

Autocorrelación espacial e indicadores locales de asociación espacial.

Importancia, estructura y aplicación

Juan Pablo Celemín*

Resumen

La autocorrelación espacial es un procedimiento intrínsecamente geográfico que nos puede decir mucho acerca del comportamiento de la información georreferenciada a diferentes escalas, en particular el tipo de asociación existente entre unidades espaciales vecinas. No obstante, a pesar de su creciente importancia en el marco del análisis exploratorio de datos espaciales, su utilización es reducida en trabajos geográficos que continúan recurriendo a los coeficientes tradicionales de correlación y a la estadística descriptiva. En este sentido, el presente estudio de índole metodológico intenta resaltar la utilidad de este procedimiento, detallando minuciosamente su estructura y funcionamiento por medio de un ejemplo aplicado a los departamentos de la Provincia de Santa Cruz. Específicamente, se concentra en el análisis de indicadores locales de asociación espacial (LISA, Local Indicators of Spatial Association) que descompone el índice global de autocorrelación y verifica en cuánto contribuye cada unidad espacial a la formación del valor general, permitiendo capturar de forma simultánea el grado de asociación espacial y la heterogeneidad resultante del aporte de cada unidad espacial.

Palabras Clave: autocorrelación espacial, indicadores locales de asociación espacial, análisis exploratorio de datos espaciales (ESDA)

Spatial Autocorrelation and Local Indicators of Spatial Association. Importance, Structure and Application

Abstract

Spatial autocorrelation is an inherently geographical statistical procedure that can tell much about the behavior of the geo-information at different scales, in particular the type of association among neighboring spatial units. However, despite its growing

* Becario CONICET – Universidad Nacional de Mar del Plata. Grupo de Estudios Sobre Población y Territorio. jpcelem@conicet.gov.ar

importance within the framework of exploratory analysis of spatial data, its use is limited in the geographical analyses that continue to use traditional correlation coefficients and descriptive statistics. In this sense, this methodological study aims at highlighting the usefulness of this procedure, by giving a thorough detail of its structure and functioning by means of an example applied to the departments of Santa Cruz Province. Specifically, this work focuses on the analysis of local indicators of spatial association (LISA) which breaks down the global index of autocorrelation and identifies how each spatial unit contributes to the formation of the global value thus allowing to capture, at the same time, the degree of spatial association and heterogeneity resulting from the contribution of each spatial unit.

Key Words: Spatial autocorrelation, local indicators of spatial association (LISA), statistics

Relevancia de la autocorrelación espacial en el estudio geográfico

Las características socioeconómicas y ambientales propias de la Geografía tienden a mostrar cierto grado de similitud, ya que a menos que existan factores de ruptura o de discontinuidad muy marcados, la situación normalmente esperable sería la de cierta homogeneidad espacial. En general se acepta que existe autocorrelación espacial (AE) siempre que haya una variación espacial sistemática en los valores de una variable a través de un mapa, es decir un patrón en el comportamiento de la variable según la ubicación geográfica del dato. Si los valores altos de una localización están asociados con valores altos en los vecinos, la autocorrelación espacial es positiva siendo la situación opuesta la de autocorrelación espacial negativa (Cepeda y Velázquez, 2005).

La AE es “especialmente indicada en fenómenos de propagación, es decir, que se adapten al modelo de difusión epidémica y en situaciones con un fuerte componente social, puesto que la población suele residir de un modo segregado” (Gamir Orueta, 1995: 216). En otras palabras,

“la utilidad de la AE se encuentra en su capacidad para estudiar la forma en que un fenómeno se irradia a través de las unidades espaciales, y si tal conducta corresponde a algún modelo de difusión conocido o bien registra la segregación espacial de alguna característica. En definitiva, refleja el grado en que objetos o actividades en una unidad geográfica son similares a los objetos o actividades en unidades geográficas próximas” (Vilalta y Perdomo, 2005: 326)

y es, quizás, después del promedio y la varianza la propiedad más importante de cualquier variable geográfica y, a diferencia de las anteriores, está explícitamente vinculada con patrones espaciales.

El concepto de AE tiene sustento en el principio de Tobler que plantea que en el espacio geográfico todo se encuentra relacionado con todo, pero los espacios más cercanos están más relacionados entre sí. Como apunta Goodchild (Goodchild, 2008) las consecuencias de este principio en el diseño de los Sistemas de Información Geográfica son profundas. Si no fuera cierto, algunos procedimientos como la interpolación espacial, tan útil en la elaboración de Modelos Digitales de Terreno, serían imposibles de realizar. Consecuentemente no habría ningún fundamento para suponer que el terreno podría ser representado como una malla de triángulos (los TIN - *Triangulated Irregular Network*-, en este caso) o que los puntos con características similares pueden ser agrupados en polígonos. Se puede ir más allá y sostener que un mundo sin este fundamento sería imposible de conocer o describir dado que cada punto sería independiente de su entorno más inmediato. Para un geógrafo es imposible imaginar un mundo sin autocorrelación espacial: no existirían las regiones dado que la variación de todo tipo de fenómenos tendría que ocurrir independientemente de la ubicación.

Como ocurre con varios de los modelos aplicados en el ámbito de la Geografía sus raíces se encuentran en otras disciplinas. El interés por la AE surge dentro de un subcampo de la economía denominado econometría espacial que se ocupa del tratamiento de la interacción espacial y la estructura espacial en modelos de regresión (Anselin, 1999). El mismo investigador la define como la colección de técnicas que lidian con las peculiaridades causadas por el espacio en el análisis estadístico de los modelos de la ciencia regional. No es llamativo la base económica de este concepto ya que como enuncia Krugman (citado por Sánchez Moral 2004: 208)

“el relativo inmovilismo de los geógrafos hizo que retrocedieran dentro del área del análisis espacial, fundamentalmente por las dificultades para expresar los trabajos en lenguaje lógico-matemático, favoreciendo los esfuerzos de la Economía para hacer efectiva la integración de la dimensión espacial”.

Desde una perspectiva histórica el origen de interés por la AE se remonta a los primeros años de la década del setenta cuando surgió la necesidad de hacer frente a modelos econométricos subregionales en los cuales a menudo las técnicas econométricas tradicionales fallaban: por ejemplo, los procedimientos regresión suelen malograrse bajo los efectos de la AE dado que los parámetros no son eficientes y los *tests* de significancia no son confiables en presencia de ese fenómeno (Miller, 1999). Es por ello que en algunos modelos de regresión, para establecer, se someten a la AE a la espera de que la configuración sea aleatoria.

Cliff y Ord, investigadores de la Universidad de Bristol en Gran Bretaña, influidos por la escuela cuantitativa de la Universidad de Washington, fueron los primeros en acuñar el término AE en el año 1968 ya que hasta entonces se la conocía como dependencia espacial, asociación espacial, interacción espacial o interdependencia espacial. En años sucesivos desarrollaron aspectos teóricos que contribuyeron significativamente a la sistematización del procedimiento Getis, (2008).

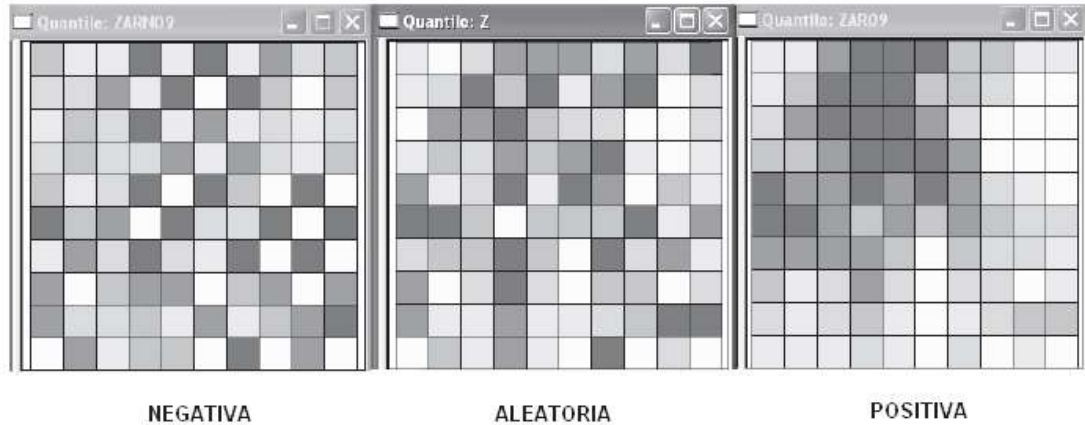
Índices de autocorrelación espacial

La medición de la correlación que una misma variable tiene en diferentes unidades espaciales contiguas en una perspectiva horizontal da lugar a una de estas tres posibilidades:

- Autocorrelación espacial positiva: las unidades espaciales vecinas presentan valores próximos. Indica una tendencia al agrupamiento de las unidades espaciales.
- Autocorrelación espacial negativa: las unidades espaciales vecinas presentan valores muy disímiles. Indica una tendencia a la dispersión de las unidades espaciales.
- Sin autocorrelación: no ocurre ninguna de las dos situaciones anteriores. Por lo tanto, los valores de las unidades espaciales vecinas presentan valores producidos en forma aleatoria.

Una forma de visualizar estos tres eventos se encuentra en la Figura 1.

Figura 1
Tipos de autocorrelación espacial



Fuente: Anselin, 2003

Los índices de AE nos permiten relacionar en forma conjunta la dependencia entre localizaciones y valores de variables o atributos que interesan y resultan muy adecuados para observar la configuración espacial fragmentada propia de nuestros tiempos. Existen distintos tipos de estadísticos o índices que permiten medir la AE cuya estructura general es la siguiente:

$$\sum_{i=1}^n \sum_{j=1}^n W_{ij} C_{ij} \quad (1)$$

donde n es el total de lugares del mapa, W_{ij} son los elementos de una matriz (matriz de conexiones, contigüidad o de pesos espaciales) cuyos valores son una función de alguna medida de contigüidad en la matriz de datos originales (*Rook*, *Bishop* o *Queen*). El valor C_{ij} es una medida de la proximidad (distancia) de los valores i y j en alguna dimensión (por ejemplo distancia euclídea, esférica, de *Manhattan*, etc.), o cualquier distancia definida por el usuario.

Generalmente en la mayoría de los análisis de AE se consideran las relaciones de vecinos próximos. Si se piensa que las áreas espacialmente referenciadas que se están analizando son cuadradas, habrá por lo menos cuatro vecinos que comparten un borde en cada lado del cuadrado. Asimismo, podría haber ocho datos espaciales para cada observación si se consideran adicionalmente aquellos puntos que limitan con los vértices del cuadrado en cuestión. Cuando se tienen en cuenta los cuatro elementos que comparten borde se habla de contigüidad tipo *Rook*. En el caso de los ocho vecinos se habla de contigüidad tipo *Queen* y si se toman solamente los vecinos contiguos por el vértice se denomina *Bishop*. Estos nombres corresponden al movimiento que realizan la torre, la reina y el alfil en un tablero de ajedrez (ver Figura 2). El método *Rook*, por su simplicidad es el más utilizado.

Figura 2
Tipos de contigüidad



Fuente: elaboración personal

En base al estadístico general se crearon diferentes índices para medir la AE. El primero fue establecido por Moran en 1950 y fue perfeccionado a través de los años y es análogo al coeficiente de correlación usual entre dos variables. A pesar de su antigüedad no fue muy utilizado debido a que el cómputo manual era sumamente laborioso, por lo que fue necesario contar con programas informáticos para poder realizarlo. Es por ello que recién en la década del noventa fue rescatado del olvido y comenzó a utilizarse en trabajos de investigación de manera consistente. El *I* de Moran no es la única medida para la AE ya que existen otras como la *C* de Geary, pero es la más utilizada ya que, como afirma Goodchild, (2008a) *I* de Moran es, esencialmente, el coeficiente de correlación de Pearson con una matriz de pesos definida por el usuario que mantiene el rango entre -1 y 1. La estructura de *I* de Moran es la siguiente:

$$I = (n / S_0) \sum_{i=1}^n \sum_{j=1}^n w_{ij} z_i z_j / \sum_{i=1}^n z_i^2 \quad (2)$$

Donde $S_0 = \sum_{i=1}^n \sum_{j=1}^n w_{ij}$ ó dicho de una manera más sencilla la suma de elementos de la matriz de pesos. Las observaciones *z* son las desviaciones de la media $(x_i - \bar{x})$ ó $(x_j - \bar{x})$ donde x_i es el valor de la variable en una unidad espacial determinada y x_j es el valor de la variable en otra localización, normalmente las vecinas a x_i . Habitualmente, en la matriz, se les asigna el valor de 1 a los vecinos de cada unidad espacial y 0 al resto. A pesar de que la estructura de este estadístico puede resultar un poco intimidatoria, como se verá más adelante, en su composición solo existen operaciones aritméticas sencillas.

Un avance más reciente es el reconocimiento de que la AE está sujeta a la heterogeneidad espacial que reconoce la unicidad intrínseca de cada observación espacial. Esto ha resultado en medidas que permiten capturar indicadores locales de autocorrelación espacial (LISA, en inglés) y el valor global de AE de forma simultánea (Miller, 2004).

En esencia, el análisis local se basa en una simple expectativa: que cualquier modelo o teoría en las ciencias sociales no cuenta perfectamente los fenómenos que se describen. Esto se debe a que los modelos aplicados en las ciencias sociales no encajan con perfección, de manera que los resultados obtenidos dependen explícitamente de los límites del área de estudio y serán diferentes si ésta es modificada.

En consecuencia, más se aprenderá si se realiza un análisis exploratorio a nivel local que al intentar hacer encajar los resultados dentro de un modelo general (Sui, 2004). En el análisis local los especialistas reconocen que cada ubicación posee un grado intrínseco de unicidad debido a su situación relativa en el espacio. En definitiva, la heterogeneidad espacial revela tanto la intensidad como los patrones de asociación espacial de cada unidad espacial analizada.

“El método LISA descompone el índice I de Moran y verifica en cuánto contribuye cada unidad espacial a la formación del valor general, permitiendo obtener un valor de significancia para cada *cluster* formado por los valores similares de cada unidad espacial y sus vecinos. Estos agrupamientos o *clusters* de especial concentración de valores extremos de una variable se conocen también como zonas calientes/frías (*hot spots/cold spots*, respectivamente) según se trate de una concentración de valores especialmente altos/bajos de una variable, correspondientemente” (Chasco Yrigoyen, 2006: 44).

Se halla la posibilidad de visualizar el comportamiento del análisis espacial local en dos tipos de mapas. En el primero, conocido como mapa de significancia, a partir de un procedimiento de aleatorización se muestra para cada unidad espacial la probabilidad de que sus relaciones de contigüidad se produzcan de manera aleatoria en base a las categorías 1 en 10.000, 1 en 1.000, 1 en 100 y 5 en 100. En el segundo, también denominado *cluster* o de agrupamiento, se observa como cada unidad espacial se diferencia de sus unidades espaciales vecinas de acuerdo al tipo de AE que posee (Buzai y Baxendale, 2006).

La forma local estandarizada para I de Moran para la observación i es la siguiente:

$$I_i = (z_i / m_2) \sum_{j=1}^n w_{ij} z_j \quad (3)$$

Donde m_2 es la varianza. Las observaciones z son las desviaciones de la media. Generalmente, en la matriz, se les asigna el valor de 1 a los vecinos de cada unidad espacial. La matriz de pesos puede ser estandarizada (es decir que la suma de las filas sume 1) para facilitar la comparación entre los diferentes conjuntos de resultados. Esta fórmula se destaca sobre otras similares porque **la suma de las unidades espaciales locales es igual al valor global**.

Visualización de las unidades espaciales en el *Moran Scatterplot*

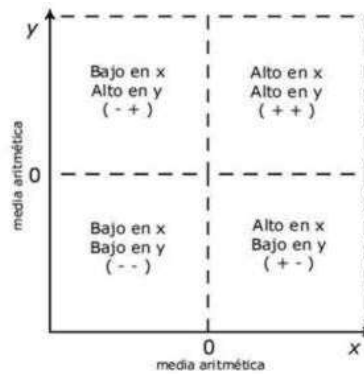
En el año 1993 Anselin presenta el *Moran Scatterplot*, una herramienta de análisis que permite observar en un gráfico de dispersión el comportamiento de cada unidad espacial. Esta novedad fue uno de los primeros pasos hacia el análisis local, es decir la desagregación del valor global de AE. El *scatterplot* se puede dividir en cuatro cuadrantes (ver Figura 3) comenzando por el primero en la parte superior derecha y siguiendo en sentido de las agujas del reloj con los siguientes. En el eje de las x aparecen los valores estandarizados de una variable para cada unidad espacial del área estudiada, y en el eje y los valores estandarizados del promedio de los valores en unidades vecinas de la misma variable (análisis univariado) u otra variable (análisis bivariado). En el cuadrante I podemos identificar las unidades espaciales con valores superiores a la media que, a su vez, también cuentan con vecinos con valores altos (situación *alto-alto*, también conocidos como *hot spots* en el *scatterplot* de Moran). La situación inversa se registra en el cuadrante III (situación *bajo-bajo*, también denominados *cold spots*). Ambos cuadrantes permiten detectar los *clusters* o agrupamientos de unidades espaciales con valores similares a los de sus vecinos. En contraparte, los *outliers* espaciales responden a contextos mixtos, en otras palabras, unidades espaciales con valores bajos (inferiores a la media) con vecinos que registran valores altos (situación *bajo-alto*) en el cuadrante IV. El escenario opuesto (situación *alto-bajo*) se encuentra en el cuadrante II. La importancia de los *outliers* radica en que no responden al principio de AE ya que el valor de las unidades espaciales que poseen a estas características se diferencia de los valores de sus vecinos. Además, por su carácter atípico sirven, inicialmente, para detectar errores en el ingreso de la información, ya sea de forma accidental o deliberada.

Tests de significancia para la autocorrelación espacial

Para definir si una AE es significativa se realiza un *test* de hipótesis nula, y así poder comprobar si la configuración espacial de la variable se produce aleatoriamente, o dicho de otra manera, si se cumplen o no los supuestos del modelo a partir de estimar si un estadístico muestral difiere significativamente de lo esperado aleatoriamente.

La forma más directa y sencilla para poner a prueba la hipótesis de AE exhibida por una muestra de n casos, es que ésta fue obtenida de una población con distribución normal donde la autocorrelación es cero, es decir, aleatoria

Figura 3
Moran Scatterplot



Fuente: Buzai, 2005

(Goodchild, 1986). Por lo tanto, si el resultado se aleja del cero, entonces, se puede decir que es significativo.

En este sentido, el valor esperado de I de Moran es:

$$E(I) = \frac{-1}{(n-1)} \quad (4)$$

Otra aproximación involucra “un proceso conocido como *randomization* o de aleatorización donde los datos de las unidades espaciales se intercambian (permutan) al azar obteniéndose diferentes valores de autocorrelación que se comparan con el valor obtenido. En este caso existe una postura apriorística, es decir que no hay una asunción previa acerca de la estructura de la distribución de tal forma que cada valor observado podría haber ocurrido por igual en todas las localizaciones “(Sánchez Moral, 2004: 217). Al realizar este tipo de *test*, inicialmente hay que definir la hipótesis nula que responde a la afirmación **H₀ la configuración espacial se produce de manera aleatoria**, y la alternativa **H_a la configuración espacial no se produce de manera aleatoria**. Luego se especifica el nivel de significancia que indica la probabilidad de rechazar la hipótesis nula siendo ésta verdadera. Por lo tanto, es la probabilidad que se está dispuesto a cometer al aceptar la hipótesis alternativa. Se suele elegir de acuerdo a la importancia del problema y generalmente es del 5 % (0.05) y 1 % (0.01) (Buzai y Baxendale, 2006).

Asimismo, el *p-valor* es el resultado que nos brinda el *test* de hipótesis. Si el nivel de significancia es superior al *p-valor*, se rechaza la hipótesis nula y se acepta la alternativa. Por el contrario, si se comprueba la hipótesis nula se puede decir que la configuración espacial se produce de forma aleatoria.

Para el análisis local, al igual que en el caso del global, una vez aplicado el procedimiento estadístico, a continuación hay que determinar su distribución. Como

este desarrollo es todavía bastante novedoso no se ha podido establecer su distribución exacta. Es por ello que en la práctica también se recomienda recurrir a procedimientos como el de aleatorización y permutación (Anselin, 1995). El valor esperado de I de Moran para una observación individual es:

$$E[I] = -w/(n-1) \quad (5)$$

Donde w es la suma de los elementos de la fila.

También, para este caso, se puede adoptar una medida más practica como la que postula O'Sullivan y Unwin (2003) que consiste en concentrarse directamente en los mapas resultantes y utilizarlos para hacer nuevas preguntas y sugerir nuevas teorías en lugar de preocuparse por procedimientos de significancia complejos que todavía no está perfeccionados.

Análisis Exploratorio de Datos Espaciales (AEDE)

Las primeras ideas de AEDE (o ESDA en inglés, *Exploratory Spatial Data Analysis*) nacieron a fines de los ochenta y se consolidaron en los noventa a través de paquetes informáticos específicos en los cuales confluyen un conjunto de técnicas destinadas a detectar esquemas de asociación espacial, concentraciones locales y regímenes espaciales presentes en un conjunto de datos para los que las características de localización resultan esenciales (Anselin, 1999). En la actualidad, las técnicas estadísticas se encuentran más fácilmente al alcance de los analistas, debido al desarrollo, abaratamiento y "amigabilidad" de los programas informáticos. Sin embargo, en lo que se refiere a su conocimiento y aplicación, el análisis estadístico-informático de los datos geográficos—espaciales—se encuentra poco desarrollado dentro de la Geografía si se la compara con otras disciplinas. En la actualidad esta situación está reflejada en los principales Sistemas de Información Geográfica (SIG) que solamente aplican una porción de los métodos de análisis espacial existentes, aunque cada vez están más presentes en este tipo de plataformas (Chasco Irigoyen, 2006).

El AEDE busca crear una interfaz intuitiva y fácil de usar para estudiar la información geográfica que aliente la exploración, permitiendo a los usuarios descubrir patrones y anomalías en los datos que de otro modo no serían aparentes. Como tal, las pruebas de su éxito parecen estar más asociadas con los programas de análisis estadístico tradicionales que con los SIG, más reconocidos por su complejidad y por el largo aprendizaje necesario para poder manejarlos. Asimismo se destacan por la capacidad de generar gráficos estadísticos y mapas que favorecen la exploración de las características del problema; por lo tanto, es capaz de mostrar simultáneamente dos "espacios": el geográfico y el matemático (Bosque Sendra, 2005).

En la actualidad el análisis preliminar de datos y el análisis exploratorio son considerados por los investigadores como técnicas esenciales para obtener resultados interpretables y evitar conclusiones superficiales y erróneas.

“El examen cuidadoso, el procesamiento, la comprobación y la representación gráfica, objeto esencial del análisis preliminar de datos, constituyen la tendencia general en el tratamiento de la información. Estas técnicas no suponen una vuelta a un empirismo ciego, sino que constituyen a un punto de vista basado en el principio de que el investigador conoce, o al menos debe conocer, más acerca de los datos que el ordenador” (Estebanéz, 1987:58).

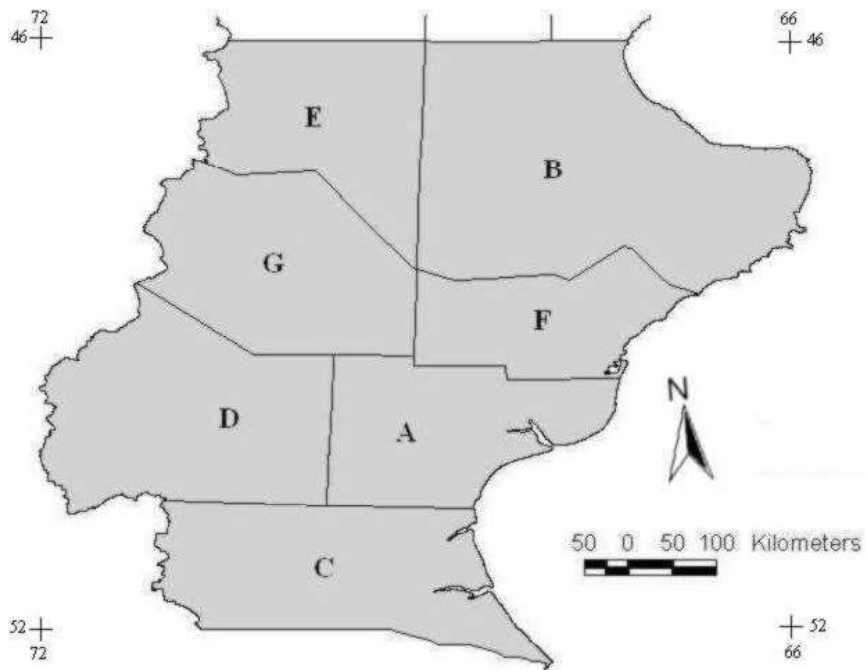
Es por ello que, considerando su incipiente uso y limitaciones, hoy en día el análisis espacial es mejor visto como una técnica de exploración, más adecuado para la generación de hipótesis, que a una estricta confirmación de la teoría dado que facilita el estudio del problema a tratar cuya finalidad es la de conocer adecuadamente sus complejidades y facilitar la formulación de presunciones que lo puedan resolver. En otras palabras, las técnicas del AEDE son muy eficaces en situaciones en las que no existe un marco formal o teoría previa acerca del fenómeno que se analiza. Estas situaciones se plantean muy a menudo en el campo de las ciencias sociales, cuando se analizan grandes bases de datos geográficos cuya distribución no tiene por qué ser conocida a priori.

Las técnicas del AEDE se encuentran incorporadas en el software GeoDa que se destaca por su sencilla interfase visual e interactiva, que no exige un conocimiento previo de la tecnología SIG, mucho más compleja. GeoDa ha sido desarrollado por la Universidad de Illinois, y tiene la ventaja de ser, hasta el momento, un producto gratuito, de libre acceso en Internet que funciona con los conocidos archivos con extensión *shape* (*.shp) y que fue utilizado en el presente trabajo. Los programas que contemplan a los AEDE también pueden ser considerados dentro de una estructura mayor denominada Sistema de Ayuda a la Decisión Espacial (SADE) que favorecen la exploración de los problemas, con la finalidad de llegar a formular hipótesis que lo resuelvan (Bosque Sendra et al, 2000).

¿Cómo se calcula la autocorrelación espacial? Un recorrido de lo local a lo global

Para ejemplificar el funcionamiento del procedimiento de AE se seleccionó una variable del último Censo Nacional de Población, Vivienda y Hogares (porcentaje de viviendas con baño de uso exclusivo) y se aplicó a los departamentos de la provincia de Santa Cruz (ver Figura 4). Esta elección se debe al escaso componente poblacional que posee esta jurisdicción que permite simplificar el examen. La finalidad es conocer la estructura y desarrollo de la AE y no un análisis detallado de la variable que solamente fue elegida para ilustrar el procedimiento.

Figura 4
Departamentos de la Provincia de Santa Cruz



Fuente: elaboración personal

Tabla 1
Porcentaje de viviendas con baño de uso exclusivo por departamento en la Provincia de Santa Cruz

<i>Departamento</i>	<i>% Baño Exclusivo</i>
A) Corpen Aike	98,92
B) Deseado	95,36
C) Guer Aike	96,35
D) Lago Argentino	97,26
E) Lago Buenos Aires	98,21
F) Magallanes	98,33
G) Río Chico	96,15
Promedio	97,23
Varianza	10,46

Fuente: CNPVyH, 2001

El primer paso es establecer la matriz de pesos a partir de la cantidad de vecinos que tiene cada departamento. Se utilizó la relación de contigüidad tipo *Rook*. Por ejemplo, el departamento A tiene cuatro vecinos a los que les corresponde un 1 en la matriz y 0 al resto.

Tabla 2
Matriz de pesos

	A	B	C	D	E	F	G	Total
A	0	0	1	1	0	1	1	4
B	0	0	0	0	1	1	0	2
C	1	0	0	1	0	0	0	2
D	1	0	1	0	0	0	1	3
E	0	1	0	0	0	0	1	2
F	1	1	0	0	0	0	1	3
G	1	0	0	1	1	1	0	4
Suma								20 ¹

Fuente: elaboración personal

A continuación se estandarizan los pesos para que cada fila suma 1. Este procedimiento no es necesario pero facilita la interpretación de las estadísticas (Anselin, 1995: 95).

Tabla 3
Matriz de pesos estandarizada

	A	B	C	D	E	F	G	Total
A	0	0	0,25	0,25	0	0,25	0,25	1
B	0	0	0	0	0,5	0,5	0	1
C	0,5	0	0	0,5	0	0	0	1
D	0,33	0	0,33	0	0	0	0,33	1
E	0	0,5	0	0	0	0	0,5	1
F	0,33	0,33	0	0	0	0	0,33	1
G	0,25	0	0	0,25	0,25	0,25	0	1

Fuente: elaboración personal

En base a la matriz estandarizada se puede aplicar la formula (3) para conocer el valor local de *I* de Moran para cada unidad espacial (departamento) y conocer como contribuye al valor global:

$$I_A = (98,92 - 97,23) / 10,46 \times [(0,25 \times -0,88) + (0,25 \times 0,03) + (0,25 \times 1,10) + (0,25 \times -1,08)]$$

$$0,1616 \times [-0,22 + 0,0075 + 0,275 + (-0,27)] = -0,0335$$

$$I_B = (95,36 - 97,23) / 10,46 \times [(0,5 \times 0,98) + (0,5 \times 1,10)]$$

$$-0,1788 \times [0,49 + 0,55] = -0,1859$$

¹ La suma de los pesos es igual a $S_0 = \sum_{i=1}^n \sum_{j=1}^n w_{ij}$ que es una parte de la ecuación global de *I* de Moran

$$I_C = (96,35 - 97,23) / 10,46 \times [(0,5 \times 1,69) + (0,5 \times 0,03)] - 0,0841 \times [0,845 + 0,015] = -0,0723$$

$$I_D = (97,26 - 97,23) / 10,46 \times [(0,33 \times 1,69) + (0,33 \times -0,88) + (0,33 \times -1,08)] + 0,0029 \times [0,5577 + (-0,2904) + (-0,3564)] = -0,0003$$

$$I_E = (98,21 - 97,23) / 10,46 \times [(0,5 \times -1,87) + (0,5 \times -1,08)] + 0,0937 \times [-0,935 + (-0,54)] = -0,1382$$

$$I_F = (98,33 - 97,23) / 10,46 \times [(0,33 \times 1,69) + (0,33 \times -1,87) + (0,33 \times -1,08)] + 0,1051 \times [(0,5577 + (-0,6171) + (-0,3564))] = -0,0437$$

$$I_G = (96,15 - 97,23) / 10,46 \times [(0,25 \times 1,69) + (0,25 \times 0,03) + (0,25 \times 0,98) + (0,25 \times 1,10)] - 0,1032 \times [0,4225 + 0,0075 + 0,245 + 0,275] = -0,0980$$

Si se desea conocer en detalle los cálculos de la primera parte de la ecuación, z_i / m_2 entonces la tabla 4 es de utilidad:

Tabla 4
Cálculo de z_i / m_2 para cada unidad espacial

Departamento	% Baño Exclusivo (x_i)	Promedio (\bar{x})	$x_i - \bar{x}$ (z_i)	Varianza (m_2)	z_i / m_2
A) Corpen Aike	98,92	97,23	1,69	10,46	0,1616
B) Deseado	95,36	97,23	-1,87	10,46	-0,1788
C) Guer Aike	96,35	97,23	-0,88	10,46	-0,0841
D) Lago Argentino	97,26	97,23	0,03	10,46	0,0029
E) Lago Buenos Aires	98,21	97,23	0,98	10,46	0,0937
F) Magallanes	98,33	97,23	1,10	10,46	0,1051
G) Río Chico	96,15	97,23	-1,08	10,46	-0,1032

La suma de los valores locales conforman el valor global de I de Moran:

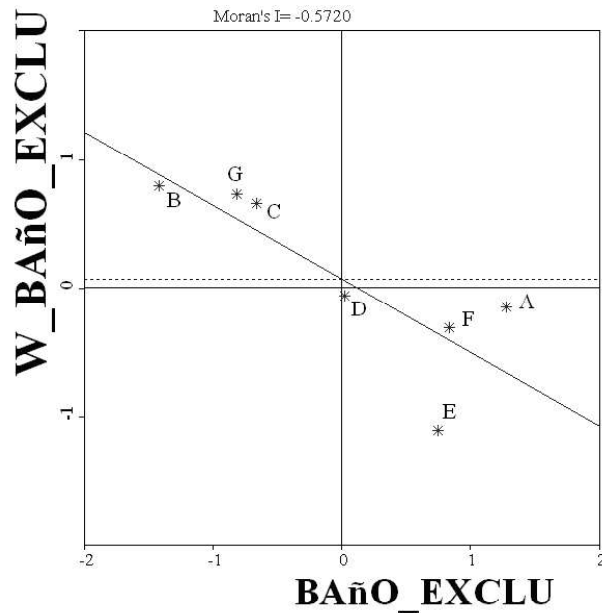
$$\sum_{i=1}^n I_i = \sum_{i=1}^n z_i \sum_{j=1}^n w_{ij} z_j$$

Es decir,

$$I = I_A + I_B + I_C + I_D + I_E + I_F + I_G = -0,5720$$

El resultado se puede visualizar en la parte superior del *scatterplot* (ver Figura 5) que, además, permite ubicar en los cuadrantes los valores estandarizados de la variable en cuestión para cada unidad espacial.

Figura 5
Moran Scatterplot para porcentaje de viviendas con baño de uso exclusivo por departamento en la Provincia de Santa Cruz



Fuente: elaboración personal en base al programa GeoDa

Como se mencionó con anterioridad en la fórmula (4), el valor esperado de I de Moran sería el siguiente:

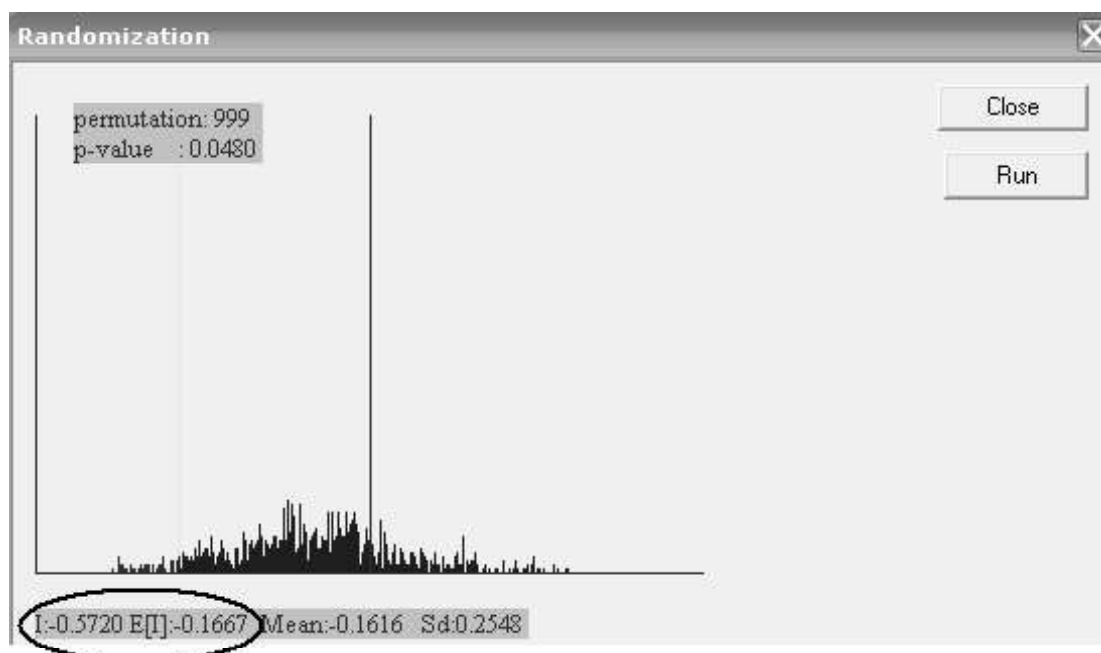
$$E[I] = -1/6 = -0,1667$$

La otra alternativa para probar si la configuración es aleatoria es por medio de las permutaciones ya explicadas y que utiliza el GeoDa. Esta prueba se basa en la idea de que si hay n observaciones sobre una región en particular, $n!$ permutaciones de los datos son posibles². El valor de I puede obtenerse por cualquiera de las permutaciones. Luego se puede comparar el valor observado de I respecto a la distribución de permutaciones para determinar si es significativo (Lloyd, 2007).

En la Figura 6 se puede observar el valor global de la I de Moran y su valor esperado, $E[I]$ calculado por medio del procedimiento de aleatorización. A través de las permutaciones se generan 999 nuevos valores de I de Moran que componen una distribución de la cual se destaca la línea que determina el promedio de la misma. En la parte inferior izquierda del gráfico se registran el valor observado I : -0,5720 y el esperado $E[I]$: -0,1667. Ambos resultados coinciden con los cálculos realizados con anterioridad.

² En este estudio existen 7! posibilidades de permutaciones, es decir, $7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1 = 5.040$

Figura 6
Aleatorización



En este caso el *p-valor* (*p-value*) es de 0.048 y por ser menor a 0.05 se rechaza la hipótesis nula y se acepta la alternativa. De todas maneras, los valores son cercanos de manera que hay que tomarlos con caución, recordando nuevamente la finalidad explicativa que tiene el presente trabajo. Cómo los resultados no son azarosos continuamos explorando la información de las unidades espaciales por medios de los mapas de significancia y *cluster*.³

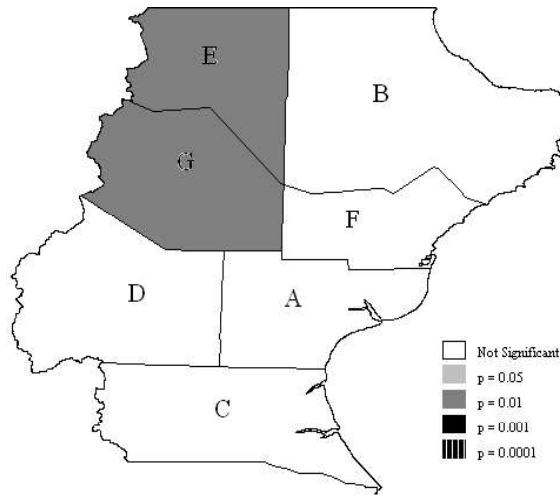
El primero (ver Figura 7) nos muestra la probabilidad de que las relaciones de contigüidad se produzcan de manera aleatoria y solamente se destacan dos significativos, el departamento E (Lago Buenos Aires) y el G (Río Chico) con un valor que denota una probabilidad de error de 0.01 (1 en 100). En el resto predomina el color blanco que denota la falta de significancia.

En el segundo (ver Figura 8) solamente se hacen presentan tres categorías de las cinco posibles. Predomina el color blanco que significa que los valores no son significativos para los departamentos que lo poseen. No se observan *clusters*, pero sí se registran dos *outliers*, el departamento E (Lago Buenos Aires) y el G (Río Chico) que ya aparecen en el mapa anterior. Se destacan porque sus valores son opuestos a los de sus vecinos, es decir que registran un comportamiento alto-bajo y bajo-alto respectivamente. Esto se observa al notar que los vecinos de E (G y B) se encuentran en un cuadrante opuesto y lo mismo sucede con los vecinos de

³ Los productos cartográficos del GeoDa carecen de la posibilidad de asignarles coordenadas, escala y norte. Un detalle menor teniendo en cuenta su potencialidad y carácter gratuito.

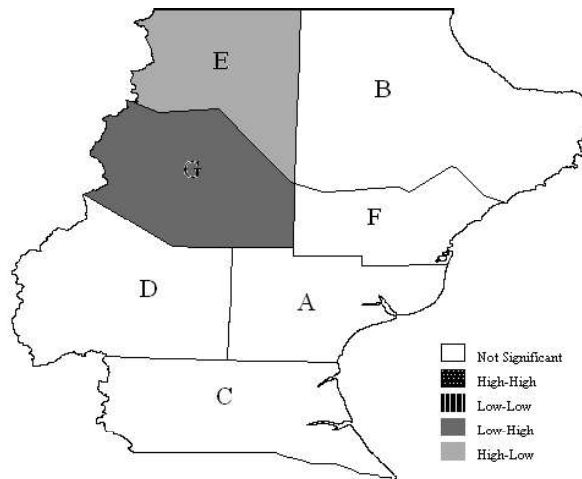
G (E, F, D y A). De esta manera “rompen” con la tendencia esperable de similitud entre las unidades espaciales propia de la autocorrelación espacial.

Figura 7
Mapa de Significancia



Fuente: elaboración personal en base al programa GeoDa

Figura 8
Mapa cluster



Fuente: elaboración personal en base al programa GeoDa

En definitiva, estos dos departamentos son los que tienen un comportamiento remarcado sobre el resto, de manera que pueden generarse nuevas hipótesis e interrogantes que traten explicar o comprender los motivos que conllevan a esta situación diferencial en base a los aportes exploratorios realizados a través del AEDE.

Reflexiones finales

El análisis de la información geográfica requiere de herramientas propias que van más allá de las convencionales técnicas estadísticas y, por tanto, de programas informáticos específicos que permitan afrontar las particularidades inherentes de los datos espaciales. Este tipo de plataformas denominadas Sistemas de Ayuda a la Decisión Espacial (SADE) con la cuales se podrá realizar un completo Análisis Exploratorio (AEDE o ESDA en inglés) proveen al geógrafo de nuevos instrumentos para sus investigaciones a través de una interfase de uso más sencillo que las presentes en los Sistemas de Información Geográfica.

En particular, tienen relevancia para el estudio de la autocorrelación espacial, un fenómeno sustancialmente geográfico que nos permite observar los patrones de distribución en el espacio de las variables. El mismo parte de una simple premisa que postula que en el espacio geográfico todo se encuentra relacionado con todo, pero los espacios más cercanos están más relacionados entre sí. Este procedimiento registró un importante avance en los últimos años a partir de la aparición del análisis de patrones locales de asociación espacial (LISA, Local Indicators of Spatial Association) que descompone el índice global de autocorrelación y verifica en cuánto contribuye cada unidad espacial a la formación del valor general, permitiendo capturar de forma simultánea el grado de asociación espacial y la heterogeneidad resultante del aporte de cada unidad espacial.

Como la autocorrelación espacial posee una estructura compleja, el presente trabajo intentó destacar su importancia y contribuir a su difusión y entendimiento a partir del examen de una variable del último censo aplicada a los departamentos de la Provincia de Santa Cruz, detallando como cada uno de ellos contribuye al valor global.

Por último, por medio de un programa de AEDE denominado GeoDa, se continuó el análisis exploratorio a través de la obtención de la significancia estadística de la autocorrelación espacial y de la elaboración de mapas que permitieron desagregar aún más la información de la variable utilizada.

Bibliografía

ANSELIN, Luc. "The Moran Scatterplot as an ESDA tool to asses instability in local association". *GISDATA*. Specialist Meeting on Gis and Spatial Analysis. The Netherlands, 1993, 23 pp.

ANSELIN, Luc. "Local Indicators of Spatial Association-LISA". *Geographical Analysis*. Ohio: Ohio State University Press, 1995, vol. 27, nº 2, pp. 93-115.

ANSELIN, Luc. *Spatial Econometrics*. Dallas, United States: School of Social Sciences University of Texas at Dallas, 1999.

ANSELIN, Luc. "Spatial Autocorrelation Refresher". *CSISS*. [En línea]. Illinois: Dept. Agricultural and Consumer Economics University of Illinois, Urbana-Champaign, 2003, sal.agecon.uiuc.edu/courses/se/pdf/w2_spauto_slides.pdf [20 de marzo de 2009].

ANSELIN, Luc. *GeoDa 0.9.5-i Release Notes*. Illinois, United States: Urbana-Champaign Spatial Analysis Laboratory. Department of Agricultural and Consumer Economics, 2004, 64 pp.

ARBIA, Giuseppe. *Spatial Econometrics*. Berlin: Springer, 2006.

BOSQUE SANDRA, Joaquín "Nuevas perspectivas en la enseñanza de las tecnologías de la información geográfica". *Serie Geográfica*. Madrid: Universidad de Alcalá de Henares, 1999, n° 8, pp. 25-34.

BOSQUE SENDRA, Joaquín. "La Ciencia de la Información Geográfica y la Geografía". En: *Actas VII Encuentro de Geógrafos de América Latina*. San Juan de Puerto Rico, 1999, 15 pp.

BOSQUE SENDRA, Joaquín; GÓMEZ DELGADO, Montserrat *et al* . "Hacia un sistema de ayuda a la decisión espacial para la localización de equipamientos". *Estudios Geográficos*, 2000, n° 241, pp. 567-598.

BOSQUE SENDRA, Joaquín. "Espacio geográfico y ciencias sociales. Nuevas propuestas para el estudio del territorio". *Investigaciones Regionales*. Sección Panorama y Debates. España: Asociación Española de Ciencia Regional, 2005, pp. 203-221.

BUZAI, Gustavo y BAXENDALE, Claudia. *Análisis Socioespacial con Sistemas de Información Geográfica*. Buenos Aires: Lugar Editorial, 2006.

BUZAI, Gustavo y BAXENDALE, Claudia. "Autocorrelación espacial univariada y bivariada. Aspectos socio-educativos en la ciudad de Luján". En: *Anuario de la División Geografía*. Luján: Departamento de Ciencias Sociales – UNLu. Luján, 2004, pp. 167-180.

BUZAI, Gustavo. "Los Sistemas de Información Geográfica y sus métodos de análisis en el continuo resolución-integración". En *Memorias X Conferencia Iberoamericana de Sistemas de Información Geográfica (X CONFIBSIG)*. San Juan de Puerto Rico, 2005, 17 pp.

CEPEDA, Rosana y VELÁZQUEZ, Guillermo. "Análisis de asociación espacial en variables de calidad de vida en Tandil". En GARCÍA, María C. (comp.) *Ciudades intermedias. Problemas de su estructura y funciones. Conflictos ambientales y sociales en los años 2000*. Tandil: CIG-REUN, 2005, pp. 53-59.

CHASCO YRIGOYEN, C. "Análisis estadístico de datos geográficos en geomarketing: el programa GeoDa". *Distribución y Consumo*, 2006, n° 2, pp. 34-45.

ESTÉBANEZ ALVAREZ, José. "La imagen de la geografía cuantitativa elaborada por sus oponentes". *Anales de geografía de la Universidad Complutense*, Madrid: Departamentos de Geografía Humana y Geografía Física-Análisis Geográfico Regional, 1987, n° 7, pp. 53-59.

GAMIR ORUETA, Agustín, SEGUÍ PONS, Joana María, *et al*. *Prácticas de análisis espacial*. Barcelona: Oikos-Tau, 1995.

GETIS, A. "A History of the Concept of Spatial Autocorrelation: A Geographer's Perspective" *Geographical Analysis*. 2008, Vol. 40, n° 3, pp. 297-309.

GOODCHILD, Michael. *Spatial Autocorrelation*. Norwich, United Kingdom: Geo Books, 1986.

GOODCHILD, Michael. "The validity and usefulness of laws in geographic information science and geography". *Annals of the Association of American Geographers*. Malden, United States: Association of American Geographers, 2004, n° 94(2), pp. 300–303.

GOODCHILD, Michael. "Spatial autocorrelation". In KEMP, Karen (Ed.) *Encyclopedia of Geographic Information Science*. Thousand Oaks, California: SAGE, 2008, pp.397-398.

KEYLOCK, Christopher y DORLING, Danny. "What kind of quantitative methods for what kind of geography?". United Kingdom: Royal Geographical Society, 2004, vol. 36, n° 4, 358–366.

LLOYD, Christopher. *Local models for spatial analysis*. United States: CRC Press, 2007.

LUCERO, Patricia y CELEMÍN, Juan "La calidad de vida de la población en la determinación de la calidad territorial. Un estudio de autocorrelación espacial aplicado a la ciudad de Mar del Plata." *GeoFocus* (Artículos). Madrid: Grupo de Tecnologías de la Información Geográfica, Asociación de Geógrafos Españoles, 2008, n° 8, pp. 94-114.

MILLER, Harvey J. "Potential contributions of spatial analysis to geographic information systems for transportation (GIS-T)". *Geographical Analysis*. Ohio: Ohio State University Press, 1999, n° 31, 373-399. MILLER, Harvey J. "Tobler's First Law and Spatial Analysis". *Annals of the Association of American Geographers*. Malden, United States: Association of American Geographers, 2004, n° 94(2), pp. 284–289.

ODLAND, John. *Spatial Autocorrelation*. Scientific Geography Series, California: SAGE Publications, 1988.

O'SULLIVAN, David y UNWIN, David. *Geographic Information Analysis*. New Jersey: Wiley, 2002.

SÁNCHEZ MORAL, Simón. "El estudio econométrico de la concentración espacial de la industria: ejemplo de aplicación en Madrid, Toledo y Guadalajara". *Anales de geografía de la Universidad Complutense*, Madrid: Departamentos de Geografía Humana y Geografía Física-Análisis Geográfico Regional, 2004, n° 24, pp. 207-227.

SEGUINOT BARBOSA, José. "Pasado, presente y futuro de los Sistemas de Información Geográfica". En SEGUINOT BARBOSA, J. (Ed.). *Geonatura. Sistemas de Información Geográfica (SIG) aplicados a las Ciencias Ambientales y de la Salud*. San Juan de Puerto Rico: 2001, pp. 120-141.

SUI, Daniel. "Tobler's First Law of Geography: A Big Idea for a Small World?". *Annals of the Association of American Geographers*. Malden, United States: Association of American Geographers, 2004, n° 94(2), pp. 269–277.

TOBLER, Waldo. "A Computer Movie Simulation Urban Growth in the Detroit Region". *Economic Geography*, 1970, vol. 46, n° 2, pp. 234-240.

TOBLER, Waldo. "On the First Law of Geography: A Reply". *Annals of the Association of American Geographers*. Malden, United States: Association of American Geographers, 2004, vol. 94(2), pp. 304–310.

VILALTA Y PERDOMO, Carlos J. "Sobre la espacialidad de los procesos electorales urbanos y una comparación entre las técnicas de regresión ols y sam". *Estudios Demográficos y Urbanos*. México D.F.: El Colegio de México, 2005, vol. 21, n° 1 (61), pp. 83-122.

VILALTA Y PERDOMO, Carlos J. "Como enseñar autocorrelacion espacial". *Economía, Sociedad y Territorio*. México D.F.: El Colegio Mexiquense, 2005, vol. V, n° 18, pp. 323-333.

UNWIN, David. "GIS, spatial analysis and spatial statistics". *Progress in Human Geography*. California: SAGE Publications, 20,4, 1996, pp. 540-551.

Fecha de recepción: 27 de marzo de 2009

Fecha de aprobación: 8 de febrero de 2010