

00381



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO  
FACULTAD DE CIENCIAS

POSGRADO EN CIENCIAS BIOLÓGICAS  
FACULTAD DE CIENCIAS

---

**Biogeografía Cuantitativa:**  
**análisis de métodos**  
**y**  
**desarrollo de herramientas**

---

**TESIS**

que para obtener el grado académico de  
Doctor en Ciencias (Biología) presenta

**Miguel Murguía Romero**

Director de Tesis:

**Dr. José Luis Villaseñor Ríos**

México, D. F., Febrero del 2005

341709



Universidad Nacional  
Autónoma de México

Dirección General de Bibliotecas de la UNAM

**Biblioteca Central**



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

*Dedico esta tesis a mi esposa e hijos:*

*Isabel,  
Isaac  
y  
Quetzalli*

Autorizo a la Dirección General de Bibliotecas de la UNAM a difundir en formato electrónico e impreso el contenido de mi trabajo recepcional.  
NOMBRE: Miguel Murguía Domínguez  
FECHA: 7-Marzo-2005  
FIRMA: Miguel Murguía

# Agradecimientos

---

El Dr. José Luis Villaseñor me acompañó durante el desarrollo de los trabajos de investigación, orientándome y escuchándome. El apoyo que el me otorgó desempeñó un papel muy importante para mi formación científica. El haber trabajado con José Luis me otorgó beneficios y satisfacciones adicionales a los productos que se presentan en esta tesis, tanto en el ámbito científico como en el meramente humano.

Mi comité tutorial me apoyó durante todo el proceso, desde el acotamiento de los objetivos, las críticas y observaciones a mis propuestas, la generación de los artículos y capítulos de libros, hasta la conformación del documento de tesis; agradezco por ello al Dr. Jorge Llorente, al Dr. Luis Eguiarte y al Dr. José Luis Villaseñor. La discusión crítica y siempre productiva con ellos, me contagió un ánimo positivo y optimista. Las sugerencias de lecturas, cursos, ejercicios y contacto con otros investigadores me facilitaron la definición de los problemas, la integración de conceptos y la búsqueda de soluciones.

La revisión de versiones previas de este documento por el Dr. Luis Eguiarte, Dra. Patricia Koleff, Dra. Amparo López, Dr. Jorge Llorente, Dr. Guillermo Ibarra, Dr. Hugo Rincón, Dr. Oswaldo Téllez y Dr. José Luis Villaseñor, fue una etapa que me ayudó a organizar mejor la presentación de mis ideas. La Dra. Patricia Koleff me indicó importantes anotaciones sobre la diversidad beta y su medición, así como observaciones sobre conceptos de bases de datos biogeográficas. La Dra. Amparo López me ayudó a precisar el aspecto informático de bases de datos. El Dr. Guillermo Ibarra hizo importantes sugerencias sobre varios conceptos biogeográficos y sobre la estructura del documento. El Dr. Hugo Rincón me aportó una visión desde el lado matemático haciéndome ver las coincidencias y diferencias en la manera de aplicar los formalismos; también me hizo notar algunos aciertos y sugerencias respecto al modelo de los caminos metodológicos. El Dr. Oswaldo Téllez me hizo importantes observaciones sobre los diferentes modelos de representación geográfica en relación al análisis biogeográfico.

Tuve algunas discusiones con la Dra. Patricia Dávila que me brindaron nuevos puntos de vista y me ayudaron a orientar mi trabajo; de igual manera, sostuve algunas pláticas con el Dr. Juan José Morrone que me aportaron visones adicionales y necesarias para mi trabajo.

La Dirección General de Asuntos del Personal Académico (DGAPA) de la UNAM me otorgó una beca para realizar los estudios de doctorado.

La revista *Annales Botanici Fennici* me otorgó permiso para reproducir los artículos de los que ella es propietaria de los derechos de copia.

# Artículos

---

Los siguientes capítulos de esta tesis incluyen las publicaciones señaladas:

Capítulo 1:

MURGUÍA, M. Y J. E. LLORENTE. 2003. **Reflexiones conceptuales en Biogeografía Cuantitativa**. En: MORRONE, J. J. Y J. LLORENTE (Eds.). **Una perspectiva Latinoamérica de la Biogeografía**. UNAM. 307p.

Capítulo 6:

MURGUÍA, M. Y F. ROJAS. 2001. **Biogeografía Cuantitativa**. En: LLORENTE, J. Y J.J. MORRONE (Ed.). **Introducción a la Biogeografía en Latinoamérica: Teorías, Conceptos, Métodos y Aplicaciones**. UNAM. 277p.

Capítulo 7:

MURGUÍA, M. Y J. L. VILLASEÑOR, 2000. **Estimating the quality of the records used in quantitative biogeography with presence-absence matrices**. *Ann. Bot. Fennici*, 37: 289-296.

Capítulo 8:

MURGUÍA, M. Y J. L. VILLASEÑOR, 2003. **Estimating the effect of the similarity coefficient and the cluster algorithm on biogeographic classifications**. *Ann. Bot. Fennici*, 40: 415-421.

# Capítulos

---

Resumen.....	1
Abstract .....	2
1. Introducción .....	3
2. Contexto de la Biogeografía Cuantitativa .....	5
3. Problemas fundamentales de la Biogeografía Cuantitativa.....	33
4. Objetivos .....	39
5. Método.....	42
6. Presentación de los métodos de la Biogeografía Cuantitativa .....	51
7. Estimación de la calidad de los registros biogeográficos .....	53
8. Efectos del índice de similitud en la clasificación biogeográfica.....	56
9. Análisis de métodos de la Biogeografía Cuantitativa .....	58
10. Desarrollo de un SIG para análisis biogeográfico .....	89
11. Discusión.....	142
12. Conclusiones .....	148
Referencias.....	154

# Contenido

---

<b>Resumen</b> .....	1
<b>Abstract</b> .....	2
<b>1. Introducción</b> .....	3
<b>2. Contexto de la Biogeografía Cuantitativa</b> .....	5
2.1. Panbiogeografía .....	5
2.2. Métodos multivariados y su aplicación en la Biogeografía, Ecología y Taxonomía .....	15
2.3. El concepto de especie .....	21
2.4. El concepto de semejanza en términos de interpretación de Biología comparada y evolutiva .....	26
2.5. Ubicación de la investigación en Taxonomía y Biogeografía en el contexto de la Biología de la conservación .....	29
<b>3. Problemas fundamentales de la Biogeografía     Cuantitativa</b> .....	33
3.1. Métodos de la Biogeografía Cuantitativa .....	33
3.2. Herramientas para la Biogeografía Cuantitativa .....	37

<b>4. Objetivos</b> .....	39
4.1. Análisis de Métodos de la Biogeografía Cuantitativa.....	39
4.2. Desarrollo de Herramientas para la Biogeografía Cuantitativa.....	41
<b>5. Método</b> .....	42
5.1. Criterios para el análisis e integración de los métodos de la Biogeografía Cuantitativa.....	43
5.2. Metodologías para el Desarrollo Informático.....	46
<b>6. Presentación de los métodos de la     Biogeografía Cuantitativa</b> .....	51
<b>7. Estimación de la calidad de los registros     biogeográficos</b> .....	53
<b>8. Efectos del índice de similitud     en la clasificación biogeográfica</b> .....	56
<b>9. Análisis de métodos de la Biogeografía     Cuantitativa</b> .....	58
9.1. Clasificación de los métodos de la Biogeografía Cuantitativa.....	58
9.2. Documentación de los métodos de la Biogeografía Cuantitativa.....	61
9.3. Un modelo de la Biogeografía Cuantitativa: Caminos metodológicos.....	83

<b>10. Desarrollo de un SIG para análisis biogeográfico</b> .....	89
10.1. Especificaciones generales de la interfaz .....	90
10.2. Diseño de la interfaz .....	95
10.3. Diseño de la Base de Datos .....	125
10.4. Modelo de Objetos .....	138
10.5. Implementación .....	140
<b>11. Discusión</b> .....	142
11.1. Fundamento teórico de los métodos de la Biogeografía Cuantitativa .....	142
11.2. Estado actual y prospectiva de las herramientas informáticas de la Biogeografía Cuantitativa .....	146
<b>12. Conclusiones</b> .....	148
12.1. Fundamento técnico y teórico de los métodos de la Biogeografía Cuantitativa .....	148
12.2. Análisis espacial y automatización de los métodos de la Biogeografía Cuantitativa .....	152
<b>Referencias</b> .....	154

# Resumen

---

Se describe el problema de la calidad de los datos usados para realizar análisis biogeográficos, proponiendo una manera de estimar la representatividad de la muestra mediante un coeficiente que involucra estimación de la riqueza en la región de estudio y estadísticas básicas de la matriz de presencia-ausencia (especies presentes en subregiones). Se realizan análisis empíricos para ejemplificar el uso del coeficiente propuesto, utilizando bases de datos a nivel nacional y a nivel local, involucrando a varios grupos de plantas con flores, tanto a nivel de familias como de géneros y especies. Se realiza un análisis de las fórmulas más comunes para calcular similitud entre subregiones a partir de matrices de presencia-ausencia, indicando algunos aspectos pragmáticos que auxilian en la decisión de su elección. Se realiza un análisis de los métodos de la biogeografía cuantitativa con el objeto de: 1) contar con un inventario de las técnicas más recurridas, 2) presentarlas de manera ordenada al biogeógrafo y 3) proveer de una herramienta informática que le auxilie directamente en sus investigaciones. Se propone un modelo integrador (“camino metodológico”) de los métodos de la biogeografía cuantitativa. Se desarrolla e implementa un diseño de un sistema de información geográfica (GIS) para análisis biogeográficos, que incluye una automatización de métodos para la clasificación de las muestras para conformar “Unidades Geográficas Operativas”, la estimación de los datos en las matrices, el cálculo de índices de similitud entre regiones a partir de las especies en común, y la clasificación biogeográfica mediante métodos de agrupamiento, como por ejemplo el algoritmo UPGMA.

Palabras Clave: Análisis espacial, Asteraceae, Conservación de la Biodiversidad, Biogeografía Cuantitativa, Calidad de muestreo, SIG, MAUP, UPGMA.

# Abstract

---

This thesis describes data quality issue in biogeographical analysis. A way to estimate sample representativity using a coefficient that estimates the species richness of the study area as some basic statistics of the presence-absence matrix (species occurrence in subregions) is presented.

Empirical analysis were performed to exemplify the coefficient proposed, using a data base at national and local levels, of several flowering plant groups at family, genus, and species level.

The most used formulae to calculate similarity between subregions based on presents-absents matrix were analyzed, showing some pragmatic aspects that might help in choosing such a formulae.

A geographical information system (GIS) design for biogeographical analysis was developed and implemented, including automation of methods to classify samples to build "Operative Geographic Units", data matrices estimations, calculation the similarity index between regions from common species and biogeographical classification using grouping methods, i.e. the UPGMA algorithm.

**Key words:** Asteraceae, Biodiversity Conservation, Quantitative Biogeography, Quality Sampling, Spatial analysis, GIS, MAUP, UPGMA.

# 1. Introducción

---

*En este capítulo se explican los conceptos de biogeografía cuantitativa, análisis espacial y análisis biogeográfico. Se explican los alcances de los Sistemas de Información Geográfica (GIS) y el estado actual de las herramientas informáticas de la biogeografía cuantitativa. Se introduce al problema de la calidad de la información en las bases de datos biogeográficas.*

Este capítulo introduce y discute varios conceptos involucrados en el quehacer de la Biogeografía Cuantitativa, como son:

1. El concepto de análisis, análisis espacial y análisis biogeográfico
2. Los Sistemas de Información Geográfica y las Bases de Datos Geográficas y su uso en la Biogeografía
3. Cuantificación de la calidad de la información biogeográfica
4. Los índices de similitud
5. El concepto de Unidad Geográfica Operativa (UGO)
6. Los antecedentes y los fundamentos de la Biogeografía Cuantitativa
7. Algunos contrastes entre el método panbiogeográfico y el de la Biogeografía Cuantitativa
8. El problema de la escala y jerarquía en el análisis de patrones biogeográficos
9. El concepto de semejanza en la sistemática filogenético, la cladística y la biogeografía

Este capítulo es íntegramente el capítulo incluido en el libro de Morrone y Llorente (2003), titulado “Reflexiones conceptuales en Biogeografía Cuantitativa”:

MURGUÍA, M. Y J. E. LLORENTE. 2003. **Reflexiones conceptuales en Biogeografía Cuantitativa**. En: MORRONE, J. J. Y J. LLORENTE (Eds.). **Una perspectiva Latinoamérica de la Biogeografía**. UNAM. 307p.

# REFLEXIONES CONCEPTUALES EN BIOGEOGRAFÍA CUANTITATIVA

Miguel Murguía y Jorge Llorente Bousquets

En un análisis intervienen factores que introducen incertidumbre en los resultados; se aplican métodos y técnicas que lo auxilian reduciendo tal incertidumbre. Sin embargo, estos métodos y técnicas pueden introducir incertidumbre adicional; cuando se tiene conciencia de ello se intenta cuantificar para indicar en los resultados el 'intervalo de confianza'. El caso particular del análisis biogeográfico no escapa a esta situación; los diversos enfoques o 'escuelas biogeográficas' se aproximan al análisis de manera diferente, pero el proceso analizado es el mismo. La biogeografía es el estudio de la distribución de los organismos vivos, la historia de cómo se forma un arreglo particular espacial-geográfico de las especies y un análisis de las posibles causas de ese arreglo (Simberloff, 1983).

En este capítulo reflexionamos sobre aspectos conceptuales de la biogeografía, con especial énfasis en la biogeografía cuantitativa. Se aplican los conceptos de biogeografía cuantitativa, análisis espacial y análisis biogeográfico. Se comentan algunos alcances de los Sistemas de Información Geográfica (GIS, por sus iniciales en inglés) y el estado actual de las herramientas informáticas de la biogeografía cuantitativa. Se introduce al problema de la calidad de la información en las bases de datos biogeográficas y a otras consideraciones en el análisis biogeográfico como son: (1) el problema de la unidad de área modificable (MAUP, por sus iniciales en inglés), (2) las escalas, (3) la jerarquía, y (4) el concepto de semejanza. También se abordan algunas diferencias epistemológicas entre la biogeografía cuantitativa y el método panbiogeográfico mediante el examen de los conceptos de 'OGU' y 'nodo'.

## Biogeografía cuantitativa

En algunos análisis biogeográficos se incorporan técnicas y métodos cuantitativos que, dado su uso generalizado, pueden considerarse como parte de la biogeografía cuantitativa (BC). La lista de técnicas y métodos que conforman la BC está en continua evolución y, por lo tanto, es importante definir el conjunto que se considera bajo estudio, ya que abarcar todas las técnicas en una discusión resultaría poco práctico.

Murguía y Rojas (2001) muestran la gama de métodos de la BC, ejemplifican su aplicación y destacan los más recurridos. Ésos son los que, para efectos de la discusión teórica y técnica en este trabajo, conforman la BC, excluyendo así de este examen a cualquier otro tipo de método citado en la literatura y considerado también como parte de la BC.

## Análisis espacial

En la sección anterior se cita el término 'técnicas de análisis', por lo que es necesario explicar que se entiende por 'análisis', específicamente por 'análisis espacial'. El análisis espacial es un conjunto de técnicas para analizar eventos geográficos, en el que los resultados dependen del arreglo espacial de los eventos (Haining, 1994). Por 'evento geográfico' se entiende una colección de puntos, líneas o áreas localizados en un espacio geográfico que tienen asociado un conjunto de valores a atributos. Por lo tanto, contrasta con otras formas de análisis en que este análisis espacial requiere de los valores de los atributos y de la localización geográfica de los objetos.

Balcy (1994) distingue 'resumen' ('summarization') de 'análisis'. El primero tiene que ver con funciones básicas, como el despliegue selec-

tivo de información espacial en áreas de interés, y con el cálculo, la tabulación y el mapeo de estadísticos básicos de esa información. Puede constituir un requisito para el análisis, pero este último, el análisis, tiene que ver con la investigación de patrones de los datos espaciales.

Sistemas de información geográfica. Un sistema de información geográfica (GIS, por sus iniciales en inglés) es un conjunto de herramientas para recolectar, almacenar, recuperar, transformar y desplegar datos del mundo real (Burrough, 1986). Los GIS son sistemas que utilizan recursos de cómputo para modelar y analizar datos referenciados geográficamente en un contexto orientado a la solución de problemas para la toma de decisiones.

La relación que guarda un GIS con los análisis biogeográficos es a través de la restricción, generalmente aceptada, que un GIS lo es sólo si incluye herramientas de análisis. Un sistema que no incluye capacidad de análisis espacial no es un GIS, caso para el que los 'geoinformáticos' reservan el término de sistema AM/FM (Automated Mapping/Facility Management), que son sistemas con mayores facilidades que los cartográficos, normalmente con buenas capacidades de despliegue (generación de mapas) y de generación de informes, pero no de análisis. Burrough (1986, 1998) establece que la principal diferencia entre los GIS y los sistemas cartográficos asistidos por computadora es la capacidad de los primeros en la transformación de los datos espaciales originales para responder a preguntas específicas del usuario. Burrough brinda una clasificación de los análisis que normalmente puede realizar un GIS. Un GIS representa a la realidad mediante un modelo de datos; los sitios de recolecta de ejemplares biológicos pueden representarse en el GIS mediante 'puntos' o mediante un polígono. No debe perderse de vista al utilizar un GIS que la información con la que se está trabajando es un modelo de la realidad y no la realidad. El modelado de datos es el que permite hacer una generalización de diversos tipos de objetos en cuanto a su representación digital-informática, es decir en la computadora. El modelo de datos, sea 'ráster' o vectorial, es el que permite una representación 'analizable' de forma genérica. En un GIS, un punto puede representar un sitio de recolecta, un pueblo o un lago.

La capacidad de adaptar los algoritmos existentes a las estructuras de datos de un GIS es un componente crítico. O'Kelly (1994) pone especial atención al respecto y considera que la transferencia de las técnicas hacia la tecnología (es decir, implementar las técnicas utilizadas por los especialistas en campos específicos en un GIS) es una de las barreras a vencer. En cuanto a la BC, se puede decir que una de las barreras a vencer es implementar sus técnicas en un GIS y adaptar los algoritmos, por ejemplo, el UPGMA, a las estructuras de datos usadas por el GIS.

Bases de datos biogeográficas. Analizar datos implica o exige tener ya conformada una base de datos confiable y consistente con algún modelo ampliamente aceptado. Por lo tanto, para el análisis biogeográfico es requisito contar con tal base de datos.

Una base de datos puede considerarse un conjunto de tablas relacionadas, donde cada tabla está compuesta de un número predeterminado de columnas (campos o atributos) y un número variable de renglones (registros). Codd (1990) define exhaustivamente lo que es un Manejador de Bases de Datos Relacional.

El modelo de una base de datos sobre biodiversidad generalmente contiene un subconjunto biogeográfico y existen muchos modelos na-

dional e internacionalmente aceptados como estándares, que en muchos casos han sido propuestos por organizaciones ambientales. Sin embargo, los GIS verticalmente orientados a la BC son desarrollos que de modo reciente han iniciado, por lo que aún existen cuestiones técnicas del modelado de datos por resolver y mejorar.

### Análisis biogeográfico

Para definir el concepto de análisis biogeográfico se puede recurrir a las consideraciones sobre el concepto de análisis espacial presentados en la sección anterior y propuestos por Haining (1994) y Bailey (1994). Al seguir la perspectiva de Haining (1994) sobre el concepto de análisis espacial, por análisis biogeográfico se entiende la transformación de datos biogeográficos con el fin de responder preguntas biogeográficas. En particular, se puede considerar como análisis biogeográfico al procesamiento de los datos de las especies, asociados a sus localidades de recolecta, para responder preguntas biogeográficas como: ¿Cuál es el área de distribución de determinada especie? ¿En qué proporción coincide la distribución de dos grupos de especies? ¿Cuáles especies existen en determinada región? ¿Cuál es la región de mayor riqueza? ¿Cuáles regiones son prioritarias para la conservación de la biodiversidad?, entre otras.

Por otro lado, siguiendo el concepto de Bailey (1994) de análisis espacial, análisis biogeográfico es la investigación de patrones biogeográficos. Desde esta perspectiva, la generación de mapas sobre distribución de especies y las 'bioestadísticas', como la diversidad alfa promedio, beta y gama, no constituyen por sí solos análisis biogeográficos, sino solamente resúmenes de la información espacial (prerrequisitos para el análisis). Tomando en cuenta las definiciones de análisis espacial de Bailey (1994) y Haining (1994), se puede decir que análisis biogeográfico es una parte del proceso de describir, visualizar, comprender, interpretar, entender, explicar y predecir la distribución de los organismos vivos sobre la superficie de la Tierra.

Ésta es la definición de análisis biogeográfico que se adopta en este trabajo y ayuda a clasificar a las técnicas de la BC en descriptivas, analíticas y de modelado.

Jerarquización de las técnicas de la biogeografía cuantitativa. Definido el concepto de análisis biogeográfico, se puede definir una frontera entre lo que es una técnica descriptiva propiamente y otra de análisis biogeográfico. De manera general se propone la siguiente clasificación (más adelante se propone otra clasificación) de las técnicas biogeográficas según sean de aspecto descriptivo, analítico y de modelación:

#### Descripción

- . Conocimiento de la biodiversidad: ¿cuáles especies y dónde están?
- . Áreas de ocupación, distribución actual y distribución potencial.
- . Cuantificación de biodiversidad: diversidad alfa, beta y gama.
- . Generalización de la biodiversidad: conformación de OGU's (Operative Geographic Units' *sensu* Crovello, 1981).

#### Análisis

- . Definición y delimitación de áreas de endemismo.
- . Regionalización de la biodiversidad: clasificación biogeográfica.
- . Descubrimiento y proposición de sitios de alta riqueza (diversidad alfa).
- . Definición y reconocimiento de áreas de alta diversidad beta.
- . Reconocimiento de diversidad gama en regiones, subregiones y provincias bióticas.

#### Modelación

- . Modelos de la biodiversidad: correlación de una especie con otros factores ambientales (*v.gr.* temperatura o altitud); diseño de áreas naturales prioritarias para la conservación.
- . Predicción espacial: dinámica de áreas de distribución, estimadores de riqueza.

Cuantificación de la calidad de la información. Ninguna base de datos sobre las especies de una región es completa en tanto que no se tiene la seguridad de que se están incluyendo a todas las especies que habitan dicha región. Varios factores influyen en la complejidad de la información biogeográfica. Un grupo de ella lo constituyen los aspectos de carácter taxonómico, incluyendo la sinonimia, la determinación y el apego al sistema de clasificación, otro incluye aspectos de georreferenciación, como la homonimia sobre la toponimia, la asignación de coordenadas geográficas y la precisión de la georreferenciación.

Además, dada una lista de especies de una región, en forma general se podría decir que las especies incluidas tienen una probabilidad alta de estar 'realmente' presentes en la región, y que muchas otras que no se incluyen podrían tener una probabilidad (no cero) de estar presentes, pues pocas regiones están tan bien recolectadas al grado de poder decir que la lista de especies es completa (Soberón y Llorente, 1993). Lo anterior sin añadir la dimensión temporal, es decir, el problema de la calidad de la información es mayor si se desea conocer la lista de especies que habitan una región en diferentes periodos de tiempo.

La cuantificación de 'qué tan bien está recolectada una región' es un aspecto fundamental que auxilia en la descripción y en la validez o robustez de los resultados de los análisis. Esa cuantificación también es necesaria para poder establecer criterios de comparación entre estudios o listas de diferentes regiones. Algunos de los factores de la calidad de la información de una base de datos biogeográfica son:

- . Validez taxonómica (determinación taxonómica correcta)
- . Escala/precisión geográfica/espacial (método de georreferenciación y precisión registrada)
- . Cobertura espacial (ubicación y distribución espacial de los sitios de recolecta)
- . Cobertura temporal (año, estación, mes y hora de las muestras)
- . Cobertura taxonómica (número adecuado de grupos taxonómicos)
- . Actualidad (la BD debe estimar lo que existe hoy y no en años o décadas pasadas)
- . Volumen (número de sitios de recolecta y de ejemplares o muestras)

Definición de OGU's mediante rejilla. Una de las operaciones comunes previas a la realización de análisis geográficos es la 'generalización', en donde los puntos de muestreo geográfico se extrapolan para obtener modelos de superficies (áreas). En biogeografía cuantitativa esa unidad es la llamada OGU (Unidad Geográfica Operativa, por sus iniciales en inglés; Crovello, 1981).

Una de las formas de construir las OGU's o subdividir un área de estudio es definiendo una retícula (*v.gr.* Birks, 1976; Ezcurrea *et al.*, 1984; McAllister *et al.*, 1994; Castro *et al.*, 1996; Soberón *et al.*, 2000) que incluya cuadrantes (o polígonos) de dimensiones similares, y tratar a cada uno como una OGU. Uno de los problemas de este método es precisamente la definición del tamaño (cuadrante o polígono).

Rapoport (1975) menciona que el método de la rejilla ha sido útil, por ejemplo, en Gran Bretaña, donde se ha elegido un tamaño de 10 x 10 km (National Grid). Pero la forma y tamaño del área geográfica de la especie dependerán, ciertamente, del grosor de la trama elegida. Griffith (1993) analiza las características de una rejilla hexagonal desde el punto de vista de complejidad estadística para interpolar datos multivariados en celdas faltantes. Llega a la conclusión que una rejilla hexagonal no es más compleja de tratar para esos fines que una rejilla cuadrada.

En general se pueden clasificar los tipos de rejillas en regulares e irregulares, según el tipo de celdas que la conformen. Kepler (1571-1630) estudió las posibles formas de llenar el plano con polígonos iguales y regulares. Solo existen tres polígonos con esas características: triángulo isósceles, cuadrado y hexágono; para cada uno de ellos sus vértices están rodeados por 3, 4 y 6 figuras respectivamente. El caso de las rejillas irregulares es interesante cuando se considera la posibilidad de tapizar todo el plano con un número de tipos de piezas finito pero además que lo hagan de una forma no periódica. Esto último ha sido estudiado con resultados elegantes por Penrose (1989) en sus diversas teselaciones.

En resumen, la definición de OGU mediante una rejilla de cuadros de igual tamaño es una estrategia bastante utilizada en análisis biogeográficos. En conservación parece ser una buena estrategia que será más usada en el futuro a nivel internacional (Miller, 1994).

**Análisis biogeográficos basados en similitud.** Una de las vertientes para realizar análisis en biogeografía es mediante la aplicación de índices de similitud (v.g. Hubálek, 1982) entre OGU (Crovello, 1981) y su tratamiento posterior para agruparlas (v.g. Hagmeier y Stults, 1966; McLaughlin, 1989; Villarreal, 1994; Nimis *et al.*, 1995).

Existen varios problemas de decisión que aún no han sido claramente solucionados. Uno de ellos es decidir qué índice de similitud utilizar, otro qué tamaño de OGU utilizar y otro más qué método de agrupamiento elegir. Pero también existe otro problema, de naturaleza más básica, que es decidir si en realidad es posible aplicar índices de similitud en el análisis.

Así como una de las condiciones para aplicar pruebas estadísticas es tener una muestra 'representativa' del universo, entonces también para aplicar índices de similitud se deberá averiguar si las OGU están estadísticamente 'bien muestreadas'. Esto significa que sean comparables en términos estadísticos, que las diferencias en el muestreo sean pocas, de forma que las conclusiones que provengan de su análisis no sean sesgadas por esas diferencias.

El problema de la calidad de muestreo espacial en biodiversidad ha sido abordado por algunos autores (v.g. Soberón y Lorente, 1993; Colwell y Coddington, 1994; Soberón *et al.*, 2000). Murguía y Villaseñor (2000) proponen unificar el problema del análisis de la validez estadística-muestral de las OGU, con la misma definición y delimitación de ellas.

**Aplicación de índices de similitud condicionada.** La facilidad que brinda la tecnología informática en la realización de análisis cuantitativos no debe de constituirse en una falta de criterios para aplicar el método, ni de ignorancia de sus fundamentos, limitaciones y posibilidades teóricas. Los índices de similitud son un parámetro útil en biogeografía para comparar la biodiversidad y su composición entre dos unidades geográficas. Sin embargo, no debe de olvidarse el hecho que los datos de riqueza y composición biótica de las regiones son estimaciones de datos parciales (ver Papaverio y Teixeira en este volumen). Lo que el investigador obtiene de los estudios de campo, bibliográficos y de colecciones es una aproximación; lo importante es conocer los límites de esa aproximación en términos estadísticos, y sus restricciones en cuanto a la capacidad de soportar deducciones o generalizaciones.

De lo anterior se sigue que una de las condiciones mínimas para aplicar índices de similitud es hacerlo cuando las unidades geográficas estén estadísticamente bien muestreadas, o bien, realizar los cálculos, pero entonces asociar el error de las estimaciones.

El problema de la satisfacción de restricciones de muestreo para aplicar índices de similitud también puede resolverse modificando a las mismas unidades de muestreo. Es decir, agrupando unidades de muestreo para obtener menor número de ellas pero mejor representadas.

**Elección de OGU estadísticamente bien muestreadas.** Uno de los problemas en análisis biogeográfico es cómo minimizar las hipótesis *a priori* sobre los patrones a descubrir, esto es cómo ser lo más neutral o imparcial en la división inicial del área a estudiar. Por ejemplo, cuando se comienza una investigación de patrones biogeográficos en BC, la elección de los límites de área de estudio ya lleva implícito cuando menos un supuesto: el que el patrón pueda ser descubierto limitando el estudio sólo a esa área; ese supuesto puede verse como una hipótesis *a priori*.

Birks (1987) propone como una de las primeras etapas en análisis biogeográfico, la definición de OGU y la escala. Considera que las OGU deben seleccionarse bajo criterios únicamente geográficos y no incluir factores biológicos o ecológicos, para evitar razonamientos circulares. También se ha hecho notar la importancia de definir el nivel taxonómico

del análisis considerando las características de los datos, sus consecuencias y limitaciones.

Ya se ha indicado que algunos parámetros en biodiversidad son estimaciones a partir de muestras, y que es importante conocer los límites de esa aproximación en términos estadísticos. Para ilustrar esta afirmación, se considerará la diferencia entre 'dato real' e 'información completa' mediante el ejemplo de la generalización de puntos hacia áreas en la estimación de áreas de distribución (ver figura 1), o sea de las 'áreas' ocupadas a las áreas de distribución. Un dato real pero que proporciona información parcial son las localidades de recolecta, mientras que la generalización en áreas de distribución a partir de esos puntos es un modelo que estima información completa.

Murguía y Villaseñor (2000) desarrollan un índice de 'Calidad de Muestreo' (Qs) para establecer la calidad de los registros biogeográficos, toman en consideración el hecho que los registros de recolectas son muestras y por lo tanto representan estimaciones subvaluadas de la diversidad de una región.

### Herramientas de la biogeografía cuantitativa

Muchos de los métodos de la BC son susceptibles de aplicarse mediante 'herramientas'. En general, las principales herramientas informáticas de la biogeografía cuantitativa las brindan sistemas del tipo:

- Manejadores de Bases de Datos (DBMS por las iniciales del término en inglés Data Bases Management Systems)
- Paquetes estadísticos
- Paquetes de Taxonomía Numérica
- Sistemas de Información Geográfica (GIS)

El biogeógrafo que aplica técnicas de la BC de modo automatizado se ve obligado a utilizar varios tipos de herramientas comunicándolas con importaciones/exportaciones. Utiliza paquetes de taxonomía numérica o de análisis estadístico (no espacial) y luego vierte sus resultados en un *software* de cartografía, AM/FM o GIS para desplegar los resultados en un mapa. Actualmente no existe una herramienta informática que integre los requerimientos para aplicar las técnicas de la BC; aunque si existen algunos desarrollos para visualización de la biodiversidad (ver por ejemplo Schalk y Troost, 1999) éstos no incluyen técnicas de análisis de la BC.

Las herramientas de los Sistemas de Información Geográfica (GIS) permiten una manipulación de la Información georreferida, contenida en bases de datos biológicos, para realizar análisis espacial. Asimismo permiten la 'cuantificación' e implementan muchos de los métodos utilizados en la biogeografía cuantitativa, pero también conducen el paso de lo cuantitativo a lo cualitativo, es decir el juego entre el análisis

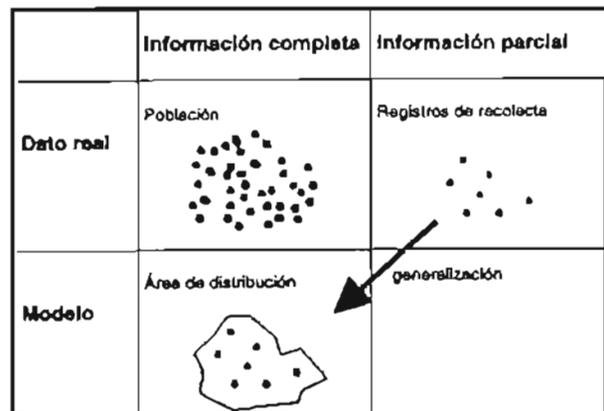


Fig. 1. Ejemplificación de la diferencia entre los conceptos de 'dato real' e 'información completa'.

sis y la síntesis. Con un GIS se pueden calcular, por ejemplo, determinados índices espaciales, como la densidad de población, pero también se pueden pintar de diferentes colores las áreas analizadas, para con ello decidir si un área está o no densamente poblada. Sin embargo, si la calidad de la información fuente es baja, la diferencia o el error puede ser un color rojo en lugar de verde, o una decisión de un 'sí' a una de un 'no'. Por ejemplo '...sí, hay que decretarla como área protegida', 'Sí, hay que invertir más en esfuerzo de recolecta', 'no, no se dará apoyo a la investigación de la biota de X región', u otras decisiones.

Desafortunadamente, aunque el campo de desarrollo de los GIS está creciendo de manera acelerada, la incorporación de las técnicas de la BC a ellos aún es muy pobre. Existen pocos desarrollos de GIS que consideren técnicas específicamente diseñadas para este campo, como el análisis de agrupamiento o la definición de áreas prioritarias para la conservación. La aplicación de los GIS a la conservación ha sido tomando las técnicas que los GIS comerciales (*v.gr.* ArcInfo, SPANS, Intergraph) implementan y que en general provienen de otros campos de aplicación.

### Antecedentes y fundamentos de la biogeografía cuantitativa

Las técnicas de la biogeografía cuantitativa de las que aquí se discute su fundamento teórico son un subconjunto de aquellas utilizadas por la biogeografía ecológica y, para precisar, son aquellas documentadas en Murguía y Rojas (2001). Por lo tanto, en esta discusión no se considera parte de la biogeografía cuantitativa a los métodos de la biogeografía histórica, como lo son el método panbiogeográfico, los métodos de la biogeografía de la vicarianza o la biogeografía de la dispersión (centros de origen); tampoco se consideran otros enfoques ecológicos, como la 'biogeografía de islas'.

La herencia de la taxonomía numérica. En general los métodos de la biogeografía cuantitativa (*sensu* Murguía y Rojas, 2001) se han heredado de la taxonomía numérica (*v.gr.* Hagmeier y Snruls, 1964; Birks y Deacon, 1973; Birks, 1976; Crovello, 1981), es decir, de los métodos expuestos por Sneath y Sokal (1963). Esta 'herencia' directa ha ocasionado una falta de sustento teórico, esto es, el sustento del uso de los 'métodos cuantitativos' en la disciplina biogeográfica. Por ejemplo, ha sido pobre o escaso el establecimiento de relaciones entre la manera de operar de los métodos multivariados y los procesos subyacentes que se desean analizar o entender, entre la aplicación de los índices de similitud y el significado de similitud biogeográfica, entre el establecimiento de las Unidades Geográficas Operativas y el significado de dividir un área ocupada por poblaciones.

La taxonomía numérica, como método de la escuela 'fenética' de la taxonomía, tiene como principio metodológico la similitud global de los caracteres morfológicos, dando un peso similar a cada uno de ellos, es decir cada carácter tiene una importancia igual a los demás. Desde el punto de vista taxonómico esta escuela ha sido bastante criticada, principalmente debido a que: *...las estimaciones basadas en similitud global mezclan caracteres que se aplican a diferentes niveles de universalidad del problema examinado (Wiley, 1981). Otro problema consiste en que es posible obtener diferentes agrupamientos a partir de una misma matriz de datos si se emplean diferentes algoritmos, por lo que no se logra la objetividad y estabilidad buscadas por los feneticistas (Scotland, 1992). [...] En la actualidad la taxonomía fenética ha sido abandonada casi completamente, salvo por su uso para resolver problemas de microtaxonomía (Siebert, 1992)... (Morrone, 2001: 18).*

Para ilustrar la carencia de sustento teórico en la utilización de las técnicas de la taxonomía numérica en biogeografía, a continuación se ejemplifica un caso sobre el tipo de justificaciones del empleo de métodos de agrupamiento. La justificación del uso de métodos multivariados para el análisis biogeográfico por la necesidad de analizar simultáneamente a todos los taxones en todas las regiones de interés, como así lo

indica Birks (1976), va acompañada de una ausencia de sustento teórico del uso de los métodos en la disciplina, es decir, establecer una relación entre la manera de operar de los métodos multivariados y los procesos subyacentes que se desean analizar o entender.

Birks y Deacon (1973) y Birks (1976) indican que los métodos de agrupamiento implican el presupuesto que las unidades a clasificar tienen una arreglo jerárquico, cosa que no siempre sucede en la realidad, razón por la que en sus análisis de clasificación biogeográfica prefieren el uso de métodos de ordenación. El argumento de Birks y Deacon para preferir métodos de ordenación a los de agrupamiento merece atención: cabe mencionar, por ejemplo, que McLaughlin (1992) prestó atención al problema, concluyendo que la jerarquización es una manera natural de clasificar a las relaciones florísticas. El cuestionar el arreglo jerárquico de una regionalización con base en datos de organismos vivos es participar en una discusión constructiva sobre la utilidad de los métodos al entendimiento y descubrimiento de los procesos biogeográficos. Sin embargo, el preferir un método por que el análogo es malo, no es una justificación bien sustentada (*v.gr.* métodos de ordenación sobre los de agrupamiento en Birks y Deacon, 1973).

Grupo I y Grupo II. Se puede decir que los métodos de la biogeografía cuantitativa heredados directamente de la taxonomía numérica (a los que se les denominará como 'grupo I') han servido como una herramienta del 'análisis exploratorio'. Desde este punto de vista, el método es un conjunto de técnicas desarticuladas, que deben de justificarse y 'reinterpretarse' en cada caso de aplicación, ya que no subyace ningún fundamento biogeográfico. Por el contrario, varias de las técnicas 'no-heredadas' de la taxonomía numérica, diseñadas específicamente para entender o analizar procesos biogeográficos ('grupo II'), no son técnicas únicamente exploratorias, sino que brindan descriptivos biogeográficos o elementos de decisión al biogeógrafo. Tal es el caso de índices como la diversidad alfa, la diversidad beta, los métodos para definir áreas prioritarias en la conservación y los índices de endemismo; hay que anotar que para este 'grupo II' de técnicas, el hecho que hayan sido creadas específicamente para el análisis de procesos biogeográficos no justifica *a priori* su aplicación, algunas de ellas pueden o no tener un sustento teórico biogeográfico.

Método formal no fundamentado. El trabajo de Hagmeier y Snruls (1964) es uno de los pioneros en proponer la aplicación a la biogeografía de las técnicas numéricas presentadas por Sokal y Sneath (1963), alertando varios de los problemas que deben de resolverse. Los trabajos de Birks (1976, 1987) y Crovello (1981) exponen a las técnicas de la biogeografía cuantitativa como un método coherente y útil en la generación y apoyo de hipótesis biogeográficas. En general, en este tipo de trabajo se observa una estructura 'cuantitativa' en apariencia congruente y claramente ligada de un paso a otro, sin embargo, los argumentos más fuertes para usar esas técnicas es que son 'cuantitativas' y 'objetivas'. Preferir un método cuantitativo a otro no-cuantitativo debe argumentarse; el carácter 'cuantitativo' por sí solo no brinda objetividad.

Birks (1976), en su figura 1, propone los estados o etapas de los estudios de 'biogeografía descriptiva', entre los que incluye: (1) la definición de los alcances del estudio; (2) la definición de los taxones y áreas de interés; (3) la compilación de datos; (4) la construcción de matrices de datos; (5) el cálculo de las medidas de (dis)similitud; (6) estructuración de datos (*v.gr.* la clasificación), y (7) el establecimiento de generalizaciones (taxones en áreas). A primera vista parece un método ordenado, ya que se da una secuencia de tareas con las que un iniciado en la ciencia diría que eso sí es un método científico por el hecho de estar ordenado y sistematizado; desafortunadamente, un método debe tener relación con los procesos que intenta descubrir o entender. En nuestro caso cada paso de la secuencia (1) a (7) debe justificarse y relacionarse con procesos biogeográficos.

En un trabajo posterior al de 1976, Birks (1987) expone (nuevamente con base en otra figura 1) un buen resumen de la metodología, pero

marcando su carácter 'exploratorio' y alerta sobre los riesgos del método, dando en casi todos los casos soluciones 'técnicas', pero no desde el punto de vista biogeográfico, es decir, del significado biogeográfico en el método. Es de notar que en la sección de 'métodos de agrupamiento' indica que el problema de elegir el mejor método es una cuestión aún sin solución, y nuevamente brinda un consejo estadístico (utilizar varios y comparar resultados), pero no en términos teórico-biogeográficos.

Crovello (1981) resume el método de la biogeografía cuantitativa en lo que llama un *Multistage Decision Process*, que básicamente es una lista de técnicas alternativas para el análisis biogeográfico.

Así, podríamos simplificar el cuadro diciendo que las técnicas documentadas en el trabajo de Sokal y Sneath (1963) y Sneath y Sokal (1973), para la taxonomía numérica, han sido trasladadas de manera automática a la biogeografía (conformando el grupo I de técnicas de la biogeografía cuantitativa). Entonces, la biogeografía cuantitativa contiene un subconjunto grande de técnicas exploratorias carentes de una consistencia teórico-biogeográfica, a lo que se podría identificar como un método formal pero no fundamentado, desde el punto de vista de que no está basado en conceptos biogeográficos.

Algunas consecuencias del método formal no fundamentado de la biogeografía cuantitativa son que las técnicas que conforman al método: (1) han sido utilizadas como herramientas del 'análisis exploratorio' en biogeografía, al permitir el análisis de información aprovechando los beneficios de los procesos automáticos y sistemáticos de los programas de computadora; (2) han conformado una metodología para el análisis de la biogeografía ecológica, en algunos casos sin un sustento teórico; (3) han facilitado la incorporación de la tecnología de los Sistemas de Información Geográfica al análisis biogeográfico; (4) han creado en la comunidad biogeográfica (ecológica) un 'sabor cuantitativo' del análisis biogeográfico; y (5) han mostrado su utilidad al aplicarse de manera extensa a conjuntos de datos empíricos ayudando en la construcción de conocimiento biogeográfico.

Alternativas para la fundamentación del método de la biogeografía cuantitativa. Desde el punto de vista de las técnicas del grupo I, debe decidirse entre: (1) seguir las utilizando únicamente como técnicas exploratorias, (2) fundamentarlas y establecer los límites de aplicación y su significado biogeográfico, y (3) desecharlas definitivamente, remplazándolas por técnicas del grupo II, existentes o nuevas, creadas *ad hoc* para la biogeografía ecológica.

Para la alternativa (2), un ejemplo en índices de similitud es el trabajo de Sánchez y López (1988), en el que las premisas para definir un 'buen índice de similitud' son biogeográficas. Aun antes de iniciar una búsqueda (entre lo existente o bien entre nuevas configuraciones de técnicas), debe definirse el 'diseño algebraico-lógico' de los 'cuantificadores' (cuantificador desde el punto de vista de que son técnicas numéricas, cuya intención es 'medir'), un ejemplo es el trabajo de Hubálek (1982), también para índices de similitud en el que las premisas para ese 'diseño' son de tipo algebraico.

La construcción de modelos nulos (*v.g.* Connor y Simberloff, 1979; Gottelli y Graves, 1996) para diferenciar entre procesos generados por causas biogeográficas de aquellos ocasionados por el azar ha sido sugerida por Birks (1987). Los modelos nulos son una alternativa para fortalecer a las técnicas de la biogeografía cuantitativa y llevarlas más allá del carácter 'exploratorio'.

### Biogeografía cuantitativa y panbiogeografía

El 'nodo' en panbiogeografía es un concepto o constructo (*sensu* Bunge, 2001) basado en el del trazo: el nodo es el punto de intersección o confluencia de dos o más trazos generalizados. El nodo en panbiogeografía puede tener varios significados biológicos (Craw *et al.*, 1999):

- . Presencia local (endemismo)
- . Ausencia local (ausencia de taxones con distribución amplia)

- . Afinidades geográficas y filogenéticas
- . Límites geográficos de taxones
- . Límites biogenéticos de taxones

Espinosa y Llorente (1993) indicaron que el nodo es quizás uno de los aportes más importantes de la panbiogeografía: los nodos son un buen recurso para especular sobre la existencia de las áreas de convergencia tectónica, además de que son un elemento primario para establecer clasificaciones biogeográficas naturales.

La secuencia localida--> trazo--> trazo generalizado--> nodo ilustra el hecho que el nodo es un constructo basado en la localida.

Lo que se puede anotar en este punto, en relación con el concepto de OGU en biogeografía cuantitativa, es que a la OGU se le da un tratamiento, en la mayoría de los casos, individual y previo al relacionarlo con otras OGUs. Por ejemplo, se pueden clasificar las OGUs por su diversidad alfa; la diversidad alfa es un atributo originado del análisis 'biogeográfico cuantitativo' previo a una consideración de relación entre las OGUs. En panbiogeografía, la filosofía es analizar al espacio-tiempo-forma de una manera integral, y es por eso que no se necesitan construir 'OGUs panbiogeográficas', y el trazo se construye directamente a partir de las localidades. Espinosa y Llorente (1993) y Craw *et al.* (1999) atribuyen importancia al nodo por ser un elemento primario para establecer clasificaciones biogeográficas naturales y porque en ciertos casos son 'hot spots' de la biodiversidad en un contexto biogeográfico. Así, el nodo se clasifica (por ejemplo, por el número de trazos que confluyen o por el número de taxones en él) pero las localidades no, solo después de haber sido consideradas nodos, construidos estos últimos vía los trazos, es decir construidos mediante relaciones espaciales.

### Escalas y jerarquías en el análisis de patrones biogeográficos

Jerarquías y escalas. Los sistemas naturales exhiben una organización jerárquica con patrones anidados y procesos que ocurren en un amplio intervalo de escalas espacio-temporales, taxonómicas y bióticas (Miller, 1994). Por lo tanto, se debe considerar la escala al realizar análisis sobre biodiversidad. Por ejemplo, a una determinada escala, un proceso puede parecer homogéneo, mientras que en otras es claramente heterogéneo. La escala del mapa puede alterar la percepción de continuidad-discontinuidad de las distribuciones. No se debe olvidar que un mapa es un gráfico de la distribución, una representación.

La palabra escala se utiliza para referirse a determinados aspectos del espacio y del tiempo, por ejemplo, tamaño del área o longitud del tiempo (Turner y Gardner, 1991), pero también se puede referir a distintos niveles taxonómicos. La teoría de jerarquías (Salthe, 1985) estudia las interacciones de los fenómenos entre diferentes escalas y está orientada hacia las 'entidades'; es decir, los límites de las entidades individuales son importantes. Se entiende por entidad a una representación teórica de una cosa como un individuo. En algunos casos se asume que la variación espacial y la temporal están correlacionadas; sin embargo, esto no es una regla general. Por ejemplo, el impacto del hombre sobre la naturaleza a través de periodos de tiempo muy cortos (Hengeveld, 1990).

La jerarquía es una estructura que se caracteriza por tener un nivel de organización ocupada por entidades de diferentes escalas. La escala es una gradación relativa basada en el tamaño y ámbito espacial de influencia, en ocasiones reflejada en la duración del tiempo del fenómeno (Salthe, 1985).

Jerarquía en biogeografía. La región biótica o bioma es un nivel de organización en la jerarquía ecológica que surge de la interacción de ecosistemas locales. Si consideramos -o suponemos- la existencia de varios 'niveles' en la estructura del ambiente natural, éstos podrían ser, en orden creciente: organismo, población, ecosistema, bioma, superficie del planeta (superficie terrestre concebida como biosfera), esto en una jerarquía ecológica. Así, el bioma se ubica dentro de esa jerarquía por

debajo del nivel de superficie de la tierra y por arriba del nivel de ecosistema (Saithe, 1985). En términos de clasificaciones naturales, esto es históricas, la jerarquía puede ser: distritos, subprovincias, provincias, dominios, subregiones, regiones y reinos.

En biogeografía histórica, la jerarquía expresa relaciones ancestrales, vicisitudes históricas comparadas, mientras que en los métodos fenéticos, como el análisis de agrupamiento, la jerarquía de los fenogramas simplemente son un artefacto del método, pero no expresa al fenómeno natural. La jerarquía en un fenograma es útil cuando se decide una línea de 'corte crítico', que divide transversalmente al fenograma y que define las categorías (agrupaciones biogeográficas, bióticas o taxones, según sea el caso).

Las principales diferencias entre la biogeografía cladística y la panbiogeografía es el apego a una jerarquía; la cladística postula el apego estricto a un arreglo jerárquico para expresar las interrelaciones históricas (Espinosa y Llorente, 1993), usa como dato básico a las genealogías que hay entre las especies endémicas. La jerarquía de un cladograma determina un orden de cambios y relaciones en el tiempo. La panbiogeografía, en su propuesta original, no utiliza el arreglo jerárquico de los organismos en sus métodos, solo ubica trazos, de longitud mínima, entre los sitios geográficos de dos grupos. Esto es, presenta relaciones espaciales que pueden servir, incluso, para mostrar o elucidar relaciones genealógicas entre biotas endémicas.

La escala biogeográfica. Las inferencias estadísticas son sensibles a la escala espacial. Cuando se realizan regresiones a partir de datos espaciales es muy importante considerar la escala, ya que los parámetros de los modelos así obtenidos no serán comparables (Wong, 1996). Se han intentado plantear varias estrategias estadísticas para detectar la escala adecuada para estudiar los patrones de distribución de los organismos (ver por ejemplo Turner *et al.*, 1991). De hecho, el problema del 'tamaño del cuadro' en análisis biogeográficos es un problema de escalas, al cual también se le han propuesto soluciones de tipo estadístico (*vg.* Phipps, 1975). Anderson y Marcus (1993) muestran que el tamaño de cuadro afecta significativamente la estimación de la densidad de especies y opinan que no hay un tamaño de cuadro que de manera intrínseca sea mejor que otros.

Varias medidas relacionadas con la biodiversidad son sensibles al tamaño de la muestra, y esto hace que también puedan serlo al cambiar de escala espacial de análisis. Por ejemplo, casi todos los índices de similitud son sensibles al tamaño de la muestra (Wolda, 1981), lo que además genera un problema de compatibilidad entre escalas al utilizar este tipo de herramientas. También los índices de rareza son sensibles al tamaño de la muestra (Gotelli y Graves, 1997), razón por la cual debe tenerse precaución al manejarlos a diferentes escalas espaciales. Las mediciones de una misma variable a diferentes escalas puede llevar a conjeturas contrarias. Por ejemplo, puede deducirse que los topos de los alfalfares son una plaga que afecta a los cultivos cuando se analizan a escala individual (que es la idea preponderante en la literatura sobre agricultura), puesto que la observación directa es que los topos se comen la alfalfa; pero al analizar el problema a una escala mayor, al nivel de población, se puede observar que son benéficos e incrementan la productividad, ya que modifican algunas condiciones en el suelo (lo aran y lo airean) que producen beneficios al cultivo (Geng, 1997).

Otros aspectos relevantes sobre las escalas en biogeografía son que en las bases de datos biogeográficas debe considerarse que: (1) las proporciones de los tipos de cobertura de una región dependen de la resolución espacial, y (2) la distribución y magnitud de las pendientes dependen de la densidad de las medidas de elevación altitudinal (Miller, 1994).

Se han propuesto varias técnicas para distinguir las diferentes escalas de variación; sin embargo, cuando no se puedan distinguir las escalas de variación puede utilizarse a la 'dimensión fractal' para estimar la variabilidad total entre las escalas, ya sean espaciales o temporales (Hengeveld, 1990). La 'dimensión fractal' (D) es un parámetro de la relación

entre la escala y el intervalo de muestreo, valores bajos de D indican alta homogeneidad del fenómeno al variar la escala, mientras que valores altos indican que el proceso es difuso al cambiar la escala.

El problema de la unidad de área modificable. El problema de la unidad de área modificable (MAUP por sus siglas en inglés) se refiere a la inconsistencia de los resultados de los análisis de un mismo conjunto de datos agrupados espacialmente de manera diferente (Wong, 1996). A un nivel de agregación de las unidades geográficas originales de datos pueden observarse algunas correlaciones, mientras que a otras no. El MAUP consiste de dos subproblemas: el efecto de escala y el de zonificación (Wong, 1996). El primero se refiere a la inconsistencia de los resultados estadísticos del análisis de los mismos datos a diferentes niveles de agregación, por ejemplo raster a diferente resolución. El efecto de zonificación se refiere a inconsistencias al agregar los datos a resoluciones similares pero en diferente forma, por ejemplo, un raster de cuadros contra un raster de hexágonos.

Cuando las unidades adyacentes se unen para formar una unidad mayor, en la que el atributo se calcula como el promedio, se disminuyen las diferencias y entonces la varianza para el área total decrece. Además, como la similitud entre las unidades de área no es homogénea en toda la región y en todas las direcciones, la agregación de forma diferente hace que la varianza también se afecte.

Jerarquía: Identificación del nivel o relación entre escalas. Al estudiar la interacción entre dos procesos, por ejemplo, distribución espacial y la evolución, se deben identificar las escalas a la cual ambos procesos interactúan; en este ejemplo, ambos procesos pueden ser el mismo: espacio-tiempo-forma inextricablemente unidos. Pero aun en un solo proceso, observado a diferentes escalas, puede ser complejo establecer la relación entre escalas adyacentes; por ejemplo, la relación de la distribución espacial de entidades a dos escalas diferentes.

Algunos autores opinan que el efecto de la escala en los análisis es predecible, mientras que otros opinan que el problema es de alta complejidad y proponen como alternativa el uso (y desarrollo) de métodos de análisis que sean insensibles a la escala. Otro enfoque para abordar el problema es detectar la escala adecuada para analizar el fenómeno bajo estudio; por ejemplo, identificar en cuál escala se da la máxima variación del fenómeno y entonces estudiarlo a ese nivel.

Finalmente, cuando un mismo fenómeno se estudia a diferentes escalas y se comparan los resultados, es frecuente encontrar contradicciones; sin embargo, mientras no se establezca el efecto de la escala en el fenómeno, se debe ser cuidadoso al confrontar ambos tipos de resultados.

## Discusión

Intuimos que, tal vez, buena parte de las críticas hacia la biogeografía cuantitativa sean por su íntima relación con la fenética y los problemas que se generan en una concepción simplificada o errónea de la semejanza (similitud, diferencia y distinción) en entidades orgánicas y su expresión espacial, a continuación se efectúan comentarios sobre ello.

Concepto de semejanza en la sistemática fenética y en la cladística. Uno de los problemas filosóficos en taxonomía consiste en decidir qué tipo de información debe utilizarse como base confiable de la clasificación biológica: las similitudes por sí solas o las inferencias sobre las relaciones filogenéticas (Llorente, 1986). El sistemático se enfrenta al problema de decidir qué significado evolutivo tienen las similitudes observadas o reconocidas.

Para el cladismo, las homólogas, los grupos monofiléticos y las relaciones filogenéticas están fuera del dominio empírico, es decir, no pueden observarse mediante la experiencia directa, pero de acuerdo con la teoría evolutiva son reconocidas y entonces pueden reconstruirse a

partir del análisis de las similitudes heredables (De Luna, 1995). Para los fenéticos solo la evidencia empírica es aceptable en la construcción del conocimiento (Sokal y Sneath, 1973).

La fenética o taxonomía numérica es un sistema de clasificación basado en la similitud total de los organismos (Sokal, 1986, en Kohlmann, 1994). Según Sokal y Sneath (1973), la similitud total entre dos entidades es una función de las similitudes individuales de cada uno de los caracteres con los que se comparan: *a priori* cada carácter tiene el mismo peso y las clasificaciones se basan en similitudes fenéticas.

Para los cladistas, quienes buscan las relaciones genealógicas, solamente los caracteres derivados compartidos o sinapomorfías proporcionan la prueba de un origen común, es decir, de monofilia (Llorente, 1986; Villaseñor y Dávila, 1992). Para el feneticismo, solo las similitudes tienen existencia real, la filogenia no puede investigarse con certidumbre; únicamente cabe estudiar la semejanza porque ésta puede descubrirse y verificarse de manera empírica (De Luna, 1995). Por esa razón ontológica y metodológica es que la clasificación fenética se basa en la similitud total de los taxones. Los fenéticos no se preguntan *a priori* sobre el porqué de la semejanza.

La taxonomía numérica también reconoce la importancia que tiene el identificar las 'homologías' en la construcción de clasificaciones, ya que la asignación de los estados de carácter a los caracteres implica el reconocimiento de dichas homologías. Se argumenta que la taxonomía numérica opera adecuadamente generando clasificaciones robustas cuando se equivoca la 'homologación' de caracteres en menos del 10% (Sokal y Sneath, 1973). La taxonomía fenética utiliza la 'homología operacional' para resolver ese problema: dos estados de carácter son el mismo si éstos son indistinguibles. En cambio, para Hennig los caracteres homólogos son aquellos que son observados como estados transformados del mismo carácter original. Feneticismo y filogeneticismo utilizan la misma palabra: homología, pero le dan connotaciones distintas.

La 'similitud total' de la fenética a menudo se mide a través de fórmulas de índices de similitud; sin embargo, quedan varios problemas por resolver a ese respecto (Baroni-Urbani y Buser, 1976). De hecho, uno de los problemas principales que tiene la taxonomía numérica es que la estimación de la similitud fenética depende del coeficiente de similitud empleado (Sokal y Sneath, 1973). Resultados artefactuales provenientes de la manipulación cuantitativa llevan a la fenética a decisiones y conclusiones lejanas a las preguntas de relación que se quieren resolver en la sistemática. Algunos autores han estudiado el comportamiento de los índices de similitud, por lo común bajo determinadas premisas algebraicas, sobre las cuales se establece el buen o mal comportamiento de las fórmulas (Hubálek, 1982; Sánchez y López, 1989; Hengeveld, 1990; Hayek, 1993).

Conceptos de semejanza en biogeografía. Los enfoques contemporáneos principales de la biogeografía se pueden clasificar en históricos y ecológicos. Dentro de los enfoques históricos se encuentra una abigarrada amalgama de escuelas del dispersionismo, la panbiogeografía y la biogeografía de la vicarianza o cladística. Dentro de los ecológicos se encuentran las escuelas de la biogeografía ecológica, propiamente, y la del equilibrio insular o biogeografía de islas, entre otras que han aparecido, prosperado y que decayeron durante el siglo XX.

El objetivo de la biogeografía histórica es mostrar y explicar la relación de biotas por su conexión histórica, entendiendo por 'relación histórica' al hecho de compartir descendientes del mismo ancestro (Andersson, 1996). La biogeografía de la vicarianza o cladística compara cladogramas de las áreas y de los grupos monofiléticos de esas regiones para reconocer qué tanta unidad histórica se manifiesta (Rosen, 1978). La panbiogeografía utiliza los 'trazos generalizados' para guiarse en la búsqueda de explicaciones de la distribución espacial de las biotas actuales, pero no condiciona su método a contar con las genealogías de los organismos, como si lo hace el cladismo. Entonces, la 'cercanía' (similitud) de las biotas, en términos de trazos generalizados, dará la explicación histórica de su distribución. En los trazos generalizados subyace

una significancia estadística (Croizat *et al.*, 1974), pues un trazo generalizado es un grupo de trazos individuales que coinciden aproximadamente; entendiendo que un trazo individual es una línea que conecta en el espacio a grupos monofiléticos (Simberloff, 1983). Por lo anterior, se podría decir que en los trazos generalizados subyace una semejanza estadística que se debe examinar con el mayor rigor posible.

Sin embargo, tanto para la biogeografía de la vicarianza como para la panbiogeografía, la correspondencia entre las relaciones genealógicas de las áreas y las especies implica homología topológica y evolutiva, topológica porque dos subconjuntos o subáreas guardan la misma posición con respecto al todo; evolutiva porque implica una misma historia de formación con un mismo origen o causa (Espínosa y Llorente, 1993). En términos geográficos es una homología espacial (Morrone, 2001). La biogeografía del dispersionismo utiliza como criterio de semejanza la ocurrencia del mismo grupo biológico actual en dos regiones distintas y esto puede ser suficiente para generar una hipótesis histórica, cuya explicación puede ser dispersión o vicarianza, pero más a menudo historias particulares de dispersión.

Con frecuencia la biogeografía ecológica estudia la distribución individual de cada especie a escalas locales y supone que las causas de los patrones de distribución operan en el corto plazo. La biogeografía histórica estudia los patrones de distribución de taxones a escala global y supone que los factores que los producen son de naturaleza histórica y actúan en tiempo evolutivo. Mientras que para las áreas de distribución de la biogeografía ecológica tiene sentido hablar de forma, tamaño y tendencias de deformación, para las áreas de endemismo sólo tiene sentido hablar de homología biogeográfica (Espínosa y Llorente, 1993).

Uno de los argumentos en contra de la biogeografía ecológica y a favor de la cladística –y que argumenta diferencias en el concepto de semejanza– es el de Nelson y Platnick (1984) acerca de la explicación de la distribución de organismos endémicos: (1) los organismos se transportan a ciertas áreas donde no son endémicos y sobreviven y prosperan, y (2) generalmente, las áreas muy separadas albergan diferentes taxones, no importando que tan similares sean en clima y ecología. Es decir, para la biogeografía cladística (y los hechos de distribución global), la similitud de condiciones climáticas y ecológicas no implica la similitud de biotas en términos de taxones, un hecho probado desde Buffon y sobre el cual descansa la biogeografía contemporánea dominante (Llorente *et al.*, 2001).

## Agradecimientos

El primer autor desea agradecer a José Luis Villaseñor las diversas discusiones que mantuvieron sobre el tema que enriquecieron mis puntos de vista útiles en la maduración de las ideas aquí presentadas; a él, a Luis Enrique Egúlarle deseo agradecer la revisión de versiones preliminares de este documento. El segundo autor ha sostenido numerosas discusiones teóricas con David Espínosa, Jorge Soberón, Nelson Papavero, y Alfredo Bueno que en cierto modo aquí cristalizan; ambos autores agradecemos especialmente a Juan J. Morrone por sus precisiones y recomendaciones. Los proyectos 32002 y 36488 de CONACYT fueron de gran apoyo.

## Referencias

- Anderson, S. y L. F. Marcus. 1993. Effect of quadrat size on measurement of species density. *J. Biogeogr.*, 20: 421-428.
- Andersson, L. 1996. An ontological dilemma: Epistemology and methodology of historical biogeography. *J. Biogeogr.*, 23: 269-277.
- Bailey, T. C. 1994. A review of statistical spatial analysis in GIS. En: Fotheringham, S. y P. Rogerson (eds.), *Spatial Analysis and GIS*, Taylor & Francis.
- Baroni-Urbani, C. y M. W. Buser. 1976. Similarity of binary data. *Syst. Zool.*, 25: 204-210.
- Birks, H. J. B. 1976. The distribution of european pteridophytes: A numerical analysis. *New Phytol.*, 77: 257-287.

- Birks, H. J. B. 1987. Recent methodological development in descriptive biogeography. *Ann. Zool. Fennici*, 24: 165-178.
- Birks, H. J. B. y J. Deacon. 1973. A numerical analysis of the past and present flora of the British Isles. *New Phytol.*, 72: 877-902.
- Bunge, M. 2001. *Diccionario de Filosofía. Siglo XXI editores*. México, D.F.
- Burrough, P. A. 1986. *Principles of Geographical Information Systems for Land Resources Assessment*. Clarendon Press, Oxford.
- Burrough, P. A., R. McDonnell y, R. A. McDonnell. 1998. *Principles of Geographical Information Systems: Spatial Information Systems and Geostatistics*. Clarendon Press, Oxford.
- Castro, I., J. C. Moreno, C. J. Humphries y P. H. Williams. 1996. Strengthening the Natural and National Park system of Iberia to conserve vascular plants. *Bot. J. Linn. Soc.*, 121: 189-206.
- Codd, E. F. 1990. *The Relational Model for Database Management*. Addison-Wesley Pub.
- Colwell, R. K. y J. A. Coddington. 1994. Estimating terrestrial biodiversity through extrapolation. *Phil. Trans. Roy. Soc. London B*, 345: 101-118.
- Connor, E. F. y D. Simberloff. 1979. The assembly of species communities: Change or competition? *Ecology*, 60: 1132-1140.
- Craw, R., J. Grehan y M. Heads. 1999. *Panbiogeography, tracking the history of life*. Oxford Univ. Press, Oxford y Nueva York.
- Croizat, L., G. Nelson y D. Rosen. 1974. Centers of origin and related concepts. *Syst. Zool.*, 23(2): 265-287.
- Crovello, T. J. 1981. Quantitative biogeography: An overview. *Taxon*, 30(3): 563-575.
- De Luna, E. 1995. Bases filosóficas de los análisis cladísticos para la investigación taxonómica. *Acta Bot. Mex.*, 33: 63-79.
- Espinosa, D. y J. Llorente. 1993. *Fundamentos de biogeografía filogenética*. UNAM, México, D.F.
- Ezcurra, E., M. Equihua, B. Kohlmann y S. Sánchez-Colón. 1984. *Métodos cuantitativos en la biogeografía*. Instituto de Ecología, México, D.F.
- Geng, S. 1997. Multi-scale influences of gophers on alfalfa yield and quality. *Field Crops Res.*, 49: 159-168.
- Gotelli, N. J. y G. R. Graves. 1996. *Null models in ecology*. Smithsonian Institution Press, Londres.
- Griffith, D. A. 1993. Advanced spatial statistics for analyzing and visualizing geo-referenced data. *Int. J. Geogr. Inf. Syst.*, 7(2): 107-123.
- Haining, R. 1994. Designing spatial data analysis modules for geographical information systems. En: Fotheringham, S. y P. Rogerson (eds.), *Spatial Analysis and GIS*, Taylor & Francis, Londres.
- Hagmeier, E.M. y C.D. Stults. 1964. A numerical analysis of the distributional patterns of North America mammals. *Syst. Zool.*, 13: 125-155.
- Hengeveld, R. 1990. *Dynamic biogeography*. Cambridge University Press, Cambridge.
- Hubálek, Z. 1982. Coefficients of association and similarity, based on binary (presence-absence) data: An evaluation. *Biol. Rev.*, 57: 669-689.
- Kohlmann, B. 1994. Algunos aspectos de la taxonomía numérica y sus usos en México. En: Llorente, J. e I. Luna (comps.), *Taxonomía biológica*, UNAM-Fondo de Cultura Económica, México, D.F.
- Llorente, J. 1986. Conceptos en Cladismo. *Ciencias (número especial)* 1: 74-87.
- Llorente, J., N. Papavero y A. Bueno. 2001. Síntesis histórica de la biogeografía. pp. 1-14. En: J. Llorente y J. J. Morrone (eds.), *Introducción a la biogeografía en Latinoamérica: Teorías, conceptos, métodos y aplicaciones*, Fac. Ciencias, UNAM, México, D.F.
- McAllister, D. E., Schueler, F. W., Roberts, C. M. y J. P. Hawkinsiller. 1994. Mapping and GIS analysis of the global distribution of coral reef fishes on an equal-area grid, pp. 155-175. En: Miller, R. I. (ed.), *Mapping the diversity of Nature*, Chapman & Hall, Oxford.
- McLaughlin, S. P. 1992. Are floristic areas hierarchically arranged? *J. Biogeogr.*, 19: 21-32.
- Miller, R. 1994. Setting the scene. En: R. Miller (ed.), *Mapping the diversity of nature*, Chapman & Hall, Londres.
- Morrone, J. J. 2001. *El lenguaje de la cladística*. UNAM, México, D.F.
- Murgula, M. y F. Rojas. 2001. Biogeografía cuantitativa. pp. 39-47. En: Llorente, J. y J. J. Morrone (eds.), *Introducción a la biogeografía en Latinoamérica: Teorías, conceptos, métodos y aplicaciones*, UNAM, México, D.F.
- Murgula, M. y J. L. Villaseñor. 2000. Estimating the quality of the records used in quantitative biogeography with presence-absence matrices. *Ann. Bot. Fennici*, 37: 289-296.
- Nelson, G. y N. Platnick. 1984. *Biogeography*. Carolina Biology Readers, no. 119.
- Nimis, L., L. Matyshev, G. Bolognini y N. Friesen. 1995. Phytogeographic diversity of the Putorana flora (N. Siberia). *Ann. Bot. Fennici*, 32: 1-17.
- O'Kelly, M.E. 1994. Spatial Analysis and GIS. En: Fotheringham, S. y P. Rogerson (eds.), *Spatial analysis and GIS*, Taylor & Francis, Londres.
- Penrose, R. 1989. *La nueva mente del emperador*. Grijalbo Mondadori, Barcelona.
- Phipps, J. B. 1975. BestBlock: optimizing grid size in biogeographic studies. *Can. J. Bot.*, 53(15): 1447-1452.
- Rapoport, E. H. 1975. *Areografía: Estrategias geográficas de las especies*. Fondo de Cultura Económica, México, D. F.
- Rosen, D. E. 1978. Vicariant patterns and historical explanation in biogeography. *Syst. Zool.*, 27: 159-188.
- Saithe, S. N. 1985. *Evolving hierarchical systems*. Columbia University Press, Nueva York.
- Sánchez, O. y G. López. 1988. A theoretical analysis of some indices of similarity as applied to biogeography. *Folia Ent. Mex.*, 75: 119-145.
- Schalk, P. H. y D. G. Troost. 1999. Computer tools for accessing biodiversity information. *Nat. Res.*, 35(3): 31-38.
- Simberloff, D. 1983. Biogeography: The unification and maturation of a science. En: Brush, A. H. y G. H. Klark (eds.), *Perspectives in Ornithology*, Cambridge University Press, Cambridge.
- Sneath, P. H. A. y R. R. Sokal. 1973. *Principles of Numerical Taxonomy*. Freeman, San Francisco.
- Soberón, J. y J. Llorente. 1993. The use of species accumulation functions for the prediction of species richness. *Conserv. Biol.*, 7: 480-488.
- Soberón, J., J. Llorente y L. Orta. 2000. The use of specimen label databases for conservation purposes: An example using Mexican Papilionid and Pierid butterflies. *Biodiv. Conserv.*, 9: 1441-1466.
- Sokal, R. R. y P. H. A. Sneath. 1963. *Numerical Taxonomy*. W.H. Freeman, San Francisco.
- Turner, M. G. y R. H. Gardner. 1991. Quantitative methods in landscape ecology: an introduction. En: Turner, M. G. y R. H. Gardner (eds.), *Quantitative methods in landscape ecology*, Springer Verlag, Nueva York.
- Villarreal, J. A., J. Valdés y J. L. Villaseñor. 1996. Corología de las asteráceas de Coahuila, México. *Acta Bot. Mex.*, 36: 29-42.
- Villaseñor, J. L. y P. Dávila. 1992. *Breve introducción a la metodología cladística*. Facultad de Ciencias, UNAM, México, D.F.
- Wolda, H. 1981. Similarity indices, sample size and diversity. *Oecologia*, 50: 296-302.
- Wong, D. 1996. Aggregation Effects in Geo-referenced Data. En: Arlinghaus, S. L. (ed.), *Practical handbook of spatial statistics*, CRC Press, Boca Raton.

## 2. Contexto de la Biogeografía Cuantitativa

---

*En este capítulo se hace una breve reseña de la Biogeografía Histórica como complemento de la Biogeografía Ecológica. Se explica el concepto de especie y se discute la relación de la Biogeografía Cuantitativa con la conservación de la riqueza biológica del planeta.*

### 2.1 Panbiogeografía

#### *Fundamentos Teóricos de la Panbiogeografía*

La teoría panbiogeográfica tiene como fundamento la consideración de que las especies y la superficie terrestre evolucionan juntas en el tiempo; Craw y Page en 1988 (en Craw *et al.* 1999) esquematizaron este hecho con la tríada 'Espacio-Tiempo-Forma', tomada de la obra de Croizat.

Craw *et al.* (1999) argumentan que la teoría panbiogeográfica resuelve el dilema/controversia 'dispersionismo-vicariancia', al considerar que los patrones de distribución actuales son el resultado de fases alternadas entre movilismo/inmovilismo:

*Primero, en la etapa de movilidad, en ausencia de barreras y bajo condiciones favorables, los organismos y biotas expanden su área de distribución, generación tras generación y a partir de sus medios usuales de supervivencia. Después, en la etapa de inmovilidad, una vez que las áreas de distribución alcanzan los límites*

*establecidos por barreras distribucionales infranqueables, dichas áreas pueden sufrir fragmentación por el surgimiento de barreras intermedias. Esto a través del tiempo da lugar a formación de nuevas especies.* (Espinosa y Llorente, 1993).

Definitivamente, la Panbiogeografía es un enfoque biogeográfico en el que el espacio (la distribución de los organismos y la superficie de la Tierra) es considerado un carácter importante y no sólo como un elemento que se analiza *a posteriori*. Los taxones se analizan vinculados a sus localidades, incorporando al espacio desde el inicio del análisis. La Panbiogeografía considera al espacio como un elemento dinámico, no estático a lo largo del tiempo, así el trazo generalizado se construye por la coincidencia de muchos trazos pero de los trazos más cortos. Aunque el ‘árbol de tendido mínimo’ es el método más recurrido para la construcción de los trazos, se han propuesto otras alternativas, como el método de compatibilidad.

El método panbiogeográfico contiene elementos tanto deductivos como inductivos en la construcción de sus hipótesis. La incorporación de la teoría de gráficas como una formalización del método panbiogeográfico se ha visto como un atributo positivo, que fortalece a la teoría debido a que las hipótesis generadas mediante el método pueden sujetarse a pruebas estadísticas (Craw, 1989).

Aunque la Panbiogeografía no puede definirse en un párrafo, las siguientes citas se exponen como definiciones de ésta:

*“La Panbiogeografía es un método biogeográfico que se centra en el análisis espacio-temporal de los patrones de distribución de los organismos. ... Es un enfoque distinto a la biogeografía fenética que investiga las similitudes entre biotas de diferentes áreas geográficas en términos de número de taxones en común”* (Craw, 1988).

*“La Panbiogeografía es una aproximación a la biología que se centra en el papel del Espacio/Localidad en el proceso evolutivo. Establece que el entendimiento de espacio/localidad es una precondition fundamental para el*

*análisis adecuado de los procesos involucrados en el cambio evolutivo” (Craw, 1989).*

*“La síntesis espacio-tiempo-forma tal y como fue propuesta por Croizat, no implica simplemente una amplia correlación entre la historia de la biota y la historia de la tierra. Lo que propone Croizat es que ambas historias son en realidad una sola. Éste es el núcleo duro de la Panbiogeografía” (Espinosa y Llorente, 1993).*

El ‘núcleo duro’ de la Panbiogeografía de Espinosa y Llorente (1993) puede apreciarse como una de las principales conclusiones de Croizat, quien de manera firme estableció con la tríada espacio, tiempo y forma en relación con la evolución orgánica: *“la evolución orgánica es función del espacio, el tiempo y la forma. De estos tres elementos, el espacio es el que concierne particularmente a la biogeografía, aunque interactúa necesariamente con el tiempo y la forma. Por ello una comprensión satisfactoria de la evolución orgánica es inseparable de la apreciación constructiva de la biogeografía” (Llorente et al., 2000).*

### ***El Método Panbiogeográfico***

Craw *et al.* (1999) hicieron énfasis en la importancia que tiene el contexto geográfico para el entendimiento de la historia de la vida y argumentaron que el trazo (‘track’ en inglés) da forma y expresión al espacio y tiempo que interviene entre localidades disyuntas.

El método panbiogeográfico descansa en la teoría de gráficas al utilizar los conceptos de ‘gráfica’, ‘árbol’, ‘vértice’, ‘nodo’ y ‘arista’ de una manera directa. El concepto de trazo en panbiogeografía es el análogo al concepto de arista. El concepto de ‘nodo’ en panbiogeografía tiene una connotación muy peculiar y en general es análogo al de teoría de gráficas en cuanto que representan vértices de alta conectividad.

El trazo es uno de los elementos fundamentales del método panbiogeográfico. Un trazo es una línea que conecta a dos localidades de un mismo grupo. La Panbiogeografía da especial importancia a la localidad, pues parte de ellas para construir sus conceptos. Un conjunto de trazos que se superponen, se consideran un ‘trazo generalizado’. Los trazos generalizados ponen en

evidencia patrones de distribución coincidentes de diferentes grupos taxonómicos; esos patrones de distribución pueden ser la fuente de hipótesis biogeográficas.

Existe un concepto panbiogeográfico que se construye a partir de los trazos: el nodo. El nodo es un concepto más elaborado que se identifica al considerar la intersección de puntos extremos de trazos (fig. 2.1).

Los centros de masa son altas concentraciones de diversidad de un grupo taxonómico y sirven para identificar o proponer 'líneas base'. Las líneas base son líneas que conectan centros de masa y que se conceptualizan como el atributo en común de los trazos generalizados; es frecuente proponer como línea base aquellas que cruzan océanos conectando dos continentes. Los trazos se 'orientan' siguiendo las líneas bases compartidas y se dirigen de los centros de masa menos diversos a los más diversos (fig. 2.2).

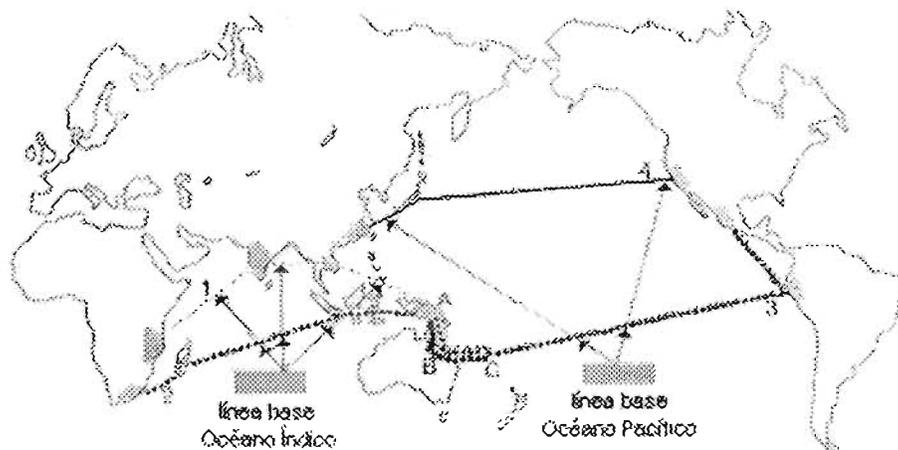


FIG. 2.1 'TRAZOS GENERALIZADOS' ASOCIADOS A UNA 'LÍNEA BASE' Y EL CONCEPTO DE 'NODO'. Los trazos 1 y 2 atraviesan o circunscriben al Océano Índico y pueden asociarse a una misma línea base. También los trazos 3 y 4 pueden asociarse a una misma línea base (Océano Pacífico). Los trazos 1, 2 y 3 se intersectan en las áreas A, B y C, que puede reconocerse como un 'nodo' (Tomado de la figura 13.4 de Craw, 1988).

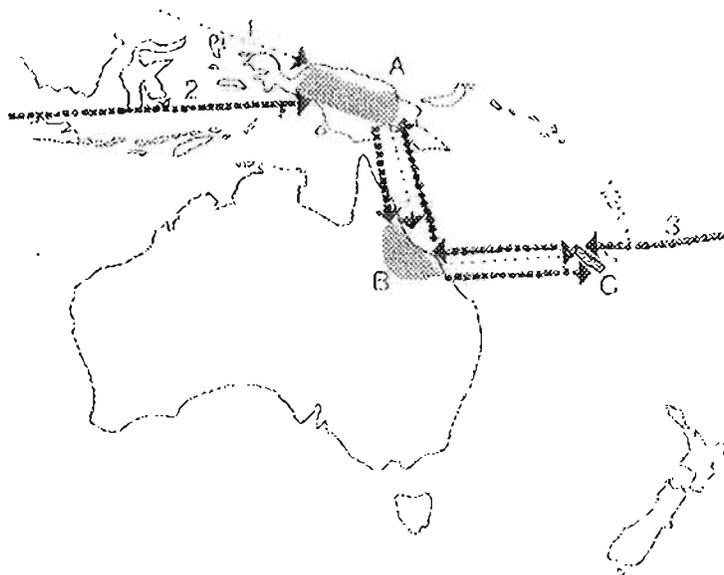


FIG. 2.2 DETALLE DE LA INTERSECCIÓN Y SOBREPOSICIÓN DE LOS TRAZOS 1, 2 Y 3 DE LA FIGURA ANTERIOR EN LAS ÁREAS A, B Y C. Los trazos 1 y 2 están orientados (flechas) en la misma dirección porque comparten la misma línea base. La polaridad combinada para los trazos 1 y 2 es de A hacia B y hacia C. El trazo 3 está orientado en la dirección opuesta de los trazos 1 y 2 porque tiene diferente línea base. (Tomado de la figura 13.5a de Craw, 1988).

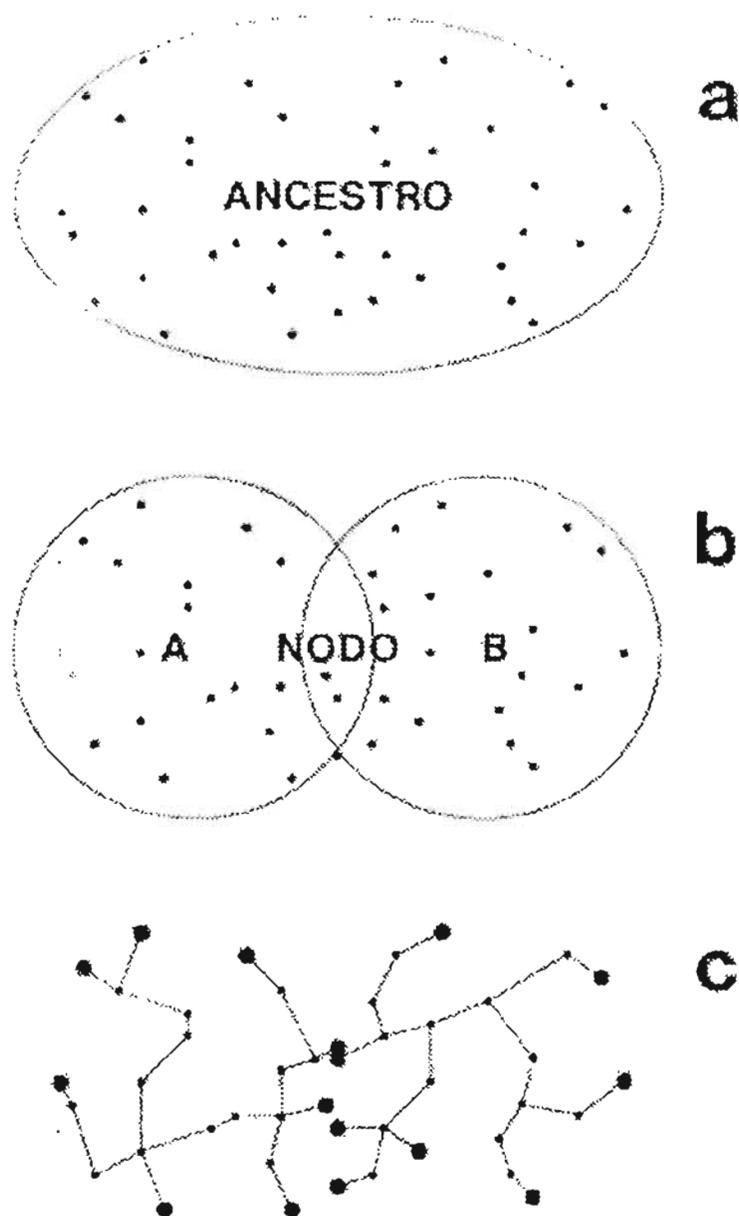


FIG. 2.3 PROCESO DE 'VICARIANZA' Y 'ÁRBOL DE TENDIDO MÍNIMO'. a) Ancestro con distribución amplia; b) Diferenciación de dos especies A y B, el centro del proceso de diferenciación es el 'nodo', c) Árboles de tendido mínimo para las dos especies. Los vértices de grado 1 se muestran en tamaño más grande y se concentran en el nodo. (Tomado de la figura 11 de Henderson, 1989)

En la figura 2.3 se aprecia por qué los nodos pueden indicar áreas de confluencia tectónica. Al generarse un proceso de vicarianza, con la consecuente aparición de dos formas a partir de una, pueden producirse zonas geográficas en donde habitan ambas formas. El 'grado de los vértices' es otro concepto de la teoría de gráficas que se ha incorporado al método panbiogeográfico. El 'grado de un vértice' es el número de trazos que confluyen en ese vértice. Al construir los trazos respectivos para ambas formas, por ejemplo mediante árboles de tendido mínimo, se aprecia que los vértices de grado 1 se ubican en los nodos.

Cuando se construyen los árboles de tendido mínimo para un género, en el que sus especies se han generado mediante vicarianza, los vértices de alto grado representan confluencia de áreas de distribución de varias especies, por lo que pueden identificarse como nodos (fig. 2.4).

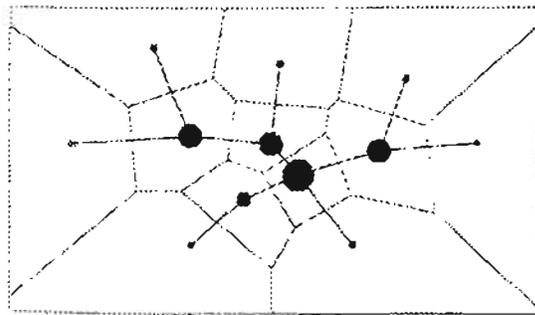


FIG. 2.4 PROCESO MÚLTIPLE DE 'VICARIANZA' Y 'ÁRBOL DE TENDIDO MÍNIMO' DE UN GÉNERO. LOS polígonos representan distribuciones de especies vicariantes. El tamaño de los círculos representa el grado de los vértices. (Tomado de la figura 12 de Henderson, 1989)

La figura 2.5 muestra el caso de la fig. 2.4 pero con datos empíricos de especies de *Trichoptera*. Se aprecia cómo los vértices de grado 1 se concentran en las costas de ambas islas donde se hipotetiza uno de los principales procesos de formación de barrera en Nueva Zelanda. Así, esa zona se puede identificar como nodo.



FIG. 2.5 DISTRIBUCIÓN DE VÉRTICES DE GRADO 1 DE LOS TRAZOS DE TODAS LAS ESPECIES DE TRICHOPTERA EN NUEVA ZELANDA. El tamaño de los círculos representa el número de vértices de grado 1 dentro de cuadros de 10 Km x 10 Km (Tomado de la figura 13 de Henderson, 1989)

En la figura 2.6 se muestran, de manera esquemática, los principales conceptos del método panbiogeográfico: Los 'trazos', los 'centros de masa', las 'líneas base' y los 'nodos'.

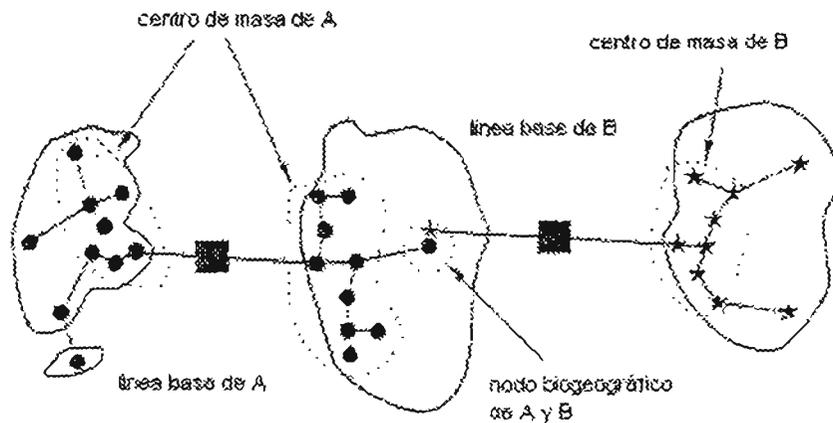


FIG. 2.6 DIAGRAMA CONCEPTUAL GENERALIZADO DE TÉRMINOS USADOS EN EL MÉTODO PANBIOGEOGRÁFICO. Las áreas encerradas representan áreas geográficas separadas por océanos o mares. (Tomado de la figura 1-6 de Craw et al., 1999)

### *Panbiogeografía y Biología de la Conservación*

Craw *et al.* (1999) indicaron que la Panbiogeografía es un enfoque útil en la definición de áreas prioritarias para la conservación, no sólo a nivel global sino regional y nacional. La Panbiogeografía, a diferencia de otros enfoques, como por ejemplo el de Vane-Wright *et al.* (1991), analiza homologías espaciales y no únicamente taxonómicas, toma en cuenta a la distribución espacial como un carácter y considera a los elementos de la biodiversidad de una manera integrada a la triada espacio-tiempo-forma; es decir, los elementos de análisis de la biodiversidad no son únicamente biológicos (taxones) sino biológico-espaciales.

Los nodos son importantes porque son puntos de confluencia donde plantas y animales pueden mostrar presencia local, como endemismos; estas propiedades de los nodos corresponden a los elementos principales que definen centros de biodiversidad; por lo tanto, los nodos son

'focos rojos' (o '*hot spots*' como se les llama en inglés) de la biodiversidad en un contexto biogeográfico (Espinosa y Llorente, 1993; Craw et al. 1999).

La aplicación del método panbiogeográfico es una alternativa novedosa para determinar las áreas a conservar, pues permite expresar, de forma simultánea, la riqueza relativa de especies y los orígenes históricos; el método permite la identificación de nodos que representan áreas complejas que tienen prioridad para su conservación (Morrone *et al.*, 1996).

## 2.2 Métodos multivariados y su aplicación en la Biogeografía, Ecología y Taxonomía

### *Clasificación de los métodos multivariados*

Los métodos multivariados son un conjunto de técnicas estadísticas para el análisis simultáneo de más de una variable independiente. Una de las clasificaciones más comunes de los métodos multivariados aplicados en taxonomía y biogeografía ecológica es aquella que los divide en ‘métodos de ordenación’ y en ‘métodos de agrupamiento’. También se incluyen dentro de las técnicas multivariadas al análisis de varianza y covarianza y la regresión y correlación múltiples.

Ambos tipos de métodos se aplican a una matriz de datos en la que se describen a las especies u otro tipo de unidades a analizar en términos de sus caracteres. En taxonomía numérica las unidades analizadas son llamadas Unidades Taxonómicas Operativas (OTUs por sus iniciales en inglés; Sneath y Sokal, 1973). En la biogeografía ecológica, la matriz de datos puede estar constituida de los taxones a analizar y una lista de variables ambientales de los sitios de recolecta; o en su lugar, la lista de sitios de recolecta o alguna otra unidad geográfica que se desee clasificar con base en la biota. En el caso particular de la biogeografía cuantitativa, estas unidades son llamadas Unidades Geográficas Operativas (OGUs por sus iniciales en inglés; Crovello, 1981). En lo sucesivo se referirá como OGTUs a las OTUs o a las OGU (Unidades Geográficas o Taxonómicas Operativas).

Los métodos de ordenación consideran un espacio multidimensional donde cada uno de los ejes representa a un carácter que en su totalidad caracterizan a las OGTUs. Los métodos de ordenación reducen ese espacio a otro de dos o tres dimensiones sin mucha pérdida de información (Kohlmann, 1994; Sneath, 1998). Uno de los métodos más comúnmente utilizados en biogeografía y sistemática es el ‘análisis de componentes principales’ (Zavala, 1986). También la regresión múltiple se considera dentro de este grupo de técnicas.

Los métodos de agrupamiento generan clases que incluyen a las unidades analizadas, mientras que los métodos de ordenación no establecen una delimitación discreta de las clases. Uno de los métodos de agrupamiento más a menudo utilizado es el ‘agrupamiento por pares

mediante ligamiento promedio' o UPGMA (por sus siglas en inglés, unweighted pair-group method using arithmetic averages).

Actualmente existen muchas herramientas informáticas para la aplicación de métodos multivariados. Palmer (1998) da una lista de *software* para la aplicación de métodos multivariados en Biología. Campbell (1989) explica la aplicación de las técnicas multivariadas en Biología mediante el paquete estadístico SPSS.

### ***Métodos de ordenación***

Los métodos de ordenación no trazan límites que definan grupos, esa tarea se deja al investigador al interpretar los resultados. Las relaciones entre las OGTUs están reflejadas por la posición en un espacio multidimensional, por ejemplo de caracteres; mientras más cercanas están, mayor relación tienen entre sí.

El análisis de componentes principales (PCA por sus iniciales en inglés) reduce la complejidad disminuyendo el número de dimensiones del espacio de las OGTUs. En la figura 2.1 se ilustra un ejemplo de resultado del PCA, en la que se puede observar la relación que guardan las OGTUs en un espacio de tres dimensiones. Esas dimensiones ya no son las dimensiones originales, por ejemplo, caracteres; son una transformación que contiene información, en diferentes grados, de todas las dimensiones originales.

El PCA realiza una transformación de los ejes originales obteniendo otro conjunto de ejes los componentes. El número de componentes es igual al número de ejes originales, pero sólo algunos de ellos, los 'principales', capturan más información de los ejes originales. Esos componentes principales son útiles para representar el espacio multidimensional original (Gauch, 1981).

En la figura 2.2 se esquematiza el algoritmo del PCA. Por razones de simplificación se supone una matriz de datos de ocho OTUs y dos caracteres. En a) se ubican a las OTUs en un espacio de dos dimensiones (carácter 1 y carácter 2); en b) se estandarizan los ejes mediante un movimiento ortogonal, y en c) se muestran los dos componentes calculados mediante PCA, pues el número de componentes siempre es igual al número de ejes originales. Además, los componentes son ortogonales entre sí, pero cada uno captura distinta cantidad de varianza.

En el análisis se obtiene para cada componente un *eigenvalor*. Los eigenvalores están relacionados con la varianza explicada por cada eje y generalmente se expresan en porcentajes; se ordenan del mayor valor al menor y se obtienen sumas acumuladas, para observar, por ejemplo, cuántos componentes principales explican el 90% de la variación.

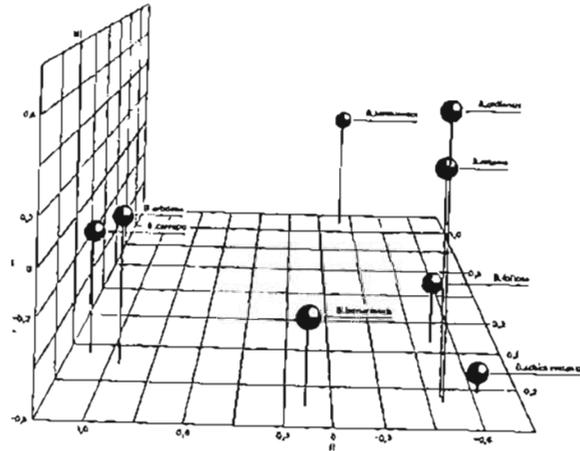


FIGURA 2.1 Análisis de componentes principales de ocho OTUs. Gráfica sobre los tres primeros componentes principales que explican el 80% de la variación total (Tomado de Crisci y López, 1983).

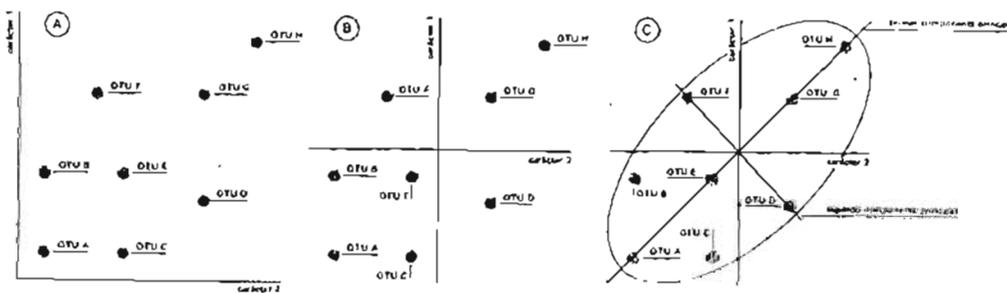


FIGURA 2.2 Esquemalización de los resultados de un análisis de componentes principales con una matriz hipotética de ocho OTUs y sólo dos caracteres. a) Ocho OTUs ubicadas en un espacio bidimensional determinado por los caracteres 1 y 2. b) estandarización de los ejes de la figura a. c) Dos componentes principales obtenidos mediante PCA (Tomado de Crisci y López, 1983).

El análisis de factores es otra técnica de ordenación, muy parecida al PCA. Una de las diferencias principales es que el análisis de factores permite que los factores no sean ortogonales (Sneath y Sokal, 1973; Gauch, 1981), lo que permite una mayor 'flexibilidad' en el acomodo espacial – hablando del multi-espacio constituido por las variables incluidas en el análisis – de los factores; la ortogonalidad es una restricción al modelo de análisis de carácter puramente técnico, la liberación de la condición de ortogonalidad es análoga a la mejora que significa el algoritmo DOMAIN frente al BOX (Carpenter *et al.*, 1993) para el cálculo de distribución potencial de especies. La figura 2.3 muestra un resumen del resultado de un análisis de factores en biogeografía ecológica.

Una lista de recursos relacionados con los métodos de ordenación, que incluye ejemplos, *software*, listas de discusión y breves explicaciones, es proporcionada por Palmer (1998).

### ***Métodos de agrupamiento***

Los métodos de agrupamiento se pueden clasificar en divisivos y aglomerativos. En ambos casos se requiere de la matriz básica de datos, pero en los aglomerativos principalmente se opera sobre la 'matriz de similitud', derivada de la matriz básica, en la que se registran las similitudes entre cada par de OGTUs.

La estrategia de los métodos divisivos es iniciar con la formación de dos grupos a partir del conjunto total de datos y repetir esto para cada uno de los dos grupos y así sucesivamente hasta llegar a grupos unitarios (Hengeveld, 1990), es decir, hasta que sólo contengan a una OGTUs.

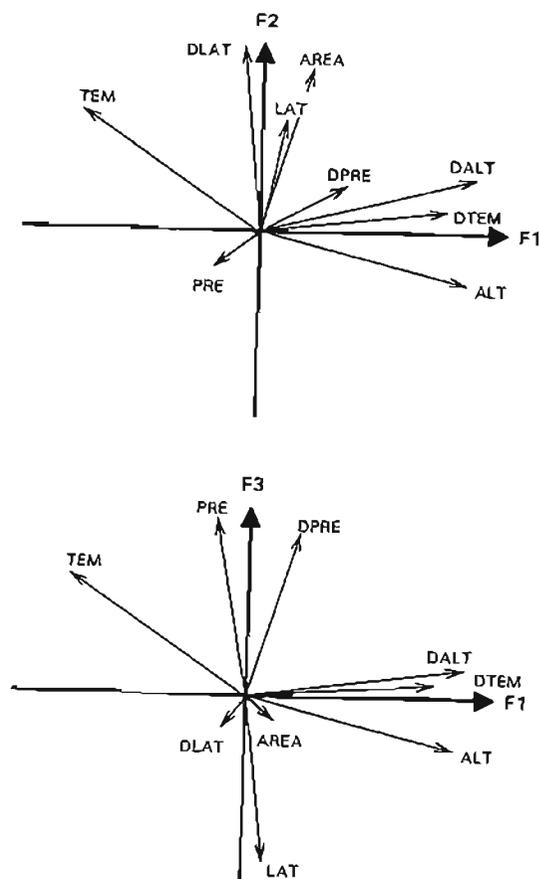


FIGURA 2.3 Análisis de factores de los mamíferos de México. El análisis se realizó tomando en cuenta nueve variables; los tres primeros factores explican el 96% de la varianza. Las nueve variables se grafican en términos de dos factores: la dirección y sentido de las flechas indica la correlación con los factores. La longitud es proporcional a la varianza explicada por cada factor (Tomado de Arita, 1993).

El criterio para dividir un grupo en dos generalmente depende de una función de optimización. Así, se exploran varias combinaciones de los posibles dos grupos y se elige la que cumpla mejor con el criterio de optimización. El análisis de información (Sneath y Sokal, 1973) se puede utilizar para la formación de grupos ya sea divisivamente o mediante agrupamiento. La idea principal es cuantificar el cambio de 'contenido de información',  $\Delta I$ , al pasar de un grupo a otros dos, originados de una partición del primero:

$$\Delta I = I(A,B) - I(A) - I(B)$$

donde  $I$  es una función que mide el contenido de información en una matriz básica de datos.

Dentro de los métodos aglomerativos los más comúnmente usados son los secuenciales-jerárquicos, dentro de los que se encuentra el ligamiento simple, el ligamiento completo y el UPGMA. Estos métodos operan sobre una matriz de similitud o de distancias de las OGTUs (matrices simétricas, con unos o ceros en la diagonal, respectivamente).

El algoritmo inicia eligiendo un par de OGTUs, aquél cuya similitud es la mayor de todos los posibles pares. La matriz de similitud es recalculada considerando ahora como una unidad al par de OGTUs elegidos. La matriz resultante contiene un renglón y una columna menos y se ha formado parte del fenograma: dos hojas unidas a un determinado nivel de similitud. En cada paso, la matriz de similitud es recalculada y va reduciéndose en un renglón y una columna. El algoritmo termina cuando todas las OGTUs han sido ubicadas en el fenograma.

La manera en que es recalculada la matriz de similitud hace la diferencia entre los tipos de ligamiento simple, completo o promedio (UPGMA). En la figura 2.4 se muestra un fenograma resultado de la aplicación del algoritmo UPGMA a una matriz de similitud calculada con el Índice de Simpson.

Ejemplos de la aplicación del UPGMA en taxonomía vegetal se encuentran en los trabajos de Villaseñor y Strother (1989) y Eguiarte (1995). Ejemplos de aplicación en biogeografía ecológica son los de León y Romo (1993), Sánchez (1993) y Villareal *et al.* (1996).

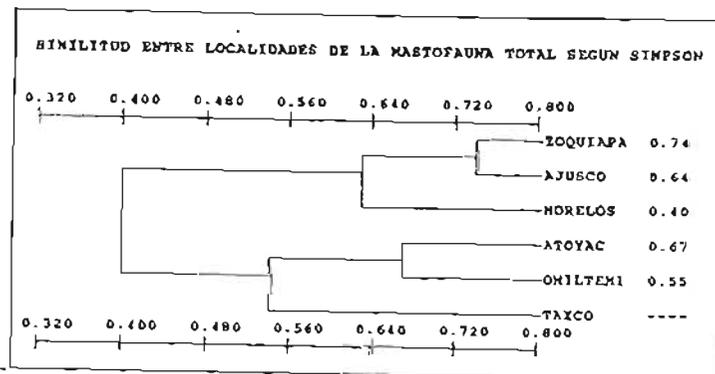


FIGURA 2.4 Fenograma de similitud (UPGMA) de la mastofauna voladora de México (Tomado de León y Romo, 1993).

## 2.3 El concepto de especie

### *La importancia del concepto de especie*

El concepto de especie es fundamental en la tarea del sistemático, del ecólogo y del biogeógrafo (Baum, 1992). La especie es la entidad sobre la que trabaja, sobre la que debe definir relaciones espaciales, temporales o de parentesco y de ancestría/descendencia, entre otras. La especie generalmente es una unidad básica de biodiversidad y conservación, por eso es importante conocer qué se entiende por ese concepto (Claridge *et al.*, 1997).

La controversia en torno al concepto de especie es una de las más ricas en el ámbito de la biología evolutiva (Claridge *et al.*, 1997). Es una de las discusiones que tiene zonas de contacto amplias entre la filosofía y la práctica, ya que tiene relación en términos empíricos con conceptos fenéticos, morfológicos, genéticos y filogenéticos, en los que subyacen conceptos filosóficos (Mayden, 1997).

Llorente y Michán (2000) discuten la validez del concepto de especie en relación con el desarrollo de inventarios biológicos y conocimiento de la biodiversidad. De particular importancia es la discusión epistemológica que hacen sobre las visiones 'realistas' y 'nominalistas' en que el concepto de especie ha sido apreciado por diversos autores. Ellos concluyen que la especie tiene una existencia real y que es una unidad válida para aproximarse al conocimiento de la biodiversidad; reconocen la naturaleza sinérgica de los ecosistemas, por lo que la especie no deja de ser una visión incompleta de 'La Biodiversidad', sin embargo aceptan que por lo pronto, considerando la crisis ecológica, es ya un avance conocer la biodiversidad a nivel especie.

### *Definiciones*

Existen varios trabajos en los que se encuentran las definiciones y discusiones de varios de los conceptos de especie. Por ejemplo, una lista de definiciones de diferentes conceptos de especie se encuentra en Crisci (1994) y en Mayden (1997); una discusión del concepto de especie evolutiva

y filogenética se encuentra en Luna (1994). A continuación se dan definiciones de algunos de los conceptos de especie recopilados por Mayden (1997).

**Especie biológica**

*Grupo de poblaciones naturales que se entrecruzan y que está aislado reproductivamente de otros grupos similares.*

(Mayr, 1969, en Mayden, 1997)

**Especie ecológica**

*Linaje o conjunto de linajes afines, que evoluciona separadamente de otros y que ocupa una determinada zona adaptativa.*

(Van Valen 1976, en Crisci, 1994)

**Especie evolutiva**

*Linaje ancestro-descendiente de poblaciones de organismos que mantienen su identidad de otros linajes y que tienen sus propias tendencias evolutivas y destino histórico.*

(Wiley, 1978, en Mayden, 1997)

**Especie fenética**

*"... el nivel de especie es aquel en el que se pueden observar diferentes grupos fenéticos".*

(Sneath, 1976, en Mayden, 1997)

**Especie filogenética**

*Grupo (basal) irreducible de organismos, diagnósticamente diferentes de otros grupos y dentro del cual hay una relación de ancestría-descendencia.*

(Cracraft, 1989, en Mayden, 1997)

**Especie morfológica**

*Grupo mas pequeño que es consistente y persistentemente distinto y distinguible por medios ordinarios.*

(Cronquist, 1978, en Mayden, 1997)

### **Especie paleontológica**

*Unidad formal de la taxonomía biológica consistente de una población de individuos de los que se presume están estrechamente relacionados entre sí y por lo general son similares en su morfología, que se distinguen de todas las demás poblaciones contemporáneas, y separadas de las poblaciones ancestras y descendientes por límites convenientes pero arbitrarios.*

(Weller, 1961)

*Serie cronológica en un solo linaje cuyos límites por definición son arbitrarios.*

(Simpson, 1961; en Crisci, 1994)

Recientemente, Wheeler y Meier (2000) en ‘Species concepts and Phylogenetic theory’, discuten y comparan los conceptos biológico, lennigiano, filogenético y evolutivo.

### ***Algunas comparaciones***

Algunos conceptos de especie se centran en los mecanismos de especiación; por ejemplo, el concepto biológico reconoce la inhabilidad para tener descendencia como el factor más importante de especiación mientras que el concepto ecológico considera a los parámetros físicos y las propiedades de los organismos (zona adaptativa) para adaptarse al medio (Luckow, 1995).

El concepto biológico de especie requiere que no exista flujo genético entre las poblaciones de especies diferentes, por lo que es difícil de aplicar en plantas, donde la hibridación es un proceso común. El concepto filogenético, que se basa en poder ‘diagnosticar’ a las poblaciones como distintas, sí admite el intercambio genético entre especies. Los dos conceptos tienen diferencias para la biogeografía: si las especies se toman como las unidades para reconocer áreas de endemismo, el concepto biológico originará áreas menos precisas que el filogenético (Cracraft, 1997).

El concepto evolutivo de especie es explícitamente temporal. Las especies son linajes que se extienden en el tiempo, mientras que el concepto fenético considera como un error a la inclusión de la dimensión temporal en el concepto de especie (Hull, 1997).

El concepto biológico es inaplicable en el ámbito paleontológico, donde es casi imposible determinar el intercambio genético. El concepto paleontológico sí considera al tiempo dentro de

su definición, pues habla de linajes y a pesar de esta gran diferencia respecto al concepto fenético, concuerdan en que el aspecto morfológico es importante.

### *Aplicación del concepto de especie*

Los científicos tienen el ideal de que sus conceptos sean lo más teóricos, generales y aplicables. Desafortunadamente, esas tres metas tienden a crear conflictos; por ejemplo, los conceptos teóricos de especie tienden a ser más difíciles de aplicarse (Hull, 1997).

La especie es una unidad real y válida para cuantificar y estudiar la biodiversidad (Llorente y Michán, 2000). Sin embargo, puesto que no hay una aceptación unánime de un concepto de especie, los sistemáticos utilizan diversos conceptos para delimitar las especies. Existe la tendencia a elegir determinados tipos de conceptos de acuerdo con el tipo de organismos, lo cual puede explicarse como poca universalidad de los conceptos en aras de la operatividad. Por ejemplo, en botánica el concepto morfológico de especie es utilizado ampliamente, mientras que en trabajos sistemáticos de zoología el concepto biológico de especie es el que más se aplica (McDade, 1995).

Algunos conceptos sólo pueden aplicarse a un número reducido de organismos, por ejemplo el concepto biológico, pero cuando se aplican forman grupos importantes de organismos. Sin embargo, como no son extensamente aplicables, existe la tendencia de no aplicarse aun cuando sean una buena opción (Hull, 1997).

En los trabajos sistemáticos es muy importante que se indique explícitamente el concepto de especie aplicado, sin embargo, esta práctica es poco frecuente (Luckow, 1995). Diferentes conceptos de especie pueden resultar en diferencias en las estimaciones de la diversidad específica (Cracraft, 1997). Por ejemplo, el concepto biológico sobreestima respecto al filogenético. Así, la integración de listas florísticas y faunísticas pueden contener unidades que nominalmente son especies, pero que por la manera en que se delimitaron pertenecen a diferentes categorías. Esto repercute en la calidad de los estudios biogeográficos. Por ejemplo, una lista de especies endémicas puede contener mayor o menor número de unidades de acuerdo con diferentes conceptos de especie.

Entonces ¿cuál concepto de especie es el adecuado para estudios biogeográficos? Para los estudios biogeográficos que integran el análisis de la distribución de varios grupos de especies,

realmente, en términos prácticos, no hay opción en la elección del concepto de especie, pues el biogeógrafo parte de listas de especies que están definidas bajo diferentes perspectivas. Así, el biogeógrafo utiliza bases de datos con información que describe la ubicación espacial de especies y aún no se tiene la información del concepto de especie con la que se definió cada grupo de especies. En realidad, existe variedad de conceptos de especie que fueron usados en la definición de los diferentes grupos de especies, hay ignorancia que se debe reconocer; el problema es de magnitudes tan grandes que es mejor aceptar esa ignorancia y trabajar sabiendo de las limitaciones de la información contenida en la base de datos, que intentar resolver ese conflicto, considerando la premura con que debe ser resuelta la ‘crisis de la bioconservación’, principalmente en lo que se refiere a la extinción masiva de poblaciones, en algunos casos abarcando a la especie, originada por la ‘avanzada’ civilización del hombre. Por eso aquí se propone el concepto de ‘especie informática’ para reflejar y hacer explícito que el biogeógrafo trabaja con una base de datos biogeográfica de las especies en la que el catálogo ESPECIE es una lista de nombres que provienen de reconocer a las especies de acuerdo con diferentes concepciones, y en general esa correspondencia no es explícita en la base de datos. Es decir, puede ser un concepto de ‘especie nominal’ (aglomerado de especies) pero con la diferencia de que se hace explícita una ignorancia en las consecuencias de trabajar y hacer análisis con ese grupo de especies.

### **Especie informática**

*Especie informática es el valor de cualquiera de los registros del catálogo de ESPECIE (nombre de la especie) de una base de datos y de la cual puede desconocerse el concepto que subyace en su definición pero se es conciente de ello.*

Esa ‘conciencia’ en la ignorancia del concepto de especie subyacente en el nombre de la especie se puede hacer explícita mediante la anexión de un campo en el catálogo ESPECIE que almacene los valores del concepto subyacente (de acuerdo con un catálogo CONCEPTO\_ESPECIE), mismo que contendría, para la mayoría de los registros, el valor NULL (*sensu* Codd, 1990).

## 2.4 El concepto de semejanza en términos de interpretación en Biología comparada y evolutiva

### Concepto de semejanza en la sistemática fenética y en la cladista

Uno de los problemas filosóficos en taxonomía consiste en decidir qué tipo de información debe utilizarse como base confiable de la clasificación biológica: las similitudes por sí solas o las inferencias sobre las relaciones filogenéticas (De Luna, 1995). El sistemático se enfrenta al problema de decidir qué significado evolutivo tienen las semejanzas.

Para el cladismo, las homologías, los grupos monofiléticos y las relaciones filogenéticas están fuera del dominio empírico, es decir, no pueden observarse mediante la experiencia directa, pero de acuerdo con la teoría evolutiva son reales y entonces pueden reconstruirse a partir del análisis de las similitudes heredables (De Luna, 1995). Para los feneticistas sólo la evidencia empírica es aceptable en la construcción del conocimiento.

La fenética o taxonomía numérica es 'un sistema de clasificación basado en la similitud total de los organismos' (Sokal, 1986, en Kohlmann, 1994). Según Sneath y Sokal (1973) la similitud total entre dos entidades es una función de las similitudes individuales de cada uno de los caracteres con los que se comparan; *a priori* cada carácter tiene el mismo peso y las clasificaciones se basan en similitudes fenéticas.

Para los cladistas solamente los caracteres derivados compartidos o sinapomorfias proporcionan la prueba de un origen común, es decir, de monofilia (Villaseñor y Dávila, 1992). Para el feneticismo, solo las similitudes tienen existencia real, la filogenia no puede investigarse con certidumbre; únicamente cabe estudiar las semejanzas porque éstas pueden descubrirse y verificarse en forma empírica (De Luna, 1995). Por esa razón es que la clasificación fenética se basa en la similitud total de los taxones.

La taxonomía numérica también reconoce la importancia que tiene el identificar las homologías en la construcción de clasificaciones, ya que la asignación de los estados de carácter a los caracteres implica el reconocimiento de dichas homologías. Se argumenta que la taxonomía numérica opera adecuadamente generando clasificaciones robustas cuando se equivoca la 'homologación' de caracteres en menos del 10% (Sneath y Sokal, 1973). La taxonomía fenética

utiliza la ‘homología operacional’ para resolver ese problema: dos estados de carácter son el mismo si éstos son indistinguibles. En cambio, para Hennig los caracteres homólogos son aquellos que son observados como estados transformados del mismo carácter original.

La ‘similitud total’ de la fenética a menudo se mide a través de fórmulas de índices de similitud; sin embargo, quedan varios problemas por resolver a ese respecto (Baroni-Urbani y Buser, 1976). De hecho, uno de los principales problemas que tiene la taxonomía numérica es que *“la estimación de la similitud fenética depende del coeficiente de similitud empleado”* (Sneath y Sokal, 1973). Algunos autores han estudiado el comportamiento de los índices de similitud, por lo común bajo determinadas premisas algebraicas, sobre las cuales se establece el buen o mal comportamiento de las fórmulas (Hubálek, 1982; Sánchez y López, 1989; Hengeveld, 1990; Hayek, 1993).

### ***Concepto de semejanza en biogeografía***

Los enfoques contemporáneos principales de la biogeografía se pueden clasificar en históricos y ecológicos. Dentro de los enfoques históricos se encuentran las escuelas del dispersionismo, la panbiogeografía y la biogeografía de la vicarianza o cladista. Dentro de los ecológicos se encuentran las escuelas de la biogeografía ecológica, propiamente, y la del equilibrio insular o biogeografía de islas, entre otros.

El objetivo de la biogeografía histórica es mostrar y explicar la relación de biotas y su conexión histórica entendiendo por ‘relación histórica’ al hecho de compartir descendientes del mismo ancestro (Anderson, 1996). La biogeografía de la vicarianza o cladista compara cladogramas de las áreas y de los grupos monofiléticos de esas regiones para observar qué tanta similitud existe entre los factores históricos (Rosen, 1978). La Panbiogeografía utiliza a los ‘trazos generalizados’ para guiar en la búsqueda de explicaciones de la distribución espacial de las biotas actuales, pero no condiciona su método a contar con las filogenias de los organismos, como si lo hace el cladismo. Entonces, la similitud de las biotas, en términos de trazos generalizados, dará la explicación histórica de su distribución. En los trazos generalizados subyace una significancia estadística (Croizat *et al.*, 1974), pues un trazo generalizado es un grupo de trazos individuales que coinciden aproximadamente; entendiendo que un trazo individual es una

línea que conecta en el espacio a grupos monofiléticos (Simberloff, 1983). Por lo anterior se podría decir que en los trazos generalizados subyace una semejanza estadística.

Sin embargo, tanto para la biogeografía de la vicarianza como para la panbiogeografía, la correspondencia entre las relaciones genealógicas de las áreas y las especies implica homología topológica y evolutiva; topológica porque dos especies de una misma área guardan la misma posición con respecto al todo; evolutiva porque implica una misma historia de formación con un mismo origen o causa (Espinosa y Llorente, 1993).

La Biogeografía del Dispersionismo utiliza como criterio de semejanza la ocurrencia del mismo grupo biológico actual en dos regiones distintas y esto puede ser suficiente para generar una hipótesis históricas, cuya explicación puede ser dispersión o vicariancia.

La Biogeografía Ecológica estudia la distribución individual de cada especie a escalas locales y supone que las causas de los patrones de distribución operan en el corto plazo. La biogeografía histórica puede considerarse como la biogeografía del endemismo, estudia los patrones de distribución a escala global y supone que los factores que los producen son de naturaleza histórica y actúan en tiempo evolutivo. Mientras que para las áreas de distribución de la biogeografía ecológica tiene sentido hablar de forma, tamaño y tendencias de deformación, para las áreas de endemismo sólo tiene sentido hablar de homología biogeográfica (Espinosa y Llorente, 1993).

Uno de los argumentos en contra de la biogeografía ecológica y a favor de la cladista –y que argumenta diferencias en el concepto de semejanza– es el de Nelson y Platnick (1984) acerca de la explicación de la distribución de organismos endémicos:

- 1) Los organismos se transportan a ciertas áreas donde no son endémicos y sobreviven y prosperan.
- 2) Generalmente, las áreas muy separadas albergan *diferentes* taxones, no importando que tan *similares* sean en clima y ecología

Es decir, para la biogeografía cladista (y los hechos de distribución global), la similitud de condiciones climáticas y ecológicas, no implica la similitud de biotas en términos de taxones.

## 2.5 Ubicación de la investigación en Taxonomía y Biogeografía en el contexto de la Biología de la conservación

### *Panorama global de la conservación*

Para poder afrontar la 'crisis de la biodiversidad' mediante estrategias de conservación puede ser necesario conocer el número de taxones y sus patrones de distribución (Villaseñor e Ibarra-Manríquez, 1998). La conformación de bases de datos biogeográficas y el desarrollo de métodos de análisis pueden ser requerimientos indispensables para el diseño de estrategias de conservación (Llorente *et al.*, 1994).

La destrucción de hábitats está reduciendo los números de especies. La magnitud con la que está ocurriendo el proceso no la sabemos, porque, para empezar no tenemos un catálogo completo, o al menos adecuado, de dicha biodiversidad (Dirzo, 1990). Aunque más del 5% de la superficie de la Tierra está designada a áreas protegidas (World Bank, 1995b), éstas no garantizan la conservación de un buen porcentaje de la biodiversidad del planeta.

El Banco Mundial cuenta con un fondo para apoyar proyectos que incluyen:

- Creación de nuevas Areas Protegidas
- Manejo de Areas Protegidas existentes
- Esfuerzos para conservar la biodiversidad existente fuera de las Areas Protegidas
- Conservación *ex-situ*
- Estudios, investigación y monitoreo
- Estrategias, planeación, políticas e instituciones

Para poder estimar la importancia del 'capital natural' de las distintas regiones del planeta, se desarrollan indicadores 'resumen' que sintetizan la información disponible. Por ejemplo, la 'Global Environment Facility' (GEF) desarrolló un método para calcular un número crudo que indique el 'capital natural', con base en las áreas naturales del país, ajustada por la biodiversidad conocida y esperada (World Bank, 1995b).

El Banco Mundial apoya proyectos para la conservación de la biodiversidad incluyendo investigación. Por ejemplo, el proyecto de Parques Nacionales de Venezuela incluye la creación y mantenimiento de bases de datos geográficas, estaciones de investigación y estudios de población y el de las Reservas Naturales de China incluye un programa de investigación de la reserva (World Bank, 1995a).

Otro proyecto global es el de la 'Agenda Sistemática 2000'; se constituye como un proyecto interinstitucional e internacional que tiene como principal objetivo crear un índice virtual de todas las especies de organismos conocidas en la Tierra y crear un marco global para que las bases de datos sobre biodiversidad puedan establecer conexión (Bisby, 1998). Está planeado para realizarse en un periodo de 25 años; sin embargo, debido a que los fondos monetarios son la sexta parte de los esperados, se requerirían 150 años para lograr sus objetivos. Entre otras recomendaciones, establece que 'en todas las universidades e institutos sobre biodiversidad se incluyan sistemáticos en sus facultades'. El proyecto de la Agenda toma como base un conocimiento actual de 250,000 plantas y 45,000 vertebrados descritos, faltando aún por describir 300,000 a 500,000 y 50,000 respectivamente (Systematics Agenda 2000, 1994).

### ***Taxonomía, biogeografía y conservación***

El primer paso para el diseño y manejo de áreas protegidas es describir, inventariar y mapear la biodiversidad, pues es crítica la localización exacta de las reservas para conservar el máximo número de especies. Así, el trabajo de conservación depende de los sistemáticos y de los biogeógrafos, ya que es indispensable conocer dónde establecer las reservas y localizar los sitios de endemismo y diversidad (Soulé, 1990).

Aún faltan por descubrir muchas especies de organismos en el planeta. Se estima que se han descrito alrededor de 1.4 millones de especies y que existen de 10 a 100 millones (Systematics Agenda 2000, 1994; Sarukhán, 1995). Sin embargo, Soulé (1990) argumenta que la estrategia de estudio de los sistemáticos debe cambiar, pues considerando la cantidad de recursos humanos y la cantidad de especies estimadas se podría tomar hasta 4000 años en recolectar, describir y clasificar a las especies faltantes.

La sistemática filogenética genera un orden de relación jerárquico, y ese orden es el que permite tratar a las especies para dar un orden de prioridad para su conservación. Por ejemplo, si se debe elegir la conservación de sólo dos especies de un grupo de tres, es razonable dar prioridad a las formas taxonómicamente más diferentes (Vane-Wright *et al.*, 1991). May (1990) ha llamado 'cálculo de la biodiversidad' a la elección de áreas de protección con base en cálculos de la prioridad de los grupos taxonómicos en la conservación, considerando sus árboles de relación. Esas ideas muestran la importancia de la sistemática en la biología de la conservación.

La sistemática y la biogeografía auxilian a la biología de la conservación proporcionando elementos para su propio marco teórico, el cual debe de reflejarse en aspectos pragmáticos. A ese respecto se puede comentar que las estructuras 'ideales' de las áreas protegidas pueden adaptarse para que puedan ser opciones viables en el 'mundo real', como por ejemplo, lo muestran Rebelo y Siegfried (1990). Varios aspectos de las técnicas de la biología de la conservación tienen que ver con cuestiones biogeográficas; por ejemplo, el algoritmo propuesto por Rebelo (1994) para proponer áreas de protección con base en las regiones de endemismo.

La estimación de las áreas de distribución de las especies a partir de los datos de los sitios de recolecta es una tarea de la biología de la conservación que tiene que ver con la teoría biogeográfica. Por ejemplo, Carpenter *et al.* (1993) describieron un procedimiento para estimar áreas de distribución potencial con base en biogeografía ecológica.

### ***Biodiversidad e investigación en México***

Una de las estimaciones más recientes acerca de la riqueza de plantas con flores en el territorio mexicano es de 25,000 especies, para la mayoría de las cuales no existe información compilada y analizada que permita determinar aspectos básicos de su biología, indispensables para la toma de decisiones de su conservación y eventual manejo (Villaseñor e Ibarra-Manríquez, 1998). Del 8 al 10% de vertebrados y plantas terrestres del planeta ocurren en el territorio de México, de hecho México es considerado uno de los 10 países de 'megadiversidad' del planeta. Se estima que de todas las especies que ocurren en el territorio mexicano, no más del 50% han sido identificadas y caracterizadas científicamente (Dirzo y Raven, 1994).

Aunque con cierta escasez, en México se cuenta con cierta base de información necesaria para la conservación: en el país existen 179 colecciones biológicas institucionales que albergan a

casi 10 millones de ejemplares. En particular, las colecciones botánicas albergan a casi 3 millones de ejemplares y sólo 61% de ellas tienen plantas vasculares (Llorente y Koleff, 1997). Sin embargo, es necesario que la información sobre biodiversidad se encuentre accesible vía bases de datos a un amplio tipo de público para que sea aprovechada (Allkin, 1998).

El gobierno federal del sexenio pasado reconoció la importancia de la investigación para poder llevar a cabo acciones de conservación:

*“Los aspectos científicos y técnicos de la lucha contra la contaminación, la conservación ecológica y la restauración, requieren de personal capacitado técnica y científicamente. ...México necesita invertir recursos en la formación de cuadros científicos e instituciones con la capacidad de realizar investigación científica creativa y original que podrá sustentar las tecnologías aplicadas a nuestras singulares condiciones ecológicas”*

(Poder Ejecutivo Federal, 1996).

Según datos del Sistema Nacional de Investigadores de México (SNI, 1994), en 1994 el número de investigadores vigentes en las áreas ‘Biológicas, Biomédicas y Químicas’ era de 1911, lo cual, evidentemente es un número reducido tomando en cuenta las consideraciones previas acerca de la riqueza y biodiversidad de nuestro país y además que la mayoría son químicos, biomédicos y biólogos experimentales. Por lo anterior, se debe tomar en consideración el cambio de estrategia en el estudio de los organismos, como así lo sugiere Soulé (1990), con la finalidad de poder afrontar con éxito el problema de la ‘crisis de la biodiversidad’.

## 3. Problemas fundamentales de la Biogeografía Cuantitativa

---

*En este capítulo se integra y estructura una lista de los problemas más importantes de la Biogeografía Cuantitativa, planteando de manera específica cada uno de ellos bajo la perspectiva de estudiarlos y generar algunas propuestas de solución.*

### 3.1 Métodos de la Biogeografía Cuantitativa

De manera superficial, *análisis* puede interpretarse como ‘entender’ o generar ‘conocimiento’. Se puede tener un mejor acercamiento al concepto de análisis si se concibe como una etapa epistemológica, es decir, como un medio de acercarse y entender al mundo.

Así, se puede dar la siguiente definición epistemológica de análisis:

*El análisis es una etapa en la construcción de abstracciones para entender el mundo, es decir, una etapa en la generación de conocimiento. El análisis es posterior a la observación y a la descripción de los hechos o fenómenos percibidos, pero anterior a la explicación del fenómeno completo (modelación).*

El análisis sólo existe en aquellos procesos en el que su entendimiento es escaso, de otra manera la observación directa basta para su explicación. El análisis surge cuando no se puede entender el mundo solo mediante la observación. Análisis implica la aceptación de la ignorancia sobre algo. En el análisis se clasifican los elementos ya conceptualizados y nombrados en las etapas de observación y descripción, estableciendo relaciones directas entre ellos. Así, uno de los principales resultados del análisis es una clasificación de las observaciones y el establecimiento de relaciones directas entre ellos (Salazar, 1979).

Al considerar el concepto expuesto sobre *análisis*, el ‘análisis de los métodos de la biogeografía cuantitativa (BC)’ puede traducirse como clasificarlos –previa a una definición y descripción de ellos– estableciendo relaciones entre ellos, y definiendo, de manera relativa, cuáles son sus propiedades. Aunque los métodos de la BC se aplican ampliamente para apoyar al análisis biogeográfico, para muchos de ellos aún se desconocen sus propiedades y, por lo tanto, la validez (mejor dicho refutabilidad) de las hipótesis construidas en su base. Las relaciones entre ellos (relaciones directas o concatenaciones establecidas en el análisis) también se desconocen o se tiene poco conocimiento.

El mismo desconocimiento de sus propiedades ha hecho que la elección de un método en particular (o subconjunto de ellos), dentro de la amplia variedad de éstas sea una tarea complicada para el biogeógrafo. En general, los autores justifican su elección en trabajos previos que a su vez tienen poca o nula fundamentación teórica.

Lo anterior obliga a pensar en términos teóricos, no sólo empíricos, en el uso de los métodos de la BC. Es decir, establecer de manera explícita y con precisión las preconcepciones sobre su objeto de estudio y la manera en que los tratan, establecer su grado de consistencia, poder de resolución y sus alcances y limitaciones. Al existir una amplia variedad de métodos, se aprecia que la tarea es de grandes dimensiones, pero es posible dividir el problema y estudiarlo por partes; este trabajo se propone abordar sólo una pequeña parte de esos problemas.

Cada método o herramienta particular de la BC requiere una revisión de sus principios y fundamentar y justificar su aplicación e integración en el contexto del análisis biogeográfico. Además de un análisis y fundamentación de cada método, se requiere un ‘hilo unificador’ diferente al atributo ‘cuantitativo’. Se requiere de igual manera para cada método particular un ‘modelo integrador’, una visión que unifique a ese conjunto de técnicas y que los criterios para

esa unificación sean distintos al sólo hecho de que se utilizan en el análisis biogeográfico y que son técnicas cuantitativas.

Así, se propone la siguiente lista de problemas fundamentales de la BC que requieren solución urgente o cuando menos descripción:

- 1) El problema de la definición del tamaño de la OGU.
- 2) El problema de la cuantificación de la calidad de la información de los registros biogeográficos.
- 3) El problema de determinar los efectos del índice de similitud en la clasificación biogeográfica.
- 4) El problema de una integración ordenada de los métodos de la BC, es decir, de establecer las relaciones entre ellos, por ejemplo la posibilidad o límites para su encadenamiento.

### ***Definición del tamaño de la OGU***

El requisito de clasificar al área de estudio en OGUs (*sensu* Crovello, 1981) para la aplicación de muchas de las técnicas de la BC, ha generado la pregunta ¿de qué tamaño y forma deben de ser las OGUs?

Phipps (1975) aborda el problema de la definición del tamaño de la OGU. Esta referencia, aunque brinda un método explícito y entendible, presenta una fundamentación algo oscura.

El problema de agregación es un problema común en el área del análisis espacial y que está ligado al de la ‘escala’ (Openshaw, 1998). En el área de los GIS se ha denominado como el ‘*problema de la unidad de área modificable*’ o MAUP, por sus iniciales en inglés (Openshaw y Taylor, 1979 en Wong, 1996).

Cabe añadir que la definición del tamaño de la OGU –y también la forma– se realizan en etapas muy tempranas del análisis biogeográfico, razón por la que su estudio adquiere especial importancia.

### ***Cuantificación de la calidad de la información***

El número de bases de datos biogeográficas y el volumen de la información que contienen actualmente es suficiente como para realizar estudios de biodiversidad de muchas regiones a varias escalas y considerando a un gran número de grupos taxonómicos. ¿Qué límites tiene la información contenida en esas bases de datos? ¿Es representativa de todas las regiones y grupos taxonómicos a las que hace referencia? ¿En qué medida?

Es claro que un problema importante es cuantificar la calidad de la información de las bases de datos biogeográficas y que al resolver ese problema, se podrá dar un ‘intervalo de confianza’ para los resultados obtenidos, es decir, abordar el estudio del efecto de la calidad de la información en los resultados del análisis. Por ejemplo, una regionalización basada en pocos registros tendrá una validez menor sobre otra basada en información con mayor número de registros, y entonces ¿qué tanto distan estos supuestos dos grados de validez?

### ***Efectos del índice de similitud en la clasificación biogeográfica***

Los índices de similitud son una herramienta de la BC muy recurrida en análisis biogeográfico. Existe una cantidad grande de ellos y es probable que continúe creciendo la lista, debido principalmente a la aplicación de los nuevos enfoques o teorías físico-tecnológico-matemáticas sobre la medida de similitud entre dos conjuntos, como lo son la lógica difusa, las redes neuronales o los fractales. Además, los índices de similitud son base de otras herramientas; la más común es el ‘análisis de agrupamiento’, el cual a su vez es un método que tiene variantes (los tipos de ligamiento). En la mayoría de los casos, al aplicar un mismo tipo de análisis de agrupamiento (por ejemplo UPGMA) a un mismo conjunto de datos, pero usando diferente índice de similitud, se obtienen resultados diferentes. Surge entonces la pregunta ¿Qué índice de similitud elegir?

En la literatura hay algunos trabajos que abordan el problema (ver por ejemplo Hubálek, 1982; Sánchez y López, 1988); sin embargo, es claro que en ellos se reconoce que son necesarios más estudios. Aquí se propone abordar el problema, con aproximaciones diferentes a las citadas en la literatura y comparar los resultados.

La pregunta particular para este problema se puede redactar como: ¿Qué consecuencias tiene la elección del índice de similitud en el análisis de agrupamiento?

### *Integración coherente de los métodos de la BC*

En las secciones anteriores se ha planteado abordar problemas que tienen que ver con aspectos particulares de la BC. Ya se ha mencionado que la BC cuenta con gran variedad de métodos y precisamente en el proceso del análisis biogeográfico se deben elegir varios de ellos. Por esa razón Crovello observa al análisis biogeográfico como un proceso de toma de decisiones. En este proceso de toma de decisiones, una visión global pero ordenada de los métodos de la BC ayudará en la elección. El orden debe establecerse con base en el proceso de análisis biogeográfico, es decir considerando que el proceso del entendimiento de la distribución espacial de las especies puede hacerse desde la descripción hasta la modelación.

## **3.2 Herramientas para la Biogeografía Cuantitativa**

Los paquetes de análisis geográfico (GIS) normalmente proveen herramientas analíticas elementales, por lo que su aplicación es generalizada. Sin embargo, los especialistas desean implementar sus propios métodos, que con frecuencia no es posible hacerlo en los paquetes comerciales (O'Kelly, 1994). El caso de la BC no es la excepción; prácticamente es imposible implementar en un GIS comercial, por ejemplo, un análisis de agrupamiento.

Lo anterior sugiere la necesidad de construir herramientas informáticas, de orientación 'vertical' hacia la BC; herramientas que integren las técnicas de la BC más recurridas por los biogeógrafos al realizar sus análisis.

## ***Desarrollo de Sistemas de Información Geográfica para la BC***

El problema de construir herramientas para la BC se plantea en este estudio como el ‘desarrollo de un GIS’ por las siguientes razones:

*Desarrollo.*- El término desarrollo hace alusión a la *construcción de sistemas informáticos* utilizando metodologías ampliamente probadas.

*GIS.*- La herramienta debe ser un GIS debido a que 1) es un sistema informático que manipula información de *entidades georreferidas*, y 2) que es una herramienta que debe permitir realizar *análisis espacial* (el concepto de *análisis espacial* se define y fundamenta en el capítulo 9 ‘*Análisis de los Métodos de la Biogeografía Cuantitativa*’).

## 4. Objetivos

---

*En este capítulo se plantean los objetivos de este trabajo como una respuesta a los problemas de la Biogeografía Cuantitativa establecidos en el capítulo anterior.*

### 4.1. Análisis de Métodos de la Biogeografía Cuantitativa

Al tomar en consideración lo planteado en el capítulo previo, el objetivo de ‘analizar los métodos de la BC’ se acota y particulariza en los siguientes ‘objetivos particulares’:

- 1) Analizar el problema de definir el tamaño de la OGU.
- 2) Analizar el problema de cuantificar la calidad de la información de los registros biogeográficos.
- 3) Analizar el problema de los efectos del índice de similitud en la clasificación biogeográfica.
- 4) Proponer una clasificación de los métodos de la BC con base en el proceso de análisis biogeográfico.

### ***Definición del tamaño de la OGU***

Para este problema también se impone una acotación: únicamente se aborda el problema del ‘tamaño’, dejando de lado el de la ‘forma’.

En un capítulo previo se dedica una sección a la discusión de la *escala* en biogeografía. Sin embargo, en este trabajo no se aborda el problema de la ‘escala’ pero sí el de la ‘resolución’, entendiendo a la *escala* como una gradación espacial en la que se pueden observar diferentes tipos de fenómenos y como *resolución* a la precisión de la información espacial. Así, la *resolución* es algo que debe determinarse después de la *escala*, ya que la *escala* tiene que ver directamente con el significado del análisis, mientras que la *resolución* tiene que ver con la precisión de la fuente de información o de los resultados.

### ***Cuantificación de la calidad de la información***

En este estudio se propone medir la calidad de la información únicamente considerando a los atributos de estudio y su ubicación espacial, es decir a los registros de ejemplares de recolecta, identificados taxonómicamente y geocodificados de manera que forman tuplas taxón-coordenadas geográficas, sin considerar a la fuente de dicha información, como lo son los periodos de recolecta, la frecuencia o intensidad de recolecta en el tiempo o el nivel y precisión de la determinación taxonómica, entre otros.

Esta cuantificación de la calidad de la información debe ayudar a estimar la confianza de los resultados de análisis surgidos a partir de bases de datos biogeográficas incompletas.

### ***Efectos del índice de similitud en la clasificación biogeográfica***

Este problema se aborda únicamente considerando tres tipos de ligamiento: simple, completo y promedio, y sólo una lista de los índices de similitud más usados en análisis biogeográficos.

Además de estudiar comparativamente a las clasificaciones generadas por diferentes índices de similitud, también se propone analizar a la matriz de similitud, que es una estructura de datos previa al fenograma.

### ***Integración coherente de los métodos de la BC***

Ya que el análisis a fondo de cada uno de los métodos de la BC es un problema poco manejable, se propone establecer un orden general mediante el análisis. En este problema se plantea al ‘análisis’ como dos conceptos diferentes y no excluyentes:

- 1) como la ‘descripción’, ‘clasificación’ y ‘establecimiento de relaciones’ de los métodos de la BC, y
- 2) como una etapa de las metodologías del desarrollo informático.

Así, se plantea el desarrollo de cada uno de los elementos de análisis con el objetivo de tener una visión global y coherente y al mismo tiempo encaminada al desarrollo de una herramienta informática.

## **4.2. Desarrollo de Herramientas para la Biogeografía Cuantitativa**

El problema de construir una herramienta para la BC se traduce en el objetivo particular:

- desarrollo de un GIS para la BC.

## 5. Método

---

*En este capítulo se explican los conceptos aplicados en el análisis de los métodos de la BC desde la perspectiva de la estructuración del conocimiento. También se explican las etapas del desarrollo de un GIS para la BC concebido como un sistema de información.*

El análisis de los métodos de la BC se realiza desde dos perspectivas. La primera es centrandó la atención en aquellos métodos que tienen que ver con los problemas fundamentales de la BC expuestos en el capítulo 3, a saber:

- 1) El tamaño de la OGU
- 2) Cuantificación de la calidad de la información
- 3) Efectos del índice de similitud en la clasificación biogeográfica

Estos problemas se abordan en los artículos incluidos en los capítulos 7 y 8 y, por lo tanto, la metodología seguida se explica en ellos. El problema del tamaño de la OGU se trata integradamente con el de la cuantificación de la calidad de la información en el capítulo 7. En el capítulo 8 se aborda el problema del efecto del índice de similitud en la clasificación biogeográfica.

La otra perspectiva del análisis es tratando de conducirlo a una integración coherente de los métodos. La metodología seguida se expone en las siguientes dos secciones; la primera observa a los métodos en un contexto cognoscitivo, pues se propone una visión de ingeniería del

conocimiento (Feigenbaum, 1977) ampliamente documentada en la rama del mismo nombre de una de las ciencias del conocimiento denominada *Inteligencia Artificial*. En particular, se resume una metodología para estructurar el conocimiento. La segunda los observa desde una perspectiva de la ingeniería de *software*. Se propone entonces, como parte de esa ‘visión integral y coherente’ su automatización de manera lógica en una herramienta de cómputo, para lo cual se aplica una metodología de desarrollo de sistemas de información de la Ingeniería de Sistemas.

## **5.1. Criterios para el análisis e integración de los métodos de la BC**

### ***Clasificación y generación de conocimiento***

Los criterios presentados en esta sección, además de la intención de delimitar el problema, también tienen el propósito de investigar aquellos aspectos biogeográficos básicos –el taxón y su ubicación espacial-; de investigar lo simple y sólo después lo más complejo, de plantear explicaciones sencillas y, sólo después de que sean refutadas, plantear explicaciones más complejas. En todo caso, si no se encuentran explicaciones coherentes con información simple, deberá replantearse el problema antes de involucrar más variables al análisis; empezar directamente un análisis es riesgoso sin antes obtener una descripción; pues se pueden aventurar conclusiones sobre el entendimiento del proceso sin antes tener bien definidos los elementos relacionados con el problema. La descripción de los elementos involucrados no da garantía de que mediante el análisis el proceso se entienda correctamente, pero sí previene y pone en claro presuposiciones.

En esta sección se explican las acotaciones impuestas al problema que reflejan puntos de vista epistemológicos del autor, la filosofía acerca del proceso de generación del conocimiento. Estos puntos de vista también se ven reflejados en la concentración de los esfuerzos de investigación al aspecto básico de la calidad de la información, tratado en el artículo de Murguía y Villaseñor (2000) incluido en el capítulo 7 de esta tesis.

A continuación se explican algunos conceptos filosóficos en los que se basa el análisis de los métodos de la BC de esta tesis.

***El conocimiento 'a priori'***. Para Kant (1781) existen dos formas de conocimiento 'a priori': el espacio y el tiempo. El término 'a priori' se refiere a que está libre de las sensaciones; no es un conocimiento empírico, no es una intuición que se refiere al objeto por medio de la sensación:

*'Es considerado, pues, el espacio como la condición de la posibilidad de los fenómenos y no como una determinación dependiente de éstos, y es una representación 'a priori', que necesariamente está en la base de los fenómenos externos' (p. 43).*

Así, investigar al taxón y su ubicación espacial es estar en la base del conocimiento biogeográfico. Al entender al espacio como una presuposición, y no como un hecho, se logra, en cierta medida, eliminar algunos aspectos subjetivos durante el análisis.

***Refutabilidad***. En la ciencia es preferible plantear explicaciones sencillas y sólo después de que sean refutadas, plantear explicaciones más complejas. Por ejemplo, en un ajuste de pares de datos se debe preferir el modelo lineal al modelo log-log, siempre y cuando el primero tenga una buena correlación y aun cuando el segundo tenga otra mejor.

La sencillez de la hipótesis tiene que ver con su refutabilidad y hacer hipótesis refutables es hacer ciencia, ya que, siguiendo a Popper (1966), los postulados científicos no son verificables sino refutables.

***Generación estructurada del conocimiento***. Es improbable empezar directamente un análisis sin antes obtener una descripción entendible de los elementos involucrados. El proceso de generación de conocimientos puede concebirse como parte del mismo conocimiento, por lo que si el conocimiento científico debe ser ordenado y sistematizado, lo mismo aplica al proceso de su generación. Salazar (1979) propuso una estructura para la generación del conocimiento científico, que en forma secuencial puede enunciarse como la secuencia de varios niveles de

abstracción: *Percepción, Descripción, Clasificación, Relaciones directas, Concatenación y Formulaciones o Emunciados.*

Cada nivel de abstracción es implicado de manera lógica por el anterior, pudiendo existir implicaciones entre niveles más alejados. Por ejemplo, la formulación de clases puede darse a través de la observación directa, sin que medie entre ellas algún proceso de descripción. Sin embargo, es una tarea ardua que pueden lograr individuos experimentados en su campo de competencia. Para la ciencia es importante documentar y hacer explícitos los estados de transición entre los diversos niveles de abstracción, ordenándolos y sistematizándolos.

### ***Criterios para integrar la lista de métodos a analizar***

Con la finalidad de trabajar con elementos ‘de un mismo conjunto’ y no mezclar elementos distintos, se propone una serie de criterios o restricciones que deben cumplir los métodos a incluir en el análisis. La lista de métodos resultante también tiene la finalidad de acotar el problema y ponerlo en dimensiones manejables. En la figura 5.1 se indican las restricciones impuestas a los métodos de la BC a incluir en el análisis.

El análisis sólo considera a aquellos métodos que:

- ⇒ son ampliamente usados por la comunidad de biogeógrafos;
- ⇒ operan sobre datos biogeográficos básicos (matrices de presencia/ausencia);
- ⇒ para su operación no es necesario contar con información del ambiente más allá de su ubicación geográfica en dos dimensiones;
- ⇒ estén dirigidos al análisis de la riqueza y no de la abundancia.

**Figura 5.1** Restricciones impuestas a la lista de métodos a incluir en el análisis.

Una vez conformada la lista de análisis y previa a su descripción, se propone construir una clasificación de los métodos de acuerdo con el nivel de abstracción en el procesamiento de

información. Lo anterior tiene la finalidad de contar con una base ordenada que permita la continuación del análisis así como establecer un orden en su presentación.

Los resultados del análisis de los métodos de la BC se exponen en el artículo '*Biogeografía cuantitativa*' y de manera detallada, en el capítulo 7 '*Análisis de Métodos de la BC*'

## 5.2. Metodologías para el Desarrollo Informático

Construir cualquier sistema de información no es trivial, es una tarea compleja y a menudo abrumadora; Por eso es que la construcción de cualquier sistema de información requiere de una metodología que encamine todos los esfuerzos a alcanzar el objetivo. Las metodologías de desarrollo de *software* permiten:

- ❑ Tener un concepto claro y limitado de lo que el *software* debe hacer.
- ❑ Planear su desarrollo:
  - estimar el tiempo requerido para su desarrollo
  - identificar los recursos a utilizar (computadoras, *software* de desarrollo y otros)
  - identificar los recursos humanos necesarios
- ❑ Evitar los riesgos del desarrollo:
  - abarcar tiempos excesivamente largos en la creación del software
  - aumentar innecesariamente la complejidad
  - elevar el costo de desarrollo
  - no cumplir con los requerimientos establecidos
  - construir interfaces confusas, poco coherentes, con desempeño pobre o inestables.

A un sistema de información se le puede considerar como un ente abstracto que tiene un ciclo de vida. Es decir, tiene un nacimiento en el cual es creado, crece a través de las

modificaciones y mejoras que se le hacen para mantenerlo al día, e inclusive muere si ya no cumple con los objetivos o es demasiado costoso modificarlo; también fenece cuando sus objetivos ya han desaparecido o simplemente se sustituye por un sistema nuevo.

### ***Ciclo de vida de un Sistema de Información***

El *ciclo de vida* de un sistema de información es un proceso por el cual los analistas de sistemas, los ingenieros de *software*, los programadores y los usuarios finales elaboran sistemas de información o *aplicaciones* informáticas.

Una *metodología* es una versión amplia y detallada de un ciclo de vida completo del desarrollo de sistemas que incluye:

- Especificación de tareas paso a paso para cada fase.
- Funciones que en cada tarea deben desempeñar los miembros del equipo de trabajo; individuales o en grupo.
- Descripción de los productos resultantes y normas de calidad para cada tarea.
- Técnicas de desarrollo.

Una *metodología* o *modelo de ciclo de vida* es un modelo prescriptivo desde la vaga concepción de un sistema hasta su declinación y muerte. Es indistinto referirse al ciclo de vida como el modelo de ciclo de vida o como una *metodología de ciclo de vida*. Para poder seleccionar el ciclo de vida más apropiado, es importante conocer los requisitos del software a desarrollar (por ejemplo, magnitud, software empotrado, de aplicación general, científico, a la medida, entre otros), los recursos humanos con que se cuentan y la experiencia de éstos, así como el tiempo y dinero disponibles.

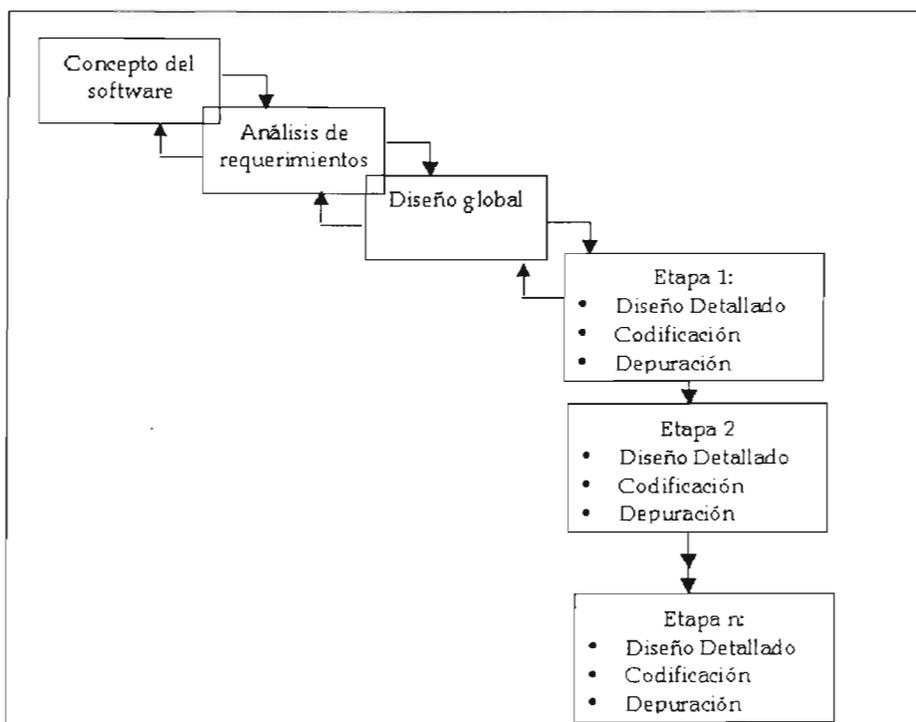
## ***Desarrollo por etapas***

Para el desarrollo del sistema de información geográfica para la BC se siguió la metodología de ‘entrega por etapas’ del desarrollo de sistemas informáticos (McConnell, 1997; Whitten *et al.*, 1997).

El desarrollo por etapas es un modelo en el que el *software* se muestra a los usuarios en etapas refinadas sucesivamente. Supone que se conoce exactamente qué es lo que se va a construir cuando se inicia el desarrollo. No se entrega el proyecto al final del desarrollo, de un solo golpe, sino que se entrega por ‘etapas’.

En la figura 5.2 se esquematizan las fases del desarrollo por etapas; a continuación se describe a cada una de ellas:

- *Concepto del software.* Establecer un plan de sistemas que tenga una visión general de la misión de la organización, con la finalidad de identificar y establecer prioridades sobre aquellas aplicaciones de los sistemas de información cuyo desarrollo ofrezca beneficios máximos para la organización.
- *Análisis de requerimientos.* Especificaciones de las necesidades de los usuarios. Se debe establecer un diálogo entre los desarrolladores y los usuarios, formalizándolo con documentos en los que ambas partes entiendan en detalle y estén completamente de acuerdo.
- *Diseño global.* Evaluación de las diferentes alternativas, así como de una primera especificación tentativa de la solución final. El primer tipo de diseño debe ser general, de manera que permita tener un esquema del diseño global, de ahí se generará un diseño detallado en el que se centran las especificaciones pormenorizadas de los componentes.



**Figura 5.2** Fases de la metodología de 'Desarrollo por Etapas'.

### *Etapas*

- *Diseño Detallado.* Éste se desglosa en dos partes: la primera se refiere al diseño externo que consiste en el conjunto de especificaciones de la interfaz del sistema con sus usuarios. La segunda parte se refiere al diseño interno, que es el conjunto de especificación del software del sistema en su estructura y lógica.
- *Codificación.* Durante esta fase se escribe el código que una vez compilado conforma la herramienta deseada.
- *Depuración.* Se realizan pruebas de desempeño con el usuario final y las correcciones necesarias como eliminar 'bugs' o quizá algunos errores de conceptos.

En el desarrollo por etapas se entregan al usuario partes funcionales antes de concluir la aplicación que le permite poder operar el *software* desde el principio. El riesgo que se corre con este modelo es el de no cumplir con los tiempos planeados. Además, pueden darse casos en el

que un módulo construido en etapas tempranas no corra hasta que se termine otro de etapas posteriores.

### ***Características generales de la herramienta***

La siguiente es una lista de características con las que inicia la idea del desarrollo de la herramienta:

- Generalización de puntos (localidades de recolecta) mediante cuadros (OGUs en forma de cuadros).
- Información base necesaria para los análisis: lista de taxones y las coordenadas de la localidad de recolecta asociada.
- Los análisis que implementa son sobre datos de riqueza (matrices de presencia ausencia) y no sobre abundancia.
- Desarrollo que requiere inversión económica moderada.
- Interfaz con el usuario muy empírica, que permita tanto análisis exploratorios como análisis formales.
- Interfaz que maneje nomenclatura biogeográfica más que informática o de GIS.
- Que permita desplegar con facilidad informes de listas de especies.

Los resultados de aplicar la metodología de desarrollo se muestran en el artículo '*Biogeografía cuantitativa*', y de manera detallada en el capítulo 8 '*Desarrollo de un GIS para análisis biogeográfico*'.

## 6. Presentación de los métodos de la Biogeografía Cuantitativa

---

*En este capítulo se presentan los resultados de integrar a los métodos de la BC en una herramienta informática (GIS para la BC). Se presentan ejemplos de la interface del sistema que muestran el resultado de desarrollo del GIS.*

Este capítulo es parte de los resultados de este trabajo en lo referente a los objetivos sobre:

- 1.4) Proponer una integración ordenada de los métodos de la BC.
- 2.1) Desarrollar un GIS para la BC.

El capítulo, en su totalidad, es el documento entregado a los editores de un libro sobre “Biogeografía”, quienes me invitaron a escribir un capítulo sobre “*Biogeografía Cuantitativa*”:

MURGUÍA, M. Y F. ROJAS. 2001. **Biogeografía Cuantitativa**. En: LLORENTE, J. Y J.J. MORRONE (Ed.). **Introducción a la Biogeografía en Latinoamérica: Teorías, Conceptos, Métodos y Aplicaciones**. UNAM. 277p.

El objetivo del capítulo dentro del libro es mostrar las herramientas de la BC y documentarlas; esta presentación se realiza, en gran medida, auxiliándose de la interfaz del sistema de información Geográfica para la BC diseñado en este trabajo.

Así, el capítulo presenta 1) los métodos de la BC de una manera integrada, y 2) la interfaz de la herramienta diseñada que resultó del análisis de esos métodos. Así, este capítulo “adelanta”, de manera superficial, algunos de los resultados que se desarrollan en el capítulo 10 sobre el desarrollo de una herramienta para la BC.

# BIOGEOGRAFÍA CUANTITATIVA

Miguel Murguía y Fidencio Rojas

La biogeografía es el estudio de la distribución de los organismos, la historia de cómo se forma un arreglo particular y el análisis de las posibles causas de ese arreglo (Simberloff, 1983). La biogeografía cuantitativa es un conjunto de técnicas de análisis que apoya a la biogeografía, por lo que la biogeografía cuantitativa constituye un método, no una ciencia en sí.

La biogeografía cuantitativa cada vez se apoya más en el uso de las computadoras. Así, mientras las computadoras aumentan su poder de procesamiento y el número de usuarios a través de interfaces agradables y ergonómicas, se estimula el uso y generación de técnicas que se incorporan a la biogeografía cuantitativa.

El análisis de datos biogeográficos sólo es un paso dentro de un estudio biogeográfico, pues antes de éste se deben tener claros los objetivos del estudio y se debe realizar una serie de decisiones en cuanto a la estructura de los datos a ser analizados (Crovello, 1981). La misma recopilación e integración de datos es un paso importante previo al análisis. Además, después del análisis, los resultados se deben interpretar de acuerdo con una escuela o enfoque particular de la biogeografía, como puede ser el histórico o el ecológico. Como uno de los últimos pasos, deben publicarse los resultados.

El número de técnicas y análisis cuantitativos que se utilizan en biogeografía aumenta día a día. En el Cuadro I se listan los análisis básicos de la biogeografía cuantitativa; aunque la lista no es exhaustiva, sí se considera básica o fundamental. Por lo general, las técnicas son operables mediante programas de computadora; con la finalidad de ilustrar de manera homogénea esa lista de técnicas, se ha diseñado una 'herramienta teórica' que instrumenta los análisis básicos de la biogeografía cuantitativa. La herramienta teórica es el resultado del diseño de un paquete de computadora para la biogeografía cuantitativa y las figuras que se presentan son parte de la interfaz de ese paquete de cómputo. Para explicar las técnicas de análisis de la biogeografía cuantitativa listadas en el Cuadro I, se utilizará un ejercicio sobre los pinos de Chihuahua registrados en el Herbario Nacional Forestal (INIF). Los registros considerados son un subconjunto de una base de datos pública (Gutiérrez, 1997). Aunque la base de datos de los pinos de Chihuahua no se considera completa, en el sentido que represente la distribución del género en el estado, es útil para la ejemplificación, independientemente de que pueda ser mejorada mediante la consulta de otros acervos y la consulta con especialistas en el género.

## Métodos de la biogeografía cuantitativa

En esta sección se explican los principales procesos y análisis de la biogeografía cuantitativa presentados en el mismo orden que se listan en el Cuadro I.

**Definición del área de estudio.** En un estudio biogeográfico es importante definir el contexto, es decir, los objetivos, los grupos taxonómicos a incluir en el análisis, así como la propia dimensión geográfica. No es lo mismo realizar una comparación entre las biotas de dos localidades que hacer un análisis de la distribución de un taxón en un continente, cada estudio requerirá de métodos y herramientas de análisis diferentes.

En la figura 1 se muestra la definición del área de estudio para el ejemplo que se sigue a lo largo de este capítulo: un área rectangular

que abarca la mayor parte del territorio del estado de Chihuahua, dentro de la que se analizará la distribución de las especies de pinos.

**Definición de OGU.** Varias de las técnicas para realizar análisis biogeográficos se basan en una generalización geográfica de los registros de las recolectas de los organismos: donde se localiza un punto de un sitio de recolecta, se supone una distribución de los organismos ahí recolectados pero en niveles geográficos superiores. Por ejemplo, en un estudio biogeográfico se puede hacer la generalización de que en el municipio donde se encuentre la recolecta de un organismo se supone la distribución de ese organismo en todo ese municipio, es decir, se realiza una generalización a partir de puntos (localidades de recolecta) para conformar áreas.

En la biogeografía cuantitativa se analizan relaciones a partir de unidades geográficas, estas unidades son llamadas Unidades Geográficas Operativas, OGU por sus iniciales en inglés (Crovello, 1981), de manera análoga a las unidades analizadas en la taxonomía numérica, que son llamadas Unidades Taxonómicas Operativas, OTUs por sus iniciales en inglés (Sneath y Sokal, 1973). Las unidades geográficas así definidas son las unidades mínimas de análisis, representan el nivel máximo de resolución en el análisis, pueden agregarse para conformar áreas de niveles superiores pero no desagregarse en unidades más pequeñas. Esa definición puede realizarse subdividiendo al área de estudio en las entidades políticas que contiene, por ejemplo los municipios, o bien en unidades de tamaño regulares, como cuadros de un grado de latitud por un grado de longitud.

En estudios biogeográficos es común la conformación de OGUs de cuadros de tamaño regular, ya que permiten obtener unidades con áreas de igual tamaño. Esto permite que su comparación en cuanto al número

**Cuadro I.** Lista de procesos y análisis básicos de la biogeografía cuantitativa.

1. Definición del área de estudio
2. Definición de las OGUs
3. Matriz de presencia/ ausencia
4. Índices de diversidad
  - 4.1. Diversidad Alfa
  - 4.2. Diversidad Alfa promedio
  - 4.3. Diversidad Beta
  - 4.4. Diversidad Gama
  - 4.5. Estimación de riqueza
5. Distribución geográfica de taxones
6. Listas de taxones de OGUs
7. Similitud biogeográfica
  - 7.1. Índices de similitud
  - 7.2. Matriz de similitud
8. Métodos de clasificación
  - 8.1. Análisis de agrupamientos
  - 8.2. Métodos divisivos
9. Métodos de ordenación
  - 9.1. Análisis de factores
10. Definición de áreas prioritarias para la conservación

de especies sea más válida que entre áreas de diferentes tamaños, pues en general, el número de especies aumenta con el tamaño del área.

Phipps (1975) propuso un método para elegir el tamaño de las OGU en estudios biogeográficos. Sin embargo, cualquiera sea el tamaño de OGU que se elija, es importante indicar que dicho tamaño afecta a la varianza de los datos que se agregan (Wong, 1996), por lo que algunos parámetros biogeográficos, como por ejemplo la diversidad alfa promedio, no son comparables si se obtienen de diferentes tamaños de OGU.

En la figura 1 se muestra el área de estudio, con la subdivisión en cuadros de un grado de lado que conforman a las OGU.

**Construcción de la matriz de presencia-ausencia.** Uno de los primeros pasos para realizar un análisis biogeográfico cuantitativo es construir una matriz de presencia y ausencia de los taxones en cada una de las OGU o subregiones del área de estudio. La estructura de la matriz de presencia-ausencia para realizar análisis biogeográficos se compone de la lista de taxones ( renglones) por la lista de OGU (columnas).

La matriz de presencia-ausencia muestra todas las combinaciones de taxones contra OGU. Cada entrada o celda de la matriz se llena con un '0' o un '1', indicando '0' que la especie está ausente en esa OGU y '1' que está presente.

En la construcción de la matriz de presencia-ausencia no se consideran datos de abundancia, y basta un registro del taxón dentro de la OGU para indicar presencia en la matriz. La matriz de presencia-ausencia es una alternativa al análisis de información cuando no se cuenta con datos de abundancia. Muchas de las herramientas de análisis biogeográfico, como los índices de diversidad, los índices de similitud o los estimadores de riqueza diferencian entre estos dos niveles de información: unas fórmulas están solo diseñadas para trabajar con datos de abundancia, mientras que otras solo requieren de datos de presencia-ausencia. A lo largo de este capítulo se presentan herramientas de análisis orientadas a aplicarse sobre datos de presencia-ausencia.

En el caso de los pinos de Chihuahua registrados en el Herbario Nacional Forestal (INIF) se cuenta con 16 especies, y el área de estudio se ha subdividido en 80 OGU de las cuales solo 28 contienen registros de pinos. Así, la matriz de presencia ausencia está compuesta de 16 renglones y 28 columnas.

**Índices de diversidad.** Existen varias propuestas para definir y cuantificar la diversidad (especies y grupos supraespecíficos). Con fines de esquematizar algunas de las propuestas de cuantificación de la diversidad se utiliza la siguiente clasificación:

a (diversidad alfa): número de taxones en una OGU.

b (diversidad beta): tasa de recambio de taxones entre las OGU.

g (diversidad gama): número de taxones en toda el área de estudio.

La cuantificación de la diversidad se utiliza para caracterizar a una región geográfica en términos de su biota, ya sea considerando uno o varios grupos taxonómicos, también es útil para poder establecer comparaciones, ya sea entre unidades geográficas, grupos taxonómicos o biotas. Pueden considerarse como algunos parámetros descriptivos con los que se pueden realizar análisis biogeográficos.

**Diversidad alfa.** La diversidad alfa se asocia con el número de especies (riqueza) en una comunidad, es decir, en un área biológicamente homogénea. De manera técnica, puede considerarse a la diversidad alfa como el número de especies en cada OGU, mientras que a la diversidad gama, como el número de especies en toda el área de estudio. En la figura 2 se muestra el mapa de la diversidad alfa, es decir, el número de especies de pinos en cada una de las OGU.

A partir de la diversidad alfa de cada OGU se puede calcular la "diversidad alfa promedio": el promedio de las diversidades alfa, que puede expresarse mediante la siguiente igualdad:

$$\alpha_{\text{prom}} = (\sum_{i=1}^n a_i) / n$$

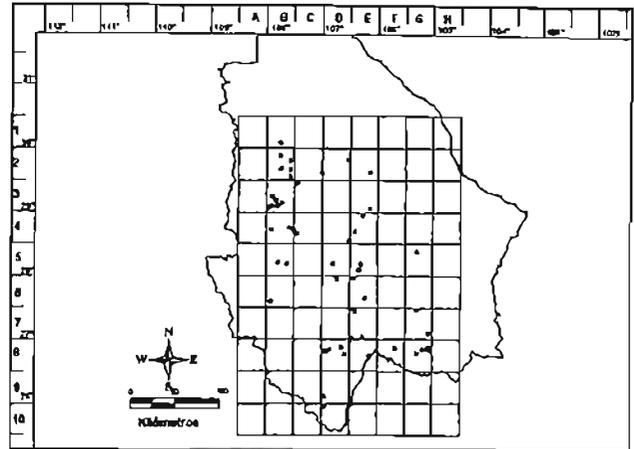


Fig. 1. Área de estudio (territorio del estado de Chihuahua) mostrando la subdivisión en cuadros de un grado por un grado y la ubicación de las localidades de recolección de pinos registradas en el herbario INIF.

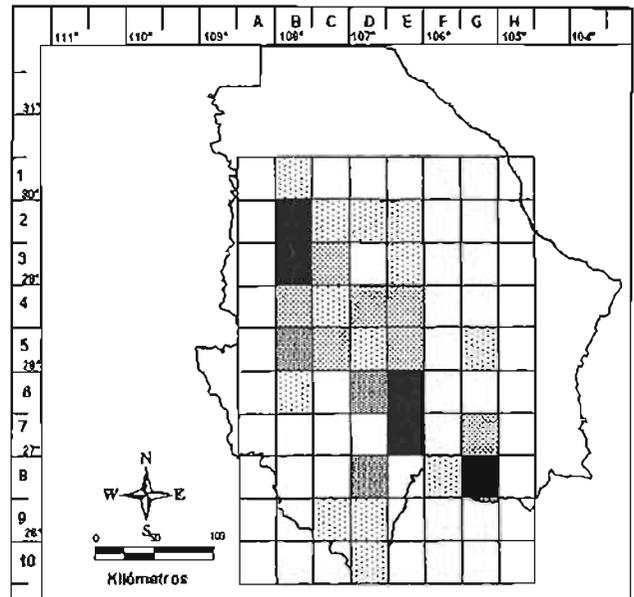


Fig. 2. Mapa temático de la diversidad alfa de cada cuadro de la figura 1.

donde  $\alpha_{\text{prom}}$  es la diversidad alfa promedio,  $a_i$  es la diversidad alfa en la  $i$ -ésima OGU y  $n$  es el número de OGU en el área de estudio. Para el ejemplo de los pinos de Chihuahua:

$$\alpha_{\text{prom}} = 64/28 = 2.29$$

esto es, en promedio existen 2.29 especies de pinos en cada OGU del área de estudio.

**Diversidad gama.** La diversidad gama es el número de especies en una región geográfica amplia (Whittaker, 1972). Para efectos técnicos, se puede considerar que el número de taxones en el área de estudio es el número de columnas de la matriz de presencia-ausencia. La diversidad alfa de las OGU siempre es menor o igual a la diversidad gama. Para el ejemplo de los pinos de Chihuahua, la diversidad gama es de 15.

**Diversidad beta.** La diversidad beta es una medida de la diferencia entre áreas de estudio en términos de la variedad de especies

que contienen (Magurran, 1988). La diversidad beta junto con la diversidad alfa constituyen una buena medida del grado de heterogeneidad biológica en el área de estudio. Una manera de observar a la diversidad beta es como una medida del remplazo de especies o del cambio biótico a lo largo de un gradiente (Wilson y Smida, 1984). Intuitivamente es posible admitir que, si las OGU contienen las mismas especies, la diversidad beta será baja, mientras que si las OGU presentan entre sí diferentes especies, la diversidad beta será alta. Existen varias propuestas para cuantificar a la diversidad beta, a continuación se explican algunas de ellas

Whittaker (1972), quien introdujo el concepto de diversidad beta, la define como el recambio de especies entre hábitats o localidades y propuso que la diversidad en un área amplia (gama) puede dividirse en dos componentes, en la diversidad alfa y beta, que puede expresarse mediante la siguiente igualdad:

$$\gamma = \beta_w \times \alpha_{prom}$$

donde  $\gamma$  es la diversidad gama y  $\alpha_{prom}$  es la diversidad alfa promedio, por lo que la diversidad beta puede expresarse como:

$$\beta_w = \gamma / \alpha_{prom}$$

Para el ejemplo de los pinos de Chihuahua,

$$\begin{aligned}\alpha_{prom} &= 2.29 \text{ especies/cuadro} \\ \gamma &= 15 \text{ especies/ 28 cuadros} \\ \beta_w &= 15 / (2.29 \times 28) = 0.23\end{aligned}$$

Otra propuesta para calcular la diversidad beta es la de Routledge:

$$\beta_r = \gamma^2 / (2r + \gamma) - 1$$

donde  $r$  es el número de pares de especies con distribuciones superpuestas. Magurran (1988) propuso calcular  $r$  como el número de pares de especies cuya distribución coincide cuando menos en una OGU. Para el ejemplo,

$$\begin{aligned}r &= 57 \\ \beta_r &= 16^2 / (2 \times 57 + 16) - 1 = 0.97\end{aligned}$$

Los valores de la diversidad beta calculados anteriormente son útiles cuando se comparan con otros análogamente calculados, por ejemplo para otro grupo taxonómico o para otra región geográfica.

**Estimadores de riqueza.** Como ya se comentó con anterioridad, la diversidad gama, técnicamente es el número de taxones en el área de estudio. En realidad, el cálculo que se basa en la definición anterior no obtiene la diversidad gama propiamente, sino que es una estimación, ya que la matriz de presencia-ausencia puede ser incompleta, es decir, pueden existir taxones en el área de estudio que no estén representados en los datos. Con la finalidad de obtener un valor de gama más aproximado al real, se han diseñado 'estimadores de riqueza', que son herramientas que permiten calcular un nuevo valor de gama a partir de los valores de las OGU.

Los estimadores de riqueza se pueden clasificar en dos grandes grupos: los paramétricos y los no paramétricos. La estrategia de cálculo de los primeros se basa en estimar tendencias, generalmente de tipo inductivo, mientras que la de los no paramétricos se basa en el establecimiento de relaciones y proporciones sin avenguar el siguiente valor posible en una serie ordenada.

Existe una amplia variedad de propuestas para estimar la riqueza. Colwell y Coddington (1994) presentaron un resumen de este tipo de técnicas. Por lo general, los estimadores presuponen que el número de especies observado es una subestimación, por lo que los resultados de

aplicar las fórmulas proporcionarán valores de gama estimados por arriba de los observados. A continuación se explican dos propuestas para estimar la riqueza, una mediante un método paramétrico y otra mediante uno no paramétrico.

Una estrategia para estimar la riqueza de una región mediante extrapolación es utilizando la ecuación de Michaelis-Menten:

$$S(n) = S_{max} \cdot n / (B + n)$$

donde  $S_{max}$  y  $B$  son constantes,  $n$  es el número de OGU y  $S(n)$  es el número de especies en las  $n$  OGU. La constante  $S_{max}$  es el estimador de riqueza y para obtenerlo es necesario realizar un ajuste lineal usando una transformación de la ecuación sobre pares de datos ( $S, n$ ), es decir, sobre una lista de número de especies observadas en  $n$  OGU, donde  $n$  varía desde 1 hasta el número de OGU del área de estudio.

El estimador no paramétrico de Chao ( $E_{Ch}$ ) se basa en el número de especies con distribución restringida dentro del área de estudio:

$$E_{Ch} = S_{obr} + Q_1^2 / (2 Q_2)$$

donde  $S_{obr}$  es el número de especies observadas en el área de estudio,  $Q_1$  es el número de especies que se encuentran en solo una OGU y  $Q_2$  es el número de especies que se encuentran en dos OGU. Para el ejemplo de los pinos:

$$\begin{aligned}S_{obr} &= 15 \\ Q_1 &= 5 \\ Q_2 &= 2 \\ E_{Ch} &= 15 + 5^2 / (2 \times 2) = 21.25\end{aligned}$$

por lo que el estimador indica que la diversidad gama del área de estudio es de 21, es decir, se estima que faltan por recolectar seis especies de pinos en la región.

**Distribución geográfica de taxones.** Una vez definidas las OGU, para el biogeógrafo es interesante observar la distribución de los taxones en términos de OGU y no simplemente de recolectas, como se muestra en la figura 1. La distribución en términos de OGU es una generalización, y visualmente permite o facilita la generación de hipótesis o preguntas al investigador. Preguntas como ¿en qué OGU se distribuye un taxón?, y ¿en qué OGU coincide la distribución de dos taxones en particular? son interesantes. En la figura 3 se muestra la distribución de *Pinus chihuahuana* en términos de OGU.

Las respuestas a preguntas biogeográficas básicas, como las enunciadas anteriormente, brindan al investigador herramientas empíricas que permiten el análisis exploratorio de datos, es decir, aunque no contribuyen directamente a contestar la pregunta biogeográfica planteada en los objetivos de la investigación, si lo hacen indirectamente. El investigador 'juega' con los datos para familiarizarse con ellos y emitir juicios sobre los resultados de los análisis, así como para conocer los límites y alcances de la información recolectada y de los propios métodos cuantitativos de análisis.

**Distribución potencial de especies.** La distribución de una especie que se deduce a partir de un muestreo, como son los registros de recolectas, en general representa un subconjunto de la distribución real de la especie. Recientemente se han propuesto métodos para definir la distribución 'potencial' de las especies, a partir de los registros de recolecta y de las variables ambientales asociadas a dichos sitios.

Un ejemplo de método para estimar distribución potencial es el DOMAIN (Carpenter *et al.*, 1993), que se basa en definir los intervalos de varios parámetros (*v. gr.* altitud, temperaturas mínimas, precipitación u otros tópicos geográficos) y reunirlos para formar un intervalo en un espacio multidimensional; es una manera de interpolar, utilizando los valores máximos y mínimos de las variables ambientales de los sitios en que se registran recolectas de la especie.

El mapeo de distribuciones potenciales de especies es una herramienta útil para el biogeógrafo ya que permite 'llenar' los vacíos que dejan los registros de recolecta.

**Listas de taxones de OGU.** ¿Qué taxones hay en determinada OGU? Ésta es una pregunta que también brinda información valiosa al biogeógrafo. La lista de especies puede deducirse a partir de la matriz de presencia/ausencia, sin embargo, el formato de presentación es una cuestión que por sí misma ayuda al investigador en la lectura de la información.

Una lista de gran importancia es la lista de las especies de toda el área de estudio. También las listas de cualquier combinación de dos o más OGU pueden tener significado valioso en el curso del análisis de la información por el investigador, por ejemplo, ¿qué especies hay en las OGU B2 y B3? (Fig. 4).

**Similitud biogeográfica.** La similitud biogeográfica se mide a través de la semejanza entre los componentes bióticos. Mientras mayor porcentaje de especies en común contienen dos áreas, mayor es su similitud biogeográfica.

Los criterios para medir la similitud biogeográfica toman en cuenta la lista de taxones entre las regiones a comparar; así, se realizan comparaciones entre dos regiones o áreas, no entre tres ni entre cuatro o más. Pueden tomar en cuenta o no la abundancia de cada uno de los taxones; como se indicó previamente, este capítulo centra la explicación en técnicas basadas en datos de presencia-ausencia, por lo que no se revisan criterios de semejanza que tomen en cuenta datos de abundancia.

**Índices de similitud.** Existen varios criterios de similitud, expresados en los 'índices de similitud', que son una manera de medir la semejanza entre dos conjuntos de taxones, áreas o biotas. Los índices de similitud se expresan mediante fórmulas. Si se concibe a una región geográfica como el conjunto de taxones que habitan en ella, entonces un índice de similitud permite medir la semejanza entre dos regiones o áreas geográficas en términos de los taxones que contienen.

En general todos siguen el criterio de que entre mayor sea el número de taxones en común entre dos áreas geográficas, mayor es su similitud. En el Cuadro II se dan las fórmulas de algunos de los índices de similitud más comúnmente usados.

Para ejemplificar la aplicación de los índices de similitud se utilizará el índice de Jaccard; este índice se calcula mediante la siguiente fórmula:

$$|A \cap B| / |A \cup B|$$

A continuación se ejemplifica la aplicación de la fórmula utilizando a los cuadros B2 y B3. Si se observa la matriz de presencia-ausencia (Cuadro II) se puede deducir que:

$|A \cap B|$  : Número de especies presentes en B2 y en B3: 4  
 $|A \cup B|$  : Número de especies presentes en B2 o en B3: 11

Así, aplicando la fórmula del índice de Jaccard:

$$I_{\text{Jaccard}} = 4/11 = 0.36$$

esto es, entre el cuadro B2 y el B3 hay una similitud, según el índice de Jaccard, de 36%.

Existe controversia sobre cuál índice usar para determinado tipo de análisis. Hubálek (1982) realiza un análisis que considera a 43 índices de similitud, y concluyen que los índices de Jaccard, Sorensen-Dice, Kulczynski y Ochiai son los que muestran un comportamiento que se ajusta más a ciertas premisas que propone como criterio de bondad.

Otro estudio comparativo realizado por Sánchez y López (1988) concluye que el índice de Simpson es el que muestra el mejor comportamiento de entre nueve considerados en el análisis. Hayek (1993) también incluye una discusión interesante sobre índices de similitud

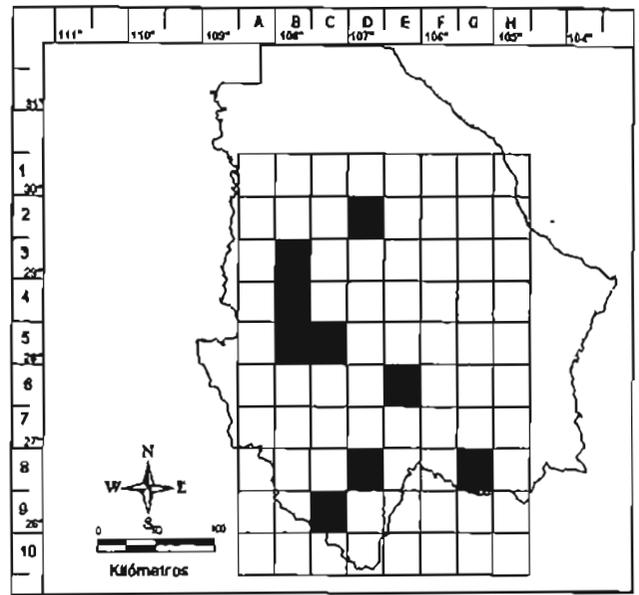


Fig. 3. Distribución de *Pinus chihuahuana* en las OGU definidas según la figura 1.

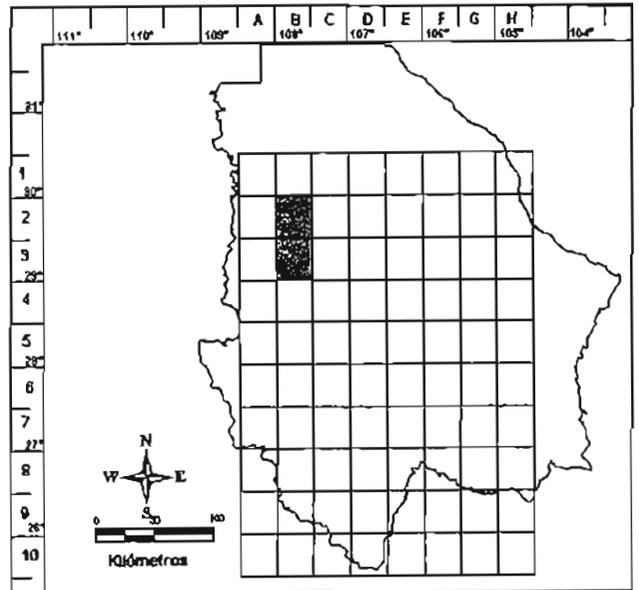


Fig. 4. Especies de las OGU B2 y B3.

**Matriz de similitud.** Los índices de similitud sirven para establecer una medida de semejanza entre dos OGU; la matriz de similitud contiene las comparaciones entre todos los pares posibles de OGU. La matriz de similitud tiene como encabezados de las columnas, así como de los renglones, a cada una de las OGU. Casi siempre, la diagonal de la matriz de similitud, que contiene la comparación entre sí de cada una de las OGU, está conformada por números '1', ya que la mayoría de los índices de similitud proporcionan valores entre '0' y '1', representando '1' la máxima similitud.

La matriz de similitud es una técnica útil para medir la semejanza entre cada par de OGU, sin embargo, también es un paso previo necesario para realizar otro tipo de análisis, como los análisis de agrupamiento.

Los análisis biogeográficos cuantitativos se pueden clasificar en dos tipos: los de 'modo Q' y los de 'modo R' (Simberloff y Connor, 1979, in Birks, 1987). Los análisis de modo Q exploran las relaciones entre OGU a partir de los taxones que en ellos se distribuyen, mientras que los de

**Cuadro II.** Fórmulas de algunos índices de similitud.

**Notación**

$A \cap B$ : Intersección (elementos en común en A y B)

$A \cup B$ : Unión (elementos que se encuentran en A o en B)

$|A|$ : Cardinalidad (número de elementos en un conjunto)

$A^c$ : Complemento (elementos del universo que no están en A)

$\max(A, B)$ : Función: máximo de dos valores enteros

$\min(A, B)$ : Función: mínimo de dos valores enteros

**Fórmulas**

Baroni-Urbani-Buser

$$\frac{\sqrt{|A \cap B|} \times \{|A \cup B|\}}{|A \cap B|}$$

$$\sqrt{|A \cap B|} \times \{|A \cup B|\}$$

Braun-Blanquet

$$|A \cap B|$$

$$\max(|A|, |B|)$$

Fager

$$|A \cap B|$$

$$(|A| \times |B|)^{1/2} \cdot \frac{1}{2} \times \max(|A|, |B|)$$

Jaccard

$$|A \cap B|$$

$$|A \cup B|$$

Kulezynski

$$|A \cap B| \times |A \cup B|$$

$$2 \times (|A| \times |B|)$$

Ochiai

$$|A \cap B|$$

$$(|A| \times |B|)^{1/2}$$

Radio de correlación

$$|A \cap B|$$

$$|A| \times |B|$$

Simpson

$$|A \cap B|$$

$$\min(|A|, |B|)$$

Sorensen-Dice

$$2 \times |A \cap B|$$

$$|A| + |B|$$

modo R exploran las relaciones entre taxones a partir de sus datos de distribución. También es útil la construcción de una matriz de similitud entre taxones (modo R), a partir de la misma matriz de presencia-ausencia, pues muestra las relaciones de las especies con base en su distribución, es decir, mientras más alto sea el valor del índice de similitud entre un par de taxones, más coinciden en su distribución geográfica.

**Técnicas de clasificación.** Los enfoques multivariados en biogeografía cuantitativa constituyen un conjunto de técnicas estadísticas para el análisis simultáneo de más de una variable independiente. Además de las técnicas de clasificación, también se incluyen dentro de los enfoques multivariados al análisis de varianza y covarianza, y a la regresión y correlación múltiples.

Una de las clasificaciones más comunes de los métodos de análisis aplicados en taxonomía numérica y biogeografía cuantitativa es aquella que los divide en 'técnicas de clasificación' y en 'técnicas de ordenación'. Ambos tipos se pueden aplicar a una matriz de datos de presencia-ausencia.

Las técnicas de clasificación generan grupos que incluyen a las unidades analizadas, mientras que las de ordenación no establecen una delimitación discreta de las clases. Una de las técnicas de clasificación más comúnmente utilizadas es el 'agrupamiento por pares mediante ligamiento promedio' o UPGMA ('unweighted pair-group method using arithmetic averages').

Las técnicas de ordenación consideran un espacio multidimensional donde cada uno de los ejes representa a un atributo que en conjunto caracterizan a las OGU; estas técnicas reducen ese espacio a otro de dos o tres dimensiones sin mucha pérdida de información (Kohlmann, 1994; Sneath, 1998). Uno de los métodos más comúnmente utilizados en biogeografía y sistemática es el 'análisis de componentes principales' (Zavala, 1986). También la regresión múltiple se considera dentro de este grupo de técnicas.

Actualmente existen muchas herramientas informáticas para la aplicación de enfoques multivariados: Palmer (1998) dio una lista de programas de cómputo para su aplicación en biología, y Campbell (1989) explicó la aplicación de los mismos mediante el paquete estadístico SPSS.

**Análisis de agrupamiento.** Las técnicas de agrupamiento se pueden clasificar en divisivas y aglomerativas. En ambos casos se requiere de la matriz de presencia-ausencia, pero en las aglomerativas además se opera sobre la matriz de similitud, en la que, como ya se indicó, se registran las similitudes entre cada par de OGU y se construye a partir de la matriz de presencia-ausencia.

**Técnicas aglomerativas.** Dentro de las técnicas aglomerativas, las más comúnmente usadas son las secuenciales-jerárquicas, dentro de las que se encuentra el ligamiento simple, el ligamiento completo y el UPGMA.

El algoritmo inicia eligiendo un par de OGU, aquel cuya similitud es la mayor de todos los posibles pares; la matriz de similitud es recalculada, al considerar ahora como una unidad al par de OGU elegidos; entonces, la matriz que resulta contiene un renglón y una columna menos y se ha formado parte del fenograma: dos hojas unidas a un determinado nivel de similitud. En cada paso, la matriz de similitud es recalculada y va reduciéndose en un renglón y una columna; el algoritmo termina cuando todas las OGU han sido ubicadas en el fenograma.

La manera en que es recalculada la matriz de similitud hace la diferencia entre los tipos de ligamientos: simple, completo o promedio (UPGMA). En la figura 5 se muestra un fenograma resultado de la aplicación del algoritmo UPGMA.

El fenograma puede 'partirse' verticalmente a un 'nivel crítico' determinado para conformar grupos de OGU con características homogéneas. Según Sánchez y López (1988), ese nivel crítico es de 66.66% si se usa el índice de Simpson; como el índice de Jaccard proporciona valores iguales o por debajo que Simpson, entonces el nivel crítico al usar Jaccard es menor. En la figura 6 se muestra el mapa de los grupos de OGU formados a partir de aplicar un nivel crítico de 20% al fenograma de la figura 5.

Ejemplos de la aplicación del UPGMA en taxonomía vegetal se encuentran en los trabajos de Eguiarte (1995) y en los de Villaseñor y Sirother (1989), ejemplos de aplicación en biogeografía ecológica en los de León y Romo (1993), Sánchez (1993) y Villareal *et al.* (1996).

**Técnicas divisivas.** La estrategia de estas técnicas es iniciar la formación de dos grupos a partir del conjunto total de datos; se repite esto para cada uno de los dos grupos y así sucesivamente hasta llegar a grupos unitarios (Hengeveld, 1990), es decir, hasta que solo contengan una OGU.

El criterio para dividir un grupo en dos, generalmente depende de una función de optimización. Así, se exploran varias combinaciones de los posibles dos grupos y se elige la que cumpla mejor con el criterio de optimización. El análisis de información (Sneath y Sokal, 1973) se puede utilizar para la formación de grupos ya sea divisivamente o mediante agrupamiento. La idea principal es cuantificar el cambio de 'contenido de información',  $I$ , al pasar de un grupo a otros dos, generados de una partición del primero, que puede expresarse mediante la siguiente igualdad:

$$\Delta I = I(AB) - I(A) - I(B)$$

donde  $I$  es una función que mide el contenido de información en una matriz básica de datos,  $I(AB)$  es el contenido de información de la matriz,  $I(A)$  el de la primera partición de la matriz e  $I(B)$  el de la segunda.

Ezcurra *et al.* (1984) propusieron utilizar la siguiente función para medir el contenido de información:

$$I = 2sn \log n - n \sum_{i=1}^s \log [a_i^2 + (n-a_i)^2]$$

donde  $s$  es el número de atributos (especies),  $n$  es el número de unidades a clasificar (OGUs), y  $a_i$  es el número de unidades que tienen el atributo  $i$ . Cuando todas las especies se encuentran presentes o ausentes en todas las OGUs  $I$  vale cero, por lo que el método intenta buscar particiones de la matriz que contengan valores  $I$  bajos. La estrategia propuesta para partir la matriz en dos submatrices es fijar un atributo (especie) y con base en él agrupar en la primera submatriz aquellas OGUs que presenten al atributo y en la segunda a aquellas OGUs que no la presenten, la partición que se elige es aquella que minimice a  $\Delta I$ .

En la figura 7 se muestra el resultado de aplicar el método divisivo propuesto por Ezcurra *et al.* (1984) a la matriz de presencia-ausencia de los pinos de Chihuahua.

Obsérvese que la raíz del árbol está etiquetada con 48.63, que es el valor de la función de contenido de información de toda la matriz de presencia-ausencia; el atributo que minimiza la suma de  $I$  es la primera especie; la submatriz definida por la presencia de la especie  $I$  tiene un valor  $I = 25.01$ , y la segunda, definida por la ausencia de la misma especie  $I$  tiene un valor  $I = 13.01$ . Así, el cambio de contenido de información  $\Delta I$  es:

$$\Delta I = 48.63 - 25.01 - 13.01 = 10.61$$

**Técnicas de ordenación.** Éstas no trazan límites que definan grupos, esa tarea se deja al investigador al interpretar los resultados. Las relaciones entre las OGUs o taxones (según sea una análisis Q o R) están reflejadas por la posición en un espacio multidimensional, por ejemplo de caracteres: mientras más cercanas están, mayor relación tienen entre sí.

El análisis de componentes principales (PCA por sus iniciales en inglés) reduce la complejidad disminuyendo el número de dimensiones del espacio de las OGUs. El PCA efectúa una transformación de los ejes originales obteniendo otro conjunto de ejes: los componentes. El número de componentes es igual al número de ejes originales, pero solo algunos de ellos, los 'principales', capturan más información de los ejes originales. Esos componentes principales son útiles para representar el espacio multidimensional original (Gauch, 1981).

En el análisis se obtiene un 'eigenvalor' para cada componente. Los eigenvalores están relacionados con la varianza explicada por cada eje y

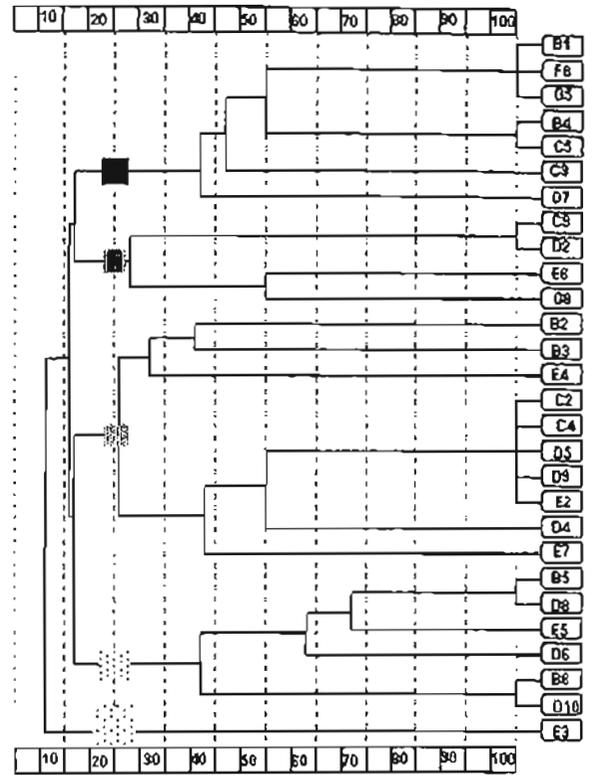


Fig. 5. Fenograma de similitud (UPGMA) de los pinos de Chihuahua.

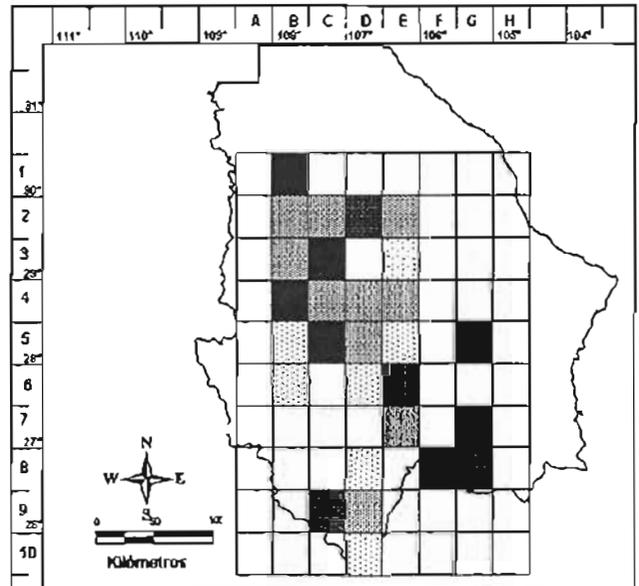


Fig. 6. Mapa mostrando los grupos de OGUs formados a partir de aplicar un nivel crítico de 20% al fenograma de la figura 5.

generalmente se expresan en porcentajes; se ordenan del mayor valor al menor y se obtienen sumas acumuladas, para observar, por ejemplo, cuántos componentes principales explican el 90% de la variación.

El análisis de factores es otra técnica de ordenación, muy parecida al PCA. Una de las principales diferencias es que el análisis de factores tiene la ventaja de permitir que los factores no sean ortogonales (Gauch, 1981; Sneath y Sokal, 1973).

Arita (1993) incluyó un análisis de factores en un estudio sobre los mamíferos de México, en el que encuentra que la riqueza aumenta con

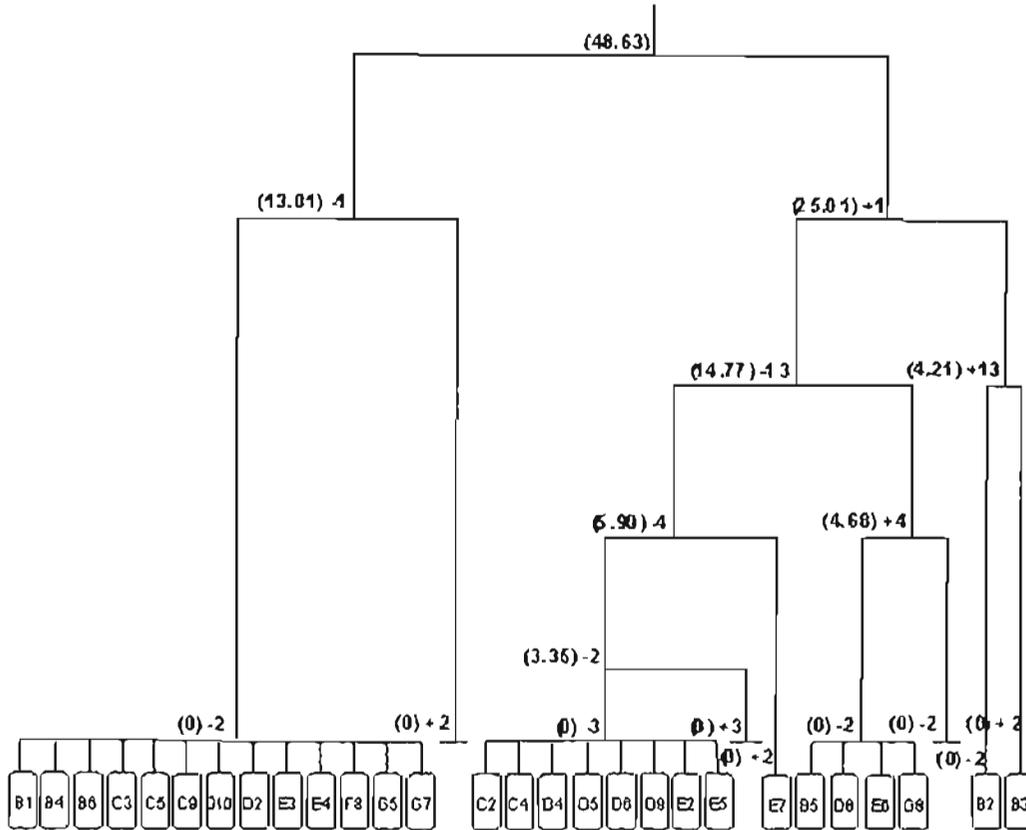


Fig. 7. Fenograma resultado de aplicar el método divisivo de Ezcurra *et al* (1984) a la matriz de presencia-ausencia.

la altitud y la temperatura e inversamente con la latitud. Palmer (1998) proporcionó una lista de recursos relacionados con las técnicas de ordenación, que incluye ejemplos, programas de cómputo, listas de discusión y breves explicaciones. Crisci y López (1983) explican didácticamente varias técnicas de ordenación usadas en taxonomía numérica.

#### Definición de áreas prioritarias para la conservación.

La definición de áreas bióticas prioritarias para la conservación es un tema que actualmente ocupa a muchos investigadores. Se han propuesto numerosos procedimientos y consideraciones para su definición, entre los que se pueden distinguir dos grupos, los taxonómicos y los cladísticos. Los primeros contribuyen en definir las áreas a proteger con base en las listas florísticas o faunísticas, al reconocer el mayor número de especies o taxones y su endemismo. Los cladistas prestan atención especial a la filogenia de los organismos, con la finalidad de proteger la mayor heterogeneidad genética. A continuación se explican dos métodos basados en listas bióticas.

Vane-Wright *et al* (1991) propusieron el siguiente método para definir la prioridad de las áreas a proteger:

1. Se selecciona la unidad (OGU) con mayor número de especies, las especies contenidas se eliminan del análisis.
2. El procedimiento se repite con las especies restantes (complemento) que no se han incluido en las unidades ya seleccionadas.
3. Cuando en alguna iteración más de una unidad tiene el mismo número de especies del complemento, se selecciona aquella con el mayor número de especies totales. Si aún así, más de una OGU cumple esa condición, se selecciona la primera, según un orden arbitrario.
4. El procedimiento termina cuando todas las especies estén incluidas en alguna OGU seleccionada (complemento igual a cero).

En la figura 8 se muestra el resultado de aplicar el procedimiento de Vane-Wright *et al* (1991) a la matriz de presencia-ausencia de la figura 9, en la cual se aplica el método en cinco ciclos. En el ciclo uno se elige a la

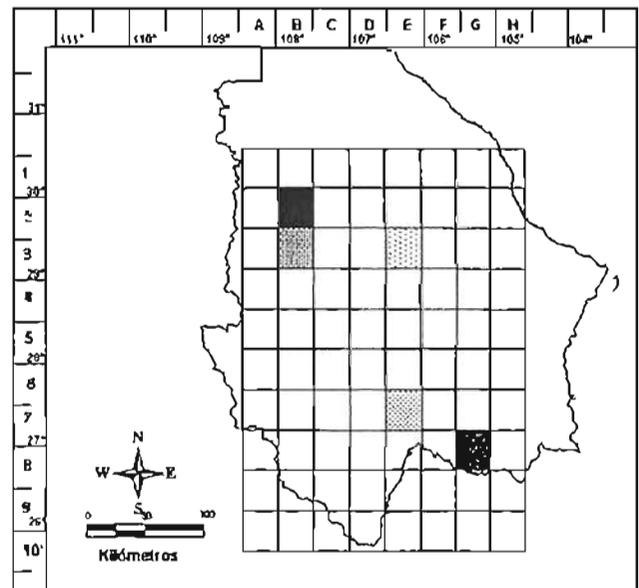


Fig. 8. Resultado del método de Vane-Wright *et al* (1991) para la definición de áreas prioritarias para la conservación.

OGU B2 como la OGU con más alta prioridad para la conservación, ya que es la que contiene el mayor número de especies (ocho); en el ciclo dos se trabaja con un subconjunto de especies resultado de eliminar las especies que ya están representadas en la OGU B2. En el ciclo dos se recalcula el número de especies en cada OGU resultando la OGU G8 la que contiene mayor número de especies (cuatro). En los ciclos 3, 4 y 5

## Paso 1

ESPECIE	ID	B1	B2	B3	B4	B5	B6	C2	C3	C4	C5	C9	D10	D2	D4	D5	D6	D8	D9	E2	E3	E4	E5	E6	E7	F8	G5	G7	G8
Pinus arizonica	1	0	1	0	1	0	1	0	1	0	0	0	0	0	1	1	1	1	1	1	0	0	1	1	1	0	0	0	1
Pinus ayacahuite	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
Pinus cembroides	3	1	0	0	1	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1
Pinus chihuahuana	4	0	1	1	1	0	0	0	0	1	1	0	1	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	1
Pinus cooperi	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Pinus douglasiana	6	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Pinus durangensis	7	0	1	0	1	1	0	0	0	0	0	1	0	0	0	1	1	0	0	0	0	1	0	0	0	0	0	0	0
Pinus engelmanni	8	0	1	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0
Pinus herrerae	9	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1
Pinus latifolia	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
Pinus lumholtzii	11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	1
Pinus pseudostrub.	12	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Pinus reflexa	13	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
Pinus strobiformis	14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
Pinus teocote	15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
		1	6	7	2	3	1	1	2	1	2	1	1	1	2	1	3	3	1	1	1	2	2	4	4	1	1	2	5

## Paso 2

ESPECIE	ID	B1	B2	B3	B4	B5	B6	C2	C3	C4	C5	C9	D10	D2	D4	D5	D6	D8	D9	E2	E3	E4	E5	E6	E7	F8	G5	G7	G8
Pinus cembroides	3	1	0	0	1	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1
Pinus chihuahuana	4	0	0	1	1	1	0	0	0	0	1	1	0	1	0	0	0	1	0	0	0	0	0	1	0	0	0	0	1
Pinus douglasiana	6	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Pinus herrerae	9	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1
Pinus latifolia	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
Pinus lumholtzii	11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	1
Pinus teocote	15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
		1	0	3	2	1	0	0	1	0	2	1	0	1	0	0	1	1	0	0	1	0	0	2	1	1	1	2	4

## Paso 3

ESPECIE	ID	B1	B2	B3	B4	B5	B6	C2	C3	C4	C5	C9	D10	D2	D4	D5	D6	D8	D9	E2	E3	E4	E5	E6	E7	F8	G5	G7	G8
Pinus douglasiana	6	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Pinus latifolia	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
Pinus teocote	15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
		0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0

## Paso 4

ESPECIE	ID	B1	B2	B3	B4	B5	B6	C2	C3	C4	C5	C9	D10	D2	D4	D5	D6	D8	D9	E2	E3	E4	E5	E6	E7	F8	G5	G7	G8
Pinus latifolia	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
Pinus teocote	15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0

## Paso 5

ESPECIE	ID	B1	B2	B3	B4	B5	B6	C2	C3	C4	C5	C9	D10	D2	D4	D5	D6	D8	D9	E2	E3	E4	E5	E6	E7	F8	G5	G7	G8
Pinus latifolia	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Fig. 9. Aplicación del método de Vane-Wright *et al.* (1993) para la definición de áreas prioritarias para la conservación (ver texto y figura 8).

las OGU B3, E3 y E7 contienen igual número de especies (uno), entonces se eligen según el número de especies totales (paso 1).

Así, de acuerdo con el método de Vane-Wright *et al.* (1991), las OGU que deberían ser elegidas como prioritarias para la conservación son B2, G8, B3, E7 y E3, en ese orden. Obsérvese que en este caso la totalidad de las especies quedan representadas en las cinco OGU elegidas.

Otro procedimiento, que fue propuesto por Rebelo (1994), considera un índice de 'endemicidad', el cual es la suma de los 'valores de rareza' de cada especie o taxón. Las unidades con mayor índice de endemicidad se seleccionan primero; después se recalculan los índices de endemicidad para las OGU restantes considerando solo a los taxones que no han sido incluidos en alguna OGU seleccionada. El procedimiento termina cuando todos los taxones han quedado incluidos en alguna OGU seleccionada como prioritaria para la conservación.

Villateñor *et al.* (1998) aplicaron el método de Rebelo a los géneros de compuestas de México, utilizando como OGU a las entidades federativas del país, resultando los estados de Baja California Norte, Chiapas y Coahuila como las tres entidades prioritarias para su conservación.

## Conclusiones

La biogeografía cuantitativa provee un conjunto de técnicas útiles en la descripción y análisis biogeográficos. Como en casi cualquier otra área del conocimiento, dentro de la biogeografía cuantitativa se tienen varios problemas por resolver, como por ejemplo, el establecimiento de criterios de comparabilidad entre técnicas del mismo tipo, *v. gr.* entre diferentes técnicas de agrupamiento, o incluso, entre diferentes índices de similitud.

Las computadoras desempeñan un papel importante, ya que proveen de programas que automatizan los métodos o parte de ellos. Los resultados de esas técnicas, que generalmente provienen de análisis numéricos, pueden ser de diversos tipos, ya que pueden expresarse, también de modo numérico o bien esquemáticamente, como los árboles que expresan relaciones de semejanza, o incluso mediante mapas, como se ejemplifica a lo largo de este capítulo.

## Agradecimientos

Agradecemos a Marcela Gutiérrez por la revisión en el Herbario Nacional Forestal para verificar la información contenida en la Base de

Datos del herbario (Gutiérrez, 1997) y a los editores Jorge Llorente y Juan J. Morrone por la revisión del manuscrito.

## Referencias

- Arita, H.** 1993. Riqueza de especies de la mastofauna de México, pp. in: Medellín, R. A. y Cevallos, G. (eds.), *Avances en el estudio de los mamíferos de México*, Asoc. Mex. de Mastozoología, México.
- Bricks, H. J. B.** 1987. Recent methodological development in descriptive biogeography. *Ann. Zool. Fennici*, 24: 165-178.
- Campbell, R. C.** 1989. *Statistics for biologists*. Cambridge University Press, Cambridge.
- Carpenter, G., A. N. Gillison y J. Winter.** 1993. DOMAIN: A flexible modeling procedure for mapping potential distributions of plants and animals. *Biodiv. Conserv.*, 2: 667-680.
- Colwell, R. K. y J. A. Coddington.** 1994. Estimating terrestrial biodiversity through extrapolation. *Phil. Trans. R. Soc. London*, B 345: 101-118.
- Crisci, J. V. y M. F. López.** 1983. *Introducción a la teoría y práctica de la taxonomía numérica*. OEA, Washington, D.C.
- Crovello, Th. J.** 1981. Quantitative biogeography: An overview. *Taxon*, 30(3):563-575.
- Eguarte, L.** 1995. Hutchinson (Agavales) vs. Huber y Dahlgren (Asparagales): Análisis moleculares sobre la filogenia y evolución de la familia Agavaceae sensu Hutchinson dentro de las monocotiledóneas. *Bol. Soc. Bot. México*, 56: 45-56.
- Ezcurra, E., M. Equihua, B. Kohlmann y S. Sánchez-Colón.** 1984. Métodos cuantitativos en la biogeografía. Instituto de Ecología, A.C.
- Gauch, H. G.** 1981. *Multivariate analysis in community ecology*. Cambridge University Press, Cambridge.
- Gutiérrez, M.** 1991. *Catálogo Florístico de México del Herbario Nacional Forestal "Biól. Luciano Vela Gálvez" (INIF)*. Publicación especial No. 4. 40p y disco de computadora.
- Hayek, L. A. C.** 1993. Analysis of amphibian biodiversity data, in: Heyer, W. R., M. A. Donnelly, R. W. McDiarmid, L. A. C. Hayek y M. S. Foster (eds.), *Measuring and monitoring biological diversity: Standard methods for amphibians*, Smithsonian Institution Press, Washington, D.C.
- Hengeveld, R.** 1990. *Dynamic biogeography*. Cambridge University Press, Cambridge.
- Hubálek, Z.** 1982. Coefficients of association and similarity, based on binary (presence-absence) data: an evaluation. *Biol. Rev.*, 57: 669-689.
- Kohlmann, B.** 1994. Algunos aspectos de la taxonomía numérica y sus usos en México, pp. in: Llorente, J. e I. Luna (eds.), *Taxonomía biológica*. UNAM-Fondo de Cultura Económica, México, D.F.
- León, L. y E. Romo.** 1993. Mastofauna de la sierra de Taxco, Guerrero, pp. in: Medellín, R.A. y Cevallos, G. (eds.), *Avances en el estudio de los mamíferos de México*, Asociación Mexicana de Mastozoología, México, D.F.
- Magurran, J.** 1988. *Ecological diversity and its measurement*. Princeton University Press, New Jersey.
- Palmer, M.** 1998. *Ordination methods for ecology*. Department of Botany, Oklahoma State Univ. <http://www.okstate.edu/artsci/botany/ordinate> (Junio/1998)
- Phips, J. B.** 1975. BestBlock: Optimizing grid size in biogeographic studies. *Can. J. Bot.*, 53: 1447-1452.
- Rebello, A. G.** 1994. Iterative selection procedures: Centres of endemism and optimal placement of reserves. *Stertia*, 1: 231-257.
- Sánchez, O.** 1993. Tendencias ecogeográficas del género *Reithrodontomys* (Rodentia: Muridae) en México, pp. in: Medellín, R. A. y G. Cevallos (eds.), *Avances en el estudio de los mamíferos de México*, Asoc. Mex. de Mastozoología, México.
- Sánchez, O. y G. López.** 1988. A theoretical analysis of some indices of similarity as applied to biogeography. *Folia Ent. Mex.*, 75: 119-145.
- Simberloff, D.** 1983. Biogeography: The unification and maturation of a science, pp. 411-473 in: Brush, A. H. y G. H. Clark (eds.), *Perspectives in ornithology*, Cambridge University Press, Cambridge.
- Sneath, P. H.** 1998. Development of computer-based systems in systematics, in: Bridge, P., P. Jeffries, D.R. Morse y P.R. Scott (eds.), *Information technology, plant pathology and biodiversity*, CAB International, Londres.
- Sneath, P. H. A. y R. R. Sokal.** 1973. *Numerical taxonomy*. W.H. Freeman, San Francisco.
- Vane-Wright, R. I., C. J. Humphries y P. H. Williams.** 1991. What to protect? - Systematics and the agony of choice. *Biol. Conserv.*, 55: 235-254.
- Villarreal, J. A., J. Valdés y J. L. Villaseñor.** 1996. Corología de las asteráceas de Coahuila, México. *Acta Bot. Mex.*, 36: 29-42.
- Villaseñor, J. L., G. Ibarra y D. Ocaña.** 1998. Strategies for the conservation of Asteraceae in Mexico. *Cons. Biol.*, 12(5): 1066-1075.
- Villaseñor, J. L. y J. L. Strother.** 1989. *Tuxtla*, a new genus for *Zexmenia pittieri* (Compositae: Heliantheae). *Syst. Bot.*, 14(4): 529-540.
- Whittaker, R. H.** 1972. Evolution and measurement of species diversity. *Taxon*, 21: 213-251.
- Wilson, M. V. y A. Shmida.** 1984. Measuring beta diversity with presence-absence data. *J. Ecol.*, 72: 1055-1064.
- Wong, D.** 1996. Aggregation Effects in Geo-referenced Data. in: Arlinghaus, S. L. (ed.), *Practical handbook of spatial statistics*, CRC Press, Boca Raton.
- Zavala, J. A.** 1986. *Introducción al enfoque multivariado en estudios de vegetación*. INIREB, Xalapa.

## Fe de erratas

Página 41:

La fórmula de la diversidad beta de Whittaker (1972) es:

$$B_w = \gamma / \alpha_{\text{prom}}$$

Sin embargo, al aplicar la fórmula en el ejemplo, se considera el número de cuadros, aplicando realmente la siguiente fórmula:

$$B = \gamma / (\alpha_{\text{prom}} \cdot N)$$

donde  $N$  es el número de cuadros. La doctora Patricia Koleff me ha indicado que en sentido estricto no es la fórmula de Whittaker. Koleff (2002) reporta una fórmula parecida:

$$B_{-1} = [B_w - 1 / (N - 1)] \quad \text{Harrison et al. (1992) en Koleff (2002)}$$

Página 41:

Dice:

$$\beta_R = 16^2 / (2 \times 57 + 16) - 1 = 0.97$$

Debe decir:

$$\beta_R = 15^2 / (2 \times 57 + 15) - 1 = 0.744$$

## 7. Estimación de la calidad de los registros biogeográficos

---

*En este capítulo se presentan los resultados de analizar el problema de estimar la calidad de los registros biogeográficos así como el de la definición del tamaño de la OGU.*

Este capítulo es parte de los resultados de este trabajo en lo referente a los objetivos sobre:

- 1.1) Analizar el problema de definir el tamaño de la OGU.
- 1.2) Analizar el problema de cuantificar la calidad de la información de los registros biogeográficos.

El capítulo, en su totalidad, es el artículo:

MURGUÍA, M. Y J. L. VILLASEÑOR, 2000. **“Estimating the quality of the records used in quantitative biogeography with presence-absence matrices”**. *Ann. Bot. Fennici*, 37: 289-296.

Cuya traducción de título al español es: *‘Estimación de la calidad de los registros usados en biogeografía cuantitativa con matrices de presencia-ausencia’*

En él se abordan, de manera conjunta, el problema de definir el tamaño de la OGU y el problema de estimar la calidad de la información contenida en los registros biogeográficos. En el artículo se presenta un índice (Qs) para estimar la calidad de los registros. El índice se construye con la filosofía de que al ganar resolución (tamaño de OGU pequeña) se pierde precisión.

En la tabla 7.1 se muestran los valores umbrales del índice Qs sobre los que se puede considerar una buena representatividad de datos biogeográficos. Se observan las variaciones en sentido positivo del nivel jerárquico en las dos dimensiones: la dimensión “nivel taxonómico” (especie, género y familia) y la dimensión “escala geográfica”, en dos situaciones: 32 OGU (estados) y 8 OGU (grupos de estados). Así, el índice Qs puede ser tomado como una medida (o función) de la precisión, pues cuando la resolución aumenta (tanto geográfica como taxonómicamente) Qs disminuye. Una medida de la precisión debe de ser reportada al realizar estudios que involucren el cambio de escala geográfica o taxonómica. Eso permitirá comparar la calidad de los registros de estudios biogeográficos de un grupo de organismo sobre regiones diferentes, o entre estudios de la misma región pero con diferentes taxones.

**Tabla 7.1** Valores umbrales (porcentajes) del índice Qs sobre los que se puede considerar una buena representatividad de datos biogeográficos.

	32 estados	8 grupos de estados
Especie	15	58
Género	50	70
Familia	60	90

Los principales resultados y conclusiones del artículo son:

- 1) El problema del “tamaño del cuadro” y del “tamaño y forma de la OGU” son una versión del llamado MAUP (de las iniciales en inglés de “Modifiable Area Unit Problem”) de los Sistemas de Información Geográfica.
- 2) La calidad de los registros biogeográficos puede ser medida mediante una comparación de la “riqueza estimada” contra la “riqueza conocida”.

- 3) La calidad de los registros biogeográficos puede ser medida además, mediante una combinación de dos factores: 1) una comparación de la “riqueza estimada” contra la “riqueza conocida”, y 2) considerando la suma de frecuencias en la matriz de presencia ausencia.
- 4) El índice Qs es una herramienta que mide la calidad de los registros biogeográficos con el criterio anterior.
- 5) Valores de Qs de 15%, 50% y 60% para especies, géneros y familias, respectivamente, se consideran como los mínimos de buenos índices de calidad para un mosaico de OGU de 32 estados de la República Mexicana.
- 6) La calidad de los registros biogeográficos está relacionada con el nivel de resolución: la calidad aumenta al disminuir la resolución (la resolución puede ser geográfica, taxonómica o de otra índole).
- 7) El índice Qs es una herramienta que permite comparar la calidad de los registros de estudios biogeográficos de un grupo de organismo sobre regiones diferentes, o entre estudios de la misma región pero con diferentes taxones.

# Estimating the quality of the records used in quantitative biogeography with presence–absence matrices

Miguel Murguía & José Luis Villaseñor

*Murguía, M. & Villaseñor, J. L., Instituto de Biología, Universidad Nacional Autónoma de México, Departamento de Botánica, Apartado postal 70-367, 04510 México, D.F. México*

*Received 13 March 2000, accepted 6 November 2000*

Murguía, M. & Villaseñor, J. L. 2000: Estimating the quality of the records used in quantitative biogeography with presence–absence matrices. — *Ann. Bot. Fennici* 37: 289–296.

We propose two new methods to estimate the error associated with data in presence–absence matrices commonly used in quantitative biogeographic analyses. The first is based on the estimated richness as compared with the known richness; the second is based on presence–absence frequencies addition in the data matrix. We propose an approach to integrate both criteria by using an index that determines the quality of the records as specified in the data matrix for each geographic region. It is assumed that the errors more commonly associated with the scoring of the biodiversity are those related to the absence of records in the subregions. On the other hand, it is also assumed that those mistakes caused by the scoring of elements in areas where they are not found are minimal. We propose to relate the quality of the records to the level of resolution or precision of the biogeographical analysis, assuming that the lower resolution of the analysis, the better the quality of the records. When a presence–absence data matrix is reordered, considering larger subregions and respectively a lower number of them, the possibility of errors due to lack of records of elements diminishes. This situation can be viewed as a particular case of the so-called modifiable areal unit problem (MAUP) of the geographic information systems.

Key words: biodiversity, biogeography, MAUP, presence–absence records, richness estimation

## INTRODUCTION

Quantitative biogeography (Crovello 1981, Bricks 1987) is an important approach in biogeographical analyses. It is based on the use of similarity or dissimilarity indices to compare the records of the biological elements in different regions or geographic units (Operational Geographic Units, or OGUs) with the ultimate goal of grouping the regions according to similarities or dissimilarities (e.g., Hagmeier & Stults 1966, McLaughlin 1989, Nims *et al.* 1995.) However, there are several decision problems that have not been successfully resolved when carrying out a biogeographical analysis of this kind (Bricks 1987). The choice of a similarity index (Hubálek 1982), the definition of the OGUs (Anderson & Marcus 1993, Wong 1996), and the choice of a clustering method (Bricks 1987, Crovello 1981) are among the decisions that should be made at the onset. Still another question to be answered is: is it really possible to apply a similarity index into the analysis? In other words, how reliable the results of the analysis based on the selected parameters will be (for example the number and shape of the OGUs and the selected similarity index).

There are several analyses concerning similarity indices behavior (*see* for example Hubálek 1982, Sánchez & López 1988). However, to our knowledge, there are no studies exploring the consequences of the application of such indices to an incomplete data set. Likewise, the definition of the form and size of the OGUs is another problem that is related to the sample representativity. When a study area is subdivided, the sampling units are re-defined, either as regular figures such as squares (e.g., McAllister *et al.* 1994) or pentagons (e.g., Griffith 1993), or as irregular areas such as political boundaries or biogeographical or biotic regions (e.g., McLaughlin 1989). Most often the definition of OGUs is carried out after sampling. Phipps (1975) proposed a method of defining OGU in a way that concentrates on maximizing the number of matches; however, he did not provide a justification for the proposed criteria.

When a similarity analysis is carried out based on a presence-absence data matrix that scores the species or taxa occurring in a set of OGUs, it is necessary to estimate the errors produced by different collecting efforts in each area. Generally,

knowledge of species richness comes from efforts carried out in different periods of time, and by different teams of researchers with different purposes. There is no single "sampling strategy" and, accordingly, it is not known how complete a given sample is.

Present estimates and analyses of biodiversity should consider the possible omissions of species and other taxonomic groups. Robust estimates should include quantitative parameters that show data on the confidence of discussed results. Taylor (1977) provided a pioneer exercise encompassing such ideas in quantitative biogeography; the author used statistical tests to diminish the error in the similitude-dissimilitude calculations among OGUs due to inaccuracies in the scoring of species.

These tests are important, especially because areas that are supposedly well-known floristically, such as North America, continue to reveal new species and genera (Erter 2000).

This paper aims to define an approach to estimate the error associated with the quality of records (sampling) that document the biodiversity of a region as is scored in a presence-absence data matrix. We also examine the effects of repeated subdivision of regions. An estimate of this kind may help to determine a "minimum limit" of error associated with the results of a biogeographical analysis.

This paper also proposes a strategy for evaluation how representative the records of a sample are in a biogeographical analysis. This is done by estimating the total richness based on the known richness, and considering the number of species presences recorded in the data matrix.

## CRITERIA TO ESTIMATE QUALITY OF THE RECORDS

To produce an index that estimates the quality of the sample records for a biogeographical analysis, two factors must be considered: the floristic or faunistic richness, and the number of incidences in the presence-absence data matrix. Non-parametric methods can be used to estimate the richness based on collecting data, assuming they represent samples of populations (Colwell & Codrington 1994). These estimates provide an ap-

proximate value of the “real richness” from the partially known richness. Thus a parameter of the quality of the records (or the sampling in the area) can be obtained from the difference between the known (sampled) and the estimated richness.

Another way to estimate the quality of the records is by presence per species in the data matrix (positive records or hints). A data matrix in which all species are recorded from all the regions can be considered complete, because there are no possible additions to the record. On the other hand, a data matrix with few species records for each region may be faulty when scoring the data for several OGUs. Thus, it is possible to define a maximum and a minimum of the presences in the data matrix and to prefer those data matrices with the high scoring.

### Richness estimators

Colwell and Coddington (1994) analyzed several proposals to estimate the total richness of a region from the known richness. They discussed the behavior of several of these richness estimators and list the formulae to calculate them.

Richness estimators require a data matrix of the species by OGUs. There are estimators that only require a presence-absence data matrix and do not require abundance data, while others require both. Among the first kind (as named by Colwell & Coddington 1994) can be cited those of Chao 2, Jackknife 1, Jackknife 2, Bootstrap or ICE (incidence-based coverage estimator), while those of Chao 1 and ACE (abundance-based coverage estimator) are of the second kind. There is evidence that estimators based only on incidences predict richness better than those based on abundance (Colwell 1996 and R. K. Colwell pers. comm).

We propose to evaluate the confidence of the data to be used in a quantitative biogeographical analysis by estimating the richness of the region under study and using the parameter  $E_s$ . This parameter measures the proportion of the known richness related to the estimated richness; it is calculated with some of the formulae that estimate the richness (see Colwell & Coddington 1994).

Let  $S_{est}$  be defined as the estimated (floristic or faunistic) richness for the whole region under study (obtained, for example, by using Chao 2 as

defined in Colwell & Coddington 1994), and  $S_{obs}$  as the known richness for the same region. Neither  $S_{est}$  nor  $S_{obs}$  represent the “real” richness of the region; strictly speaking both are estimates. The error  $E_s$  of the known richness  $S_{obs}$  with respect to the estimated one can be calculated according to the following formula:

$$E_s = [(S_{est} - S_{obs})/S_{est}] - 1 = 1 - S_{obs}/S_{est} \quad (1)$$

$E_s$  takes values from 0 to 1, because the known richness  $S_{obs}$  (we assume) should never surpass the estimated richness  $S_{est}$ ; accordingly, the quotient  $S_{obs}/S_{est}$  will never be larger than 1.

The  $E_s$  parameter encompasses all the regions (OGUs) and, because a biogeographical analysis may include multiple comparisons within the data matrix (among all the OGUs), the error in the results will be higher, perhaps a potency of  $E_s$ :

$$\text{Error} = C(E_s)^k \quad (2)$$

with  $C$  and  $k$  constants.

To obtain the  $E_s$  value the Chao 2 estimator formula can be used:

$$S_{est} = S_{obs} + Q_1^2/(2Q_2) \quad (3)$$

where  $S_{est}$  is the estimated richness according to the Chao criterion,  $Q_1$  is the number of species (or taxa) that occur in only one sample (i.e., in only one region) and  $Q_2$  is the number of species (or taxa) that occur in two samples or regions. Chao 2 is one among several proposed estimators; however, we have used it because is easy to interpret, it provides a better estimate of richness than others and it uses only presences and absences, not requiring abundance data, as Chao 1 (Colwell & Coddington 1994).

$E_s$  is a value that determines the robustness of the results from different studies; the lower the  $E_s$ , the higher the quality of the results. In general, it is expected that the confidence of the results is greater when spatial resolution diminishes, that is, when the OGUs include larger areas.

### Presence-absence data matrix frequencies addition

A second criterion to evaluate the confidence of the results of a similarity analysis is to consider the way species distribute among the sampled ar-

cas or OGUs, and the kind of mistakes most commonly made when a presence-absence data matrix is assembled. In terms of information content, two kinds of mistakes can be produced: (1) a particular species is scored when it really does not occur (errors of commission), and (2) a particular species is not scored when it does occur (errors of omission.)

These two cases are viewed in the data matrix as "presences" (or 1's) where they should be "absences" (or 0's) and vice versa. When all species are present in all OGUs, it can be considered that a good sampling representation has been made, because there is no evidence of missed scores for any species. On the other hand, if there are restricted distributions, for example, if certain species are only present in one OGU, the probability that presences have not been scored for several OGUs increases. In the first case, results that can be derived from the similarity analysis are extremely poor, whereas in the second case results should seem very informative. Thus, the more robust the results are, the less informative, and vice versa. In this way, we consider that the errors most common in a presence-absence data matrix are of the latter kind.

A way to evaluate the confidence of the distributions of species in a presence-absence data matrix is by analyzing their frequency number. Consider an histogram of species frequencies in  $j$  areas (i.e., in 1, 2, 3, ...,  $m$  areas). The sum ( $F$ ) of frequencies of all classes multiplied by all classes is defined as follows:

$$F = \sum(jQ_j); j = 1, \dots, m \quad (4)$$

where  $m$  is the total number of areas (OGUs),  $Q_j$  is the number of species found in  $j$  areas.

$F$ -values increase when OGUs become smaller; in other words, when spatial resolution is greater. If it is assumed that estimating error of richness is 0, then  $F$  has its upper limit at  $S_{\text{obs}}m$  and its lower limit is either  $S_{\text{obs}}$  or  $m$ ; that is, when not one species is shared between any pair of sites (restricted distribution).

If the data matrix has only 1's, then  $F = S_{\text{obs}}m$ . Assuming that neither a column nor a row have only 0's, then the minimum possible of 1's is  $S_{\text{obs}}$  or  $m$ , and strictly the maximum of  $S_{\text{obs}}$  and  $m$  corresponds to  $\max(S_{\text{obs}}, m)$ . If there are no mistakes in the calculation of  $F$ , that is, if it is assumed that

$S_{\text{obs}}$  is correct, then the maximum  $F$ -value corresponds to  $S_{\text{obs}}m$ . This means that there are no missing species in the studied area because all of them are present in every OGU.

In this way, a measure of the quality of the records ( $Q_s$ ) can be defined as:

$$Q_s = F/[S_{\text{obs}}m - \max(S_{\text{obs}}, m)] \quad (5)$$

This quotient can take values between 0 and 1, because  $F$  is divided by the interval of values it can take.

If there are no mistakes in the calculation of  $S_{\text{obs}}$ ,  $F$  may be the source of error; that is, may be all species of the region are scored but not in each OGU where they occur. The former assuming no mistakes are produced when scoring presences (the 1's are well placed in the data matrix) and the source of error is only the lack of records (could be 1's where there are 0's). On the other hand, an error in  $S_{\text{obs}}$  very probably causes, as a consequence, an error in  $F$ ; if a row is absent in the data matrix, there will be no 1's to add.

Figure 1 shows different forms an histogram of frequencies of species in  $j$  sites can take, assuming a constant  $S_{\text{obs}}$  and a variable  $F$ . If the data matrix is re-built in order to have a smaller number of OGUs, the trend is to get histograms like those on the right-hand side in Fig. 1. Although the absolute value of  $F$  diminishes, its relative value with respect to  $m$  will probably increase.

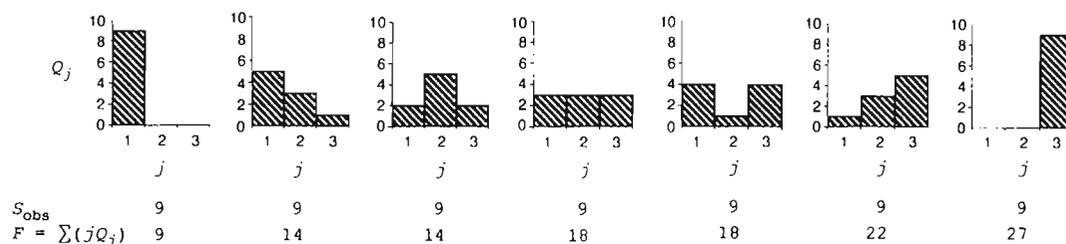
#### COMBINING RICHNESS ESTIMATORS AND SUM OF FREQUENCIES

An additional alternative to evaluate the quality of the records in presence-absence data matrices is combining the  $E_s$  and  $F$  parameters. The limits of  $F$  can be re-written both in terms of the estimated richness ( $S_{\text{est}}$ ) and in terms of the error  $E_s$  of the richness estimation.

From Eq. 1 above, we have that

$$S_{\text{est}} = S_{\text{obs}}/(1 - E_s) \quad (6)$$

As discussed above,  $F$  has a minimum value of  $\max(S_{\text{obs}}, m)$  and a maximum value of  $S_{\text{obs}}m$ . However, if it is considered that the richness can take values between  $S_{\text{obs}}$  and  $S_{\text{est}}$ , and the former is substituted by  $S_{\text{obs}}/(1 - E_s)$  as the upper limit of  $F$ , then the new values the  $F$  range can take are



**Fig. 1.** Seven forms an histogram of number of species by regions with constant known richness ( $S_{\text{obs}}$ ) and variable  $F$  can take ( $F = \sum (jQ_j)$ ;  $j = 1, \dots, m$ ;  $m =$  total number of OGU's and  $Q_j =$  frequencies of species in  $j$  sites). To the right the histogram forms with higher  $F$ -values are illustrated. The first histogram has nine taxa, all of them occurring in only one site, while the last histogram also has nine taxa, but all of them occur in three sites. The former has a low  $F$ -value ( $9 = 9(1)$ ) while the latter has a high  $F$ -value ( $27 = 9(3)$ ).  $F$ -values increase when OGU's become smaller, that is when resolution is greater. Histograms on the right are from the data matrices with most 1's and a low error in recording data is suspected. The histograms on the left are from the data matrices with less 1's and a larger error in recording data is suspected.

defined as the minimum value  $\max(S_{\text{obs}}, m)$  and as the maximum value  $S_{\text{obs}}m/(1 - E_s)$ .

If  $F$  is divided by the interval  $S_{\text{obs}}m/(1 - E_s) - \max(S_{\text{obs}}, m)$  a number between 0 and 1 is obtained. A number close to 1 indicates a lesser probability of error in the species sampling, and a number close to 0 indicates a greater probability of error. Based on this, an estimate of the quality of the sampling records ( $Q_s$ ) is:

$$Q_s = F / (S_{\text{obs}}m / (1 - E_s) - \max(S_{\text{obs}}, m)) \quad (7)$$

#### A PARTICULAR CASE OF MODIFIABLE AREAL UNIT PROBLEM (MAUP)

There is a problem when one attempts to describe the behavior of  $F$  when a particular area is subdivided into smaller areas of different sizes. In general, statistical variance of values associated with geographic units rises when an area is fragmented into smaller ones. This fact has been defined as the modifiable areal unit problem (MAUP, *see* Wong 1996). MAUP implies that while a correspondence (if it does occur) between variance and size of geographic units is found the results of studies in the same region using different sized units are not comparable.

Biogeographical data are also subject to MAUP because statistical variance is larger when smaller OGU's are considered.  $Q_s$  defines the effect of changing size and/or shape of OGU's, helping to decide what configuration to use in order to carry out a quantitative analysis.

#### A CASE STUDY USING DATA OF SELECTED GROUPS OF THE FLORA OF MEXICO

To evaluate the formula for estimating the quality of the sampling records ( $Q_s$ ) in a presence-absence data matrix, we analyzed the information on the distribution of different floristic groups of the flora of Mexico. We used the information from several databases: (1) the families of flowering plants (Magnoliophyta), (2) the genera of Asteraceae, Fabaceae, Malvaceae, Poaceae, and Rhamnaceae, and (3) the species of the genera *Ageratina* (Asteraceae) and *Desmanthus* (Fabaceae). The first two data sets come from unpublished information compiled by the junior author (except for Malvaceae, obtained from Fryxell (1988) and Rhamnaceae, obtained from Fernández (1993)). The data for *Ageratina* were taken from Turner (1997) and for *Desmanthus* from Luckow (1993). Accordingly, we analysed data for three different taxonomic levels (families, genera, and species). Tables 1 and 2 summarize the information of the number of taxa considered (*see*  $S_{\text{obs}}$ ).

To define the OGU's, Mexico was first subdivided into its 32 political states (Fig. 2 and Table 2) and then those 32 states were re-ordered into 8 regions (Table 2). Accordingly, two sets of OGU's were used, each one having its own data matrix.

Different parameters used to estimate the  $Q_s$  values for each taxonomic group in the two sets of OGU's are presented in Table 1. As there is a hierarchical arrangement among the groups analyzed, it is natural to see higher  $Q_s$  values at

**Table 1.** Parameters calculated to determine the quality of records ( $Q_s$ ) for different angiosperm taxa of the flora of Mexico (number of OGU: 32/number of OGU: 8).  $Q_1$  = number of species (or taxa) that occur in only one sample (i.e., in only one region);  $Q_2$  = number of species (or taxa) that occur in two samples or regions;  $S_{obs}$  = the known richness;  $F$  = sum of frequencies of all classes multiplied by all classes of the histogram of species frequencies by sample;  $S_{est}$  = the estimated richness according to the Chao 2 criterion;  $E_s(\%) = 100[(S_{est} - S_{obs})/S_{est}]$ ;  $Q_s(\%)$  = the proposed quality of the sampling records index.

Plant group	$Q_1$	$Q_2$	$S_{obs}$	$F$	$S_{est}$	$E_s(\%)$	$Q_s(\%)$
Magnoliophyta	7/7	9/16	257/257	4864/1633	259.7/258.3	1.0/1.0	60.4/90.2
Asteraceae	68/84	32/45	369/369	3977/1537	441.2/447.4	16.4/17.5	28.9/47.9
Fabaceae	11/14	8/8	139/139	1980/761	146.6/151.2	5.2/8.1	43.5/71.1
Malvaceae	4/7	2/6	53/53	660/252	57.0/57.1	7.0/7.1	37.3/62.4
Rhamnaceae	0/1	1/0	11/11	230/76	11.0/11.0	0.0/0.0	67.4/98.7
Poaceae	19/12	6/19	159/159	1856/807	189.1/162.8	15.9/2.3	31.5/70.6
<i>Ageratina</i>	44/3	28/51	138/138	720/498	172.6/138.9	20.0/0.1	13.4/51.5
<i>Desmanthus</i>	6/1	2/5	18/18	100/74	27.0/18.1	33.3/0.5	12.0/58.4

**Table 2.** Political division of Mexico and the region into which each state was assigned. See Fig. 2.

	State	Region
AGS	Aguascalientes	4
BCN	Baja California	1
BCS	Baja California Sur	1
CAM	Campeche	8
CHIS	Chiapas	7
CHIH	Chihuahua	2
COAH	Coahuila	3
COL	Colima	4
DF	Distrito Federal	6
DGO	Durango	2
GTO	Guanajuato	6
GRO	Guerrero	7
HGO	Hidalgo	5
JAL	Jalisco	4
MEX	México	6
MIC	Michoacán	6
MOR	Morelos	6
NAY	Nayarit	4
NLE	Nuevo León	3
OAX	Oaxaca	7
PUE	Puebla	7
QRO	Querétaro	6
QROO	Quintana Roo	8
SLP	San Luis Potosí	5
SIN	Sinaloa	2
SON	Sonora	2
TAB	Tabasco	8
TAM	Tamaulipas	3
TLAX	Tlaxcala	6
VER	Veracruz	5
YUC	Yucatán	8
ZAC	Zacatecas	4

higher taxonomic levels, for example families, and lower  $Q_s$  values at lower taxonomic levels.

At the family level, 257 families of Magnoliophyta occurring in Mexico were analyzed. It is considered that taxonomic and floristic knowledge at this level is adequate; therefore, results may serve to calibrate the upper limits of expected  $Q_s$  values at lower taxonomic levels. When a set of 32 OGUs (Fig. 1 and Table 2) is used, a  $Q_s$  of 60.4% is obtained; with 8 OGUs (Table 2), the  $Q_s$  value rises to 90.2%.

At the genus level, the three families with the largest number of species in Mexico (Asteraceae, Fabaceae, and Poaceae) plus two additional taxonomically well-known families (Malvaceae and Rhamnaceae) were evaluated.  $Q_s$  at the genus level was, as expected, intermediate between those at family and species level (Table 1). The  $Q_s$  values ranged from 29% to 67% when 32 OGUs were used, and from 47.9% to 98.7% when a set of 8 OGUs was employed. At the species level the two genera analyzed, with contrasting numbers of species, had  $Q_s$  values ranging from 12% to 13.4% with 32 OGUs, and from 51.5% to 58.4%, when the set of OGUs was 8.

The results point out that  $Q_s$  values will be influenced by the size of the taxonomic group as well as by the quality of taxonomic and distributional knowledge. Those groups with a large number of species are taxonomically less well known and thus show lower  $Q_s$  values (for example Asteraceae) than those with fewer species and

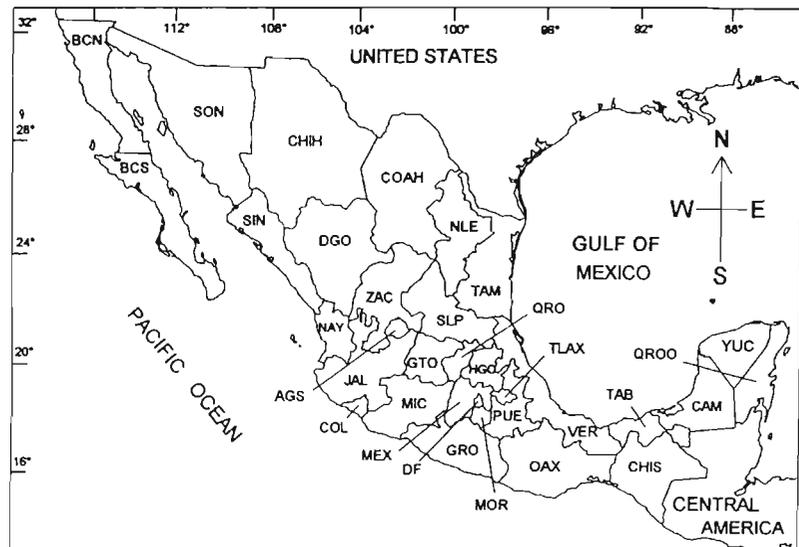


Fig. 2. Political states of Mexico (abbreviations as in Table 2.)

taxonomically better known (for example Rhamnaceae).

At the genus level, where the expected  $Q_s$  values are intermediate between family and species levels, estimated values can be proposed either for taxonomically well-known groups or to poorly-known ones. The Asteraceae have  $Q_s$  values lower than those obtained at species level (Table 1); on the other hand the Rhamnaceae have values higher than those obtained for all the families. It can then be proposed that when Mexico is divided into 32 OGUs (state level), values higher than 50% indicate a good "representativity" of the data set at the genus level, whereas values lower than 30% indicate a poor data set to be analyzed. On the other hand, when the country is subdivided into 8 OGUs, intervals of  $Q_s$  values could be established as follows: 50% or less are "poor" (poor data quality in the data matrix), 50% to 70% are "good", and more than 80% are "very good".

## DISCUSSION AND CONCLUSION

Biogeographical analyses based on similarity coefficients must be performed with the realization that they are based on incomplete data. Therefore, it is important to estimate the quality of the data matrix in order to assign an error to the results obtained. In this paper two different criteria

are proposed. Also, a combination of both is proposed, resulting in a formula applicable to any data matrix based on presence-absence data. It has the advantage of comparing the quality of the records, either from different regions or taxonomic groups or from data matrices from the same region and the same taxonomic groups but with different OGUs configuration.

Several elements affect the quality of biogeographical data represented in a presence-absence data matrix. Among the most important are the size of the taxonomic group, the taxonomic level, and the kind and size of the OGUs. To define the quality of the data, it is important to consider all of these factors.  $Q_s$  coefficient can be useful to estimate how complete the data are, in order to make comparable the values between areas.

The  $Q_s$  coefficient provides a way to compare the status of biodiversity knowledge from different regions. The higher the  $Q_s$  values, the better known the region. Likewise, it provides a way to estimate the biodiversity knowledge of a region at different times. An increase in this knowledge is not only the scoring of new taxa in a region; it includes their scoring in different places.

An index to estimate the records' quality is useful because it provides an indirect way to decide the resolution level of biogeographical analyses. Our  $Q_s$  parameter may be useful to measure how this quality diminishes when resolution increases.

ACKNOWLEDGEMENTS: We thank David Charlet, Claudio Delgadillo, Jorge Llorente, Luis E. Eguiarte, Juan José Morrone, Lawrence Kelly, and one anonymous reviewer for careful reviews of the manuscript and for their valuable comments. The Comisión Nacional para el Conocimiento y Uso de la Biodiversidad (CONABIO) and the Asociación de Biólogos Amigos de la Computación, A. C. (ABACo, A.C.) provided the hardware and software facilities. The senior author thanks the Dirección General de Asuntos del Personal Académico (DGAPA-UNAM) for support through a Doctoral Fellowship.

## REFERENCES

- Anderson, S. & Marcus, L. F. 1993: Effect of quadrat size on measurements of species density. — *J. Biogeogr.* 20: 421–428.
- Bricks, H. J. B. 1987: Recent methodological development in descriptive biogeography. — *Ann. Zool. Fennici* 24: 165–178.
- Colwell, R. K. 1996: User's guide to the richness estimator program, EstimateS. — Univ. Connecticut, MS available via the internet at <http://viceroy.eeb.uconn.edu/EstimateS>.
- Colwell, R. K. & Coddington, J. A. 1994: Estimating terrestrial biodiversity through extrapolation. — *Phil. Trans. Royal Soc. London B* 345: 101–118.
- Crovello, T. J. 1981: Quantitative biogeography: an overview. — *Taxon* 30: 563–575.
- Erter, B. 2000: Floristic surprises in North America north of Mexico. — *Ann. Missouri Bot. Garden* 87: 81–109.
- Fernández, R. 1993: *La familia Rhamnaceae en México*. — Ph.D. Thesis. Escuela Nacional de Ciencias Biológicas, I.P.N. México, D.F. 345 pp.
- Fryxell, P. A. 1988: Malvaceae of Mexico. — *Syst. Bot. Monogr.* 25: 1–522.
- Griffith, D. A. 1993: Advanced spatial statistics for analysing and visualizing geo-referenced data. — *Int. J. Geogr. Inf. Syst.* 7: 107–123.
- Hagmeier, E. M. & Stults, C. D. 1966: A numerical analysis of the distribution patterns of North American mammals. — *Syst. Zool.* 15: 125–155.
- Hubálek, Z. 1982: Coefficients of association and similarity, based on binary (presence-absence) data: an evaluation. — *Biol. Rev.* 57: 669–689.
- Luckow, M. 1993: Monograph of *Desmanthus* (Leguminosae-Mimosoideae). — *Syst. Bot. Monogr.* 38: 1–166.
- McAllister, D. E., Schueler, F. W., Roberts, F. W., Roberts, C. M. & Hawkinsiller, J. P. 1994: Mapping and GIS analysis of the global distribution of coral reef fishes on an equal-area grid. — In: Miller, R. I. (ed.), *Mapping the diversity of nature*: 155–175. Chapman & Hall, Oxford.
- McLaughlin, S. P. 1989: Natural floristic areas of the western United States. — *J. Biogeogr.* 16: 239–248.
- Nimis, L., Malyshev, L., Bolognini, G. & Friesen, N. 1995: Phytogeographic diversity of the Putorana flora (N. Siberia). — *Ann. Bot. Fennici* 32: 1–17.
- Phipps, J. B. 1975: BestBlock: optimizing grid size in biogeographic studies. — *Can. J. Bot.* 53: 1447–1452.
- Sánchez, O. & López, G. 1988: A theoretical analysis of some indices of similarity as applied to biogeography. — *Folia Entomológica Mexicana* 75: 119–145.
- Taylor, D. W. 1977: Floristic relationships along the Cascade-Sierran Axis. — *Am. Midl. Natur.* 92: 333–349.
- Turner, B. L. 1997: The Comps of Mexico. Vol. 1. Eupatorieae. — *Phytologia Mem.* 11: 1–272.
- Wong, D. 1996: Aggregation effects in geo-referenced data. — In: Arlinghaus, S. L. (ed.), *Practical handbook of spatial statistics*: 83–106. CRC Press, New York.

## 8. Efectos del índice de similitud en la clasificación biogeográfica

---

*En este capítulo se presentan los resultados de analizar la interacción entre dos tipos de herramientas de la BC: los índices de similitud y los tipos de ligamiento, abordando también algunos aspectos de la influencia del índice de similitud elegido sobre la matriz de similitud.*

Este capítulo es parte de los resultados de este trabajo en lo referente a los objetivos sobre:

1.3) Analizar el problema de los efectos del índice de similitud en la clasificación biogeográfica.

El capítulo, en su totalidad, es el artículo:

MURGUÍA, M. Y J. L. VILLASEÑOR, 2003. “**Estimating the effect of the similarity coefficient and the cluster algorithm on biogeographic classifications**”.  
*Ann. Bot. Fennici*, 40:.

Cuya traducción de título al español es: ‘*Estimación del efecto del índice de similitud y del algoritmo de agrupamiento en las clasificaciones biogeográficas*’

El artículo aborda de manera empírica el efecto de la elección del índice de similitud sobre los fenogramas de áreas así como con un enfoque combinatorio, considerando las posibles matrices de presencia ausencia de tamaño pequeño.

El análisis se realiza guiado por la necesidad de conocer la “comparabilidad” de los resultados de estudios que utilizan diferentes índices de similitud (v. gr. índice de Simpson, índice de Jaccard o índice de Sorensen) con el mismo o diferente tipo de ligamiento (v. gr. ligamiento simple, completo o promedio).

Los principales resultados y conclusiones del artículo son:

- 1) El índice de similitud puede influir hasta el 70% en las diferencias de las clasificaciones biogeográficas.
- 2) El algoritmo de agrupamiento puede influir hasta el 53% en las diferencias de las clasificaciones biogeográficas, y muy probablemente arriba del 16%.
- 3) Las clasificaciones biogeográficas realizadas con diferente índice de similitud no son comparables.
- 4) Las clasificaciones biogeográficas realizadas con diferente algoritmo, aunque con el mismo índice de similitud, no son comparables.
- 5) Si no hay un argumento en favor de utilizar un índice de similitud en particular, el índice de Jaccard es una buena opción, debido a que a) las clasificaciones biogeográficas realizadas con diferente algoritmo de agrupamiento muestran pocas diferencias (por lo tanto son comparables), y b) muestra alto porcentaje de definición (ver texto del artículo y concepto de “topología altamente informativa”).

## Estimating the effect of the similarity coefficient and the cluster algorithm on biogeographic classifications

Miguel Murguía<sup>1</sup> & José Luis Villaseñor<sup>2</sup>

<sup>1</sup> *Asociación de Biólogos Amigos de la Computación (ABACo, A. C.), Av. San Jerónimo 507, Col. San Jerónimo Lídice, 10200 México, D.F. México (e-mail: mmr@ciencias.unam.mx)*

<sup>2</sup> *Instituto de Biología, Universidad Nacional Autónoma de México, Departamento de Botánica, Apartado postal 70-237, 04510 México, D.F. México (e-mail: vrios@mail.ibiologia.unam.mx)*

*Received 25 Sep. 2002, revised version received 4 June 2003, accepted 13 June 2003*

Murguía, M. & Villaseñor, J. L. 2003: Estimating the effect of the similarity coefficient and the cluster algorithm on biogeographic classifications. — *Ann. Bot. Fennici* 40: 415–421.

Similarity matrices obtained using a null model and nine similarity coefficients based on an exhaustive and hypothetical set of presence/absence data matrices are generated and compared. Likewise, the biogeographic classifications obtained from an empirical set of data (the genera of Asteraceae of Mexico) and from the application of the same nine similarity coefficients and three cluster methods are compared. It is concluded that differences in the classifications generated from different similarity coefficients can be of almost 50% with the set of hypothetical matrices and more than 70% with the empirical data. The kind of clustering method (single, complete, or average) also generates differences in the classification topologies, even when using the same similarity coefficient. The empirical data produced similar topologies in 51% to 85% of the cases. Due to the dependence among the similarity coefficients, the cluster method used, and the generated classifications, it is concluded that classifications obtained through the use of different similarity coefficients or cluster methods are not comparable. The most similar classification topologies were obtained from the use of the Jaccard and Sorensen-Dice similarity coefficients. They also showed the lowest number of poorly informative structures.

Key words: Asteraceae, cluster analyses, null models, quantitative biogeography, similarity coefficients

### Introduction

Each day more precise biogeographic data are available. They have also been ever more frequently used to classify areas based on knowledge of the distribution patterns of their biota.

Numerous tools are currently available for applying numerical methods to biogeographic analyses. They have contributed to background

knowledge that better describes and explains the geographical distribution patterns of organisms. However, analyses carried out with different numerical methods do not always arrive at the same result, mostly due to the variables selected (Crovello 1981, Hubálek 1982). Therefore, understanding the behavior of a particular method using a given set of variables is a crucial issue.

Crovello (1981) pointed out that quantitative biogeographic analysis is a decision making process. Among the decisions to be made are the definition of the operative geographical units (OGUs), and the selection of a similarity coefficient or cluster method. Each of these decisions may lead to different results from the same data set.

A general methodology to carry out a quantitative biogeographical analysis uses a data set arranged as a matrix that scores the presence-absence of taxa (e.g., species) for each OGU selected (*see* Bricks 1987: fig. 1). The data matrix is transformed into a similarity (or dissimilarity) matrix to compare how similar (or different) each pair of OGUs are. Finally, the similarity matrix is used to carry out a cluster analysis, where one of several cluster algorithms must be chosen to generate a classification of the OGUs. The latter will help in proposing a geographical regionalization that supposedly has biological significance (Crovello 1981, McLaughlin 1986, Bricks 1987).

Different criteria to evaluate the similarity between pairs of OGUs, as seen in the similarity coefficients, will lead to different classifications. To what extent does the selection of one or another similarity index influence the resulting biogeographical classifications? This is a question that requires a response in view of the increasing number of analyses using the general methodology outlined above.

The literature documents the need to incorporate null models as a test of biogeographical hypotheses (Simberloff 1983, Craw 1989, Gotelli & Graves 1996). The null probability concept as the basis for constructing up null models has been discussed by Simberloff (1983), who has applied it to different biogeographical school methodologies (e.g. Simberloff 1978). Gotelli and Graves (1996) summarize the methods used to construct up null models.

The aim of this paper is to determine how the selection of a similarity coefficient affects resulting biogeographical classifications when using methods of quantitative biogeography.

## Methods

The classification of OGUs into higher level biogeographical units can be considered a process

that selects one of many possible combinations. Naturally, the more OGUs being analyzed, the larger the number of possible combinations. In this paper we analyze with a combinatorial approximation the different possible classifications resulting from the use of different similarity coefficients.

Two sources of information were used to investigate how the similarity coefficients affect biogeographical classifications. The first is a null model approach (Gotelli & Graves 1996) using an exhaustive set of hypothetical matrices of three OGUs and three, four, and five attributes (species or taxa), and comparing the classification topologies obtained after the cluster analysis. The second uses an empirical data set, a presence/absence data matrix of the genera of Mexican Asteraceae (368 genera), and the 32 political states (OGUs) of Mexico.

## Possible classifications

Part of the analysis was carried out on a set of all possible classifications of three OGUs. A presence/absence data matrix of three OGUs and three taxa can be ordered in 512 ( $2^9$ ) possible combinations (each of the nine cells can be equally scored as an absence or a presence). Likewise, a data matrix of three OGUs and four taxa will have 4096 ( $2^{12}$ ) possible combinations, and one with three OGUs and five taxa 32 768 ( $2^{15}$ ) possible combinations.

Next, any of these equally possible datasets is transformed into a similarity matrix by using a similarity coefficient. This coefficient will transform the raw presence/absence data into values that measure how similar two OGUs are based on the number of taxa shared. The transformed similarity matrix is now a triangular matrix, because both halves contain the same data (Sneath & Sokal 1973).

Finally, the similarity matrix is used to construct a tree graph or dendrogram that will depict the relationships among the OGUs, based on the similarity values. This dendrogram (the topology used for the classification) is used to group the OGUs in higher level units that finally will be useful to propose a regionalization of the study area.

To compare the classifications obtained, we used the triplets method, similar to the quartet method (Estabrook *et al.* 1985). This method determines a criterion when a triplet has the same arrangement in two dendrograms. For a binary tree with  $n$  terminal leaves (OGUs) there are  $n(n-1)(n-2)/6$  triplets (Page 1993). Thus, for example the data matrix for the genera of Asteraceae of Mexico, which includes 32 OGUs, each classification includes 4960 triplets. The triplets method is more relaxed than that of quartets, and we consider it more appropriate in that it reduces overestimates of the differences among the classificatory topologies obtained in the dendrograms.

### Matrix structures

In addition to analyzing the classification topologies (dendrograms) obtained from different similarity coefficients, the structure of the similarity matrices is likewise analyzed for the possible arrangement of values the matrices can produce (this arrangement is here referred to as the 'matrix structure', and can show equal, larger, or lower relationships among three pairs of OGUs). For example, a three-OGUs data matrix results in a similarity matrix that can produce 13 different equally possible hierarchical values (Table 1).

Each arrangement (matrix structure, Table 1) was then associated with one or more classification topologies, using a cluster algorithm (Fig. 1). Thus the most helpful structures to select the similarity coefficients that produce a single dendrogram were determined.

We employed the following rules to evaluate the arrangements of the similarity values in a data matrix:

- a. Poorly informative structures: those with the three pairs of OGUs equal in values ( $AB = BC = AC$ ) or those in which the similarity coefficient cannot be applied because the denominator is zero.
- b. Half informative structures (enclosing structures): those that produce several dendrograms due to equal numerical values between two pair of OGUs in the similarity matrix (for example  $AB = AC > BC$ .)

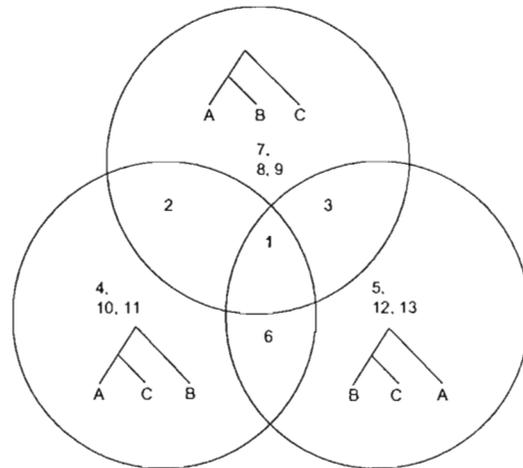


Fig. 1. Correspondence between matrix structures and classification topologies. The numbers correspond to matrix structures in Table 1.

- c. Highly informative structures: structures with different values among the three pairs of OGUs that thus produce a single dendrogram (for example  $AB > AC > BC$ .)

The present study emphasizes the influence of the similarity coefficients in the construction of dendrograms. However, based on the former

**Table 1.** The 13 equally possible hierarchical values (matrix structures) that can produce a similarity matrix obtained from an original data matrix of three OGUs (A, B, and C) (=, > or < indicate the relative order of the similarity values, that is, equal, larger than, or lower than.) The structures are arranged according to their information content.

#### Poorly informative structures:

1.  $AB = BC = AC$

#### Half informative structures:

2.  $AB = AC > BC$
3.  $AB = BC > AC$
4.  $AB = BC < AC$
5.  $AB = AC < BC$
6.  $BC = AC > AB$
7.  $BC = AC < AB$

#### Highly informative structures:

8.  $AB > AC > BC$
9.  $AB > BC > AC$
10.  $AC > BC > AB$
11.  $AC > AB > BC$
12.  $BC > AB > AC$
13.  $BC > AC > AB$

rules, it is possible to explore criteria for evaluating coefficients behavior.

It is important to point out that several matrix structures are associated with just one of the classification topologies (highly informative structures), while others can be associated with more than one (poorly or half informative structures) (Fig. 1). This arrangement is independent of the clustering method used (either single or complete linkages, or UPGMA).

For each data matrix, nine similarity matrices were produced by applying nine similarity coefficients (Table 2). Each similarity matrix was defined as poorly informative, half informative, or highly informative.

### Empirical analysis

In addition to evaluating the behavior of the similarity coefficients relative to the biogeographic classifications depicted as dendrograms based on theoretical matrices of three OGUs and three, four and five taxa, we conducted a similar analysis with empirical data: the genera of Asteraceae of Mexico. A presence/absence data matrix of 368 genera in the 32 political states of Mexico was transformed into nine similarity matrices using nine different coefficients (Table 2). Likewise, the dendrograms obtained from the use of three cluster algorithms (single linkage, complete linkage, and UPGMA) were analyzed. The analyses were made following the same procedure as with the theoretical data sets, that is, the dendrograms were compared with

**Table 2.** Similarity coefficients used in the analysis.

Simpson	$a/\min\{(a+b), (a+c)\}$
Jaccard	$a/(a+b+c)$
Braun-Blanquet	$a/\max\{(a+b), (a+c)\}$
Sorensen-Dice	$a/[a+0.5(b+c)]$
Kulczynski 1	$a/(b+c)$
Kulczynski 2	$0.5[a/(a+b) + a/(a+c)]$
Fager	$a/[(a+b)(a+c)]^{0.5} - 0.5\{\max\{(a+b), (a+c)\}\}$
Otsuka	$a/[(a+b)(a+c)]^{0.5}$
Correlation ratio	$a^2/[(a+b)(a+c)]$

$a$  = number of taxa (attributes) present in both OGUs  
 $b$  = number of taxa present only in the first OGU  
 $c$  = number of taxa present only in the second OGU

the triplets method to evaluate the differences resulting from the use of different similarity coefficients and different cluster algorithms, as well as the same coefficient but different cluster algorithms.

### Results

Table 3 shows the percentages of matrix structures generated with each data matrix analyzed using the nine similarity coefficients cited in Table 2. In the three data matrices, the coefficients with the largest percentages of poorly

**Table 3.** Percentages of structures, arranged according to their information content, obtained from the use of nine similarity coefficients in three different group matrices (Poorly = poorly informative; Half = Half informative; Highly = Highly informative).

Similarity coefficient	Poorly	Half	Highly
<b>Three taxa three OGUs group matrices</b>			
Simpson	50.6	42.4	7.0
Jaccard	15.0	67.4	17.6
Braun-Blanquet	18.5	70.9	10.5
Sorensen-Dice	15.0	67.4	17.6
Kulczynski 1	27.3	55.1	17.6
Kulczynski 2	57.8	35.2	7.0
Fager	36.7	45.7	17.6
Otsuka	36.7	45.7	17.6
Correlation ratio	43.7	49.2	7.0
<b>Four taxa three OGUs group matrices</b>			
Simpson	34.8	22.1	43.1
Jaccard	8.4	14.5	77.1
Braun-Blanquet	12.8	20.6	66.6
Sorensen-Dice	8.4	14.5	77.1
Kulczynski 1	13.2	17.7	69.1
Kulczynski 2	37.3	22.1	40.6
Fager	21.2	24.7	54.1
Otsuka	21.2	14.5	64.3
Correlation ratio	26.6	18.6	54.8
<b>Five taxa three OGUs group matrices</b>			
Simpson	23.0	24.7	52.3
Jaccard	7.5	13.1	82.4
Braun-Blanquet	8.3	20.9	70.9
Sorensen-Dice	7.5	13.1	82.4
Kulczynski 1	6.2	14.3	79.5
Kulczynski 2	22.1	23.6	54.3
Fager	11.6	21.9	66.4
Otsuka	11.6	14.7	73.6
Correlation ratio	14.4	20.5	65.1

informative structures generated were Simpson and Kulczynski 2. On the other hand, the coefficients that generated in the three data matrices the largest percentages of highly informative structures were Jaccard and Sorensen-Dice.

Because different classification topologies may be produced by one similar structure, it is important to point out that the choice of a coefficient will affect qualitatively the results of the biogeographical analysis. Only two coefficients (Jaccard and Sorensen-Dice) produced identical results (matches) both in the matrix structures and the classification topologies obtained. Those with lower percentages of matches were Fager and Braun-Blanquet with 38.0%, 36.1% and 39.2% of identical matrix structures for three,

four and five taxa matrices, and 50.2%, 56.0% and 60.5% of identical triplets for three, four and five taxa matrices. The results point out that two equally feasible biogeographical classifications may be obtained from the same data matrix by using different similarity coefficients, ranging from 50.2% to 100%.

The analysis of the genera of Asteraceae of Mexico also showed contrasting results, parallel to those obtained from the theoretical analysis. Table 4 compares the classification topologies obtained from the use of three different clustering methods. Percentages of similar topologies showed differences among the coefficients used. They go from 29.5% (Braun-Blanquet-Correlation Ratio, single linkage) to 100% (for example

**Table 4.** Percentages of identical classification topologies (triplets) obtained from the use of nine different similarity coefficients and three cluster methods, applied to a presence/absence data matrix of the genera of Asteraceae (368) occurring in the states of Mexico (32).

	1	2	3	4	5	6	7	8	9
<b>Complete linkage</b>									
(1) Simpson	100.0								
(2) Jaccard	67.5	100.0							
(3) Braun-Blanquet	60.6	87.4	100.0						
(4) Sorensen-Dice	67.5	100.0	87.4	100.0					
(5) Kulczynski 1	67.5	100.0	87.4	100.0	100.0				
(6) Kulczynski 2	82.4	73.7	69.8	73.7	73.7	100.0			
(7) Fager	68.5	61.2	56.0	61.2	61.2	77.6	100.0		
(8) Otsuka	67.5	100.0	87.8	100.0	100.0	73.7	61.2	100.0	
(9) Correlation ratio	71.2	66.6	62.7	66.6	66.6	71.6	66.4	66.6	100.0
<b>Single linkage</b>									
(1) Simpson	100.0								
(2) Jaccard	42.6	100.0							
(3) Braun-Blanquet	35.4	79.7	100.0						
(4) Sorensen-Dice	42.6	100.0	79.7	100.0					
(5) Kulczynski 1	68.9	100.0	79.7	100.0	100.0				
(6) Kulczynski 2	67.8	52.0	46.6	52.0	52.0	100.0			
(7) Fager	41.7	39.1	30.1	39.1	39.1	53.6	100.0		
(8) Otsuka	54.2	94.9	81.9	94.9	94.9	55.0	37.1	100.0	
(9) Correlation ratio	100.0	29.8	29.5	29.8	29.8	50.4	31.5	32.4	100.0
<b>UPGMA</b>									
(1) Simpson	100.0								
(2) Jaccard	61.0	100.0							
(3) Braun-Blanquet	55.7	72.1	100.0						
(4) Sorensen-Dice	61.1	100.0	72.1	100.0					
(5) Kulczynski 1	54.5	83.6	66.2	83.6	100.0				
(6) Kulczynski 2	74.8	72.3	64.0	77.3	69.6	100.0			
(7) Fager	74.7	66.2	50.2	66.2	55.9	75.4	100.0		
(8) Otsuka	62.7	88.1	72.7	88.1	78.3	69.5	56.7	100.0	
(9) Correlation ratio	62.5	45.6	52.2	45.6	39.4	53.4	59.7	49.5	100.0

Jaccard–Sorensen–Dice in the three clustering methods).

Table 5 shows the percentages of similar classification topologies of Table 4, comparing the same similarity coefficient but differing in clustering method. In all cases similar topologies never matched totally; values ranged from 51.0 (Complete linkage vs. Single linkage, based on Fager) to 84.1% (Complete linkage vs. UPGMA, based on Jaccard and Sorensen–Dice). The Complete Linkage and UPGMA methods produced the greatest number of matching classification topologies, no matter the similarity coefficient used (Table 5). The Complete and Single Linkage methods produced the lesser matching classification topologies. The differences are expected because the UPGMA method is intermediate among the Complete Linkage and the Single Linkage methods, which use respectively the maximum and minimum similarity values.

## Discussion and conclusion

The greatest number of similar classification topologies were obtained from the use of the Jaccard and Sorensen–Dice similarity coefficients. They also showed the least number of poorly informative structures. The best behavior of these coefficients agree with the results presented by Hubálek (1982), who concluded that these are the best coefficients based on a series of specific criteria.

**Table 5.** Percentages of identical classification topologies (triplets) obtained with the three cluster methods applied to a presence/absence data matrix of the genera of Asteraceae (368) occurring in the states of Mexico (32).

	Complete vs. Single	Complete vs. UPGMA	Single vs. UPGMA
Simpson	52.7	68.2	60.5
Jaccard	66.2	84.1	71.9
Braun-Blanquet	69.7	79.9	73.3
Sorensen-Dice	66.2	84.1	71.9
Kulczynski 1	66.2	78.3	76.6
Kulczynski 2	55.2	73.4	61.1
Fager	51.0	73.0	58.5
Otsuka	64.8	77.7	69.2
Correlation ratio	47.1	64.7	62.3

Our results do not agree with that of Sánchez and López (1988), who concluded that Simpson's coefficient was adequate for biogeographical studies. This coefficient, along with Kulczynski 2 produced lower percentages of highly informative structures among nine similarity coefficients analyzed (Table 3). Also, they showed the lowest percentage values when applied to the empirical data (Tables 4 and 5). Based on our results, it is surprising that Hubálek (1982) included Kulczynski 2 as a "good" coefficient, along with Jaccard and Sorensen–Dice.

Our analysis of empirical data resulted in classificatory topologies as different as 51% (Table 4), depending on the clustering method used. The results indicate that care should be used in selecting a clustering method. The clustering method definitely influences strongly the classification's topology.

Although there are no solid arguments in favor of a particular similarity coefficient, the analyses of their behavior and properties (*see* also Hubálek 1982, Sánchez & López 1988) help guide the choice of a coefficient. However, if several coefficients are to be used, additional criteria such as those evaluated in this paper can be used. Our results suggest that based on the number of poorly informative structures generated, the best index to use is Jaccard or Sorensen–Dice, followed by Braun-Blanquet or Kulczynski 1.

The biogeographical classifications produced by clustering methods rely strongly on the similarity coefficient and the clustering methods used. Thus, as our results indicate, the classifications obtained from different data sets employing different similarity coefficients and clustering methods are not comparable. These classifications obtained by quantitative methods, should be used as heuristic guides to define biogeographical regions. They must be compared with alternative forms of analysis, for example, using phylogenetic principles as in PAE analyses (Rosen 1988) or panbiogeographic proposals (Craw 1989) to avoid relying only on a single interpretation.

In this paper we use null models design principles in similarity coefficients and clustering methods, in an equivalent way to the equiprobable cladograms of Simberloff (1983). As did

Simberloff, we quantified the differences and probabilities shown by similar methods, independent of the data used in a biogeographical study. Our analysis could be considered an additional null model that explores the properties, goodness and failures of different similarity coefficients and clustering methods to biogeographical presence/absence data.

### Acknowledgements

The senior author acknowledges the Dirección General de Asuntos del Personal Académico of the Universidad Nacional Autónoma de México, the economic support through a doctoral fellowship. The careful reading by Drs. Claudio Delgadillo, Luis Eguiarte, Jorge Llorente, and Mark Olson improved the ideas and content of the paper.

### References

- Bricks, H. J. B. 1987: Recent methodological developments in descriptive biogeography. — *Ann. Zool. Fennici* 24: 165–178.
- Connor, E. F. & Simberloff, D. 1978: Species number and compositional similarity of the Galapagos flora and avifauna. — *Ecol. Monogr.* 48: 219–248.
- Craw, R. 1989: Quantitative panbiogeography: introduction to methods. — *New Zealand J. Zool.* 16: 485–494.
- Crovello, T. J. 1981: Quantitative biogeography: an overview. — *Taxon* 30: 563–575.
- Estabrook, G. F., McMorris, F. R. & Meacham, C. A. 1985: Comparison of undirected phylogenetic trees based on subtrees of four evolutionary units. — *Syst. Zool.* 34: 193–200.
- Gotelli, N. J. & Graves, G. R. 1996: *Null models in ecology*. — Smithsonian Inst. Press, Washington, D.C.
- Hubálek, Z. 1982: Coefficients of association and similarity, based on binary (presence-absence) data: an evaluation. — *Biol. Rev.* 57: 669–689.
- McLaughlin, S. P. 1986: Floristic analysis of the southwestern United States. — *Great Basin Naturalist* 46: 46–65.
- Page, R. D. M. 1993: *Component. Version 2.0. User's guide*. — Nat. Hist. Mus., London.
- Rosen, B. R. 1988: From fossils to earth history: applied historical biogeography. — In: Myers, A. A. & Giller, P. S. (eds.), *Analytical biogeography*. 437–481. Chapman & Hall, New York.
- Sánchez, O. & López, G. 1988: A theoretical analysis of some indices of similarity as applied to biogeography. — *Folia Entomológica Mexicana* 75: 119–145.
- Simberloff, D. 1978: Using island biogeographic distributions to determine if colonization is stochastic. — *Am. Nat.* 112: 713–726.
- Simberloff, D. 1983: Biogeography: the unification and maturation of a science. — In: Brush, A. H. & Klark, G. H. (eds.), *Perspectives in ornithology*: 411–455. Cambridge Univ. Press, Cambridge.
- Sneath, P. H. A. & Sokal, R. R. 1973: *Numerical taxonomy*. — W. H. Freeman, San Francisco.

## 9. Análisis de métodos de la Biogeografía Cuantitativa

---

*En este capítulo se exponen los resultados del análisis de los métodos de la BC así como su documentación y su integración en un modelo operativo. La documentación también es un resultado necesario que especifica los requerimientos del sistema y las especificaciones detalladas para su automatización en un Sistema de Información, mediante la codificación en un lenguaje de computadora de alto nivel.*

### 9.1. Clasificación de los métodos de la BC

#### *Clases de métodos de la BC*

Con el fin de establecer un orden en la secuencia de descripción de los métodos de la BC, en esta sección se construye una clasificación, justificándola en los niveles de abstracción del conocimiento y en clasificaciones previas de métodos implementados en los GIS.

En la figura 9.1 se proponen las clases en las que puede ubicarse cada método de la BC, indicando para cada una los niveles de abstracción del conocimiento.

Clase de Método de la BC	Nivel de abstracción del conocimiento
Métodos Descriptivos	Descripción
Métodos de Despliegue e Informes	Clasificación; Relaciones Directas
Métodos Analíticos	Concatenación
Métodos de Modelación	Formulación o Enunciados

Figura 9.1 Clases de métodos de la BC.

Las clases están inspiradas en la formulación propuesta por Salazar (1979), sobre el proceso de generación del conocimiento, expuesta de manera muy resumida en el capítulo 5, y en la división que marca Bailey (1994) entre el *resumen espacial* (*'spatial summarization'*) y el *análisis espacial* (*'spatial analysis'*) ya referida en la sección 'Análisis espacial' de la introducción. La categoría de 'resumen espacial' de Bailey puede identificarse con las dos primeras clases de la figura 9.1, y la categoría de 'análisis espacial' con las últimas.

Haining (1994) propuso una clasificación del tipos de preguntas que un GIS debe responder; inspirada en esa clasificación se propone una serie de 'Preguntas biogeográficas' que se deben incluir en el GIS para la BC. De la misma manera que Haining observa que de datos espaciales surgen preguntas espaciales (*v.gr.*, dónde se distribuye una especie) y no espaciales (*v.gr.* cuáles especies se distribuyen en una región), se definen preguntas biogeográficas en las dos categorías.

Dada esa clasificación y la correspondencia ya mostrada con la de Salazar, la de Bailey y la de Haining, en la etapa de análisis del desarrollo de la herramienta surgen tres conceptos fundamentales para agrupar de manera técnica algunos métodos de la BC:

- El concepto de *Bioestadísticas* (a partir de la *summarization* de Bailey)
- El concepto de *Preguntas biogeográficas* (a partir de *preguntas espaciales* y *no espaciales* de Haining)
- El concepto de *Camino metodológico* (a partir de las *formulaciones* de Salazar)

## Esquema de clasificación

Al considerar las clases identificadas en la sección anterior (figura 9.1) y una lista de métodos – que cumplen con los acotamientos explicados-, en la figura 9.2 se muestra la clasificación propuesta de los métodos de la BC. En la figura 9.2, en la sección ‘Despliegues e Informes’, se ha puesto entre paréntesis una abreviación de cada método que servirá como identificador o ‘nombre corto’; se han utilizado las iniciales ‘UG’ (Unidad Geográfica) en lugar de las de OGU: UG se utiliza para referirse a la implementación informática de las OGUs, es decir, asocia el concepto biogeográfico OGU con un concepto informático.  $P_i$  connota la clase de UGs con riqueza  $i$ ;  $Q_j$  connota la clase de taxones con frecuencia  $j$ .

<b>Descripción</b>
Cuantificación de la Diversidad
Estimadores de Riqueza
Calidad del muestreo
<b>Despliegue e Informes</b>
Ubicación de localidades de recolecta (Localidades)
¿Dónde está un determinado conjunto de especies? (Especies → Areas)
¿Qué especies hay en un determinado conjunto de UGs? (Areas → Especies)
¿Cuáles UG's tienen riqueza $r$ ? ( $P_i$ ; UG)
¿Cuáles especies se distribuyen en $n$ UG's? ( $Q_j$ ; Especies)
¿Cuál es la similitud entre dos conjuntos de UG's? (Entre UG's)
<b>Análisis</b>
Índices de similitud
Análisis de agrupamiento
Elección de áreas prioritarias para la conservación
<b>Modelación</b>
Caminos Metodológicos

Figura 9.2. Clasificación de los Métodos de la BC considerados para su análisis.

Las acotaciones justifican el que no se incluyan en el análisis a los estimadores de riqueza con base en abundancia, sino solo a aquellos con base en incidencias. Los métodos multivariados de la estadística, como el análisis de componentes principales y las regresiones lineales múltiples, son ampliamente usados por la comunidad, pero no se incluyen debido a que: 1) existen paquetes estadísticos (*v.gr.* SPSS o SAS) que los implementan, 2) normalmente requieren de variables ambientales para su aplicación (*v.gr.* altitud, tipo de vegetación o clima).

Seguramente el lector biogeógrafo no encontrará algunos o varios de los métodos a los que habitualmente recurre en sus estudios, pero quizá sí pueda ubicarlos en alguna de las clases propuestas. Se debe tener presente que el propósito de la clasificación propuesta no es dar una lista exhaustiva de métodos, sino de construir una herramienta para análisis biogeográficos que integre, de manera ordenada, técnicas para las que aún no existe una versión automatizada e integrada en un paquete biogeográfico.

## 9.2. Documentación de los métodos de la BC

Uno de los principales resultados de la etapa de análisis del desarrollo de la herramienta es la documentación técnica de los métodos a automatizar. La documentación además tiene el objeto de facilitar la etapa de codificación en un lenguaje de alto nivel, razón por la que no sólo se incluyen las fórmulas, procedimientos y algoritmos de los métodos, sino también una descripción de la salida o resultado y un ejemplo del despliegue gráfico (mapa) de resultados por el sistema al aplicar ese método.

En la descripción de los métodos se manejan los siguientes conceptos, que también son resultado del análisis:

1. **Área de Estudio (AE):** Es el área definida por los dos puntos extremos del rectángulo. Es en sí la región geográfica sobre la que se aplican los análisis.
2. **Unidad Geográfica (UG):** Es la denominación que aquí se le asigna a las OGU. Son cada una de las áreas geográficas en la que se divide el AE. Las UGs son cuadriláteros de la misma área y contienen la información de presencia/ausencia de las especies recolectadas.

3. **Rejilla (Rj):** Cuando se define el tamaño de la AE y el tamaño de las UGs. El AE es dividida en tantas UGs como pueda contener (no importando si dicha UG se considera o no dentro de los análisis).
4. **Geoframe (GF):** Rejilla en la cual se han seleccionado las UG's que participarán en los cálculos de los análisis. Una definición de rejilla puede tener varias definiciones de *Geoframe*. La funcionalidad del *geoframe* es la posibilidad de excluir UGs, ya sea porque no tienen datos de recolecta y/o por algún criterio del usuario.
5. **Localidad:** Se refiere al punto geográfico de la localidad de recolecta. Permiten crear la información de presencia/ausencia de las especies para cada rejilla definida. Se refiere aquí como 'punto geográfico' a las coordenadas geográficas asignadas a la localidad de recolecta, independientemente del nivel de resolución que pueda significar un punto geométrico dentro del análisis espacial

**Métodos Descriptivos**

**Cuantificación de la Diversidad**

**Notación:**

$m$ : Número de UG's en el *Geoframe*, *v.gr.* 10

$m_i$ :  $i$ -ésima UG

$LSp_i$ : Lista de especies en la  $i$ -ésima UG; *v.gr.* (*especie 1, especie 2, especie 3, ..., especie n*)

$|LSp_i|$ : Número de especies en la  $i$ -ésima UG (cardinalidad). *v.gr.* 5

**Diversidad Alfa ( $\alpha$ )**

Es el número de especies de determinada UG.

- **Fórmula:**  $\alpha_i = |LSp_i| \forall m_i$  donde  $i \in [1, m]$  ( $m_i$  es la  $i$ -ésima UG)
- **Entrada:** a)  $LSp_i$  de cada  $m_i$
- **Salida:** a) Un número entero para cada  $m_i$   
b) Mapa temático de  $C$  colores (intervalos de valores  $\alpha$ )

**Ejemplo:**

5	4	2	5
1		3	1
	5		1

	6-4
	3-2
	1
	N/A

*Diversidad alfa promedio ( $\alpha$ ):*

Es el promedio de todas las diversidades alfa.

- **Fórmula:**  $\bar{\alpha} = \frac{(\sum_{i=1}^m \alpha_i)}{m}$ .
- **Entrada:** a)  $\alpha_i \forall i \in [1, m]$  (Un número entero por cada UG)
- **Salida:** a) Un número entero para toda el AE

*Diversidad gama ( $\gamma$ )*

Es el número total de especies en el AE.

Fórmula :  $\gamma = \left| \prod_{i=1}^m LSp_i \right|$ .

- **Entrada:** a)  $LSp_i \forall m_i \mid i \in [1, m]$  (las listas de las especies de cada UG)
- **Salida:** a) Un número entero para toda el AE

*Diversidad beta:*

Es la tasa de cambio de las especies con respecto a las UG's.

*(Whittaker, 1960 en Wilson & Shmida, 1984)*

- **Fórmula:**  $\beta = \frac{\gamma}{\bar{\alpha}} - 1$ .
- **Entrada:** a)  $\bar{\alpha}$   
b)  $\gamma$
- **Salida:** a) Un número real para toda el AE

(Whittaker, 1972 en Mourelle & Ezcurra, 1997)

- **Fórmula:**  $\beta_w = \frac{k}{\ln 2}$ ,

donde k es la constante del ajuste a la curva exponencial del tipo:  $y = be^{-kx}$

- **Procedimiento:** Se obtiene la matriz de similitud, dado un criterio ( $S_{ij}$ )

<i>l</i>	<i>i</i>	<i>m</i>		
<i>A1</i>	<i>A2</i>	<i>A3</i>	<i>A4</i>	
<i>l</i>	$S_{1,2}$	$S_{1,3}$	$S_{1,4}$	<i>A1</i>
	<i>l</i>	$S_{2,3}$	$S_{2,4}$	<i>A2</i>
		<i>l</i>	$S_{3,4}$	<i>A3</i>
			<i>l</i>	<i>A4</i>

Se obtiene la matriz pero ahora de distancias entre todas la UG's (centros geométricos).

<i>l</i>	<i>i</i>	<i>m</i>		
<i>A1</i>	<i>A2</i>	<i>A3</i>	<i>A4</i>	
<i>l</i>	$X_{1,2}$	$X_{1,3}$	$X_{1,4}$	<i>A1</i>
	<i>l</i>	$X_{2,3}$	$X_{2,4}$	<i>A2</i>
		<i>l</i>	$X_{3,4}$	<i>A3</i>
			<i>l</i>	<i>A4</i>

Se tabulan de la siguiente forma:

$m_{ij}$	$S_{ij}$	$X_{ij}$
$m_{1,1}$	$S_{1,1}$	$X_{1,1}$
$m_{1,2}$	$S_{1,2}$	$X_{1,2}$
....	...	...
$m_{i-1,j-1}$	$S_{i-1,j-1}$	$X_{i-1,j-1}$

Considerando que la columna  $S_{ij}$  es la variable dependiente de la columna  $X_{ij}$ , se ajusta a una curva exponencial a través de una regresión lineal mediante mínimos cuadrados ( $S = be^{-kx}$ )

- **Entrada:** a) Criterio de similitud a utilizar
  - b) Matriz de similitud completa
  - c) Matriz de distancias entre UG's
- **Salida:** a) Un número real para toda el AE

## Estimadores de Riqueza

Los estimadores de riqueza son una manera de calcular una aproximación al número total de especies presentes en AE.

### Notación :

$Sp_{infr}$ : Número de especies infrecuentes (aquellas presentes en 10 o menos UG).

$Sp_{rec}$ : Número de especies frecuentes (aquellas presentes en 11 o más UG).

$S_{obsf}$ : Número de especies en  $m_i$ .

$Q_j$ : Número de especies que están en  $j$  UG's ( $Q_1$  Es el número de especies que se colectaron en una sola UG;  $Q_2$  el número de especies que se recolectaron en dos UGs; ...).

$m$ : Número de UGs en el análisis.

$m_i$ :  $i$ -ésima UG.

$m_{infr}$ : Número de UG que tienen cuando menos una especie infrecuente.

$p_k$ : Proporción de UG que contienen a la especie  $k$ -ésima.

$B$ : Constante en la fórmula del estimador MIMMean

- Salida para todos los estimadores: Un número real para toda el AE.

### ICE:

- **Fórmula:**  $S_{est} = Sp_{rec} + \frac{Sp_{infr}}{C_{ice}} + \frac{Q_1}{C_{ice}} \gamma_{ice}^2$ ,

donde:

$$C_{ice} = 1 - \frac{Q_1}{N_{infr}}$$

$$\gamma_{ice}^2 = \max \left\{ \frac{Sp_{infr}}{C_{ice}} \frac{m_{infr}}{(m_{infr} - 1)} \frac{\sum_{j=1}^{10} j(j-1)Q_j}{(N_{infr})^2} - 1, 0 \right\}$$

$$N_{infr} = \sum_{i=1}^m (Q_i * i)$$

- **Entrada:** Cálculo de los parámetros en cada caso

*Chao 2:*

- **Fórmula:**  $S_{est} = S_{obs} + \frac{Q_1^2}{2Q_2}$
- **Entrada:** Cálculo de los parámetros.

*Jack 1:*

- **Fórmula:**  $S_{est} = S_{obs} + Q_1 \frac{m-1}{m}$
- **Entrada:** Cálculo de los parámetros

*Jack 2 :*

- **Fórmula:**  $S_{est} = S_{obs} + Q_1 \left( \frac{2m-3}{m} \right) - Q_2 \left( \frac{(m-2)^2}{m(m-1)} \right)$
- **Entrada:** Cálculo de los parámetros

*Boot:*

- **Fórmula:**  $S_{est} = S_{obs} + \sum_{k=1}^{\alpha} (1-p_k)^2$
- **Entrada:** Cálculo de los parámetros

*MMMean:*

- **Fórmula:**  $S(m) = \frac{S_{max} m}{B + m}$

- **Entrada:**  $\alpha_i$  para cada  $m_i$  donde  $i \in [1, m]$

$m_1$	$\alpha_1$
$m_2$	$\alpha_2$
....	...
$m_i$	$\alpha_i$

- **Procedimiento:** Transformar los pares  $(m_i, \alpha_i)$  en  $(\frac{1}{m_i}, \frac{1}{\alpha_i})$ ; realizar ajuste de mínimos cuadrados ( $y = ax + b$ ), donde:

$$\text{Pendiente} = \frac{B}{S_{\max}}$$

$$\text{Ordenada\_al\_origen} = \frac{1}{S_{\max}}$$

## Calidad de Muestreo (Qs)

• **Fórmula:** 
$$Q_s = \frac{I}{S_{obs} \cdot m / (1 - E_s) - \max(\alpha_{min})}$$

donde:

$$E_s = \frac{S_{est} - S_{obs}}{S_{est}} \qquad I = \sum_{j=1}^m j \cdot q(j)$$

$q(j)$  es el número de especies que están en  $j$  áreas

$I$  es el número de incidencias ( $I_s$ ) en la matriz de presencia/ausencia)

- **Entrada:** a) Cálculo de los parámetros; el usuario debe seleccionar un índice de estimación de riqueza
- **Salida:** a) Un número real para toda la AE

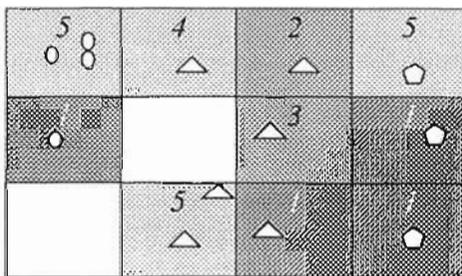
*Despliegue e Informes*

Ubicación de Localidades de recolecta

- **Entradas:** Localidades contenidas en UG'
- **Salidas:** Mapa con marcadores clasificados (por forma) según los siguientes criterios:
  - Número de especies.
  - Número de visitas.
  - Número de colectores.
- **Comentarios:** Los puntos no son sensibles al click. Se puede ver el mapa de diversidad alfa, en estos momentos las UG's son dinámicas, pero una vez definida la rejilla ya no lo son.

**Ejemplo:**

Mapa de localidades por número de visitas (sobre Mapa Alfa)



Alfa      Loca-      Número  
 Alfa      lidad      de visitas

	6-5	0	5
	3-2	△	4 a 3
	1	◻	2 a 1
	N/A		

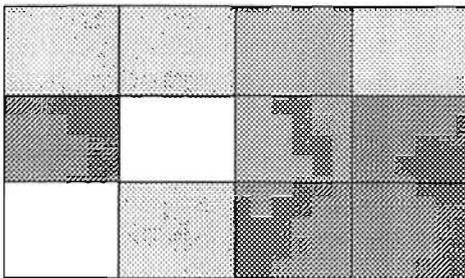
## Especies → Áreas

¿Donde está un determinado conjunto de especies? (Área de distribución)

- **Entradas:** a) Lista de  $k$  especies |  $k \in [1, \gamma]$   
 b) Criterio lógico AND/OR (AND =Presencia/ausencia, OR=Grados de Presencia del conjunto de especies).
- **Salidas:** AND: Mapa temático de tres colores (presencia, ausencia, no aplica) ó  
 OR: Mapa temático de  $n$  colores (grado de presencia)

### Ejemplo:

Mapa de distribución de un conjunto de especies (OR)



% de especies presentes

	100%	<i>Lista de Especies:</i> <i>Especie 1</i> <i>Especie 2</i> <i>Especie 3</i> <i>Especie n</i>
	99-1%	
	0%	
	N/A	

## Áreas → Especies

*¿Qué especies hay en un determinado conjunto de UGs?*

a) *En un conjunto de UG*

- **Entradas:** a) Lista de  $j$  UGs  $| j \in [1, m]$   
 b) Criterio lógico AND/OR (AND=cada especie está en cada UG, OR=cada especie está en alguna UG).
- **Salidas:**  $LSp_{\{0,n\}}$ , Lista de especies de un intervalo de UG's (unión o intersección)

**Ejemplo:**

Mapa de las UG's seleccionadas




<p><i>Lista de Especies:</i></p> <p><i>Especie 1</i></p> <p><i>Especie 2</i></p> <p><i>Especie 3</i></p> <p><i>Especie n</i></p> <p><i>Operador: AND</i></p>
--

b) *En toda el área*

- **Entrada:**  $LSp_i | i \in [1, m]$ .
- **Salida:**  $\prod_{i=1}^m LSp_i$ , donde  $\left| \prod_{i=1}^m LSp_i \right| = \gamma$ .

**Ejemplo:**

<i>Lista de Especies:</i>
<i>Especie 1</i>
<i>Especie 2</i>
<i>Especie 3</i>
...
<i>Especie <math>\gamma</math></i>

## UG

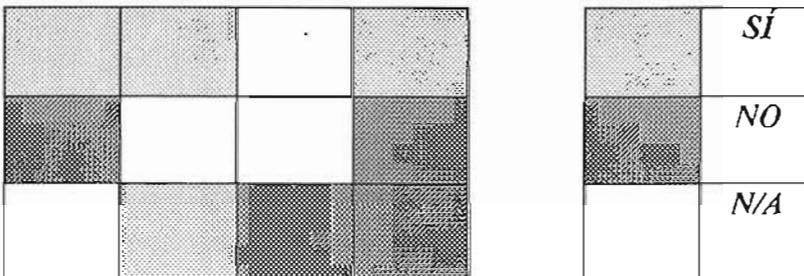
### ¿Cuáles UG's tienen riqueza $r$ ? ( $P_i$ )

Cuáles UG contienen  $n$  especies, donde

- **Entrada:**  $r \mid r \in [1, \gamma]$ .
- **Salida:** Mapa temático de tres colores (sí/no/NA).

**Ejemplo:**

*Mapa de las UG's con solo dos especies ( $r=2$ )*



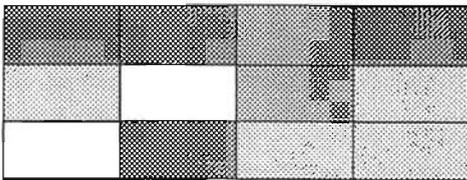
## Especies

¿Cuáles especies se distribuyen en  $n$  UG's? ( $Q_n$ )

- **Entradas:**  $n \mid n \in [1, m]$ .
- **Salida:**
  - a) Lista de especies  $Q_n$ .
  - b) Mapa temático de  $Q_n$  en función de cuántas especies de la clase  $Q_n$  contiene.

**Ejemplo:**

Mapa de riqueza de especies  $Q_2$



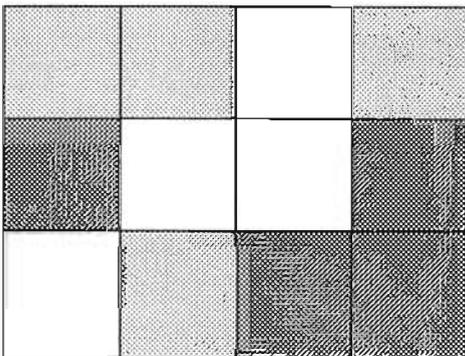
7 a 5	<b>Lista de Especies</b> $Q_2$ : Especie 1 Especie 2 Especie 3 ... Especie $k$
4 a 1	
0	
N/A	

## Entre UG's

¿Cuál es la similitud entre dos conjuntos de UG's?

- **Entrada:** a) Dos conjuntos de UG's  
b) Selección de criterio de similitud
- **Salida:** a) Valor de la similitud (número real)  
b) Lista de especies comunes y totales (intersección y unión)

Ejemplo:



Similitud entre UG's grises y negras = 75%.

<b>Intersección</b>	<b>Lista de especies</b> <b><u>INTERSECCIÓN</u></b> Especie 1 ... Especie <i>k</i>
<b>Unión</b>	<b><u>UNIÓN</u></b> Especie <i>k+1</i> ... Especie <i>n</i>

Criterio de Similitud: Fager.

**Métodos Analíticos**

**Índices de Similitud**

Permiten obtener mapas de similitud de UG's, además de ser utilizarse en los agrupamientos.

**Notación:**

- a: *Entero:* número de taxones (atributos) presentes en ambas OGU's
- b: *Entero:* número de taxones en sólo la primera OGU
- c: *Entero:* número de taxones en sólo la segunda OGU
- max(b, c): *Función:* máximo de dos valores enteros
- min(b, c): *Función:* mínimo de dos valores enteros
- / *División*
- X<sup>n</sup> *Raíz cuadrada de X*

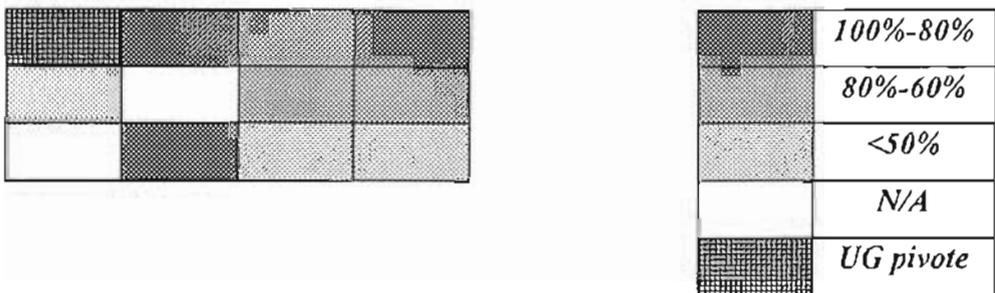
<i>Simpson</i>	$a/\min\{(a+b),(a+c)\}$
<i>Jaccard</i>	$a/(a+b+c)$
<i>Braun-Blanquet</i>	$a/\max\{(a+b),(a+c)\}$
<i>Sorensen-Dice</i>	$a/[a+1/2(b+c)]$
<i>Kulezynski 1</i>	$a/(b+c)$
<i>Kulezynski 2</i>	$1/2[a/(a+b) + a/(a+c)]$
<i>Fager</i>	$a/[(a+b)(a+c)]^{1/2} - 1/2[\max\{(a+b),(a+c)\}]$
<i>Otsuka</i>	$a/[(a+b)(a+c)]^{1/2}$
<i>Correlation Ratio</i>	$a^2/[(a+b)(a+c)]$

**Matriz de Similitud:**

- **Entrada:**  $LSp_i$  para todo  $m_i$ ; donde  $i \in [1,m]$ . Para observar un mapa de similitud con respecto a una UG hay que seleccionar la UG pivote (un renglón de la matriz de similitud)
- **Salida:**
  - a) Puede ser un mapa temático de  $n$  colores que represente un renglón de la matriz, dado un criterio
  - b) Toda la matriz a modo de informe
  - c) Matriz coloreada

**Ejemplo:**

Salida a) Mapa de similitud respecto a una UG pivote



Salida b) Matriz Informe

UG 1	UG 2	UG 3	UG 4	UG 5	UG 6	UG 7	UG 8	UG 9	UG 10	UG 11	UG 12	
100	0	0	0	0	0	0	0	0	0	0	0	UG 1
	100	0	0	0	50	50	0	100	50	33	0	UG 2
		100	50	0	50	0	50	0	0	33	0	UG 3
			100	0	33	33	100	0	33	25	0	UG 4
				100	0	0	0	0	0	0	0	UG 5
					100	33	33	50	33	60	0	UG 6
						100	33	50	100	25	0	UG 7
							100	0	33	25	0	UG 8
								100	50	33	0	UG 9
									100	33	0	UG 10
										100	33	UG 11
											100	UG 12

Salida c) Matriz Coloreada

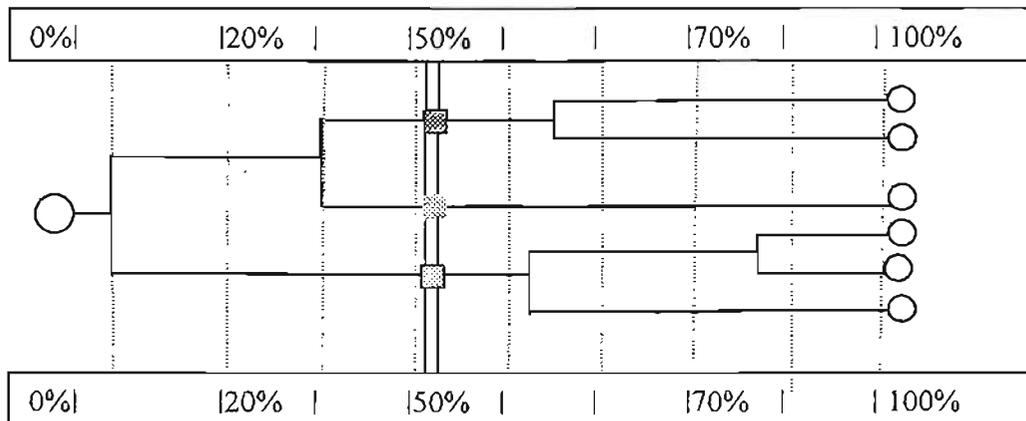
UG 1	UG 2	UG 3	UG 4	UG 5	UG 6	UG 7	UG 8	UG 9	UG 10	UG 11	UG 12	
100	0	0	0	0	0	0	0	0	0	0	0	UG 1
	100	0	0	0	50	50	0	100	50	33	0	UG 2
		100	50	0	50	0	50	0	0	33	0	UG 3
			100	0	33	33	100	0	33	25	0	UG 4
				100	0	0	0	0	0	0	0	UG 5
					100	33	33	50	33	60	0	UG 6
						100	33	50	100	25	0	UG 7
							100	0	33	25	0	UG 8
								100	50	33	0	UG 9
									100	33	0	UG 10
										100	33	UG 11
											100	UG 12

ESTA TESIS NO SALE  
DE LA BIBLIOTECA

## Análisis de Agrupamiento

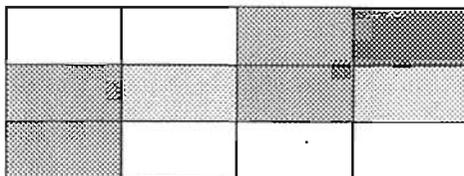
- **Entrada:**
  - a) Índice de similitud
  - b) Nivel de corte (opcional)
- **Salida:**
  - a) Un fenograma en caso de no especificar nivel de corte
  - b) Un mapa temático de agrupamiento de UG dado un nivel de corte

### Salida a) Fenograma:

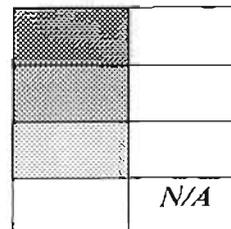


### Salida b) Mapa temático de agrupamiento de UG dado un nivel de corte

Mapa de UG's por clase; *nivel de corte = 47%*



*Clases*



- **Ligamiento promedio:** La fusión de los renglones (o columnas) correspondientes a cada UG se recalcula mediante el promedio de las similitudes.
- **Ligamiento completo:** La fusión de los renglones (o columnas) correspondientes a cada UG se recalcula mediante la mayor de las similitudes.
- **Ligamiento simple:** La fusión de los renglones (o columnas) correspondientes a cada UG se recalcula mediante la menor de las similitudes.

## Elección de áreas prioritarias para la conservación

- **Procedimiento:**

Método de Rebelo (1994). Consiste en elegir una lista del mínimo de UGs que contenga a la mayor cantidad de especies de la región.

Supóngase la siguiente matriz de presencia/ausencia (las letras representan las UGs y los números a las especies):

Sp <sub>i</sub> , m <sub>i</sub>	UG 1	UG 2	UG 3	UG 4	UG 5
SP 1	X	X	X	X	
SP 2		X	X	X	
SP 3			X		
SP 4				X	X

Se procede a calcular un número de importancia de cada especie a partir del número de UG en que se encuentra:

Importancia  
de cada SP

Sp <sub>i</sub> , m <sub>i</sub>	UG 1	UG 2	UG 3	UG 4	UG 5	Suma	(S/Suma)
SP 1	X	X	X	X		4	1.25
SP 2		X	X	X		3	1.67
SP 3			X			1	5.0
SP 4				X	X	2	2.5

Se le asigna una importancia a cada UG: suma de las importancias de especies que contiene :

*Iteración 1*

UG 1	UG 2	UG 3	UG 4	UG 5	No. SP = 5
1.25	2.92	7.92	5.42	2.5	$\Sigma(S/Suma)$

UG elegida para la reserva: UG 3 (7.92)

En la siguiente iteración solo se suman las importancias de las especies que no están en una UG elegida como reserva (no se suman las especies 1, 2 y 3 porque ya se encuentran en la UG 3), en este caso sólo cuenta la especie 4, que tiene un valor de importancia de 2.5.

Importancia  
de cada SP

Sp./m.	UG 1	UG 2	UG 3	UG 4	UG 5	No. SP = 1
SP 4				2.0	2.0	$\Sigma(4/Suma)$

En este caso las dos UGs (4 y 5) tienen el mismo valor de importancia (2.0), por lo que se deja la decisión al “usuario” o experto cuál de las dos elegir.

UG elegida para la reserva: ITERACIÓN 1: UG 3 (7.92)

ITERACIÓN 2: UG 4 (2.0) o UG 5 (2.0)

Se propone conservar a dos UGs (UG 3 y UG 4; o UG 3 y UG 5) pues en estas dos UGs se captura toda la diversidad de la región.

- **Entradas:** Datos solicitados y el número de la iteración de paro (opcional)
- **Salida:** Mapa temático con un orden cardinal.

**Ejemplo:**

Mapa de UG's por orden de prioridad: *Prioridad*

1	2
2	2
	2

	1
	2
	N/A

### 9.3. Un modelo de la BC: Caminos metodológicos

En esta sección se explica un modelo que integra a los métodos de la BC desde el punto de vista de la secuencia de su aplicación en el análisis biogeográfico. Por razones didácticas, primero se expone propiamente el modelo y después la manera en que construyó, al generar una correspondencia – de entre varias posibles - con los niveles de abstracción en la generación del conocimiento propuestos por Salazar (1979).

#### *El Modelo: caminos metodológicos*

El fundamento del modelo es permitir el encadenamiento de algunos métodos, es decir, los resultados de un análisis pueden ser parámetros de otro método. La figura 9.3 muestra esta idea de manera gráfica.

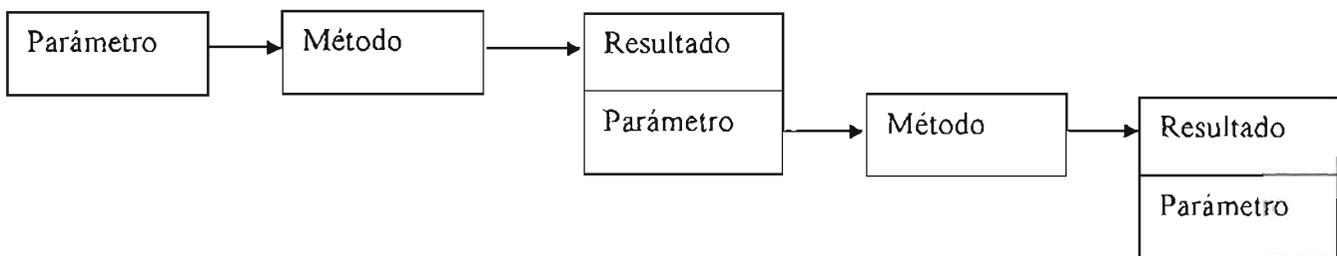


Figura 9.3 Esquemización del encadenamiento entre métodos para constituir ‘caminos metodológicos’.

El encadenamiento es simple de concebir si se consideran solo a aquellos métodos que tienen como entrada y como salida, ya sea una lista de especies o una lista de UGs. La figura 9.4 indica, para cada método, el tipo de entrada y el tipo de salida, además de un símbolo que lo identifica y que se utiliza en la *gramática* de los caminos metodológicos. La figura 9.5 muestra los métodos agrupados por el tipo de entrada y el tipo de salida. Los métodos con salida tipo UG son los que siguiendo a Haining (1994) se llaman ‘preguntas espaciales’, mientras que los que

tienen una salida tipo SP son las ‘no espaciales’. Se observa que, como indica Haining, ambos tipos de preguntas consideran datos espaciales y deben responderse por sistemas geográficos, no pudiendo hacerlo los sistemas no espaciales.

Símbolo	Método	Tipo de Entrada	Tipo de Salida
Alfa	Mapa Alfa	UG	UG
Sim	Renglón de la Matriz de Similitud	UG	UG
Liga	Ligamiento	UG	UG
Rebelo	Elección de áreas para la conservación	UG	UG
SpUg	Dónde está determinado conjunto de especies (Especies → Areas)	SP	UG
UgSP	Qué especies hay en un conjunto de UGs (Areas → Especies)	UG	SP
TotUgSp	Lista de especies de toda el área	UG	SP
UgnSpUg	Cuáles UGs contienen n especies (UGs)	UG,n	UG
SpnUgSp	Cuáles especies están en n UGs (Especies)	SP,n	SP
UgvsUgSp	Cuál es la similitud entre dos conjuntos de UGs	UG,UG	SP

Figura 9.4 Simbología usada por la matriz ‘gramática’ para caminos metodológicos.

entrada→salida	Total	Métodos
UG→UG	5	Alfa, Sim, Liga, Rebelo, UgnSp
UG→SP	3	UgSp, TotSp, UgvsUGSp
SP→UG	1	SpUg
SP→SP	1	SpnUGSp

Figura 9.5 Métodos agrupados según el tipo de entrada y el tipo de salida.

Para establecer la correspondencia entre las estradas y salidas de los métodos, se especifican las posibilidades del encadenamiento mediante una gramática. La gramática (figura 9.6) se representa mediante una matriz, en la que los encabezados de las columnas representan los métodos de los que provienen los parámetros y los renglones representan los métodos que reciben esos parámetros. En cada celda de la matriz se indica si es o no posible el encadenamiento de ambos métodos (columna -> renglón).

		Entrada										
		Alfa	Sim	Liga	Rebelo	SpUg	UgSp	TotUgSp	UgnSpUg	SpnUgSp	UgvsUgSp	
Salida		GU	GU	GU	GU	GU	SP	SP	GU	SP	SP	
	Alfa	GU	S									
	Sim	GU		S								
	Liga	GU			S							
	Rebelo	GU				S						
	SpUg	SP					NA					
	UgSp	GU						NA				
	TotUgSp	NA							NA			
	UgnSpUg	GU								NA		
	SpnUgSp	SP									NA	
	UgvsUgSp	UG										NA

	Es Posible
NA	No es posible la recursividad
	No es posible
S	Recursividad Posible

Figura 9.6 Matriz-gramática para los caminos metodológicos.

Como se observa, la gramática permite la recursividad para algunos métodos, es decir, los resultados de un método pueden ser parámetros del mismo método. El significado para esos casos se esquematiza en la figura 9.8

	<b>Alfa</b>	<b>Sim...</b>	<b>Liga</b>	<b>Rebelo</b>
<b>Alfa</b>	Reclasificación- Selección de las UG basándose en su riqueza			
<b>Sim</b>		Navegación por la Matriz de Similitud		
<b>Liga</b>			Incremento en el nivel de corte del fenograma	
<b>Rebelo</b>				Refinamiento en el análisis de las áreas prioritarias, incluyendo sólo a UGs con cierta categoría de importancia

Figura 9.8 Esquemización del significado de la recursividad en los caminos metodológicos.

Los caminos metodológicos son una opción para integrar de manera formal a los métodos de la BC, permitiendo documentar al análisis biogeográfico de manera homogénea y transferir entre grupos de investigación metodologías de una manera eficiente. Estos caminos también sirven para profundizar en el análisis e integración de los métodos de la BC y para generar un esquema de diseño informático que permite su automatización.

### ***Correspondencia de los niveles de abstracción en la generación del modelo***

Como parte del análisis de los métodos de la BC que conduzca a una visión integral y coherente de ellos, se aplicó la metodología para la generación estructurada del conocimiento propuesta por Salazar (1979). La generación del modelo de los 'caminos metodológicos' surge del establecimiento de una correspondencia entre los niveles de abstracción de la generación estructurada del conocimiento (percepción, descripción, clasificación, relaciones directas, concatenación y formulaciones o enunciados) con los métodos de la BC.

El mapeo establecido es:

- 1) *Percepción*: Conformar una lista de métodos a incluir en el análisis.
- 3) *Descripción*: Describir de manera homogénea, práctica y acotada a cada uno de ellos.
- 4) *Clasificación*: Identificar, con base en la descripción, elementos en común a los métodos que permita integrarlos o concatenarlos.
- 5) *Relaciones directas*: Identificar algunas relaciones directas entre los métodos
- 6) *Concatenación*: Con base en las relaciones directas identificadas, intentar establecer relaciones transitivas entre ellas.
- 7) *Formulaciones*: A partir de la concatenación propuesta en el paso anterior, intentar generar una fórmula operacional de los métodos de la BC.

Los resultados de aplicar esa correspondencia para cada nivel de abstracción se resumen como:

#### ***1) Percepción***

Se generó una lista de métodos de la BC, acotándola mediante restricciones que permitieron su análisis dadas las dimensiones y recursos disponibles del estudio.

#### ***3) Descripción***

Se generaron fichas para cada uno de los métodos que los describen de una manera homogénea y práctica. Las fichas se ordenan de acuerdo con una clasificación, siguiendo

también el criterio de niveles de abstracción y otras aplicadas para los procesos en el área de los Sistemas de Información Geográfica.

#### **4) Clasificación**

Se identificó cierta homogeneidad en la estructura de las entradas y de las salidas, que permitió clasificarlos (identificarlos) según respondan a preguntas ‘espaciales’ y ‘no espaciales’

#### **5) Relaciones directas**

Como relación directa entre cada par de métodos se estableció que el resultado de uno puede ser la entrada del otro, para lo que también se construyó una gramática que especifica si se dá este tipo de relación y, en su caso, el tipo de relación entre cada par de métodos (UG-> UG, UG->SP, SP->UG; y SP->SP).

#### **6) Concatenación**

Estableciendo relaciones transitivas, se observó que es posible concatenarlos, a lo que se le denominó ‘caminos metodológicos’.

#### **7) Formulaciones**

Se propuso una manera de integrar a los caminos metodológicos como herramienta de modelación en un SIG para la BC, permitiendo su operación como un proceso formal.

### ***Implementación de los caminos metodológicos en el GIS-BC***

El diseño del GIS para la BC considera la implementación de los caminos metodológicos; cada comando del *software* que aplica un método de la BC es representado internamente como un método o función que recibe determinado tipo de parámetro (SPs o UGs) y da como resultado otro (SPs o UGs)

El usuario puede guardar los caminos metodológicos en la base de datos; se almacenan bajo un nombre definido por el usuario y contienen asociada la lista-secuencia de métodos que lo conforman, así como las listas de resultados-parámetros (SPs o UGs) asociados a cada paso de la cadena.

Cada camino está asociado al *geoframe* en que se generó; un *geoframe* puede contener varios caminos metodológicos.

# 10. Desarrollo de un SIG para análisis biogeográfico

---

*En este capítulo se muestran los resultados del desarrollo de un SIG para la BC. Se esquematiza la manera en la que el sistema presenta ante el usuario los resultados y mostrando la flexibilidad del sistema para aceptar preguntas y peticiones. Se presenta la estructura de la base de datos y la estructura general del código en un lenguaje de programación de alto nivel.*

El desarrollo de la herramienta contempla un análisis de requerimientos, en donde se definen las características generales del sistema que guían en las etapas de diseño de:

- a) la interfaz
- b) la base de datos
- c) la estructura de código del programa

El diseño de la interfaz se realizó considerando las aplicaciones Windows como un estándar (menús “Pull-down”, diálogos, ventanas, uso de controles como listas, combos y botones) y su estructura lógica útil para la comunicación del sistema con los usuarios.

La base de datos se diseñó considerando el paradigma de las bases de datos relacionales de Codd (1990).

El diseño de la estructura del código se realizó considerando los paradigmas de la Programación Orientada a Objetos (POO) con las limitantes y ventajas que presenta el lenguaje de programación Visual Basic.

## 10.1. Especificaciones generales de la interfaz

Se definieron como requerimientos generales de la interfaz con el usuario, la inclusión de los siguientes elementos:

- Menú
- Ventana del Mapa
- Ventana de la Matriz de Similitud
- Ventana del Fenograma

### Menús

La estructura del menú debe agrupar de manera ordenada a las operaciones que permite realizar el sistema. Los comandos se agrupan en ocho submenús:

1. Archivo
2. Edición
3. Ver
4. Análisis
5. Preguntas
6. Herramientas
7. Ventana
8. Ayuda

En la figura 10.1 se muestra la estructura de los comandos en el menú.

<b>1) Archivo</b>		<b>2) Edición</b>		<b>3) Ver</b>	
Nueva AE...		Nombre de la UG		Acercar	
Abrir AE...		Copiar		Alejar	
Abrir Geoframe...				Ver retícula	
Importar Datos				Ver reglas	
Exportar Datos				Desplazamiento	
Guardar				Ver toda el AE	
Guardar como...				Información de la UG	
Imprimir					
Salir					
<b>4) Análisis</b>				<b>5) Preguntas</b>	
Mapa Alfa				Localidades	
Mapa de Similitud				Áreas de Distribución	
Matriz de Similitud->		Matriz coloreada		Lista de Especies ->	
Bioestadísticas		Matriz %		UG's por Riqueza	
Agrupamiento				Especies por Distribución	
Áreas Prioritarias				Similitud entre UG's	
				Limpiar	
<b>6) Herramientas</b>				<b>7) Ventana</b>	
Opciones				Mapa	
Paletas y Rangos				Mapa y simbología	
Definir Rejilla				Lista de especies	
Seleccionar UG ->		Selección manual			
Definir Geoframe		Selección por rangos			
Camino Metodológico		Seleccionar toda el AE			
Usuarios					

**Figura 10.1** Estructura del menú del SIG para la BC. Siguiendo la notación más usada, los tres puntos suspensivos (...) indican que al seleccionar el comando se desplegará un cuadro de diálogo. La flecha (->) indica que los sub-comandos a la derecha son desplegados al seleccionar ese comando.

## Ventana del Mapa

En esta parte se especificó un conjunto de operaciones gráficas sobre el mapa, así como su comportamiento y contenido. A continuación se detallan los requerimientos:

A) Operaciones gráficas del mapa:

- Acercar/Alejar (“Zoom”).
- Movimiento dentro del mapa (“Pan”).
- Cada UG Sensible al botón izquierdo para seleccionarla.
- Todo el mapa sensible al botón derecho para menús contextuales.
- Sensible a la posición del ratón para el despliegue de coordenadas y análisis por cada UG.

B) Capacidad para imprimir su contenido.

C) Los mapas que despliega deben de generarse con base en paletas de colores seleccionadas por el usuario.

D) Se debe poder visualizar la regla.

E) Debe aparecer algún elemento gráfico que denote la escala.

F) Debe aparecer algún elemento gráfico que denote el norte.

G) Se debe permitir el encendido/apagado de la retícula.

H) Se deben manejar todos los cuadrantes de la Tierra (despliegue UTM).

I) Capacidad de despliegue de cartografía (segunda versión del software).

J) Importación de archivos DXF (segunda versión del software).

## Ventana de la Matriz de Similitud

Las operaciones que se deben poder realizar sobre la matriz de similitud son:

A) Operaciones acercarse y alejarse (“Zoom”).

B) Desplazamiento (“Pan”).

C) Paleta de colores predeterminada.

## Ventana del Fenograma

Las operaciones que se deben poder realizar sobre el fenograma son:

- A) Operaciones de acercarse y alejarse (“Zoom”).
- B) Desplazamiento (“Pan”).
- C) Regla (valores de similitud) visible.
- D) Movimiento del nivel de corte dinámico (interactivo).

## Arreglo de las ventanas del ambiente

Se definieron 2 tipos generales de arreglos de las ventanas:

Pantalla Salida Tipo I: que incluye las ventanas del Mapa, Simbología y Lista de Especies

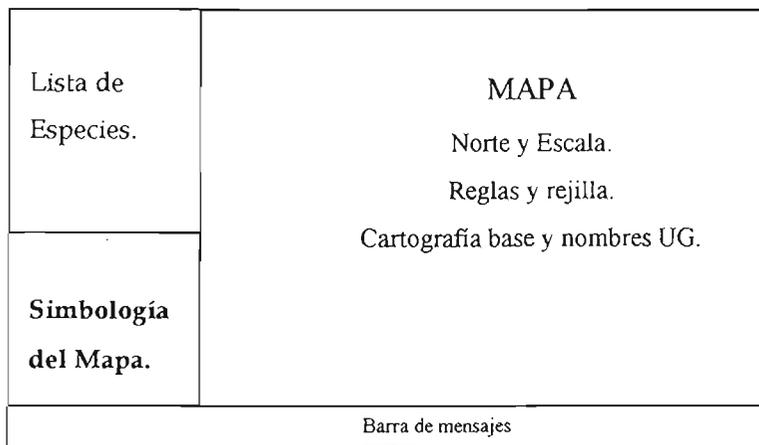
Pantalla Salida Tipo II: que incluye las ventanas del Mapa, Simbología y Fenograma

La pantalla Salida Tipo I contiene los subcasos:

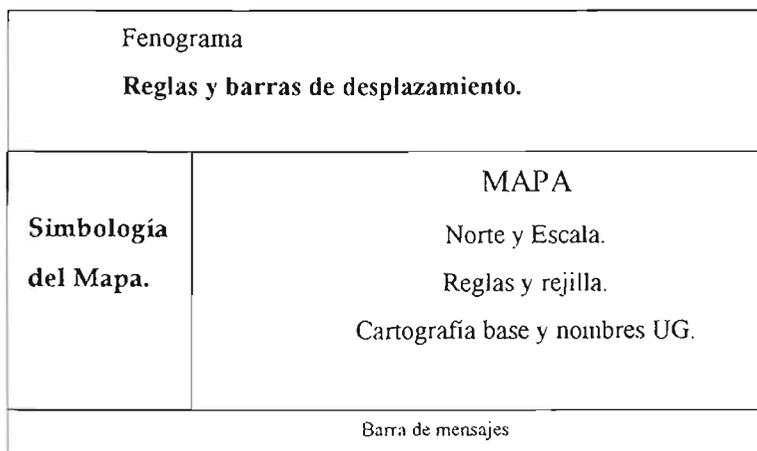
1. Maximizar mapa.
2. Maximizar mapa y simbología.
3. Maximizar lista especies.

La Pantalla de Salida Tipo II contiene los subcasos:

1. Maximizar mapa.
2. Maximizar mapa y simbología.
3. Maximizar fenograma.



a) Arreglo de ventanas Tipo I: Mapa, Simbología y Lista de especies.



b) Arreglo de ventanas Tipo II: Mapa, Simbología y Fenograma.

Figura 10.2 Tipos de arreglos de ventanas del ambiente del usuario

## 10.2. Diseño de la Interfaz

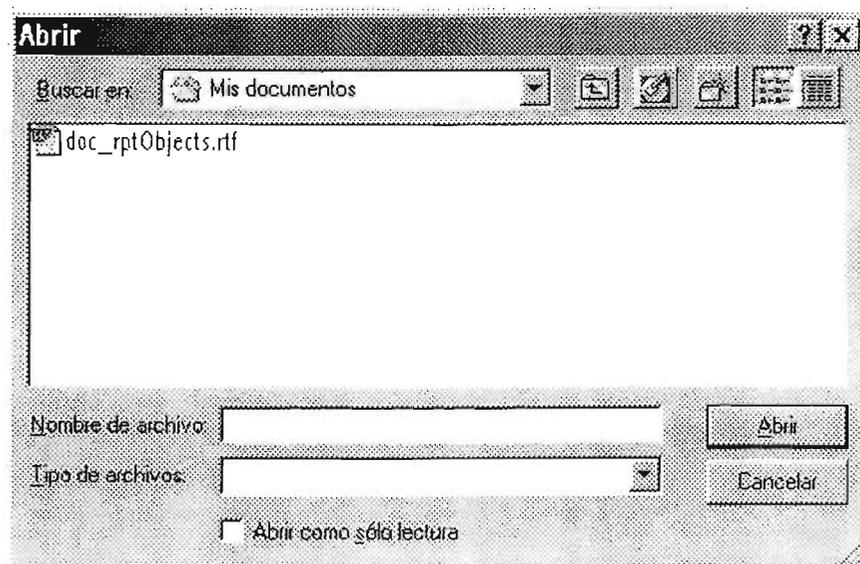
### Comandos del submenú *Archivo*

**Guardar** Es un menú que despliega un cuadro de diálogo que depende del contexto, se aclara en el menú herramienta.

#### **Nueva AE y Abrir AE:**

Despliega el cuadro de diálogo del tipo *Common Dialog* que permite navegar entre los archivos para crear o abrir una Base de Datos.

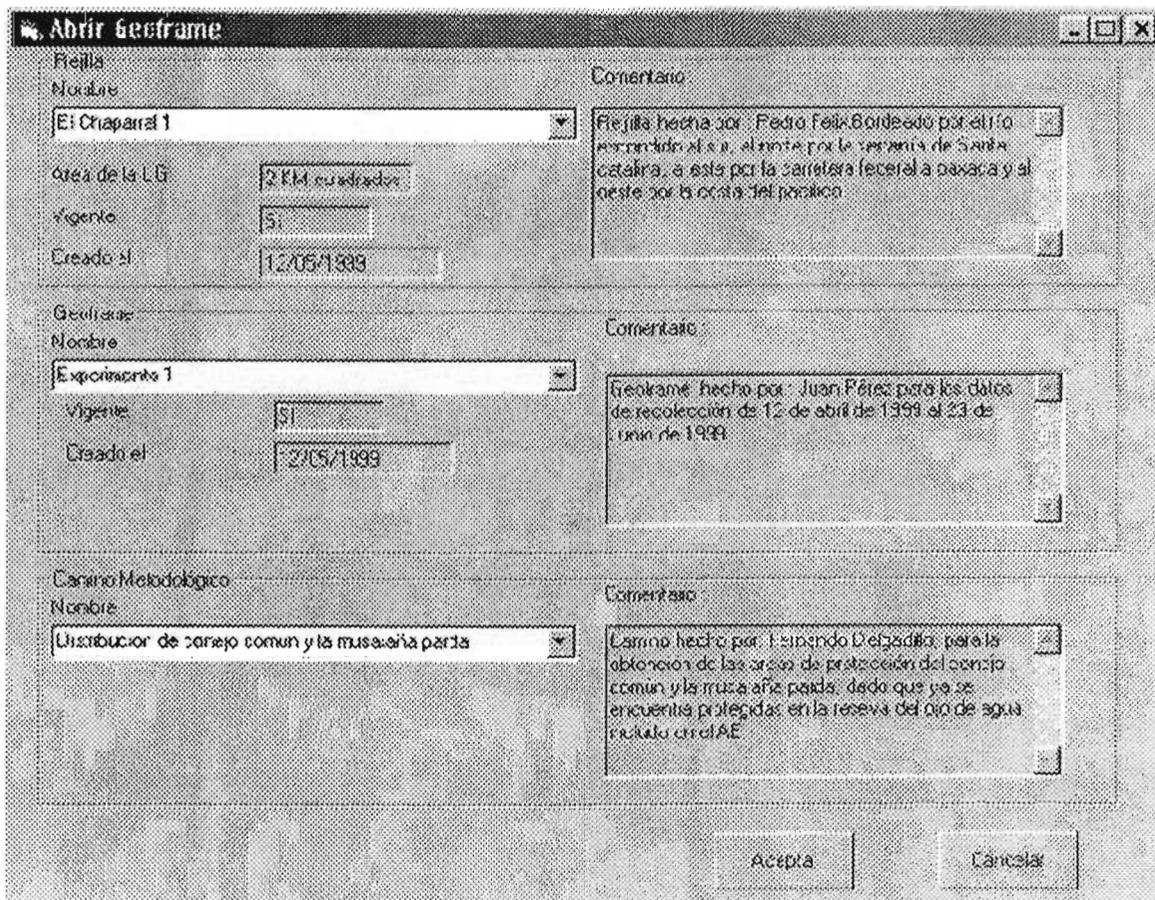
- No hay entradas previas.
- No hay análisis.
- **Abrir:** Da como salida un mensaje de AE abierta o error en caso de no poderla abrir.
- **Nueva** Da como resultado una BD datos completa.



### Abrir Geoframe

Cuadro de diálogo que permite abrir una rejilla, geoframe y el resultado de un camino metodológico.

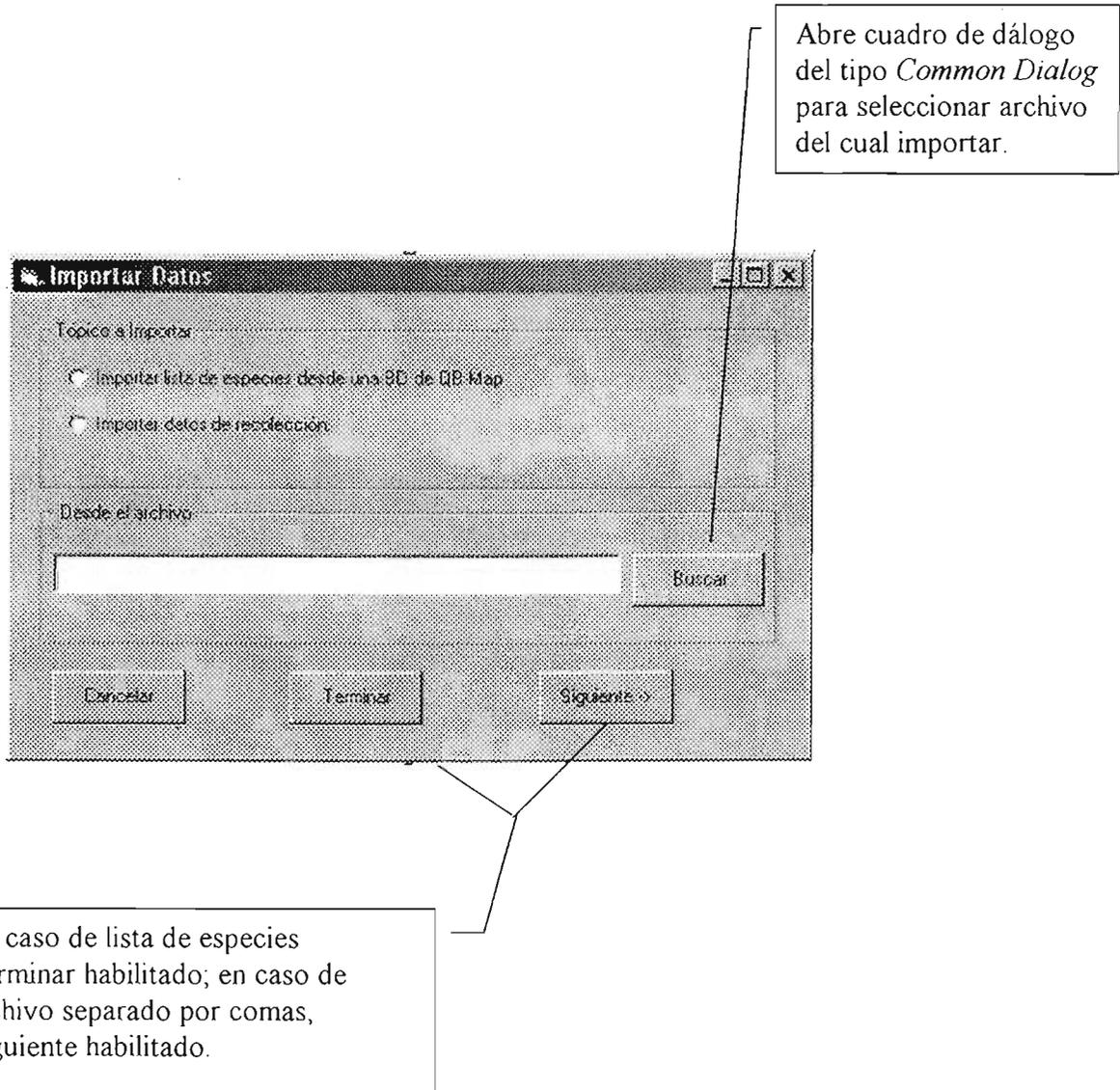
- No hay entradas previas.
- Da como salida una salida tipo I.



## Importar Datos

- Debe de estar abierta una rejilla, geoframe o camino metodológico.
- Da como salida un mensaje de importación exitosa o error en la importación.
- Se despliega en varios pasos dependiendo de las opciones que se escogen.

### Diálogo 1



Diálogo 2:

The dialog box 'Asignación de Campos' is used for mapping fields from a source file to a destination database table. It features two main tables: 'Campo del archivo fuente' and 'Campo Destino'. The 'Campo del archivo fuente' table has columns 'Campo' and 'Ejemplo\*', with rows for 'Especie', 'Loc', 'Lat g', 'Lat min', and 'Lat seg'. The 'Campo Destino' table has columns 'Campo', 'Fuente', and 'Ejemplo', with rows for 'Especie', 'Lat G', 'Lat M', 'Lat S', and 'Localidad'. A 'Concatenar' button is located between the tables. At the bottom, there are buttons for 'Cancelar', 'Terminar', and 'Continuar', along with a checkbox for 'Revisar Integridad'.

1ª fila del archivo de texto separado por comas ;contiene los nombres de los campos.

Lista de Esquemas de importación, guardados en la BD.

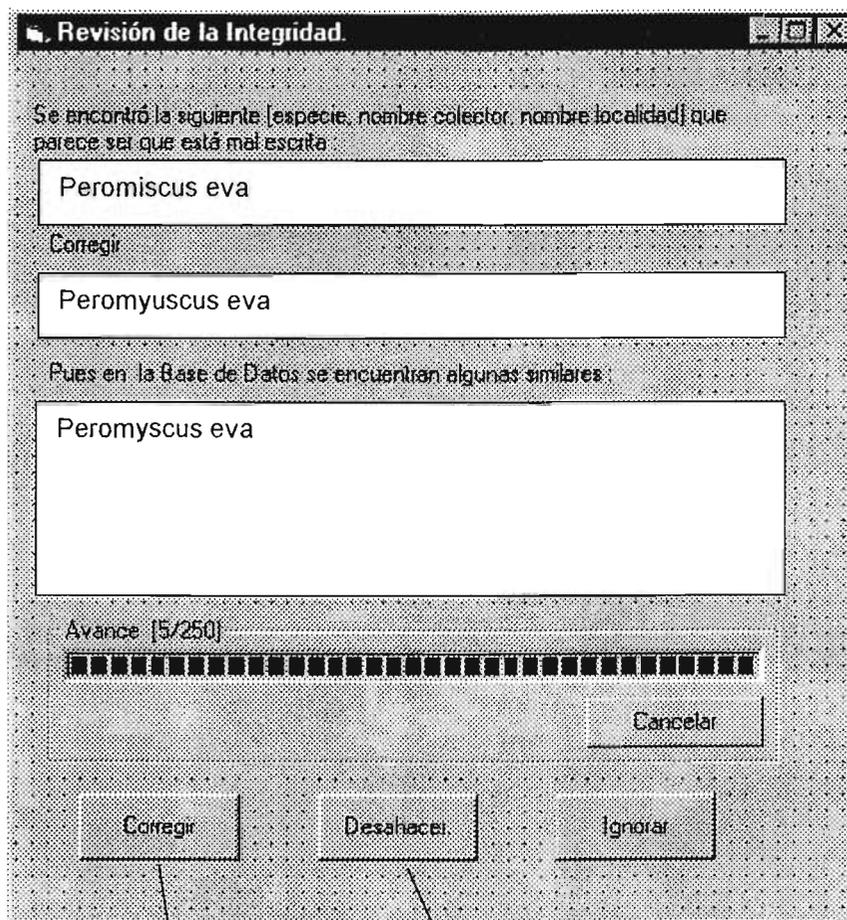
Correspondencia entre campos entre archivo de importación y base de datos.

Terminar habilitado si no escogió "revisión de integridad", se inicia importación y se muestra error en caso de haberlo. En caso de revisar continuar está habilitado.

Permite la concatenación: unión de 2 campos fuente en uno destino.

Permite revisar la integridad, e caso afirmativo crea tablas temporales y aparece tantas veces sea

Diálogo 3 (cuadro revisión):

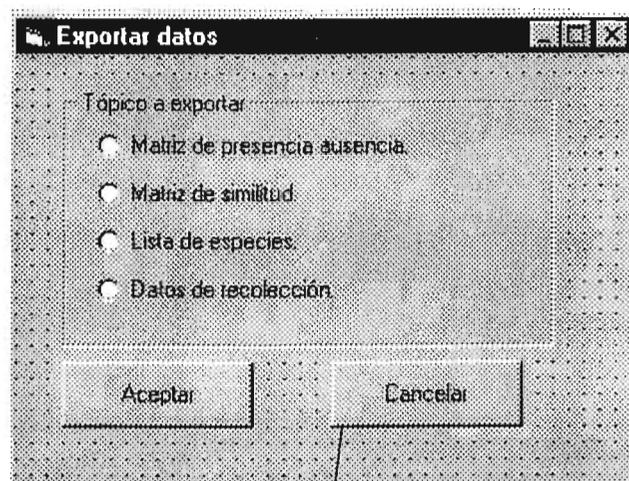


Vuelve a la palabra anterior.

Corrige con sugerencia, captura manual. Pasa al siguiente error detectado.

## Exportar Datos

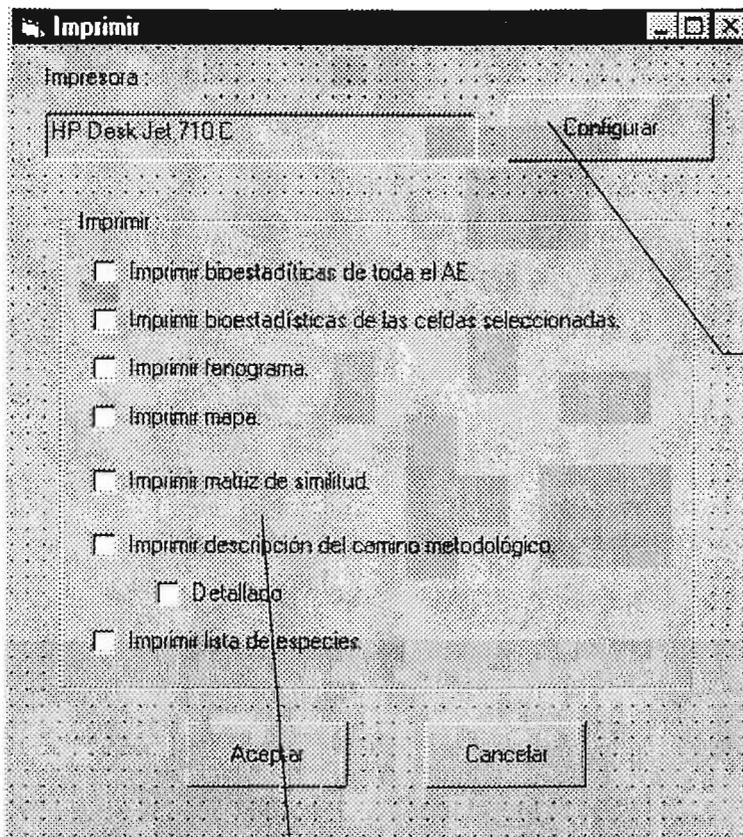
- Debe de haber una rejilla, geoframe o camino metodológico abierto.
- Da como salida un archivo separado por comas, con separador de regiones.



Al presionar aceptar se abre un cuadro de diálogo del tipo *Common Dialog* para seleccionar el archivo destino.

## Imprimir

- Es sensible al contexto, pues las opciones se habilitan y deshabilitan según lo que haya en la forma principal.
- Da como salida una impresión en papel.



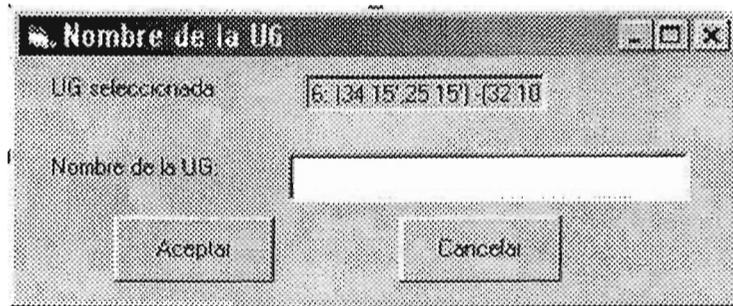
Despliega cuadro de diálogo de configurar impresora del tipo *Common dialog*.

- Bioestadísticas del AE siempre habilitado.
- Bioestadísticas de las UG seleccionadas sólo cuando haya más de una celda seleccionada.
- Fenograma sólo cuando esté visible.
- Mapa siempre que esté visible.
- Matriz de similitud siempre.
- Camino metodológico, sólo cuando esté activo uno.
- Detalle indica todos los parámetros y salidas.
- Especies: todas en caso de no haber ningún análisis, y las del resultado, en caso de tenerlas.

## Comandos del submenú *Edición*

### Nombre de la UG

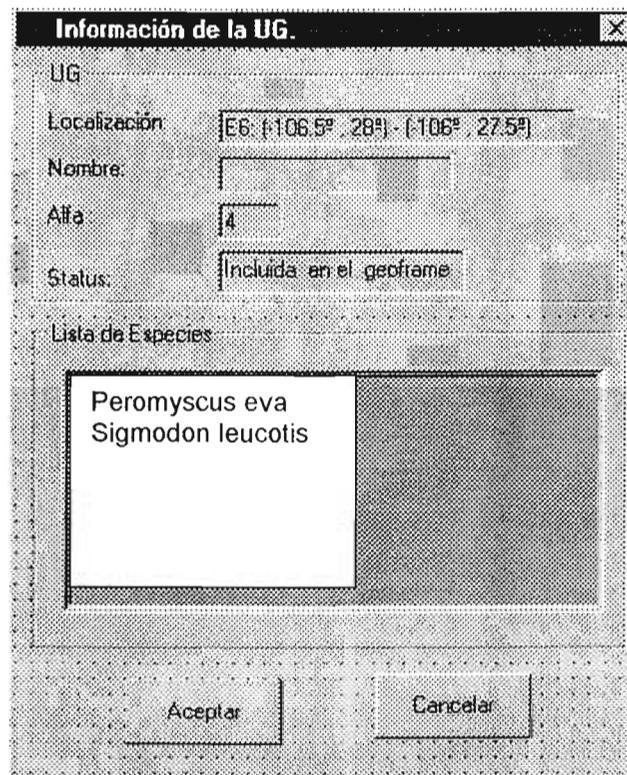
- Debe de haber seleccionada sólo una UG.



## Comandos del submenú Ver

### Información de la UG

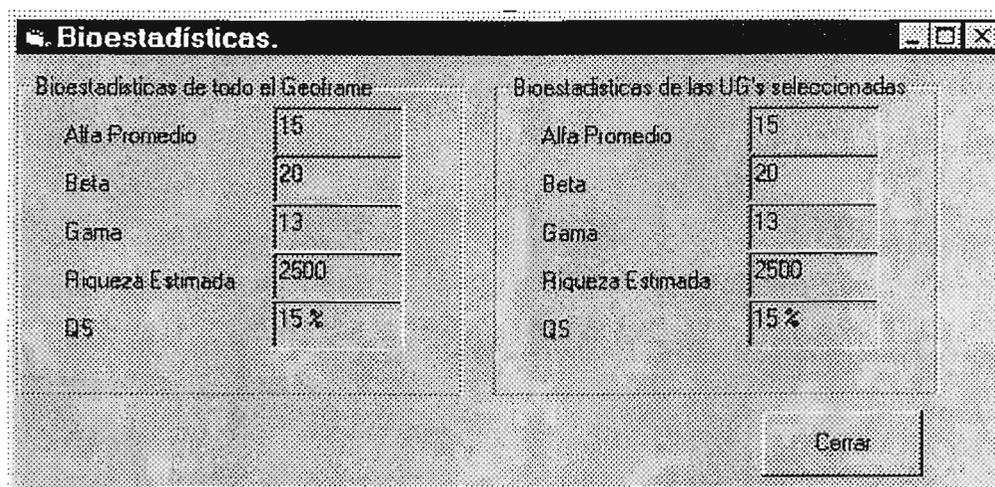
- Debe de haber una UG seleccionada, ya sea por selección normal con “seleccionar”, o aparece directamente al presionar el botón de la barra de herramientas de información y dar un click en la UG deseada. De esta manera se puede pedir información de una UG que no está incluida en el Geoframe.
- Se puede pedir información de UG's fuera del AE.
- Utiliza el análisis de alfa y lista de especies por UG.



## Comandos del submenú *Análisis*

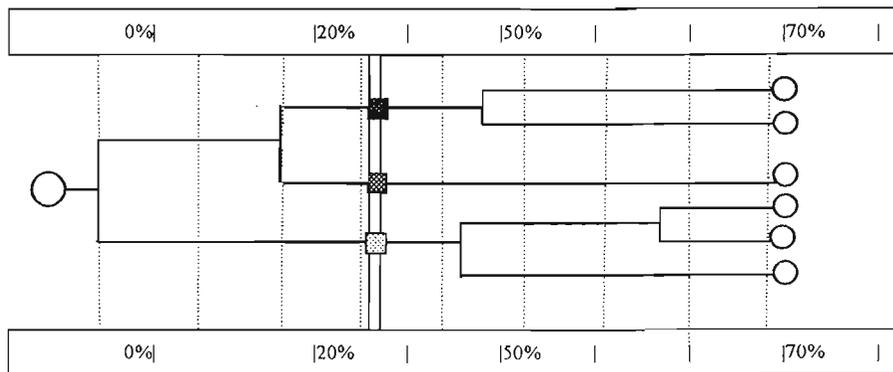
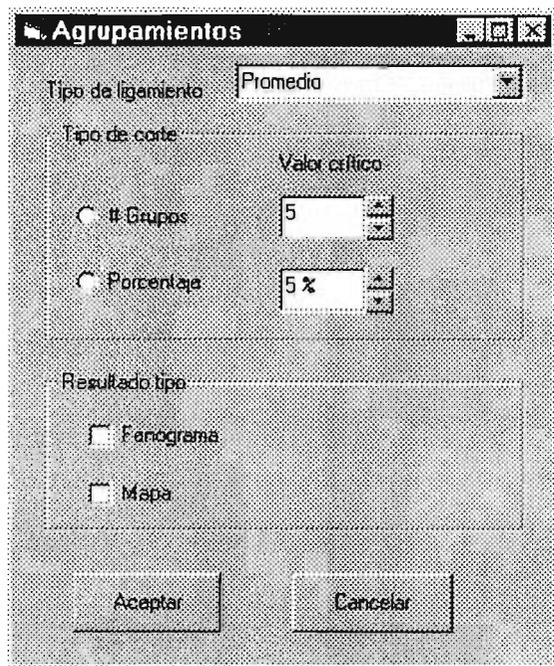
### Bioestadísticas

- Debe estar cargado un geoframe.
- Si hay más de una UG seleccionada se muestran las bioestadísticas del área seleccionada.
- Utiliza los análisis que se muestran en la figura.
- Depende del diálogo de opciones para seleccionar los índices a utilizar (para Beta y riqueza estimada).

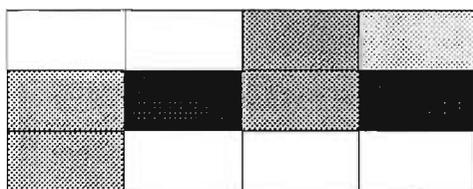


### Agrupamientos

- Debe haber un geoframe abierto.
- Genera el mapa con el fenograma y sus subcasos.
- El nivel crítico se puede modificar con un control de desplazamiento del fenograma.



Ejemplo del mapa (Nivel de corte =47%)



Leyenda

	<i>Gpo. 1</i>
	<i>Gpo. 2</i>
	<i>Gpo. 3</i>
	<i>N/A</i>

### Áreas prioritarias para la conservación

- Debe de haber un geoframe abierto.
- Genera un mapa con UG protegidas por iteración.
- Es iterativo, se corre paso a paso.



Ejemplo de mapa

1	2
2	2
	2

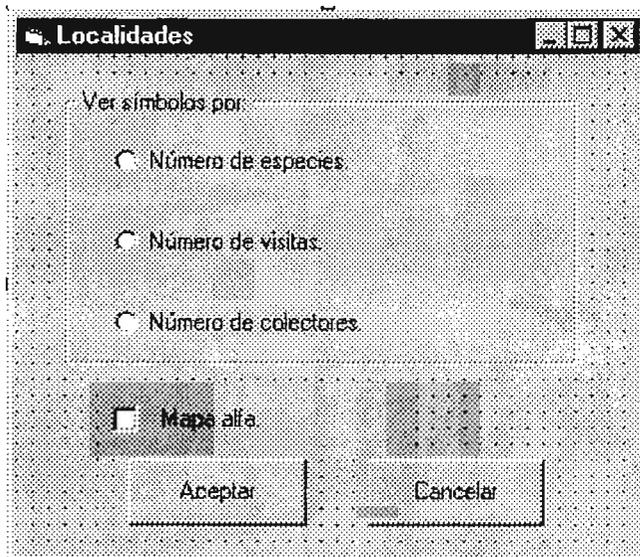
Leyenda

	1
	2
	N/A

## Comandos del submenú *Preguntas*

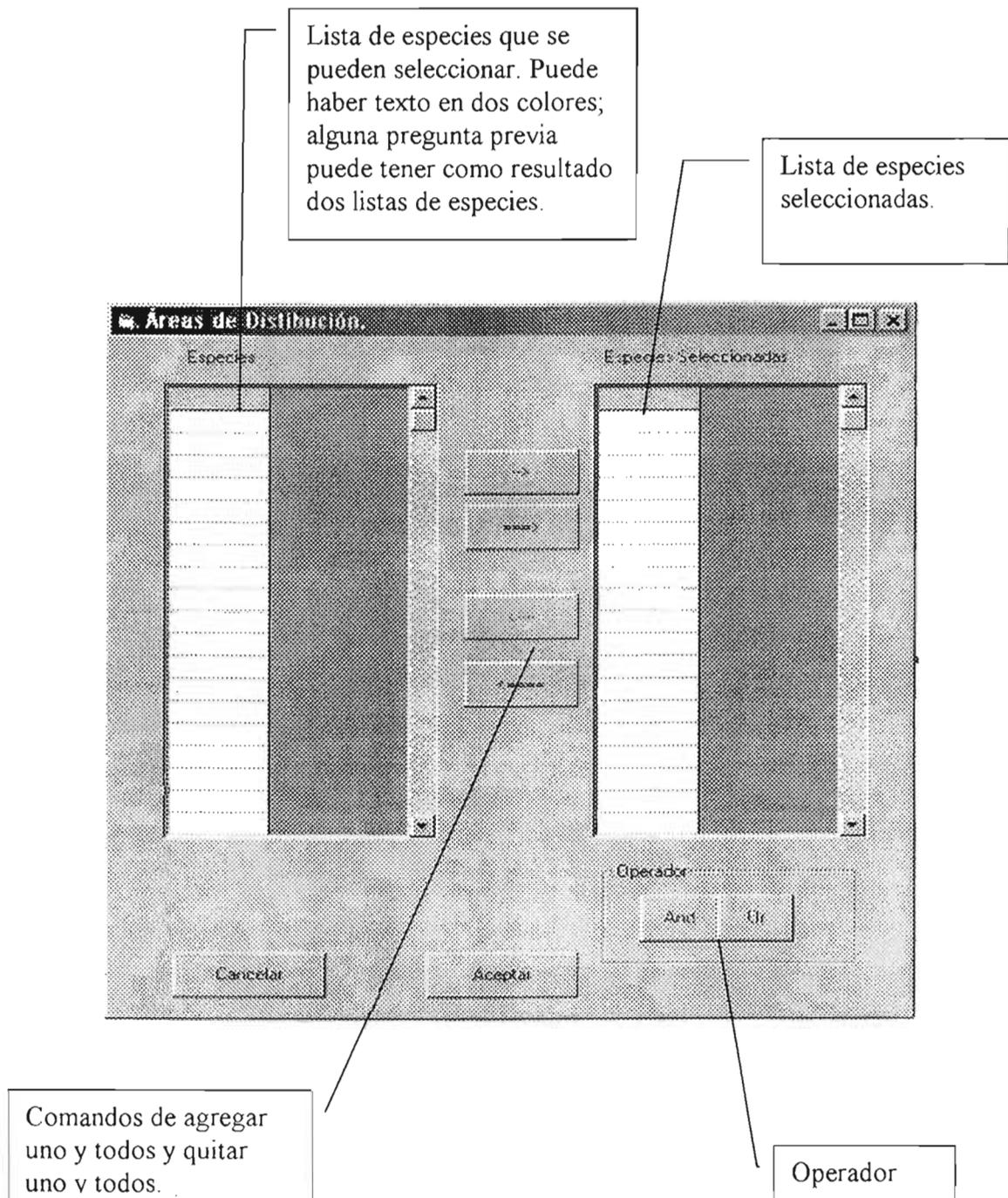
### Localidades

- Se permite este análisis aunque no se haya definido el geoframe, pero sí la rejilla.
- Se puede atrás el mapa alfa.
- Muestra un mapa de localidades.



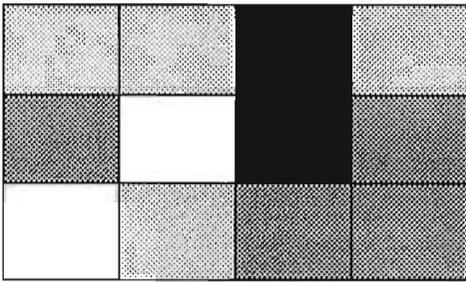
## Áreas de Distribución

- Debe de haber un geoframe abierto.
- Las especies son de toda el AE o de un análisis previo que haya dado como resultado especies.
- Muestra un mapa de UG's.



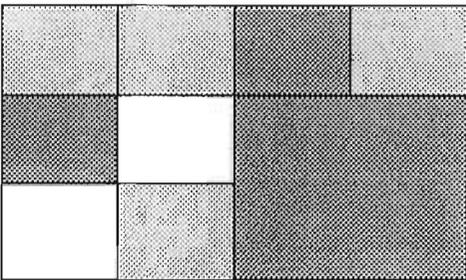
Ejemplo de mapa de salida:

En caso de haber seleccionado OR:



	100%	<i>Lista de Especies:</i>  <i>Especie 1</i> <i>Especie 2</i> <i>Especie 3</i> <i>Especie n</i>
	1-99%	
	0%	
	N/A	

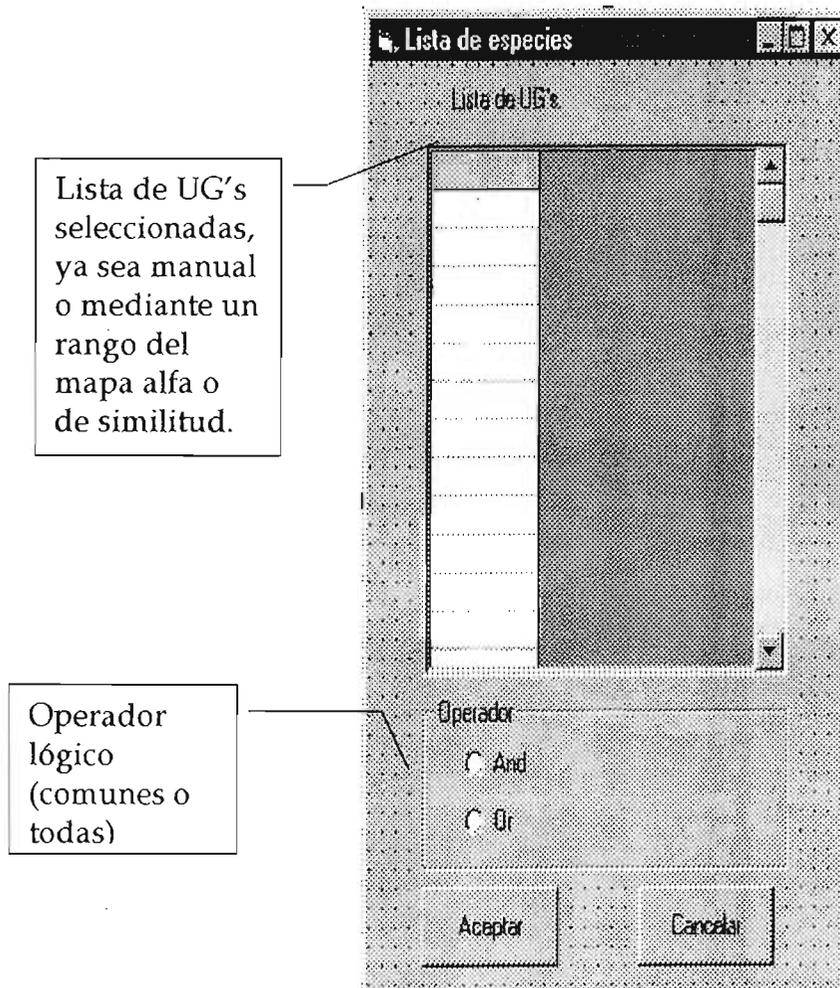
En caso de haber seleccionado AND.:



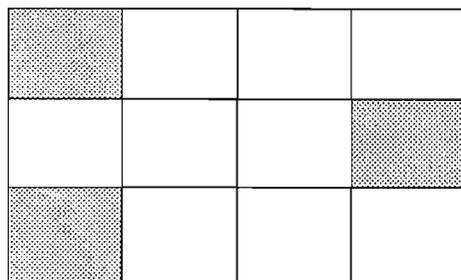
	<i>presentes</i>	<i>Lista de Especies:</i>  <i>Especie 1</i> <i>Especie 2</i>  <i>Especie 3</i> <i>Especie n</i>
	<i>ausentes</i>	
	N/A	

### Lista de Especies

- Se puede pedir una lista de especies de: a) un mapa por rangos (análisis previo) o b) una selección manual previa de UG's.
- Genera una lista de especies con un mapa resaltando las UG seleccionadas para el análisis.
- Con el menú de "ver" se puede "seleccionar toda el AE", y al pedir este análisis se despliega la lista de especies de toda el AE.



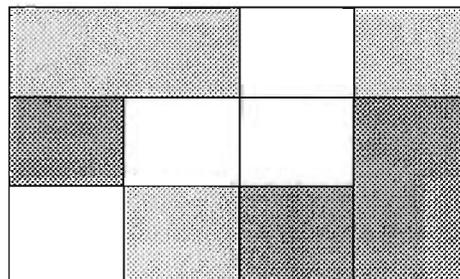
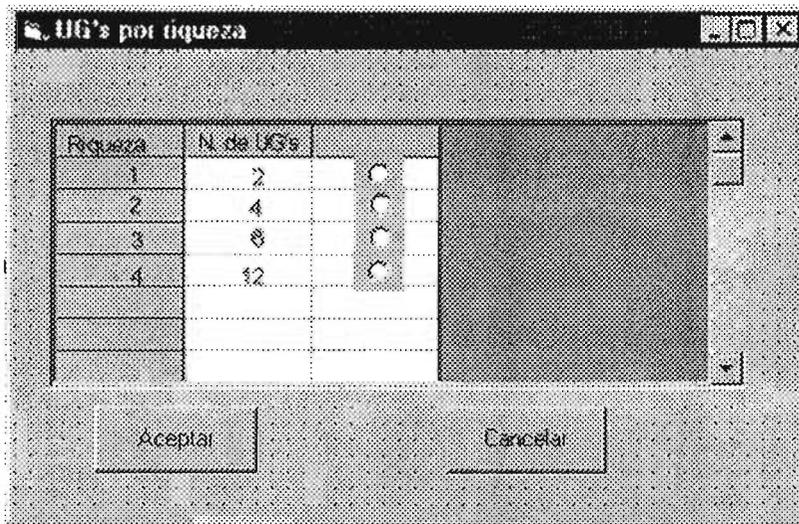
Ejemplo de mapa de salida:



<i>Lista de Especies:</i>
<i>Especie 1</i> <i>Especie 2</i>
<i>Especie 3</i> <i>Especie n</i>
<i>Operador : AND</i>
<i>UG seleccionada</i>

### UG's por Riqueza

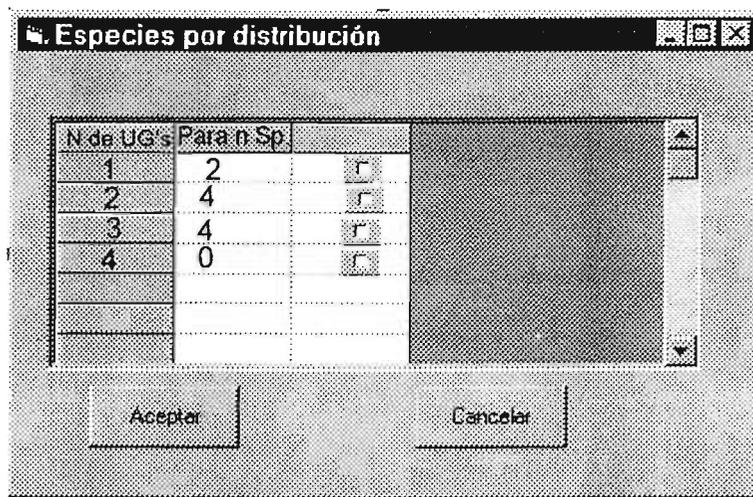
- Debe de haber un geoframe cargado o algunas UG seleccionadas.
- La salida es un mapa de UG's de 3 colores.
- Sólo se puede seleccionar un renglón.



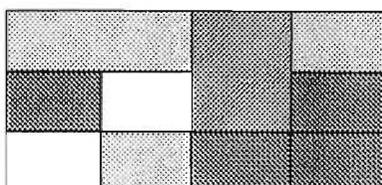
	<i>SÍ</i>
	<i>NO</i>
	<i>N/A</i>

### Especies por Distribución

- Debe de estar abierto un Geoframe.
- Genera un mapa y una lista de especies que están en n UG's.
- Especies de distribución restringida.



Ejemplo de Mapa:

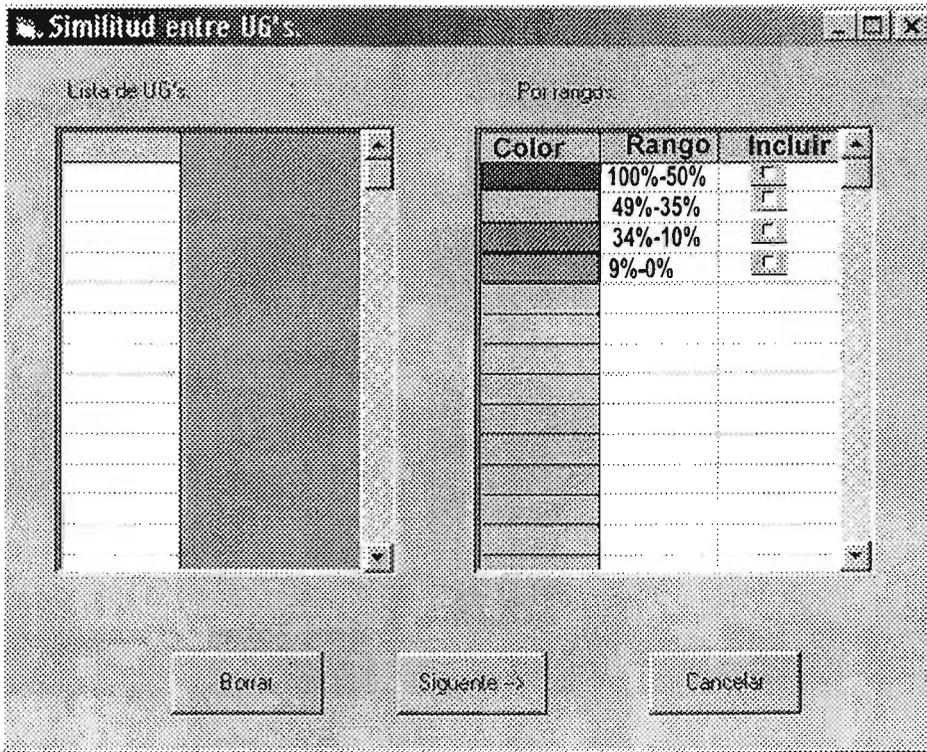


Leyenda:

	7 a 5	Lista de Especies Q <sub>2</sub> :
	4 a 1	Especie 1
	0	Especie 2
	N/A	Especie 3
		...
		Especie n

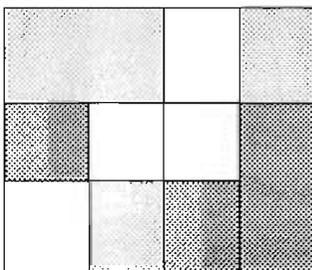
### Similitud entre UG's

- Debe de haber una selección de dos conjuntos de UG's; primero debe haber una selección previa y este cuadro permite la segunda selección.
- Se puede seleccionar por rangos dentro de la misma forma, por lo que debe de haber un mapa ya generado.
- Resultado: Numero real y una lista de especies comunes y totales (Unión e Intersección).



Similitud entre ambos conjuntos: 75%.  
Índice: Fager

Ejemplo de mapa



Leyenda

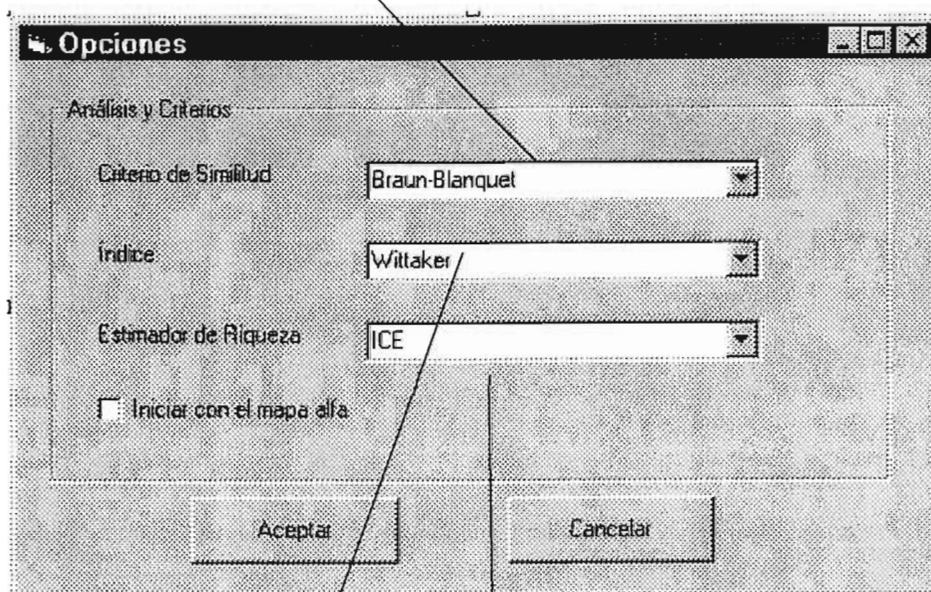
Intersección	Lista de especies Especie 1
Unión	Especie 2 Especie 3 ... Especie n

## Comandos del submenú *Herramientas*

### Opciones

- Permite seleccionar los índices a utilizar en los análisis.
- Debe de haber un geoframe cargado.

Braun-Blanquet
Fager
Jaccard
Kulezynski 1
Kulezynski 2
Otsuka
Simpson
Sorensen-Dice
Radio de correlación

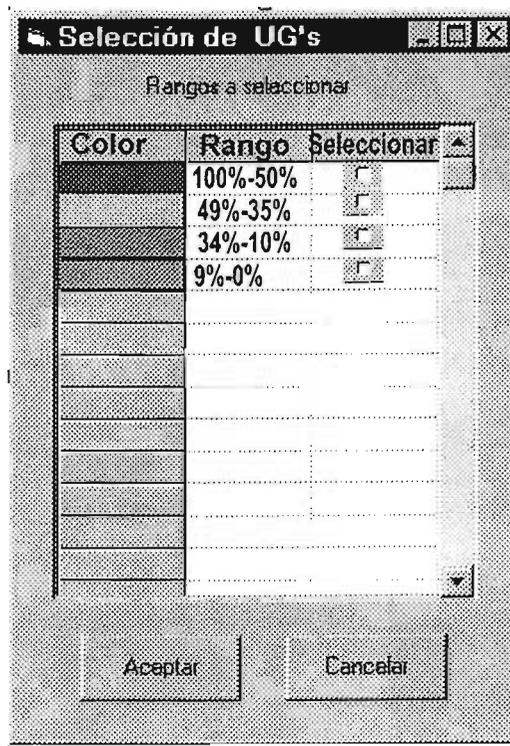


Whitaker
Regresión

ICE
Chao 2
Jack 1
Jack 2
Boot
MMMean

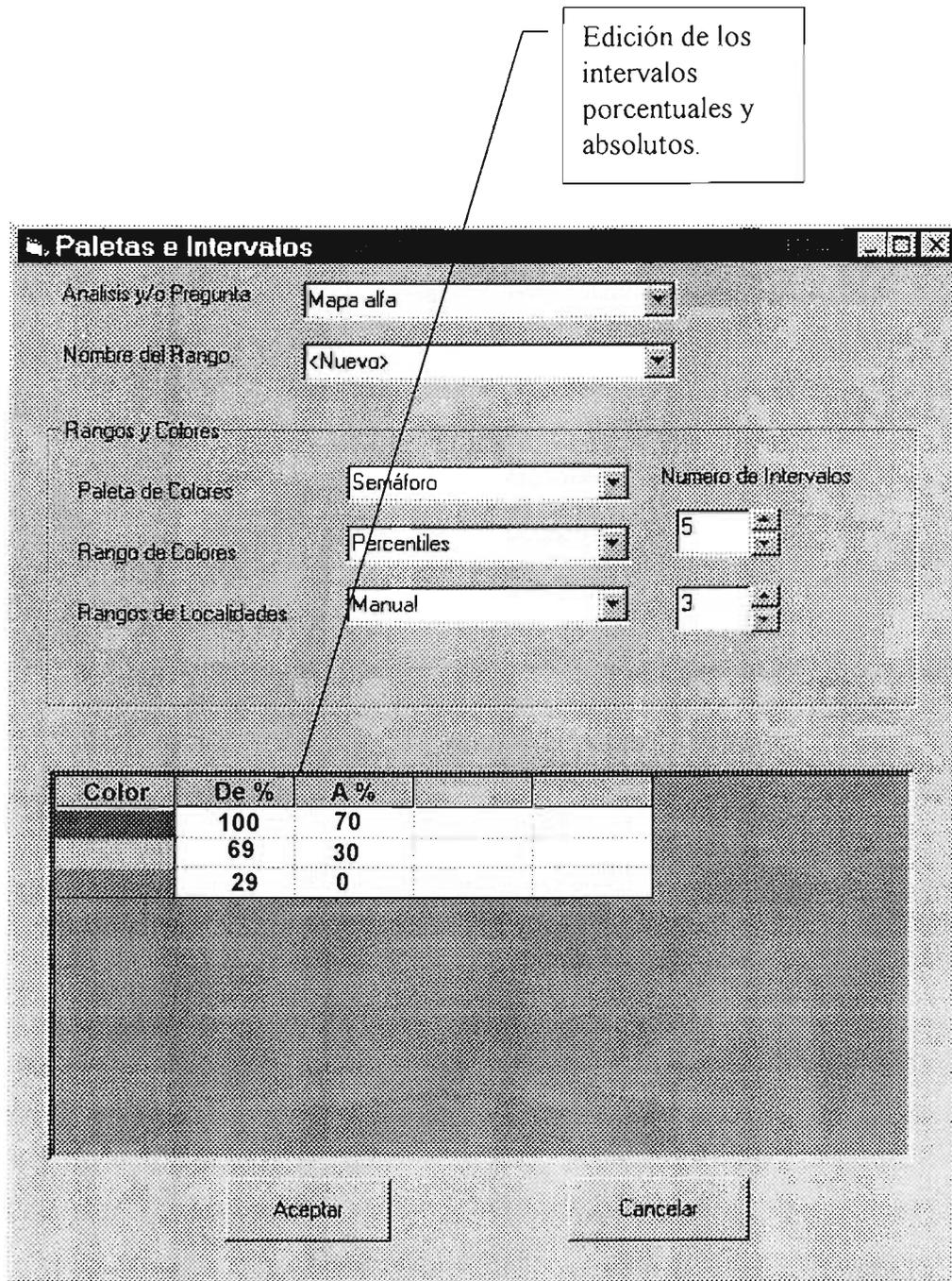
### Seleccionar UG por rangos.

- Debe de haber un mapa de rangos en el mapa de UG's.
- Marca las celdas seleccionadas del mapa.
- Puede haber celdas seleccionadas anteriormente y las respeta.



### Paletas y rangos

- Cada análisis tiene definido un rango con un nombre.
- Hay asignación automática opcional de nombres, en caso de los intervalos constantes y percentiles.



### Definir rejilla

- Debe de haber una base de datos abierta.
- Se puede ver el mapa alfa y los análisis de localidades.
- Se prevé la posibilidad de regresar al mapa a través del menú y el sistema debe de recordar los valores pasados.

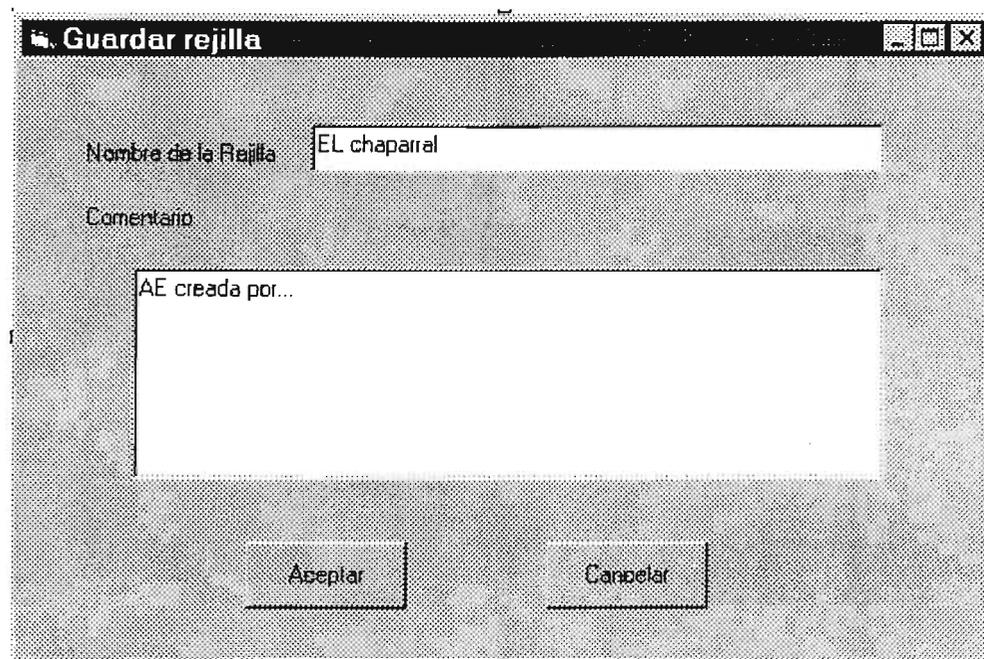
Definición del rectángulo que define el AE, con los puntos más al NE y SW.

Tipo de definición de cada UG.

Permite prever la rejilla, y que estén habilitados el mapa alfa y el de localidades.

Guardar con un nombre al Geoframe

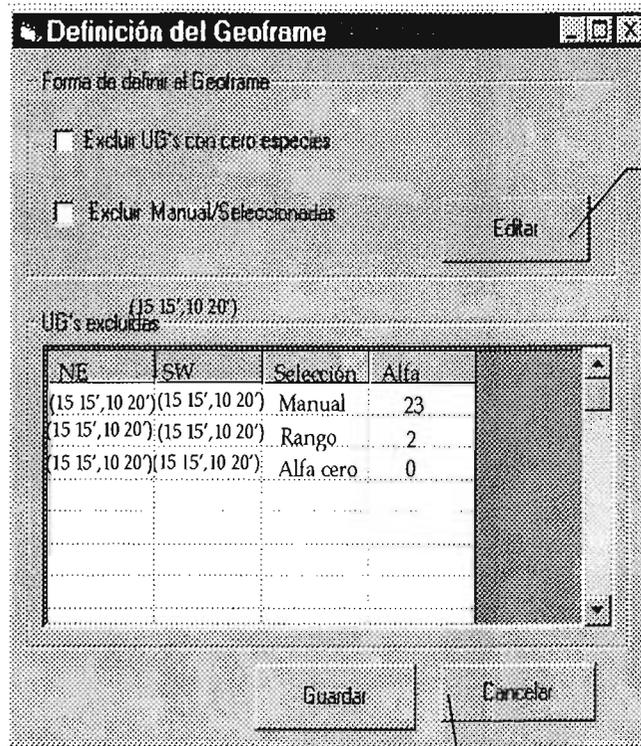
Interfaz para guardar rejilla (aparece también al seleccionar el comando “guardar” por primera vez durante una sesión).



The image shows a dialog box titled "Guardar rejilla". It has a standard Windows-style title bar with minimize, maximize, and close buttons. The main area contains a text input field labeled "Nombre de la Rejilla" containing the text "EL chaparral". Below it is a text area labeled "Comentario" containing the text "AE creada por...". At the bottom of the dialog, there are two buttons: "Aceptar" on the left and "Cancelar" on the right.

### Definir geoframe

- Debe de haber una rejilla cargada o un geoframe base.
  - Permite ver el mapa alfa.
  - Se puede minimizar para selección manual o por rangos.
    - El cursor cambia a la forma de selección.
    - Se selecciona con un click y se de-selecciona con otro.
- Las UG seleccionadas son excluidas y se resaltan

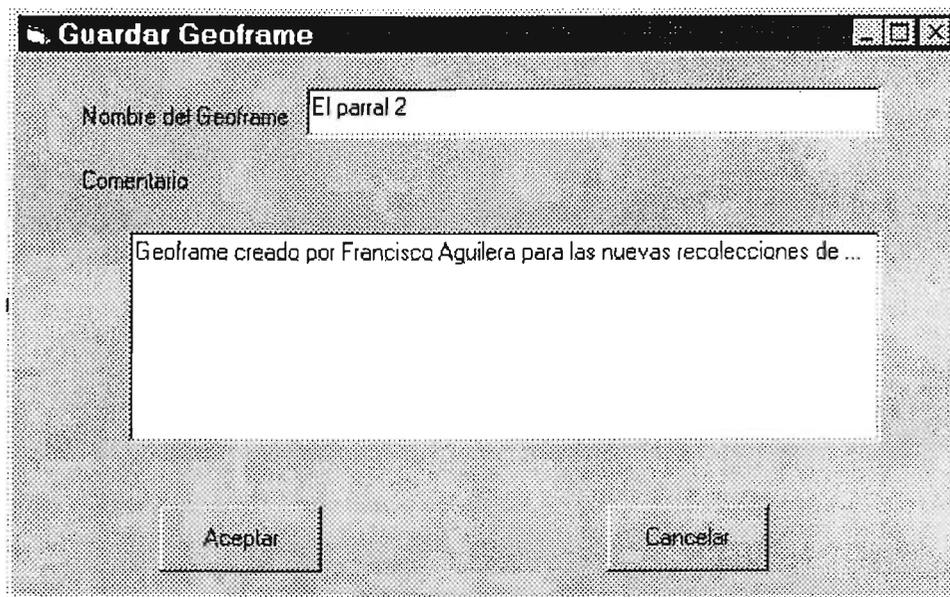


Permite editar el geoframe con selección de UG, ya sea manual o por rangos.

Despliega el diálogo "guardar geoframe".

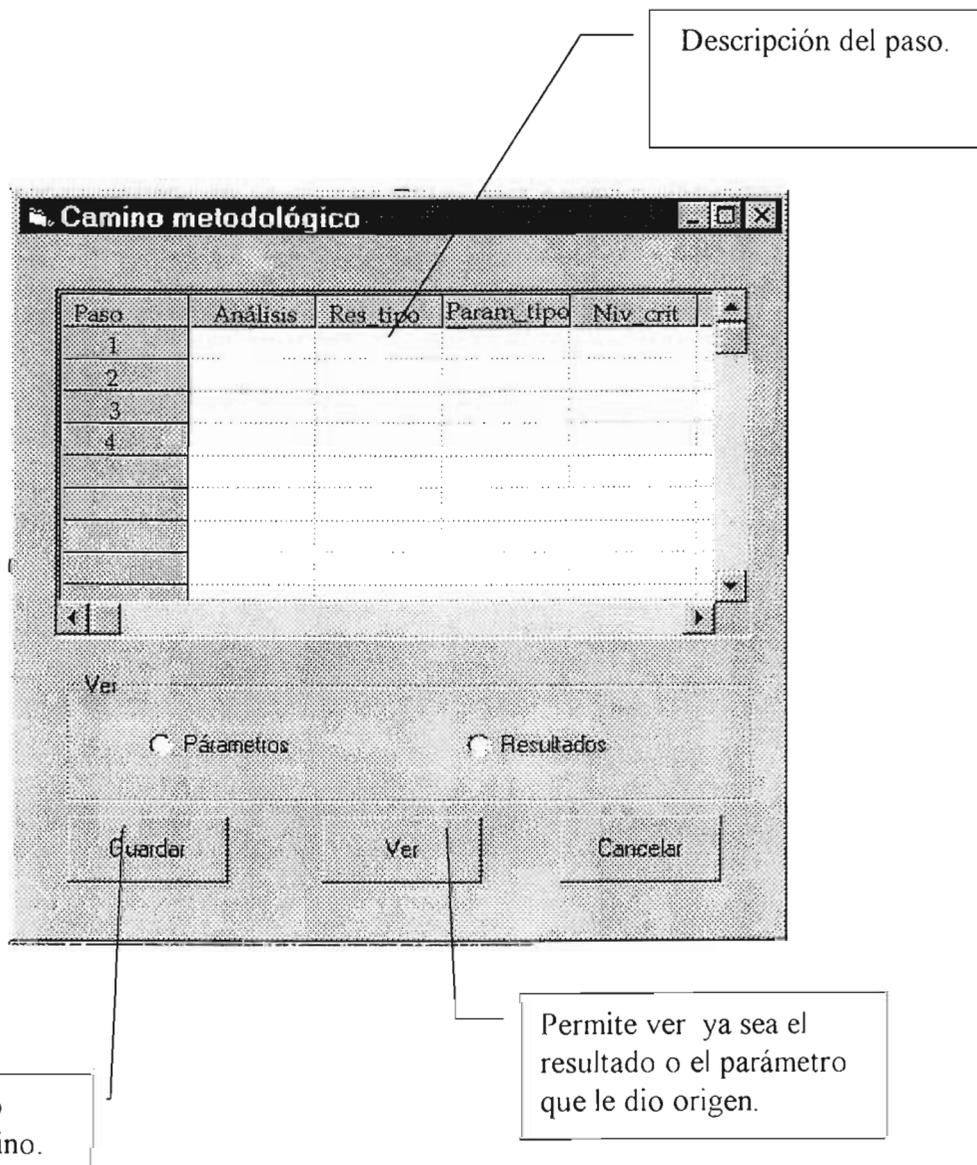
### Guardar Geoframe

- Interfaz para guardar Geoframe (aparece también si al editar el geoframe se selecciona el comando de “guardar” por primera vez.
- Se puede indicar varias veces “guardar” mientras se define el geoframe, pero en el momento en que se pida un análisis diferente al mapa alfa o de localidades, ese menú asume “guardar camino metodológico”).

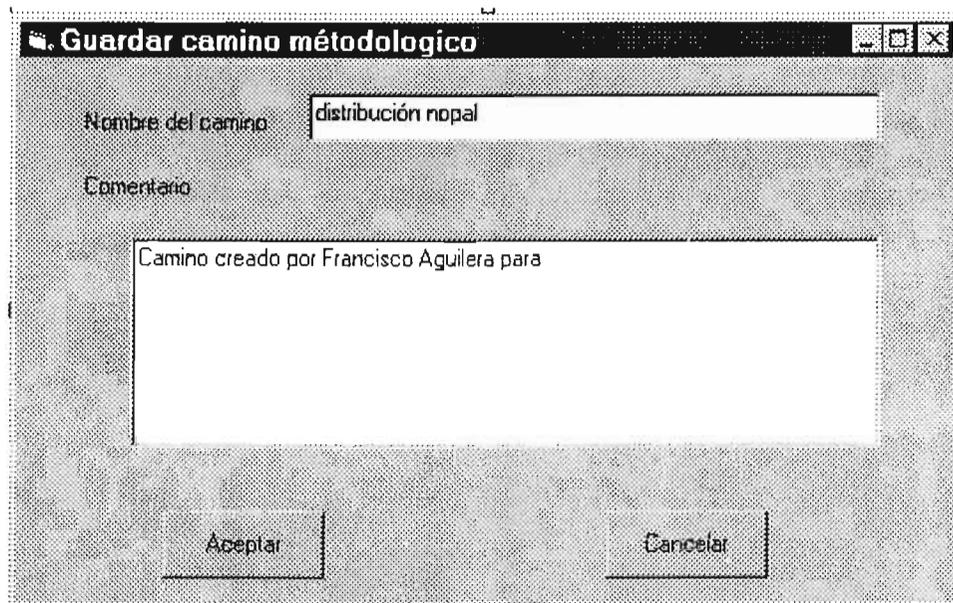


### Camino Metodológico

- Proporciona la descripción corta del camino:
  - los parámetros numéricos ingresados,
  - rango utilizado (nombre),
  - tipo resultado,
  - tipo parámetro.
- Permite ver cualquier paso del camino:
  - tanto parámetros como
  - resultados.
- En pasos anteriores no se permite ya la edición, no así en pasos siguientes.

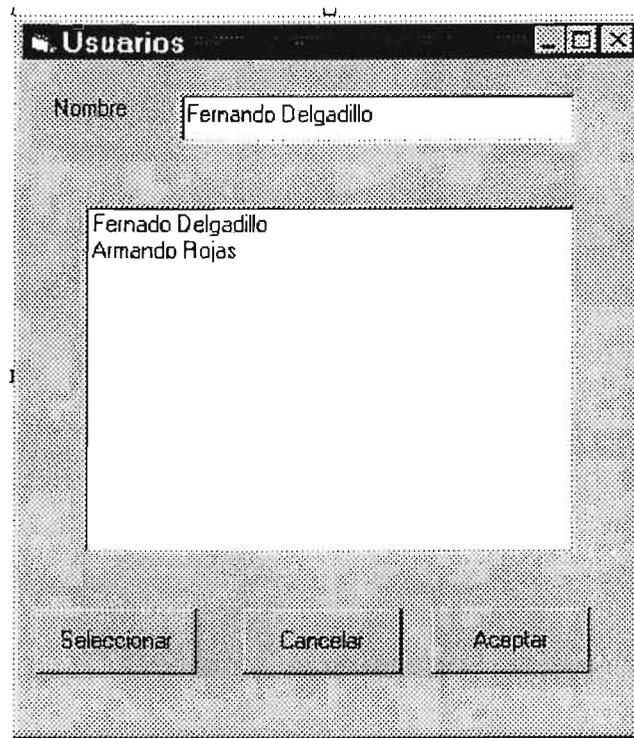


Interfaz de guardar camino (aparece por primera vez en el comando guardar)



## Usuarios

- Permite control opcional de las modificaciones y trabajos que se han efectuado



### 10.3. Diseño de la Base de Datos

El esquema de la base de datos es el principal resultado del diseño, sin embargo hay otros elementos importantes que se definieron en la etapa de diseño, entre ellos:

1. *Diseño abierto*, es decir, la documentación final debe contener el modelo completo, así como el diccionario de datos para que a los usuarios les sea posible modificar el diseño agregando campos e inclusive tablas.
2. *El diseño original está implementado en MS Access*, sin embargo se debe permitir la conexión a otras BD a través de ODBC (Open Data Base Connect) en versiones subsecuentes.
3. *Los procesos de importación* contemplados son
  - a. Importación de nombres de especies de otra BD del mismo programa.
  - b. Importación de datos de recolección por medio de un archivo de texto separado por comas (\*.csv) único.
4. *Se contempla la implementación de procesos para revisión de integridad* (opcional), que evite, en la medida de lo posible, información duplicada de datos de recolección (nombre de especies, localidades, etc.) por medio de algoritmos de comparación de cadenas.
5. *La interfaz de importación* debe permitir hacer el mapeo de campos desde el archivo de texto hacia los campos que contienen la información, permitiendo la concatenación de campos cuando sea necesario (v.gr. GENERO, ESPECIE hacia TAXON).

## ***Esquema de la Base de Datos***

La BD está dividida conceptualmente en 5 grandes grupos de tablas:

### **Los Datos de Recolecta**

Esta parte de la base de datos contiene los datos de recolecta (taxon, localidad, fecha, etc.).

Es la tabla base para poder generar la rejilla.

### **Área del Geoframe**

Esta parte de la BD contiene todas las tablas que conforman la rejilla y el geoframe:

1. Datos para su creación, por ejemplo, puntos más al SW y NE de el AE y de cada uno de las UGs y sus centroides.
2. Datos asociados a cada UG incluyendo a las especies que contiene.
3. Resultados de los análisis que se apliquen a todo el geoframe, para incrementar el desempeño del programa, por ejemplo, algunas bioestadísticas.
4. Matriz de similitud para cada criterio.
5. Su vigencia, pues para nuevas recolectas será necesaria la definición de una nueva rejilla (colectas fuera del AE actual) o geoframe (colectas dentro del AE actual), para mantener su consistencia con los caminos metodológicos.

### **Caminos Metodológicos**

Este grupo de la BD contendrá toda la información requerida para recuperar cualquier paso del encadenamiento de análisis que efectúe el usuario, permitiéndole continuar a partir del último análisis de la cadena. En este conjunto de tablas se almacena:

1. El nombre dado por el usuario y el orden de los análisis aplicados.
2. Los parámetros y resultados de los análisis involucrados.
3. Los árboles de similitud requeridos.
4. La información de los rangos utilizados.

### **Tablas de Rangos y Colores**

En esta parte de la BD estará contenida la información de los rangos y paletas de colores que se pueden usar dentro de la aplicación, así como almacenar los diseñados por el usuario:

1. La información de los rangos:
  - 1.1.El análisis en el que puede utilizarse.
  - 1.2.El tipo de rango (manual, percentiles o intervalos constantes).
  - 1.3.El número de intervalos.
  - 1.4.Los límites de cada uno de los rangos.
  
2. La información de las paletas de colores; el programa debe proveer de varias paletas de colores predeterminadas, permitiendo al usuario la selección de cualquiera de ellas, ahorrándole así tiempo.
  - 2.1.La definición del color (en formato RGB).
  - 2.2.El orden de los colores dentro del rango.

### **Tablas del Sistema**

Esta parte de la BD almacena información para la administración del sistema como es:

1. Tablas de las plantillas de importación, en las que se almacenan configuraciones sobre el mapeo de campos en la importación de datos hacia el sistema.
2. Información de los usuarios, que de manera opcional permitirán diversos niveles de acceso los datos.

Diagrama de Relaciones

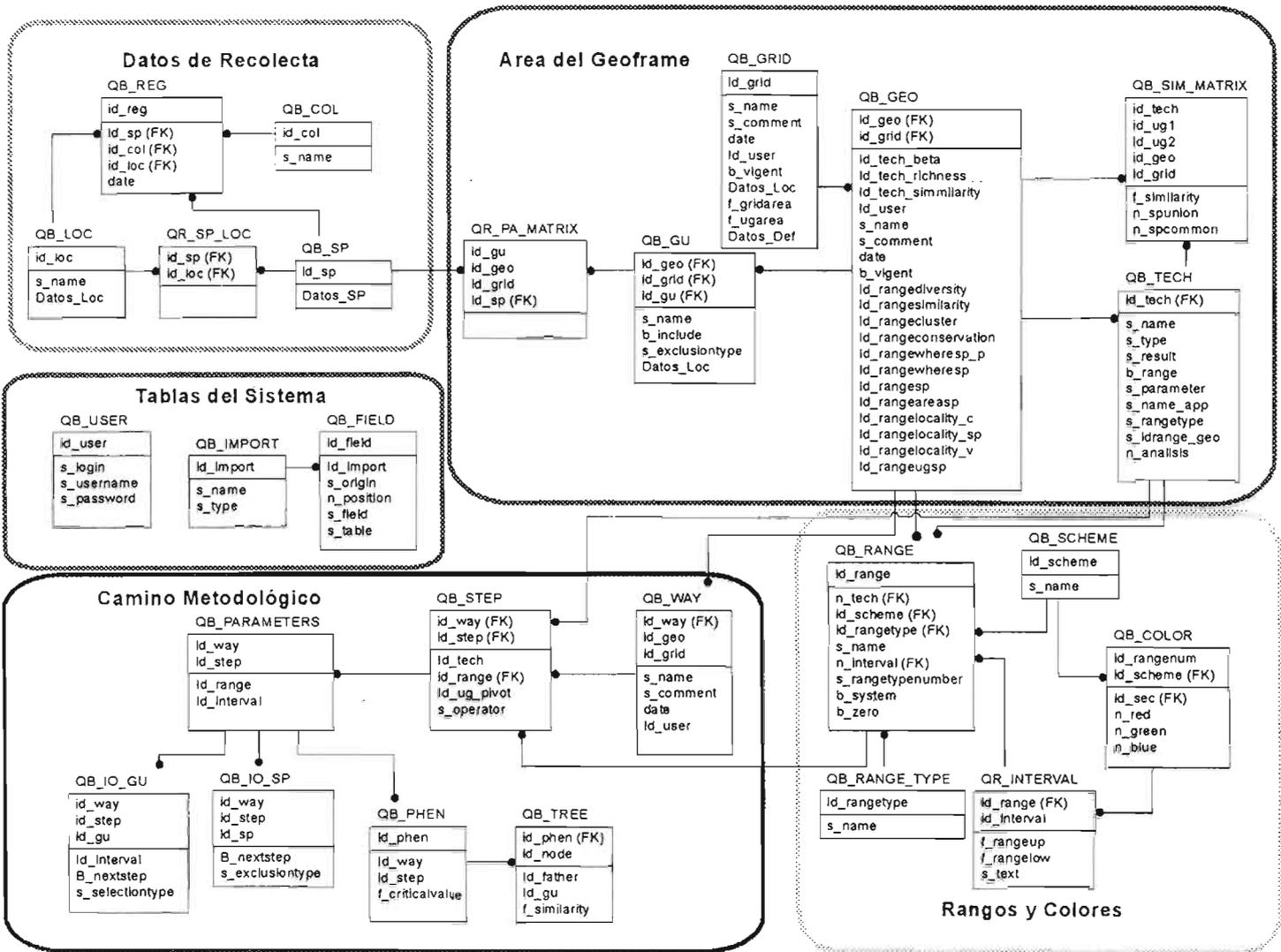


Figura 10.4 Diagrama de Relaciones.

## *Notación de la Base de Datos*

El esquema está regido por criterios de nomenclatura, entre ellos:

### 1. Nombres de tablas

1.1. Deben tener el prefijo QB\_ (v. gr. QB\_SP).

1.2. Abreviaciones en inglés.

1.3. En singular.

1.4. En mayúsculas.

1.5. Las tablas que sirvan de relación tendrán el prefijo QR\_, seguido por los nombres de las tablas que relaciona (v. gr. QR\_SP\_Loc).

### 2. Nombres de Campos

2.1. Los nombres de los campos deben ser en minúsculas.

2.2. Los campos llaves deben tener el prefijo id\_ y el nombre debe ser el mismo que el de la tabla en donde son llave primaria.

2.3. Los campos de cadena deben tener el prefijo de s\_.

2.4. Los campos numéricos deben tener el prefijo n\_ para los enteros y f\_ para los flotantes.

2.5. Los campos booleanos deben tener el prefijo b\_.

*Catálogo de la Base de de Datos*

TABLA	DESCRIPCIÓN
QB_CAT_FIGURE	Definición de las figuras.
QB_COL	Nombres de los colectores.
QB_COLOR	Definición de los colores y sus atributos.
QB_FIELD	Configuración de la importación de datos.
QB_FIGURE	Definición de los colores y sus atributos.
QB_GEO	Geoframe y sus bioestadísticas.
QB_GRID	Definición del grid.
QB_GU	Todas las GU's.
QB_IMPORT	Nombres de los formatos de importación.
QB_IO_GU	Entrada/salida de un paso del camino de GU.
QB_IO_SP	Entrada/Salida de un paso en el camino de SP.
QB_LOC	Datos de las localidades.
QB_PARAMETERS	Parámetros del paso (uno a muchos).
QB_PHEN	Fenogramas del camino.
QB_RANGE	Rangos.
QB_RANGE_TYPE	Tipos de rangos.
QB_REG	Registros de recolección.
QB_SCHEME_C	Esquemas de color.
QB_SCHEME_F	Esquemas de color.
QB_SIM_MATRIX	Matriz de similitud para cada criterio.
QB_SP	Nombres de las especies (y alguna categoría supraespecífica).
QB_STEP	Secuencia de pasos de un camino.
QB_TECH	Lista de los análisis y preguntas.
QB_TOWN	Lista de ciudades y pueblos (nombre y localización).
QB_TREE	Nodos del fenograma.
QB_USER	Nombres de usuarios de una instalación de QB-MAP.
QB_WAY	Caminos metodológicos.
QR_GEO_TOWN	Almacena las claves de las ciudades a desplegar en el geoframe.
QR_INTERVAL	Intervalos.
QR_PA_MATRIX	Matriz de Presencia/Ausencia, tabla de relación entre UG y SP.
QR_SP_LOC	Relación entre localidades y especies.

## *Diccionario de Datos de la Base de Datos*

TABLA/Campo	Tipo	Tam.	Descripción
<b>QB_CAT_FIGURE</b>			
o_figure	Objeto OLE	-	Mapa de bits.
s_dsc	Texto	20	Descripción de la figura.
id_figure	Número (largo)	4	Llave principal.
<b>QB_COL</b>			
id_col	Número (largo)	4	Llave principal.
s_name	Texto	100	Nombre del colector.
<b>QB_COLOR</b>			
n_red	Número (Byte)	1	Componente roja en RGB.
n_green	Número (Byte)	1	Componente verde en RGB.
id_sec	Número (largo)	4	Secuencia del color en el esquema.
id_rangenum	Número (largo)	4	Número de rangos que soporta
id_color	Número (largo)	4	Llave principal.
id_schemec	Número (largo)	4	Llave foránea de QB_SCHEMEC.
n_blue	Número (Byte)	1	Componente azul en RGB.
<b>QB_FIELD</b>			
s_table	Texto	50	Tabla destino.
id_field	Número (largo)	4	Llave primaria.
id_import	Número (largo)	4	Llave foránea de QB_IMPORT.
s_origin	Texto	50	Nombre del campo origen (1a. columna).
n_position	Número (largo)	4	Posición que ocupa el origen.
s_field	Texto	50	Campo destino.
<b>QB_FIGURE</b>			
id_figure	Número (largo)	4	Llave foránea en QB_CAT_FIGURE.
id_sec	Número (largo)	4	Secuencia de las figuras en el esquema.
id_schemec	Número (largo)	4	Llave foránea de QB_SCHEMEC.
id_rangenum	Número (largo)	4	Número de rangos que soporta.
<b>QB_GEO</b>			
f_lastzoomy1	Número (doble)	8	Estado de acercamiento en que se quedó el geoframe, Y1.
f_richness_chao	Número (doble)	8	Riqueza estimada Chao 2.
f_richness_jack1	Número (doble)	8	Riqueza estimada Jack1.
f_richness_jack2	Número (doble)	8	Riqueza estimada Jack2.
f_richness_boot	Número (doble)	8	Riqueza estimada Boot.
f_richness_mmmmean	Número (doble)	8	Riqueza estimada MMMean.
f_quality_sampling	Número (doble)	8	Calidad del muestreo.

f_lastzoomx1	Número (doble)	8	Estado de acercamiento en que se quedó el geoframe, X1.
f_richness_ice	Número (doble)	8	Riqueza estimada ICE.
f_lastzoomx2	Número (doble)	8	Estado de acercamiento en que se quedó el geoframe, X2.
f_lastzoomy2	Número (doble)	8	Estado de acercamiento en que se quedó el geoframe, Y2.
b_grid	Sí/No	1	Grid prendida.
b_rule	Sí/No	1	Regla prendida.
n_min_pop_town	Número (largo)	4	Tamaño mínimo de población de las ciudades a desplegar.
b_alphamap	Sí/No	1	Iniciar con el mapa alfa.
id_user	Número (largo)	4	Llave foránea de QB_USER para el nombre del usuario.
f_beta1	Número (doble)	8	Valor de beta (gama entre alfa promedio).
n_gamma	Número (largo)	4	Valor de gamma.
id_grid	Número (largo)	4	Llave foránea de QB_GRID.
id_tech_beta	Número (largo)	4	Llave foránea de ID_TECH para la beta seleccionada.
id_tech_similarity	Número (largo)	4	Llave foránea de ID_TECH para el criterio de similitud.
id_geo	Número (largo)	4	Llave principal.
s_name	Texto	50	Nombre del Geoframe.
s_comment	Memo	-	Comentario del geoframe.
date	Texto	10	Fecha de creación del geoframe.
b_vigent	Sí/No	1	Si el geoframe sigue siendo vigente no ha habido actualizaciones de recolección.
f_avg_alpha	Número (doble)	8	Valor de alfa promedio.
f_beta2	Número (doble)	8	Valor de beta regresiones.
id_tech_richness	Número (largo)	4	Llave foránea de ID_TECH para el criterio de estimación de riqueza.
<b>QB_GRID</b>			
f_deltax	Número (doble)	8	Delta X en caso de definición lineal.
n_selon_second	Número (Byte)	1	Longitud del punto más al suroeste, segundos.
n_selon_minute	Número (Byte)	1	Longitud del punto más al suroeste, minutos.
f_gridarea	Número (doble)	8	Area total.
f_ugarea	Número (doble)	8	Area de cada UG.
s_ugdefinition	Texto	50	Tipo de definición del grid (LINEAR para una definición en distancia y RADIAL para una definición en grados).

s_units	Texto	10	Tipo de unidades utilizadas en caso de definición lineal ("km" kilómetros, "m" metros, "milles" millas, otras).
f_deltay	Número (doble)	8	Delta Y en caso de definición lineal.
n_deltax_degree	Número (largo)	4	Delta X en caso de definición radial, grados.
n_deltax_minute	Número (largo)	4	Delta X en caso de definición radial, minutos.
n_deltax_second	Número (largo)	4	Delta X en caso de definición radial, segundos.
n_deltay_degree	Número (largo)	4	Delta Y en caso de definición radial, grados.
n_deltay_second	Número (largo)	4	Delta Y en caso de definición radial, segundos.
date	Texto	10	Fecha de creación.
n_selon_degree	Número (Byte)	1	Longoitud del punto más al suroeste, grados.
n_deltay_minute	Número (largo)	4	Delta Y en caso de definición radial, minutos.
s_comment	Memo	-	Comentario del Grid.
b_vigent	Texto	1	Si el Grid sigue vigente (caso en que las nuevas recolectas amplíen geográficamente el AE).
n_selat_second	Número (Byte)	1	Latitud del punto más al suroeste, segundos.
s_name	Texto	50	Nombre del Grid.
id_user	Número (largo)	4	Llave foránea de QB_USER para el nombre del usuario.
id_grid	Número (largo)	4	Llave principal.
n_quadrant	Número (Byte)	1	Cuadrante terrestre: lon,lat (1: +,+) (2:+,-) (3:-,+ ) (4:-,-)
n_nwlat_degree	Número (Byte)	1	Latitud del punto más al noreste, grados.
n_nwlat_second	Número (largo)	4	Latitud del punto más al noreste, segundos.
n_nwlon_degree	Número (Byte)	1	Longoitud del punto más al noreste, grados.
n_nwlon_minute	Número (Byte)	1	Longitud del punto más al noreste, minutos.
n_nwlon_second	Número (Byte)	1	Longitud del punto más al noreste, segundos.
n_nwlat_minute	Número (Byte)	1	Latitud del punto más al noreste, minutos.
n_selat_minute	Número (Byte)	1	Latitud del punto más al suroeste, minutos.
n_selat_degree	Número (Byte)	1	Latitud del punto más al suroeste, grados.
<b>QB_GU</b>			
n_loncenterpoint_second	Número (Byte)	1	Longitud del punto central, segundos.
n_selat_minute	Número (Byte)	1	Latitud del punto más al sureste, minutos.
n_selat_second	Número (Byte)	1	Latitud del punto más al sureste, segundos.
n_selon_degree	Número (Byte)	1	Longoitud del punto más al sureste, grados.

n_selon_minute	Número (Byte)	1	Longitud del punto más al sureste, minutos.
n_selon_second	Número (Byte)	1	Longitud del punto más al sureste, segundos.
n_latcenterpoint_minute	Número (Byte)	1	Latitud del punto central, minutos.
n_loncenterpoint_minute	Número (Byte)	1	Longitud del punto central, minutos.
n_latcenterpoint_degree	Número (Byte)	1	Latitud del punto central, grados.
n_selat_degree	Número (Byte)	1	Latitud del punto más al sureste, grados.
b_include	Texto	1	Incluido en el Geoframe.
n_latcenterpoint_second	Número (Byte)	1	Latitud del punto central, segundos.
n_loncenterpoint_degree	Número (Byte)	1	Longitud del punto central, grados.
n_nwlon_second	Número (Byte)	1	Longitud del punto más al noreste, segundos.
n_nwlon_minute	Número (Byte)	1	Longitud del punto más al noreste, minutos.
n_nwlon_degree	Número (Byte)	1	Longitud del punto más al noreste, grados.
n_nwlat_second	Número (Byte)	1	Latitud del punto más al noreste, segundos.
n_nwlat_minute	Número (Byte)	1	Latitud del punto más al noreste, minutos.
n_nwlat_degree	Número (Byte)	1	Latitud del punto más al noreste, grados.
s_exclusionstype	Texto	10	Tipo de exclusión de la UG (ZERO por alfa cero, MANUAL por selección manual)
n_alpha	Número (largo)	4	Cantidad de especies que contiene.
s_name	Texto	50	Nombre de la GU.
id_gu	Número (largo)	4	Llave principal.
id_grid	Número (largo)	4	Llave foránea de QB_GRID.
id_geo	Número (largo)	4	Llave foránea de QB_GEO.
n_quadrant	Número (Byte)	1	Cuadrante terrestre: lon,lat (1: +,+) (2:+,-) (3:-,+ ) (4:-,-)
<b>QB_IMPORT</b>			
s_type	Texto	10	RECORDS o SPECIES.
s_name	Texto	50	Nombre de la importación.
id_import	Número (largo)	4	Llave primaria.
<b>QB_IO_GU</b>			
id_way	Número (largo)	4	Llave foránea de QB_WAY.
id_step	Número (largo)	4	Llave del paso.
id_gu	Número (largo)	4	Llave foránea de QB_GU.
id_interval	Número (largo)	4	Llave foránea de QR_INTERVAL que indica el intervalo en la que cayó la UG.
b_nextstep	Texto	1	Indica que en caso de ser resultado sí pasó como parámetro.

s_selectiontype	Texto	10	Tipo de selección de la UG, en caso de ser parámetro (MANUAL, ALL para todos, RANGE para caso de un rango).
<b>QB_IO_SP</b>			
b_nextstep	Texto	1	Indica si la especie paso como parámetro del siguiente análisis.
id_way	Número (largo)	4	Llave foránea de QB_WAY.
id_step	Número (largo)	4	Llave del paso.
s_exclusiontype	Texto	10	Tipo de Exclusión (MANUAL, COMMON, UNION).
id_sp	Número (largo)	4	Llave foránea de QB_SP.
<b>QB_LOC</b>			
n_lat_second	Número (Byte)	1	Latitud de la localidad, segundos.
n_lon_second	Número (Byte)	1	Longitud de la localidad, segundos.
id_loc	Número (largo)	4	Llave principal.
n_lon_degree	Número (Byte)	1	Longitud de la localidad, grados.
s_name	Texto	255	Nombre de la Localidad.
n_quadrant	Número (Byte)	1	Cuadrante terrestre: lon,lat (1: +,+) (2:+,-) (3:-,+ ) (4:-,-)
n_lat_degree	Número (Byte)	1	Latitud de la localidad, grados.
n_lon_minute	Número (Byte)	1	Longitud de la localidad, minutos.
n_lat_minute	Número (Byte)	1	Latitud de la localidad, minutos.
<b>QB_PARAMETERS</b>			
id_range	Número (largo)	4	Llave foránea de QB_RANGE.
id_step	Número (largo)	4	Llave foránea de QB_STEP.
id_interval	Número (largo)	4	Llave foránea de QR_INTERVAL que indica los intervalos excluidos.
id_way	Número (largo)	4	Llave foránea de QB_WAY.
<b>QB_PHEN</b>			
id_phen	Número (largo)	4	Llave principal.
f_criticalvalue	Número (doble)	8	Valor crítico para conformar grupos.
id_step	Número (largo)	4	Llave foránea de QB_STEP.
id_way	Número (largo)	4	Llave foránea de QB_WAY.
<b>QB_RANGE</b>			
n_intervalcolor	Número (largo)	4	Número de intervalos en color.
s_name	Texto	50	Nombre del rango.
n_intervalfigure	Número (largo)	4	Número de intervalos en figuras.
id_range	Número (largo)	4	Llave principal.
id_rangetype	Número (largo)	4	Llave foránea de QB_RANGE_TYPE.
id_schemec	Número (largo)	4	Llave foránea de QB_SCHEME_C.
id_tech	Número (largo)	4	Llave foránea de QB_TECH.

<b>QB_RANGE_TYPE</b>			
id_rangetype	Número (largo)	4	Llave Primaria.
s_name	Texto	20	Nombre del tipo de rango (MANUAL, PERCENTILES, EQUALI, ...).
<b>QB_REG</b>			
id_loc	Número (largo)	4	Llave foránea de QB_LOC.
date	Texto	10	Fecha de la recolección.
id_col	Número (largo)	4	Llave foránea de QB_COL.
id_sp	Número (largo)	4	Llave foránea de QB_SP
id_reg	Número (largo)	4	Llave principal.
<b>QB_SCHEME_C</b>			
s_name	Texto	50	Nombre del esquema.
id_schemec	Número (largo)	4	Llave principal.
<b>QB_SCHEME_F</b>			
id_schemec	Número (largo)	4	Llave principal.
s_name	Texto	50	Nombre del esquema.
<b>QB_SIM_MATRIX</b>			
n_sunion	Número (largo)	4	Número de especies en total de las 2 UG's.
f_similarity	Número (doble)	8	Similitud entre los 2 UG's.
id_geo	Número (largo)	4	Llave foránea de QB_GEO.
id_ug2	Número (largo)	4	LLave foránea de QB_GU.
id_ug1	Número (largo)	4	Llave foránea de QB_GU.
n_spcommon	Número (largo)	4	Número de especies en común de las 2 UG's.
id_tech	Número (largo)	4	Llave foránea de QB_TECH.
<b>QB_SP</b>			
id_sp	Número (largo)	4	Llave principal.
s_genus	Texto	40	Nombre del género.
s_sp	Texto	40	Nombre de la especie.
s_author	Texto	50	Nombre del autor
s_supra	Texto	30	Nombre de categoría taxonómica supraespecífica.
<b>QB_STEP</b>			
id_range	Número (largo)	4	Llave foránea de QB_RANGE.
id_way	Número (largo)	4	Llave foránea de QB_WAY.
id_tech	Número (largo)	4	Llave foránea de QB_TECH.
id_ug_pivot	Número (largo)	4	Llave foránea de QB_UG indicando cuál fue la UG Pivote en caso de un análisis de similitud.
s_operator	Texto	5	Tipo de operador (OR, AND).
id_step	Número (largo)	4	Llave principal.

<b>QB_TECH</b>			
s_parameter	Texto	10	Tipo de parámetro (UG, SP, INTEGER, FLOAT).
id_tech	Número (largo)	4	Llave principal.
s_name	Texto	50	Nombre del análisis.
s_type	Texto	20	Nombre del grupo al que pertenece (DIVERSITY, RICHNESS, SIMMILARYTY, QUALITY, CLUSTER, CONSERVATION, LOCALITY, WHERE, SP, AREA SP, UG, SP, BETWEEN UG)
s_result	Texto	10	Tipo de resultado (UG, SP, INTEGER, FLOAT).
<b>QB_TOWN</b>			
n_lat_degree	Número (Byte)	1	Latitud grados.
n_population	Número (largo)	4	Tamaño de la población (número de habitantes).
id_town	Número (largo)	4	Llave principal.
n_lat_second	Número (Byte)	1	Latitud segundos.
s_name	Texto	50	Nombre de la ciudad.
n_lat_minute	Número (Byte)	1	Latitud minutos.
n_lon_degree	Número (Byte)	1	Longitud grados.
n_lon_minute	Número (Byte)	1	Longitud minutos.
n_lon_second	Número (Byte)	1	Longitud segundos.
<b>QB_TREE</b>			
id_phen	Número (largo)	4	Llave foránea de QB_PHEN.
id_node	Número (largo)	4	Llave del nodo.
id_father	Número (largo)	4	Llave del nodo padre.
id_gu	Número (largo)	4	Llave foránea de QB_GU.
f_similarity	Número (doble)	8	Grado de similaridad.
<b>QB_USER</b>			
id_user	Número (largo)	4	Llave primaria.
s_login	Texto	15	Login.
s_username	Texto	50	Nombre del usuario.
s_password	Texto	10	Password encriptado.
<b>QB_WAY</b>			
id_geo	Número (largo)	4	Llave foránea de QB_GEO.
date	Texto	10	Fecha de creación (día/mes/año).
s_comment	Memo	-	Comentario del camino.
id_user	Número (largo)	4	Llave foránea de QB_USER para el nombre del usuario.
id_way	Número (largo)	4	Llave principal.
s_name	Texto	50	Nombre del camino.

id_grid	Número (largo)	4	Llave foránea de QB_GRID.
<b>QR_GEO_TOWN</b>			
id_geo	Número (largo)	4	Llave en QB_GEO.
id_town	Número (largo)	4	Llave en QB_TOWN.
<b>QR_INTERVAL</b>			
f_rangeup	Número (doble)	8	Límite superior (si es nulo indica el rango los menores que f_rangedown).
s_text	Texto	20	Texto a desplegar (v.g. "50 a 100", "prioritaria")
f_rangelow	Número (doble)	8	Límite inferior (si es nulo indica el rango los mayores que f_rangeup).
s_rangetype	Texto	10	Indica el tipo de rango (PERCENT, ABSOLUTE).
id_interval	Número (largo)	4	Llave del intervalo.
id_range	Número (largo)	4	Llave foránea de QB_RANGE.
n_secuence	Número (entero)	2	Secuencia del rango.
<b>QR_PA_MATRIX</b>			
id_gu	Número (largo)	4	Llave foránea de QB_GU.
id_geo	Número (largo)	4	Llave foránea de QB_GEO.
id_grid	Número (largo)	4	Llave principal.
id_sp	Número (largo)	4	Llave foránea de QB_SP.
<b>QR_SP_LOC</b>			
id_loc	Número (largo)	4	Llave foránea de QB_LOC.
id_sp	Número (largo)	4	Llave foránea de QB_SP.

## 10.4 Modelo de Objetos

El modelo de objetos resultante considera los siguientes aspectos:

1. Definición para cada clase de las funciones públicas, sus argumentos y su tipo
2. Cómo debe de construirse, copiarse y destruirse un objeto de la clase.
3. Definición de un conjunto de operaciones que requiere el concepto representado por la clase.
4. Identificación de las funciones que son comunes a todas las clases.

En la figura 10.5 se muestra la organización de las clases del diseño de objetos.

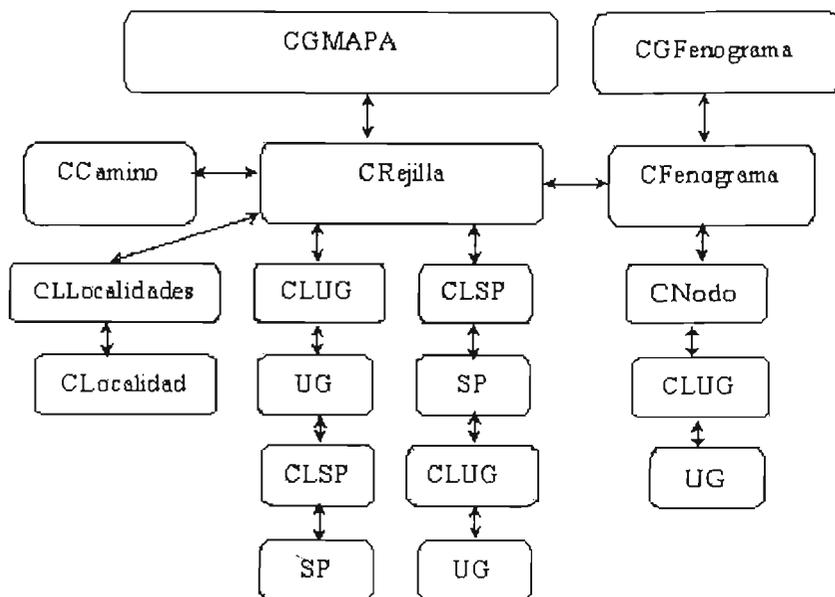


Figura 10.5 Principales clases y sus relaciones fundamentales del modelo de objetos

### Clases

El diseño del sistema consiste de varias clases, que se puede agrupar en tres tipos:

Clases gráficas

Clases para administrar al sistema

Clases para los análisis biogeográficos

A continuación se listan y explican brevemente cada una de las clases de los análisis biogeográficos:

- CUG: Clase que representa una UG.
- CSP: Clase que representa una especie.
- CLocalidad: Clase que representa una localidad de recolección.
- CLLocalidades: Clase lista de localidades.

- CLUG: Clase lista de UGs.
- CLSP: Clase lista de especies
- CNode: Clase que representa un nodo del fenograma.
- CFenograma: Clase que se encarga de la construcción y las operaciones propias del fenograma.
- CRejilla: Clase que contiene una serie de listas tanto de UGs como de especies, operando con ellas para poder realizar los análisis requeridos, es en realidad una serie de clases:
  - CRejillaAbre: Clase encargada con todas las operaciones de lectura y despliegue de la rejilla, geoframe y camino metodológico.
  - CRejillaSimilitud: Clase encargada de la creación y manipulación de la matriz de similitud, así como del mapa de similitud.
  - CRejillaAlfa: Clase encargada de la creación del mapa alfa.
  - CRejillaRiqueza: Clase encargada de todos los cálculos de estimación de la riqueza.
  - CCamino: Clase encargada del control del camino metodológico.

## 10.5 Implementación

El diseño resultante del sistema se implementó hasta un 70% de funcionalidad, según un plan previamente establecido de las “etapas”. El plan de implementación es el resultado de una evaluación de los módulos que considera la utilidad para el usuario, la interrelación con otros módulos y la complejidad o costo de su implementación.

Otros aspectos de la implementación, además del orden y definición de las etapas, es la elección de las herramientas de programación. En este caso se decidió, después de una evaluación, utilizar Visual Basic de Microsoft como lenguaje de programación así como un conjunto de librerías públicas de OPEN-GL, que facilitan el manejo gráfico. Se comenzó una implementación en Visual C++ (también de Microsoft), sin embargo se abandonó y decidió cambiar a Visual Basic debido a 1) definitivamente los tiempos de desarrollo en Visual C++ son

más largos, pues existe una mayor complejidad, por ejemplo, en la sola construcción de diálogos y 2) la poca disponibilidad de programadores en ese lenguaje hace elevar el costo y riesgo del proyecto. Algunas de las consecuencias negativas de abandonar a Visual C++ son:

1. No contar con una base sólida para la programación orientada a objetos (POO) (pues Visual Basic definitivamente implementa menos aspectos de la POO)
2. Se sacrifica desempeño, por ejemplo en rapidez.
3. Se tiene menor flexibilidad tanto en el diseño de la interfaz como en la misma implementación de las clases.

# 11. Discusión

---

*En este capítulo se discuten los resultados obtenidos en este trabajo sobre el análisis de los métodos de la BC y el desarrollo de la herramientas informática para la BC. Se discute la validez epistemológica del concepto de OGU.*

## 11.1 Análisis de Métodos de la Biogeografía Cuantitativa

### **Métodos descriptivos de la Biogeografía Cuantitativa (Calidad de los registros biogeográficos)**

En este trabajo se ha propuesto un índice (Qs) para estimar la calidad de los registros utilizados en biogeografía cuantitativa; para su cálculo, este índice considera básicamente aspectos ‘estructurales’ de la matriz de presencia-ausencia (especies x OGUs) y de ahí mismo se derivan algunas ventajas y desventajas. Entre las ventajas se puede mencionar: a) la información necesaria para calcularlo normalmente está disponible ya en forma estructurada (la matriz de presencia-ausencia), y b) relaciona directamente al problema de la calidad de la información con el problema del tamaño de la OGUs mediante la aproximación de que al aumentar la resolución disminuye la precisión, ayudando a elegir el tamaño de las OGUs. Entre las desventajas o debilidades se puede mencionar el hecho de que el índice Qs basa su estimación en *qué es lo que contiene la base de datos* y no *cómo se obtuvo esa información*, además no considera atributos biológicos

propiamente; eso es, en parte, una consecuencia de que incorpora a los índices no paramétricos para estimación de la riqueza.

La manera de abordar al problema de la calidad de la información tampoco considera el problema de la escala, que es un aspecto inseparable de cualquier estudio espacial. Debe continuar la investigación para el establecimiento de estadísticas que aproximen al parámetro estimando con precisión adecuada, de manera particular la estimación del error en los mismos datos espaciales. Los estimadores de riqueza son una propuesta de estimación del error de la riqueza observada en una región particular, pero no dicen nada acerca de la distribución espacial de la riqueza. El índice  $Q_s$  es una manera de ‘extrapolar’ al ámbito espacial la estimación de error de la riqueza; la invitación es a proponer mejores maneras de conocer la calidad de los registros biogeográficos, por ejemplo, proponiendo índices que estimen con menores márgenes de error a la calidad de los datos o bien, que incorporen otros aspectos (índices más integrales) relacionados con la temática del ámbito espacial, como lo es el concepto de la escala.

## **Epistemología del concepto OGU**

Llorente y Michán (2000) argumentan la validez del concepto de especie como unidad de análisis de la biodiversidad; se inclinan por una concepción ‘realista’ y no sólo nominalista de la especie. Para la Panbiogeografía la localidad de recolecta es una unidad válida para fundamentar el análisis biogeográfico. Los conceptos sobre diversidad alfa, beta, gama, calidad de muestreo y área potencial de distribución y OGU, entre otros manejados por la biogeografía cuantitativa, deben revisarse a la luz de las preconcepciones espacio-temporales-biológicas. Es decir, la vida se da en el tiempo y en el espacio; el tiempo y el espacio, a su vez, son presuposiciones que necesitan ser revisadas en conjunto con los conceptos de una teoría biogeográfica.

La dimensión espacial en relación con la vida, puede acotarse mediante el estudio de la distribución de las especies en la superficie de la Tierra, es decir, tomar como unidad biológica a la especie. Así como la panbiogeografía toma como unidad espacial a la localidad de recolecta, la biogeografía cuantitativa toma como unidad espacial a la OGU. En este trabajo se ha mostrado una manera de aproximarse al estudio de la validez de las OGU como unidades espaciales, combinando la dimensión biológica y la espacial. La biogeografía aún no ha consolidado una

unidad espacial ampliamente aceptada (como sí lo es el concepto de especie en taxonomía biológica) para el estudio espacial de la vida en la Tierra. Sin embargo, considero que existen algunas restricciones que debe de cumplir tal unidad supuesta, algunas de ellas incorporadas en el estudio sobre 'la calidad de los registros biogeográficos' en este trabajo. De hecho el diseño de construcción de OGU's en biogeografía cuantitativa es una búsqueda continua de una unidad espacial válida. Tales restricciones son:

- Considerar la escala espacial (resolución).
- Considerar la precisión geográfica.
- Considerar el error generado al hacer discretas a las variables continuas involucradas.
- Considerar los fenómenos geológicos relevantes al nivel de la escala espacial elegida.
- Considerar móvil o fija la rejilla

La OGU tiene una analogía epistemológica al concepto de especie, mientras que la población tiene una analogía epistemológica al de localidad de recolecta. La primera mediante lo que Leibniz llama verdad de razón y la segunda, verdad de hecho.

Las verdades de razón constituyen una manera práctica de acercarse al mundo, ya que las verdades de hecho pueden ser infinitas. Esa es una perspectiva de porqué la especie es un concepto valioso en biología.

La especie es una entidad con un nivel de abstracción superior al de población, así también la OGU tiene un nivel de abstracción arriba al de localidad recolecta. Por esa razón, es importante considerar a la OGU como un concepto válido para el estudio de la dimensión espacial de la vida. Así como la teoría de la evolución provee un marco conceptual para el concepto de especie, la biogeografía debe proveer ese marco para la unidad espacial. En este trabajo se ha avanzado en ese sentido proponiendo un método (el índice Qs) para la medición de los atributos de la OGU.

## **Métodos analíticos de la BC**

Además de las propiedades de las herramientas descriptivas, también deben de investigarse mejor aquellas de las herramientas analíticas. En este estudio se abordó una de ellas: el efecto del índice de similitud sobre la clasificación biogeográfica, es decir se abordó el estudio de los índices de similitud en relación con: 1) la propia matriz de similitud generada y 2) los tipos de ligamientos más comúnmente usados en la clasificación biogeográfica. En este trabajo se han obtenido resultados que aunque útiles en la toma de decisiones sobre qué herramientas utilizar, no son definitivos. Los índices de similitud son una herramienta útil y muy recurrida en análisis biogeográfico, sin embargo, aún el problema de qué índice utilizar permanece abierto. Análogamente al problema de decisión sobre el índice de similitud, existen otros, por ejemplo, el problema del tipo de ligamiento, o de manera más general, sobre el tipo de clasificación (aglomerativa o divisiva). Las herramientas de análisis de la BC son útiles, pero no existe información suficiente sobre ellos que guíe de manera objetiva al biogeógrafo en su elección.

## **Descripción de los Métodos de la BC**

La descripción que se presenta en este trabajo de los métodos de la BC como resultado del 'análisis informático' provee de una base sobre la que se pueden realizar análisis más profundos. Es necesario analizar a los métodos, en forma individual y de forma integrada, es decir, a una gran parte de ellos estudiando sus relaciones, no sólo entre un par de ellos, como se hizo aquí con los índices de similitud y los tipos de ligamientos, sino entre los elementos de un conjunto que represente aquel conjunto de herramientas que el biogeógrafo normalmente utiliza en sus análisis. Ese 'análisis integral' deberá de observar los avances y aproximaciones de las áreas de los Sistemas de Información Geográfica y en general del Análisis Espacial, que en años recientes está surgiendo ya como una disciplina científica consolidada, muestra de esto es la aparición de libros en los se integran capítulos con la intención de esa consolidación (*v.gr.* Arlinghaus, 1996; Fotheringham y Rogerson, 1994), otro ejemplo es el cambio de nombre de la revista '*International Journal of Geographical Information Systems*' al de '*International Journal of Geographical Information Science*'.

## 11.2 Desarrollo de Herramientas para la Biogeografía Cuantitativa

El diseño de la herramienta presentada en este trabajo provee un marco conceptual del método de la BC, una ordenación de las técnicas y algunas de las maneras en que pueden interactuar en el proceso de transformación de la información. Aunque las herramientas de desarrollo, tanto las de diseño como las de programación, han ido mejorando sustancialmente, desarrollar sistemas es una tarea compleja y más aún sistemas de información que tengan que ver con análisis espacial. El análisis informático de los métodos de la BC realizado como una etapa previa e indispensable para el desarrollo del sistema, presenta un marco estructurado de cierta parte del conocimiento de la BC.

Existen varias herramientas que permiten la implementación de algunas de las técnicas de la BC (v.gr. ESRI, 1996; INTERA TYDAC, 1993; INTERA TYDAC, 1993a); sin embargo, por el momento esas herramientas son de difícil acceso al biogeógrafo debido a 1) el elevado costo de ellas, y 2) el entrenamiento necesario para su uso, tanto específicamente de la herramienta como de algunos conocimientos informáticos sobre programación. La ‘popularización’ de herramientas informáticas de análisis para el biogeógrafo provocará mayor uso y necesidades más específicas que podrán ser cubiertas mediante a) el desarrollo de herramientas, o bien, b) el uso más intensivo y profundo de herramientas GIS.

Así como un índice de un libro presenta su contenido de conocimiento de manera ordenada y resumida, los menús ‘*pull-down*’, entre otras interfaces gráficas, de los sistemas de información presentan los procesos disponibles en el paquete. La interfaz del software es una manera de presentar las posibles transformaciones automáticas de la información que en muchos casos pueden constituir conocimiento. De esta manera, el *software* contiene conocimiento científico, ya que contiene conocimiento ordenado y sistematizado. El *software* es una manera de transferir, automatizar y generar conocimiento científico, así también el desarrollo de *software* científico es una manera de generar conocimiento científico. Para el caso particular de la BC, el análisis informático de los métodos de la BC - realizado como una de las etapas del desarrollo de un GIS para la BC – constituyó generación de conocimiento científico; el diseño de la herramienta

(por ejemplo los cuadros de diálogo de la interfaz con el usuario y el menú de comandos) también es conocimiento científico y no sólo una manera de presentarlo.

# 12. Conclusiones

---

*En este capítulo se presentan las conclusiones del análisis de los métodos de la BC, haciendo énfasis en la necesidad de conocer mejor, técnica y teóricamente, a las herramientas disponibles. Se concluye también que el desarrollo de herramientas informáticas para la BC es una necesidad del biogeógrafo aún no cubierta.*

## 12.1. Fundamento técnico y teórico de la Biogeografía Cuantitativa

En los capítulos previos se ha abordado el ‘análisis de los métodos de la BC’; el término ‘analizar’ implica *describir y clasificar estableciendo relaciones entre objetos concatenándolos*. Las tecnologías de la información reservan un significado particular al ‘análisis’ como una etapa del desarrollo informático en la que se dedican recursos a entender el problema a automatizar. La aplicación de las metodologías de análisis de la informática a los métodos de la BC ha dado como resultado en este trabajo una integración coherente, una descripción homogénea en la que los elementos analizados pueden compararse y clasificarse. Esa descripción, obtenida del análisis, constituyó la base sobre la que se construyó un modelo estructurado del conocimiento de los métodos de la BC.

La tecnología informática ya está integrada a la ciencia, como método y como conocimiento en sí. La tecnología de la información ha generado una base sólida de conocimientos para estructurar a la información de manera ordenada y sistematizada, ese conjunto

de conocimientos es, a su vez, ordenado y sistematizado rompiendo con la línea que dividía a la ciencia de la tecnología.

La BC, entendida como un conjunto de métodos 'cuantitativos' para realizar análisis biogeográfico, se ha ido construyendo a la par de la disposición de métodos cuantitativos y el desarrollo de la informática y ha carecido de una *integración estructurada* que la soporte como *metodología formal* de análisis. El modelo de los 'caminos metodológicos' propuesto en este trabajo *integra* a los métodos pero no los *fundamenta*. Ese modelo facilita la aplicación y a su vez el entendimiento de los métodos, falta aún el análisis profundo de cada uno de los métodos que brinde conocimiento sobre sus límites, sobre las condiciones necesarias para su uso y sobre la cuantificación de la calidad de sus resultados. El índice Qs es una propuesta para cuantificar la calidad de la información y auxilia en el proceso de generación de OGUs. El análisis del efecto del índice de similitud en la clasificación biogeográfica cuantifica las posibles diferencias en los resultados únicamente debidas al índice de similitud y tipo de ligamiento elegidos.

Así, en este trabajo se ha abordado, en menor o mayor grado, el análisis de los métodos de la BC que generalmente se ocupan en etapas tempranas del análisis biogeográfico:

Conformación de OGUs

Índice de similitud

Matriz de similitud

Ligamientos

Aunque se han obtenido resultados útiles, pragmáticos, que sirven al biogeógrafo que trabaja en una región particular con problemas concretos sobre la distribución de organismos, esos resultados están lejos de presentar a la BC como una metodología sólida. Falta abundar, generar mayor número de propuestas de análisis de cada uno de los métodos, evaluarlas y proponer paradigmas que hagan de la BC una metodología formal, y sobre todo críticamente fundamentada.

En general, el objetivo de la aplicación de cada método es claro, pero no las consecuencias de la elección de sus parámetros. Por ejemplo, la conformación de OGUs es necesaria para la generalización del muestreo de los registros biogeográficos y sus parámetros son el tamaño y la forma; las consecuencias de elegir un tamaño y forma particular frente a otras opciones no son

claras. Otro ejemplo: la matriz de similitud mide la semejanza entre cada par de OGU, pero cómo se altera su estructura al cambiar de índice de similitud es algo que no se sabe.

Crovello (1981) ve al análisis biogeográfico mediante técnicas de la BC como un proceso de toma de decisiones, el problema es que aunque las alternativas sean claras, no hay suficiente información que permita elegir entre ellas. Cuando se toma una decisión se debe de estar informado de sus consecuencias, comparadas contra las que se tendrían si se hubiera elegido otro camino; en la BC no se conocen esas consecuencias. El índice Qs construido en este trabajo es una propuesta para que el biogeógrafo conozca **algunas** de las consecuencias de cuando menos **uno** de los pasos iniciales en el análisis biogeográfico: algunos aspectos en la conformación de las OGU. El índice Qs permite al biogeógrafo elegir, conociendo algunas consecuencias de esa elección, la conformación de las OGU.

En la construcción del índice Qs se tomó en cuenta la precisión contra la resolución; la escala se entendió solo como un cambio en la resolución (tamaño de OGU), pero la escala también puede considerarse como las dimensiones espaciales en las que se observan diferentes fenómenos. Así, al aumentar la resolución puede disminuir la precisión, pero sólo para aquellos fenómenos que se observan a una escala determinada. Al cambiar la escala y por lo tanto la posibilidad de observar otros fenómenos, la relación inversa entre precisión y resolución no aplica. Quattrochi y Goodchild (1996), dentro de un contexto de percepción remota y SIG, proponen el establecimiento de una 'ciencia de la escala' que proporcione una estructura formal para el manejo de la escala y consideran como un punto importante la:

*habilidad para cambiar la escala: ¿qué tipos de transformaciones de la escala están disponibles para agregar o desagregar datos de manera lógica, rigurosa y bien fundamentada teóricamente? ¿Se puede desarrollar un conjunto genérico de métodos para desagregar y agregar datos de manera que sean compatibles con los procesos del sistema Tierra?*

Se puede observar que el proceso y problemática de generar OGU es una visión particular de esa 'ciencia de la escala'; el lugar adecuado para integrar el problema es en el ámbito de los SIG.

El índice Qs tiene límites de aplicación que no han sido establecidos con claridad, y como se explicó anteriormente, los resultados que surjan al aplicarlo a escalas diferentes en las que se

observen fenómenos diferentes no son comparables; únicamente pueden compararse valores de Qs para tamaños de OGU en las que se observa el mismo fenómeno a estudiar.

Los límites de Qs y nuestra ignorancia de ellos es solo una muestra, esto es, un ejemplo de los aspectos que faltan por conocer de los métodos de la BC. El análisis del ‘efecto del índice de similitud sobre la clasificación biogeográfica’ es otro ejemplo que muestra la falta de conocimiento de las consecuencias de la aplicación de los métodos de la BC.

Conocer mejor al conjunto de los métodos de la BC es conocer mejor al propio proceso biogeográfico. El descubrimiento de patrones de distribución de las especies, así como la proposición de sus posibles causas se puede realizar desde diferentes aproximaciones filosóficas, la BC es una de esas aproximaciones, un mejor fundamento de ella es un avance en el conocimiento científico.

Se pueden resumir las conclusiones de los trabajos presentados, respecto a la medición de la calidad de los registros biogeográficos, en la siguiente lista:

- 1) El problema del “tamaño del cuadro” y del “tamaño y forma de la OGU” son una versión del llamado MAUP (de las iniciales en inglés de “Modifiable Area Unit Problem”) de los Sistemas de Información Geográfica.
- 2) La calidad de los registros biogeográficos puede ser medida mediante una comparación de la “riqueza estimada” contra la “riqueza conocida”.
- 3) La calidad de los registros biogeográficos puede ser medida además, mediante una combinación de dos factores: 1) una comparación de la “riqueza estimada” contra la “riqueza conocida”, y 2) considerando la suma de frecuencias en la matriz de presencia ausencia.
- 4) El índice Qs es una herramienta que mide la calidad de los registros biogeográficos con el criterio anterior.
- 5) Valores de Qs de 15%, 50% y 60% para especies, géneros y familias, respectivamente, se consideran como los mínimos de buenos índices de calidad para un mosaico de OGU de 32 estados de la República Mexicana.

- 6) La calidad de los registros biogeográficos está relacionada con el nivel de resolución: la calidad aumenta al disminuir la resolución (la resolución puede ser geográfica, taxonómica o de otra índole).
- 7) El índice Qs es una herramienta que permite comparar la calidad de los registros de estudios biogeográficos de un grupo de organismo sobre regiones diferentes, o entre estudios de la misma región pero con diferentes taxones.

La siguiente lista es una lista que resume las conclusiones de los trabajos presentados respecto al efecto del índice de similitud en las clasificaciones biogeográficas:

- 1) El índice de similitud puede influir hasta el 70% en las diferencias de las clasificaciones biogeográficas.
- 2) El algoritmo de agrupamiento puede influir hasta el 53% en las diferencias de las clasificaciones biogeográficas, y muy probablemente arriba del 16%.
- 3) Las clasificaciones biogeográficas realizadas con diferente índice de similitud no son comparables.
- 4) Las clasificaciones biogeográficas realizadas con diferente algoritmo, aunque con el mismo índice de similitud, no son comparables.
- 5) Si no hay un argumento en favor de utilizar un índice de similitud en particular, el índice de Jaccard es una buena opción, debido a que a) las clasificaciones biogeográficas realizadas con diferente algoritmo de agrupamiento muestran pocas diferencias (por lo tanto son comparables), y b) muestra alto porcentaje de definición (ver texto del artículo y concepto de “topología altamente informativa”).

## 12.2 Análisis espacial y automatización de los métodos de la Biogeografía Cuantitativa

La discusión actual en la literatura sobre los SIG (*v. gr.* Bailey, 1994; Rogerson y Fotheringham, 1994; Quattrochi y Goodchild, 1996) y en general sobre análisis espacial (Wong, 1996;

Openshaw, 1998) provee un marco conceptual con el que la BC puede tener gran relación. El desarrollo de los SIG han potenciado el uso del análisis espacial haciendo peticiones cada vez más precisas y útiles que ayuden a explicar fenómenos espaciales; también ha precisado el significado de 'análisis espacial' y ha puesto de manifiesto que el desarrollo de herramientas teóricas para el análisis ha sido impulsado por la posibilidad de automatizarlos, es decir, por el propio desarrollo de los SIG.

La BC es un caso particular de esa problemática; necesita que el desarrollo de herramientas se incremente, que el biogeógrafo se involucre aún más, no sólo con el uso de esas herramientas de análisis, sino con su propio diseño.

La necesidad del biogeógrafo de contar con herramientas informáticas de análisis aún no se ha cubierto satisfactoriamente, más bien, comienza a definirse con mayor claridad esa necesidad.

La siguiente lista es una lista que resume las conclusiones de los trabajos presentados respecto a la biogeografía cuantitativa como herramienta de análisis espacial en biogeografía:

- 1) Los métodos de la biogeografía cuantitativa se pueden clasificar en dos tipos: a) los métodos heredados directamente de la taxonomía numérica; y b) los métodos diseñados especialmente para analizar procesos biogeográficos.
- 2) El desarrollo de herramientas teóricas de análisis espaciales ve favorecido por el desarrollo de técnicas automatizadas, en especial en el desarrollo de sistemas de información geográfica.
- 3) Una aproximación de análisis a los métodos de la biogeografía cuantitativa es la aplicación de metodologías de análisis de los sistemas de información.
- 4) El modelo de los 'camino metodológicos' propuesto en esta tesis es una herramienta para analizar y describir las posibles interacciones entre diversos métodos de la biogeografía cuantitativa.
- 5) El concepto de 'especie informática' propuesto en esta tesis es útil en la construcción del conocimiento biogeográfico pues marca la necesidad de especificar el concepto de especie para cada valor almacenado en una base de datos.

# Referencias

---

- ALKIN, R. 1998. **Effective management and delivery of Biodiversity information**, pp. 87-102. En: Bridge, P., P. Jeffries, D. R. Morse y P. R. Scott (eds.). **Information Technology, Plant Pathology & Biodiversity**. CAB International. Londres.
- ANDERSON, L. 1996. **An ontological dilemma: epistemology and methodology of historical biogeography**. *J. Biogeog.*, 23:269-277.
- ARITA, H. 1993. **Riqueza de especies de la mastofauna de México**. pp. 109-128. En: Medellín, R. A. y Cevallos, G. (eds.). **Avances en el estudio de los mamíferos de México**. Asoc. Mex. de Mastozoología. México.
- ARLINGHAUS, S. L. (ED.). 1996. **Practical handbook of spatial statistics**. CRC Press.
- BAILEY, T. C. 1994. **A review of statistical spatial analysis in GIS**. pp. 13-44. En: Fotheringham, S. y P. Rogerson (eds.). **Spatial Analysis and GIS**. Taylor & Francis.
- BAUM, D. 1992. **Phylogenetic species concepts**. *Tree*, 7(1):1-2.
- BARONI-URBANI, C. Y M. W. BUSER. 1976. **Similarity of binary data**. *Syst. Zool.*, 25:204-210.
- BISBY, F. A. 1998. **Putting names to things and keeping track: the Species 2000 Programme for a coordinated catalogue of life**, pp. 59-68. En: Bridge, P., P. Jeffries, D.R. Morse y P.R. Scott (eds.). **Information Technology, Plant Pathology & Biodiversity**. CAB International. Londres.
- BIRKS, H. J. B. 1976. **The distribution of european pteridophytes: a numerical analysis**. *New Phytol.*, 77: 257-287.
- CAMPBELL, R. C. 1989. **Statistics for Biologist**. 3ª ed. Cambridge Univ. Press. 446 p.

- CARPENTER, G., A. N. GILLISON Y J. WINTER. 1993. **DOMAIN: a flexible modelling procedure for mapping potential distributions of plants and animals.** *Biodiversity and Conservation*, 2: 667-680.
- CLARIDGE, M. E., H. A. DAWAH Y M.R. WILSON. 1997. **Practical approach to species concepts for living organisms.** En: Claridge, M. F., H. A. Dawah y M. R. Wilson. (eds.). **Species. The units of biodiversity.** Chapman & Hall. Londres
- CODD, E. F. 1990. **The Relational Model for Database Management.** Addison-Wesley Pub.
- CRACRAFT, J. 1997. **Species concepts in systematics and conservation biology.** En: Claridge, M. F., H. A. Dawah y M.R. Wilson. 1997. **Species. The units of biodiversity.** Chapman & Hall. Londres.
- CRAW, R., 1988. **Panbiogeography: method and synthesis in biogeography.** pp: 405-435. En: Myers, A.A. y P. Giller (eds.). *Analytical Biogeography.* Chapman and Hall. Londres.
- CRAW, R. 1989. **Quantitative Panbiogeography; Introduction to methods.** *New Zealand J. Zool.*, 16: 485 - 494.
- CRAW, R. J. GREHAN Y M. HEADS. 1999. **Panbiogeography, tracking the history of life.** *Oxford Univ. Press*, Oxford. 229 p.
- CRISCI, J. 1994. **La especie: realidad y conceptos.** En: Llorente, J. e I. Luna (Comps.). *Taxonomía Biológica, un libro de texto universitario.* UNAM-Fondo de Cultura Económica. México.
- CRISCI, J. V. Y M. F. LÓPEZ. 1983. **Introducción a la teoría y práctica de la taxonomía numérica.** OEA. Washington, D.C. 132 p.
- CROIZAT, L., G. NELSON Y D. ROSEN. 1974. **Centers of origin and related concepts.** *Systematic Zool.*, 23(2):265-287.
- CROVELLO, TH. J. 1981. **Quantitative biogeography: an overview.** *Taxon*, 30(3):563-575.
- DE LUNA, E. 1995. **Bases filosóficas de los análisis cladísticos para la investigación taxonómica.** *Acta Bot. Mex.*, 33:63-79.

- DIRZO, R. 1990. **La biodiversidad como crisis ecológica actual ¿qué sabemos?**  
*Ciencias*, 4: 48-55.
- DIRZO, R. Y P. H. RAVEN. 1994. **Un inventario biológico para México.** *Bol. Soc. Bot. México*, 55: 29-34.
- EGUARTE, L. 1995. **Hutchinson (Agavales) vs. Huber y Dahlgren (Asparagales): análisis moleculares sobre la filogenia y evolución de la familia Agavaceae sensu Hutchinson dentro de las monocotiledóneas.** *Bol. Soc. Bot. México*, 56: 45-56.
- ESPINOSA, D. Y LLORENTE, J. 1993. **Fundamentos de biogeografías filogenéticas.** UNAM. México. 133 p.
- ESRI. 1996. **Avenue. Customization and application development for ArcView.** ESRI Inc. 260 p.
- FEIGENBAUM, E. A. 1977. **The art of Artificial Intelligence: themes and case studies of knowledge engineering.** *IJCAI*. 5: 1014-1029.
- FOTHERINGHAM, S. Y P. ROGERSON (EDS.). 1994. **Spatial Analysis and GIS.** Taylor & Francis. 281 p.
- GAUCH, H. G. 1981. **Multivariate analysis in community ecology.** Cambridge Univ. Press. Cambridge. 198 p.
- HAINING, R. 1994. **Designing spatial data analysis modules for geographical information systems.** En: Fotheringham, S. y P. Rogerson (eds.). **Spatial Analysis and GIS.** pp. 45-63. Taylor & Fancis.
- HAYEK, L. A. C. 1993. **Analysis of amphibian biodiversity data.** En: Heyer, W. R., M. A. Donnelly, R. W. McDiarmid, L. A. C. Hayek y M. S. Foster (eds.). **Measuring and monitoring biological diversity. Standard methods for amphibians.** Smithsonian Inst. Press.
- HENDERSON, L. M. 1989. **Quantitative Panbiogeography; an investigation into concepts and methods.** *New Zeland J. Zool.* 16: 495 -510.
- HENGEVELD, R. 1990. **Dynamic biogeography.** Cambridge Univ. Press. Cambridge. 249 p.

- HUBÁLEK, Z. 1982. **Coefficients of association and similarity, based on binary (presence-absence) data: an evaluation.** *Biol. Rev.* 57: 669-689.
- HULL, D. L. 1997. **The ideal species concept - and why we can't get it.** En: Claridge, M. F., H. A. Dawah y M. R. Wilson (eds.). 1997. **Species. The units of biodiversity.** Chapman & Hall. London.
- INTERA TYDAC. 1993. **SPANS Modeling handbook.** INTERA TYDAC Technologies Inc. 136 p.
- INTERA TYDAC. 1993a. **SPANS Technical manual.** INTERA TYDAC Technologies Inc. 7 capítulos.
- KANT, M. 1781 [1991]. **Crítica de la razón pura.** Porrúa. México. 375 p.
- KOHLMANN, B. 1994. **Algunos aspectos de la taxonomía numérica y sus usos en México.** En: Llorente, J. e I. Luna (Comps.). 1994. *Taxonomía Biológica, un libro de texto universitario.* UNAM-Fondo de Cultura Económica. 656 p.
- KOLEFF, P. 2002. **Spatial species turnover: patterns, determinants, and implications.** Tesis de Doctorado. University of Sheffield, London. 195 p.
- LEÓN, L. Y E. ROMO. 1993. **Mastofauna de la sierra de Taxco, Guerrero.** pp. 45-64. En: Medellín, R. A. y Cevallos, G. (eds.). **Avances en el estudio de los mamíferos de México.** Asoc. Mex. de Mastozoología. México. 45-64 p.
- LUCKOW, M. 1995. **Species concepts: assumptions, methods and applications.** *Syst. Bot.* 4(20): 589-605.
- LUNA, I. 1994. **Los conceptos de especie evolutiva y filogenética.** En: Llorente, J. e I. Luna (Comps.). 1994. *Taxonomía Biológica, un libro de texto universitario.* UNAM-Fondo de Cultura Económica. México.
- LLORENTE, J. Y P. KOLEFF. 1997. **La actividad taxonómica en México.** *Biodiversitas.* CONABIO. 13: 11-14.
- LLORENTE, J., I. LUNA, J. SOBERÓN Y L. BOJÓRQUEZ. 1994. **Biodiversidad, su inventario y conservación: teoría y práctica en la taxonomía alfa contemporánea.** En: Llorente, J. e I. Luna (Comps.). 1994. *Taxonomía Biológica, un libro de texto universitario.* UNAM-Fondo de Cultura Económica. México. 656 p.

- LLORENTE, J. Y L. MICHÁN. 2000. **El concepto de especie y sus implicaciones para el desarrollo de inventarios y estimaciones en biodiversidad**. Monografías Tercer Milenio. SEA Zaragoza. 1: 87-93.
- LLORENTE, J., J. J. MORRONE, A. BUENO, R. PÉREZ-HERNÁNDEZ, A. VILORIA, D. ESPINOSA. 2000. **Historia del desarrollo y la recepción de las ideas panbiogeográficas de León Croizat**. Rev. Acad. Colomb. Cienc., 24: 549-577.
- MAY, R. M. 1990. **Taxonomy as destiny**. *Nature* 347: 129-130.
- MAYDEN, R. L. 1997. **Hierarchy of species concepts: the denouement in the saga of the species problem**. En: Claridge, M. F., H. A. Dawah y M. R. Wilson (eds.). 1997. **Species. The units of biodiversity**. Chapman & Hall. London. 439 p.
- MCCONNELL, S. 1997. **Desarrollo y gestión de proyectos informáticos**. McGraw Hill.
- MCDADE, L. A. 1995. **Species concepts and problems in practice: insight from botanical monographs**. *Syst. Bot.* 4(20): 606-622.
- MORRONE, J. J., D. ESPINOSA Y J. LLORENTE. 1996. **Manual de biogeografía histórica**. Universidad Nacional Autónoma de México. México. 155 p.
- NELSON, G. Y N. PLATNICK. 1984. **Biogeography** (J.J. Head ed.) No. 119. Carolina Biology Readers. 16 p.
- O'KELLY, M.E. 1994. **Spatial Analysis and GIS**. pp. 65-80. En: Fotheringham, S. y P. Rogerson (Ed). **Spatial Analysis and GIS**. Taylor & Francis.
- OPENSHAW, S. 1998. **Optimizing scale**. School of Geography, Univ. of Leeds <http://www.geog.leeds.ac.uk/research/presentations/98-10/sl1001.htm> (Noviembre/2000).
- PALMER, M. 1998. **Ordination Methods for ecology**. Department of Botany, Oklahoma State Univ. <http://www.okstate.edu/artsci/botany/ordinate> (Junio/1998).
- PHIPPS, J. B. 1975. **BestBlock: optimizing grid size in biogeographic studies**. *Canadian Journal of Botany* 53(15): 1447-1452.
- PODER EJECUTIVO FEDERAL. 1996. **Programa de Medio Ambiente 1995-2000**. Gobierno de los Estados Unidos Mexicanos. México. 172 p.
- POPPER, C. 1966. **Conjectures and refutations: the growth of scientific knowledge**. Harper and Row. Neva York, USA.

- QUATTROCHI, D. A. Y M. F. GOODCHILD. 1996: **Scale, multiscaling, remote sensing, and GIS**. En: Quattrochi, D. A. y M. F. Goodchild (ed.). **Scale in remote sensing and GIS**. Lewis Pub. 406 p.
- REBELO, A. G. 1994. **Iterative selection procedures: centres of endemism and optimal placement of reserves**. *Sterlizia* 1: 231-257.
- REBELO, A. G. Y W. R. SIEGFRIED. 1990. **Protection of fynbos vegetation: ideal and real-word options**. *Biological Conservation* 54: 15-31.
- ROGERSON, P. A. Y A. S. FOTHERINGHAM. 1994. **GIS and spatial analysis: introduction and overview**. pp. 1.-10. En: Fotheringham, S. y P. Rogerson (eds.). **Spatial Analysis and GIS**. Taylor & Francis. 281 p.
- ROSEN, D. E. 1978. **Vicariant patterns and historical explanation in biogeography**. *Syst Zool.* 27: 159-188.
- SALAZAR, E. J. 1979. **Modelos esquemáticos para la elaboración de planes de estudio en la educación superior**. ANUIES. México, 195 p.
- SALTHER, S. N. 1985. **Evolving hierarchical systems**. Columbia Univ. Press. New York. 343 p.
- SÁNCHEZ, O. 1993. **Análisis de Algunas tendencias ecogeográficas del género *Reithrodontomys* (Rodentia: Muridae) en México**. pp. 25-44. En: Medellín, R. A. y Cevallos, G. (eds.). **Avances en el estudio de los mamíferos de México**. Asoc. Mex. de Mastozoología. México. 464 p.
- SÁNCHEZ, O. Y G. LÓPEZ. 1988. **A theoretical analysis of some indices of similarity as applied to biogeography**. *Folia Entomológica Mexicana* 75: 119-145.
- SARUKHÁN, J. 1995. **Diversidad biológica**. *Universidad de México. Revista de la UNAM* 536/537: 3-10.
- SIMBERLOFF, D. 1983. **Biogeography: The unification and maturation of a science**. pp. 411-473. En: Brush, A. H. y G. H. Klark (eds.). **Perspectives in Ornithology**. Cambridge Univ. Press.
- SNEATH, P. H. A. Y R. R. SOKAL. 1973. **Numerical Taxonomy**. W.H. Freeman. San Francisco. 573 p.

- SNEATH, P. H. A. 1998. **Development of computer-based systems in systematics.** pp. 15-26. En: Bridge, P., P. Jeffries, D. R. Morse y P. R. Scott (ed.). **Information Technology, Plant Pathology & Biodiversity.** London. CAB International. 478 p.
- SNI. 1994. **Datos del SNI en 1994.** *Ciencia y Desarrollo* 119: 18-23.
- SOULÉ, M. E. 1990. **The real work of systematics.** *Ann. Missouri Bot. Gard.* 77: 4-12.
- SYSTEMATICS AGENDA 2000. 1994. **Systematics Agenda 2000: Charting the Biosphere.** Technical Report. New York. 34 p.
- VANE-WRIGHT, R. I., C. J. HUMPHRIES Y P. H. WILLIAMS. 1991. **What to protect? - Systematics and the agony of choice.** *Biol. Conserv.*, 55: 235-254.
- VILLARREAL, J. A., J. VALDÉS Y J. L. VILLASEÑOR. 1996. **Corología de las asteráceas de Coahuila, México.** *Acta Bot. Mex.*, 36: 29-42.
- VILLASEÑOR, J. L. Y P. DÁVILA. 1992. **Breve introducción a la metodología cladística.** Facultad de Ciencias, Universidad Nacional Autónoma de México. México. 48 p.
- VILLASEÑOR, J. L. Y G. IBARRA-MANRIQUEZ. 1998. **La riqueza arbórea de México.** *Boletín IBUG* 5(3): 95-105.
- VILLASEÑOR, J. L. Y J. L. STROTHER. 1989. **Tuxtla: a new genus for Zexmenia pittieri (Compositae: Heliantheae).** *Syst. Bot.* 14(4): 529-540.
- WELLER, J. M. 1961. **The species problem.** *J. Paleontology* 35(6): 1181-1192.
- WHITTEN, J. L., I.D.BENTLEY Y V. M. BARLOW. 1997. **Análisis y diseño de sistemas de información.** McGraw Hill.
- WONG, D. 1996. **Aggregation Effects in Geo-referenced Data.** pp. 86-103. En: Arlinghaus, S. L. (ed.). **Practical handbook of spatial statistics.** CRC Press. 307 p.
- WORLD BANK. 1995a. **Mainstreaming Biodiversity in Development.** Environment Department Paper No. 29. Washington, D.C. 45 p.
- WORLD BANK. 1995b. **Monitoring Environmental Progress.** A report on work in progress. Environment Department. Washington, D. C. 82 p.
- ZAVALA, J. A. 1986. **Introducción al enfoque multivariado en estudios de vegetación.** Instituto Nacional de Recursos Biológicos (INIREB). Xalapa, México. 58 p.