

MÉTODOS GRÁFICOS DEL ANÁLISIS EXPLORATORIO DE DATOS ESPACIALES

Chasco Yrigoyen, Coro

E-mail: coro.chasco@uam.es

Instituto L.R. Klein-Dpto. de Economía Aplicada

Universidad Autónoma de Madrid

Palabras clave: Análisis exploratorio de datos espaciales, econometría espacial, autocorrelación espacial, heterogeneidad espacial, enfoque reticular “lattice”, GIS.

Abstract

El Análisis Exploratorio de Datos Espaciales (AEDE) constituye un campo muy novedoso e interesante para el análisis económico y de las ciencias sociales, en general. El AEDE podría definirse como el grupo de herramientas estadístico-gráficas que describen y visualizan las distribuciones espaciales, identificando localizaciones atípicas, descubriendo formas de asociación (autocorrelación espacial) que, a su vez, pueden ser de carácter global o local, y sugiriendo estructuras en el espacio geográfico (heterogeneidad espacial). El AEDE puede ser aplicado desde una perspectiva univariante, sobre uno o varios indicadores geográficos, o multivariante, como una etapa propia del proceso econométrico espacial, previa tanto al análisis confirmatorio (estimación y contrastes) como a todo ejercicio de predicción espacial. Cuando el AEDE es abordado desde los presupuestos conceptuales de la econometría espacial, los métodos utilizados siguen el esquema reticular (“lattice”) propio de esta disciplina, que se diferencian de la perspectiva geoestadística. En este documento, se realiza una revisión de la literatura existente sobre los principales métodos gráficos del AEDE reticular o “lattice” univariante, con algunos ejemplos ilustrativos, así como comentarios y propuestas de futuro en este campo.

1. INTRODUCCIÓN

En muchas ocasiones, el análisis de datos se realiza con el objetivo de contrastar determinadas hipótesis o modelos como, por ejemplo, cuando se quiere conocer a través de una encuesta si la edad de las personas está relacionada positivamente con su capacidad de asumir riesgos o si, en situaciones más complicadas, se desea conocer el tipo de relación existente entre un determinado número de variables (lineal o no lineal). En estos casos, el investigador no dispone de una hipótesis definida a contrastar, la naturaleza del estudio es “exploratoria” y se deben utilizar ciertas técnicas especialmente diseñadas para encontrar algún “sentido” en los datos, es decir, para la detección en ellos de estructuras o conglomerados (“clusters”).

En general, el análisis exploratorio de datos, a diferencia de los contrastes de hipótesis tradicionales diseñados para la verificación a priori de conexiones entre variables, se utiliza para identificar **relaciones sistemáticas** entre variables cuando no existen expectativas claras sobre la naturaleza de estas relaciones.

Cuando se trabaja con series geográficas o datos espaciales deben considerarse explícitamente las características propias de esta información: **georreferenciación, multidireccionalidad y multidimensionalidad**. Efectivamente, el espacio geográfico tiene una naturaleza georreferenciada que exige conocer la posición relativa o absoluta donde se producen los fenómenos que se están analizando, siendo las relaciones que se establecen entre las distintas observaciones espaciales de tipo multidireccional, a diferencia de la unidireccionalidad (pasado-presente-futuro) de las relaciones entre los datos temporales.

Estas peculiaridades de los datos geográficos suelen producir en los mismos los fenómenos, conocidos como **efectos espaciales**, de dependencia o autocorrelación espacial y heterogeneidad espacial. La dependencia espacial podría ser definida como la existencia de una relación funcional entre lo que ocurre en un punto determinado del espacio y lo que ocurre en otro lugar (Moreno y Vayá, 2000). En cuanto a la heterogeneidad espacial, se trata de un efecto relacionado con la diferenciación espacial o regional y viene definido por la ausencia de estabilidad en el espacio del comportamiento humano o de otras relaciones en estudio.

Esta peculiaridad de los datos espaciales motiva el nacimiento de una subdisciplina del análisis exploratorio de datos denominada Análisis Exploratorio de Datos Espaciales (AEDE)¹, diseñada para el tratamiento específico de los datos geográficos.

El AEDE puede ser aplicado desde una perspectiva univariante, sobre una (o varias) variable geográfica, o como una etapa propia del proceso econométrico espacial, previa tanto al análisis confirmatorio (estimación y contrastes), como a todo ejercicio de predicción espacial (Anselin, 2001B; Chasco, 2003). Por otro lado, cuando el AEDE es abordado desde los presupuestos conceptuales de la econometría espacial, los métodos utilizados siguen el **esquema reticular o “lattice”**, llamado así porque es más adecuado para el tratamiento de unidades poligonales (reticulares) continuas, que son las propias de la econometría espacial, a diferencia de la geoestadística, que suele trabajar con puntos (no polígonos) en el espacio (unidades espaciales discretas).

En este documento, se realiza una revisión de la literatura existente sobre los principales métodos del AEDE reticular o “lattice” univariante, con algunos ejemplos ilustrativos. Para ello, en el siguiente apartado, se presenta el concepto de AEDE como parte del más general análisis exploratorio de datos, así como algunas importantes aplicaciones y los paquetes informáticos especializados en el desarrollo de estas técnicas. En el capítulo tercero, se desarrollan los principales métodos gráficos del AEDE univariante, con ilustraciones y ejemplos que ayudan a su comprensión. Finalmente, el apartado de bibliografía cierra este trabajo.

2. ANÁLISIS EXPLORATORIO DE DATOS ESPACIALES (AEDE)

El AEDE se considera como una sub-disciplina perteneciente al denominado **Análisis Exploratorio de Datos (AED)**, que consiste en el tratamiento y comparación de una batería de variables utilizando técnicas que identifican en ellas formas estables. Según Tukey (1977), autor que hizo posible la extensión de este tipo de análisis multivariante y del “software” estadístico moderno, el AED podría definirse como “el conjunto de herramientas gráficas y descriptivas utilizadas para el descubrimiento de patrones de

¹ Para todas las referencias en lengua inglesa, este término es conocido como “Exploratory Spatial Data Analysis” o por sus siglas iniciales (ESDA).

comportamiento en los datos y el establecimiento de hipótesis con la menor estructura posible”.

De este modo, es posible obtener una estructura explicativa de los datos a través de unas técnicas que combinan herramientas de la estadística básica (descriptivos, correlaciones, tablas de frecuencias o de correlación cruzada) con análisis multivariante avanzado, especialmente diseñado para identificar formas en grandes bases de datos (análisis cluster, escalas multidimensionales, análisis logit, correlación canónica, análisis de correspondencias, árboles de clasificación, etc.).

El AEDE, en palabras de Anselin (1999), puede definirse como el conjunto de técnicas que describen y visualizan las distribuciones espaciales, identifican localizaciones atípicas o “atípicos espaciales” (“spatial outliers”), descubren esquemas de asociación espacial, agrupamientos (“clusters”) o puntos calientes (“hot spots”) y sugieren estructuras espaciales u otras formas de heterogeneidad espacial. Por tanto, estos métodos tienen un carácter descriptivo (estadístico) más que confirmatorio (econométrico), aunque la detección de estructuras espaciales en las variables geográficas, hace posible la formulación de hipótesis previas para la modelización econométrica y, en su caso, posible predicción espacial de nuevos datos.

Es decir, el AEDE debe constituir la fase previa a toda modelización econométrica espacial, sobre todo cuando no exista un marco formal o teoría previa acerca del fenómeno que se pretende explicar, como es el caso de algunos análisis interdisciplinares realizados en las ciencias sociales, sobre todo en el campo de la economía regional, así como en los ejercicios de predicción-extrapolación de datos, en los que puede no existir una identificación entre relaciones económicas establecidas en diversos ámbitos o escalas territoriales (Arbia, 1989).

Tal como se desprende de los últimos trabajos realizados con AEDE (ver Tabla 1), los fenómenos analizados dentro del campo de las ciencias sociales abarcan temas propios de la economía regional, como la convergencia regional en términos de renta (Bosch *et al.*, 2002; Rey, 2001; Chasco y Vicéns, 2000; Moreno y Vayá, 2000) o los efectos “spillover” en los modelos de datos espaciales (Bosch *et al.*, 2002; Yilmaz *et al.*, 2002; Van der Kruk, 2001), entre otros.

Tabla 1. Algunos ejemplos recientes de AEDE aplicado a las ciencias sociales

Materia	Fenómeno explorado
Criminología	Ceccato <i>et al.</i> (2001), Baller <i>et al.</i> (2001)
Economía regional	Bosch <i>et al.</i> (2002), Gezici y Hewings (2002), Yilmaz <i>et al.</i> (2002), Button y Kulkarni (2001), López <i>et al.</i> (2001), Rey (2001), Stough <i>et al.</i> (2001), Van der Kruk (2001), Chasco y Vicéns (2000), LeSage y Kelley (2000), Moreno y Vayá (2000),
Educación	Fotheringham <i>et al.</i> (2001)
Sanidad	Wall y Devine (2000)

Fuente: Elaboración propia.

Además, algunos autores han avanzado en el AEDE integrado en el entorno de los Sistemas de Información Geográfica (GIS), como Anselin *et al.*, 2002, Anselin, 1998, Anselin y Bao, 1997; Wise *et al.*, 1997), así como en la solución de problemas aún pendientes de resolver (Anselin, 2001A).

En los últimos años, los esfuerzos desarrollados por la investigación en este campo del AEDE se han venido centrando en la conexión de los GIS disponibles en el mercado (Grassland, ESRI, MapInfo) con paquetes estadísticos tradicionales (S-Plus, SPSS) o específicos de AEDE. De este modo, a la potente capacidad de visualización y análisis de los GIS se le une la especialización propia del análisis espacial estadístico y gráfico (ver Bao *et al.*, 2000).

En la Tabla 2, se presentan algunos importantes programas de “software” relacionados con la perspectiva de retícula o “lattice”, es decir, el análisis de localizaciones geográficas discretas de puntos o polígonos².

Debe destacarse la marcada orientación hacia el análisis econométrico espacial (exploratorio y confirmatorio) del “software” SpaceStat y sus extensiones para AcView (Anselin). Otros programas, como Econometrics Toolbox (LeSage), Sage (Ma, Haining y Wise), SpaceStatPack y Spatial Statistics Toolbox (Pace y Barry), S+SpatialStats

² Los datos reticulares o “lattice” pueden definirse como aquéllos referidos a regiones de formas no regulares en el espacio (provincias, municipios, etc.), tal como se expone en los primeros trabajos de Cressie (1993) o recientemente en Bao *et al.* (2000), Haining *et al.* (2000), Anselin (2001A), Anselin *et al.* (2001), Wall y Devine (2000). En Okabe y Funamoto (2000) pueden encontrarse referencias sobre otros paquetes del AEDE desarrollados para el campo de la ingeniería o sanidad. En Heba *et al.* (2002), se dan a conocer los avances de nuevos módulos informáticos que permiten implementar las técnicas del AEDE en programas ya existentes.

(MathSoft) y S-Plus para ArcView (Insightful Corporation), permiten también la estimación, contraste y validación de diversos tipos de modelos espaciales de regresión lineal, modelos de variables dependientes limitadas (logit, probit, tobit), modelos de expansión o estructuras espaciales, modelos bayesianos, espacio-temporales, etc.

Tabla 2. Paquetes informáticos especializados en el análisis reticular o “lattice” de datos espaciales exploratorio y confirmatorio

Empresa/autor	Página web	Paquete informático
TerraSeer/L. Anselin	www.spacestat.com	SpaceStat SpaceStat para ArcView DynESDA DynESDA2.0 (en desarrollo)
D. Swayne, D. Cook y A. Buja	www.xplore-stat.de/tutorials/rpcframe7.html	XGobi/ ArcView/ XploRe
Insightful Corporation	www.insightful.com	S-Plus para ArcView GIS
J. Ma, B. Haining y S. Wise	www.shef.ac.uk/~scgisa	Sage
J.P. LeSage	www.spatial-econometrics.com	Econometrics Toolbox
K. Pace y R. Barry	www.spatial-statistics.com	SpaceStatPack Spatial Statistics Toolbox
MathSoft, Inc.	www.mathsoft.com	S+SpatialStats

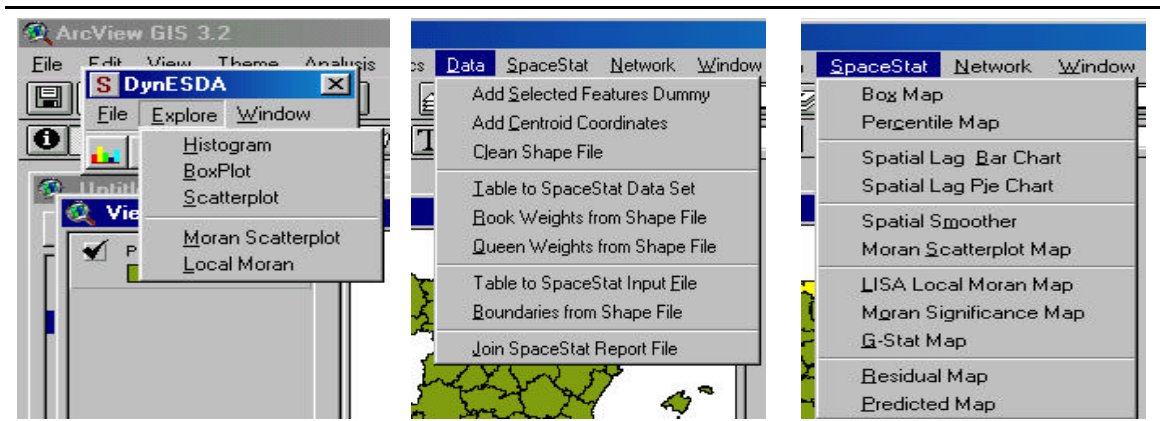
Fuente: Elaboración propia.

El programa **SpaceStat**, desarrollado por Luc Anselin (1992, 1995A), como ya se ha indicado, destaca en todo este panorama por su dedicación específica a los métodos exploratorios y confirmatorios propios de la econometría espacial. SpaceStat no es un paquete estadístico amplio, en el sentido tradicional del término, dado que se han omitido intencionadamente muchos tests y métodos fácilmente disponibles en programas actualmente en el mercado (SPSS, E Views, SAS, SCA, etc.). El objetivo de este programa es centrarse en aquellas técnicas estadísticas espaciales que implican una inversión considerable en términos de programación, en vez de incluir las que ya se encuentran disponibles. Este programa se encuentra escrito y compilado en GAUSS, que es un lenguaje matricial para técnicas estadísticas de Aptech Systems, Inc., de forma que puede desarrollarse tanto sobre el propio programa GAUSS como sobre un módulo o “runtime” del mismo incluido en la propia versión de SpaceStat.

Además, Anselin ha desarrollado dos extensiones para el GIS ArcView, la extensión de SpaceStat para ArcView propiamente dicha y **DynESDA** (Anselin, 1999, 2000), así como una nueva versión **DynESDA2.0**, para MapObjects de ESRI (Anselin *et al.*, 2001, 2002). Estas aplicaciones tienen como objetivo facilitar el intercambio y visualización de datos de algunos resultados obtenidos por SpaceStat susceptibles de representación geográfica, haciendo especial hincapié en los métodos del AEDE.

La extensión de SpaceStat para ArcView consiste en un conjunto de librerías (DLL) escritas en lenguaje C++ que se presenta al usuario en la barra superior de ArcView, en forma de dos menús adicionales denominados “Data” y “SpaceStat” (ver Figura 1).

Figura 1. Extensión de SpaceStat para el GIS ArcView y DynESDA



Fuente: Elaboración propia a partir de la extensión de SpaceStat para ArcView.

Por su parte, la extensión DynESDA para ArcView ha dotado de **dinamicidad** el AEDE, de forma que un mapa constituya una de las varias “vistas” posibles de una base de datos, junto con una tabla, histograma, gráfico de caja u otro gráfico de carácter estadístico. Estas vistas se encuentran conectadas de forma que cualquier observación señalada en una de ellas queda simultáneamente destacada en el resto. Además de gráficos estadísticos tradicionales, DynESDA incluye un visualizador para el fenómeno de asociación espacial que permite el recálculo interactivo de los estadísticos de asociación para subconjuntos de datos, así como contrastes de influencia y atípicos espaciales. Esta aplicación ha sido diseñada para el tratamiento de datos reticulares o “lattice” (unidades poligonales), en lugar de puntos, como sucede en la perspectiva geoestadística.

3. MÉTODOS GRÁFICOS DEL AEDE

En los últimos años, se han propuesto gran cantidad de métodos gráficos para el AEDE aunque, como ponen de manifiesto Haining *et al.* (2000), existen pocos estudios que valoren la utilidad y efectividad de todos ellos. En línea con Wise *et al.* (1999), tal como propone Tukey (1977) para el AED, podría afirmarse que un buen método gráfico de AEDE es aquél capaz de analizar y representar dos características fundamentales en toda distribución espacial: alisado (“smooth”) y asperezas (“rough”).

Por un lado, la propiedad de alisado, que en el campo temporal serían la tendencia central de la variable (medida a través de la mediana) y su dispersión (recorrido intercuartílico), en las variables geográficas incluiría elementos globales, referidos a todo el mapa, como la tendencia espacial y autocorrelación espacial global. Por otro, la propiedad que podría denominarse como asperezas (“rough”) de la distribución se refiere a aquellos datos localizados a cierta distancia del elemento de alisado (mediana), como los llamados atípicos (“outliers”), situados bajo/sobre el primer/tercer cuartil de un diagrama de caja. Esta última propiedad es local, a diferencia de la propiedad de alisado que es global, por lo que incluye, por ejemplo, casos que se revelan como muy distintos de sus correspondientes valores vecinos en el mapa (atípicos espaciales), regiones que se encuentran agrupadas en forma de valores altos o bajos de una variable produciendo autocorrelación espacial local (agrupamientos o “clusters” de zonas calientes o frías) o incluso, líneas de discontinuidad geográfica (heterogeneidad espacial).

En la Tabla 3, se presentan las principales técnicas del AEDE que, como puede apreciarse, contemplan los dos citados elementos de alisado (global) y asperezas (local) desde una perspectiva reticular o “lattice”.

Esta taxonomía de métodos del AEDE se inspira en otra similar propuesta por Anselin (1998) y recogida por Moreno y Vayá (2000), en la que se diferencian los métodos de representación de distribuciones espaciales, por un lado y, por otro, los específicos de visualización del fenómeno de asociación espacial (global, local y multivariante), a los que se han añadido otros gráficos y el análisis exploratorio propio del efecto de heterogeneidad espacial. Estos métodos han sido también implementados en varios de los programas informáticos expuestos en la Tabla 2.

Tabla 3. Técnicas del análisis exploratorio de datos espaciales (AEDE), según la perspectiva reticular o “lattice”

Distribución espacial	Univariante	<ul style="list-style-type: none"> • Diagrama/mapa de caja
	Multivariante	<ul style="list-style-type: none"> • Diagrama dispersión-caja
Asociación espacial	Global	<ul style="list-style-type: none"> • Mapa de contigüidades espaciales • Gráfico del retardo espacial • Diagrama/mapa de dispersión de Moran
	Local	<ul style="list-style-type: none"> • Puntos atípicos en el diagrama de dispersión de Moran • Mapas LISA • Diagrama de caja LISA
	Multivariante	<ul style="list-style-type: none"> • Diagrama de dispersión multivariante de Moran
Heterogeneidad espacial		<ul style="list-style-type: none"> • Mapa del histograma de frecuencias • Diagrama de dispersión

Fuente: Elaboración propia.

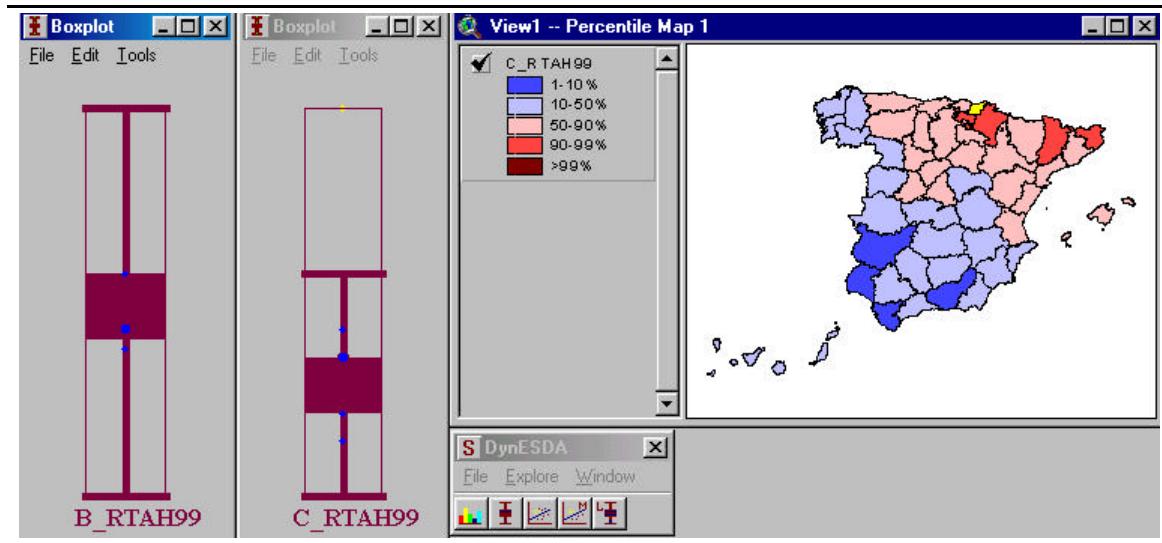
4.1. Métodos de representación de distribuciones espaciales

Las técnicas de visualización de distribuciones geográficas son muy parecidas a las herramientas de representación cartográfica, aunque el punto de partida es ligeramente distinto: para las segundas, el mapa es el elemento central, mientras que los procedimientos del AEDE se basan en elementos gráficos de la estadística clásica. En la primera fila de la Tabla 3, se presentan el diagrama/mapa de caja como los métodos de representación univariante más destacados del enfoque reticular o “lattice”. Además, se han incluido algunos métodos novedosos de representación multivariante como el diagrama de dispersión-caja, para el que se ha utilizado el programa Cvd 1.1, especializado en el AEDE multivariante.

El diagrama/mapa de caja³ es un método de representación univariante utilizado por la perspectiva reticular o “lattice” que considera que los datos espaciales son realizaciones individuales de un proceso estocástico (como en el análisis de series

temporales), haciendo posible la identificación de puntos atípicos o “outliers” espaciales. En el caso de trabajar en un sistema de AEDE dinámico, el usuario puede seleccionar varios puntos atípicos en un diagrama de caja dando lugar a zonas destacadas en el mapa.

Figura 2. Diagrama y mapa de caja de la variable renta disponible por habitante de las provincias españolas en 1999



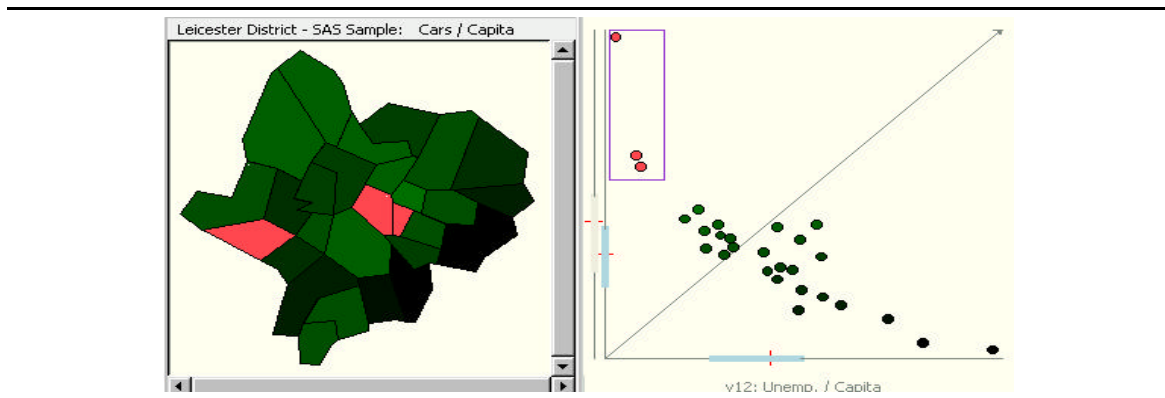
Fuente: Elaboración propia a partir de las extensiones SpaceStat y DynESDA para ArcView

La construcción del diagrama y mapa de caja parte del cálculo de los cuartiles y la media de una variable, así como de la obtención de las llamadas cotas o valores adyacentes superior e inferior, que se obtienen, a su vez, como 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 situados por encima (o por debajo) de dichas cotas (en la Figura 2, se trata del diagrama de la variable C_RTAH99). Un criterio un poco más estricto que el anterior consistiría en multiplicar por tres el recorrido intercuartílico para la fijación de las cotas (B_RTAH99). En el ejemplo de la Figura 2, el único valor atípico (superior) de renta disponible por habitante sólo se produce en el caso menos estricto (C_RTAH99) correspondiéndose con la provincia de Guipúzcoa (destacada en color amarillo tanto en el gráfico como en el mapa), no existiendo atípicos por debajo de la cota inferior.

³ Los mapas de caja representan mediante distintos colores las unidades geográficas cuyos datos en una variable coinciden con la mediana, el rango intercuartil y valores atípicos.

Por su parte, el **diagrama de dispersión/caja** es una técnica propia del AEDE multivariante que consiste en un diagrama de dispersión para dos variables en el que se representan también los correspondientes diagramas de caja en los ejes, indicándose los valores de las medianas (en la Figura 3 con una línea roja) y los cuartiles (caja azul). En esta Figura, en el diagrama de dispersión, que muestra una fuerte relación negativa entre las variables tasa de paro y tasa de automóviles, se han destacado tres puntos atípicos, con altas tasas de paro y bajas tasas de automóviles.

Figura 3. Diagrama de dispersión/caja (derecha) de las variables tasa de paro y tasa de automóviles de los distritos de Leicester



Fuente: Elaboración propia a partir del programa cvd 1.1.

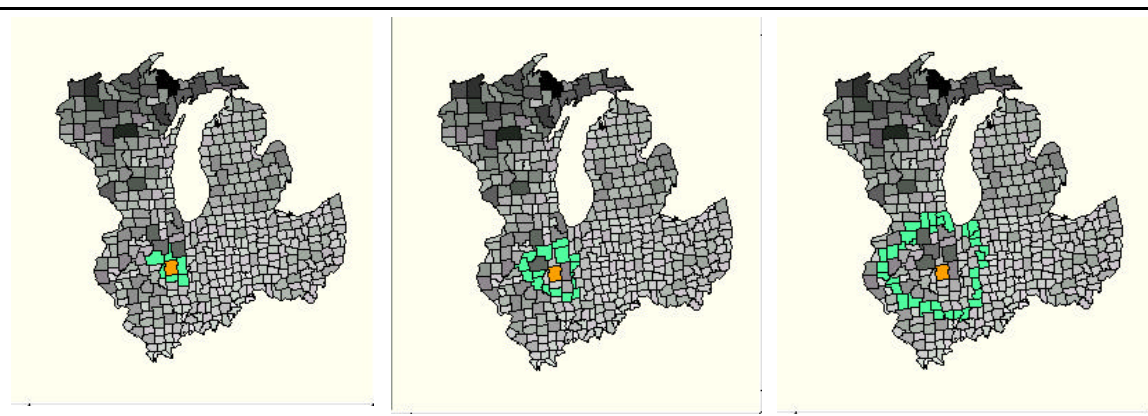
4.2. Representación del fenómeno de dependencia espacial

Los otros tres grupos de técnicas expuestas en la Tabla 3 se encuentran dentro del concepto de asociación, dependencia o autocorrelación espacial (global, local y multivariante). En el Apartado 2, la dependencia espacial era definida como el efecto que consiste en la existencia de una relación funcional entre lo que ocurre en un punto determinado del espacio y en lugares cercanos o vecinos. Es decir, una variable se encontrará espacialmente autocorrelacionada cuando los valores observados en un punto o región dependan de los valores observados en regiones vecinas, de forma que se produzca una cierta continuidad geográfica en la distribución de esta variable, por ejemplo, sobre un mapa.

Las principales técnicas del AEDE de asociación espacial, pertenecientes a la perspectiva reticular o “lattice”, se encuentran en paquetes informáticos como SpaceStat, siendo posible también el dinamismo de dicho análisis con un “software” especial de vinculación con ArcView, como DynESDA. Estas técnicas son las siguientes: mapa de contigüidades espaciales, gráfico del retardo espacial, diagrama y mapa de dispersión de Moran (asociación global), puntos atípicos en el diagrama de dispersión de Moran, mapas LISA, diagrama de caja LISA (asociación local) y diagrama de dispersión multivariante de Moran (asociación multivariante).

Los contrastes de dependencia o autocorrelación espacial propios del enfoque reticular o “lattice” se basan en la noción de contigüidad binaria entre las unidades espaciales, según la cual dos unidades espaciales se consideran como contiguas si poseen una frontera común de longitud no nula. Esta definición de contigüidad requiere obviamente de la existencia de un mapa, a partir del cual se puedan obtener las fronteras entre unidades espaciales. De acuerdo con esta definición, podrían considerarse varios órdenes de contigüidad de forma recursiva, tal como se expone en el **mapa de contigüidades** de la Figura 4.

Figura 4. Representación de las relaciones de contigüidad de orden 1, 2 y 4 para el condado de Champaign, en el medio oeste americano



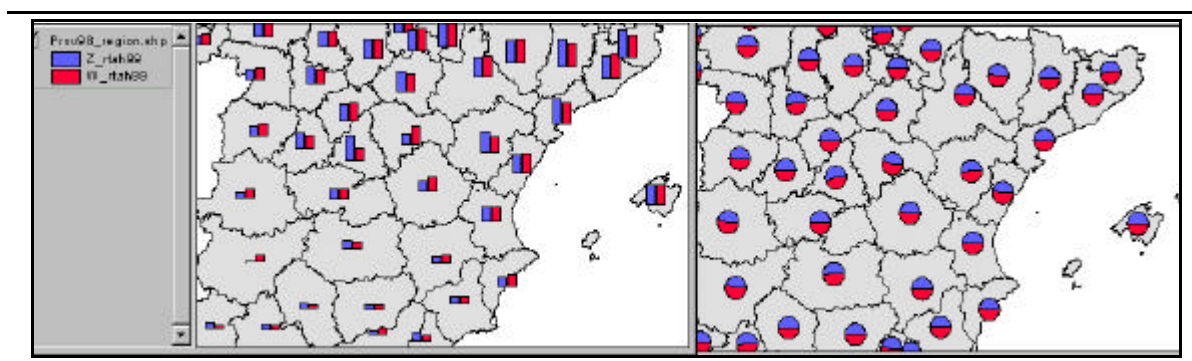
Fuente: Elaboración propia a partir del programa cvd 1.1.

Así, por ejemplo, en la Figura 4, se representan varios órdenes de contigüidad para el condado americano de Champaign (señalado en naranja): los condados contiguos de primer orden al considerado (en color aguamarina), serían los que comparten con él una

frontera común (Mclean, Ford, Vermilion, Edgar, Douglas y Piatt), mientras que los 14 condados contiguos de segundo orden (Woodford, Livingstone, Kankakee, Iroquois, Benton, Warren, Vigo, Clark, Coles, Moultrie, Macon, De Witt, Logan y Tazewell), son los que, manteniendo una frontera común con los seis anteriores, no son contiguos de primer orden con Champaign.

En el **gráfico del retardo espacial** el valor que adopta una variable en una determinada región se representa mediante un diagrama de barras o sectores, frente a su retardo espacial⁴, de forma que el predominio de barras de igual altura, dos a dos, es signo claro de fuerte autocorrelación espacial y, por el contrario, alturas desiguales en las barras resultan indicativas de atípicos espaciales.

Figura 5. Gráfico del retardo espacial de la renta per capita provincial para 25 provincias españolas



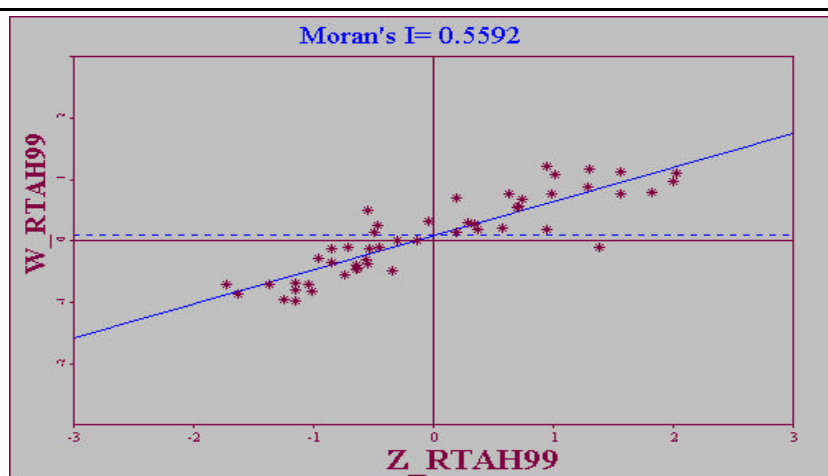
Fuente: Elaboración propia a partir de la extensión de SpaceStat para ArcView.

En la Figura 5, se han representado dos gráficos del retardo espacial de la renta disponible per cápita de una selección de provincias españolas, advirtiéndose bastante equilibrio entre la altura de las barras dentro de cada par de variables, excepto en el caso de las provincias de Madrid y Guadalajara que constituyen atípicos espaciales.

⁴ Las herramientas del AEDE reticular (“lattice”) se basan en la representación gráfica de la asociación existente entre variables y retardos espaciales, entendiéndose por retardo espacial el promedio ponderado de los valores que adopta una variable en el subconjunto de observaciones vecinas a una dada. Por ejemplo, el retardo espacial de la variable renta per cápita de la provincia de Madrid podría obtenerse como una media aritmética simple de los valores de renta per cápita en las provincias limítrofes (Segovia, Ávila, Toledo, Cuenca y Guadalajara).

En el **diagrama de dispersión de Moran**, se estandariza la variable a analizar y se obtiene el retardo espacial de dicha variable estandarizada, representándose ambos valores en un eje cartesiano. La pendiente de la recta de regresión es el valor del estadístico I de Moran de autocorrelación espacial global⁵, de forma que cuanto mayor sea el ángulo que forme ésta con el eje de abscisas, más fuerte será el grado de autocorrelación espacial, y viceversa.

Figura 6. Diagrama de dispersión de Moran de la renta per capita provincial



Fuente: Elaboración propia a partir de las extensiones de SpaceStat y DynESDA para ArcView.

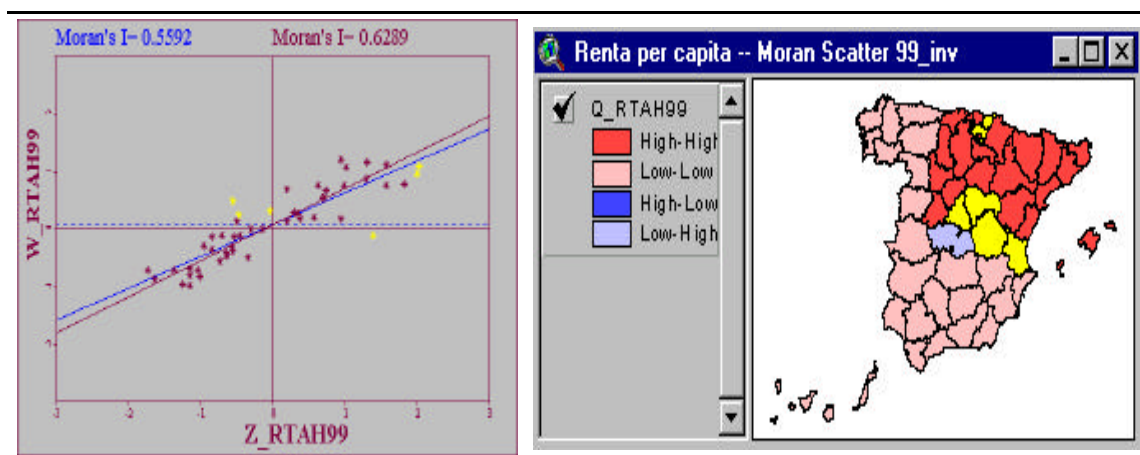
Este diagrama de dispersión suele dividir el tipo de asociación espacial en cuatro categorías: dos para autocorrelación espacial positiva (valores altos de una variable rodeados de valores altos o valores bajos rodeados de valores bajos) y dos para autocorrelación espacial negativa (valores altos rodeados por valores bajos, y viceversa). Las categorías de asociación espacial positiva se corresponden con los cuadrantes I y III (ver gráfico de la Figura 6), en los que se presentan aquellas provincias que, con una renta disponible por habitante superior/inferior a la media nacional (valor estandarizado positivo/negativo de la variable) se encuentran rodeadas de provincias que también disponen de rentas superiores/inferiores a la media nacional (retardo espacial positivo/negativo), respectivamente. Por el contrario, las categorías de asociación negativa vienen dadas por

⁵ Para un mejor conocimiento de las medidas estadísticas de autocorrelación espacial que se mencionan en este Apartado, ver las ya mencionadas referencias de Cliff y Ord (1973, 1981), Anselin (1988), Ord y Getis

los cuadrantes II y IV de este diagrama, en los que se representan las provincias con valores bajos/altos de renta per cápita rodeadas por provincias con valores altos/bajos de dicha variable, respectivamente.

Estas cuatro categorías dan lugar a una cierta forma de alisado espacial que puede visualizarse fácilmente en un **mapa** de la Figura 7 (Anselin y Bao, 1997). Debe advertirse, además, que la regresión del retardo espacial sobre la variable correspondiente se realiza con todas las hipótesis clásicas del análisis de regresión. Así, la interpretación del test I de Moran claramente permite conocer el grado en que este estadístico resume la estructura global de asociación lineal existente en un fenómeno espacial que, en el caso que se expone en la Figura 6, para el total de la distribución, sería del 56% (valor del estadístico I de Moran en términos porcentuales), aunque si se excluyen los casos atípicos situados en el primero, segundo y cuarto cuadrantes (puntos destacados en color amarillo, en la Figura 7), se eleva al 63%.

Figura 7. Diagrama y mapa de dispersión de Moran de la renta provincial



Fuente: Elaboración propia a partir de las extensiones de SpaceStat y DynESDA para ArcView.

Dado que los valores de la variable se encuentran estandarizados en el diagrama de dispersión de Moran, es posible conseguir información tanto de la asociación espacial global (pendiente de la recta de regresión) como de la local. De este modo, los valores en el diagrama de dispersión de Moran situados a más de dos unidades del origen, según el

(1995) y, en español, Moreno y Vayá (2000) y Chasco (2003).

llamado criterio “2-sigma”, pueden considerarse como **puntos atípicos en el diagrama de dispersión de Moran**, es decir, importantes “agujeros” locales de no estacionariedad espacial que, en el caso de la Figura 7, se trataría de las provincias de Guipúzcoa y Álava (en color amarillo y en el primer cuadrante). Se da además la circunstancia de que algunas provincias se localizan en los cuadrantes de asociación espacial negativa, como Cuenca, Guadalajara y Valencia/València (segundo cuadrante), que poseen un valor de renta per cápita inferior a la media pero están rodeadas de provincias con valores superiores (“Low-high”) y, en el cuarto cuadrante, la provincia de Madrid, que presentando un valor alto de renta per cápita, se encuentra rodeada de valores de dicha variable ligeramente inferiores a la media nacional (“High-low”).

En cuanto a los **mapas LISA** (“Local Indicator of Spatial Association”), se trata de mapas en los que se representan aquellas localizaciones con valores significativos en indicadores estadísticos de asociación espacial local (Getis y Ord, 1992; Anselin, 1995B; Ord y Getis, 1995; Unwin, 1996), poniendo así de manifiesto la presencia de puntos calientes (“hot spots”) o atípicos espaciales, cuya mayor o menor intensidad dependerá de la significatividad asociada de los citados estadísticos.

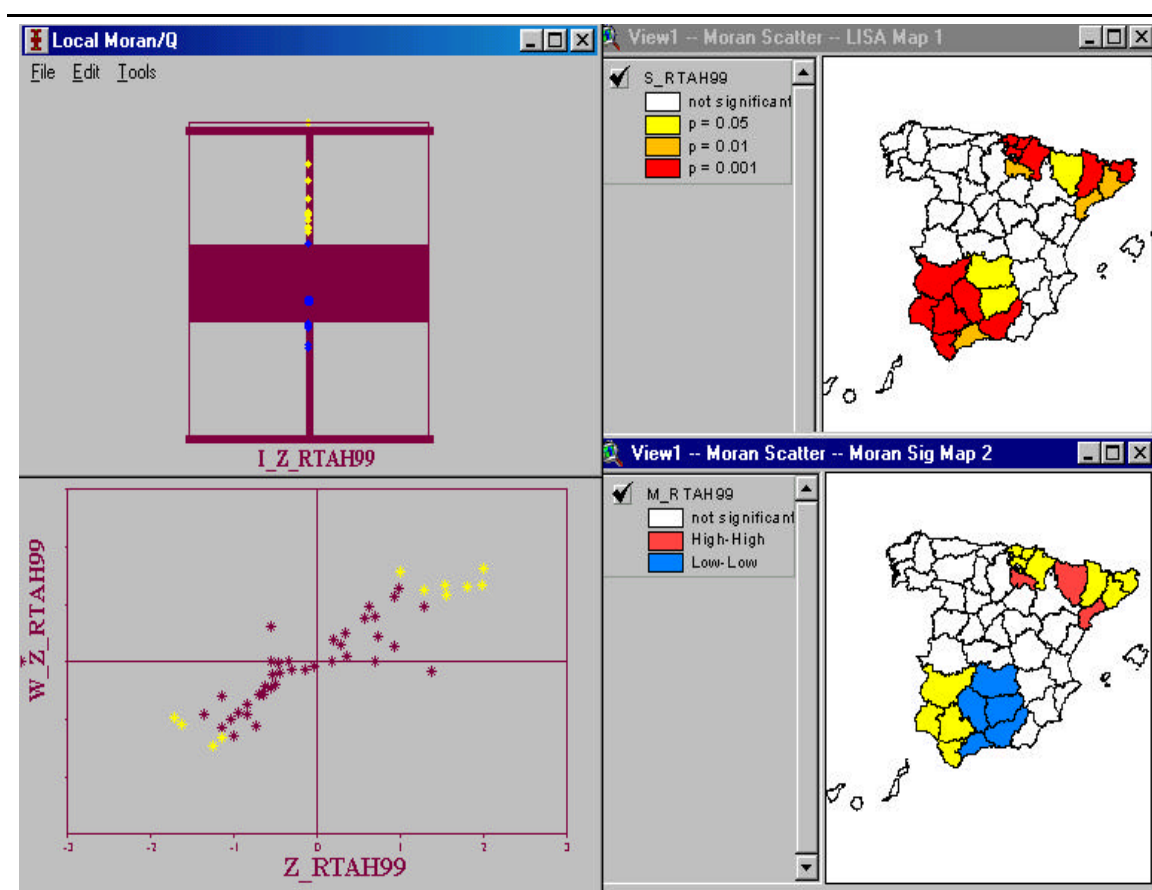
En la Figura 8, se han representado dos mapas LISA: en el superior, a través de una gradación de colores, se representan los distintos niveles de significación del estadístico de asociación espacial I local de Moran para las provincias con alta concentración de valores de renta per cápita alta rodeadas de vecinas también con renta alta (en el diagrama de dispersión de Moran, zona “High-high”), y viceversa para valores bajos de renta per cápita (“Low-low”).

En el mapa LISA inferior de la Figura, se representan las zonas calientes de alto valor (color rojo) y bajo valor (azul) de la distribución de la renta disponible per cápita provincial. La zona caliente de alta renta disponible per cápita está formada por las provincias vascas, catalanas, Navarra, La Rioja y Huesca), mientras que la zona caliente de bajo valor de renta se localiza en el suroeste peninsular y comprende las provincias andaluzas (excepto Almería), Badajoz y Ciudad Real.

En la Figura 8, también se ha representado el **diagrama de caja LISA** para la distribución del estadístico I local de Moran de asociación espacial. En un entorno de AEDE dinámico, se han destacado (en color amarillo) las provincias con un nivel alto en

este estadístico (superior al segundo cuartil), que también aparecen destacadas en el diagrama de dispersión de Moran inferior. Es decir, se trataría de aquellas provincias en las que se produce una especial concentración de valores extremos (altos rodeados de altos o bajos rodeados de bajos) de renta per cápita. Así, en la zona caliente, el centro de la misma se localizaría en las provincias más septentrionales (excepto Huesca), mientras que el “núcleo duro” de la zona fría vendría definido por las provincias suroccidentales de Cádiz, Huelva, Sevilla y Badajoz.

Figura 8. Mapas LISA de la renta disponible per capita provincial (derecha) y diagrama de dispersión multivariante de Moran (izquierda)



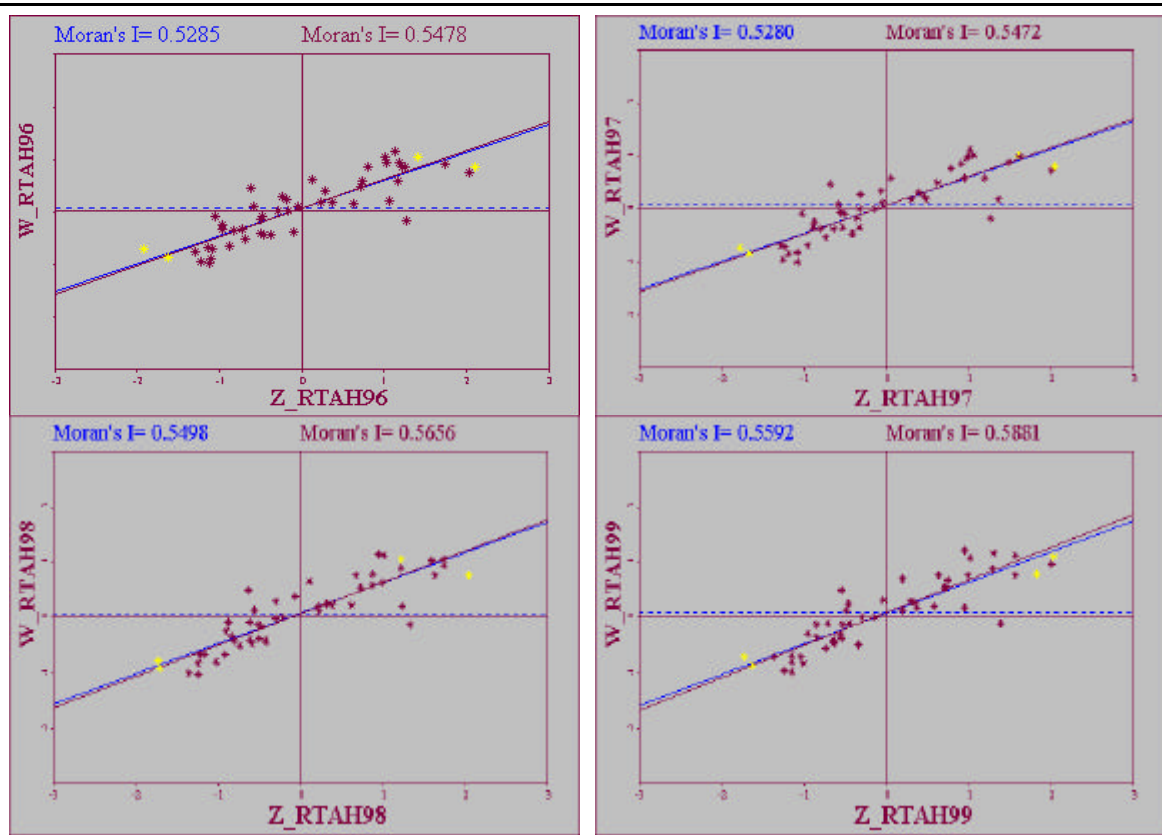
Fuente: Elaboración propia a partir de las extensiones de SpaceStat y DynESDA para ArcView.

Por último, el **diagrama de dispersión multivariante de Moran** es una técnica de exploración de asociación espacial multivariante derivada del clásico estadístico I de Moran de asociación espacial. Este diagrama multivariante, implantado en un entorno dinámico,

permite comparar el comportamiento del fenómeno de asociación espacial en varios indicadores.

Como puede apreciarse en la Figura 9, durante el período 1996-99, la variable renta disponible por habitante de las provincias españolas mantiene un nivel alto de autocorrelación espacial positiva, siendo máximo en el último año considerado.

Figura 9. Diagrama de dispersión multivariante de Moran de las variables de renta disponible per capita provincial para los años 1996-1999.



Fuente: Elaboración propia a partir de las extensiones de SpaceStat y DynESDA para ArcView.

Además, este método permite analizar el comportamiento de determinadas unidades geográficas que, por ejemplo, se presentan con valores atípicos o extremos. Este último sería el caso de las provincias de Guipúzcoa y Álava, de altos niveles de renta per capita (destacadas en amarillo en el cuadrante primero), que han ido avanzando posiciones en el ranking de renta per capita en los cuatro años considerados, mientras que, por el contrario

(ver cuadrante tres), las provincias de Cádiz y Badajoz se han mantenido siempre, a lo largo de este período, en los últimos puestos.

La existencia de asociación espacial (global y/o local), puesta de manifiesto por las técnicas del AEDE en una variable que va a ser explicada en un modelo de regresión (así como en alguna de las variables explicativas o en el propio residuo), pone de manifiesto la necesidad de contrastar estadísticamente este fenómeno, a través de las diversas medidas estadísticas de autocorrelación espacial que se han propuesto en la literatura. De confirmarse esta hipótesis, el fenómeno de asociación espacial deberá ser considerado explícitamente a través de modelos de regresión específicos, como el modelo del retardo espacial (“spatial lag model”) y el modelo del error espacial (“spatial error model”)⁶.

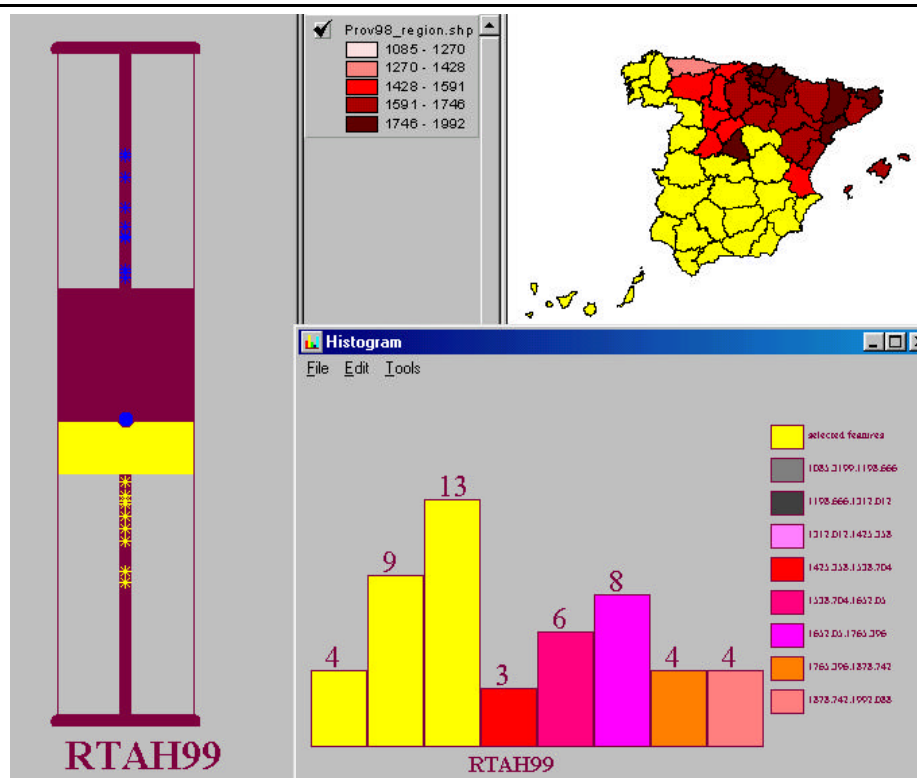
4.3. Representación del fenómeno de heterogeneidad espacial

La heterogeneidad espacial es, junto con el fenómeno de asociación espacial, uno de los efectos espaciales, definido por la ausencia de estabilidad en el espacio de la variable en estudio, lo que suele ocurrir con situaciones del tipo centro-periferia, norte-sur, este-oeste, etc. Como ya se puso de manifiesto en el capítulo anterior, esto implicará que, en los modelos espaciales, las formas funcionales y los parámetros variarán con la localización geográfica, no siendo homogéneos para toda la matriz de datos. En este caso, aunque no suele ser un tema muy citado en la literatura, la utilización del AEDE dinámico resulta también particularmente útil para su exploración. Anselin (1999) sugiere la utilización conjunta del mapa del histograma de frecuencias y el diagrama de dispersión, en el modo que se presenta a continuación.

El **mapa del histograma de frecuencias** consiste en la selección geográfica de las unidades situadas en las “barras” o intervalos de un histograma que, a su vez, puede ser complementado con un diagrama de caja, como en el caso de la Figura 10. La interacción dinámica existente entre el histograma y el mapa permitirán seleccionar el número más adecuado de intervalos de cara a una mejor identificación de diferentes estructuras o zonas de comportamiento en una variable.

⁶ Para un análisis más profundo sobre modelización econométrica espacial, es muy recomendable, por su alto valor pedagógico, la obra de Anselin, especialmente sus referencias de 1988 y 1995A y, en español, Moreno y Vayá (2000) y Chasco (2003).

Figura 10. Mapa del histograma de la variable renta disponible por habitante de las provincias españolas en 1999.

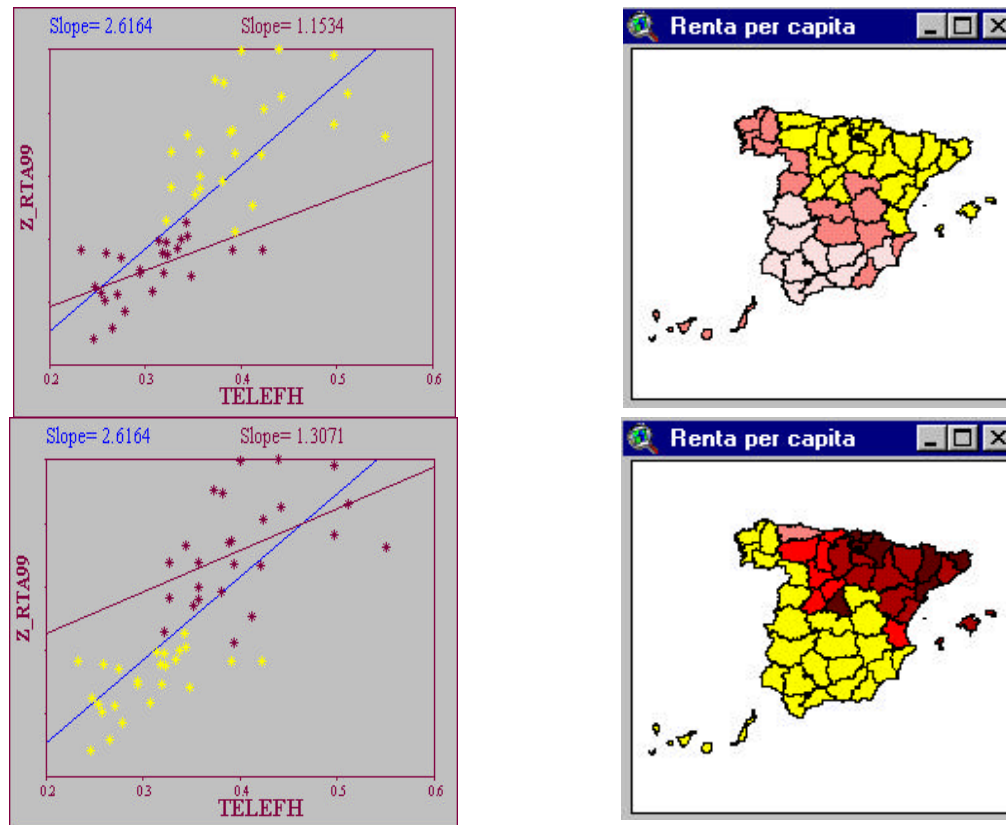


Fuente: Elaboración propia a partir de las extensiones de SpaceStat y DynESDA para ArcView.

En concreto, en esta Figura se ha dividido la distribución de la renta disponible por habitante en 8 intervalos, de forma que la selección de los tres primeros intervalos del histograma de frecuencias de la variable de renta per cápita de 1999 (con valores inferiores a la mediana, como se pone de manifiesto en el diagrama de caja), destaca en color amarillo las provincias localizadas en la mitad oeste-centro-sur de España, pudiéndose trazar una línea inclinada de discontinuidad geográfica en la distribución de esta variable.

Además, Anselin (1999B) propone comprobar la existencia de relaciones o coeficientes diferentes entre la variable a explicar en un modelo y sus explicativas, en las dos estructuras detectadas, lo que podrá llevarse a cabo mediante un **diagrama de dispersión**

Figura 11. Diagrama de dispersión de la variable teléfonos por habitante vs renta disponible por habitante de las provincias españolas en 1999.



Fuente: Elaboración propia a partir de las extensiones de SpaceStat y DynESDA para ArcView.

Efectivamente, como se observa en la Figura 11, el valor del coeficiente estimado para la variable explicativa de un modelo de regresión lineal en las estructuras espaciales anteriormente detectadas no varía significativamente: la regresión de la variable de líneas telefónicas por habitante sobre la renta disponible por habitante, que estima un coeficiente de 2,62 para el conjunto global de provincias españolas, obtiene un coeficiente de 1,15 para la muestra de datos correspondiente a la mitad occidental-centro-sur de España (valores bajos de renta per capita), siendo el coeficiente correspondiente a la más rica mitad nororiental de 1,31. En cualquier caso, estos resultados estarían indicando que la elasticidad de la variable de teléfonos por habitante sobre la renta per cápita es ligeramente superior en la mitad nororiental que en la occidental-centro-sur.

Cuando el AEDE pone de manifiesto la posible hipótesis de heterogeneidad espacial en un modelo, deberían aplicarse los contrastes estadísticos propuestos en la literatura que,

de confirmar esta hipótesis, conducirán a la utilización de modelos específicos que contemplen adecuadamente este problema, como los modelos de análisis de la varianza espacial (SANOVA) y de estructuras espaciales, para situaciones de heterogeneidad espacial discreta, o los modelos de superficie tendencial y expansión espacial, en los casos de heterogeneidad espacial continua.

4. BIBLIOGRAFÍA

- . ANSELIN, L. (1988A), "*Spatial econometrics: methods and models*". Kluwer Academic Publishers.
- . ANSELIN, L. (1992), "*Space Stat tutorial. A workbook for using SpaceStat in the analysis of spatial data*". Technical Report S-92-1, National Center for Geographic Information and Analysis, University of California. Santa Barbara, CA.
- . ANSELIN, L. (1995A), "*Space Stat version 1.80: Users' guide*". Regional Research Institute, West Virginia University, Morgantown, WV.
- . ANSELIN, L. (1995B), "*Local Indicators of Spatial Association-LISA*". *Geographical Analysis*, vol. 27(2); pp. 93-115.
- . ANSELIN, L. (1998), "*Exploratory spatial data analysis in a geocomputational environment*". *Actas de la Conferencia de GeoComputación 1998*, Bristol (UK), 17-19.
- . ANSELIN, L. (1999), "*Spatial Data Analysis with SpaceStat™ and ArcView®. Workbook (3rd Edition)*". Department of Agricultural and Consumer Economics, University of Illinois, Urbana.
- . ANSELIN, L. (2000), "*Computing environments for spatial data analysis*". *Journal of Geographical Systems* 2 (3), pp. 201–225.
- . ANSELIN, L. (2001A), "*Spatial effects in econometric practice in environmental and resource economics*". *American Journal of Agricultural Economics*, 83 (3), pp. 705-710.
- . ANSELIN, L. (2001B), "*Spatial econometrics. 2001 Summer course*". University of Illinois, Urbana-Champaign, <http://www.spacestat.com>.
- . ANSELIN, L. y S. BAO (1997), "*Exploratory Spatial Data Analysis*". En "Recent developments in spatial analysis" (Eds. Fischer y Getis), Springer-Verlag, Berlín; pp.35-59.
- . ANSELIN, L., I. SYANBRI, O. SMIRNOV y Y. REN (2001), "*Visualizing spatial autocorrelation with dynamically linked windows*". Discussion Paper de la Regional Economics Applications Laboratory REAL 01-T-10, noviembre.

- . ANSELIN, L., I. SYABRI y O. SMIRNOV (2002), “*Visualizing multivariate spatial correlation with dynamically linked windows*”. Documento de la Regional Economics Applications Laboratory, REAL 02-T-8, julio 2002.
- . ARBIA, G. (1989), “*Spatial data configuration in statistical analysis of regional economics and related problems*”. Dordrecht: Kluwer.
- . BALLER, R., L. ANSELIN, S. MESSNER y D. HAWKINS (2001), “*Structural covariates of U.S. county homicide rates: incorporating spatial effects*”. Criminology (próxima publicación).
- . BAO, S., L. ANSELIN, D. MARTIN y D. STRALBERG (2000), “*Seamless integration of spatial statistics and GIS: the S-Plus for ArcView and the S+Grassland links*”. Journal of Geographical Systems 2 (3), pp. 287–306.
- . BOSCH, M., P. AROCA, I. FERNÁNDEZ, C. AZZONI (2002), “*Growth dynamics and space in Brazil*”. 42º Congreso de la Asociación Europea de Ciencia Regional (ERSA). Dortmund (Alemania). CD-ROM.
- . BUTTON, K. y R. KULKARNI (2001), “*Spatial and distance statistics of the trucking and warehousing industries using GIS tools*”. 41º Congreso de la Asociación Europea de Ciencia Regional (ERSA). Zagreb. CD-ROM.
- . BRUNDSON C. (1998), “*Exploratory spatial data analysis and local indicators of spatial association with XLISP-STA.*”. The Statistician 47, pp. 471-484.
- . CECCATO, V., R. HAINING y P. SIGNORETA (2002), “*Exploring offence statistics in Stockholm city using spatial analysis tools*”. Annals of the American Association of Geographers, 92 (1); pp. 29-51.
- . CHASCO, C. (2003), “*Econometría espacial aplicada a la predicción-extrapolación de datos microterritoriales*”. Consejería de Economía e Innovación Tecnológica de la Comunidad de Madrid (próxima publicación, en abril).
- . CHASCO, C. y J. VICÉNS (2000), “*EU-membership impacts in the Spanish province income convergence: a spatial autocorrelation perspective*”. Anales de Economía Aplicada XIV Reunión de ASEPELT-España, Oviedo.
- . CLIFF, A. y J. ORD (1973), “*Spatial autocorrelation*”. London: Pion.
- . CLIFF, A. y J. ORD (1981), “*Spatial processes, models and applications*”. London: Pion.
- . COX, K. R. (1969), “*The voting decision in a spatial context*”. Progress in Geography, 1. Eds. C. Board, R. J. Chorley, P. Hagget, D. R. Stoddart. Arnold, London; pp. 81- 117.
- . CRESSIE, N. (1993), “*Statistics for spatial data..Revised edition*”. New York: Wiley.

- . FOTHERINGHAM, A., M. CHARLTON y C. BRUNSDON (2001), "*Spatial variations in school performance: A local analysis using geographically weighted regression*". Geographical & Environmental Modelling, vol. 5, nº 1; pp. 43-66.
- . GETIS, A. y J. ORD (1992), "*The analysis of spatial association by use of distance statistics*". Geographical Analysis, 24; pp. 189-206.
- . GEZICI, F. Y G. HEWINGS (2002), "*Spatial Analysis of Regional Inequalities in Turkey*". Discussion Paper de la Regional Economics Applications Laboratory REAL 02-T-11.
- . HEBA, I., E. MALIN y C. THOMAS-AGNAN (2002), "*Exploratory spatial data analysis with GEOCP*". Actas del 42º Congreso de la ERSA (Asociación Europea de Ciencia Regional), Dortmund (Alemania). CD-ROM.
- . HAINING, R., S. WISE y P. SIGNORETTA (2000), "*Providing scientific visualization for spatial data analysis: Criteria and an assessment of SAGE*". Journal of Geographical Systems, 2; pp.121-140.
- . LESAGE, J. y R. KELLEY PACE (2000), "*Using matrix exponentials to explore spatial structure in regression relationships*". Discussion Paper de la Regional Economics Applications Laboratory REAL, octubre.
- . LÓPEZ, F.A., M.A. PALACIOS y M. RUIZ (2001), "*Modelos explicativos del desempleo en términos de localización. Una aplicación a las provincias españolas.*". Anales de Economía Aplicada XV Reunión de Asepelt-España, Santiago de Compostela.
- . MORENO, R. y E. VAYÁ (2000), "*Técnicas econométricas para el tratamiento de datos espaciales: la econometría espacial*". Edicions Universitat de Barcelona, colecció UB 44, manuals.
- . OKABE, A. y S. FUNAMOTO (2000), "*An exploratory method for detecting multi-level clumps in the distribution of points - a computational tool, VCM (variable clumping method)*". Journal of Geographical Systems, 2; pp. 111-120.
- . ORD, J.K. y A. GETIS (1995), "*Local spatial autocorrelation statistics: distributional issues and an application*". Geographical Analysis, 27.4; pp. 286-306.
- . REY, S. (2001C), "*Spatial analysis of regional income inequality*". Discussion Paper de la Regional Economics Applications Laboratory REAL 01-T-9, octubre.
- . SPSS Inc. (1999), "*Tutorial del programa SPSS para Windows, vs. 10.0.6*". 27 de noviembre de 1999.
- . STOUGH, RR., R. KULKARNI y J. PAELINCK (2001), "*Spatial-structural relations among technology industrial clusters: A comparative analysis of metropolitan regions in the U.S.*". Actas del 41º Congreso de la Asociación Europea de Ciencia Regional (ERSA). Zagreb. CD-ROM.
- . TOBLER, W. (1979), "*Cellular Geography*". En "Philosophy in Geography", ed. S. Gale y G. Olsson, Dordrecht: Reidel; pp. 379-86.

- . TUKEY, J.W. (1977), "*Exploratory Data Analysis*". Reading: Addison-Wesley.
- . UNWIN, A. (1996), "*Exploratory spatial analysis and local statistics*". *Computational Statistics*, 11; pp. 387-400.
- . UNWIN, A. (2000), "*Using your eyes- making statistics more visible with computers*". *Computational Statistics & Data Analysis*, 32; pp. 303-312.
- . VAN DER KRUK, R. (2001), "*Economic Impacts of wetland amenities: A spatial econometric analysis of the Dutch housing market*". 41º Congreso de la ERSa (Asociación Europea de Ciencia Regional), Zagreb. CD-ROM.
- . WALL, P. y O. DEVINE (2000), "*Interactive analysis of the spatial distribution of disease using a geographic information system*". *Journal of Geographical Systems*, 2; pp. 243–256.
- . WISE, S., R. HAINING y P. SIGNORETTA (1999), "*Scientific visualization and the exploratory analysis of area based data*". *Environment and Planning A*.
- . YILMAZ, S., K. HAYNES y M. DINE (2002), "*Geographic and network neighbours: Spillover effects of telecommunications infrastructure*". *Journal of Regional Science*, vol. 42; pp. 339-360.