09 Goodchild y Haining 5/7/05 13:14 Página 175

© *Investigaciones Regionales. 6 – Páginas 175 a 201*
Sección PANORAMA Y DEBATES

Investigaciones Regionales

SIG y análisis espacial de datos: perspectivas convergentes*

Michael F. Goodchild[1] y Robert P. Haining[2]

RESUMEN: En este artículo se identifican algunos de los desarrollos más importantes experimentados por los SIG y el análisis espacial de datos desde los inicios de los 50. Aunque tanto los SIG como el análisis espacial de datos comenzaron como dos áreas de investigación y aplicación más o menos separadas, han crecido unidos estrechamente a lo largo del tiempo. En el trabajo se mantiene que estas dos disciplinas se unen en el terreno de la Ciencia de la Información Geográfica, proporcionando cada una de ellas apoyo o añadiendo valor a la otra. El artículo comienza proporcionando una visión crítica retrospectiva de los desarrollos que han tenido lugar en los últimos cincuenta años. A continuación, se reflexiona acerca de los desafíos actuales y se especula sobre el futuro. Por último se comenta el potencial de convergencia del desarrollo de los SIG y del análisis espacial de datos bajo la rubrica de la Ciencia de la Información Geográfica (o SIGciencia).

Clasificación JEL: C1, C31, C8.

Palabras clave: SIG, análisis espacial de datos, modelos espaciales, geoestadística, procesos de puntos.

GIS and spatial data analysis: converging perspectives

ABSTRACT: This article identifies some of the important developments in GIS and spatial data analysis since the early 1950s. Although GIS and spatial data analysis started out as two more or less separate areas of research and application, they have grown closer together over time. We argue that the two areas meet in the field of geographic information science, with each supporting and adding value to the other.

---

* Traducción del artículo «GIS and spatial data analysis: Converging perspectives», publicado en Papers in Regional Science 83, n.° 1, 2004, realizada por A. Andrea Caviedes Conde, Master en Administración y Gerencia Pública (INAP) y doctoranda de la Universidad de Alcalá (Madrid).

[1] National Center for Geographic Information and Analysis, and Department of Geography, University of California, Santa Barbara, CA 93106-4060, USA (e-mail: good@geog.ucsb.edu).

[2] Department of Geography, University of Cambridge, Downing Site, Cambridge CB2 3EN, England (e-mail: rph26@cam.ac.uk).

09 Goodchild y Haining 5/7/05 13:14 Página 176

The article starts off providing a critical retrospective of developments over the past 50 years. Subsequently, we reflect on current challenges and speculate about the future. Finally, we comment on the potential for convergence of developments in GIS and spatial data analysis under the rubric of geographic information science (GISscience).

JEL classification: C1, C31, C8.

Key words: GIS, spatial data analysis, spatial modelling, geostatistics, point processes.

1. Introducción

En este trabajo se toma como punto de partida la situación de los Sistemas de Información Geográfica (SIG) y del análisis espacial de datos hace cincuenta años atrás cuando la Ciencia Regional surgía como un novedoso campo de investigación. En los últimos años de la década de los 50 y primeros de la de los 60 los avances en la tecnología informática posibilitaron formas automatizadas de cartografía que, a su debido tiempo, condujeron al desarrollo de los SIG. Aunque pocos hubiesen imaginado en aquellos momentos que el complejo contenido gráfico de los mapas se vería facilitado en su tratamiento por los ordenadores y que otros muchos objetivos pudiesen conseguirse también con ellos- el desarrollo de los escáners y las impresoras en los años sesenta, junto a los rápidos avances en el software, comenzó a abrir posibilidades apasionantes, incluso en la etapa verdaderamente inicial del desarrollo de la informática (Foresman 1998; Maguire *et al*.,1991).

Al mismo tiempo, también aparecieron trabajos pioneros en los campos de las matemáticas y la estadística, que serían fundamentales para el desarrollo del análisis espacial de datos. El trabajo seminal de Whittle (1954) amplió la consideración de los modelos autorregresivos, fundamentales para el análisis de la variación de las series temporales (Kendall 1976) en su aplicación a los datos espaciales. Este grupo de modelos autorregresivos espacialmente fue de los primeros que aparecieron en la literatura estadística para representar formalmente las variaciones espaciales. Al menos para algunas clases de datos, se haría posible ir más allá de la simple comprobación de la existencia de autocorrelación espacial. Las primeros estadísticos (Geary 1954; Krishna Iyer 1949; Moran 1948,1950) probaron la hipótesis nula de no existencia de autocorrelación espacial (esto es de no estructura espacial) frente a hipótesis alternativas no especificas. Con el trabajo de Whittle se posibilitó, al menos para datos regulares de *retícula*, especificar la forma de hipótesis alternativas (o lo que es similar, especificar la representación formal de determinados tipos de estructura espacial), comprobando la significación del modelo asi como la evaluación de la bondad del ajuste.

Más o menos simultáneamente a los anteriores desarrollos que tuvieron lugar en Inglaterra, Matheron en Francia y su escuela para la industria minera desarrollaron el método del «Kriging]», denominado así en homenaje al sudafricano D.G. Krige (Matheron, 1963). Su trabajo extendió la teoría de la predicción de procesos estocásticos de Wiener - Kolmogorov al caso de los procesos espaciales definidos en un es-

09 Goodchild y Haining 5/7/05 13:14 Página 177

pacio geográfico continuo. Su estudio se llevó a cabo para hacer frente a las necesidades prácticas de la industria minera de predecir los rendimientos del sector basándose en muestreos dispersos (oportunistas). Con el paso del tiempo, se desarrolló el campo científico de la geoestadística.

La tercera área que se convertiría en una de las piedras angulares de la estadística espacial fue la teoría de procesos de puntos. En los años 50 se habían desarrollado muchas estadísticas basadas en la distancia y el análisis de cuadrados para examinar la aleatoriedad espacial en los mapas de puntos, aplicándolas a la ecología y al ámbito forestal. El trabajo de Greig-Smith (1952) extendería la contabilidad de cuadrados al examen simultáneo de estructuras en diversas escalas. Sin embargo, por entonces no se había producido un desarrollo equivalente de la teoría de procesos de puntos que correspondiera al que Whittle había llevado a cabo en el área de los datos de *retícula*.

Este artículo examina algunos de los desarrollos clave en estas áreas y la evolución de la Ciencia de la Información Geográfica como un campo unificado que estudia los SIG y los métodos de estadística espacial considerándolos como un apoyo teórico esencial para el análisis espacial de datos. Recordemos a este respecto que el desarrollo de toda ciencia depende de que se reúnan varias condiciones. La primera es que se disponga de datos de buena calidad. La segunda, que existan hipótesis bien formuladas, que puedan formalizarse matemáticamente y someterse a comprobación empírica. En tercer lugar, que exista una metodología rigurosa que permita que al analista obtener inferencias y conclusiones válidas respecto a las cuestiones examinadas partiendo de los datos. Ello incluye la capacidad de formular modelos que puedan utilizarse para probar la hipótesis acerca de los parámetros de interés. La cuarta y última condición es la disponibilidad de una tecnología que permita que en la práctica se lleve a cabo la investigación y con niveles de precisión aceptables.

La evolución de la Ciencia de la Información Geográfica (SIGciencia como la definen Duckham y otros 2003; Goodchild 1992) debe mucho a los desarrollos experimentados en los SIG y en el dominio del análisis de datos espaciales. Las investigaciones en los SIG han mejorando nuestra capacidad técnica para manejar datos de referencia espacial. Además, han estimulado las reflexiones sobre la relación entre lo que puede denominarse en términos muy generales como «realidad geográfica» y la conceptualización y representación de esa realidad en formas digitales finitas, es decir, con cifras que pueden expresarse en forma de puntos, líneas y áreas en el espacio de dos dimensiones.

El análisis de datos espaciales se refiere a aquellas ramas de análisis de datos en los que la referencia geográfica de los objetos contiene información importante. En muchas áreas de la recogida de datos, y en especial en algunas de las ciencias experimentales, los índices que distinguen los diferentes casos pueden intercambiarse sin pérdidas de información. Toda la información pertinente para la comprensión de las variaciones en los conjuntos de datos se contiene en las observaciones mientras que la no relevante se contiene en los índices. En el caso de datos espaciales los índices (de localización y de tiempo) pueden contener informaciones de importancia crucial. Una definición del análisis espacial (del que el análisis de datos espaciales es solo un elemento) es que representa un conjunto de técnicas y modelos que utilizan explícita-

09 Goodchild y Haining 5/7/05 13:14 Página 178

mente la referencia espacial de cada caso de datos. El análisis espacial requiere establecer supuestos o sacar conclusiones sobre los datos que describen las relaciones espaciales o las interacciones espaciales entre casos. Los resultados de cualquier análisis no serán los mismos con una reordenación de la distribución espacial de los valores o bajo una reconfiguración de la estructura espacial (Chorley 1972; Haining 1994).

Los SIG y el análisis espacial de datos entran en contacto, dicho sea en términos coloquiales, en la matriz de datos espaciales. A nivel de concepto, esta matriz esta formada por filas y columnas, referidas las filas a casos y las columnas a atributos medidos para cada uno de los casos. Las últimas columnas proporcionan la referencia espacial. Al nivel más sencillo, pueden existir dos columnas finales que contengan pares de coordenadas: latitud y longitud; aunque algunos sistemas coordinados emplean χ e γ . Hoy, la tecnología de las bases de datos permite una sola columna conceptual que contiene una representación compleja de la geometría o de la forma espacial del caso.

La matriz conceptual sería así parte de un cubo más grande, en el que otro eje es el tiempo. A nivel práctico, la matriz espacial de datos es el depósito de los datos recogidos por el investigador. En términos operativos, la estructura y el contenido de la matriz es el resultado final de los procesos de conceptualización y representación a través de los cuales se recogen algunos segmentos de la realidad geográfica. En cierto sentido es el resultado del proceso de capturar digitalmente el mundo. En otro sentido, la matriz es el punto de salida o el input para el analista de datos espaciales.

Quienes están especialmente preocupados con el análisis de datos necesitan prestar especial consideración al grado de certidumbre con que la matriz de datos refleja la realidad geográfica que subyace en el problema y las implicaciones (para conclusiones interpretativas) de las elecciones de su representación. La cuestión de lo que falta en la representación —la incertidumbre que la representación deja en la mente de los usuarios sobre el mundo que representa (Zhang and Goodchild, 2002)— en algunas aplicaciones puede ser tan importante como el contenido. Por la misma razón, aquellos especialmente preocupados por la representación digital de espacios geográficos necesitan conocer el poder de la metodología estadística para poner de manifiesto la comprensión y la percepción de los datos si se dispone de los mismos en la forma adecuada y se someten a métodos de análisis apropiados.

Este artículo identifica algunos de los desarrollos importantes de los SIG y del análisis espacial de datos desde los primeros 50. Los dos comenzaron como áreas separadas de investigación y aplicación operativa, pero han acabado desarrollándose conjuntamente. Al menos, puede decirse que ambos se encuentran en el dominio de la SIGciencia, apoyándose y añadiéndose valor mutuamente. En el siguiente epígrafe se proporciona una perspectiva histórica de los desarrollos registrados en los últimos cincuenta años. En el epígrafe tercero se reflexiona sobre los retos actuales y se especula acerca del futuro. Por último, en el cuarto epígrafe se comentan las posibilidades de convergencia bajo la rúbrica de la SIGciencia.

09 Goodchild y Haining 5/7/05 13:14 Página 179

2. Una critica retrospectiva

2.1. SIG

2.1.1. Motivaciones iniciales

El desarrollo inicial de los SIG se registró en varias áreas claramente separadas, no siendo hasta los años 80 cuando comenzó a aparecer un aparente consenso sobre el tema. Las motivaciones más importantes para los SIG pueden resumirse de la manera siguiente:

1. Las dificultades prácticas y tediosas para obtener mediciones precisas de los mapas así como la simplicidad para obtener dichas medidas por medio de una representación digital. Concretamente, la relación de una cierta área se convirtió en un tema principal de los informes de inventarios de tierras y condujo al gobierno de Canadá a realizar una gran inversión en el Sistema de Información Geográfica de dicho país (SIG), que comenzó a mediados de 1960 (Tomlinson y otros, 1976).
2. La necesidad de integrar datos acerca de múltiples tipos de características (censos, análisis de zonas de tráfico, calles, hogares, lugares de trabajo, etc.) y las relaciones entre las mismas en proyectos de grandes dimensiones tales como los Estudios sobre Transportes del Área de Chicago en los años sesenta.
3. Los problemas prácticos asociados a la edición de mapas durante el proceso de producción cartográfica, que conducirían a los primeros sistemas de edición de mapas informatizados, nuevamente en los citados años sesenta.
4. La necesidad de integrar múltiples niveles («layers») de información al valorar los impactos ecológicos de los proyectos de desarrollo, lo que condujo a comienzos de los 70 a llevar a cabo esfuerzos por informatizar el método de McHarg de transparencias superpuestas (McHarg, 1992).
5. Los problemas a los que se enfrentó la Oficina del Censo de EE.UU. para gestionar grandes volúmenes de estadísticas censales y repartirlas de formas concretas entre zonas informativas como son las áreas y subáreas censales.

A principios de los setenta, la comunidad investigadora empezaba a ver los beneficios que suponía el software integrado para manejar la información geográfica. Como ocurriría con otros tipos de información, podían obtenerse importantes economías de escala una vez que se hubieran construido los principios para tratar la información geográfica, puesto que pueden añadirse rápidamente a los mismos nuevas funciones y capacidades con mínimos esfuerzos de programación. Los primeros SIG comercialmente viables se presentaron en los primeros años de la década de los 80, cuando la aparición de los microprocesadores hizo posible conseguir suficiente potencia informática dedicada a estas tareas dentro del presupuesto de una empresa o de un departamento del sector público; y ello a medida que el software de las bases de datos relativas al tema evitaba la necesidad de construir funciones sofisticadas para el manejo de datos partiendo de los primeros principios básicos.

09 Goodchild y Haining 5/7/05 13:14 Página 180

Las economías de escala que subyacen en la viabilidad de los SIG comerciales también plantearon otra condición importante: la necesidad de atender los requerimientos de varios tipos de aplicaciones y la de dedicar una mayor atención a segmentos de mercado más grandes. El análisis espacial de datos puede ser la más sofisticada y convincente de las aplicaciones de los SIG pero no es en modo alguno la más significativa comercialmente (Longley y otros 2001). En efecto, las aplicaciones más usuales tienden a responder a necesidades más sencillas de búsquedas y de inventarios. En la gestión de servicios públicos, que se ha convertido en una de las más lucrativas de las aplicaciones, el valor principal de los SIG reside más en su capacidad para rastrear localizaciones y estatus de las instalaciones que en llevar a cabo sofisticados análisis y modelos. Las aplicaciones de los SIG a la investigación científica han tenido siempre que competir con demandas de otros tipos presentadas a la atención de quienes desarrollan los sistemas. Aunque esto pudiera parecer beneficioso en el sentido de que el análisis espacial de datos no ha de sufragar el coste total del desarrollo de los SIG, tiene sin duda también su parte de perjuicio.

2.1.2. Primeros pasos en falso

Desde el punto de vista del análisis espacial de datos, no resulta difícil identificar varias cuestiones y decisiones en el comienzo del diseño de los SIG que hoy, con visión retrospectiva, parecen haber sido desacertadas o como mínimo no productivas. Dentro del espíritu de una revisión crítica, este epígrafe examina algunos de esos primeros pasos en falso.

En primer lugar, los diseños de los SIG miden típicamente las localizaciones de puntos de la superficie terráquea en términos absolutos en relación con la estructura geodésica de la tierra o respecto al Ecuador y al meridiano de Greenwich. Esto tiene sentido si se supone que resulta posible conocer perfectamente la localización absoluta de tales puntos y cuando los instrumentos de medida mejoran la exactitud de la posición las bases de datos pueden también mejorarse. Ahora bien, dos principios fundamentales actúan en sentido contrario a esta posición. En primer lugar, resulta imposible medir con perfección absoluta la localización en la superficie de la tierra y en segundo término las localizaciones relativas pueden medirse con mayor precisión que las absolutas. Debido tanto a los movimientos del eje de la tierra y a las incertidumbres acerca de la forma precisa del globo como a las imprecisiones de los instrumentos de medición no resulta posible medir la localización absoluta sin errores menores a 5 metros. Sin embargo, resulta posible medir de forma mucha más precisa las localizaciones relativas. La distancia entre dos puntos separados por 10 kilómetros puede medirse con precisión de centímetros, utilizando instrumentos estandarizados. Con una visión retrospectiva, habría sido más adecuado construir bases de datos de los SIG registrando localizaciones relativas y haber derivado rápidamente de ellas las localizaciones absolutas cuando hubiese sido necesario (Goodchild, 2002).

En segundo lugar y relacionándolo directamente con el punto anterior la industria de los SIG ha sido muy lenta en la incorporación de métodos que resolvieran las incertidumbres asociadas a todos los aspectos de los SIG. El mundo real es infinitamente complejo. De ello se deduce que es imposible crear una representación per-

09 Goodchild y Haining 5/7/05 13:14 Página 181

fecta del mismo. Zhang y Goodchild (2002) pasan revista a dos decenios de investigaciones sobre la medición, la caracterización, la creación de modelos y la propagación de la incertidumbre. Todavía muy pocas partes de su investigación se han incorporado a productos comerciales, en parte porque ha existido muy poca presión de los usuarios y también porque la incertidumbre representa una especie de talón de Aquiles para los SIG, un problema del que se reconoce plenamente que puede romper la construcción de mapas (Goodchild, 1998).

En tercer lugar, los sistemas de gestión de las relaciones de bases de datos (RDBMS) introducidos en los primeros años 80 proporcionaron una excelente solución para un problema urgente, como es la necesidad de separar el software de los SIG de la gestión de las bases de datos, aunque la solución fuera sólo parcial. En los sistemas híbridos que florecieron entre 1980 y 1995 los atributos de los rasgos variables se almacenaban en los RDBMS, pero la estructura y la localización de tales rasgos se recopilaban en bases de datos separadas, construidas a propósito, que además no permitían emplear el software de gestión normal para las bases de datos. La razón era muy sencilla: las estructuras de los rasgos no podían definirse en la forma tabular que requerían los RDBMS. No fue sino hasta los últimos noventa cuando la adopción generalizada de sistemas de gestión de bases de datos orientadas hacia el objetivo preferente de los RDBMS solucionó finalmente el problema y permitió que todos los datos fuesen tratados por medio de un separado sistema genérico, de gestión de bases de datos.

2.1.3. Dos visiones del mundo

Muchos de los fundamentos de los SIG residen en el trazado de mapas y sin duda la metáfora del mapa aún guía el diseño y la aplicación de los SIG. Ahora bien, muchas de las aplicaciones que interesan a los estudiosos de la Ciencia Regional y a investigadores de otros ámbitos no requieren mapas y en tales casos la metáfora tiende a restringir las ideas. Los SIG funcionan bien cuando se aplican a datos estáticos y menos bien cuando los estudios exigen el análisis de series temporales, de datos detallados dinámicos de transacciones. Funcionan bien con datos bidimensionales y peor cuando resulta importante una tercera dimensión espacial. Se confía en que los mapas proporcionen una cobertura completa de un área geográfica pero las tecnologías de los SIG no han sido adaptadas para tratar aquellos casos en los que faltan cantidades importantes de datos.

El énfasis de los SIG en el conocimiento de las posiciones geográficas absolutas de los casos y de su geometría precisa resulta apropiado para el trazado de mapas, pero no para muchas aplicaciones de las ciencias sociales en las que resulta importante la posición o situación relativa. Muchas teorías acerca del funcionamiento de procesos sociales o económicos en el espacio suponen que no existe variación respecto a la ubicación, el reflejo o la rotación, pero no a las que se registran en las posiciones relativas. De igual forma muchos métodos de la estadística espacial o del análisis espacial no varían y requieren solo el conocimiento de la matriz de interacciones espaciales entre casos, no sus coordenadas exactas. Al respecto, las literaturas tanto sobre interacciones espaciales como sobre redes sociales utilizan el símbolo *W* para

09 Goodchild y Haining 5/7/05 13:14 Página 182

denotar la matriz de interacción. En el caso del espacio, los elementos de *W* pueden colocarse siguiendo una función decreciente de la distancia o de alguna longitud de fronteras comunes entre los mismos y pueden designarse con un 1 si los casos tienen una frontera común y 0 si no es así.

Como expresión de las relaciones espaciales la matriz *W* es compacta y conveniente. Los SIG, con su vasto y más complejo aparato para representar las propiedades espaciales, resultan útiles para calcular los elementos de partiendo de distancias o fronteras comunes y para exponer los resultados del análisis en forma de mapas. Sin embargo, el análisis real puede requerir y encontrar un conjunto de software que sólo necesite reconocer la matriz cuadrada *W* de interacciones y la matriz de datos espaciales normalmente rectangular referida en el epígrafe 1. En los últimos años se ha avanzado mucho por medio del acoplamiento de análisis de software a los SIG.

En esencia, los SIG y la matriz representan dos formas distintas de considerar el mundo: los primeros proporcionan una representación completa y continua de la variación espacial mientras que la segunda suministra una representación mucho más abstracta y discreta. Con esta segunda óptica, las preocupaciones por conseguir representaciones precisas de la posición absoluta que ha caracterizado el desarrollo de los SIG sólo resulta importante en la preparación de datos y en el trazado de mapas de los resultados.

2.1.4. Infraestructura para compartir datos

La explosión de Internet de los últimos años de la década de los noventa produjo un cambio substancial en las perspectivas sobre los SIG y cambió de sentido muchos esfuerzos del desarrollo en este campo. Así, en los Estados Unidos, el Gobierno Federal ha sido tradicionalmente la fuente principal de la información geográfica. De acuerdo con las leyes estadounidenses tales datos tienen la característica de bien público, que puede compartirse libremente sin restricciones de ninguna norma de propiedad intelectual. Desde 1995, la comunidad de los SIG ha realizado una masiva inversión en infraestructura para apoyar que la información geográfica se comparta rápida y fácilmente. Hoy existen petabytes[1] de información geográfica disponible gratuitamente en numerosas páginas web (véase como ejemplo la Alexandria Digital Library (http:www.alexandria.ucsb.edu). Se han desarrollado nuevas normas por parte del Comité Federal de Datos Geográficos de los Estados Unidos (htp://www. tgdc.gov) y del consorcio abierto de los SIG (http://wwwopengis.org) para formatos de datos, la descripción de los mismos (metadata) y su comunicación. Se han desarrollado igualmente nuevas tecnologías que permiten a los usuarios de los SIG interactuar de forma transparente con sitios web remotos por medio de un software comercial eliminando cualquier necesidad de transformar datos como respuesta a las diferencias en formatos, proyecciones o en los datos geodésicos.

Un esfuerzo paralelo se emprendió mucho antes en las Ciencias Sociales, a través de organizaciones como el Archivo de Datos de Essex en el Reino Unido o el Consorcio Interuniversitario de la Universidad de Michigan para la investigación política y

[1] N de la T: petabytes es una medida que supone 10^5 bytes.

09 Goodchild y Haining 5/7/05 13:14 Página 183

social. Se han desarrollado métodos muy sofisticados para describir las matrices de datos, muchos de los cuales incluyen informaciones espaciales. La Iniciativa de Documentación de Datos (DDI) (http://www.icpsr.umich.edu/DDI/) es un esfuerzo internacional para intentar definir una norma aplicable a las Ciencias Sociales respecto a los metadatos. Se están haciendo grandes esfuerzos para integrarla con las normas para los metadatos aplicados a los datos de los SIG, haciendo posible así que los investigadores de ambos tipos puedan llevar a cabo sus trabajos de forma simultánea.

2.1.5. La situación actual de los SIG

La industria de los SIG constituye hoy un sector multimillonario en el que se gastan anualmente miles de millones de dólares en la adquisición y difusión de datos y el desarrollo de software y de aplicaciones. Ha penetrado virtualmente en todas las disciplinas que tratan de alguna forma con la superficie o con las proximidades a la superficie de la tierra, desde la ciencia atmosférica y la oceanográfica a la criminología y la historia. Decenas de miles de personas siguen cursos anualmente sobre los SIG y millones se sirven de los SIG por medio de servicios como el Mapquest (http://www. mapquest.com). También son millones quienes usan los Servicios Globales de Posicionamiento (GPS) y los diseños sencillos de navegación (para automóviles o buques).

El enfoque orientado al objetivo domina ahora la modelización de datos de los SIG. Su primer principio es que todo rango o atributo de la superficie terrestre constituye un ejemplo de una clase y el segundo es que tales clases pueden ser especificaciones de clases más generales. Los autores, por ejemplo, son tipos de la clases «seres humanos masculinos» que constituyen una especialización de la clase más general «seres humanos». A su vez los seres humanos pueden ser considerados como parte de una jerarquía de generalidad creciente: «mamíferos», «vertebrados», «animales» y «organismos», en ese orden. Las clases especificas asumen todas las propiedades de las clases más generales y añaden propiedades diferenciales propias.

Los SIG permiten un amplio abanico de operaciones basadas en su aproximación a la presentación. La mayoría de los métodos conocidos de análisis espacial pueden encontrarse aplicados a productos normales ofertados por los vendedores comerciales de los SIG o en las extensiones de tales productos ofrecidos por terceros. También puede disponerse de una gran variedad de productos y extensiones de los SIG, software libre o *freeware* en comunidades académicas y otras organizaciones similares. La industria de los SIG ha adoptado recientemente un enfoque hacia el software basado en los componentes, descomponiendo lo que eran anteriormente paquetes monolíticos en agregados de componentes susceptibles de volver a ser utilizados. Esto reporta grandes ventajas para la integración de los SIG con otras formas de software que empleen las mismas normas, especialmente los paquetes destinados al análisis estadístico (Ungerer, M.J. y Goodchild, 2002)

En los últimos años se han registrado grandes progresos para apoyar la representación de variaciones en el espacio-tiempo y en las tres dimensiones espaciales. Pero el cambio hacia los modelos de normas industriales orientadas por el objeto ha fallado en un aspecto al abordar diversas formas muy importantes de los datos geográ-

09 Goodchild y Haining 5/7/05 13:14 Página 184

ficos. Mientras que el concepto de objetos discretos resulta apropiado para los seres humanos, los vehículos, los edificios o los objetos manufacturados; el mismo resulta mucho menos compatible con muchos fenómenos del mundo geográfico que son fundamentalmente continuos: ríos, carreteras o terrenos son ejemplo evidentes. En las Ciencias Sociales tiende a pensarse en la densidad de población como un campo con variación continua, expresada matemáticamente en función de la localización y de forma similar muchas otras variables sociales son concebidas mejor como campos (Angel y Hyman, 1976). Los campos pueden representarse en los SIG pero solo atribuyendo su variación a conjuntos de objetos discretos, tales como los puntos de la muestra o extensiones de calles entre secciones administrativas adyacentes o zonas a que se refieren los informes. En este último caso, las consecuencias de atribuir variaciones continuas a espacios arbitrarios están bien estudiadas como el problema de una unidad modificable de áreas (Openshaw, 1983). Pero en tanto que la orientación hacia el objeto continúe siendo la vía dominante de pensamiento sobre datos en la industria informática y en los SIG, los científicos sociales tendrán que emplear los mecanismos creados para representar variaciones continuas de la manera indicada.

2.2. Progresos en el análisis de los datos espaciales

La revisión que se hace en este epígrafe del progreso en el análisis de datos espaciales desde los primeros cincuenta se ha dividido en dos partes. En la primera se pasa revista a los desarrollos de la teoría estadística en el análisis de datos espaciales. Después sigue una revisión de las nuevas técnicas y métodos. Un aspecto del proceso realizado ha consistido en la aparición de nuevos campos de aplicación que estudiamos en el epígrafe 3 en el contexto del futuro desarrollo del análisis de datos espaciales.

2.2.1. Desarrollo de la teoría

Una visión general de la teoría aplicable al análisis de datos en los últimos sesenta y en los primeros setenta muestra que los desarrollos hasta esos momentos habían sido claramente desiguales. No se encuentran textos que consideren conjuntamente los avances en ese campo y que traten de presentarlos dentro de una estructura teórica más amplia. No ha existido un campo de la estadística espacial como tal, aunque el término análisis espacial haya penetrado en la literatura geográfica (Berry y Marble, 1968). Lo más cercano a un texto de carácter general fue el trabajo realizado en los sesenta por Matern en su monografía «Spatial Variation» (Matern, 1986). Esta situación subsistió en agudo contraste con el análisis de series temporales donde sí aparecieron varios textos teóricos de importancia (Anderson, 1971; Grenander y Rosenblatt, 1957; Kendall,1976; Wold, 1954), igualmente como los que surgieron en el terreno de textos econométricos que trataban del análisis de series temporales económicas (Johnston, 1963; Malinvaud, 1970). Si bien en muchos textos y monografías, se mencionaba el análisis de datos espaciales, su cobertura resultaba generalmente escasa, con la excepción de la monografía de Bartlett de 1955 «Stochastic Processes».

09 Goodchild y Haining 5/7/05 13:14 Página 185

Fue ese primer trabajo de Bartlett el que proporcionó la base para presentar una nueva clase de modelos espaciales para datos de área, que fueron desarrollados en el trabajo seminal de Besag aparecido en 1974. Se desarrollaron modelos que satisfaccieran una versión bidimensional de la propiedad de Markov, propiedad que subyace en muchos de los modelos más importantes que representan variaciones temporales. Como consecuencia del trabajo de Besag se hizo posible especificar una clase general de modelos espaciales para variables discretas y continuas definidas en el espacio bidimensional.

La diferencia entre los tipos de modelos de Whittle y de Basag, puede ilustrarse utilizando el caso de observaciones [$z(1)$, ..., $z(T)$] obtenidas de una distribución normal y generalizando la discusión al caso de un sistema irregular de áreas (por ejemplo, los distritos censales de una ciudad). La propiedad espacial de Markov que subyace en los modelos de Besag generaliza la propiedad de primer orden de Markov en la modelización de series temporales. La propiedad de Markov en el momento (t) establece que únicamente el pasado más reciente determina la probabilidad condicional del presente o sea:

$$prob\{Z(t) = z(t) / Z(l) = z(l), ..., Z(t-1) = z(t-1)\} = \\ = prob\{Z(t) = z(t) / Z(t-1) = z(t-1)\} \qquad [2.1]$$

En la que *prob* { } indica la probabilidad.

El tiempo tiene un orden natural. Una expresión análoga a la de [2.1] para datos espaciales se basa en la definición de una estructura de gráficos en el conjunto de unidades espaciales. Para cada área (o lugares) se define un conjunto de áreas vecinas, como pueden ser, por ejemplo, todas las áreas que comparten una frontera común o cuyos centroides se encuentran dentro de una determinada distancia del centroide del área. La forma análoga a la recogida en (2.1) sería para este caso:

$$prob\{Z(i) = z(i) / Z(j) = z(j)\},\ j \in D, j \neq i\} = \text{Prob}\ \{Z(i) \\ = z(i) / Z(j) = z(j)\}, j \in N(i)\} \qquad [2.2]$$

en la que D representa el conjunto de todas las áreas (o sitios) en el estudio de la región y $N(i)$ indica el conjunto de áreas vecinas al sitio i.

En el análisis espacial este enfoque se ha denominado como la aproximación condicional para especificar modelos de campo aleatorios de datos regionales. La ecuación [2.2] y las generalizaciones de orden superior proporcionan la base natural para construir modelos espaciales que satisfacen una forma espacial de la propiedad de Markov (Haining, 2003).

Un modelo normal espacial multivariante que satisface la expresada propiedad de primer orden de Markov puede definirse como sigue (Besag, 1974; Cressie, 1991; pág. 407):

$$\mathrm{E}[Z(i) = z(i) / \{Z(j) = z(j)\}\ j \in N(i)\}] = \mu(i) + \sum_{j \in N(i)} \kappa(i,j)[Z(j) - \mu(j)]$$

y

$$Var[Z(i) = z(i) / \{Z(j) = z(j)\}\ j \in N(i)\}] = \sigma(i)^2 \qquad [2.3]$$

09 Goodchild y Haining 5/7/05 13:14 Página 186

$i = 1, ..., n$
donde $\kappa(i,i) = 0$ y $\kappa(i,j) = 0$ a menos que $j \in N(i)$. A ello se le denomina el modelo autonormal o el modelo autoregresivo condicional (modelo CAR —Conditional AutoRegressive—). Las propiedades no condicionales del modelo expresadas en notación matricial serían:

$$E[Z] = \mu \quad \text{y} \quad Cov[(Z - \mu), (Z - \mu)^T] = (I - K)^{-1} M \qquad [2.4]$$

En la que E[ ] representa el valor esperado (esperanza) y Cov[ ] la covarianza. κ es una matriz $n \times n\{\kappa(i,j)\}$ y $\mu = \{\mu(1), ... \mu(n)^T\}k$ es especificada exógenamente M es una matriz diagonal en la que $m(i,i) = \sigma(i^2)$ puesto que $(1 - k)^{-1} M$ es una matriz de la covarianza debe ser simétrica y definida positiva. Se deduce de ello que $\kappa(i,j)\sigma(j)^2 = \kappa(j,i)\sigma(i)^2$. Nótese que si las varianzas condicionales son todas iguales entonces $\kappa(i,j)$ debe ser lo mismo que $\kappa(j,i)$.

Si el analista de datos regionales no concede importancia al cumplimiento de la propiedad de Markov (Haining, 2003) puede disponer de otra opción, la llamada aproximación simultanea a la especificación de un modelo de campo aleatorio. Este es el enfoque que Whittle desarrolló en un trabajo de 1954 y puede ser fácilmente generalizado a unidades de área irregular. Sea una variable independiente normal $IN(O,\sigma^2 I)$ y $e(i)$ la variable asociada con i (= 1, ... n). Definimos la expresión:

$$Z(i) = \mu(i) + \sum_{j \in N(i)} s(i,j)[Z(j) - \mu(j)] + e(i) \qquad i = 1, ... n \qquad [2.5]$$

en la que $s(i,i) = 0$. Aunque no sea una condición que $s(i,j) = s(j,i)$, en la práctica tal supuesto se adopta a menudo (Cressie, 1991; pp. 410). S representa una matriz $n \times n$ especificada exógenamente $\{s(i,j)\}$. La misma esta de nuevo basada en la especificación de las vecindades de cada lugar. Sea de nuevo $\mu = [\mu(1), ... \mu(n)]^T$. En términos matriciales la expresión 2.5 se convierte en:

$$\text{E}[Z] = \mu \text{ y } Cov[(Z - \mu), (Z - \mu)^T] = \sigma^2 (I - S)^{-1} (I - S^T)^{-1} \qquad [2.6]$$

Los años 70 y los inicios de los 80 fueron unos años que conocieron avances significativos en la que empezaba a ser conocida como estadística espacial. Cressie (1991) señaló que en el caso del análisis espacial de modelos de punto «un momento de cambio importantísimo para el desarrollo moderno en este campo fue el artículo de Ripley (1977)». Este autor propuso el empleo de la función K, un instrumento descriptivo y de modelización para los datos de proceso de punto que había sido sugerido originalmente por Bartlett. La ventaja de la función K es que proporciona una descripción de un modelo de punto (aleatorio, regular o de cluster) para una serie de escalas. La función K puede utilizarse también para ajustar modelos espaciales adecuadamente especificados a datos de punto.

Sea $N(A)$ la cantidad de puntos en un área A que proporciona la base natural para analizar procesos de punto. La covarianza entre $N(A)$ y $N(B)$ para dos áreas (A y B) puede ser reducida a una función creciente no negativa tal como:

09 Goodchild y Haining 5/7/05 13:14 Página 187

$$K(d) = \lambda^{-1}E$$
[Número de puntos adicionales ≤ distancia d un punto escogido al azar] [2.7]

En la anterior expresión [2.7] λ es la densidad de puntos (número de puntos por unidad de área) y E{ } representa el valor esperado de la expresión entre corchetes.

Ignorando los efectos de fronteras (límites) introducidos por las vecindades de *A*, la estimación de la función *K* [2.7] para cualquier distancia es:

$$1^{-1}\left[(\lambda|A|)^{-1}\sum_{I}\sum_{m(\neq I)} I_d(d_{I,m})\right] \qquad [2.8]$$

donde $|A|$ es el área de la región a estudio; $I_d(d_{I,m})$ es la función indicada que toma el valor de 1 si la distancia entre los puntos *I* y *m* es menor o igual a *d*, 0 en caso contrario. La expresión entre corchetes de [2.8] es la cifra media de casos adicionales dentro de la distancia *d* de un caso observado, en el que la media se computa para todos los casos. El estimador del parámetro de densidad λ es el número de casos *n* dividido por el área de la región de estudio $|A|$. Resulta preciso introducir una función ponderada en la función indicador para corregir los efectos de frontera (Gatrell y otros, 1996; pp. 262-263). El significado de los picos y simas de la impresión gráfica de $\{K(d),d\}$ puede establecerse por medio de la simulación de Montecarlo.

Los primeros años 80 vieron la aparición de una monografía que reveló la forma del campo que estaba surgiendo. En el libro de Ripley (1981) «Spacial Statistics» se proporcionaba una visión general de este dominio, incluyendo los avances en Geoestadística. El tratamiento de la Geoestadística se incorporaba en un capítulo del mismo sobre el ajuste espacial y la interpolación espacial. La importancia de esta área y el significado más amplio del trabajo de Matheron, más allá de los intereses de geólogos e ingenieros de minas, se hacía más evidente a medida que empezaba a expandirse el terreno del trazado automático de mapas. Otra monografía significativa de aquella época es el trabajo de Diggle (1983) sobre el análisis de mapas de puntos. Este relacionaba la función que se utiliza para describir estructuras de modelos de punto con modelos, que representan variaciones de los mismos permitiendo su utilización para comprobar hipótesis espaciales especificas. La monografía de Stoyan y otros (1987) trataba de procesos de puntos y geometría estocástica.

En 1991 la publicación del volumen de Cressie, de 900 páginas, «Statistics for Spacial Data» anunciaba la última ocasión en que los estadísticos proporcionaban una visión general de todo este campo de trabajo. Su texto incluye métodos y modelos para el análisis de datos geoestadísticos, datos de retícula (incluyendo los modelos de Whittle y Besag), así como datos de modelos de punto y de objetos.

2.2.2. Técnicas

Como se indicó en el epígrafe 1, entre las primeras técnicas del análisis espacial figuraron las desarrolladas para evaluar la autocorrelación espacial de retículas regulares. Para la ciencia regional que trata frecuentemente datos de unidades de área irregular esto significaba que las evaluaciones eran de valor limitado. Cliff y Ord (1973,

09 Goodchild y Haining 5/7/05 13:14 Página 188

1981) fueron los primeros en desarrollar criterios de evaluación para la autocorrelación espacial en unidades de área irregular, generalizando los tests anteriores y examinando con detalle la teoría de la inferencia asociada a los mismos. En muchas áreas no experimentales (observables) de la ciencia los datos tienen mayores niveles de incertidumbre asociados a su medición que en el caso de las ciencias experimentales. Cressie (1984) introdujo técnicas para un análisis espacial consistente en Geoestadistíca, esto es técnicas más robustas (que los métodos tradicionales) en el tratamiento de los errores en los datos y atípicos *(outliers).* También desarrolló métodos de posible utilización para destacar posibles errores en los conjuntos de datos espaciales. Haining (1990) amplió varios de los métodos de Cressie a los datos de área.

Muchos de los primeros métodos de análisis de los mapas de puntos eran limitados respecto al alcance de sus aplicaciones, principalmente a causa del espacio subyacente a través del cual podrían tener lugar homogeneamente los sucesos. Muchos de los primeros métodos se desarrollaron en ecología para estudiar distribuciones de plantas en su área de estudio. Pero para estudiar por ejemplo clusters de casos de enfermedad o estructuras de delitos era necesario acudir a supuestos de falta de homogeneidad en los colectivos que presentaban riesgos. Diggle y Chetwynd (1991) emplearon funciones K para desarrollar un test para la agrupación de casos en una, estructura de puntos no homogéneos y así poder comprobar la hipótesis de «etiquetado aleatorio» de casos de enfermedad en una población distribuida en el espacio.

Muchas de las técnicas del análisis espacial han subrayado las propiedades medias presentes en los datos espaciales. Los tests para agrupación de casos o la autocorrelación espacial anteriormente descritos proporcionan pruebas respecto a las que se han denominado propiedades de «mapa entero». En 1990 existió un interés especial en desarrollar tests localizados que reconociesen la heterogeneidad de las propiedades de los mapas a través del área de estudio. Además de los tests para indicar clusters (una propiedad de los mapas enteros) que pueden probar la hipótesis de si los casos de una enfermedad tienden a presentarse conjuntamente, se precisan también tests para detectar la presencia de clusters (localizados). Tales tests pueden proporcionar evidencias de donde pueden encontrarse concentraciones locales, lo que puede usarse para dirigir esfuerzos sobre el alcance en los que las condiciones locales pueden ser las responsables de los casos planteados.

Los tests para detectar clusters (Besag y Newell, 1991; Openshaw y otros, 1987; Kulldorff y Nagarwalla, 1995) plantean problemas como el sesgo de la preselección y los efectos de pruebas múltiples, que necesitan tratarse para construir una teoría inferencial fiable. Otros procedimientos localizados de tests incluyen los indicadores locales de Anselin (1995) de asociación espacial, o LISAs (Local Indicator Spatial Association) que comprueban estructuras localizadas de autocorrelación espacial. Tales métodos numéricos han sido complementados con el desarrollo de métodos de visualización para datos espaciales (Haining, 2003). La interacciones de métodos científicos de visualización con avances en cartografía automatizada y los SIG han sido una de las áreas más dinámicas del análisis espacial en los 90 (McEachren y Monmonier, 1992). El desafío ha sido dejar de considerar los mapas como el producto final de un programa de investigación y que fuesen desde entonces empleados sólo para exponer resultados. Esto requiere la integración del tra-

09 Goodchild y Haining 5/7/05 13:14 Página 189

zado automático de mapas dentro de la agenda más amplia de los análisis de exploración de datos, sugiriendo hipótesis y también líneas para investigaciones adicionales. El trazado de mapas es un elemento esencial en la visualización científica de los datos espaciales.

3. Desafíos actuales y direcciones del futuro

3.1. Respecto a los SIG

3.1.1. Representaciones y fuentes de datos

A pesar de las limitaciones y de los resbalones a que se ha hecho referencia en el epígrafe 2.1 los SIG representan hoy una base sólida para el análisis espacial de datos. Los productos básicos de venta comercial proporcionan un conjunto de técnicas para el análisis y la visualización, además de las funciones esenciales de orden interno de transformación, proyección de cambios y de nuevos muestreos. Con frecuencia se añaden técnicas adicionales de análisis por medio de extensiones en lenguajes escritos y crecientemente en lenguajes tales como el Visual Basic for Application, desarrollados en muchas ocasiones por terceros no interesados directamente. Diversas estrategias de acoplamiento también se han utilizado para relacionar los SIG con códigos más especializados para aplicaciones específicas.

Pasado el tiempo, los investigadores sobre los SIG han llegado a situar su mayor esfuerzo en las cuestiones de representación, esto es, en la tarea fundamental de representar el mundo geográfico real dentro del alfabeto binario de la informática digital. La investigación se ha centrado en la representación del tiempo, la cual tiende a ser mucho más compleja que la simple agregación de una tercera (o cuarta) dimensión a las dos (o tres) de las representaciones espaciales. Peuquet (2002) proporciona un análisis de los fundamentos filosóficos y conceptuales del problema y Langran (1992) presenta una perspectiva más práctica mientras que Frank y otros (2001) analizan los temas relativos a la información sobre las zonas fronterizas que cambian en el tiempo. Otras investigaciones se han dirigido a la representación de objetos cuyos límites son inciertos (Burrough y Frank, 1996), un tema común en datos ambientales aunque quizá lo sea menos en datos sociales.

No obstante, los problemas sobre el tiempo son claramente más que una simple cuestión de representación. Todo el aparato de la recogida de datos censales se ha aproximado al tiempo a través de fotografías decenales, cambiando los límites de las zonas fronterizas de información en cada ciclo y creando problemas complejos para quienes desean analizar los datos censales en series longitudinales. No sólo hay poco disponible en la línea de los datos espacio-temporales sino que nuestro aparato analítico y estructuras teóricas son igualmente limitados. La llegada de mayores cantidades de datos de rastreo (tracking) en los años recientes, por medio del empleo del sistema GPS conectando a muestras de vehículos e individuos es posible que estimule un interés renovado por esta área descuidada de los SIG y del análisis espacial de datos.

09 Goodchild y Haining 5/7/05 13:14 Página 190

Se va disponiendo además de nuevas fuentes de datos por medio del despliegue de sistemas de imágenes de alta resolución proporcionadas por satélites. Resulta ahora posible obtener imágenes a todo color de la superficie de la tierra con resoluciones más ajustadas que 1 metro, lo que abre una serie de nuevas posibilidades para el análisis de la morfología urbana y de las formas construidas (Liverman y otros, 1998). Pueden utilizarse instrumentos basados en la red web para investigar datos y también para detectar y analizar referencias a la localización en los textos, creando otra nueva fuente de datos de interés potencial para las ciencias sociales. La tecnología también está creando nuevos tipos de comportamiento humano espacial en tanto las personas adquieren conciencia de las capacidades de los teléfonos móviles de tercera generación para determinar las localizaciones exactas y para modificar las informaciones que suministran de acuerdo con esto (Economist, 2003).

3.1.2. La agenda de investigación dela SIGciencia

Se han desplegado numerosos esfuerzos para identificar direcciones futuras en los SIG y para delinear la correspondiente agenda de investigación para la SIGciencia. Hasta bien entrados los años 90, la idea prevaleciente respecto a los SIG era considerarlos como una ayuda a la investigación, que conducía a tareas que el usuario encontraba demasiado tediosas, complejas o tan consumidoras de tiempo como hacerlo manualmente. A medida que avanzaba el tiempo, se suponía que los SIG se iban convirtiendo en un instrumento poderoso que ponía en funcionamiento proporciones cada vez mayores de las técnicas conocidas de análisis espacial de datos.

La llegada de la web y la popularización de Internet condujeron a una reconsideración importante de la visión anterior y amplió considerablemente la agenda de investigación al mismo tiempo. En vez de ser un asistente personal, los SIG en este nuevo contexto, se consideraban como una especie de medium, una vía de comunicación de lo que se conoce acerca de la superficie de la tierra. Esto es compatible con la discusión anterior en el epígrafe 2.1.4 sobre datos compartidos, pero va mucho más lejos y ofrece una visión muy diferente acerca del futuro de los SIG. De acuerdo con esta opinión, el análisis de datos espaciales viene proporcionado por un sistema de servicios, algunos locales y otros situados en lugares remotos, algunos libres y otros de pago. Los datos pueden ser igualmente locales o pueden provenir de algún servidor remoto. La localización adquiere así diversos significados: la localización del tema sujeto al análisis, la localización del usuario, la localización del almacenaje de los datos o la localización que proporciona el servicio de análisis.

Quizá el esfuerzo más extenso y continuado para definir la agenda de investigación de la SIGciencia y al mismo tiempo el futuro de los SIG haya sido el organizado por el Consorcio Universitario para la Ciencia de la Información Geográfica (UCSIG) en su agenda inicial de diez puntos (UCGIS, 1996). En ella se incluyeron una serie de temas fundamentales de investigación como son los avances de datos espaciales con SIG, ampliaciones a las representaciones con SIG e incertidumbre. La lista ha sido ampliada posteriormente y se puede encontrar en el sitio Web de la UCSIG

09 Goodchild y Haining 5/7/05 13:14 Página 191

(http://www.ucgis.org) Se ha agregado a la lista la visualización como lo ha hecho la ontología, interpretada en este caso como el estudio de lo más abstracto.

3.1.3. Continuando con los temas metodológicos

El crecimiento de los SIG ha conducido a una popularización masiva de los métodos espaciales y a una apreciación mucho más amplia del valor de las más antiguas disciplinas de la Geografía y la Ciencia Regional. El público corriente ahora encuentra soluciones a los problemas del camino más corto entre puntos o las direcciones para la conducción proporcionadas por sitios como el MapQuest y traza rutinariamente mapas de áreas locales a través de sitios que implementan un software para el trazado de mapas con base a la red Web. En muchas formas los SIG constituyen vías de encuentro del análisis de datos espaciales con las preocupaciones de la sociedad y en que los resultados de la investigación científica se convierten en decisiones y en políticas (Craig y otros, 2002). Poderosos instrumentos de los SIG se hallan ahora en menos de cientos de miles de usuarios. Muchas de estas personas no han tenido mayor exposición al análisis teórico que el que se encuentra en el análisis de datos espaciales. Como resultado, es fácil encontrar ejemplos de utilizaciones e interpretaciones desacertadas y casos de conflictos entre el empleo de los SIG y los principios de la ciencia.

Ya anteriormente se hizo mención a la cuestión de la incertidumbre. Resulta sencillo para los diseñadores de sistemas expresar los datos numéricos con precisión redoblada llegando a las catorce cifras decimales, pudiendo expresarse los resultados de los SIG con la misma precisión. Pero en la realidad apenas existen casos en que se requieren las catorce cifras decimales para lograr una precisión real de datos o cálculos. El respetado principio, de que los resultados deben presentarse con una precisión numérica determinada por su exactitud es fácil de ignorar y el hecho de que los resultados aparezcan en un ordenador digital da a menudo una falsa sensación de autoridad.

Otro respetado principio se refiere a la replicabilidad[2] y exige que los resultados se presenten con el suficiente detalle que permita a terceros replicarlos. Este principio entra todavía en conflicto con la complejidad de muchos análisis SIG, la naturaleza propia de buena parte del software de los SIG, y el aburrimiento de escribir una documentación extensa y rigurosa.

En estos y otros aspectos los SIG se encuentran en un extremo de la ciencia, en un área gris compartida entre un pensamiento científico que quiere ser preciso y objetivo y el mundo subjetivo del discurso humano. Ello origina beneficios potenciales por su capacidad de enlazar con el público general y con quienes han de tomar decisiones, pero al mismo tiempo plantea riesgos de interpretaciones y de falsos usos. En algunos terrenos, como el de la supervisión la respuesta histórica ha sido la profesionalización – dar licencias a los practicantes y otras formas de acceso restrictivo– pero hasta el momento los esfuerzos por profesionalizar la práctica de los SIG no han tenido éxito, en parte porque el uso de los SIG en la investigación científica exige apertura y no restricciones.

---

[2] N de la T: falsabilidad en la denominación de Popper.

09 Goodchild y Haining 5/7/05 13:14 Página 192

3.2. En el análisis de datos espaciales

Se proporciona a continuación una visión general de algunas de las áreas actuales de interés para la investigación en la metodología del análisis de datos espaciales. Se continúa con una discusión acerca de las nuevas aplicaciones basadas en la modelización espacial y en sus aplicaciones para el desarrollo del análisis de datos espaciales.

3.2.1. Metodología y desarrollo tecnológico

Algunos modelos de regresión espacial han dominado hasta el momento la literatura sobre ciencia regional: el modelo de regresión con errores espacialmente correlacionados, el modelo de regresión con variables de respuesta aleatoria espacialmente ponderadas (o retardadas). Cada modelo representa alguna forma de desviación respecto al modelo de regresión usual y tradicional en el que las relaciones entre una respuesta y sus covariantes se definen verticalmente, entendiendo por vertical que la respuesta en el área es sólo una función del nivel de predictores en . No existen, sin embargo razones para presumir que exista esta estructura vertical de las relaciones en el caso de los datos espaciales. Esto es así porque los procesos subyacentes no reconocen ordinariamente las unidades espaciales artificiales (los distritos censales) en cuyos términos se han recogido las observaciones, distinto, por ejemplo, a una serie de réplicas independientes y separadas de una situación experimental. La continuidad inherente al espacio geográfico en relación con la estructura espacial empleada para capturar las variaciones y las formas en que los procesos se llevan a cabo en el espacio geográfico ha conducido a un mayor interés hacia las versiones «espacializadas» del modelo de regresión tradicional, así como hacia otras variantes (Haining, 2003).

Para ilustrar algunos desarrollos recientes se considera el siguiente ejemplo. El modelo de regresión con una variable respuesta espacialmente retardada ha sido empleado para modelizar resultados bajo el supuesto de que los niveles de la variable respuesta en un área pueden ser función de los niveles de la variable respuesta en áreas vecinas (este modelo ha sido utilizado para analizar efectos de competencia y algunos tipos de procesos de difusión e interacción). Modelos de este tipo se ilustran en [2.3] y en [2.5]. Sin embargo, esta aproximación plantea problemas para algunas clases de datos, especialmente para algunos de los de recuento en los que las restricciones de los parámetros impiden la aplicación de modelos a situaciones reales (Besag, 1974). Pero existe además un segundo problema. Los parámetros de interés (la media en el caso de un proceso de Gauss; la probabilidad de un suceso en el caso de un proceso logístico o binomial; el parámetro de intensidad en el caso de un proceso de Poisson) se supone que son desconocidos pero constantes. El problema es obtener la mejor estimación del parámetro y calcular intervalos de confianza en torno a él. (En la práctica no es la media o la intensidad del parámetro lo que naturalmente interesa sino los parámetros de la función de regresión, cada una de ellos asociado a una covarianza). ¿Por qué deberían tratarse tales parámetros como si fuesen valores fijos a estimar en lugar de cómo cantidades aleatorias?

En la modelización de las enfermedades, el parámetro de interés para cualquier conjunto de áreas se asocia generalmente con los riesgos relativos subyacentes de la

09 Goodchild y Haining 5/7/05 13:14 Página 193

incidencia, la prevalencia o la mortalidad de alguna enfermedad especifica. En los modelos de datos delictivos el parámetro de interés es el área especifica subyacente de riesgo relativo de ser victima de algún delito especifico (robo, por ejemplo).

Supongamos que $Z(i) = O(i)$ representa el número de muertes observadas en el área *i* durante un período determinado de tiempo provocadas por una enfermedad rara pero no infecciosa. $O(i)$ se obtiene de una distribución de Poisson con un parámetro de intensidad $\lambda(i) = E(i)r(i)$ indica el número esperado de fallecimientos debido a la enfermedad en el área *i*, dadas la edad y la distribución de sexos en el área y $r(i)$ es el riesgo relativo de morir en el área especifica como consecuencia de la enfermedad en el área *i*, siendo el parámetro de interés.

Una aproximación para estimar $\{r(i)\}$ es suponer que se obtiene de un distribución de probabilidades denominada en la terminología bayesiana la distribución preferente. A menudo se consideran dos tipos de efectos aleatorios: uno en el que los efectos aleatorios de $\{r(i)\}$ no están estructurados espacialmente y otro en el que sí lo están (Mollié, 1996).

Consideramos el caso en que el modelo se incluyen efectos aleatorios espacialmente estructurados y no estructurados. Suponemos la expresión:

$$\log [r(i)] = \mu + v(i) + e(i) \qquad [3.1]$$

en la que $v(i)$ es la variación aleatoria espacialmente estructurada y $e(i)$ es la no estructurada. La $\{e(i)\}$ es un proceso gausiano de ruido blanco. La $\{v(i)\}$ y por tanto la $\{\log[r(i)]\}$se supone que pueden obtenerse de un campo aleatorio preferente gaussiano de Markov como en [2.3]. Para este tipo de modelos «la distribución condicional del riesgo relativo en el área *i,* dados valores de riesgo relativo en todas las otras áreas $j \neq i$, depende de los valores de riesgo relativo en las áreas vecinas $N(i)$ del área *i*. De esta forma, en este modelo los riesgos relativos tienen una estructura localmente dependiente de probabilidades preferentes (Mollié, 1996; pp. 365). Nótese que la dependencia espacial está especificada para los parámetros de interés $\{\log[r(i)]\}$ y no en los datos observados $\{O(i)\}$, aunque estos últimos heredaran la dependencia espacial debido a su dependencia de $\{\log[r(i)]\}$.

Un modelo para una variación espacialmente estructurada de tipo puro viene proporcionada por la autoregresión gaussiana intrínseca, una forma limitada de un modelo condicional autoregresivo como el [2.3], con un solo parámetro de dispersión κ^{-2} (Besag y otros, 1991). En este modelo si se supone una conectividad binaria o una matriz de interacción (W) entonces ($w(i,j) = 1$ si *i* y *j* son vecinos o 0 en otro caso). Entonces:

$$E[v(i)|\{(v(i)\}\, j \in N(i),\, k^{-2}] = \sum_{j=1,\ldots n} w^*(i,j)\, v\,(i) \qquad [3.2]$$

$$\text{var}[\text{v(i)}|\{\text{v(j)}\}\, j \in N(i),\, k^2] = k^2 / \sum_{j=1,\ldots n} w(i,j) \qquad [3.3]$$

en la que W^* representa la forma normalizada en filas de W. De esta forma el valor condicional esperado del logaritmo del riesgo relativo del área *i* es la media del loga-

09 Goodchild y Haining 5/7/05 13:14 Página 194

ritmo de los riesgos relativos de las áreas vecinas. La varianza condicional es inversamente proporcional al número de áreas vecinas. Si escribimos además:

$$\log [r(i)] = \mu(i) + v(i) + e(i) \qquad [3.4]$$

$$\mu(i) = \beta_0 + \beta_1\chi_1(i) + \beta_2\chi_2(i) + ... + \beta_k\chi_k(i) \qquad [3.5]$$

entonces la variación espacial en el riesgo relativo (expresado logaritmicamente) se modela también como una función de los predictores $(\chi_1, \chi_2, ... \chi_k)$. También pueden utilizarse otros modelos espaciales para representar los efectos aleatorios espacialmente estructurados en la distribución de los $\{r(i)\}$.

La aproximación que se usa en el ajuste y en la inferencia de estos modelos es la simulación de Montecarlo de las cadenas de Markov (MCMC). La MCMC es un tipo de algoritmo de simulación para la obtención de muestras desde unas distribuciones posteriores requeridas de parámetros modelo. Se necesitan muestras grandes para su distribución posterior y poder obtener luego propiedades tales como la media o la moda o los cuantiles mediante la integración de Montecarlo en las distribuciones marginales de los parámetros de interés. La MCMC que utiliza el caso especial de la muestra de Gibbs se usa en el software WinBUGS. Casella y George (1992) proporcionan una introducción a la muestra de Gibbs, y Gelman y otros (1995, cap. 11) comparan los diferentes algoritmos.

Es posible que en el futuro se disponga de un conjunto más amplio y más rico de modelos estadísticos por medio de los cuales se representen las variaciones espaciales. Esto será muy importante. El propósito de la modelización estadística es facilitar a los investigadores la evaluación de sus hipótesis introduciendo las mismas en representaciones válidas de las propiedades estadísticas de los datos. Todos los modelos, en razón de las simplificaciones que introducen, son erróneos, pero algunos modelos son útiles. Modelos pobres no proporcionan tests válidos de las hipótesis y no son, por tanto, útiles. Hasta el momento, quienes quieren construir modelos espaciales sólo han dispuesto de un conjunto muy limitado de modelos para trabajar con ellos. En la medida en que los tipos de modelos espaciales sean más numerosos, los analistas podrán evaluar hipótesis por medio de modelos que representen mejor las variaciones subyacentes en los datos.

Consideremos las dos áreas de desarrollo tecnológico que tienen consecuencias para la futura evolución del análisis espacial de datos. En primer término, en la medida en que la resolución de los datos espaciales aumente como resultado de mejoras en la recogida y almacenamiento de los datos espaciales, se estarán creando oportunidades para efectuar agregaciones espaciales más flexibles y para crear estructuras espaciales que sean de mayor relevancia para los problemas sometidos a estudio. Pero una mayor precisión espacial no significa automáticamente una mayor precisión estadística y en consecuencia tampoco mejores representaciones de algunas formas de variación espacial por medio de mapas (tales como las tasas de riesgo relativo, las tasas de desempleo, las tasas de delito o los parámetros medioambientales). Esto se deriva de los efectos del problema de los pequeños-números (cuando se registran datos

09 Goodchild y Haining 5/7/05 13:14 Página 195

a escalas de muchos detalle espacial) y de sus implicaciones sobre el trazado de mapas, la descripción de las variaciones espaciales y la inferencia subyacente.

La metodología anteriormente descrita es una de las formas que se están empleando de forma creciente para proporcionar mejores mapas basados en datos de área pequeña y posibilitados por medio de avances en la recogida de datos, sean estos obtenidos por muestreo o por realización de censos. La especificación de la distribución preferente (en $\{r(i)\}$ por ejemplo) determina como es preciso tomar prestada la información de otras fuentes con el propósito de fortalecer las estimaciones de área pequeña con parámetros de interés. Cuanto más depurados los datos más importante resulta encontrar vías para tratar con niveles más elevados de variaciones estocásticas en los datos. Esta es una de las formas en que los avances originados por una recogida de datos mejor y más depurada y a través de los SIG como un *medium* para el almacenamiento de datos plantea retos a la realización del análisis estadístico.

En segundo lugar, la visión de la e-ciencia (la ciencia electrónica), que incluye a las Ciencias Sociales y Medioambientales, es un tema de una amplia comunidad de investigadores y de su necesidad de recursos, que interactúan a través de co-laboratorios virtuales. Una consecuencia de esto es la necesidad de una tecnología de la información para apoyarla, en la que los usuarios y servicios de análisis estén espacialmente distribuidos (véase al respecto el epígrafe 3.1.2). Ello incluiría también accesos compartidos a mayores recursos informativos, a mejores archivos de datos, a técnicas de visualización de datos, y al acceso desde lugares remotos a estos instrumentos especializados. La implicación del análisis espacial no es tan sólo para el desarrollo de adecuados instrumentos numéricos y de visualización que puedan manejar series de datos muy grandes, sino también para sistemas inteligentes de apoyo que ayuden al analista en todos los aspectos del manejo de datos y faciliten la obtención de informaciones útiles. El análisis espacial pasará de tratar sólo con regiones pequeñas o con escalas locales de análisis (o de regiones mayores pero con resolución en los datos menos finas) a mayores escalas de análisis. Esto a su vez hace aparecer serios problemas de no estacionalidad y complejos modelos de asociación que requieren ser considerados. En pocas palabras, unos SIG con un intensivo del ordenador precisan de métodos analíticos que soporten esta escala de análisis. La modelización de complejas estructuras espaciales requiere un análisis de datos cuantitativos espaciales de multinivel y de multitextura. Ello, a su vez exige métodos de modelización de la estadística espacial y de la matemática espacial así como formas complejas de simulación que sean facilitadas por las mejoras en la potencia de los ordenadores.

3.2.2. Aplicaciones y aproximaciones a la modelización

Los métodos del análisis espacial de datos se han desarrollado y puestos en práctica en contextos muy diferentes. En sus primeras aplicaciones, los modelos espaciales fueron utilizados para valorar los efectos de la competencia, midiéndolas por el rendimiento entre plantas creciendo a una distancia fija separadamente. Mead (1967) Whittle (1954) y Cliff y Ord (1973, 1981), utilizaron estadísticas de correlación para probar el ajuste de modelos de difusión espacial del tipo Hägerstrand a datos reales.

09 Goodchild y Haining 5/7/05 13:14 Página 196

Haining (1987) utilizó autoregresiones espaciales unilaterales para estimar multiplicadores de población y de renta en ciudades con organización basada en un sistema de lugar central. Anselin (1988), al tratar este terreno como una rama de la econometría (econometría espacial) desarrolló una estrategia de modelización estadística con un software para implementar la metodología (Spacestat) que sigue la estrategia empleada en ciertas formas de modelización econométrica de las series temporales. Encontramos numerosos ejemplos de utilización de la modelización de regresión espacial en una amplia variedad de dominios (Haining, 1990, 2003).

La modelización espacial están modificando su énfasis y apuntando a nuevos retos para el análisis de datos espaciales como el de valorar la correspondencia entre resultados de modelos y datos reales. Para aclarar estos hagamos referencia a dos ejemplos.

Las aproximaciones para comprender el crecimiento regional de los 60 y 70 se centraron en el papel del sector exportador. La modelización se basaba en regiones definidas previamente entre las cuales los factores de producción se moverían de forma similar a los flujos de bienes. La modelización econométrica regional y el input-output fueron las estructuras analíticas que implementaron un enfoque arriba-abajo (top-down) en el cual las relaciones inter e intrarregiones se especificaban normalmente en términos de un gran número de parámetros.

Los economistas de la «Nueva Geografía Económica» se preocuparon del crecimiento regional y de entender como el funcionamiento de la economía a escala regional afecta a los resultados de una economía nacional (Krugman 1995; Porter, 1988) y de su comercio (Krugman, 1991). Un rasgo central del modelo de Krugman es el «tirón de guerra entre fuerzas que tienden a promover la concentración geográfica y las que tienden a oponerse a ella, o en otras palabras, entre fuerzas centrípetas y centrífugas» (Krugman, 1996). Las últimas incluyen deseconomías externas como la congestión y los costes de polución, rentas de la tierra y factores no móviles. Las centrípetas incluyen economías externas tales como el acceso a los mercados y las ventajas naturales. En el centro de los modelos de esta Nueva Geografía Económica hay una consideración de la economía del espacio como una estructura compleja, autoorganizada y adaptativa. Es compleja en el sentido de existencia de un gran número de productores y consumidores individuales. Es autoorganizada a través de la mano invisible de los procesos de mercado. Es adaptativa en el sentido que los consumidores y productores responden a cambios en los gustos, estilos de vida y tecnología, por ejemplo. La Nueva Geografía económica se basa en rendimientos crecientes, de los que la estructura espacial es una propiedad emergente (Waldrop, 1992). Los resultados de los modelos se caracterizan por bifurcaciones, de forma que los cambios de una estructura espacial a otra pueden provenir de pequeños cambios en los parámetros subyacentes.

Los modelos deterministas de Krugman parecen compartir, conceptualmente, un campo común con los modelos de multi-agentes utilizados en la modelización urbana. En los modelos multi-agentes, los agentes autónomos activos interactúan y cambian su localización así como sus propios atributos. Los individuos responden no sólo a la información local sino también a la global o de la dimensión del sistema. La

09 Goodchild y Haining 5/7/05 13:14 Página 197

estructura espacial en la distribución de individuos es una propiedad emergente y los modelos multi-agentes, a diferencia de los del enfoque regional usados en la modelización urbana desarrollados en los setenta y ochenta no se basan en zonas predefinidas y utilizan muchos menos parámetros (Benenson, 1998).

Estos modelos estocásticos han sido utilizados para simular la conducta residencial de los individuos en una ciudad. Han evolucionado desde los enfoques de modelización de autómatas celulares a las estructuras urbanas, pero describen una visión dinámica de las pautas de interacción humana y de las conductas espaciales (Benenson, 1998; Xie, 1996). En el modelo de Benenson la probabilidad de emigración de una familia es función de la tensión económica local o de la disonancia cognitiva que la misma experimenta en su localización actual. La probabilidad de trasladarse a una casa vacía es una función de los nuevos niveles de tensión económica o de la disonancia cognitiva que experimentaría en la nueva localización.

Un punto de interés tanto para los modelos multi-agentes o de autómatas celulares es cómo estructuras complejas y modificaciones de tales estructuras pueden surgir desde procesos espaciales bastante sencillos y con pocos parámetros (Batty, 1998; Benenson, 1998; Portugali y otros, 1994; White y Engelen, 1994). La inclusión de una interacción espacial puede conducir a resultados fundamentalmente diferentes acerca de la existencia y estabilidad de equilibrios que repitan las conductas de la fase de transición en algunos procesos físicos (Follmer, 1974; Haining, 1985). Es la posibilidad de producir una estructura espacial de nuevas y limitadas formas, junto con el hecho de que la introducción de relaciones espaciales en los modelos familiares pueden proporcionar nuevas percepciones, en algunos casos sorprendentes, lo que subyace en el interés actual por el espacio en ciertas áreas de las ciencias sociales. Este interés, como señaló Krugman (1996) se sostiene en nuevas áreas de las matemáticas y moderna informática que hace posible analizar estos sistemas.

Las interacciones a escala local entre unidades elementales fijas, estén definidas en términos de individuos o de pequeñas áreas, pueden afectar tanto a las propiedades locales como a las de todo el sistema. Este efecto se muestra también en ciertos modelos de comercio intraurbano en los que la fijación de precios en un lugar responde a estrategias de precios de vecinos competidores (Haining y otros, 1996; Sheppard y otros, 1996). La modelización multi-agente añade otro ejemplo a nivel de sistema al conjunto de interacciones, haciendo posible que los individuos puedan emigrar en el espacio respondiendo a condiciones globales y locales en distintos segmentos del espacio.

Todas estas formas de modelización plantean cuestiones acerca de cómo deben compararse las expectativas generadas por los modelos con los datos observados al objeto de conseguir la validación del modelo, de manera que la validación se base en algo más que el grado de correspondencia visual con las pautas reales La comparación de la estructura espacial generada por la simulación del modelo con las estructuras espaciales observadas supone otro aspecto que requiere directamente métodos de análisis de datos espaciales.

En términos de la tecnología de adquisición de datos, en los de almacenamiento y exposición de los datos que utilizan los SIG y en términos de modelizar los paradig-

09 Goodchild y Haining 5/7/05 13:14 Página 198

mas hay un creciente interés por la microescala. En términos de como se comporta la ciencia hay un creciente interés por la interdisciplinariedad, la colaboración y la comunicación más allá de los limites tradicionales. El análisis espacial de datos tendrá que responder en los años venideros a los cambios en este gran marco.

4. Conclusión

Este trabajo de revisión se inició con comentarios acerca de las diferencias entre las dos visiones del mundo (una la que representan los SIG y otra la que proporciona el análisis de datos espaciales) y con una discusión acerca de sus diferentes raíces. Los últimos 40 años han mostrado una amplia evidencia de convergencia, de forma que las dos disciplinas han reconocido su complementariedad esencial. A medida que la ciencia se mueve hacia una nueva era de colaboración basada en la tecnología y en la ciberinfraestructura, explotando instrumentos que han llegado a ser crecientemente esenciales para una ciencia preocupada por la comprensión de sistemas complejos, parece claro que los SIG y el análisis de datos espaciales se necesitan uno a otro. Comparten la misma relación existente entre los paquetes estadísticos y la estadística o entre los procesadores de texto y la escritura. Es más difícil analizar las grandes cantidades de datos de que disponen los científicos regionales y comprobar nuevas teorías e hipótesis sin una infraestructura informática. A la vez, la existencia de tales infraestructuras abre posibilidades para diversas clases de teorías y modelos enteramente nuevos y para nueva clases de datos.

También las dos perspectivas han hecho mucho para estimular el pensamiento en cada una de ellas. Los SIG son más ricos a consecuencia de las demandas de análisis de datos espaciales y este último también es más rico gracias al enfoque que los SIG han iniciado sobre cuestiones de representación y ontología. La preocupación por los mapas completos, el enfoque nomotético de la ciencia está dando paso a un nuevo enfoque, centrado en los lugares en el que las variaciones en la superficie de la tierra son tan potencialmente interesantes como la uniformidad. La visión de los SIG como un asistente inteligente está dando paso a una nueva consideración de los SIG como un *medium* de comunicación en el que el análisis espacial de datos es una de las formas de realzar el mensaje.

Los SIG constituyen el resultado de adaptación de tecnologías genéricas a las necesidades específicas de datos espaciales. En tal sentido su futuro está asegurado puesto que no hay falta de nuevas tecnologías en el crisol. Las nuevas tecnologías han estimulado también la nueva ciencia dado qué los investigadores han comenzado a reflexionar acerca de las implicaciones de las nuevas fuentes de datos o de las nuevas actividades basadas en la tecnología. Estas a su vez han estimulado nuevas clases de métodos analíticos y nuevas hipótesis acerca del mundo geográfico. El proceso de estímulo y convergencia que se iniciara en los años sesenta entre los SIG y el análisis espacial de datos está lejos de haberse completado. La interacción entre ambos probablemente seguirá siendo interesante y productiva durante muchos años del futuro.

09 Goodchild y Haining 5/7/05 13:15 Página 199

Bibliografía

Anderson, T.W. (1971): *The statistical analysis of time series*. Wiley, New York.

Angel, S. y Hyman, G.M. (1976): *Urban fields: A geometry of movement for regional science*. Pion, London.

Anselin, L. (1988): *Spatial econometrics: Methods and models*. Kluwer Academic, Dordrecht.

Anselin, L. (1995): «Local indicators of spatial association - LISA». *Geographical Analysis,* 27:93-115.

Bartlett, M.S. (1995): *An introduction to stochastic processes*. Cambridge University Press, Cambridge.

Batty, M. (1998): «Urban evolution on the desktop: Simulation with the use of extended cellular automata». *Environment and planning* A, 30: 1943-1967.

Benenson, I. (1998): «Multi-agent simulations of residential dynamics in the city. Computing», *Environment and Urban Systems,* 22: 25-42.

Berry, B.J.L., Marble, D.F. (1968): *Spatial analysis*. Prentice Hall, Englewood Cliffs, NJ.

Besag, J.E. (1974): «Spatial interaction and the statistical analysis of lattice systems». *Journal of the Royal Statistical Society,* B36:192-225.

Besag, J.E., York, J., Mollié, A. (1991): «Bayesian image restoration with two applications in spatial statistics». *Annals of the Institute of Statistical Mathematics,* 43: 1-21.

Besag, J., Newell, J. (1991): «The detection of clusters in rare diseases». *Journal of the Royal Statistical Society* A, 154: 143-155.

Burrough, P.A., Frank, A.U. (eds) (1996): *Geographic objects with indeterminate boundaries*. Taylor and Francis, New York .

Casella, G., George, E.I. (1992): «Explaining the Gibbs sampler». *The American Statistician,* 46:167-174.

Chorley, R.J. (1972): *Spatial analysis in geomorphology*. Methuen, London.

Cliff, A.D., Ord, J.K. (1973): *Spatial autocorrelation*. Pion, London.

Cliff, A.D., Ord, .JK. (1981): *Spatial processes*. Pion, London.

Craig, W.J., Harris, T.M., Weiner, D. (eds) (2002): *Community participation and geographic information systems*. Taylor and Francis, London.

Cressie, N. (1984): «Towards resistant geostatistics». In: Verly G, David M, Journel AG, Marechal A (eds) *Geostatistic for natural resourses characterization*. Reidel, Dordrecht, pp. 21-44.

Cressie, N. (1991): *Statistics for spatial data*. Wiley, New York.

Diggle, P. (1983): *Statistical analysis of spatial point patterns*. Academic Press, London.

Diggle, P.J., Chetwynd, A.D. (1991): «Second-order analysis of spatial clustering for inhomogeneous populations». *Biometrics,* 47: 1155-1163.

Duckham, M., Goodchild, M.F., Worboys, M.F. (2003): *Fundamentals of geographic information science*. Taylor and Francis, New York.

Economist (2003): The revenge of geography. 15 March.

Follmer, H. (1974): «Random economies with many interacting agents». *Journal of Mathematical Economics,* 1: 51-62.

Foresman, T.W. (1998): *The history of GIS: Perspectives from the pioneers*. Prentice Hall PTR, Upper Saddle River, NJ.

Frank, A.U., Raper, J., Cheylan, J.P. (eds) (2001): *Life and motion of socio-economic units*. Taylor and Francis, New York.

Gatrell, A.C., Bailey, T.C., Diggle, P.J., Rowlingson, B.S. (1996): «Spatial point pattern analysis and its application in geographical epidemiology». *Transactions*, Institute of British Geographers, 21: 256-274.

Geary, R.C. (1954): «The contiguity ratio and statistical mapping». *The Incorporated Statistician,* 5:115-145.

Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B. (1995): *Bayesian data analysis*. Chapman and Hall, London.

Goodchild, M.F. (1992): «Geographical information science». *International Journal of Geographical Information Systems*, 6:31-45.

Goodchild, M.F. (1998): «Uncertainty: The Achilles heel of GIS?» *Geo Info Systems* (November): 50-52.

Goodchild MF (2002): Measurement-based GIS. In Shi W, Fisher PF, Goodchild MF (eds) *Spatial data quality*. Taylor and Francis, New York, pp. 5-17.

09 Goodchild y Haining 5/7/05 13:15 Página 200

Greig-Smith, P. (1952): The use of random and contiguous quadrats in the study of the structure of plant communities. *Annals of Botany,* 16: 293-316.

Grenander, U., Rosenblatt, M. (1957): *Statistical analysis of stationary time series*. Wiley, New York.

Haining, R.P. (1985): «The spatial structure of competition and equilibrium price dispersion». *Geographical Analysis,* 17:231-242.

Haining, R.P. (1987): «Small area aggregate income models: Theory and methods with an application to urban and rural income data for Pennsylvania». *Regional Studies,* 21:519-530.

Haining, R.P. (1990): *Spatial data analysis in the social and environmental sciences*. Cambridge University Press, Cambridge.

Haining, R.P. (1994): Designing spatial data analysis modules for GIS. In: *Fothering gham* AS, Rogerson P (eds) Spatial analysis and GIS. Taylor and Francis, London, pp. 45-63.

Haining, R.P. (2003): *Spatial data analysis: Theory and practice*. Cambridge University Press, Cambridge.

Haining, R.P., Plummer, P., Sheppard, E. (1996): «Spatial price equilibrium in interdependent markets: price and sales configuration». *Papers of the Regional Science Association,* 75:41-64.

Johnston, J. (1963): *Econometric methods*. McGraw Hill, New York. Existe traducción española de J. Mª Otero: Métodos de Econometría. Vicens Vives. Barcelona 1975.

Kendall, M.G. (1976): *Time series.* Griffin, London.

Krishna, Iyer, P.V.A. (1949): «The first and second moments of some probability distributions arising from points on a lattice and their applications». *Biometrika,* 36:135-141.

Krugman, P. (1991): *Geographic and trade*. MIT Press, London.

Krugman, P. (1995): *Development, geography and economic theory*. MIT Press, London. Existe traducción española de Adelina Comas: Desarrollo, Geografía y Teoría Económica. Antonio Bosh. Barcelona 1997.

Krugman, P. (1996): «Urban concentration: The role of increasing returns and transport costs». *International Regional Science Review,* 19: 5-30.

Kulldorff, M. y Nagarwalla, N. (1995): «Spatial disease cluters: Detection and inference». *Statistics in Medecine,* 14: 799-810.

Langran, G. (1992): *Time in geographic information systems*. Taylor and Francis, New York.

Liverman, D. (ed.) (1998): *People and pixels: Linking remote sensing and social science*. National Academy Press, Washinton, DC.

Longley, P.A., Goodchild, M.F., Maguire, D.J., Rhind, D.W. (2001): *Geographic information systems and science*. Wiley, New York.

Manguire, D.J., Goddchild, M.F., Rhind, D.W. (eds) (1991): *Geographical information systems: Pinciples and applications*. Longman Scientific and Technical, Harlow.

MacEachren, A.M., Monmonier, M. (1992): «Introduction (to special issue on geographic visualization)». *Cartography and Geographic Information Systems,* 19: 197-200.

Malinvaud, E. (1970): *Statistical methods of econometrics*. North Holland Publishing Company, Amsterdam.

Matern, B. (1986): *Spatial variation* 2nd ed. Springer, Berlin Heidelberg New York.

Matheron, G. (1963): Principles of geostatistics. *Economic Geology,* 58: 1246-1266.

McHarg, I.L. (1992): *Design with nature*. Wiley, New York.

Mead, R. (1967): «A mathematical model for the estimation of inter-plant competition». *Biometrics,* 23: 189-205.

Mollié, A. (1996): Bayesian mapping of disease. In: Lawson, A., Biggeri, A., Bohning, D., Lesaffre, E., Viel, J.F., Bertollini, R. (eds) *Markov Chain Monte Carlo in practice: Interdisciplinary statistics*. Chapman and hall, London, pp. 359-379.

Moran, P.A.P. (1948): «The interpretation of statistical maps». *Journal of the Royal Statistical Society,* B10: 243-251.

Moran, P.A.P. (1950): Notes on continuous stochastic phenomena. *Biometrika,* 37: 17-23.

Openshaw, S. (1983): The modifiable a real unit problem. *Concepts and Techniques in Modern Geography* 38. GeoBooks, Norwich.

09 Goodchild y Haining 5/7/05 13:15 Página 201

Openshaw, S., Charlton, M., Wymer, C., Craft, A. (1987): «A Mark 1 geographical analysis machine for the automated analysis of point data sets». *International Journal of Geographical Information Systems,* 1: 335-358.

Peuquet, D.J. (2002): *Representations of space and time*. Guilford, New York.

Porter, M.E. (1998): *The competitive advantage of nations*. MacMillan, London. Existe traducción española: La Ventaja Comparativa de las naciones. Plaza y Janes, Barcelona, 1991.

Portugali, J., Benenson, I., Omer (1994): Sociospatial residential dynamics: Stability and instability within a self-organizing city. *Geographical Analysis,* 26: 321-340.

Ripley, B.D. (1977): «Modelling spatial patterns». *Journal of the Royal Statistical Society* B39:172-192.

Ripley, B.D. (1981): *Spatial Statistics*. Wiley, New York.

Sheppard, E., Haining, R., Plummer, P. (1992): «Spatial pricing in interdependent markets». *Journal of Regional Science,* 32: 55-75.

Stoyan, D., Kendall, W.S., Mecke, J. (1989): *Stochastic geometry and its applications*. Wiley, New York.

Tomlinson, R.F., Calkins, H.W., Marble, D.F. (1976): *Computer handling of geographical data: An examination of selected geographic information systems*. UNESCO Press, Paris.

Ungerer, M.J. y Goodchild, M.F. (2002): «Integrating spatial data analysis and GIS: A new implementation using the Component Object Model (COM)». *International Journal of Geographical Information Science,* 16: 41-54

University Consortium for Geographic Information Science (1996): Research priorities for geographic information science. *Cartography and Geographic Information Science,* 23: 115-127.

Waldrop, M.W. (1992): *Complexity*. Simon and Schuster, New York.

White, R., Engelen, G. (1994): Cellular dynamics and GIS: Modeling spatial complexity. *Geographical Systems,* 1: 237-253.

Whittle, P. (1954): On stationary processes in the plane. *Biometrika,* 41: 434-449.

Wold, H. (1954): *A study in the analysis of stationary time series*. Almqvist and Wiksells, Uppsala.

Xie, Y. (1996): A generalized model for cellular urban dynamics. *Geographical Analysis,* 28:350-373.

Zhang, J.X., Goodchild, M.F. (2002): *Uncertainty in geographical information*. Taylor and Francis, New York.

09 Goodchild y Haining 5/7/05 13:15 Página 202