Ecología Austral: 5:75-80,1995
Asociación Argentina de Ecología

P.An.D.A. Programa para el análisis de datos areográficos. Desarrollo y aplicaciones

Edmundo M. Lopasso[1] y Adriana Ruggiero[2]
1 Centro Atómico Bariloche. Av. Bustillo 9.5 km. (8400) Bariloche, Río Negro, Argentina
2 Depto. de Ecología. Centro Regional Univer sitario Bariloche, C. C. 1336 (8400) Bariloche, Río Negro, Argentina

Resumen. *Se presenta un programa (P.An.D.A.), desarrollado según los standards del FORTRAN 77. El programa aplica técnicas de dimensionamiento variable y de almacenamiento compacto de los datos que le permiten procesar bases de datos extensas en computadoras de tipo PC. Dada una matriz básica de datos que registra la presencia-ausencia de especies en cada celda de un mapa grillado, el programa permite estimar valores de riqueza de especies, diversidad taxonómica y similitud biótica en cada celda. El P.An.D.A. ha sido empleado para analizar la distribución geográfica de 533 especies de mamíferos, archivadas como datos de presencia-ausencia en cada una de las 481 celdas de un mapa grillado de América del Sur. Probó ser altamente eficiente para manejar matrices que sobrepasarían la capacidad de memoria de computadoras de tipo PC sin expansiones de memoria.*

Abstract. *We introduce P.An.D.A., a computer program developed according to FORTRAN 77 standards. The application of variable dimensioning techniques and the compact storage of data allow the program to handle large data bases in limited personal computers. Given a basic data matrix that records the presence-absence of species in each cell of a grid, the program can estimate species richness, taxonomic diversity, and biotic resemblance for each cell. P.An.D.A. was used to analyze the geographical distribution of 533 species of mammals recorded as presence-absence data in 481 cells of a grid map of South America. It proved to be highly efficient while managing matrices that otherwise would overwhelm the memory capacity of computers that lack memory expansions.*

Introducción

El uso generalizado de computadoras abre nuevas posibilidades para realizar estudios biogeográficos cuantitativos. Numerosos trabajos han mostrado la conveniencia de archivar numéricamente la información geográfica de las especies para constituir una base de datos susceptible de ser estudiada por métodos analíticos (e.g. Peters 1968, Crovello 1981, Wartenberg 1985, Berry 1987). Aunque usualmente el análisis de mapas de distribución geográfica por técnicas numéricas no involucra una dificultad matemática importante, el procesamiento de la información en computadoras de tipo PC-IBM compatible puede complicarse seriamente cuando la base de datos es grande, debido a restricciones en la capacidad de memoria. En este trabajo se presenta un programa de cálculo y se describen las técnicas de computación utilizadas para estimar valores de (1) riqueza de especies (2) diversidad taxonómica (=jerárquica) (ver p. ej., Pielou 1975, Rabinovich y Rapoport 1975) y (3) similitud biótica (Rapoport 1975, 1979), a partir de datos de distribución geográfica de especies que han sido archivados numéricamente (i.e., registrados como datos de presencia-ausencia sobre una superficie grillada). El P.An.D.A. produce archivos de salida que pueden utilizarse como matrices de entrada a sistemas de información geográfica (e.g., PCMAP version 2.0; SURFER version 4.0, Golden Software Inc. 1989).

Indices Biogeográficos

Dada una matriz básica de datos (MBD), que registra la presencia-ausencia (1-0) de n_s especies en cada una de las n_c celdas de un mapa grillado, y conociendo el arreglo taxonómico de los datos (i.e., la distribución de las especies estudiadas en géneros y familias), el P.An.D.A. permite calcular, en cada celda, los siguientes estimadores:

(1) Riqueza de especies, géneros o familias: número de especies, géneros o familias registrados en cada celda.

(2) Diversidad taxonómica ($H'_{spp/gén}$ $H'_{spp/fam}$ y $H'_{gén/fam}$): aplicación de la fórmula de Shannon-Weaver (Shannon 1948): $H' = -\Sigma p_i \log_2 p_i$, donde p_i, representa, respectivamente, la distribución de especies en géneros, de especies en familias y de géneros en familias.

(3) Similitud biótica(sensu Rapoport 1975, 1979): dado un punto geográfico designado como "muestra central" (m_c), se cuenta el número de especies presentes en ese sitio: luego se calcula, para cada una de las restantes celdas del mapa, la relación: (número de especies compartidas con m_c/número total de especies presentes en m_c) x 100. Esto constituye una medida de la probabilidad de encontrar cualquier taxón que habita la muestra central, en cada una de las restantes celdas del mapa.

Descripción del Programa

La característica más notable de la programación corresponde al aprovechamiento de la memoria, lograda con la utilización de la técnica de dimensionamiento variable (Stamm'ler y Abbate 1983) y con el almacenamiento compacto de los datos. Esto último permite que el programa conserve en la memoria sólo el registro de las dos celdas límite de un rango dc presencias de la MBD. Dado un mapa grillado numerado correlativamente de izquierda a derecha y de arriba hacia abajo (Figura 1), se denomina rango de presencias a un conjunto de celdas correlativas donde una especie se encuentra presente. Por ejemplo, la especie A de la figura posee un rango de presencias entre las celdas 85 y 89. El programa archiva la información de esta especie usando tres números (1 85 89): el primero indica la cantidad de rangos de presencias y, los dos siguientes, las celdas límite de ese rango. La especie B (Figura 1) ocupa cinco rangos de presencias, porque las celdas donde este taxón está presente no corresponden a celdas correlativas en la grilla. La distribución geográfica de esta especie se archiva usando 11 lugares de memoria (5 9 9 19 19 29 29 39 39 49 49): el primer número indica la cantidad de rangos de presencia, y cada par siguiente, las celdas límite de cada rango. Nótese que aunque las especies A y B ocupan la misma cantidad de celdas, se requiere mayor cantidad de memoria para almacenar la información de la especie B. La especie C representa a un taxón con distribución geográfica disyunta. Posee cinco rangos de presencias y se archiva utilizando 11 lugares de memoria (5 13 14 22 25 33 35 52 53 61 63). En total, el P.An.D.A. usa 25 lugares de memoria para almacenar la información geográfica de las especies A, B y C. De no utilizarse el almacenamiento compacto de los datos, se requerirían 300 lugares de memoria para archivar la misma información (i.e., 100 celdas donde se registra la presencia-ausencia de cada especie x 3 especies). El almacenamiento compacto de los datos hace que el requerimiento de espacio para almacenar todos los datos de un problema sea mínimo, aun cuando un taxón presente múltiples rangos. Las características mínimas requeridas para el uso del programa en una máquina de tipo PC son: 640 Kb de memoria RAM, disco rígido o dos unidades de disco flexible de 360 Kb cada una. Una opción deseable aunque no imprescindible es que posea coprocesador matemático. El P.An.D.A. puede utilizarse para procesar datos de distribución geográfica de especies que representan distintos tipos de problema (i.e., matrices de distinto tamaño). Los interesados pueden solicitar el programa a los autores.

P. An. D. A. no posee capacidades de graficación, ni modo interactivo en pantalla. Es sólo una herramienta para el cálculo de estimadores biogeográficos específicos a partir de una MBD numérica (1-0), y sólo cuando es utilizado en combinación con otros programas (e.g., SURFER 4.0) pueden obtenerse mapas de isoriqueza de especies, isodiversidad taxonómica o isodensidad biótica

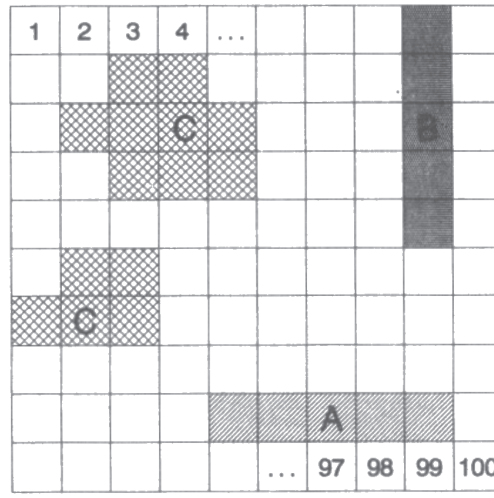

Figura 1. Ejemplo que muestra la distribución geográfica de tres especies en una grilla compuesta por 100 celdas. Las celdas están numeradas correlativamente de izquierda a derecha y de arriba hacia abajo. La aplicación de la técnica de almacenamiento compacto de los datos permite archivar la información geográfica de las tres especies usando únicamente 25 lugares de memoria (ver explicación detallada en el texto).
Figure 1. An example that shows the geographic distribution of three species in a grid composed of 100 cells. Cells are correlatively numbered from left to right and from up to down. The application of compact storage of data technique allows to record the geographic distribution of these three species by using only 25 memory places (see text for a detailed explanation).

(=equiprobabilísticos) (ver referencias citadas en la introducción del presente trabajo). Importando, por ejemplo, los archivos de salida del P.An.D.A. al sistema SURFER (version 4.0, Golden Software Inc. 1989), se pueden efectuar interpolaciones entre los valores obtenidos para cada celda del mapa, graficar curvas de nivel (isolíneas) y, superponiendo un contorno digitalizado del continente o país estudiado, obtener el mapa de isolíneas deseado. Es necesario destacar que el tipo de grilla (o "grano" de análisis), como también las coordenadas (x,y) de cada celda, no son definidas mediante el uso del P. An. D. A., sino que deben estar previamente establecidas por el usuario al archivar numéricamente la información de base.

Eficiencia del programa en el empleo de la memoria

El P.An.D.A. ha sido empleado para analizar la distribución geográfica de 533 especies de mamíferos sudamericanos con el objeto de describir la variación geográfica de los índices de resistencia ambiental (R_{50}) y anisotropía (A_{50}) propuestos por Rapoport (1975, 1979), y analizar los patrones geográficos en la riqueza de especies y diversidad taxonómica de los distintos taxones estudiados (para una descripción detallada de las hipótesis propuestas y material y métodos utilizados, consultar Ruggiero 1993). En ese trabajo, se superpuso una grilla compuesta por 481 celdas de 200 km de lado sobre un mapa de América del Sur, registrándose la presencia-ausencia de cada especie en cada una de las celdas. Con estos datos, se pudo comprobar la eficiencia del programa en el uso de la memoria, utilizando una computadora PC-AT 286 de 25 MHz con disco rígido de 40 Mb, memoria de 640 Kb, coprocesador matemático y tarjeta de video tipo Hércules. La Tabla 1 ejemplifica algunas de las corridas realizadas para estimar valores de riqueza de especies, diversidad taxonómica y similitud

Tabla 1. Eficiencia del P.An.D.A. en el empleo de la memoria de una computadora de tipo PC al procesar los datos de presencia-ausencia de 533 especies de mamíferos registrados en una grilla de 481 celdas superpuesta a un mapa de América del Sur. Los datos fueron reagrupados en conjuntos que constituyen matrices de diferente tamaño. S=número de especies. I=número de índices biogeográficos calculados. Mp(bytes) =espacio de memoria potencialmente necesario si se utilizara la matriz completa. Me (bytes)= memoria efectivamente utilizada. Eficiencia= 100 (1- Me/Mp). * indica que el espacio de memoria requerido excedería el disponible en las computadoras sin expansión de memoria.
Table 1. P.An.D.A.'s efficiency in the utilization of a PC-computer memory while procesing presence-absence data of 533 species of mammals recorded in a grid of 481 cells overlaid on a South American map. Data were arranged in groups to compose matrices of different size. S=number of species. I=number of biogeographic indices that were calculated. Mp(bytes) =memory needed to store the whole matrix. Me(bytes)=memory effectively used by the program. Efficiency=100 (1- Me/Mp). * cases in which the capacity of a PC without memory expansion would be exceeded.

S	I	M_p	M_e	Eficiencia
12	8	38480	18608	52
13	8	40404	18664	54
14	8	42328	18636	56
18	5	44252	13392	70
28	8	69264	20336	71
43	1	84656	12032	86
43	5	92352	17804	81
44	8	100048	21804	78
54	3	109668	16980	85
54	8	119228	20860	83
77	8	163540	23152	86
122	8	250120	23852	90
122	6	246272	19972	92
187	5	369428	30900	92
187	15	388648	44416	89
187	8	375180	36704	90
187	6	371332	32848	91
533	26	1075516	98272	91*
533	15	1054352	77020	93*
533	25	1073592	96344	91*
533	18	1060124	82968	92*
533	20	1063972	86704	92*

faunística, sobre la base del reagrupamiento de los datos en MBD de distinto tamaño. Puede verse que la cantidad de registros potencialmente necesarios para almacenar la información puede exceder en uno o dos órdenes de magnitud a la cantidad efectivamente utilizada por el P.An.D.A, siendo muy eficiente para manejar la información de matrices grandes (e.g., 533 x 481), las cuales quedan compactadas a alrededor de un 10% de su tamaño original (Tabla l). Esta medida, sin embargo, no debe tomarse como un indicador absoluto ya que la eficiencia del programa para compactar la información depende de la estructura de los datos que componen la MBD. Similarmente, es difícil estimar un límite absoluto para su capacidad de almacenamiento de datos ya que ésta depende de las siguientes variables: (1) el tamaño de la grilla; (2) el tipo de distribución geográfica de las especies:

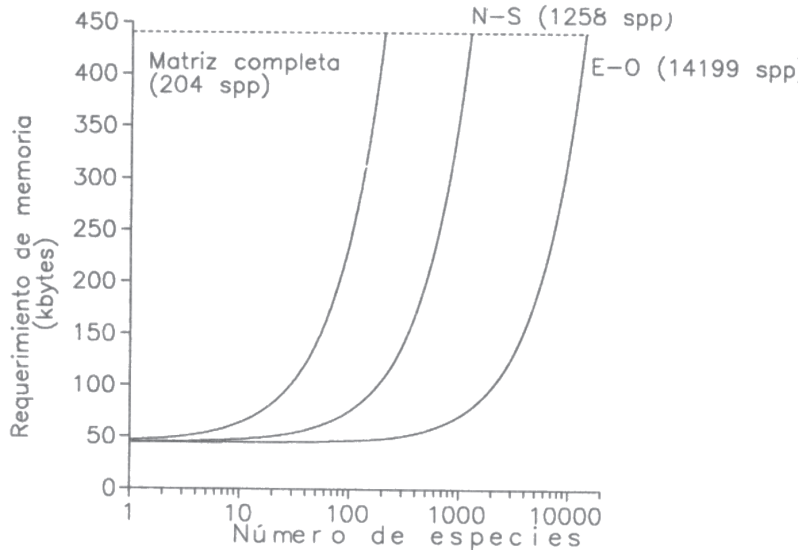

Figura 2. Estimación de un posible límite de memoria del programa sobre la base de considerar: una grilla compuesta por 481 celdas, NESP=NGEN=NFAM, valor en abscisas, NCELD=481, CALC=20. Para ejemplificar el caso de las especies distribuidas en sentido longitudinal (E-O) se supuso que cada especie ocupa un bloque de presencias (similar a la especie A de la Figura 1), con lo cual BLOQUES= NESP. Para ejemplificar el caso de las especies distribuidas en sentido latitudinal (N-S) se supuso que cada especie se distribuye a lo largo de la máxima extensión del mapa que corresponde a 37 celdas verticales (similar a la especie B de la Figura 1), con lo cual BLOQUES= 37*NESP.
Figure 2. Estimation of a possible memory limit based on the following assumptions: a grid composed of 481 cells. NESP=NGEN=NFAM, abscissa value, NCELD=481, CALC=20. To illustrate the case of species having a E-W geographic distribution each species was supposed to occupy a single presence block (similar to species A in Figure 1); as a consequence, BLOQUES=NESP. To illustrate the case of species having a N-S geographic distribution each species was supposed to be distributed along the maximum latitudinal extension of the map which is represented by a total of 37 vertical cells (similar to species B in Figure 1); as a consequence: BLOQUES=37*NESP.

aquellas especies extendidas mayormente en sentido longitudinal (E-O) ocuparán menos lugares de memoria (i.e., podrán compactarse mejor) que aquellas especies extendidas mayormente en sentido latitudinal (N-S). Esto es así por la forma en que el P. An. D. A. almacena los rangos de presencia (ver sección "Descripción del programa"); (3) el número de estimadores calculados en una corrida; (4) La estructura taxonómica de los datos: dados dos conjuntos de datos, compuestos por igual número de especies, pero que difieren porque en uno las especies están distribuídas en pocos géneros y familas mientras que en el otro existen muchos géneros y/o familias representados por una única especie, el P.An.D.A. utilizará más lugares de memoria para almacenar la información de este último caso.

El requerimiento de memoria para un caso general (MREQ) se calcula mediante la fórmula: MREQ=1+NESP+NGEN+NFAM+2*NCELD+MAX(NCELD,NESP+NGEN)+2*BLOQUES +(NCELD+1)*CALC, donde NCELD, NESP, NGEN y NFAM representan el número de celdas, especies, géneros y familias respectivamente. MAX(NCELD,NESP+NGEN) indica el valor máximo entre NCELD y NESP+NGEN. BLOQUES indica el número total de bloques de presencias y se calcula sumando la cantidad de bloques de presencias de todas las especies. CALC es igual a NISO+RESP+RGEN+RFAM+MED+2*DTEG+2*DTEF+2*DTGF. NISO es la cantidad de celdas utilizadas como muestra central en el cálculo de valores de similitud biótica. Los siguientes parámetros adquieren el valor 1 si se calcula el estimador, y 0 si no se lo calcula: RESP (riqueza de especies), RGEN (riqueza de géneros), RFAM (riqueza de familias), DTEG (diversidad taxonómica especie-género), DTEF (diversidad taxonómica especie-familia) y DTGF (diversidad taxonómica

género-familia).

A través de un ejercicio numérico relativamente simple puede estimarse un límite posible de memoria del programa para el caso de una grilla compuesta por 481 celdas (Figura 2). En esta aproximación se ha supuesto que cada especie representa a un único género y a una única familia. El ejemplo permite observar cómo influye el tipo de distribución geográfica de las especies sobre el límite de capacidad 'de almacenamiento de datos (línea punteada en la Figura 2). Se observa que el número de especies que el P. An. D. A. puede analizar en una corrida puede incrementarse en un orden de magnitud si las especies están distribuidas en sentido longitudinal. Por otra parte, la Figura 2 remarca las ventajas del almacenamiento compacto de los datos ya que si la información geográfica de las especies se archivara en forma de una matriz completa o expandida (i.e., almacenando todos los datos de presencia y ausencia (1 y 0) registrados en cada una de las celdas), sólo podrían analizarse 204 especies bajo los supuestos del presente ejemplo. Este número contrasta con las 1255 o las 14199 especies que pueden archivarse mediante el almacenamiento compacto de los datos. Todas las corridas realizadas para analizar la distribución geográfica de los mamíferos sudamericanos representan casos ubicados bien por debajo de este límite de memoria del programa (ver Tabla 1). Se comprueba que el P. An. D. A. es un programa de cálculo eficiente para el manejo de bases de datos extensas utilizando computadoras de tipo PC-IBM compatible de configuraciones mínimas.

Agradecimientos. Los comentarios de Alejandra C. Guasp y Mónica de Torres Curth contribuyeron a mejorar el manuscrito. La publicación del presente trabajo ha sido autorizada por CAB/CNEA.

Bibliografía

Berry, J. K. 1987. A mathematical structure for analyzing maps. Environmental Management 11:317-325.

Crovello, T. J. 1981. Quantitative biogeography: An overview. Taxon 30:563-575.

Peters, J. A. 1968. A computer program for calculating degree of biogeographical resemblance between areas. Systematic Zoology 17:64-69.

Pielou, E. C. 1975. Ecological Diversity. Wiley, New York, 165 pp.

Rabinovich, J. E. y E. H. Rapoport.. 1975. Geographical variation of diversity in Argentine Passerine birds. Journal of Biogeography 2:141-157.

Rapoport, E. H. 1975. Areografía. Estrategias geográficas de las especies. México: Fondo de Cultura Económica, 214 pp.

Rapoport, E. H. 1979. Movimiento, dispersión y barreras: hacia un modelo predictivo del curso de plagas. En (J. E. Rabinovich y G. Halffter, eds) Tópicos Selectos de Ecología Contemporánea. Fondo de Cultura Económica, México: 47-53.

Ruggiero, A. 1993. Areografía de América del Sur. Un análisis basado en el estudio de sus mamíferos. Tesis Doctoral. Facultad de Ciencias Naturales y Museo. Universidad Nacional de La Plata.

Shannon, C. E. 1948. The mathematical theory of communication. Bell System Technical Journal 27:379, 623.

Stamm'ler R. J. J. y M.J. Abbate. 1983. Methods of Steady-State Reactor Physics in Nuclear Design. Academic Press. London, 506 pp.

Wartenberg, D. 1985. Canonical trend surface analysis: A method for describing geographic patterns. Systematic Zoology 34:259-279.

Recibido: 20/10/93
Aceptado: 2/10/94