómeno de interacción que se estudia, de forma que se estime un número menor de parámetros de autocorrelación espacial.

Esta simplificación es posible gracias al concepto de *retardo espacial* que, aunque similar, no es equiparable al más conocido del retardo temporal. En efecto, aunque la dependencia espacial es un fenómeno similar a la dependencia en las series temporales, ambas se diferencian en la naturaleza *multidireccional* de la dependencia en el espacio frente a la clara situación unidireccional del tiempo. En el campo de las series temporales, las relaciones entre valores contiguos de una misma variable  $Y$  son de carácter lineal y, por tanto, unidireccional: de presente a pasado ( $y_t, y_{t-1}$ ) o de presente a futuro ( $y_t, y_{t+1}$ ). Por el contrario, la relación de contigüidad (o vecindad) entre una observación de la variable  $Y$  en el espacio ( $y_{i,j}$ ) debe extenderse a todas las posibles direcciones existentes en el plano.

Por eso, antes de definir el retardo espacial, debe aclararse qué se entiende por *vecindad geográfica*. La definición de vecindad geográfica requiere de la existencia de un mapa en el que se encuentren representadas las fronteras entre unidades espaciales poligonales o, en el caso de ser puntos, las distancias entre ellos. Cuando se trabaja con unidades geográficas de carácter poligonal, el criterio de vecindad más intuitivo (y utilizado) se basa en el concepto de *frontera* común. Es decir, dos unidades geográficas de tipo poligonal serán consideradas como vecinas si tienen una frontera común. De este modo, las provincias vecinas a Madrid serán las cinco que tienen una frontera político-administrativa en común. Cuando las unidades espaciales son puntos, como las ciudades en un país, el concepto de vecindad se define sobre la noción de *distancia crítica*, *radio de distancia* o *camino más corto*.

La estructura de vecindad espacial puede expresarse matemáticamente a través de una matriz denominada *matriz de interacciones, pesos, ponderaciones, distancias o contactos espaciales* (de aquí en adelante, "matriz  $W$  de pesos espaciales"). Se trata de una matriz cuadrada de orden  $n$  (tantos elementos como unidades espaciales tenga el sistema), en la que cada unidad geográfica se representa a la vez mediante una fila y una columna. En cada fila, los elementos ( $w_{ij}$ ) no nulos de las columnas se corresponden con las unidades vecinas. Es decir,  $w_{ij} = 1$  cuando las unidades espaciales  $i, j$  son vecinas (según el criterio de vecindad previamente fijado) y  $w_{ij} = 0$  cuando dichas unidades  $i, j$  no son vecinas. Además, por convenio,  $w_{ij} = 0, \forall i = j$ ; es decir, una unidad espacial nunca se considera vecina de sí misma, por lo que la diagonal principal de la matriz estará formada por ceros. Además del criterio de vecindad basado en la frontera común, se han propuesto otras formulaciones más complejas (CHASCO & FERNÁNDEZ-AVILÉS, 2009).

La matriz de pesos espaciales permite definir el retardo espacial como un promedio ponderado de los valores de una variable en las localizaciones vecinas a una dada, con unos pesos fijos y dados de forma exógena. Es decir, el valor espacialmente retardado de una variable en una localización  $i$  sería la suma ponderada de todos los valores de dicha variable en las localizaciones vecinas a  $i$ . Los térmi-

<sup>2</sup> GRIFFITH (2006) indica que el signo del efecto de autocorrelación espacial puede cambiar según la escala territorial de análisis. Así, si una variable se ha medido en unidades territoriales de ámbito muy desagregado (por ejemplo, secciones censales o distritos) y presenta autocorrelación espacial negativa, puede ver cambiado el

signo de este efecto si se mide en ámbitos espaciales más agregados (municipios, provincias). En este último caso, el efecto de agregación promediará los valores de la variable que, por este motivo, serán más parecidos entre sí, produciendo autocorrelación espacial positiva.

nos de una variable espacialmente retardada se obtienen multiplicando las observaciones en cuestión ( $y_i$ ) por sus correspondientes pesos de la matriz de ponderaciones  $W$  del modo siguiente:

$$W y_i = \sum_{j \in J} w_{ij} y_j \quad ; \quad \forall j \in J_i$$

donde  $W$  es la matriz de pesos espaciales (el equivalente al operador retardo en series temporales),  $j$  es el subíndice correspondiente al conjunto  $J_i$  de unidades relacionadas con  $i$ , según un criterio de vecindad determinado,  $w_{ij}$  son las ponderaciones espaciales.

En el caso de una matriz de vecindad basada en el criterio de tener una frontera común, el retardo espacial no sería un promedio ponderado, sino la suma de los valores de una variable en las localizaciones vecinas. Por eso, hay autores que recomiendan transformar la matriz  $W$  para calcular variables espacialmente retardadas, sobre todo cuando van a ser utilizadas en modelos de regresión espacial (*vide* ANSELLIN, 1992: capítulo 16.1). Esta transformación se denomina *estandarización por filas*. Una matriz  $W^*$  estandarizada por filas es aquella en la que cada elemento  $w_{ij}^*$  se divide por la suma de los pesos de cada fila, del modo siguiente:

$$w_{ij}^* = \frac{w_{ij}}{\sum_{j \in J} w_{ij}}$$

En este tipo de matrices, la variable espacialmente retardada será el valor suavizado (o alisado) de los valores vecinos a cada localización, siendo la suma de los elementos de cada fila igual a la unidad. Además, como  $i = 1, 2, \dots, n$ , la suma de todos los pesos de la matriz  $W$  es igual al tamaño de la muestra ( $n$ ).

### 3.3. Métodos de representación y contraste de la dependencia espacial

En el apartado anterior vimos que el fenómeno de dependencia espacial en una variable suele expresarse matemáticamente como una relación funcional entre lo que ocurre en un punto determinado del espacio y lo que sucede en lugares cercanos (vecinos) al mismo. Es por eso que este fenómeno de dependencia espacial se conoce también como autocorrelación, asociación o interacción espacial, dado que expresa las relaciones de interrelación existen-

tes en los fenómenos socio-económicos sobre el espacio geográfico. Cuando una variable tiene autocorrelación espacial, su distribución sobre un mapa presenta continuidad geográfica, es decir, formas como de "mancha de aceite". Sin embargo, la simple visualización de una variable no basta y será necesario contrastar este efecto espacial para comprobar su grado de significación estadística.

#### 3.3.1. Diagrama de dispersión de Moran

Una conocida técnica de AEDE es el diagrama de dispersión de Moran. Se trata de un diagrama de dispersión que representa en el eje de abscisas una variable, previamente estandarizada, ( $y_i$ ) y en el eje de ordenadas, su retardo espacial ( $w_{ij} y_i$ ). Este diagrama permite distinguir los dos signos propios del fenómeno de dependencia espacial. Las situaciones de dependencia espacial positiva se corresponden con los cuadrantes I y III, mientras que el fenómeno de dependencia espacial negativa se encuentra en los cuadrantes II y IV.

Por ejemplo, en la FIG. 3(a) se ha elaborado el diagrama de dispersión de Moran utilizando una matriz de pesos espaciales,  $W$ , basada en el criterio de vecindad de la frontera común. En el cuadrante I, se han seleccionado aquellas secciones censales del centro de Madrid con un porcentaje de extranjeros, superior a la media que, a su vez, se encuentran rodeadas por secciones que también disponen de porcentajes de extranjeros superiores a la media de la zona considerada. Estas secciones aparecen destacadas en el mapa de la FIG. 3(b) que, como puede observarse, se localizan fundamentalmente en las secciones del sur y oeste de la "Almendra Central" (distritos Arganzuela, Centro y Tetuán). Por su parte, en el cuadrante III del diagrama de dispersión de Moran se encuentra el mismo fenómeno pero con valores opuestos: secciones con valores bajos en el porcentaje de extranjeros que están rodeadas de secciones también con valores bajos de esta variable. Es decir, el diagrama de dispersión de Moran ofrece una visión más detallada del mapa de la tasa de extranjeros en el centro de la ciudad, poniendo de manifiesto las zonas con mayor o menor concentración de esta población. En cuanto a las categorías de dependencia espacial negativa, éstas vienen dadas por los cuadrantes II y IV, en los que se representan las secciones con valores altos o bajos de porcentaje de extranjeros que están rodeadas por secciones con valores bajos o altos, respectivamente.

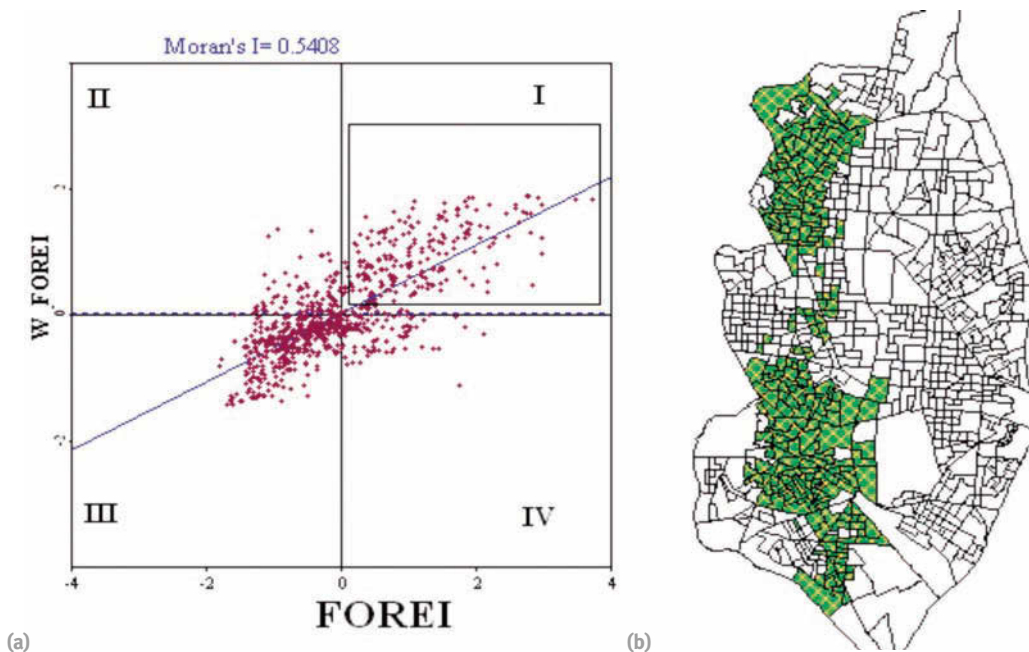


FIG. 3/ (a) Diagrama de dispersión de Moran (b) mapa con selección del cuadrante I del porcentaje de extranjeros en las secciones censales del centro de Madrid

Fuente: Elaboración propia a partir de los datos del Censo 2001 (INE)

### 3.3.2. Estadístico / de Moran

El estadístico *I* de Moran permite conocer el grado de dependencia espacial existente en una variable, es decir, hasta qué punto existe alguna tendencia o concentración de valores de una variable en determinadas zonas de la geografía. Este estadístico se expresa como función de una variable (*Y*) en desviaciones a la media, considerada en los puntos del espacio (*i, j*), y los elementos de una matriz de pesos espaciales *W* que, si está estandarizada por filas, da lugar a la siguiente expresión:

$$I = \frac{n}{S_0} \times \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{\sum_{i=1}^n (y_i - \bar{y})^2} = \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij}^* (y_i - \bar{y})(y_j - \bar{y})}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

donde  $w_{ij}$  es el elemento de la matriz de pesos espaciales *W* correspondiente al par de locali-

zaciones (*i, j*);  $S_0 = \sum_i \sum_j w_{ij}$ , es decir, la suma de los elementos de la matriz de pesos espaciales;  $\bar{y}$  es el valor medio o esperado de la variable *Y*; *n* es el número de observaciones o tamaño muestral y  $w_{ij}^*$  es el elemento de la matriz de pesos espaciales estandarizada por filas. Como puede observarse, cuando la matriz *W* está estandarizada por filas, el estadístico *I* queda reducido al cociente del producto espacial cruzado de los valores de la variable partido por la varianza.

Este estadístico es similar al coeficiente de autocorrelación entre dos variables: el término del numerador es una medida de la covarianza entre valores de la variable *Y* en dos puntos distintos (*i, j*) y el denominador expresa la varianza de *Y* en el punto *i*. En efecto, del estadístico *I* de Moran podría derivarse la siguiente expresión matricial:

$$I = \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij}^* (y_i - \bar{y})(y_j - \bar{y})}{\sum_{i=1}^n (y_i - \bar{y})^2} = \frac{\dot{y}' W^* \dot{y}}{\dot{y}' \dot{y}} \Rightarrow \dot{y}' W^* \dot{y} = I \cdot \dot{y}' \dot{y}$$



siendo  $\hat{y}$  el vector de observaciones en desviaciones a la media y el producto  $W^*\hat{y}$  el retardo espacial asociado a la variable  $Y$  en desviaciones a la media.

En esta expresión, el estadístico  $I$  resulta funcionalmente equivalente al segundo coeficiente a estimar en un análisis de regresión lineal de  $W^*\hat{y}$  sobre  $\hat{y}$  (no de  $\hat{y}$  sobre  $W^*\hat{y}$ , que sería la forma más natural de especificar un proceso autorregresivo espacial). La interpretación del estadístico  $I$  de Moran como *pendiente de una línea de regresión* es lo que permite visualizar la relación lineal entre una variable (en desviaciones a la media) y su correspondiente retardo espacial en forma de diagrama de dispersión bivalente de  $W^*\hat{y}$  sobre  $\hat{y}$  (diagrama de dispersión de Moran). Así, el estadístico  $I$  de Moran puede ser utilizado, por un lado, como indicador del grado de ajuste en la nube de puntos y, por otro, resulta también útil para la detección de posibles valores atípicos de inestabilidad espacial. Cuanto mayor sea el valor de  $I$ , es decir, cuanto mayor sea el ángulo que forme la recta de regresión con el eje de abscisas, más fuerte será el grado de dependencia o autocorrelación espacial en la variable, y viceversa. Obviamente, este estadístico estará muy afectado por aquellos puntos vecinos sensiblemente distintos de la media de la variable en estudio (valores atípicos).

En la FIG. 3(a), sobre la nube de puntos, se incluye el valor del estadístico  $I$  de Moran (*Moran's I* = 0,5408), que resume el ajuste de la variable espacialmente retardada del porcentaje de extranjeros ( $W\_FOREI$ ) sobre el porcentaje de extranjeros ( $FOREI$ ). El valor del estadístico  $I$  coincide con el valor de la pendiente del siguiente modelo de regresión estimado por mínimos cuadrados ordinarios, cuyo término independiente es nulo (por estar las variables expresadas en desviaciones a la media):

$$W\_FOREI = 0,5408 \cdot FOREI$$

El resultado de este ajuste permitiría conocer las consecuencias que tendría en la zona centro de Madrid un incremento en el porcentaje de extranjeros, debido al "efecto llamada" (autocorrelación espacial) que suele producirse en estos casos. Por ejemplo, si Ayuntamiento se plantea realojar a un grupo significativo de extranjeros en viviendas de protección oficial, las autoridades locales sabrán que, por cada porcentaje añadido de extranjeros en una de las secciones censales del centro se producirá, por término medio, un incremento del 54% del porcentaje de extranjeros en las secciones contiguas.

A continuación, se nos plantea saber hasta qué punto el valor estimado para la pendiente es o no estadísticamente significativo (o distinto de cero). Para ello, hay que construir un test estadístico en el que la hipótesis nula será ausencia de autocorrelación espacial o, lo que es lo mismo, la nulidad del parámetro de la pendiente de la regresión en la expresión (5). El *contraste de bondad del estadístico I*, en cuanto pendiente de un modelo de regresión, no debe realizarse con el conocido test  $t$ , basado en la distribución  $t$  de Student, ya que el estadístico  $I$  de Moran no se encuentra centrado en el valor cero (vide CLIFF & ORD, 1981: 43-44). De hecho, la media teórica del estadístico  $I$  de Moran es el cociente:  $-1/(n - 1)$ . En otras palabras, el valor esperado del estadístico  $I$  de Moran mayor que su valor esperado sería indicativo de autocorrelación espacial positiva, mientras que un valor de  $I$  inferior a la media pondría de manifiesto la existencia de autocorrelación espacial negativa.

ANSELIN (1988) demuestra que, cuando la variable objeto de estudio no se distribuye como una normal, cosa muy habitual en la práctica, debemos aproximar los momentos del estadístico  $I$  de forma empírica, siguiendo el *método de las permutaciones*, que es muy habitual en la estadística moderna, siendo la hipótesis nula  $I = 0$ . En el caso de la FIG. 3, en que  $I = 0,5408$ , se trata de calcular el nivel de confianza que existe para el rechazo de esa hipótesis en favor de la hipótesis alternativa de autocorrelación espacial positiva. Tras la generación de 999 muestras diferentes con los datos de la variable (porcentaje de extranjeros en las secciones censales del centro de Madrid), la media de los valores obtenidos para el estadístico  $I$  en cada muestra es  $E[I] = -0,0013$  y la desviación típica  $DT(I) = 0,0122$ . Según esta distribución de valores del estadístico  $I$ , el valor obtenido para el porcentaje de extranjeros de Madrid es muy superior al resto de valores de  $I$  calculados en configuraciones de datos diferentes de la original. En concreto, se obtiene un  $p$ -valor = 0,001, lo que implica que es posible rechazar la hipótesis nula en favor de la hipótesis alternativa de autocorrelación espacial positiva con casi el 100% de confianza  $(1 - 0,001) * 100$ . Este elevado nivel de confianza avalaría, por ejemplo, las conclusiones que puedan derivarse de la expresión (6).

#### 4. Detección de “clusters” y otras estructuras espaciales

El estadístico  $I$  de Moran es una medida de dependencia espacial global, en cuanto que se refiere a todo el conjunto geográfico sometido al análisis. Por eso, no es capaz de detectar la presencia de zonas concretas de inestabilidad (zonas calientes o frías), en las que una variable tenga un comportamiento diferente de la seguida por la tendencia general. Cuando esto se produce, decimos que la variable es no estacionaria, es decir, que la intensidad (y el signo) de la dependencia espacial es significativamente diferente en algunas subzonas del espacio global considerado.

La hipótesis de estacionariedad es la base sobre la que se asienta el análisis estadístico de la autocorrelación espacial. En el caso de existencia de dependencia espacial local, este supuesto se incumple y, por eso, el análisis global no sería válido. El fenómeno de la inestabilidad espacial o dependencia espacial local puede plantearse desde dos puntos de vista:

- En primer lugar, es posible que el estadístico  $I$  de Moran no detecte la presencia de dependencia espacial en la distribución de una variable sobre un espacio dado, pero que en su lugar, existan *clusters* o agrupamientos de valores similares (o diferentes) en ciertas sub-zonas, que sean de interés para el análisis.
- En segundo lugar, también sería posible que el test  $I$  detectara la presencia de dependencia espacial en una variable, pero que no todas las unidades del espacio considerado contribuyan con igual peso en dicho indicador global.

Es decir, dentro del espacio considerado, pudieran coexistir zonas en las que una variable no esté espacialmente autocorrelacionada con otras zonas en las que el fenómeno de dependencia espacial sea muy importante. A continuación, se presentan algunas herramientas del AEDE que permiten detectar este tipo de estructuras.

##### 4.1. Puntos atípicos en el diagrama de dispersión de Moran

El diagrama de dispersión de Moran es capaz de ofrecer no sólo una impresión visual global de la existencia de dependencia espacial, sino también una perspectiva más desagregada de autocorrelación espacial, en cuanto positiva (aquellos puntos situados en los cuadrantes

primero y tercero) y negativa (cuadrantes segundo y cuarto). Además, como los valores de la variable  $Y$  están estandarizados, es posible obtener información de los valores atípicos, es decir, aquéllos situados a más de dos unidades del origen (según el criterio estadístico “2-sigma”). Estos valores atípicamente altos/bajos podrían corresponderse con estructuras de dependencia espacial local y/o con *puntos atípicos*, en cuanto valores que no se ajustan bien a las estructuras de dependencia espacial de la variable.

En la FIG. 4, hemos destacado los valores atípicos del tercer cuadrante del diagrama de dispersión de Moran, en cuanto aquellas secciones censales del centro de Madrid claramente alejadas de la nube principal de puntos. Se trata de secciones con un claro agrupamiento espacial de valores inferiores de la variable (porcentaje de edificios construidos antes de 1970). Como puede observarse, se trata de secciones localizadas en los extremos norte, este y sur de la “Almendra Central”, cercanas a la M-30. Estas secciones censales forman un “cluster” espacial caracterizado por la concentración de edificaciones más modernas (menos antiguas), respecto del conjunto de edificios de la zona centro de Madrid.

##### 4.2. El estadístico $I_i$ local de Moran y el mapa de “clusters”

El estadístico  $I_i$  local de Moran calcula, para cada unidad  $i$ , la existencia de un agrupamiento significativo de valores similares (altos o bajos) alrededor de dicha unidad geográfica del modo siguiente (ANSELIN, 1995):

$$I_i = \dot{y}_i \sum_{j \in J_i} w_{ij} \dot{y}_j$$

siendo  $\dot{y}_i$ ,  $\dot{y}_j$  los valores de la variable  $Y$ , en desviaciones a la media, en las unidades espaciales  $i, j$ ;  $w_{ij}$  son los elementos de la matriz de pesos espaciales  $W$ ;  $J_i$  es el conjunto de unidades relacionadas con  $i$ .

Como se deriva de esta expresión, un valor positivo de  $I_i$  indica la existencia de un agrupamiento espacial de valores parecidos (o bien altos o bien bajos) alrededor de la unidad espacial  $i$ , mientras que un valor negativo indica el agrupamiento de valores muy diferentes (altos junto a bajos o bajos junto a altos) alrededor de  $i$ . También puede comprobarse que la suma de estadísticos locales  $I_i$  es el estadístico global  $I$  de Moran:

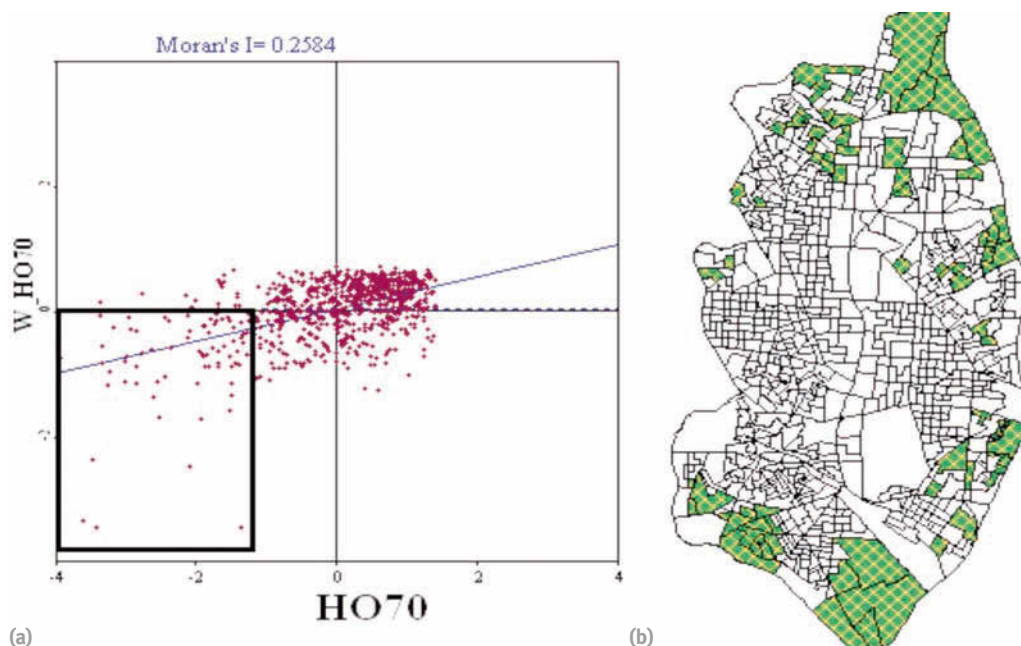


FIG. 4/ Diagrama de dispersión de Moran del porcentaje de edificios construidos antes de 1970 en las secciones censales de Madrid centro: (a) selección de valores atípicos en el tercer cuadrante, (b) mapa con selección

Fuente: Elaboración propia a partir de los datos del Censo 2001 (INE)

$$\sum_{i=1}^n I_i = \sum_{i=1}^n \dot{y}_i \sum_{j \in J_i} w_{ij} \dot{y}_j = \sum_{i=1}^n \sum_{j \in J_i} w_{ij} \dot{y}_i \dot{y}_j$$

Esta suma también podría expresarse como una función lineal del test  $I$  global:

$$\sum_{i=1}^n I_i = \frac{1}{S_0 m_2} I = \gamma \cdot I$$

donde  $S_0 = \sum_i \sum_j w_{ij}$ ;  $\gamma = S_0 m_2$  un factor de proporcionalidad, para:

$$m_2 = \sum_i \dot{y}_i^2$$

Como en el caso del estadístico  $I$  global, aquí también es necesario realizar un proceso inferencial, basado en el enfoque de las permutaciones, para conocer el nivel de significación estadística de cada indicador local  $I_i$  como “máximo contribuyente” al valor único del indicador global  $I$ . La representación gráfica de los índices  $I_i$  da lugar al *mapa LISA*<sup>3</sup> de “clusters” espaciales (*LISA Cluster Map*). La representación de los valores significativos  $I_i$  se lleva a

cabo a través de distintos colores, uno para cada cuadrante del diagrama de dispersión de Moran. De este modo, es posible saber alrededor de cuál de los puntos del diagrama se produce una concentración espacial significativa de datos parecidos o diferentes de una variable, así como el signo de la dependencia espacial.

En la FIG. 5, se ha representado el mapa LISA de clusters correspondiente a la variable de densidad del PIB (PIB por kilómetro cuadrado en 2006), de las NUT-3 de la Unión Europea (UE)<sup>4</sup>. En color rojo (valores altos rodeados de valores altos), este mapa pone de manifiesto la existencia de un núcleo duro de concentración de la producción localizado en las zonas urbanas policéntricas de Alemania (ejes Rhine-Ruhr y Rhine-Main-Neckar), en el sur de Inglaterra e Ile-de-France. Aunque posiblemente el hallazgo más impactante de este mapa es la existencia de una extensa periferia, es decir, un gran “cluster” de valores bajos en la densidad de la producción. Se trata de las periferias norte (regiones aisladas de Irlanda, Escocia, países escandinavos y bálticos), sur (Grecia y regiones interiores de España y Portugal) y este de la UE.

<sup>3</sup> LISA es el acrónimo de *Local Indicators of Spatial Association*.

<sup>4</sup> NUT: son las siglas en francés de la Nomenclatura de las

Unidades Territoriales utilizadas por la Unión Europea con fines estadísticos. En España, las NUT-3 coinciden con las provincias.

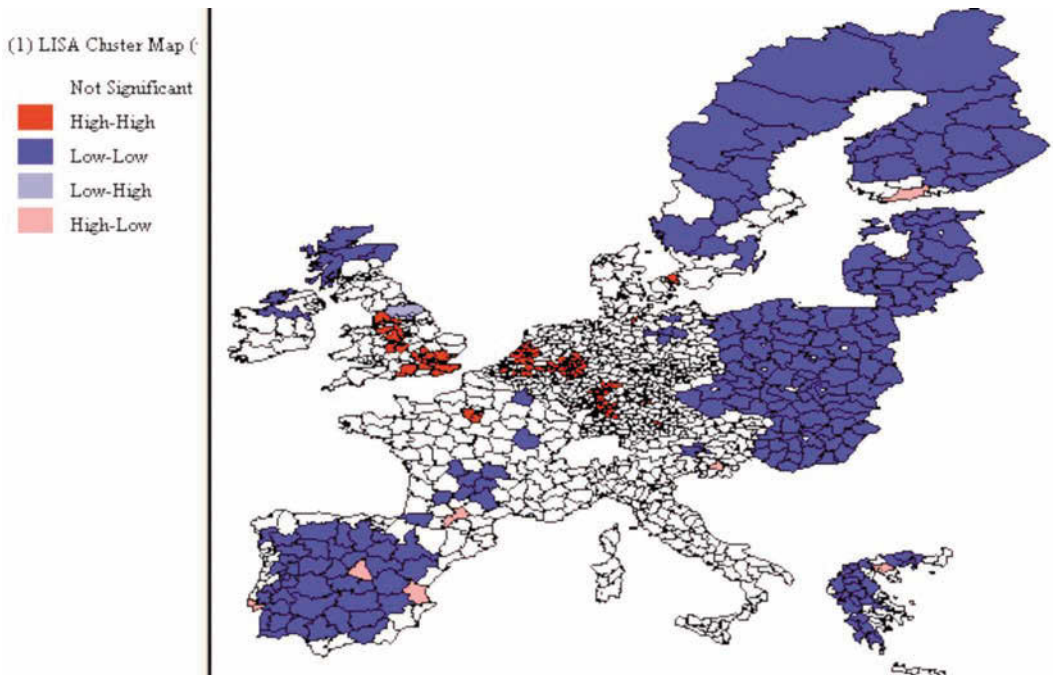


FIG. 5/ Mapa LISA de la densidad del PIB en las regiones NUT3 de la UE

Fuente: Elaboración propia a partir de la "European Regional Database" (Cambridge Econometrics)

### 4.3. El estadístico local $G_i(d)$

GETIS & ORD (1992) sugirieron un enfoque diferente a partir de los estadísticos de distancias, que requieren de la definición de una distancia crítica  $d$  en torno a cada punto  $i$ . De este modo, si se definen diferentes bandas de distancia para cada punto, deben también construirse distintas matrices de pesos espaciales para cada radio de distancia,  $W(d)$ . El estadístico  $G_i(d)$  permite contrastar la hipótesis alternativa de existencia de autocorrelación espacial en el conjunto de valores de la variable  $Y$  asociados a los  $j$  puntos incluidos en un radio  $d$  de de distancia en torno a un punto inicial dado  $i$ .

$$G_i(d) = \frac{\sum_{j=1}^N w_{ij}(d)y_j}{\sum_{j=1}^N y_j} \quad ; \quad \text{para } j \neq i$$

donde  $[w_{ij}(d)]$  son los elementos de una matriz simétrica y binaria (0-1), tal que  $w_{ij} = 1$  para todos los vínculos existentes entre aquellos puntos  $j$  situados en un radio  $d$  del punto  $i$ , y  $w_{ij} = 0$  para el resto de vínculos, incluida la relación del punto  $i$  consigo mismo. En el numerador se

consideran únicamente los valores  $y_j$  situados en un radio  $d$  del punto  $i$  (excepto  $y_i$ ) mientras que, en el denominador, se consideran todos los valores de la variable, excepto  $y_i$ . Obsérvese que, para cada punto del espacio  $i$ , se obtiene un valor de  $G_i(d)$  que, dependiendo de la distancia considerada, será también distinto.

Una variante de este test sería el estadístico Nuevo  $G_i^*$ , que mide la asociación espacial en aquellos casos en los que se también se considera la relación  $i = j$ . Tanto éste como el anterior, son estadísticos que sólo pueden ser aplicados al caso de variables naturales (no enteras) positivas, así como para matrices  $W$  binarias y simétricas (por tanto, no estandarizadas por filas). Para superar estas limitaciones, ORD & GETIS (1995, 2001) presentaron, más tarde, nuevas variantes de estos estadísticos.

Los resultados del cálculo del Nuevo  $G_i^*$  se presentan en la FIG. 6. Aunque los resultados se parecen mucho a los ofrecidos por el mapa LISA, el detalle del nivel de significación asociado a cada uno de los valores locales del test nos permite concluir que la máxima concentración, en términos de densidad de producción, no se produce en los valores elevados de la variable, sino en los inferiores (las periferias de la UE).



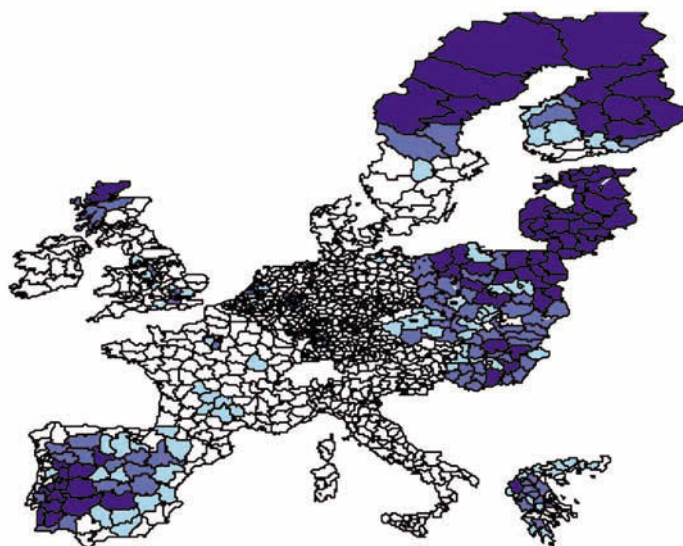


FIG. 6/ Mapa LISA de la densidad del PIB en las regiones NUT3 de la UE

Fuente: Elaboración propia a partir de la "European Regional Database" (Cambridge Econometrics)

#### 4.4. Fenómenos de difusión y concentración espacio-temporal

Cuando se dispone de un corte temporal para una base de datos geográficos, ambos pueden combinarse y trabajar con el panel completo. De este modo, pueden abordarse temas como el crecimiento o convergencia (en el ámbito de la economía regional y urbana) o el lanzamiento de nuevos productos y su difusión a través del espacio geográfico (en geomarketing).

El diagrama de dispersión bivalente de Moran representa una variable  $X$  en el eje horizontal y el retardo espacial de otra variable  $Y$  (con ayuda de una matriz de pesos espaciales) en el eje vertical. Cuando en el eje horizontal se representan los valores de la variable  $Y$  referida a un momento del tiempo  $t$  ( $y_{i,t}$ ) y en el eje vertical se representan los valores de dicha variable  $Y$  espacialmente retardada, y referida a otro momento del tiempo  $t-k$  ( $Wy_{i,t-k}$ ), entonces se trata de un caso particular, que podría denominarse *diagrama de dispersión espacio-temporal de Moran*. En este caso, la pendiente de la recta ajustada a la nube de puntos será el *estadístico I espacio-temporal de Moran* (CHASCO Y LÓPEZ, 2008), que vendrá definido como:

$$STI = I_t = \frac{Z'_{t-k} W^* z_t}{Z'_{t-k} z_{t-k}}$$

siendo  $z_t$ ,  $z_{t-k}$  los valores de variable  $Y$  estandarizada referidos a los momentos  $t$  y  $(t - k)$ , respectivamente y  $W^*$  la matriz de pesos espaciales estandarizada por filas. Por tanto, el estadístico STI (en inglés, *Space-Time I*) mide la influencia que produce un cambio en una variable en un lugar del espacio  $i$  y un momento del tiempo pasado  $(t - k)$  sobre su entorno geográfico en el presente ( $t$ ). El valor del estadístico STI es la pendiente de la línea de regresión de  $Wz_t$  sobre  $z_{t-k}$ . Cuando el retardo temporal es nulo ( $k = 0$ ), este índice coincide con el índice  $I$  global de Moran. La inferencia sobre el índice STI se puede calcular como en el caso univariante, siguiendo el enfoque de permutaciones. Si el estadístico STI resulta significativo en una variable espacial para varios desfases temporales ( $k = 1, 2, \dots$ ) puede concluirse que en dicha variable existe un proceso de difusión espacio-temporal, es decir, *dependencia espacial no contemporánea*.

El test STI permite conocer la existencia de *agrupamientos espaciales en el proceso de crecimiento* de un fenómeno. Por ejemplo, si la variable fuera la evolución de la renta per cápita en un período temporal, la detección de *clusters* de tipo espacio-temporal demostraría la existencia de nichos o "clubes" permanentes de desigualdad económica en el espacio total. En estos casos, el índice  $I$  global (univariante) de Moran se irá incrementando con el paso del tiempo produciendo una dinámica divergente de crecimiento economi-



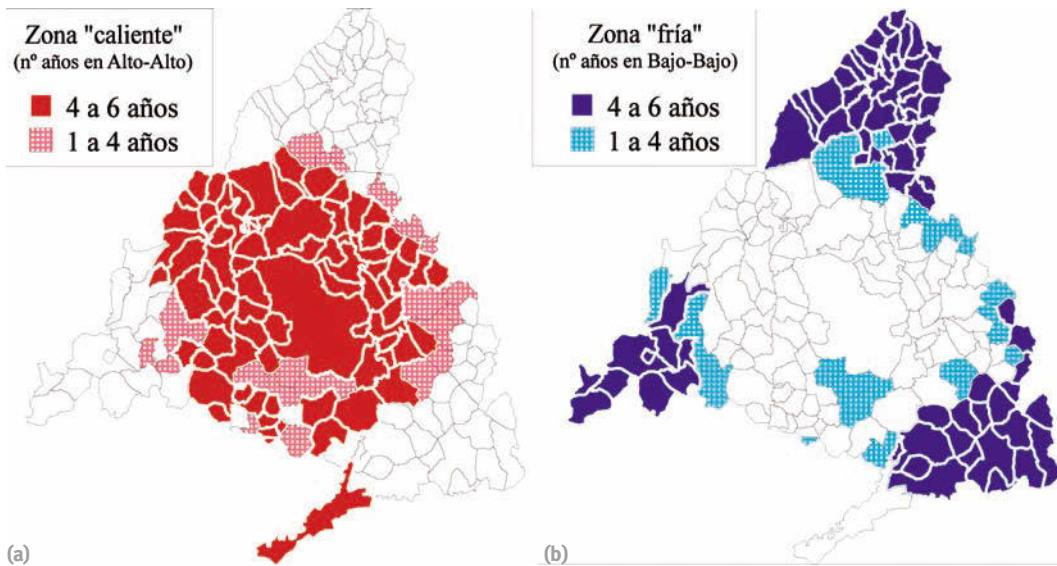


FIG. 7/ Mapas LISA de clusters de la tasa de líneas ADSL por habitante de los municipios de la Comunidad de Madrid: (a) zona caliente (cluster de valores elevados de la variable), (b) zona fría (cluster de valores bajos)

Fuente: Elaboración propia a partir del Instituto Lawrence R. Klein (Universidad Autónoma de Madrid)

co entre las distintas unidades espaciales (países, regiones, etc). Si la variable cuya evolución se analiza fuera la implantación de un producto de consumo, como las líneas ADSL a finales del siglo XX, la disminución del grado de dependencia espacial con el paso de los años demostraría la existencia de un proceso de reparto equitativo de dicho bien en el territorio (convergencia entre las unidades geográficas).

Siguiendo la lógica de apartados anteriores, será posible obtener herramientas locales de análisis espacio-temporal, como el diagrama de dispersión espacio-temporal de Moran y el mapa LISA espacio-temporal de "clusters". En la FIG. 7, se analiza la tasa de las líneas ADSL por habitante de los municipios de la Comunidad de Madrid, con objeto de conocer la existencia de fenómenos de difusión espacial en el período de implantación de este nuevo producto (años 2000-2006). En la FIG. 7(a), se presenta en color rojo la asociación "alto-alto", es decir, aquéllos municipios que, en un número determinado de años, han registrado tasas de ADSL per cápita superiores a la media regional, cumpliendo, a su vez, la condición de estar rodeado de municipios que, por término medio, también tenían tasas de ADSL per cápita superiores a la media. Estos municipios constituyen una zona "caliente" de negocio de este producto, situada en Madrid capital y las periferias norte y oes-

te, sobre todo (6 ó 7 años en situación "alto-alto"), con algunos municipios del Corredor del Henares. Por su parte, en el mapa de la FIG 7(b) se representa el número de veces que un municipio ha tenido una tasa de ADSL per cápita inferior a la media regional habiendo estado también rodeado de municipios con tasas inferiores a la media regional (asociación "bajo-bajo"). Se trata ésta de una zona "fría" para el negocio del ADSL, donde aún podría existir un mercado sin conquistar. Se trata de los municipios rurales de la Sierra Norte y esquinas este y oeste de la Comunidad de Madrid.

## 5. Conclusión

En este artículo, se han presentado las principales técnicas del AEDE, al menos, las más utilizadas en el ámbito de la econometría espacial. Tal y como se ha expuesto en las páginas anteriores, el análisis de los fenómenos geográficos no puede llevarse a cabo sin la ayuda de un mapa, así como de gráficos y estadísticos especialmente diseñados para el tratamiento de los datos espaciales. El carácter de multidireccionalidad de las relaciones en el territorio, y los efectos espaciales de dependencia y heterogeneidad, son las principales características que diferencian las variables espaciales de las series temporales y de otras distribuciones de corte transversal.

Sin embargo, en muchas ocasiones, la sola utilización de mapas digitales o gráficos estadísticos básicos no suele ser suficiente para lograr una buena detección de “clusters” y estructuras subyacentes en los datos. Para ello, el AEDE provee de estadísticos más avanzados, como el estadístico  $I$  de Moran, los gráficos de dispersión de Moran, los tests de dependencia local y estadísticos espacio-temporales. Todos

ellos, y otros más, que no se han podido detallar en este trabajo, se han ido formulando, desde distintas disciplinas, como respuesta a problemas comunes propios de la ciencia regional y el análisis urbano. Esperamos haber ofrecido al lector interesado algo más de luz sobre estos aspectos, así como haberle también proporcionado las referencias oportunas para un estudio futuro más en profundidad.

## 6. Bibliografía

- ANSELIN, L. (1988): *Spatial econometrics: Methods and Models*, Kluwer Academic Publishers.
- (1995): “Local indicators of spatial association-LISA”, en *Geographical Analysis*, 27: 93-115.
- (1999): “The future of spatial analysis in the social sciences”, en *Geographic Information Sciences*, 5: 2: 67-76.
- (2001a): “Spatial econometrics”, en *A companion to theoretical econometrics*: 310-330, Basil Blackwell, Oxford.
- (2001b): “Spatial effects in econometric practice in environmental and resource economics”, en *American Journal of Agricultural Economics*, 83: 3: 705-710.
- (2002): *SpaceStat Tutorial: A Workbook for Using SpaceStat in the Analysis of Spatial Data*. University of Illinois, Urbana-Champaign Urbana, IL.
- (2005): *Exploring spatial data with GeoDa<sub>TM</sub>: a workbook*. Spatial Analysis, Laboratory y Center for Spatially Integrated Social Science, Urbana-Champaign.
- (2006): “How (not) to lie with spatial statistics”, en: *American Journal of Preventive Medicine*, 30: 2S: 3-6, Elsevier Inc.
- ANSELIN, L. & R. J. G. M. FLORAX & Sergio J. REY (eds.) (1995): *New directions in spatial econometrics*. Springer-Verlag, Berlín.
- ANSELIN, L. & R. J. G. M. FLORAX & S. REY (eds.) (2004), *Advances in spatial econometrics. Methodology, tools and applications*. Springer-Verlag, Berlín.
- ANSELIN, L. & N. LOZANO-GRACIA (2008): “Errors in variables and spatial effects in hedonic house price models of ambient air quality”, en *Empirical Economics*, 34: 5-34.
- ANSELIN, L. & S. REY, eds. (2010): *Perspectives on spatial data analysis*. Springer-Verlag, Series “Advances in Spatial Sciences”, Berlín.
- ANSELIN, L. & A. VARGA & Z. ACS (2000): “Geographic and sectorial characteristics of academic knowledge externalities”, en *Papers in Regional Science*, 79: 4: 435-443.
- ARBIA, G (2006): *Spatial Econometrics: Statistical Foundations and Applications to Regional Convergence (Advances in Spatial Science)*, Spriger, Berlín.
- BALLER, R. D. & L. ANSELIN & S. F. MESSNER & G. DEANNE & D. F. HAWKINS (2001): “Structural covariates of U.S. county homicide rates: incorporating spatial effects”, en *Criminology*, 39: 3: 561-588.
- CECCATO, V. & R. HAINING & P. SIGNORETTA (2002): “Exploring offence statistics in Stockholm city using spatial analysis tools”, en *Annals of the American Association of Geographers*, 92: 1: 29-51.
- CHASCO, C. & G. FERNÁNDEZ-AVILÉS (2009): *Análisis de datos espacio-temporales para la economía y el geomarketing*, NetBiblo, A Coruña
- CHASCO, C. & F. LÓPEZ (2008): “Is spatial dependence an instantaneous effect? Some evidence in economic series of Spanish provinces”, en *Estadística Española*, 50: 167: 101-118.
- CLIFF, A. D. & J. K. ORD (1973): *Spatial autocorrelation*, Pion, London.
- (1981): *Spatial processes, models and applications*, Pion, London.
- COX, K. (1969): “The voting decision in a spatial context”, en *Progress in Geography*, 1: 81-117: Edward Arnold, London.
- FISCHER, M. M. (2006): “Spatial analysis and geocomputation: selected essays” in Manfred M. FISCHER Springer,
- & A. GETIS (eds.) (1997): *Recent developments in spatial analysis*, Springer-Verlag, Series “Advances in Spatial Sciences”, Berlín.
- (eds.) (2010): *Handbook of applied spatial analysis*, Springer-Verlag, Berlín.
- FISCHER, M. M. & A. VARGA (2003): “Production of knowledge and geographically mediated spillovers from universities: special econometric perspective and evidence from Austria”, en *Annals of Regional Science* 37: 2: 303-322.
- GETIS, A. & J. K. ORD (1992): “The analysis of spatial association by use of distance statistics”, en *Geographical Analysis*, 24: 189-206.
- GRIFFITH, D. (2006): “Hidden negative spatial autocorrelation”, en *Journal of Geographical Systems*, 8: 4: 335-355.
- HAINING, R. & S. WISE & P. SIGNORETTA (2000): “Providing scientific visualization for spatial data analysis: Criteria and an assessment of SAGE”, en *Journal of Geographical Systems*, 2: 121-140.
- JENKS, G. F. & F. C. CASPALL (1971): “Error on choroplethic maps: definition, measurement, reduction”, en *Annals of the Association of American Geographers* 61: 2: 217-244.
- JIANG, B & X. YAO (eds.) (2010): *Geospatial analysis and modelling of urban structure and dynamics*. Springer, Dordrecht, Heidelberg, London, New York.
- LE GALLO, J. & C. CHASCO (2008): “Spatial analysis of urban growth in Spain, 1900-2001”, en *Empirical Economics*, 34: 1: 59-80.
- PÁEZ, A. & J. LE GALLO & R. BULIUNG & S. DALL'ERBA (eds.) (2010): *Progress in spatial analysis*. Springer-Verlag, Series “Advances in Spatial Sciences”, Berlín.

- ORD, J. K. & A. GETIS (1995): "Local spatial autocorrelation statistics: distributional issues and an application", en *Geographical Analysis*, 27.4: 286-306.
- (2001): "Testing for local spatial autocorrelation in the presence of global autocorrelation", en *Journal of Regional Science*, 41(3): 411-432.
- REY, S. & B. MONTOURI (1999): "U.S. regional income convergence: A spatial econometric perspective. *Regional Studies*, 33, 143-156.
- SMITH, M. J & M. F. GOODCHILD & P. A. LONGLEY (eds.) (2007): *Geospatial analysis*.
- UNWIN, A. (2000): "Using your eyes- making statistics more visible with computers", en *Computational Statistics & Data Analysis*, 32: 303-312.