

GEOGRAPHIC DATA BASES AND SPACE TO IDENTIFY AUTOCORRELATION SPATIAL DISTRIBUTION PATTERNS

BASES DE DATOS GEOGRAFICAS Y AUTOCORRELACIÓN ESPACIAL PARA IDENTIFICAR PATRONES DE DISTRIBUCIONES ESPACIALES

Recibido: 12 de Diciembre 2013- aceptado: 30 de Enero 2014

Yair Rivera Julio¹
Corporación Universitaria Americana

Keywords:

SIG; índice de Moran.

Abstract

The Geographic Information Systems (GIS) methods combined with space-time statistics are an effective tool for studying exploratory analysis of spatial variables, especially one that has an influence with herself near or adjacent areas. Through spatial correlation method Auto Moran and aid the iterative technology systems and mobile devices, it was determined that all the phenomena of expansion and concentration can be analyzed in real time, following a statistical-mathematical model, which allow to simulate nature and mobility of the variable in different areas of direct and simple way, all this to identify spatial patterns over time. All this mathematical analysis in the near future allow the creation of intelligent mobile web services supported using location-based services and GIS for monitoring real-time spatial variables.

Palabras clave:

SIG; índice de Moran.

Resumen

Los sistemas de información geográfica combinados con los métodos de estadística espacio-temporal, constituyen una efectiva herramienta de análisis exploratorio para el estudio de variables espaciales, en especial aquella que tiene una influencia consigo misma en zonas próximas o adyacentes. A través del método de Auto correlación espacial de Moran y con ayudas de las tecnologías de los sistemas iterativos y dispositivos móviles, se determinó que todos los fenómenos de expansión y concentración pueden ser analizados en tiempo real, siguiendo un modelo estadístico-matemático, el cual permitiría simular la naturaleza y movilidad de la variable en diferentes zonas de manera directa y sencilla, todo esto para la identificación de patrones espaciales a través del tiempo. Toda este análisis matemático permitiría en un futuro cercano la creación de servicios móviles inteligentes soportados en la web con ayuda de servicios basados en la localización y GIS para el monitoreo de variables espaciales en tiempo real.

1. Telemática y Telecomunicaciones, Uninorte 2012. Docente investigador de ingeniería, Corporación Universitaria Americana. Barranquilla, Colombia. E-mail: yproenza@uci.cu.

* Plataforma Informática Interactiva Basada en Neo-geografía y la Correlación Espacial para Identificar Patrones de Distribuciones Espaciales

I. INTRODUCCION

En la actualidad las variables geo-espaciales juegan un papel protagonista como referencia caracterizadora de grupos, de ahí la importancia y análisis de su estudio, pues nos permite obtener un mayor conocimiento a través de diferentes capas organizadas de información, este procesamiento integrado de datos nos lleva descubrir patrones y tendencias que con métodos tradicionales sería muy difícil descubrir [1], más aun sería más relevante si a todo lo anterior le sumamos el desarrollo del internet, métodos estadísticos, representaciones gráficas y dispositivos móviles los cuales nos ayudan a incorporar una nueva dimensión en el análisis de los datos en tiempo real para el desarrollo de sistemas inteligentes basados en la geolocalización, como la telemedicina y el geomarketing [2], [3].

II. IMPORTANCIA DE LA AUTOCORRELACIÓN ESPACIAL (AE) EN LOS SISTEMAS GEOGRÁFICOS [4].

Desde el contexto espacial se puede explorar el comportamiento de una variable a través del tiempo y el espacio; a través de un SIG se puede modelar este comportamiento determinando sus similitudes o diferencias; es decir, qué tanto varía su valor en el plano espacial. En síntesis, la auto correlación espacial refleja el grado en que objetos y actividades de una unidad geográfica son similares a otros objetos o actividades próximas entre sí; por ejemplo, la delincuencia suele ser menor en las cercanías de las comisarías de policía, denotando una auto correlación espacial negativa entre casos de delincuencia y presencia policial.

La dependencia o auto correlación espacial aparece como consecuencia de la existencia de una relación funcional entre lo que ocurre en un punto determinado del espacio y lo que ocurre en otro lugar [4]; es decir, el valor que toma una variable en una región no viene explicado únicamente por condiciones internas sino también por el valor de esa misma variable en otras regiones vecinas o próximas [5], [6].

La auto correlación espacial nos muestra como una variable está explícitamente relacionada en una distribución espacial; es decir todo está relacionado con todo pero los espacios adyacentes tienen una relación más íntima [7], esta relación nos permite asociar patrones de comportamientos en diferentes áreas geográficas, de tal manera que la variación de cualquier fenómeno depende de su ubicación y de las áreas contiguas. Podemos definir así un modelo matemático

que definen la homogeneidad o la heterogeneidad de una variable con función al espacio, disminuyendo al mismo tiempo el factor de aleatoriedad [8].

Vemos como se relacionan los diferentes espacios adyacentes según una misma variable a considerar

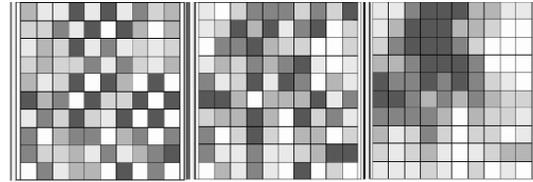


Figura 1. Ejemplos y tipos de auto correlación espacial

Fuente: ASELIN 2003[9] NEGATIVA ALEATORIA POSITIVA

La AE (auto correlación espacial) nos permitiría entonces analizar una variable a correlacionar en forma conjunta, teniendo en cuenta la dependencia entre su localización y su valor [10]. Éste es un modelo bastante característico para observar una configuración espacial fragmentada. Por ejemplo, el número de accidentes en áreas determinadas adyacentes, identificación de aglomeraciones y clústeres, etc. [11].

Calculo de la Auto correlación Espacial.

$$\text{Ec. (1)} \sum_{i=1}^n \sum_{j=1}^n W_{ij} \cdot C_{ij} \quad (12)$$

Donde n es el total de lugares o secciones en que se divide el área y W_{ij} son los elementos de una matriz (matriz de conexiones, contigüidad o de pesos espaciales) cuyos valores se caracterizan por una medida de contigüidad (Rook, Bishop o Queen) o tipo de relación adyacente (ver fig. 2). El valor C_{ij} es una medida de la proximidad (distancia) o contigüidad (áreas vecinas adyacentes) de los valores i y j, o cualquier distancia definida por el usuario [12]. Para nuestro estudio emplearemos una medida de contigüidad tipo QUEEN donde las zonas vecinas son las zonas adyacentes [13].

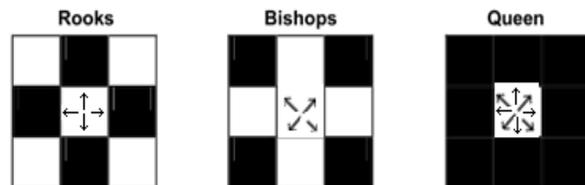


Figura 2. Tipos de contigüidad o medida de adyacencia [14]

Esta medida de adyacencia se materializa a través de una matriz de pesos espaciales o matriz de contigüidad [15]:

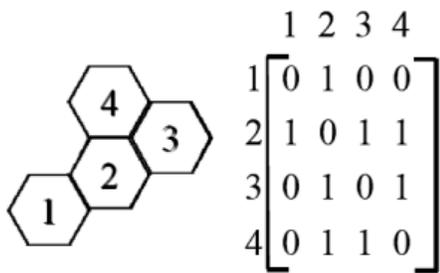


Figura 3. **Matriz de contigüidad [16]**

La matriz de contigüidad nos define los pesos o número de vecinos adyacentes de una zona determinada, como vemos en la figura 3 cada pentágono representa un área. Si 2 pentágonos comparten un vértice o un punto en común se escribe un 1 en la matriz de contigüidad; en caso contrario se anota 0, es decir el pentágono 1 tiene como adyacentes al pentágono 2 por tal motivo en la matriz de contigüidad habrá un 1 en la casilla (1,2)[17].

Nota: Ninguna de las áreas es vecina es adyacente consigo misma, por tal motivo la diagonal principal de la matriz de adyacencia siempre será 0.

Teniendo definida la matriz de contigüidad se procede a desarrollar el índice de correlación espacial para determinar la influencia entre las áreas, este será un indicador que permite capturar que tanto contribuyen cada unidad espacial local a la formación del valor general, o las variables que la componen en el contexto global [18], obteniendo así un valor de significancia para cada sub área, es decir nuestra variable global era el promedio del número de visitas que recibía el zoológico por día y la variable local el promedio del número de visitas que recibía cada sección del zoológico, su relación y valor determina un tipo de dependencia espacial visible.

El índice de Moran está definido a través de la siguiente fórmula:

$$I_i = \frac{n}{S_0} \frac{\sum_{i=1}^n \sum_{j=1}^n W_{ij} Z_i Z_j}{\sum_{i=1}^n Z_i^2} \quad S_0 = \sum_{i=1}^n \sum_{j=1}^n W_{ij} \quad (2)$$

Donde o Sumatoria de pesos de contigüidad de la matriz de contigüidad, Z representa las desviaciones de la media $(x_i - \bar{x})$ o también $(x_j - \bar{x})$, donde x_j es la variable a correlacionar en las diferentes áreas. Este índice define cuanto puede aportar cada área como unidad espacial correlacionada a la integración del valor global o general [19]; Es decir La suma de los índices locales conforman el valor global de I de Moran observado:

Valor global de I

$$E(I) = \sum_{i=1}^n I_i$$

El valor esperado de este índice individual está definido

por: $I = \frac{-W}{n-1}$ Donde W es la suma de los elementos de la fila [20].

Finalmente podemos decir que los índices locales I_i permiten explorar visualmente patrones de aglomeraciones en función de la variable correlacionada y las áreas adyacentes. De igual manera se visualiza la heterogeneidad o homogeneidad de los patrones según el valor del índice [21], como un coeficiente de correlación, varía entre +1 y -1, aunque fácilmente puede superar ambos límites, para el primer caso tenemos una auto correlación positiva (Ver gráfica 1) (valores similares entre puntos próximos, sean valores altos o sean valores bajos no los distingue), para el segundo caso los valores son totalmente diferentes y/o poco semejantes [22].

ARQUITECTURA DEL SISTEMA E IMPLEMENTACIÓN

Para la modelación del sistema en tiempo Real (ver fig. 5), el sistema de información es alimentado desde los dispositivos móviles (sistema android), en donde cada móvil a través de una red de datos celular e inalámbrica GSM envía su ubicación espacial (latitud y longitud) hacia una base de datos relacional geográfica (postgres y postgis), al mismo instante un motor estadístico (PL/R) instalado en la base de datos, determina en tiempo real la distribución espacial geográfica de las zonas adyacentes [23], correlacionando el número de visitantes que tiene el Zoológico con sus respectivas áreas o secciones, con el fin de determinar el tipo de patrones de visitas o clústeres formados.

Nota: PL/R. Es un proyecto de software libre (también multiplataforma) que constituye un Conjunto integrado

de utilidades, en este caso librerías para manipulación de datos, cálculos estadísticos y visualización gráfica.

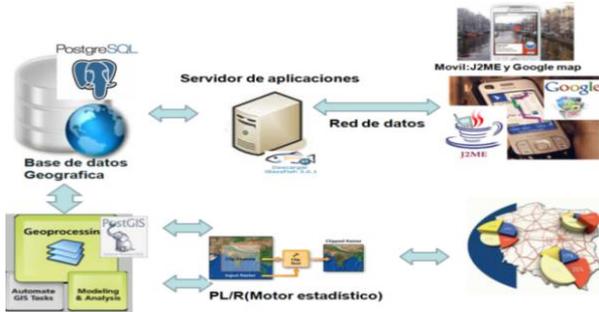


Fig. 5. Arquitectura del sistema

Para la arquitectura del sistema se han usado las siguientes Herramientas: Postgres, Postgis, Netbeans, J2ME, OpenJum, PL/R y JAVA (Ver fig. 5). Fuente: Msc. Yair Rivera Julio

Este software alojado en un servidor web-apache realiza una conexión a través de una API de Google MAP, jscrip y php para graficar los niveles de correlación respectiva (Ver fig. 6)

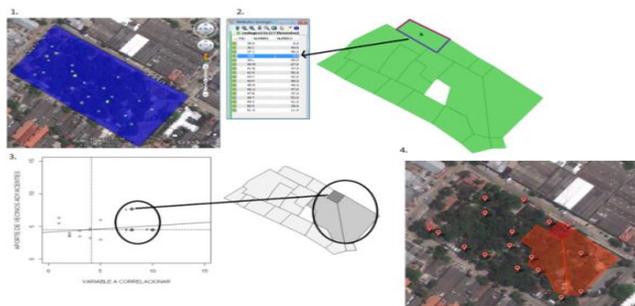


Fig. 6. Vista final desde el navegador

Para el desarrollo de esta arquitectura se tuvo en cuenta los siguientes aspectos necesarios:

1. Digitalizar de coordenadas del zoológico (Google Earth). Se hace solo una vez y sirve para tomar las coordenadas geográficas del área del zoológico y hacer las divisiones necesarias, en que se componen dichas áreas.
2. Almacenamiento de la cartografía en la base de datos: obtenidas e importadas las secciones del zoológico en unidades espaciales (Open jump, Postgres y postgis) en la base de datos geográficas. Cada sub área es identificada por un índice, acompañada por un campo adicional el cual corresponde al número de visitas actuales de cada sub área en tiempo real.

Aplicación del índice de Moran para la correlación espacial de las áreas adyacentes en PL/R(“Motor estadístico”) y Postgre SQL, para esto se desarrolló una consulta la cual nos sirve de interfaces entre estos 2 motores de dato:

```
CREATE OR REPLACE VIEW vistazoo AS
SELECT
zoologico.nombre, count(zoologico.nombre) AS número
FROM
zoologico, puntos
WHERE
st_contains(zoologico.the_geom, puntos.the_geom)
GROUP BY
zoologico.nombre
ORDER BY
zoologico.nombre;
```

A través de una consulta en SQL y las funciones de postgis procedemos a seleccionar el número de puntos que encierra cada unidad de área geográfica (Ver Fig. 7.1), contando así el número de visitantes por unidad de área, posteriormente estos datos son enviados a un procedimiento almacenado hecho en PL/R, el cual determina las áreas adyacentes y con base a esta información aplica el índice de correlación de Moran dando como resultado los clústeres o agrupaciones de las áreas correlacionadas (Ver Fig. 7.3 y 7.2), para esto último se desarrolló el siguiente código:

```
getinfo.shape("C:/SALIDA/zoologico.shp")
sids<-readShapePoly("C:/SALIDA/zoologico.shp")
class(sids)
proj4string(sids)<-CRS("+proj=longlat ellps=WGS84")
sids_NAD<-spTransform(sids, CRS("+init=epsg:3358"))
sids_SP<-spTransform(sids, CRS("+init=ESRI:102719"))
sids_nbq<-poly2nb(sids)
sids_nbq_w<-nb2listw(sids_nbq,style="W")
sids_nbq_wb<-nb2listw(sids_nbq, style="B")
dist<-nbdists(sids_nbq, coordinates(sids_SP))
moran.test(sids_NAD$NUMERO,
listw=sids_nbq_w,alternative="two.sided")
```

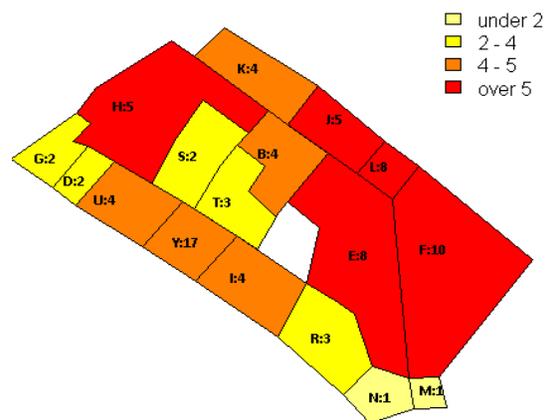


Figura 7.1. Diagrama de densidad

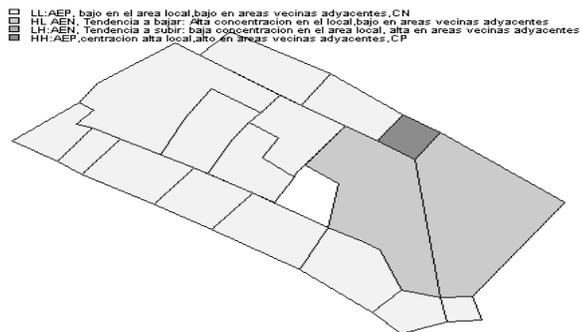


Figura 7.2. Diagrama de correlación espacial [25]

Nota: como resultado cada sub área posee un número que identifica su grado de correlación a nivel global.

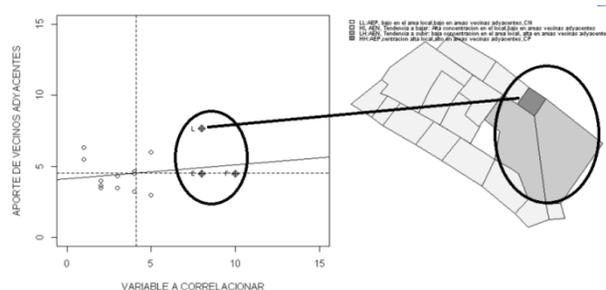


Figura 7.3. Correlalograma y Representación espacial

- Finalmente los valores de correlación son recogidos en la misma consulta SQL y enviados al método que hizo la petición de forma remota desde un servidor web Apache, un script php grafica el diagrama de densidad con ayuda de un API de google MAP.

Finalmente la información es graficada en el navegador de la siguiente manera:



Figura 7.4. Vista final desde el navegador

V. Conclusiones

Teniendo en cuenta todo el análisis anterior concluimos que:

Un fenómeno espacial no puede ser analizado sin tener en cuenta las influencias que el mismo ejerce sobre sus zonas más próximas y sin considerar la forma en que estas últimas lo condicionan.

Todos los fenómenos de difusión y concentración turística pueden ser analizados mediante la contrastación de la hipótesis de auto correlación espacial (positiva y negativa, respectivamente). Para ello, además de los valores observados de la variable en las unidades espaciales analizadas, es necesario considerar también una matriz de interacciones espaciales que determine qué unidades espaciales son vecinas y cuáles no lo son.

Dependiendo de que el análisis de la auto correlación se refiera a la totalidad del territorio analizado o únicamente a determinadas partes del mismo, es posible considerar tanto índices de auto correlación espacial global como índices de auto correlación espacial local.

Además de los índices antes referidos, existen también métodos de visualización que permiten apreciar de forma directa y sencilla la presencia de auto correlación espacial y la ubicación de los clúster y atípicos espaciales. Entre estos métodos se destaca el diagrama de dispersión de Moran [2].

Las áreas que no representan una auto correlación, lo que se sugiere es establecer una reubicación de las exhibiciones adyacentes, ya que el comportamiento espacial sin auto correlación en estas zonas sería aleatorio.

La identificación de los patrones espaciales, en función del número de visitantes, puede definir el estado actual y ayudar a definir un futuro, el estado de cada exhibición de forma individual.

Es efectiva técnica de análisis exploratorio espacial que ayuda a determinar la fuerza de la dependencia espacial, describiendo sus cambios a través del tiempo y ayudando a detectar valores atípicos, puntos calientes y fríos en el espacio.

Los sistemas de información geográfica combinados con los métodos de estadística espacio-temporal, constituyen herramientas de gran utilidad para la resolución de diferentes problemas, entre ellos los relacionados con la identificación de riesgos, tanto naturales como entrópicos.

Se creó una metodología con nuevas perspectivas para aportar un enfoque pragmático y escalable para la creación de geo-procesamiento en flujos de trabajo generados por el usuario con contenido dentro del contexto para dar como resultado una colección de servicios basados en la localización.

La movilidad y el comportamiento estarían definidos por un modelo espacial de correlación, sentando así las bases para una estrategias de despliegue como en el geo-marketing, el cual contaría con una caracterización espacial de los visitantes en tiempo real, y que al mismo tiempo se almacenarían en un sistema persistente, lo que permite a posteriori un filtrado de datos.

Referencias

[1] J. Chen, R. Roth, A. Naito, E. Lengerich, A. Maceachren, Geovisual analytics to enhance spatial scan statistic interpretation: an analysis of U.S. cervical cancer mortality. *International Journal of Health Geographics* 7 (57): 1-18. 2008.

[2] M. Craglia, R. Haining, P. Wiles. A comparative evaluation of approaches to urban crime pattern analysis. *Urban Studies* 37 (4): 711–729. 2000.

[3] S. Merrall, GIS for spatial analysis of fire incidence; Identification of social, economic and environmental risk indicators. In: A. Hirschfield & K. Bowers (Eds.), *Mapping and analysing crime data Lessons from research and practice*: 156–183. Taylor and Francis, London and New York. 2001.

[4] AE, Autocorrelación espacial e indicadores locales de asociación espacial. Importancia, estructura y aplicación, *Revista Universitaria de Geografía*, versión impresa ISSN 0326-8373, Rev. Univ. Geogr. v.18 n.1 Bahía Blanca 2009, Juan Pablo Celemín*,* Becario CONICET - Universidad Nacional de Mar del Plata. Grupo de Estudios Sobre Población y Territorio. jpcelemín@conicet.gov.ar.

[5] J. Corcoran, G. Higgs, Brunson CH, Ware A, Normal P. the use of spatial analytical techniques to explore patterns of fire incidence: A South Wales case study. *Computers, Environment and Urban Systems* 31 (6): 623-647. 2007

[6] R. Moreno, E. Vayá, Técnicas econométricas para el tratamiento de datos espaciales: La econometría espacial. España: Ediciones de la Universidad de Barcelona. 2000.

[7] Análisis espacial de datos y turismo: nuevas técnicas para el análisis turístico. Una aplicación al caso extremeño Sánchez Rivero, Marcelino (Universidad de Extremadura) *Revista de Estudios Empresariales*. Segunda época. Número: 2 Páginas: 48 – 66. 2008.

[8] M. Dodge, The visualisation and analysis of fire incidents using GIS. In: *Joint European conference and exhibition on geographical information*, Barcelona. 1996.

[9] L. Anselin, GeoDa 0.9.5-i Release Notes. Illinois, United States: Urbana-Champaign Spatial Analysis Laboratory. Department of Agricultural and Consumer Economics, pp 64, 2004.

[10] S. Chandler, A. Chapman, S. Hallington, Fire incidence, housing and social conditions – The urban situation in Britain. *Fire Prevention* 172: 15–20. 1984.

[11] Association for Geographic Information (AGI), 1997. Emergency services GI survey. AGI, London. Bivand, Roger, Pebesma, Edzer. *Applied Spatial Data Analysis with R Springer*: 21-268. 2008

[12] Likelihood estimation of very large spatial autoregressive models: A characteristic polynomial Approach. *Computational statistics and Data Analysis* 35: 301-319.

[13] Autocorrelación espacial e indicadores locales de asociación espacial. Importancia, estructura y aplicación, Juan Pablo Celemín, Becario CONICET - Universidad Nacional de Mar del Plata. Grupo de Estudios Sobre Población y Territorio. jpcelemín@conicet.gov.ar

[14] Las ubicaciones con autocorrelación local alta-alta y baja-baja (autocorrelación positiva) son referidas como agrupaciones espaciales y las ubicaciones con autocorrelación alta-baja y baja-alta (autocorrelación negativa) lo son como valores espaciales atípicos.

Mientras que los valores atípicos son ubicaciones individuales por definición, este no es el caso para los clusters (Anselin, 2005).

[15] Stationary Moran Scatter Plot: Spatial Associations with indian literacy data, Hazem Mohamed, 30 March , 2012.

[16] L. Anselin, A. Bera, Spatial dependence in linear regression models with an introduction to spatial econometrics. In: Ullah, A. and Giles, D. E. Editors, Handbook of Applied Econometrics: 237-289. Marcel Dekker, New York. 1998.

[17] M. Ducanson, A. Woodward, P. Reid, Socio-economic deprivation and fatal unintentional domestic fire incidents in New Zealand 1993-1998. Fire Safety Journal 37: 165-179. 2002.

[18] E. Pebema, K. Jong, D. Briggs, Interactive visualization of uncertain spatial and spatio-temporal data under different scenarios: an air quality example. International Journal of Geographical Information Science 21 (5): 515-527. 2007.

[19] Influence of Geomarketing on the Rollout of New Telecom Network Infrastructure KoenCasier, Jan Van Ooteghem, ManonSikkema, SofieVerbrugge, Didier Colle, Mario Pickavet, Piet Demeester Internet Based Communication Networks and Services (IBCN) Department of Information technology, Ghent University – IBBT Gaston Crommenlaan 8 (Bus 201), B-9050 Gent, Belgium {koen.casier, jan.vanooteghem, sofie.verbrugge, didier.colle, mario.pickavet, piet.demeester@intec.ugent.be CTTE 2011 · 16-18 May, Berlin, Germany, 2011.

[20] L. Anselin, Local Indicators of Spatial Association-LISA. Geographical Analysis, 1996 27: 93-115.

[21] J. P. Celemín., Becario CONICET - Universidad Nacional de Mar del Plata. Autocorrelación espacial e indicadores locales de asociación espacial. Importancia, estructura y aplicación, Grupo de Estudios Sobre Población y Territorio. jpcelemín@conicet.gov.ar, 1c plan de investigaciones. Revista Universitaria de Geografía, versión impresa ISSN 0326-8373, Rev. Univ. Geogr. v.18 n.1 Bahía Blanca. 2009.