Procedimientos informáticos para el análisis de datos geográficos

M. J. VIDAL DOMÍNGUEZ

J. BOSQUE SENDRA

Universidad Complutense

A. MORENO JIMÉNEZ

Universidad Autónoma de Madrid

*Introducción **

En el quehacer diario de la investigación geográfica se pone de manifiesto la existencia de una serie de tareas eminentemente no científicas (o paracientíficas), por cuanto consisten simplemente en labores mecánicas de las que el estudioso podría, en muchos casos, desentenderse sin menoscabo de la investigación en sí. De ellas podrían enumerarse las siguientes:

1. Recogida de datos.
2. Realización de recuentos y tabulaciones.
3. Ejecución de cálculos numéricos: que abarca desde la obtención de índices simples (porcentajes, coeficientes, etc.) hasta la evaluación de relaciones complejas (cálculos estadísticos y multivariados, ajuste y validación de modelos, etc.).
4. Representación gráfica de los resultados: histogramas, diagramas, mapas, etc.

No puede negarse que el contacto directo del investigador con estas tareas le proporciona ciertas enseñanzas, relativas sobre todo a la fiabilidad y peculiaridades de las fuentes; sin embargo, es asimismo evidente que, cuando el volumen de información y datos a manejar crece mucho, los trabajos arriba mencionados pueden absorber una gran parte del tiempo total destinado al estudio. De ello resultan casi ineludiblemente dos consecuencias negativas: por un

* Los números entre paréntesis de los cuadros hacen referencia a la bibliografía final.

lado, una baja «productividad», en términos de la relación cantidad del factor trabajo/producto, y, por otro, una desviación o distracción de la principal actividad de la investigación (el pensamiento genuinamente creativo) hacia operaciones secundarias; no se debiera perder de vista al respecto que si en Geografía, como en cualquier otra disciplina, es importante describir y comparar, el énfasis ha de ponerse siempre en la explicación, y que es en este punto donde recae en última instancia el juicio valorativo del trabajo y su aportación desde una óptica científica.

Son de sobra conocidas las limitaciones que han impedido y todavía impiden llevar a cabo en nuestro país una investigación geográfica que libere, aunque sólo fuera parcialmente, al estudioso de las tareas mecánicas que se le presentan, y que reducen indudablemente su rendimiento. El propósito último de esta nota consiste en mostrar las posibilidades y la ayuda concretas que para la ejecución de algunas de esas tareas amanuenses pueden hallarse en un instrumento y una técnica cada vez más difundidos: el ordenador y el proceso de datos. Quisiéramos advertir al respecto, y aunque pueda parecer ocioso, que el empleo de la Informática no equivale a decir «Geografía cuantitativa» (aunque lo contrario sea una relación obligada); como podrá verse a lo largo de estas líneas, el ordenador presta actualmente unos servicios que rebasan el mero ámbito de la cuantificación en Geografía.

El objetivo concreto perseguido aquí es ampliar la información ofrecida por uno de nosotros (Moreno, 1979) relativa a los programas de ordenador de interés geográfico, tanto los desarrollados por la propia comunidad científica de los geógrafos para resolver sus peculiares problemas científicos, como los elaborados en otros campos pero que por su flexibilidad y asepsia pueden ser empleados, y de hecho así se hace con éxito, en la realización de tareas y cálculos con datos geográficos. Debe darse por supuesto que las referencias aquí recogidas no cubren con mucho la totalidad de las existentes; es más, somos conscientes de la magnitud de las lagunas, pero, tras haber realizado una labor de búsqueda de información en algunos de los temas que nos han parecido prioritarios, resulta aconsejable difundirla, aunque sea con la restricción antedicha y de forma sintética, de cara a ponerla en conocimiento de los geógrafos interesados en el tema.

La exposición se ha organizado en torno a tres elementos distintos: en primer lugar, la relativa a programas que son de aplicación general, es decir, aquellos que procedentes de origen no geográfico permiten por su propia naturaleza ser transplantados a nuestra disciplina con éxito; como puede suponerse, la información que presentamos en este apartado está muy seleccionada, dada la cantidad exis-

tente de ella. En segundo lugar, los programas elaborados por y para geógrafos o geólogos, que poseen un mayor interés por cuanto se orientan a aplicaciones específicas de nuestra disciplina; y en tercer lugar, las publicaciones periódicas de las que se pueden extraer los programas que habitualmente incluyen.

Antes, sin embargo, quisiéramos hacer una precisión terminológica, diferenciando los sistemas de programas de los programas *sensu strictu*, ya que el manejo de unos u otros suele diferir bastante. Se considera un sistema de programas a un conjunto de ellos estructurados de forma coherente y compacta, dotados de una sintaxis de acceso específica y simple, y asequibles, por tanto, a usuarios con mínimos o nulos conocimientos de Informática. El sistema suele admitir y aportar mensajes en un lenguaje muy conversacional y fácilmente inteligible, siendo ésta la única forma que habitualmente relacionará al usuario con los programas integrados en el sistema. Por el contrario, un programa *sensu strictu* se presenta al usuario normalmente como un paquete de sentencias escritas en un lenguaje de programación habitual (por ejemplo, FORTRAN, ALGOL, etc.) cuya aplicación a la resolución de problemas, y por supuesto su manipulación interna, exige inicialmente conocimientos sobre dicho lenguaje, si bien una vez puestos a punto pueden ser usados con relativa facilidad; de todas maneras en este caso la «conversación» entre el usuario y el programa se reduce al mínimo y la información ofrecida por éste, a excepción de los resultados, suele ser bastante críptica y su interpretación exige conocimientos especializados.

Sistemas de programas de aplicación general

El número de sistemas de programas de ordenador disponibles para realizar aplicaciones estadísticas y matemáticas, de posible interés general para diversas ciencias, es muy importante. En el cuadro I se recogen los más difundidos y útiles de los que conocemos.

Las bibliotecas BMD y BMDP, elaboradas ambas por la Universidad de California en Los Angeles (USA), son dos generaciones de programas que, aunque pensadas para un uso más específico en las Ciencias Biomédicas, tienen multitud de usos generales. Están centradas en los tratamientos estadísticos de la información, incluyendo desde la estadística descriptiva elemental hasta diversos métodos multivariados muy sofisticados. Las diferencias entre BMD y BMDP estriban esencialmente en la mayor simplicidad de la sintaxis del BMDP y la mayor perfección y expresividad de sus salidas gráficas (aparte de los programas específicos de representación gráfica reseñados en

CUADRO I

Tipos de análisis	Programas	Bmdp ⁷	Bmd ⁴	Sas ²⁴	Spss ²⁰	Ost-ris ²²	Ssp ²⁶	Tsp ²⁷	Ni-Svs ²¹	Md. (*) ¹⁴	Ad-dad 1.10	Le-bar ¹¹ Fen-elon
Estadística	Descriptiva	1D, 2D, 4D, 9D	01D, 04D, 06D, 07D	Freq, Means, Standard, Summary, Univariate	Condescriptive Frequences, Aggregate	×	×		×			
	Inferencial paramétrica	3D, 6D, 7D, 8D, 1F, 1V, 2V	02D, 03D, 08D, 09D, 10D, 11D, 12D, 13D, 01R, 02S, 01V a 12V	Anova, Autoreg, Corr, Glm, Nested, Nparlway, Plans, T-Test	Crosstabs Breakdown, T-Test, Partial Corr, Portial Corr, Anova, Oneway	×	×		×			×
	Infr. no paramétrica	1F, 3S		Duncan	Crosstabs, Nonpar Corr	×	×					×
	Análisis de series temporales		02T a 05T				×	×				
	A. Factorial	4M	01M, 08M, 02M	Factor, Var-comp	Factor	×	×		×			×
Análisis multivariado	A. Clasificación	1M, 2M, 3M		Cluster		×			×	×	×	
	A. Discriminante	7M	04M, 05M, 07M	Discrim, Neighbor	Discriminant		×		×			
	Corr. Canónica	6M	09M	Cancorr	Cancorr		×					
	Reg. Múltiple	1R a 6R	02M, 02R a 07R	Glm, Nlin, Rsquare, Score, Stepwise, Sys-reg	Regression	×	×					×
	A. Correspondencias										×	×
Análisis matemático	Otros		10M	Guttman	Guttman Scale	×			×	×		
	Cálculo numérico y matricial	1S	09S, 10S, 12S, 13S	Matrix			×		×			
	Modelos		01S, 03S, 11S	Probit, Spectra								
Gráficos y diagramas	5D, 6D	05D	Char, Plot	Scattergram				×		×		

(*) Los acrónimos dentro del cuadro especifican los programas que ejecutan los análisis correspondientes en cada sistema de programas. Cuando por diversas circunstancias no se pudo hacer así, hemos indicado con una cruz la posibilidad de realizar los análisis.

el cuadro citado, casi todos los restantes programas de análisis producen diagramas bastante útiles).

El sistema SPSS (Statistical Package for Social Sciences), redactado en la Universidad de Chicago (USA), está dotado de un lenguaje de acceso muy cercano al lenguaje natural (por supuesto inglés), y se ha elaborado pensando en las Ciencias Sociales (Sociología especialmente). Añade a los procedimientos estadísticos más conocidos, muy semejantes a los contenidos en las anteriores bibliotecas BMD y BMDP, los métodos de escalogramas múltiples de Guttman.

Por fin, el SAS (Statistical Analysis System), de la Universidad norteamericana de Carolina del Norte, es un reciente sistema de programas que intenta ser el más completo, flexible y general procedimiento de análisis estadístico para todo tipo de datos científicos. El SAS asume todas las ventajas del SPSS y los propios programas del BMDP pueden ser integrados como un subprograma más del SAS. Admite salidas gráficas mucho más amplias y elaboradas, incluyendo el empleo del «plotter», incluso en colores, y la pantalla de rayos catódicos. Además, contiene varios procedimientos de análisis estadístico nuevos, no integrados en los sistemas antes citados. Realmente el SAS, dotado además de una sintaxis muy simple, es el sistema de programas estadísticos con mayor futuro en este momento.

Menos ambiciosos son el resto de los sistemas de programas que mencionamos a continuación. El TSP (Time Series Processor), creado en la Universidad de Lovaina (Bélgica), está diseñado específicamente para el análisis y descripción de series temporales, especialmente desde el punto de vista de la Econometría. MDS(x) y ADDAD son dos conjuntos de programas, de menos envergadura, desarrollados, respectivamente, por la Universidad de Edimburgo (Gran Bretaña) y la Association pour le Developpement et le Difussion de l'Analyse de Données. El ADDAD incluye análisis de correspondencias, la versión del análisis factorial elaborado en Francia. Los paquetes OSIRIS y NT-SYS son grupos de programas muy semejantes a los anteriores, pero de mayor antigüedad y menor versatilidad. El NT-SYS está orientado hacia el tratamiento de problemas de taxonomía numérica (clasificación), ofreciendo una amplia gama de modelos y posibilidades al investigador; inicialmente está pensado para aplicaciones en Biología y Ecología.

El SSP (Scientific Subroutine Package) de IBM no constituye exactamente un sistema de programas en el sentido de la definición dada; es más bien un conjunto de subprogramas (subrutinas) utilizables para elaborar programas complejos; estas subrutinas resuelven una multitud de cálculos básicos, tanto de estadística como de cálculo numérico.

CUADRO II

<i>Programas</i>		<i>Leeds</i> ³⁰	<i>GPE</i> ⁸	<i>Baker</i> ²	<i>Keyfitz</i> ¹³	<i>Gino</i> ²⁹
<i>Tipos de análisis</i>						
<i>G. Física</i>	Geomorfología	Hidrodat, Topmodel, Floodn, Oxbala	Climat, Koppen			
	Climatología					
<i>G. Humana</i>	G. de la población	Enfoque demográfico	Simple		×	
		Enfoque espacial	Multipop, Dame			
	G. Transportes e interacción espacial		Entropy, Indiff, SPA, Presloc, Trans, Weberi	Network, Transal		
	G. Agrario G. Urbano y áreas de influencia		Landuse Luam, Sires, Solup	Market		
<i>Regionalización</i>			Engrp	Grpward		
<i>Análisis espacial</i>	Descripción e interplación		Centro, Geneb, Geodis, Gridmentro, Intrapol, Locate, Naybor	Locatin, Point, Stdist		
	Modelos de ajuste espacial		Dyad, Mave, Double, Geopart, Polyfit, Singre, Trend, Spectr			
<i>Modelos matemáticos espaciales</i>	Gravedad Gravitatorios		Gravty (u), Locpot Equal, Scope, Trendl	Gravity Potent		
	Difusión Simulación: C. Markov		Simlit Markov 1 y 2, Noncel	Extrap		
	Modelos de asignación de localización		Alloc, Multilap, Maptran, Inpmed, Torn, Altern			
<i>Estadística</i>		Multireg	Covar, Norm, Optreg, Ridreg			
<i>Análisis multivariado</i>		IDD	Mdisc, Canon, Cenda, Engrp, Losine, Disgrp			
<i>Representación gráfica</i>	Cartografía plana		Cart 1, Cart 2, Camp, Entour, Mona, Maplot, Contr (APL), Cpletm, Gipsy, Intrap (PA)	Gridmap		×
	Diagramas		Nscat, Plots, Dendron, Heragon, Poppyr			×
	Salidas tridimensionales		Azmat, Burge, Blck, Pers			×
<i>Gestión de coordenadas: Base de datos espaciales</i>			Coord, Sdis, Dstaz, Turgz			

Finalmente, libros y publicaciones de carácter general que editen programas son abundantes y no añaden con frecuencia nuevas prestaciones a las ofrecidas por los sistemas hasta aquí enumerados. De todas formas merece la pena citar los libros de Lebart y Fenelon (1972) y Veldman (1967).

En resumen, este conjunto de programas de aplicación general resuelven con rigor y precisión gran número de las tareas de análisis y tratamiento de la información requeridos en la investigación geográfica. No obstante, sus planteamientos básicos, o muy generales o muy específicos de otras ciencias (Medicina, Biología y Sociología), pueden resultar algo inadecuados para los concretos requerimientos de la Geografía como ciencia espacial. Por ello cada vez más se plantea la necesidad de elaborar programas de ordenador que atiendan las citadas exigencias específicas de la Geografía.

Programas de ordenador de orientación geográfica

Hace ya varios años que la necesidad antes aludida ha comenzado a ser cubierta por el trabajo de geógrafos e informáticos, existiendo en la actualidad una amplia serie de programas que abarcan abundantemente muchos de los problemas más específicos de la investigación geográfica. No obstante, podemos señalar todavía la falta de un sistema de programas completo diseñado específicamente para la Geografía y que cubra la mayor parte, si no la totalidad de sus requerimientos de investigación.

Entre los programas ya disponibles y más difundidos destaca, en primer lugar, los realizados en el Laboratorio de Cartografía Automática de la Universidad de Harvard (USA), en concreto el conocido SYMAP, además de otros semejantes y más recientes, que utilizan el «plotter» y la pantalla de rayos catódicos interactiva para la representación cartográfica. En esta misma línea se puede mencionar el GIPSY (ver cuadro II).

Otras instituciones científicas americanas, como el Kansas Geological Survey, son el origen de varios programas pensados específicamente para el análisis de datos espaciales, en principio de la Geología, aunque se podrían emplear igualmente para la Geografía, tanto Física como Humana, es el caso del llamado SURFACE II, en el que se contienen procedimientos modernos de la estadística espacial: variograma, krigeage.

La Universidad inglesa de Leeds, por su parte, se ha centrado en dos temas: programas de estudio de la población e Hidrología. En los primeros están incluidos análisis y proyecciones de la población en varias regiones.

En el GPE (Geography Program Exchange) se ha establecido un procedimiento de intercambio de programas de ordenador redactados por los diferentes Departamentos de Geografía, especialmente anglosajones, que resulta de gran utilidad. El GPE reúne más de 100 programas, que tocan los más diversos temas y cuestiones de interés geográfico: representación gráfica y cartográfica, análisis estadístico, uso de modelos matemáticos geográficos (difusión espacial, gravedad, transporte, entropía, etc.).

La bibliografía especializada constituye otra fuente de programas de ordenador interesantes. En muchas obras de Geografía cuantitativa, junto al desarrollo de los procedimientos estadísticos y matemáticos, se contienen listados de programas de ordenador adecuados para utilizar tales técnicas de análisis. Entre los libros de este tipo más interesantes destacan los de Mather, centrado en la estadística multivariada; Davis, que incluye algunos elementos de estadística espacial, y el de McDougall, que analiza en profundidad métodos de representación gráfica y sus problemas de programación.

En último lugar, no debemos olvidar los sistemas de programas de origen español: el SIGNA, una base de datos para tratamiento y recuperación de datos espaciales, diseñada en el IGN, y la BSG (Biblioteca de Subprogramas Gráficos), que está siendo desarrollada en el Centro de Cálculo de la Universidad Complutense de Madrid, mediante la fructífera colaboración de geógrafos e informáticos, que contiene programas de cartografía automática plana y en tres dimensiones, además de diagramas varios.

Publicaciones periódicas

Además de lo hasta aquí mencionado existen revistas o series más o menos irregulares que publican programas de ordenador; de ellas hemos realizado una selección basada en el criterio de su orientación geográfica o parageográfica (cuadro III), aunque algoritmos difundidos en las revistas especializadas de Informática puedan eventualmente ser de utilidad en nuestra disciplina.

Información específica sobre el tema de las representaciones gráficas, cubriendo no sólo el aspecto de los programas sino también el de la maquinaria *ad hoc*, sistemas y servicios, puede obtenerse en el Harvard Directory of Computer Graphics Suppliers, publicado anualmente en Sudbury (Massachusetts, 01776, P.O. Box 89).

Conclusión final

La amplitud y complejidad de todas las referencias aquí recogidas evidentemente plantean un grave problema de acceso para los geó-

CUADRO III

<i>Denominación</i>	<i>Organismo editorial</i>
1. <i>Computer Application in the Natural and Social Sciences.</i>	Department of Geography, Univ. of Nottingham (GB).
2. <i>Computers and the Humanities.</i>	Association for Computers and the Humanities, Queens College (Nueva York, USA).
3. <i>Computer Contributions</i> (publicación interrumpida).	State Geological Survey, Univ. of Kansas (USA).
4. <i>Geocom Programs.</i>	Geosystems, Londres (GB).
5. <i>Environment and Planning, A.</i>	Ed. Pion, Londres (GB).
6. <i>Programmation et Sciences de l'Homme.</i>	Ecole Normale Supérieure, París (Francia).
7. <i>Lund Studies in Geography, Series C</i>	Department of Geography, The Royal Univ. of Lund (SW).
8. <i>Technical Reports.</i>	Computer Institute for Social Science Research, Michigan State University (USA).
9. <i>Working Papers.</i>	School of Geography, Univ. of Leeds (GB).

grafos españoles interesados en estas cuestiones. Una posible solución consistiría en la organización de algo semejante al GPE a nivel español, en el que todos los trabajos de programación, cada vez más numerosos entre los geógrafos de nuestro país, puedan ser fácilmente difundidos entre los diferentes Centros de Cálculo y Departamentos de Geografía.

En este sentido queremos aportar un primer paso, enumerando los programas actualmente disponibles en los Centros de Cálculo de uso público de Madrid. En concreto son los siguientes: ADDAD, BMD, BMDP, BSG, GIPSY, MATHER, MDS(x), NT-SYS, OSIRIS, SAS, SIGNA, SPSS, SYMAP.

Evidentemente, como se puede comprobar en la lista anterior, todavía las aportaciones propias de los geógrafos españoles son muy reducidas, y casi todo lo actualmente disponible son trabajos elaborados en otros países, muchos de ellos cubiertos por el Copyright, lo que impide su fácil difusión; sin embargo, su uso en los Centros de Cálculo que los poseen suele ser sencillo y asequible.

Esperamos que todas estas noticias ayuden a la Geografía española a crear una nueva línea de trabajo, que nos acerque rápidamente a las posiciones alcanzadas ya en otros países.

BIBLIOGRAFIA

1. ADDAD: Association pour le Development et la Diffusion de l'Analyse des Données, Laboratoire de Statistique, Univ. P. et M. Curie, París.
2. BAKER, L.: *A selection of geographical computer programs*, Geographical Papers, n.º 6, London School of Economics and Political Science, Dep. of Geography, 1974.
3. BAXTER, R. S.: *Computer and statistical techniques for planners*, Methuen, Londres, 1976, 340 p.
4. *BMD. Biomedical Computer Programs*. Univ. of California Press, 1973.
5. DAVIS, J. C.: *Statistics and data analysis in Geology*, John Wiley and Sons, Nueva York, 1973, 550 p.
6. DAVIS, J. C., y McCULLAGH, M. J. (dir.): *Display and analysis of spatial data*. John Wiley and Sons, Londres, 1975, 378 p.
7. DIXON, W. J.: *Biomedical Computer Programs (Series P)*, Univ. of California Press, Berkeley, 1975, 792 p.
8. *Geography Program Exchange*. Computer Institute for Social Science Research, Michigan State University, East Lansing, Michigan.
9. HERRERO, R.; BOSQUE, J., y CEBRIÁN, J. A.: «Nuevas tendencias en la investigación geográfica: el sistema de información del Instituto Geográfico Nacional (SIGNA)», *Est. Geo.*, núm. 161, 1980, pp. 447-466.
10. JAMBU, M., *Classification automatique pour l'analyse des données*, Dunod, París, 1978, 2 vols.
11. KADMON, N.: «KOMPLOT: 'Do-it-yourself' computer cartography», *Cartographic Journal*, 8, 1971, pp. 139-144.
12. KANSAS GEOLOGICAL SURVEY: *FORTRAN IV. Programs-Computer Contributions*, Lawrence, The Univ. of Kansas.
13. KEYFITZ, N., y FLIEGER, W.: *Demografía: métodos estadísticos*, Marymar, Buenos Aires, 1975, 609 p.
14. KRUSKAL, J. B.: «Multidimensional scaling: A numerical method», *Psychometrika*, 29, 1964, pp. 1-27.
15. LIBART, L., y FENELON: *Statistique et Informatique appliquées*, Dunod, París, 1972.
16. McDOUGALL, E. B.: *Computer programming for spatial problems*, Edward Arnold, Londres, 1976, 158 p.
17. MATHER, P. M.: *Computational methods of multivariate analysis in Physical Geography*, John Wiley and Sons, Londres, 1976, 532 p.
18. MONGE, P. R., y CAPPELLA, J. M.: *Multivariate techniques in human communication research*, Academic Press, Nueva York, 1980.
19. MORENO, A.: «La geografía española y la Informática», *Est. Geo.*, núms. 156-157, pp. 499-501, 1979.
20. NIE, W. J. (dir.): *SPSS. Statistical Package for Social Sciences*, McGraw-Hill, Nueva York, 1975.
21. *NT-SYS: Numerical Taxonomy*. Department of Ecology and Evolution, The State University of New York (EE. UU.).
22. *OSIRIS III. System and program description*. Institute for Social Research, The Univ.
23. SAMPSON, R. J.: «The SURFACE II graphics system», en DAVIS y McCULLAGH (1975), pp. 244-267.
24. *SAS: User's Guide* (ed. 1979). SAS Institute Inc. Raleigh, North Caroline, Estados Unidos.
25. SCHMIDT, A. H., y ZAFFT, W. A.: «Programs of the Harvard University Laboratory for computer graphics and spatial analysis», en DAVIS y McCULLAGH (1975), pp. 231-244.

26. *SSP: Scientific Subroutine Package*. GH20-0205-4, Sistem 360, Programmer's Manual. IBM.
27. *TSP. Time Series Processor. User's manual*. Institut des Sciences Economiques, Univ. Catholique de Louvain.
28. VELDMAN, D. J.: *FORTRAN programming for the behavioral sciences*, Holt-Rinehart and Winston, Nueva York, 1967.
29. WOODSFORD, P. A.: «The design and implementation of the GINO 3D graphics software package», *Software: Practice and Experience*, 1, 1971, pp. 335-365.
30. *Working Papers*. School of Geography, Univ. of Leeds (publicación no periódica).