

Una breve revisión de técnicas de análisis multivariantes aplicables en Fitosociología

ADRIÁN ESCUDERO*, ROSARIO GAVILÁN*
& AGUSTÍN RUBIO**

* Departamento de Biología Vegetal II. Facultad de Farmacia
Universidad Complutense de Madrid. E-28040

** Departamento de Silvopascicultura. ETSI Montes
Universidad Politécnica de Madrid. E-28040

Resumen:

ESCUADERO, A.; GAVILÁN, R. & RUBIO, A. 1994. Una breve revisión de técnicas multivariantes aplicables en Fitosociología. *Botanica Complutensis*, 19: 9-38.

Se presenta una breve semblanza de las técnicas multivariantes que se están utilizando actualmente en los estudios de vegetación, intentando enfocarla a la resolución de problemas fitosociológicos. Las técnicas se han agrupado en dos grandes bloques, las denominadas técnicas de gradiente y las de clasificación numérica. Previo al estudio de dichas técnicas se comentan algunos conceptos generales como son la transformación y la estandarización de los datos, así como la medida de la semejanza. Las técnicas de gradiente han sido desglosadas para su estudio en técnicas de regresión, de calibración, de ordenación y de ordenación constreñida. Las de clasificación numérica se han dividido en métodos jerárquicos y divisivos. Finalmente, se tratan los procedimientos numéricos más avanzados, como son las técnicas de comparación de particiones o las de generación de clasificaciones consensuadas, entre otras.

Palabras clave: Análisis Multivariantes, Clasificación, Fitosociología, Ordenación.

Abstract:

ESCUADERO, A.; GAVILÁN, R. & RUBIO, A. 1994. A brief review of multivariate methods with Phytosociological applicability. *Botanica Complutensis*, 19: 9-38.

A brief review of multivariate methods used in Phytosociological studies is presented. These methods have been grouped in two items: gradient techniques and numerical classifications. Some general concepts like transformation and standarization of data and similarity measures have been previously commented. Gradient methods have been classified in regression, calibration, ordination and constrained ordination techniques and, in the other hand, we have considered hierarchical and divisive methods under classification procedures. Finally, advanced numerical approaches, like partitions comparisons or consensus classifications, are treated.

Keywords: Multivariate Analysis, Classification, Phytosociology, Ordination.

INTRODUCCIÓN

Aunque el valor de la Fitosociología como ciencia capaz de describir la vegetación es incuestionable (BRAUN-BLANQUET, 1964; MUCINA & AL., 1993; WESTHOFF & VAN DER MAAREL, 1978), también se ha puesto de manifiesto la necesidad de resolver diversas lagunas y problemas de tipo metodológico (VAN DER MAAREL, 1975; GOODALL, 1978; FISCHER & BEMMERLEIN-LUX, 1989; LEPS & HANDICOVA, 1992). El desarrollo, a partir de los años sesenta, de métodos numéricos útiles para el estudio de la vegetación ha revolucionado, por lo menos en el contexto europeo, los estudios de vegetación (KENT & BALLARD, 1988) y, por extensión, los fitosociológicos (MUCINA & VAN DER MAAREL, 1989).

La Sintaxonomía Numérica, como ciencia auxiliar de la Fitosociología, fue definida por VAN DER MAAREL (1981) como consecuencia del desarrollo de la Fitosociología Numérica (LAMBERT & DALE, 1964; VAN DER MAAREL, 1969; DALE & WEBB, 1975), ya que permitía resolver algunos de los problemas de la Fitosociología clásica, como son: la objetivación del proceso de obtención de la jerarquía de tipos, la objetivación de las similitudes absolutas o relativas entre los inventarios, el cálculo de la homogeneidad a cada nivel jerárquico o la posibilidad de relacionar los niveles sintaxonómicos con los obtenidos en las clasificaciones y la información sintaxonómica con la jerárquica de una forma objetiva (FOLI & LAUSI, 1981).

Pese al fuerte desarrollo de la Fitosociología española (PEINADO & RIVAS-MARTÍNEZ, 1987), las aproximaciones numéricas realizadas hasta ahora no han sido demasiado numerosas (TARAZONA, 1984; BUENO & FERNÁNDEZ-PRIETO, 1991; MORENO & AL., 1990; TARAZONA & AL, 1991; ESCUDERO & PAJARÓN, 1994). En este trabajo pretendemos dar una visión panorámica de las técnicas numéricas más ampliamente utilizadas en estudios fitosociológicos, intentando sintetizar las revisiones aparecidas en los últimos años sobre distintos aspectos y aproximaciones (ORLÓCI, 1978; GAUCH, 1982; PIELOU, 1984; DIGBY & KEMPTON, 1986; WILDI & ORLÓCI 1990; LEGENDRE & LEGENDRE 1983). En el ámbito español también destacan los trabajos pioneros desarrollados por DÍAZ PINEDA & GONZÁLEZ BERNÁLDEZ (1978) y DÍAZ PINEDA & AL. (1979) sobre la descripción automática de la vegetación.

Estas técnicas las hemos agrupado en dos bloques independientes, por un lado los **análisis de gradiente**, que se encargarían de describir las relaciones entre los datos de vegetación con los datos ambientales, y por otro, las **clasificaciones numéricas**, que permitirían estructurar esos datos, agrupándolos, de forma que los grupos resultantes sean lo más homogéneos posible, y que estudiaremos en primer lugar.

Primero revisaremos algunos conceptos previos relacionados con la transformación de los datos, de importancia capital tanto en el contexto de las clasificaciones como en el del análisis de gradiente. Posteriormente estudiaremos la medida de la semejanza y a continuación ambos bloques de técnicas, en primer lugar las técnicas de gradiente y en segundo lugar las ordenaciones.

TRANSFORMACIÓN Y ESTANDARIZACIÓN DE DATOS

Las matrices de datos ecológicos frecuentemente deben ser manipuladas en diferentes sentidos con el fin de neutralizar algún fenómeno que no es de interés (NOY-MEIR, 1973). En algún caso, estas transformaciones pueden ir implícitas en un análisis en concreto, caso de las ordenaciones relacionadas con el CA («weighted averaging» –media ponderada–), aunque generalmente es una decisión del investigador. De forma sintética, dichas manipulaciones se pueden agrupar en:

- * *Transformaciones.* Son alteraciones de los datos individuales en el conjunto de los mismos. Se reemplazan los datos brutos por otros, aplicando a todos la misma función algebraica. Entre las funciones más utilizadas, podemos resaltar la raíz cuadrada, las funciones logarítmicas o las funciones exponenciales. Los efectos tanto de la transformación logarítmica como de la raíz cuadrada se relacionan con la atenuación de la influencia de las especies dominantes, mientras que la función exponencial enfatiza dicha influencia (VAN TONGEREN, 1987).

La escala de abundancia-dominancia utilizada en el método fitosociológico hace necesaria la transformación de los datos a una escala de tipo ordinal para poder aplicar sobre ellos cualquier tipo de análisis numérico. Existen diversos métodos de transformación de dicha escala (SCHWICKERATH, 1931, 1938, 1940; TÜXEN & ELLEMBERG, 1937), de los cuales algunos son de tipo cualitativo, basados únicamente en datos de presencia/ausencia (WILLIAMS & LAMBERT, 1959; VAN DER MAAREL, 1966), o pseudocualitativos, en los que los valores de cobertura más bajos se toman como ausencias (DAGNELIE, 1960). Estudios clásicos comparativos de unos y otros se pueden encontrar en JENSEN (1978), CAMPBELL (1978) o VAN DER MAAREL (1979). WESTHOFF & VAN DER MAAREL (1978) proponen una de las transformaciones más utilizadas hoy en día, la cual consiste en una escala numérica del 1 al 9, en la que el valor 2 de la escala fitosociológica se divide en tres partes (atendiendo a los porcentajes de la cobertura), de acuerdo con los criterios de BARKMANN & AL. (1964).

- * *Estandarizaciones.* Son alteraciones que se realizan de forma independiente a cada uno de los vectores (inventarios o especies) que compo-

nen la matriz de datos, brutos o previamente transformados. Hemos incluido bajo este epígrafe el centrado y la normalización como casos particulares de la estandarización (CLIFFORD & STEPHENSON, 1975). El centrado de los datos consiste en restar a cada uno de los elementos que forman un vector determinado el valor medio de dicho vector, mientras que en la normalización se divide cada uno de los elementos del vector por la longitud de éste (LEGENDRE & LEGENDRE, 1983).

Uno de los métodos más comunes es el denominado *estandarización por totales* («standardization by totals»), que consiste en determinar la proporción que existe entre un elemento determinado del vector y el total del vector en cuestión, expresando los datos como proporciones. Se puede realizar en los dos sentidos de la matriz, bien por objetos (inventarios), bien por descriptores (especies). En el primer sentido, hay que tener cuidado con las modificaciones inducidas tanto por la medida de semejanza, que puede ser sensible a aspectos cualitativos de los datos, como a las generadas por los tamaños de los inventarios, sobre todo si éstos son muy diferentes. En cuanto a la estandarización por totales de las especies, se sobrepesan las raras y se infravaloran las comunes, por lo que sólo se recomienda utilizarla si las frecuencias de las especies en la tabla no varían excesivamente (VAN TONGEREN, 1987).

Otro método muy utilizado es la estandarización por el máximo valor, es decir, dividiendo cada una de las abundancias por la máxima. Esta estandarización realizada en un doble sentido –por inventarios y especies– fue muy utilizada por BRAY & CURTIS (1957).

Finalmente, la aplicación de algunos índices de semejanza suponen la estandarización implícita de los datos como es el caso de la distancia de la cuerda («chord distance») (ORLÓCI, 1967, 1978).

LA MEDIDA DE LA SEMEJANZA

Es la etapa base del tratamiento numérico de los datos, ya que examinando la matriz de semejanza se puede elucidar la estructura de los datos e incluso responder a alguna de las cuestiones propuestas. Es, por ello, la primera etapa a realizar en análisis numéricos, o quizá la «única» (LEGENDRE & LEGENDRE, 1983). Se efectúa midiendo la similitud entre los objetos, o bien entre las variables que los describen. Existe una gran variedad de índices de semejanza, aunque muchos de ellos son meras variantes de otros y tienen sus mismas propiedades (ORLÓCI, 1978).

Una matriz de datos puede estudiarse desde dos puntos de vista (CATTELL, 1952), bien observando las relaciones entre los objetos (inventarios), o bien entre los descriptores (especies). Si además incluimos el tiempo como una tercera dimensión a estudiar, se pueden definir seis tipos de análisis

(CATTELL, 1966), de los cuales en este apartado sólo se van a tratar las ya mencionadas, es decir: 1) la semejanza entre objetos y 2) entre descriptores. El primero incluye los *coeficientes de semejanza* y los *coeficientes de distancia*; el segundo, los *coeficientes de dependencia*.

a) Coeficientes de semejanza

Se utilizan para medir la asociación entre los objetos. Estas medidas no son nunca métricas —en oposición a las medidas de distancia— y, por tanto, las semejanzas no pueden ser utilizadas para posicionar los objetos unos frente a otros en un espacio métrico. Generalmente se utilizan como base de los análisis de agrupamiento, aunque éstos admiten indiferentemente un tipo u otro de medidas (LEGENDRE & LEGENDRE, 1983).

Fueron desarrollados para datos binarios de tipo presencia/ausencia. En este tipo de datos, uno de los principales factores a tener en cuenta es la presencia de los **dobles ceros**. Esta situación es muy frecuente en ecología, ya que muestras realizadas sobre un número de especies determinadas pueden darnos una matriz de datos con muchos dobles ceros, sobre todo en aquellas muestras más pobres. En estas últimas, y dependiendo del coeficiente utilizado, podemos obtener semejanzas muy altas, aunque las muestras originales sean muy diferentes entre sí en lo que a descriptores se refiere. Por ello, estos coeficientes de semejanza se dividen entre aquellos que excluyen de su cálculo los dobles ceros y los que no.

- * *Coeficientes binarios que incluyen los dobles ceros.* La semejanza se basa en la presencia/ausencia de determinadas características en las dos muestras. Algunos de ellos no tienen en cuenta la diferencia entre el doble cero y el doble uno, como el de SOKAL & MICHENER (1958). Otros le dan un peso mayor a las semejanzas que a las diferencias (SOKAL & SNEATH, 1963), o viceversa (ROGERS & TANIMOTO, 1960).
- * *Coeficientes binarios que no incluyen los dobles ceros.* Entre ellos tenemos los que dan igual peso a los descriptores considerados (JACCARD, 1908) o los que dan un peso doble o triple a la doble presencia (SORENSEN, 1948). Por el contrario, existen otros que dan mayor peso a las diferencias (SOKAL & SNEATH, 1963). Algunos son similares a los coeficientes que incluían los dobles ceros, pero excluyendo la parte correspondiente a éstos (OCHIAI, 1957). En estos últimos, las menores semejanzas dan valores a los índices positivos o negativos pero cercanos a cero, mientras que semejanzas muy grandes dan valores cercanos a la unidad, pudiendo llegar a infinito (CLIFFORD & STEPHENSON, 1975).

Posteriormente se desarrollaron coeficientes para datos cuantitativos e incluso para datos mixtos. De todos ellos podemos destacar:

- * *Coefficientes cuantitativos que incluyen los dobles ceros.* Todos los coeficientes binarios se pueden extrapolar aquí. De entre todos ellos podemos citar el índice de GOWER (1971), muy laborioso de calcular, en el que no se realiza la comparación entre dos objetos cuando falta uno de ellos, y sí se realizará en el caso en que ambos estén presentes o ausentes. De análogas características es el de ESTABROOK & ROGERS (1966), aunque la medida de la similitud es diferente.
- * *Coefficientes cuantitativos que excluyen los dobles ceros.* También se pueden incluir aquí los coeficientes binarios. De entre éstos podemos comentar el índice de JACCARD (1908) modificado, el cual se suele utilizar, sobre todo, cuando la abundancia de las especies está codificada bajo un pequeño número de clases y se quieren estudiar las diferencias entre ellas. En el resto de estos casos su empleo conlleva una pérdida parcial de la información contenida en los descriptores de abundancia.

Las medidas de semejanza típicamente cuantitativas se pueden dividir en dos grupos, según que usen datos brutos o datos normalizados o estandarizados. Entre las primeras, las hay que comparan la más pequeña de las abundancias respecto a la media de las dos muestras tomadas en consideración (CZEKANOWSKI, 1909, 1913); y entre las segundas tenemos variaciones de algunos índices binarios y cuantitativos que excluyen los dobles ceros.

b) Coeficientes de distancia

Comprenden una serie de índices que tienen en común dar un valor máximo a dos objetos completamente distintos y un valor mínimo a dos objetos idénticos, respecto a los descriptores utilizados en el análisis. Las medidas de distancia se dividen en dos grupos: las *métricas*, que obedecen a una serie de propiedades (ver LEGENDRE & LEGENDRE, 1983) y las *semimétricas*, denominadas por algunos autores como de *disimilitud*, en contraposición a las métricas, que son consideradas como las únicas medidas reales de distancia en sentido estricto. La distancia y la semejanza están íntimamente relacionadas a través de un conjunto de fórmulas matemáticas, siendo $D=1-S$ (D: distancia; S: semejanza) la más sencilla de entre todas las existentes.

Entre los coeficientes más utilizados destaca la *Distancia Euclídea*, basada en el teorema de Pitágoras, que define la relación existente entre dos puntos que se encuentran en un espacio euclídeo de n dimensiones. El problema de esta medida es que no tiene límite superior y aumenta a medida que se incrementa el número de descriptores. También depende de la escala de éstos,

ya que al cambiar la escala (por ejemplo, con transformaciones), las distancias se modifican. Este problema se resuelve empleando variables centradas o matrices de datos dimensionalmente homogéneas.

También puede ocurrir que dos muestras sin ninguna especie en común puedan tener una medida de distancia euclídea más pequeña que otro par de muestras con las mismas especies, debido a la presencia de muchos ceros. Este problema se puede resolver transformando o normalizando los datos originales. Para evitar los inconvenientes de la Distancia Euclídea se han propuesto algunas modificaciones, como es la *distancia de la cuerda* («Chord Distance»), propuesta por ORLÓCI (1967, 1978), en la que los vectores son normalizados y los dobles ceros excluidos.

c) Coeficientes de dependencia

El más utilizado es la *correlación de Pearson*, también denominado *correlación producto-momento (r)*, definida como la covarianza calculada sobre los datos centrados y reducidos (LEGENDRE & LEGENDRE, 1983). Ha sido utilizado como medida de semejanza en taxonomía para aquellas situaciones en las cuales los caracteres cualitativos de una muestra (animales o plantas) son más importantes que los cuantitativos (SOKAL & MICHENER, 1958). El uso de este índice debe ser restringido sobre todo cuando se estudian matrices de datos donde existe un 50% de ceros o más, es decir, con muchas especies raras o poco abundantes que originan unos valores de *r* muy pequeños, aunque las especies dominantes en las muestras sean las mismas (CLIFFORD & STEPHENSON, 1975).

ANÁLISIS DE GRADIENTE

1. Planteamiento

El valor de estas técnicas desde el punto de vista fitosociológico estriba fundamentalmente en la posibilidad de caracterizar ecológicamente tanto las plantas como las unidades de vegetación, aunque, como veremos, también pueden ser valiosas herramientas de clasificación.

Según TER BRAAK & PRENTICE (1988), recogido también por OKLAND (1990) y JONGMAN & AL. (1987), la expresión **ANÁLISIS DE GRADIENTE** hace referencia a técnicas que permiten la interpretación de las comunidades en términos de la respuesta de las especies a los gradientes ambientales, entendiendo esta respuesta en un sentido muy amplio. Estos autores señalan cuatro clases de técnicas de análisis de gradiente: 1. Análisis de gradiente directo o **regresión**; 2. Análisis de gradiente directo invertido o **calibración**;

3. Análisis de gradiente indirecto u **ordenación**, y 4. **Ordenación forzada** («constrained ordination»).

Al querer contrastar la abundancia de unas especies, en unas localidades concretas con variables ambientales medidas en esos mismos lugares, se puede recurrir a un amplio abanico de técnicas, y cada una de ellas se puede relacionar con alguna de las cuatro clases a las que hemos hecho referencia.

2. Regresión

Sin duda, la primera aproximación y, seguramente, la más sencilla se puede hacer mediante técnicas de regresión, las cuales describen la función que mejor ajusta los valores de abundancia de una especie (variable dependiente) frente a una variable ambiental (variable independiente) en relación con un modelo concreto; éste sería el caso de una *regresión simple*. Esta primera aproximación es lo que se conoce como *análisis de gradiente directo* (WHITTAKER, 1967, 1978). Una posibilidad todavía más simple consistiría en la comparación directa de los valores de la variable frente a los de abundancia del taxon considerado (WHITTAKER, 1967). En TER BRAAK & LOOMAN (1986), podemos encontrar un estudio exhaustivo de las técnicas de regresión en relación con diferentes modelos estadísticos.

Sin embargo, cuando el interés se centra en el estudio de las relaciones de las especies con un conjunto de variables que actúan sobre ellas, es necesario recurrir a otro tipo de técnicas.

De forma casi intuitiva se nos ocurriría utilizar técnicas basadas en *coeficientes de correlación*, como puede ser el *análisis de correlación canónica* (COR) (GAUCH & WENTWORTH, 1976; GITTINS, 1985) o técnicas relacionadas con éste, como el *análisis de variación canónica* («Canonical Variate Analysis», CVA) (PIELOU, 1984; GITTINS, 1985). Pero éstas presentan una doble limitación, por un lado, el hecho de asumir un modelo lineal de respuesta especies-variables, salvable mediante la introducción de modelos polinomiales y, por otra parte, el hecho de que el número de especies deba ser menor que el de localidades inventariadas, condición que raramente se cumple en estudios de vegetación. En este tipo de análisis, las especies y las variables ambientales son consideradas de forma simétrica (TER BRAAK, 1987); sin embargo, TSO (1981) desarrolló un modelo asimétrico en el que las especies son explicadas por las variables. Esta última variante tiene un comportamiento semejante al *análisis de redundancia* (RDA), que se comentará más adelante dentro de las técnicas de ordenación, pero del que se diferencia por la asunción diferente del tipo de errores.

Otra alternativa más elaborada estriba en la utilización de técnicas de *regresión múltiple*. Pese a las indudables ventajas teóricas de dichas técnicas,

que se han aplicado con éxito en algunas ocasiones (YARRANTON, 1970; AUSTIN, 1971), nunca han llegado a ser populares en estudios de vegetación. Según TER BRAAK (1987), las razones han sido: 1. Cada especie requiere un análisis individualizado, por si el número de especies no es muy reducido, el esfuerzo a invertir puede ser considerable; 2. Los datos de vegetación suelen ser cualitativos o, si son cuantitativos, poseen muchos ceros, y en ninguno de los dos casos es posible asumir una distribución normal del error; 3. Las relaciones entre especies y variables ambientales son generalmente no lineales, salvo que el segmento ambiental sea lo suficientemente corto; 4. Las variables ambientales están a menudo fuertemente correlacionadas, por lo que en muchas ocasiones es muy difícil separar su efecto conjunto (POTVIN & ROFF, 1993). Tal como afirman AUSTIN & AL. (1984), y TER BRAAK & LOOMAN (1986), algunos de estos problemas (2 y 3) se podrían solucionar mediante la utilización de modelos obtenidos por técnicas basadas en el *Modelado Lineal Generalizado* (GLM) (MCCULLAGH & NELDER, 1983), o sobre todo modelos desarrollados a partir de técnicas de *Modelado Aditivo Generalizado* (GAM), que son una extensión de las anteriores (YEE & MITCHELL, 1991). Éstas permiten considerar distribuciones de los errores distintas de la normal, como las que asuman las de *Máxima Probabilidad* («Maximum Likelihood»). Probablemente esta vía sea la más adecuada para realizar un análisis riguroso con el que determinar la forma de la curva de respuesta en relación a un número de factores considerados simultáneamente. En cualquier caso, estos problemas pueden ser solventados mediante la utilización de técnicas de «ordenación forzada», como veremos más adelante.

3. Calibración

Es el proceso inverso de la regresión, es decir el procedimiento que permite inferir valores para las variables ambientales a partir de la composición florística de cada localidad. Los tres métodos utilizados más habitualmente (TER BRAAK, 1987) son:

- 1) Análisis basado en las *funciones de respuesta* obtenidas previamente mediante regresiones.
- 2) *Regresión invertida*, en la que las especies son las variables independientes (TER BRAAK & PRENTICE, 1988).
- 3) Análisis basado en los valores indicadores de cada especie y posterior ponderación («*Weighted averaging*») –ver ELLEMBERG (1979) para su confirmación–. TER BRAAK & BARENDREGT (1986) detallan el valor de este último procedimiento y su capacidad de emular modelos no lineales, condición que limita notablemente a los procedimientos anteriores.

4. Ordenación

Bajo este término se agrupan varias técnicas multivariantes que presentan el denominador común de reducir la dimensionalidad de una matriz de datos, en nuestro caso florísticos, mediante la extracción de ejes y sin que intervengan las variables ambientales (GAUCH, 1982; PIELOU, 1984; TER BRAAK, 1986; OKLAND, 1990; FEOLI & ORLÓCI, 1991). La interpretación de estos ejes en términos ecológicos se puede realizar en una segunda fase; es el denominado **ANÁLISIS DE GRADIENTE INDIRECTO** (WHITTAKER, 1967).

Tal como señala OKLAND (1990), conceptualmente existen tres aproximaciones diferentes a las técnicas de ordenación:

En primer lugar, *geométricamente*, como métodos de escalado (PRENTICE, 1977). En este sentido, estas técnicas permiten sintetizar la información mediante la producción de un espacio de ordenación de pocas dimensiones donde la disimilitud entre los inventarios en la matriz inicial es preservada (MINCHIN, 1987; TER BRAAK & PRENTICE, 1988). La bondad del ajuste se mide mediante una función de estrés (PRENTICE, 1977). Dependiendo de la función de disimilitud utilizada, del tratamiento que se dé a estos valores, así como de dicha función de estrés, podemos dividir las técnicas de ordenación en *técnicas de escalado métrico* («metric scaling techniques»), también denominadas *Análisis de Coordenadas Principales* (PCoA), si se utilizan directamente los valores de disimilitud para calcular el estrés, y *técnicas de escalado no métrico* («non-metric scaling techniques»), si lo que se utiliza es el rango de dichos valores. Una tercera posibilidad es lo que se denominan *técnicas de escalado híbrido* («hybrid scaling techniques»), y que combina tanto el valor de disimilitud como su rango (FAITH & AL., 1987). Las funciones de similitud también pueden ser utilizadas en este sentido (PODANI, 1989a).

La segunda aproximación que se puede realizar es de tipo *estadístico*, considerando los ejes de ordenación como variables latentes o hipotéticas, obtenidas de manera que se optimice el ajuste de los datos de las especies respecto a un modelo estadístico (lineal, unimodal, etc.) de respuesta de la abundancia de las especies a un gradiente dado (TER BRAAK & PRENTICE, 1988; TER BRAAK, 1987). En este sentido, la técnica de ordenación busca la variable hipotética que mejor se ajusta de forma global a los datos de vegetación de donde partimos. Esta hipotética variable es el primer eje de ordenación. Tal como señalan TER BRAAK & PRENTICE (1988), las técnicas de escalado métrico pueden ser explicadas mediante ciclos de regresión y calibración respecto de un modelo estadístico concreto. La conocida técnica *Análisis de Componentes Principales* (PCA) puede ser interpretada como un caso especial de PCoA cuando la función utilizada es la Distancia Euclídea, ajusta los datos según un modelo lineal; mientras que el *Análisis de Correspondencias* (CA) se ajusta heurísticamente a un modelo unimodal (TER BRAAK, 1986).

La tercera aproximación es de tipo *algebraico*. Así, las técnicas de ordenación métrica pueden ser consideradas como técnicas capaces de obtener los vectores propios («eigenvectors») —que pueden ser tomados como variables latentes— y los correspondientes valores propios («eigenvalues») —que dan una idea de la bondad del ajuste entre los datos y las nuevas variables latentes (ORLÓCI, 1978; PIELOU, 1984).

Aunque no las vamos a desglosar aquí, han surgido nuevas técnicas de ordenación, entre las que destacan las relacionadas con el desarrollo de los conceptos esgrimidos por la «Fuzzy Set Theory» (ROBERTS, 1987) y que se basan en clasificaciones (FEOLI & ZUCCARELLO, 1986), o las aproximaciones más avanzadas de IHM & GROENEWOUD (1975), GOODALL & JONHSON (1982), etc.

5. Ordenaciones constreñidas

De forma sintética podemos afirmar que los análisis de regresión sirven para encontrar la variable ambiental que mejor explica la variación de la cobertura de las especies en un conjunto de datos; mientras que en las ordenaciones, los ejes son las variables hipotéticas que optimizan el ajuste de los datos florísticos a un modelo estadístico dado. Una tercera cuestión, intermedia entre las anteriores, que también puede ser resuelta, es qué combinación de variables ambientales puede explicar dichos datos de una forma más eficaz. Las *ordenaciones canónicas* u ordenaciones constreñidas (TER BRAAK, 1986, 1987) pueden resolver este problema (TER BRAAK & PRENTICE, 1988). Estas técnicas permiten conocer la combinación de variables que explica mejor la variación en la abundancia de las especies, optimizando el ajuste de la abundancia de especies a un conjunto dado de variables ambientales (TER BRAAK, 1986; TER BRAAK & PRENTICE, 1988).

En cualquier caso, una aproximación de este tipo se puede realizar recurriendo a análisis de regresión múltiple, como ya indicamos en el apartado correspondiente. La elección entre una regresión múltiple y una ordenación canónica depende de si consideramos ventajoso el hecho de analizar simultáneamente todas las especies o no. En el caso de que se utilicen todas, como ocurre en la ordenación canónica, se asume que existe un modelo de respuesta común, de manera que se extrae un reducido número de gradientes, denominados ejes —en este sentido, es una ordenación normal—, para todas las especies; evidentemente, en los análisis de regresión múltiple no ocurre así, en dicho caso es necesario construir un gradiente particular para cada especie. Por otro lado, la construcción de buenos modelos no lineales con técnicas de regresión no es sencilla, dado que la consideración de combinaciones lineales entre las variables complica mucho el proceso. Sin embargo, en este tipo de ordenaciones la posible no linealidad entra de la mano de los distintos mode-

los que subyacen en cada una de las técnicas de ordenación: PCA asume modelos lineales; CA, modelos unimodales, y el resto de las técnicas MDS («Metric Dimensional Scaling»), diferentes modelos según la función de partida. En definitiva, las técnicas canónicas requieren menos datos y son más fáciles de aplicar que las de regresión, además sus resultados dan una visión mucho más global. De todas formas, si lo que se pretende es una descripción más detallada y precisa de las relaciones entre las plantas y las variables ambientales, es necesario recurrir a técnicas de regresión múltiple (TER BRAAK, 1988).

Se han desarrollado las correspondientes versiones forzadas o constreñidas para cada una de las técnicas básicas de ordenación que hemos estudiado en el apartado anterior. La técnica de *ordenación forzada* correspondiente al muy conocido *Análisis de Componentes Principales* (PCA) se denomina *Análisis de Redundancia* (RDA) (RAO, 1964); la del *Análisis de Correspondencias* (CA) se denomina *Análisis Canónico de Correspondencias* (CCA) (TER BRAAK, 1986, 1987), y las correspondientes a las técnicas NMDS («Non Metric Dimensional Scaling») han sido desarrolladas por MEULMANN & HEISER (1984). Sin duda, hoy en día el CCA es la técnica más utilizada por los investigadores para resolver problemas relacionados con los análisis de gradiente directo (ZHANG & OXLEY, 1994). El auge del CCA y este tipo de técnicas ha sido impresionante; BIRKS & AUSTIN (1992) recogen 165 trabajos relacionados con el CCA y técnicas afines entre 1986 y 1991.

Además, pequeñas modificaciones de estas técnicas permiten obtener nueva información. En primer lugar, los *análisis canónicos híbridos* permiten conocer la variación florística que permanece después de que el efecto conjunto de las variables haya quedado ajustado (DAUDIN, 1980; YANAI, 1986; TER BRAAK, 1988). En general, el número máximo de ejes «forzados» es igual al de variables ambientales consideradas en el análisis, pero siempre es posible hacerlo sobre menos ejes (TER BRAAK, 1986; 1987), obteniendo a continuación ejes de ordenación parciales (TER BRAAK, 1988). Estos ejes representan la dirección principal de variación después de que la información atribuible a las variables consideradas ha sido extraída. La interpretación de estos ejes responde a la siguiente regla: en caso de que la información de estos ejes fuera mucho mayor que la acumulada y explicada por los anteriores, habría que pensar que las variables estudiadas no son eficaces a la hora de explicar la variabilidad florística de los datos (TER BRAAK & PRENTICE, 1988).

En segundo lugar, las *ordenaciones canónicas parciales* extraen la variabilidad introducida por variables ambientales que no presentan interés directo para el investigador –denominadas covariables o variables concomitantes– (DAVIES & TSO, 1982). De esta forma, los ejes de ordenación representarán únicamente el efecto atribuible a las variables de interés (TER BRAAK, 1988). BORCARD (1992) y OKLAND & EILERTSEN (1993) comentan el interés de

estas técnicas en estudios de vegetación, sobre todo a la hora de reducir la variabilidad debida a los patrones espaciales (TER BRAAK, 1987; LEGENDRE & FORTIN, 1989), aunque también sirve para eliminar ruidos relacionados con el muestreo (ESCUADERO, 1992).

Para la interpretación de los ejes extraídos se pueden utilizar los coeficientes canónicos o/y las correlaciones intragrupo («intra-set correlation»). Los coeficientes canónicos definen los ejes de ordenación como combinación lineal de las variables ambientales. Las correlaciones intragrupo son los correspondientes coeficientes entre las variables y los ejes de ordenación. Se puede comprobar la importancia relativa en la extracción de cada eje por parte de cada variable sopesando ambos índices. Los valores de los coeficientes canónicos sólo tienen un valor exploratorio, ya que no pueden ser probados (TER BRAAK, 1988; NANTEL & NEUMANN, 1992). Además, cuando un grupo de variables está muy correlacionado, los coeficientes de correlación se hacen muy inestables. Este problema se denomina multicolinealidad, y es semejante al que se presenta en el análisis de regresión múltiple (MONTGOMERY & PECK, 1982; TER BRAAK, 1990). En general, este problema se presenta con variables nominales («dummy variables», MONTGOMERY & PECK, 1982). Por ello, siempre resulta recomendable estudiar con antelación el conjunto de variables ambientales para retirar las que presenten correlaciones muy significativas. Posteriormente, estas variables que presentan problemas de colinealidad pueden ser introducidas de forma pasiva en los diagramas de ordenación, de manera que también puedan intervenir en la explicación de los diagramas (TER BRAAK, 1988; TONTERI & AL., 1990; FRANKLIN & MERLIN, 1992).

Por último, queremos señalar que estas técnicas pueden dar diagramas de difícil interpretación cuando la correspondencia entre ambos conjuntos de datos, florísticos y ambientales, es muy escasa (OKSANEN & HUTTUNEN, 1989). Este problema se puede solucionar gracias a una correcta elección de las especies.

6. Premisas para trabajar con diagramas de ordenación

En este apartado vamos a indicar algunas de las consideraciones que se deben tener en cuenta al utilizar estas técnicas numéricas. Una posible rutina de trabajo podría ser la indicada por TER BRAAK (1988), de manera que en primer lugar se realiza una inspección de los datos mediante un CA, para que se pueda ver hasta qué punto nuestros datos son unimodales o si, por el contrario, se ajustan a un modelo lineal; como ya hemos indicado, este punto es crucial a la hora de decidimos por una técnica de ordenación u otra. Para ello, HILL & GAUCH (1980) propusieron el uso de la desviación estándar (SD, «standard units») de las curvas de respuesta de las especies respecto al gra-

diente determinado por dicho eje. Si esta longitud es superior a 3 unidades SD, es recomendable utilizar métodos no lineales (TER BRAAK & PRENTICE, 1988). Inversamente, si la longitud es inferior o igual a 2 SD, será necesario recurrir a métodos lineales. Un ejemplo de la determinación de la longitud del gradiente mediante DCA lo podemos encontrar en TONTERI & AL. (1990).

Recientemente, EILERTSEN & AL. (1990) han estudiado los efectos de las manipulaciones habituales sobre los datos, como retirar especies de baja presencia («downweighting») y modificar el peso de las especies características, comprobando que se producen fuertes modificaciones de la longitud del gradiente, lo cual cuestiona en cierta medida la validez de este estimador. Por otro lado, JOHN & DALE (1990) han realizado con éxito algunas aplicaciones de CCA con gradientes muy cortos; ellos recurrieron a una simulación de Montecarlo para ver si el primer eje extraído era significativo. TER BRAAK (1986) ya había insinuado que, pese a todas las restricciones teóricas, el CCA podría funcionar bien con gradientes cortos, aunque es evidente que si los vectores propios son muy cortos, los valores óptimos de muchas de las especies caerían fuera de la región inventariada y el comportamiento frente a las variables empleadas puede ser considerado prácticamente lineal, por lo que en cierta medida ambas aproximaciones se pueden hacer coincidentes. Sin duda, las ventajas de los métodos lineales, tanto de PCA como de RDA, en este contexto, estriba en que sus «biplots» son mucho más informativos que los de (D)CA y (D)CCA, pero evidentemente esto se torna en inconveniente cuando los datos son difíciles de aproximar a modelos lineales. También queremos señalar que la tolerancia de las especies es cercana a 1 SD, por lo que cada curva de respuesta (modelo gaussiano) puede alcanzar aproximadamente 4 SD (TER BRAAK, 1987), aunque MINCHIN (1987) habla de 6 SD. Por lo tanto, es razonable pensar que muestras que presentan más de esta longitud no tienen especies en común. Por supuesto, es necesario asumir que todas las especies presentan un comportamiento gaussiano frente a los gradientes extraídos. En cualquier caso podemos encontrar fuertes reticencias, de manera que algunos autores consideran que éste no es el único modelo de respuesta especies-variables (MINCHIN, 1989; AUSTIN & SMITH, 1989), argumento de nuevo utilizado por estos autores cuando cuestionan la utilidad de las técnicas de ordenación canónica, muy eficaces a la hora de aproximar los datos a dicho modelo, aunque sea de una forma heurística.

Respecto a las técnicas que asumen modelos lineales, MINCHIN (1987) –también recogido por OKLAND (1990)– llega a afirmar que no existe ninguna justificación para utilizar un PCA con el propósito de realizar un análisis de gradiente indirecto. En cualquier caso, y dado que el espacio resuelto por el PCA es de tipo euclídeo, su interpretabilidad, tanto en términos fitosociológicos como ecológicos, puede ser mayor al haber una adecuación entre el

espacio original y los nuevos espacios generados (FEOLI & ORLÓCI, 1991); por ello, algunos autores (MAZZOLENI & AL., 1992; CASADO & AL., 1989) consideran que su aplicación puede ser muy ventajosa. En el extremo opuesto podemos encontrar autores como FAITH & AL. (1987), que estiman que el espacio euclídeo permite medidas muy pobres de las distancias ecológicas y por consiguiente desaconseja el uso de un PCA.

Por otro lado estas técnicas de ordenación también pueden ser utilizadas para detectar inventarios desviantes («outliers»). Para localizar este tipo de inventarios desviantes generalmente se han utilizado técnicas de clasificación (GAUCH, 1982), sin embargo, coincidimos con WILDI (1989) en que es mucho más sencillo mediante el uso de técnicas de ordenación. El CA es muy sensible a los inventarios pobres en especies, situando estos inventarios en el extremo del primer eje (GAUCH, 1982). En este sentido, es interesante comentar que un comportamiento muy desviante en estas ordenaciones no tiene por qué reflejarse en las ordenaciones canónicas, debido a que las variables ambientales que caracterizan el inventario no tienen por qué ser desviantes, por lo que la detección de estos inventarios debe hacerse con técnicas de ordenación sencillas.

Para finalizar este apartado, queremos hacer referencia a los valores porcentuales de la varianza total que explica cada eje. Aunque tradicionalmente estos valores han sido muy utilizados para explicar la cantidad de información acumulada por los ejes extraídos, hoy en día tal interpretación está en desuso, ya que la información que pueden aportar es muy críptica. TER BRAAK (1986, 1987) afirma que ese tipo de información sólo se puede obtener a partir del tamaño de los vectores directores, que varían de 0 a 1 en las técnicas derivadas de CA (JONGMAN & AL., 1987), o mejor aún, utilizando la desviación estándar de dichos ejes con un test de significación; además, valores porcentuales muy bajos pueden dar diagramas muy informativos, y viceversa (GAUCH, 1982). De todas formas, hay que tener en cuenta que no es posible obtener valores del 100% de varianza explicada, dado que siempre hay un ruido de fondo debido a muy diferentes factores (GAUCH, 1982).

7. Interpretación de los diagramas de ordenación

Las ordenaciones originadas por cada una de las técnicas que hemos comentado no se pueden interpretar siguiendo un único esquema. En el caso del PCA y técnicas relacionadas metodológicamente con él, tanto los inventarios como las especies pueden ser representados en un mismo espacio (generalmente se utilizan espacios bidimensionales o planos), de manera que si unimos la posición de una especie dada y el origen (0,0), tenemos lo que se ha denominado «biplot» (representación simultánea de especies-localidades; GABRIEL, 1971), en el que la dirección y módulo de dichos vectores nos da

una idea de la dirección de la máxima variación de la abundancia de una especie dada, y la longitud es proporcional a la intensidad del cambio (TER BRAAK & PRENTICE, 1988). De forma sintética podemos afirmar que las especies más alejadas del centro son las más importantes a la hora de diferenciar las localidades (inventarios); en cualquier caso, existen reglas de carácter cuantitativo para interpretar los diagramas de ordenación de un PCA (TER BRAAK, 1983).

En el caso de los diagramas de ordenación obtenidos con un CA o técnicas relacionadas con ésta, las especies y las localidades son representadas por puntos, siendo, en el caso de las localidades, el centro de gravedad de las especies que allí aparecen y el de las especies una aproximación a su óptimo (modelo de respuesta unimodal). Las especies más alejadas del origen no podemos saber si su óptimo cae dentro o no de nuestros diagramas, de manera que en muchos sentidos estas especies pueden llevarnos a interpretaciones erróneas, ya que no podemos estar seguros de si su posición es consecuencia de su preferencia por condiciones ecológicas extremas dentro del tramo de gradiente analizado o si simplemente obedece a su baja representatividad.

En TER BRAAK (1986; 1987; 1988) se explican detalladamente los procedimientos de obtención e interpretación de los diagramas producidos en un CCA. De forma sintética, las especies y localidades se representan por puntos; las variables cuantitativas, por vectores (interpretación semejante a la de los biplots de un PCA), y las variables nominales, por puntos que representan sus centroides.

La interpretación de estos diagramas puede ser implementada mediante la utilización de nuevas técnicas, apoyadas en el desarrollo del *software* necesario. En este sentido destacan las técnicas de interpolación espacial (HAUSER & MUCINA, 1991; TER BRAAK, 1987; BURGAZ & AL., 1993); así como las basadas en la utilización de elipses de máxima probabilidad (FEOLI & LAGONEGRO, 1979).

8. El efecto de arco

A la hora de interpretar los diagramas, uno de los problemas más importantes que nos podemos encontrar es consecuencia del llamado *efecto de arco* (HILL & GAUCH, 1980). Si los datos de las especies se ajustan a un modelo unidimensional, el segundo eje extraído generalmente presenta una relación parabólica con el primer eje; esto ocurre en la práctica cuando el gradiente del primer eje es mucho mayor que el del siguiente. En definitiva, los valores de las especies para el segundo eje van a formar un arco; característica ésta que nos permite identificar el problema sobre los diagramas de ordenación, ya que las especies de comportamiento más antagónico van a situarse muy cerca en el plano de ordenación. HILL & GAUCH (1980) desarrollaron el

DCA como una modificación heurística del CA, diseñada para solventar este problema. Se trata de un reescalado no lineal de los ejes, de manera que se intenta igualar la varianza interna en todos los puntos a lo largo del eje de ordenación; los ejes son divididos en segmentos y se los expande o contrae según sea necesario. Este método de «detrending» ha sido criticado, por lo menos para ciertas situaciones (OKSANEN, 1988; KENKEL & ORLÓCI, 1986; MINCHIN, 1987; PIELOU, 1984), y ha sido muy poco utilizado en aplicaciones fuera del ámbito de la ecología (GREENACRE, 1984). TER BRAAK & PRENTICE (1988) y TER BRAAK (1988) consideran mucho más ventajoso hacer un «detrending» por el método polinomial, método que también fue esbozado por HILL & GAUCH (1980). En cualquier caso, también podemos encontrar autores reticentes a la hora de aplicar estas técnicas. KNOX (1989) afirma que no mejora sustancialmente a la técnica primitiva. Esta situación es la que hace comentar a PIELOU (1984) y WARTENBERG & AL. (1987) que quizá sea mucho más adecuado recurrir exclusivamente a CA y detectar los auténticos gradientes, pese a los inconvenientes que se puedan presentar. Aunque no hemos hablado de ello, esta problemática es extensiva a todo tipo de ordenaciones forzadas, aunque, según PALMER (1993), el CCA es inmune a la mayoría de los problemas que presenta el DCA.

CLASIFICACIONES NUMÉRICAS

Bajo esta denominación se reúnen un conjunto de técnicas denominadas de «agrupamiento» capaces de generar clases diferentes unas de otras, cuyos elementos son comunes dentro de cada una de ellas (DALE, 1988). Estas técnicas permiten establecer una división de los conjuntos iniciales de datos de una forma más objetiva y menos intuitiva. En cualquier caso, la elevada cantidad de decisiones que el investigador debe tomar (tipo de algoritmo a utilizar, función de semejanza, transformaciones de los datos, etc.) puede cuestionar esta afirmación (KOVAR & LEPS, 1986; GOODALL & FEOLI, 1988).

En estudios de vegetación y fundamentalmente en el contexto fitosociológico estas técnicas han sido utilizadas con profusión (MUCINA & VAN DER MAAREL, 1989; KENT & BALLARD, 1988), pese a las reticencias manifestadas por algunos autores, que cuestionan la concepción determinista y no-continua implícita en la mayor parte de las mismas (MINCHIN, 1987; AUSTIN & SMITH, 1989).

En este capítulo vamos a presentar esquemáticamente las técnicas de agrupamiento y comentaremos brevemente algunos conceptos más avanzados, como son la determinación del nivel óptimo de partición, la generación de clasificaciones consensuadas, la comparación de dendrogramas y particiones, etc. (Ver ANDERBERG, 1973; CLIFFORD & STEPHENSON, 1975; PIELOU,

1984; WESTHOFF & VAN DER MAAREL, 1978; LEGENDRE & LEGENDRE, 1984; ORLÓCI, 1978; SNEATH & SOKAL, 1973; MARDIA & AL., 1979; EVERITT, 1980).

1. Análisis numérico de agrupamiento

La diversidad de técnicas de agrupamiento existentes en la literatura puede resumirse atendiendo a la estrategia clasificatoria que desarrollan en el proceso. Siguiendo el criterio de WILLIAMS (1971), que usa una clave dicotómica para diferenciar las distintas técnicas, la primera disyunción se establece entre *técnicas de solapamiento* y de *no solapamiento*. En las primeras, una entidad puede aparecer en más de un grupo, razón por la cual han sido poco utilizadas en ecología (YARRANTON & AL., 1972), aunque a partir del desarrollo de las técnicas relacionadas con la *Teoría de los Conjuntos Difusos* («Fuzzy Set Theory»), su uso se está incrementando notablemente (ROBERTS, 1986; MARSILI-LIBELLI, 1989).

A su vez, las técnicas de no solapamiento se dividen en *extrínsecas* o *intrínsecas*. Las primeras utilizan datos bióticos y abióticos, tratando de relacionar ambos, mientras que las clasificaciones intrínsecas han sido empleadas para formar grupos únicamente a partir de sus atributos. Estas últimas, a su vez, han sido divididas en *jerárquicas* y *no jerárquicas*. Según WILLIAMS (1971), las jerárquicas siempre buscan una ruta entre la población entera y el conjunto de grupos individuales de que ésta se compone, optimizando para ello la relación existente entre dos grupos cercanos. Por el contrario, en las estrategias no jerárquicas se optimiza la estructura del grupo individual, que es todo lo homogénea que sea posible, no existiendo un camino definido entre los grupos y la población completa. Su valor estriba en que son capaces de producir clasificaciones óptimas como consecuencia de la compactación individual de los grupos (CORMACK, 1971) y de la separación que consiguen entre los mismos (VAN TONGEREN, 1987). Además, la multidimensionalidad inicial de los datos puede ser poco compatible con la dimensión única que se establece en una clasificación jerárquica (WHITTAKER, 1962; GAUCH & WHITTAKER, 1981).

Los métodos jerárquicos han sido muy utilizados por los taxónomos, de una parte por su paralelismo con las teorías evolutivas, y de otra porque son capaces de organizar el conocimiento adquirido (CLIFFORD & STEPHENSON, 1975) de manera que la jerarquización de los grupos permite establecer y conocer las relaciones que existen entre ellos (GAUCH, 1982).

Finalmente, los métodos jerárquicos pueden dividirse en *aglomerativos* y *divisivos*. Los métodos divisivos pueden ser *monotéticos* o *politéticos*, mientras que los aglomerativos únicamente son *politéticos*. Los métodos divisivos toman el conjunto de los datos como una entidad única, dividiéndolo con

posterioridad; cada uno de los fragmentos obtenidos sufre el mismo proceso. Los métodos aglomerativos comienzan a unir los objetos de forma individual, hasta llegar a tenerlos todos unidos. Tradicionalmente, las técnicas divisivas han sido consideradas más eficaces que las aglomerativas, por lo menos cuando se utilizaban con datos ecológicos (HILL, 1979; GAUCH & WHITTAKER, 1981); sin embargo, algunos autores han puesto en duda, recientemente, la eficacia de algunas de las técnicas divisivas de uso más extendido (TWINSPAN) (VAN GROENEWOUD, 1992; BELBIN & McDONALD, 1993), mientras que otros resaltan la capacidad clasificatoria de las técnicas aglomerativas (PODANI, 1989A; PIELOU, 1984).

Las técnicas monotéticas utilizan como criterio clasificatorio la presencia o ausencia de una especie, mientras que las politéticas utilizan criterios más numerosos. En general, se suelen utilizar más profusamente estas últimas, debido a que la presencia-ausencia azarosa de algunas especies puede introducir importantes desviaciones (GOODALL, 1978), sobre todo si el número de especies o de inventarios es pequeño (HILL, 1977).

2. Métodos jerárquicos aglomerativos

El número de técnicas incluíbles bajo este epígrafe es enorme, pero prácticamente la totalidad se deriva de la solución combinatoria propuesta por LANCE & WILLIAMS (1966, 1967), tal como expone PODANI (1989a), quien además agrupa estas técnicas entre las que buscan la distancia mínima entre grupos («clusters») (*d-SAHN*) y las que optimizan la homogeneidad intra-grupo (*h-SAHN*; «Sequential», «Agglomerative», «Hierarchical», «Non overlapping»). Entre los primeros, podemos destacar:

2.1. *Enlace de grupos por el vecino más cercano* («single linkage clustering»)

Toma como distancia entre dos grupos aquella existente entre dos pares de puntos, uno de cada grupo, cuya distancia sea la más pequeña de todas las posibles.

No es frecuente su uso en ecología, debido al problema de «encadenamiento» que suele presentar, es decir, la tendencia a generar un crecimiento desde la base a partir de un pequeño grupo (de dos inventarios, por ejemplo), al cual se le une una serie de objetos, uno después de otro, a modo de cadena. En un dendrograma de estas características es muy difícil para el usuario definir grupos (PIELOU, 1984). Sin embargo, debido a este efecto de encadenamiento es muy utilizado para detectar objetos desviantes («outliers»), ya que serán los que presenten mayores distancias frente al resto de los objetos y, por tanto, se unirán a la parte final del dendrograma, donde podrán ser detectados (WILDI, 1989).

2.2. Enlace de grupos por el vecino más lejano («complete linkage clustering»)

La distancia entre dos grupos se basa en la que existe entre los dos puntos más alejados de cada uno de los grupos que se van a fusionar.

Tiene la ventaja de que los grupos obtenidos tienen un tamaño parecido, aunque su principal inconveniente estriba en que dos grupos vecinos que aparecen en el dendrograma separados a una gran distancia pueden ser muy similares. Por ello, es difícil detectar elementos desviantes. Según LANCE & WILLIAMS (1967), el proceso conlleva un alejamiento entre sí de algunos grupos y un acercamiento de otros, produciendo una sensación muy marcada de dilatación espacial en el conjunto de datos.

2.3. Agrupamiento por la unión media («average linkage clustering»)

La distancia entre dos grupos se puede definir de cuatro formas diferentes, que dan lugar a cuatro tipos de agrupamiento: la *distancia media con o sin carga* («unweighted/weighted average distance», UPGMA/WPGMA), la *distancia entre centroides* («centroid distances», UPGMC) y la *distancia entre medianas* («median distance»- WPGMC). La distancia entre centroides puede considerarse sin carga, a diferencia de la distancia media. Este término de **carga** se refiere al peso con el que cada objeto contribuye: en los análisis sin carga, cada objeto es vinculado al grupo con igual peso, aunque el peso de un grupo es tratado de forma proporcional al número de elementos que contiene. Resultado de todo ello es que el centroide de dos grupos no es el punto medio entre los centroides, sino que se acerca más al grupo que contenga mayor número de miembros (PIELOU, 1984).

Entre las técnicas de las denominadas técnicas h-SAHN, podemos destacar la siguiente:

2.4. Agrupamiento por varianza mínima («minimum variance clustering»)

Su formulación algebraica fue dada por BURR (1970). Se basa en la dispersión dentro de los grupos, que no es más que la suma de los cuadrados de la distancia entre cada punto y el centroide del grupo.

Es considerado, junto a los recogidos en el apartado anterior, como el que produce la dicotomía de forma más eficaz de entre todos los métodos aglomerativos y el que mejor clasifica los inventarios de vegetación que varían de forma continua (FEOLI-CHIAPELLA & FEOLI, 1977; WILDI, 1989).

3. Métodos divisivos

Cuentan con menos inconvenientes que los métodos aglomerativos, ya que en éstos, al empezar a unir los objetos desde abajo, se corre el riesgo de

que en los primeros pasos del proceso de agrupamiento se realicen **fusiones erróneas** con influencia negativa en las posteriores. A pesar de ello, algunos autores, como PIELOU (1984), consideran estas técnicas menos eficaces que las técnicas aglomerativas descritas anteriormente.

Las aproximaciones más ampliamente utilizadas comienzan realizando una ordenación de los datos, dividiendo después los espacios resultantes y generando las correspondientes particiones. Este procedimiento es conocido como **ordenación en espacio divisivo**. Existe un gran número de métodos para realizar la ordenación inicial, así como para situar los límites de las particiones (GAUCH, 1982).

La ventaja de los métodos divisivos consiste en que podemos parar el proceso de división en cualquier momento y no necesitamos continuarlo hasta tener los objetos aislados completamente. Sin embargo, esto conlleva un inconveniente, y es que se necesita una regla para decidir en qué momento se detiene el proceso, lo cual no deja de ser arbitrario. Sin embargo, también se ha trabajado para desarrollar algunos criterios con los que decidir cuándo la división ha llegado lo suficientemente lejos (PIELOU, 1977; DALE, 1988).

A continuación, veremos algunos de los métodos más utilizados, haciendo especial hincapié en el tercero de ellos, por su amplia utilización.

3.1. MÉTODO DE LA MÍNIMA EXTENSIÓN («Minimum spanning tree»)

Se basa en el método aglomerativo de unión por el vecino más cercano («single linkage clustering»). Debe cumplir dos reglas: 1) cada par de puntos se une, exclusivamente, por una línea; y 2) ninguna de las trayectorias puede formar un conjunto cerrado. Previamente a la aplicación del método clasificativo, se realiza una ordenación.

Tiene una gran utilidad cuando se realizan ordenaciones con reducción de la dimensionalidad, ya que dos objetos que en un espacio de s -dimensiones (tantas como especies) aparecen alejados, al reducir la dimensión a dos pueden aparecer situados muy cerca, de ahí que la sobreimpresión de un MST en este tipo de diagramas sea de gran ayuda para detectar semejanzas no aparentes o diferencias enmascaradas.

3.2. PARTICIÓN DE UNA ORDENACIÓN POR PCA

- a) El *Método partitivo de Lefkovitch* consiste en realizar un PCA centrado. Esto implica que el origen de los ejes está en el centroide de la nube de datos. La primera división se hará, por lo tanto, rompiendo por el primer eje de la ordenación, y la segunda por el segundo eje, así hasta finalizar el método. Finalmente, se obtiene un dendrograma de si-

milares características a los aglomerativos. En cualquier caso, y dadas las características intrínsecas de los biplots de un PCA, puede que dicho problema no resulte tan severo.

Su principal desventaja es que puede inducir la dispersión de un grupo que debería quedar unido, ya que puede ser descolocado por alguno de los ejes principales. LEFKOVITCH (1976) sugirió varios métodos para realizar la verificación del método partitivo.

- b) El *Método partitivo de Noy-Meir* (1973) es parecido al anterior, pero a diferencia de éste, el punto de corte se escoge en cada rotura. Esto se realiza haciendo que la varianza dentro de los grupos de las coordenadas del eje principal para un grupo de puntos determinado sea lo más pequeña posible. Evita los defectos del método de Lefkovitch. Sin embargo, tiene muchas similitudes con el agrupamiento de mínima varianza, anteriormente comentado, aunque no es la inversa de dicho método, ya que examinar todas las posibles divisiones de los puntos en dos grupos, con objeto de encontrar con cuál se obtiene la mínima suma de la varianza dentro de los grupos, sería impracticable.

3.3. PARTICIÓN DE ORDENACIONES CA Y DCA

Uno de los métodos clasificatorios (politético, jerárquico y divisivo) más comúnmente empleado (GANDULLO & AL., 1991; RETUERTO & CARBALLLEIRA, 1991; FRANKLIN & MERLIN, 1992) es el TWINSPAN (*Two-way Indicator SPecies ANalysis*-HILL, 1979; HILL & AL., 1975). El TWINSPAN fue diseñado en principio para el tratamiento de matrices de datos florísticos, clasificando simultáneamente parcelas y especies. En este método, los valores de las parcelas y los valores de las especies se estiman por cálculo de la *media recíproca* (R.A.O., «Reciprocal Averaging Ordenation»). Es un método ideado a partir del denominado *análisis de especies indicadoras* (ISA-HILL & AL., 1975), en el que se seleccionan cinco especies **indicadoras** con las que, por un lado, se elabora una clave sencilla que puede ser usada en el campo para asignar otras parcelas a sus clases apropiadas; y que además, y como aspecto más interesante, reproducen fielmente la ordenación completa inicial con todas las especies. Con este método, un grupo de parcelas se puede singularizar mediante una combinación de especies características o diferenciales (OKLAND, 1990). El concepto cualitativo de especies diferenciales es transferido a datos cuantitativos mediante el desarrollo del concepto de **seudoespecies** (HILL & AL., 1975). Éstas se definen por una abundancia mínima de una especie. A dicho mínimo se le denomina **nivel de corte**. Así pues, el comportamiento medio de una especie responde al gradiente mantenido por las seudoespecies.

En este método, la clasificación de las parcelas emplea los primeros ejes de una ordenación de análisis de correspondencia (CA) de las parcelas; primeramente lleva a cabo una clasificación en dos grupos, aproximadamente a partir de la mitad de la ordenación, que se refina por una identificación primera de **especies indicadoras** para cada polo del eje, con los que se obtiene un primer valor preferencial para cada especie. Para cada una de las parcelas se calcula el peso medio de los valores preferenciales, y éste sirve como base para una reordenación de las parcelas (**ordenación refinada**) y para obtener una dicotomía provisional. Repitiendo el proceso se consigue una **ordenación indicadora** de los dos primeros grupos. Cuando se fija la dicotomía, estos grupos se dan por definitivos y sobre cada uno de ellos se vuelve a repetir todo el proceso (HILL, 1979). Los grupos, en todos los niveles jerárquicos, son ordenados a lo largo del gradiente mayor de todas las especies, puesto que las ordenaciones realizadas en cada nivel dependerán de niveles superiores (OKLAND, 1990).

La clasificación de las especies por el TWINSPAN se lleva a cabo de la misma manera que la clasificación de las parcelas, y basándose en ella.

Esta metodología presenta una gran versatilidad, ya que permite, siguiendo los criterios de DIGBY & KEMPTON (1987), obtener una clasificación exhaustiva e imparcial, al menos inicialmente. Es posible asignar el mismo peso a los diferentes niveles de seudoespecies y a los valores de los potenciales indicadores para los niveles de corte. Sin embargo, también permite fijar diferentes pesos a los niveles de seudoespecies y a los valores de los potenciales indicadores, sin modificar notablemente los resultados obtenidos previamente, pero consiguiendo cambios que mejoran los resultados.

Finalmente, quisieramos comentar las técnicas de clasificación derivadas de la denominada *Teoría de los conjuntos difusos* («Fuzzy Set Theory», ZIMMERMAN, 1984), las cuales están adquiriendo gran importancia en los estudios de vegetación (ROBERTS, 1986; FEOLI & ZUCCARELLO, 1986, 1988; MARSILI-LIBELLI, 1989; ESCUDERO & PAJARÓN, 1994). Se trata de técnicas que permiten el solapamiento de los grupos de manera que los elementos pueden pertenecer en mayor o menor medida a los diferentes grupos (DE PATTA PILLAR & ORLÓCI, 1991), y por consiguiente, el establecimiento de los límites entre grupos no es tan severo como en las particiones que hemos estudiado con anterioridad.

La elección entre el abanico de técnicas de las que el investigador dispone puede ser bastante complicado. Tal como señalan PIELOU (1984) o GOODALL (1973), la elección debe basarse en las características intrínsecas de cada una de ellas, o de las combinaciones de algoritmos, funciones de semejanza y posibles transformaciones de los datos, aunque quizá la mejor elección se relaciona con la interpretabilidad de los resultados (MUCINA & VAN DER MAAREL,

1989; MAZZOLENI & AL., 1992). En cualquier caso, hoy en día todavía son necesarios estudios sobre medidas y métodos (NOY-MEIR & VAN DER MAAREL, 1987). Existen técnicas multivariantes de uso generalizado para comparar tanto dendrogramas como particiones (PODANI & DICKINSON, 1984; PODANI, 1989C; 1989D), cuyos resultados pueden ser utilizados para establecer las relaciones entre ellos (PAULE & GÖMÖRY, 1987; PODANI, 1989B). Igualmente, es posible generar soluciones consensuadas a partir de diferentes dendrogramas y/o particiones (ROHLF, 1982; NEUMANN & NORTON, 1986; PODANI, 1989D; ESCUDERO ET AL., 1994). En cuanto a la evaluación de las clasificaciones y la determinación de las particiones óptimas, existe un gran número de trabajos en la literatura (DALE, 1988; VAN DER MAAREL, & AL., 1978; POPMA & AL., 1983; ESCUDERO & PAJARÓN, 1994).

AGRADECIMIENTOS

El trabajo se ha beneficiado notablemente por los comentarios y consejos realizados por el Dr. Fernández-González tras revisar las primeras versiones del manuscrito.

REFERENCIAS BIBLIOGRÁFICAS

- ANDERBERG, M. R. 1973. *Cluster analysis applications*. Academic Press. London.
- AUSTIN, M.P. 1971. Role of regression analysis in plant ecology. *Proc. Ecol. Soc. Austr.*, 6: 63-75.
- AUSTIN, M. P., CUNNINGHAM, R. B. & FLEMING, P. M. 1984. New approach to direct gradient analysis using environmental scalars and statistical curve-fitting procedures. *Vegetatio*, 55: 11-27.
- AUSTIN, M. P. & SMITH, T. M. 1989. A new model for the continuum concept. *Vegetatio*, 83: 35-47.
- BELBIN, L. & McDONALD, C. 1993. Comparing three classification strategies for use in ecology. *J. Veg. Sci.*, 4(3): 341-348.
- BIRKS, H. J. B. & AUSTIN, H. A. 1992. An annotated bibliography of CCA and related constrained ordination methods. 1986-1991. University of Bergen. Norway.
- BORCARD, D. 1992. Partialling out the spatial component of ecological variation. *Ecology*, 73(3): 1045-1055.
- BRAUN BLANQUET, J. 1964. *Pflanzensoziologie. Grundzüge der vegetationskunde*. Springer-Verlag. Wien.
- BRAY, J. R. & CURTIS, J. T. 1957. An ordination of the upland forest communities of southern Wisconsin. *Ecol. Monogr.*, 27: 325-349.
- BUENO, A. & FERNÁNDEZ-PRIETO, J. A. 1991. Acebuchales y lauredales de la costa cantábrica. *Lazaroa*, 12: 273-201.
- BURGAZ, A. R., FUERTES, E. & ESCUDERO, A. 1993. Ecology of cryptogamic epiphytes and the communities in deciduous forests in mediterranean Spain. *Vegetatio*, 112: 73-86.

- CAMPBELL, B. M. 1978. Similarity coefficients for classifying relevés. *Vegetatio*, 37: 101-109.
- CASADO, M. A., ABBATE, G., BLASI, C. & PINEDA, F. D. 1989. Pattern diversity analysis of clearing in a *Quercus cerris* wood. *Vegetatio*, 79: 143-149.
- CATTELL, R. B. 1952. *Factor analysis - an introduction and manual for the psychologist and social scientist*. Harper. New York.
- CATTELL, R. B. 1966. The data box: its ordering of total resources in terms of possible relational systems. In Cattell, R. B. (ed.) *Handbook of multivariate experimental psychology*: 67-128. Rand McNally & Co. Chicago.
- CLIFFORD, H. T. & STEPHENSON, W. 1975. *An introduction to numerical classification*. Academic Press. London.
- CORMACK, R. M. 1971. A review of classification. *J. Roy. Statist. Soc., Ser. A*, 134: 321-367.
- CZEKANOWSKI, J. 1909. Zur differentialdiagnose der neandertalgruppe. *Korrespondenzblatt deutsch. Ges. Anthropol. Ethnol. Urgesch.*, 40: 44-47.
- CZEKANOWSKI, J. 1913. Zarys metod statystycznych w zastosowaniu do antropologii. *Travaux de la société des Sciences de Varsovie. III. Classe des sciences mathématiques et naturelles, n° 5*.
- DALE, M. B. 1988. Knowing when to stop: cluster concept-concept cluster. *Coenoses*, 1: 11-31.
- DALE, M. B. & WEBB, L. J. 1975. Numerical methods for the establishment of associations. *Vegetatio*, 30: 77-87.
- DAUDIN, J. J. 1980. Partial association measures and an application to qualitative regression. *Biometrika*, 67: 581-590.
- DAVIES, P. T. AND TSO, M. K. S. 1982. Procedures for reduced-rank regression. *Appl. Statist.*, 31: 244-255.
- DE PATJA PILLAR, V. & ORLÓCI, L. 1991. Fuzzy components in community level comparisons. In Feoli, E. & Orlóci, L. (eds.) *Computer assisted vegetation analysis*: 87-93. Kluwer Academic Publishers. Dordrecht.
- DÍAZ PINEDA, F. & GONZÁLEZ BERNÁLDEZ, F. 1978. Descripción automática de la vegetación II. Análisis factorial de datos cualitativos obtenidos en muestreos estratificados, *Anal. Edaf. Agr.*, 37: 539-549.
- DÍAZ PINEDA, F., GONZÁLEZ BERNÁLDEZ, F. & DE NICOLÁS, P. 1979. Descripción automática de la vegetación III. Clasificación y Ordenación simultánea de datos cualitativos. *Anal. Edaf. Agr.*, 38: 2207-2224.
- DIGBY, P. G. N. & KEMPTON, R. A. 1986. *Multivariate analysis of ecological communities*. Chapman & Hall. London
- EILERTSEN, O., OKLAND, R. H, OKLAND, T. & PEDERSEN, O. 1990. Data manipulation and gradient length estimation in DCA ordination. *J. Veg. Sci.*, 1: 261-270.
- ELLEMBERG, H. 1979. Zeigerwerte der Gefässpflanzen Mitteleuropas. *Scr. Geobot.*, 9: 1-121.
- ESCUADERO, A. 1992. *Estudio fitoecológico de las comunidades rupícolas y glerícolas del macizo del Moncayo*. Tesis doc. U. C. M. Madrid.
- ESCUADERO, A., GAVILÁN, R. & PAJARÓN, S., 1994. Saxicolous communities of the Sierra del Moncayo (Spain). A classificatory study. *Coenoses* 9: 15-24.
- ESCUADERO, A. & PAJARÓN, S. 1994. Numerical Syntaxonomy of *Asplenietalia petrarchae* in the Iberian Peninsula. *J. Veg. Science*, 5(2): 205-214.
- ESTABROOK, G. F. & ROGERS, D. J. 1966. A general method of taxonomic description for a computed similarity measure. *Bioscience*: 16: 789-793.
- EVERITT, B. S. 1980. *Cluster analysis*. 2nd Edition. Heinemann. London.
- FAITH, D. P., MINCHIN, P. R. & BELBIN, L. 1987. Compositional dissimilarity as a robust measure of ecological - distance. *Vegetatio*, 69: 57-68.
- FEOLI, E. & LAGONEGRO, M. 1979. Intersection analysis in phytosociology: computer program and application. *Vegetatio*, 40: 55-59.

- FEOLI, E. & LAUSI, D. 1981. The logical basis of syntaxonomy in vegetation science. In Dierschke, H. (ed.) *Syntaxonomie*, 35-42. J. Cramer. Vaduz.
- FEOLI, E. & ORLÓCI, L. 1991. The properties and interpretation of observations in vegetation study. In Feoli, E. & Orlóci, L. (eds.) *Computer assisted vegetation analysis*: 3-13. Kluwer academic publisher. Dordrecht.
- FEOLI, E. & ZUCCARELLO, V. 1986. Ordination based on classification: yet another solution? *Abs. Bot.*, 10: 203-219.
- FEOLI, E. & ZUCCARELLO, V. 1988. Syntaxonomy: a source of useful fuzzy sets for environmental analysis? *Coenoses*, 3: 141-147.
- FEOLI-CHIAPELLA, L. & FEOLI, E. 1977. A numerical phytosociological study of the summits of the Majella massive (Italy). *Vegetatio*, 34: 21-39.
- FISCHER, H. S. & BEMMERLEIN, F. A. 1989. An outline of data analysis in phytosociology: past and present. *Vegetatio*, 81(1-2): 17-28.
- FRANKLIN, J. & MERLIN, M. 1992. Species-environment patterns of forest vegetation on the uplifted reef limestone of Atiu, Mangaia, Ma'uke and Miti'aro, Cook Island. *J. Veg. Sci.*, 3(1): 3-15.
- GABRIEL, K. R. 1971. The biplot display of matrices with application to principal components analysis. *Biometrika*, 58: 453-467.
- GANDULLO, J. M., BAÑARES, A., BLANCO, A., CASTROVIEJO, M., FERNÁNDEZ LÓPEZ, A.; MUÑOZ, L.; SÁNCHEZ PALOMARES, O. & SERRADA, R. 1991. *Estudio ecológico de la laurisilva canaria*. ICONA, Colección Técnica. Madrid.
- GAUCH, H. G. 1982. *Multivariate analysis in community ecology*. Cambridge Uni. Press. Cambridge.
- GAUCH, H. G. & WENTWORTH, T. R. 1976. Canonical correlation analysis as an ordination technique. *Vegetatio*, 33: 17-22.
- GAUCH, H. G. & WHITTAKER, R. H. 1981. Hierarchical classification of community data. *J. Ecol.*, 69: 537-557.
- GITINS, R. 1985. *Canonical analysis. A review with applications in ecology*. Springer-Verlag. Berlin.
- GOODALL, D. W. 1973. Samples similarity and species correlation. In Whittaker, R. H. (ed.) *Ordination and classification of communities*: 105-156. Junk. The Hague.
- GOODALL, D. W. 1978. Numerical classification. In Whittaker, R. H. (ed.) *Classification of plant communities*: 247-286. Junk. The Hague.
- GOODALL, D. W. & FEOLI, E. 1988. Application of probabilistic methods in the analysis of phytosociological data. *Coenoses*, 1: 1-9.
- GOODALL, D. W. & JONHSON, R. W. 1982. Non linear ordination in several dimensions. A maximum likelihood approach. *Vegetatio*, 48 : 197-208.
- GOWER, J. C. 1971. A general coefficient of similarity and some of its properties. *Biometrics*, 27: 857-871.
- GREENACRE, M. J. 1984. *Theory and applications of correspondence analysis*. Academic Press. London.
- HAUSER, M. & MUCINA, L. 1991. Spatial interpolation methods for interpretation of ordination diagrams. In Feoli, E. & Orlóci, L. (eds.) *Computer assisted vegetation analysis*: 299-316. Kluwer Academic Publishers. Dordrecht.
- HILL, M. O. 1977. Use of simple discriminant functions to classify quantitative phytosociological data. In Diday, E., Lebert, L., Pages, J.P. & Tomassone, R. (eds.) *First international symposium on data analysis and informatics*; 181-199. Institut de recherche d'informatique et d'automatique. Le Chesnay.
- HILL, M. O. 1979. TWINSPLAN A FORTRAN program for arranging multivariate data in an ordered two-way table by classification of the individual and attributes. Cornell Univ. Ithaca. New York.

- HILL, O. M., BUNCE, R. G. H. & SHAW, M. W. 1975. Indicator species analysis. A divisive polythetic method of classification, and its applications to a survey pinewoods in Scotland. *J. Ecol.*, 63: 597-613.
- HILL, M. O. & GAUCH, H. G. 1980. Detrended correspondence analysis: an improved ordination technique. *Vegetatio*, 42: 47-58.
- IHM, P. & GROENEWOUD, H. 1975. A multivariate ordering of vegetation based in Gaussian type gradient response curves. *J. Ecol.*, 63: 767-777.
- JACCARD, P. 1908. Nouvelles recherches sur la distribution florale. *Bull. Soc. vaudoise Sci. nat.*, 44: 223-270.
- JOHN, E. A. & DALE, M. R. T. 1990. Environmental correlates of species distributions in a saxicolous lichen community. *J. Veg. Sci.*, 1: 385-392.
- JONGMAN, R. H. G., TER BRAAK, C. J. F. & VAN TONGEREN, O. F. R. 1987. *Data analysis in community and landscape ecology*. Pudoc, Wageningen.
- KENKEL, N. C. & ORLÓCI, L. 1986. Applying metric and nonmetric multidimensional scaling to ecological studies some new results. *Ecology*, 67: 919-928.
- KENT, M. & BALLARD, J. 1988. Trends and problems in the application of classification and ordination methods in plant ecology. *Vegetatio*, 78: 109-124.
- KNOX, R. G. 1989. Effects of detrending and rescaling on correspondence analysis: solution stability and accuracy. *Vegetatio*, 83: 129-136.
- KOVARS, P. & LEPS, J. 1986. Ruderal communities of the railway station. Česká Trebová (eastern Bohemia, Czechoslovakia), remarks on the application of classical and numerical methods of classification. *Preslia*, 58: 141-163.
- JÉNSEN, S. 1978. Influences of transformation of cover values on classification and ordination of lake vegetation. *Vegetatio*, 37: 19-31.
- LAMBERT, J. M. & DALE, M. B. 1964. The use of statistics in phytosociology. *Adv. Ecol. Res.*, 2: 59-99.
- LANCE, G. N. & WILLIAMS, W. T. 1966. A generalized sorting strategy for computer classifications. *Nature*, 212: 218.
- LANCE, G. N. & WILLIAMS, W. T. 1967. A general theory of classificatory sorting strategies. I. Hierarchical systems. *Computer J.*, 9: 373-380.
- LEFKOVITCH, L. P. 1976. Hierarchical clustering from principal coordinates: an efficient method for small to very large numbers of objects. *Math. Biosci.*, 31: 157-174.
- LEGENDRE, P. & FORTIN, J. M. 1989. Spatial pattern and ecological analysis. *Vegetatio*, 80: 107-138.
- LEGENDRE, L. & LEGENDRE, P. 1983. *Numerical ecology*. Elsevier. Amsterdam.
- LEPS, S. & HANDICOVA, V. 1992. How reliable are our vegetation analyses. *J. Veg. Sci.*, 3: 119-124.
- MARDIA, K. V., KENT, J. T. & BIBBY, J. M. 1979. *Multivariate analysis*. Academic Press. London.
- MARSILI-LIBELLI, S. 1989. Fuzzy clustering of ecological data. *Coenoses*, 2: 95-106.
- MAZZOLENI, S., FRENCH, D. D. & MILES, J. 1992. A comparative study of classification and ordination methods on successional data. *Coenoses*, 7: 23-44.
- MCCULLAGH, P. & NELDER, J. A. 1983. *Generalized linear models*. Chapman and Hall. London.
- MEULMANN, J. & HEISER, W. J. 1984. Constrained multidimensional scaling: more directions than dimensions. COMPSTAT. Physica-Verlag. Vienna.
- MINCHIN, P. R. 1987. Simulation of multidimensional community patterns: towards a comprehensive model. *Vegetatio*, 71: 145-156.
- MINCHIN, P. R. 1989. Montane vegetation of the Mt. Field massif, Tasmania: a test of some hypotheses about properties of community patterns. *Vegetatio*, 83: 97-110.
- MONTGOMERY, D. C. & PECK, E. A. 1982. *Introduction to Linear Regression Analysis*. John Wiley, New York.

- MORENO, J. M.; PINEDA, F. D. & RIVAS-MARTÍNEZ, S. 1990. Climate and vegetation at the eurosiberian-mediterranean boundary in the Iberian Peninsula. *J. Veg. Sci.*, 1(2): 233-244.
- MUCINA, L., RODWELL, J. S., SCHAMINÉE, J. H. J. & DIERSCHKE, H. 1993. European vegetation survey: current state of some national programmes. *J. Veg. Sci.* 4: 429-438.
- MUCINA, L. & VAN DER MAAREL, E. 1989. Twenty years of numerical syntaxonomy. *Vegetatio*, 89(1-2): 1-16.
- NANTEL, P. & NEUMANN, P. 1992. Ecology of ectomycorrhizal-basidiomycete communities on a local vegetation gradient. *Ecology*, 73(1): 99-117.
- NEUMAN, D. A. & NORTON, V. T. 1986. Clustering and isolation in the consensus problem for partitions. *J. Class.*, 3: 281-297.
- NOY-MEYR, I. 1973. Data transformations in ecological ordination. I. Some advantages on non-centering. *J. Ecol.*, 61: 329-341.
- NOY-MEYR, I. & VAN DER MAAREL, E. 1987. Relations between community theory and community analysis in vegetation science: some historical perspectives. *Vegetatio*, 69: 5-15.
- OCHIAI, A. 1957. Zoogeographic studies on the soleis fishes found in Japan and its neighbouring regions. *Bull. Jap. Soc. Sci. Fish.*, 22: 526-530.
- OKLAND, R. H. 1990. Vegetation ecology: theory, methods and applications with reference to Fennoscandia. *Sommerfeltia*, suplement 1: 9-233.
- OKLAND, R. H. & EILERTSEN, O. 1994. Canonical Correspondence Analysis with variation partitioning: some comments and an application. *J. Veg. Sci.*, 5(1): 127-138.
- OKSANEN, J. 1988. A note on the occasional instability of detrending in correspondence analysis. *Vegetatio*, 74: 29-32.
- OKSANEN, J. & HUTTUNEN, P. 1989. Finding a common ordination for several data sets by individual differences scaling. *Vegetatio*, 83: 137-145.
- ORLÓCI, L. 1967. An agglomerative method for classification of plant communities. *J. Ecol.*, 55: 193-205.
- ORLÓCI, L. 1978. *Multivariate analysis in vegetation research*. 2nd. ed. Junk, The Hague.
- PALMER, M.W. 1993. Putting things in even better order: the advantages of canonical correspondence analysis. *Ecology*, 74(8): 2215-2230.
- PAULE, L. & GÓMORY, D. 1987. Comparison of different clustering methods applied to matrices of genetic distances based on allelic frequencies. *Abstr. Bot.*, 11: 43-51.
- PEINADO, M. & RIVAS-MARTÍNEZ, S. 1987. *La vegetación de España*. Publicaciones de la Universidad de Alcalá de Henares. Madrid.
- PIELOU, E. C. 1977. *Mathematical Ecology*. Wiley, New York.
- PIELOU, E. C. 1984. *The interpretation of ecological data: a primer on classification and ordination*. Wiley & Sons, New York.
- PODANI, J. 1989a. New combinatorial clustering methods. *Vegetatio*, 81: 61-77.
- PODANI, J. 1989b. Comparison of Fuzzy Classifications. *Coenoses*, 1: 17-21.
- PODANI, J. 1989c. Comparison of classifications and ordinations of vegetation data. *Vegetatio*, 83: 111-128.
- PODANI, J. 1989d. A method for generating consensus partitions and its application to community classification. *Coenoses*, 1: 1-10.
- PODANI, J. & DICKINSON, T. A. 1984. Comparison of dendrograms: a multivariate approach. *Can. J. Bot.*, 62: 2765-2778.
- POPMA, J., MUCINA, L., VAN TONGEREN, O. F. R. & VAN DER MAAREL, E. 1983. On the determination of optimal levels in phytosociological classification. *Vegetatio*, 52: 65-75.
- POTVIN, C. & ROFF, D. A. 1993. Distribution-free and robust statistical methods: viable alternatives to parametric statistics? *Ecology*, 74(6): 1617-1628.
- PRENTICE, I. 1977. Non-metric ordination methods in ecology. *J. Ecol.* 65: 85-94
- RAO, C. R. 1984. The use and interpretation of principal components analysis and applied research. *Sankhya*, 26 : 329-358.

- RETUERTO, R. & CARBALLEIRA, A. 1991. Defining phytoclimatic units in Galicia, in Spain, by means of multivariate methods. *J. Veg. Sci.*, 2(5): 699-710.
- ROBERTS, D. W. 1986. Ordination on the basis of fuzzy set theory. *Vegetatio*, 66: 123-131.
- ROGERS, D. J. & TANIMOTO, T. T. 1960. A computer program for classifying plants. *Science* (Wash. D.C.) 132: 1115-1118.
- ROHLF, F. J. 1982. Consensus indices for comparing classifications. *Mathematical biosciences*, 59: 131-144.
- SCHWICKERATH, M. 1931. Die gruppenabundanz (gruppenmächtigkeit); ein beitrag zur begriffsbildung in der pflanzensoziologie. *Englers Bot. Jahrb.*, 64: 1-16.
- SNEATH, P. H. A. & SOKAL, R. R. 1973. *Numerical taxonomy*. Freeman, San Francisco.
- SOKAL, R. R. & MICHENER, C. D. 1958. A statistical method for evaluating systematic relationships. *Univ. Kansas Sci. Bull.*, 38: 1409-1438.
- SOKAL, R. R. & SNEATH, P. H. A. 1963. *Principles of numerical taxonomy*. WH. Freeman, San Francisco.
- SORENSEN, T. 1948. A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application to analysis of the vegetation on Danish commons. *Biol. Skr.*, 5: 1-34.
- TARAZONA, T. 1984. *Estudio florístico, ecológico y fitosociológico de los matorrales del sector ibérico-soriano*. Pub. Inst. Nac. Inv. Agrarias. Colección Tesis Doctorales, 45: 1-355. Madrid.
- TARAZONA, T.; CALVO, P.; ZALDÍVAR, P. & MOREIRO, S. 1991. Tipos biológicos en la estructura de los hayedos españoles. *Doc. Phytosoc.*, 13: 247-280.
- TER BRAAK, C. J. F. 1983. Principal components biplots and alpha y beta diversity. *Ecology*, 64: 454-462
- TER BRAAK, C. J. F. 1986. Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis. *Ecology*, 67: 1167-179.
- TER BRAAK, C. J. F. 1987. The analysis of vegetation-environment relationships by canonical correspondence analysis. *Vegetatio*, 69: 69-77.
- TER BRAAK, C. J. F. 1988. Partial canonical correspondence analysis. In Bock, H.H. (ed.). *Classification methods and related methods of data analysis*: 551-558. Amsterdam.
- TER BRAAK, C. J. F. 1990. Update notes: CANOCO version 3.10. Wageningen.
- TER BRAAK, C. J. F. & BARENDREGT, L. G. 1986. Weighted averaging of species indicator values: its efficiency in environmental calibration. *Math. Biosc.*, 78: 57-72.
- TER BRAAK, C. J. F. & LOOMAN, C. W. N. 1986. Weighted averaging, logistic regression and the gaussian response model. *Vegetatio*, 65: 3-11.
- TER BRAAK, C. J. F. & PRENTICE, I. 1988. A theory of gradient analysis. *Adv. Ecol. Res.*, 18: 271-317.
- TONTERI, T., MIKKOLA, K. & LAHTI, T. 1990. Compositional gradients in the forest vegetation of Finland. *J. Veg. Sci.*, 1: 691-698.
- TSO, M. K. S. 1981. Reduced-rank regression and canonical analysis. *J. Roy. Statist. Soc. B.*, 43: 183-189.
- TÜXEN, R. & ELLEMBERG, H. 1937. Der systematische und ökologische gruppenwert. Ein beitrag zur begriffsbildung und methodik der pflanzensoziologie. *Mitt. Flor. Soz. Arbeitsgem.*, 3: 171-184.
- VAN DER MAAREL, E. 1966. *Over vegetatiestructuren, -relaties en- systemen, in het bijzonder in de duingraslanden van Voorne* (With summary). Thesis Utrecht, 170 pp.
- VAN DER MAAREL, E. 1969. The use of ordination models in phytosociology. *Vegetatio*, 19: 21-46.
- VAN DER MAAREL, E. 1975. The Braun-Blanquet approach in perspective. *Vegetatio*, 30(3): 213-219.
- VAN DER MAAREL, E. 1981. Some perspectives of numerical methods in syntaxonomy. In Dierschke, H. (ed.) *Syntaxonomie*, 77-93. J. Cramer. Vaduz.

- VAN DER MAAREL, E., JANSEN, J. G. M & LOUPPEN, J. M. W. 1978. TABORD, a program for structuring phytosociological tables. *Vegetatio*, 38: 143-156.
- VAN GROENEWOUD, H. V. 1992. The robustness of correspondence, detrended correspondence, and TWINSpan analysis. *J. Veg. Sci.*, 3: 239-246.
- VAN TONGEREN, O. F. R. 1987. Cluster analysis. In Jongman, R. H. G., Ter Braak, C. J. F. & van Tongeren, O. F. R. (eds.). *Data analysis in community and landscape ecology*. Pudoc. Wageningen.
- WARTENBERG, D., PERSON, S. & ROHLF, F. J. 1987. Putting things in order: a critique of detrended correspondence analysis. *Amer. Nat.*, 129: 434-448.
- WESTHOFF, V. & VAN DER MAAREL, E. 1978. The Braun-Blanquet approach. In Whittaker, R. H. (ed.). *Ordination and Classification of communities*, 617-726. Junk, The Hague.
- WHITTAKER, R. H. 1962. Classification of natural communities. *Bot. Rev.*, 28: 1-239.
- WHITTAKER, R. H. 1967. Gradient analysis of vegetation. *Biol. Rev.*, 49: 207-264.
- WHITTAKER, R. H. 1978. Direct gradient analysis. In Whittaker, R. H. (ed.) *Ordination of plant communities*. Junk. The Hague 7-50
- WILDI, O. 1989. A new numerical solution to traditional phytosociological tabular classification. *Vegetatio*, 81: 95-106.
- WILDI, O. & ORLÓCI, L. 1990. *Numerical exploration of community patterns*. SPB Academic Publishing, The Hague.
- WILLIAMS, W. T. 1971. Principles of clustering. *Annu. Rev. Ecol. Syst.*, 2: 303-326.
- WILLIAMS, W. T. & LAMBERT, J. M. 1959. Multivariate methods in plant communities. I. Association analysis in plant communities. *J. Ecol.*, 47: 83-101.
- YANAI, H. 1986. Some generalizations of correspondence analysis in terms of projection operators. In Diday, E. & al. (eds.) *Data Analysis and Informatics 4*: 193-207. Amsterdam.
- YARRANTON, G. A. 1970. Towards a mathematical model of lime-stone pavement vegetation III. Stimulation of the determinants of species frequencies. *Can. J. Bot.*, 48: 1387-1404.
- YARRANTON, G. A., Beasleigh, W. J., Morrison, R. G. & Shafi, M. I. 1972. On the classification of phytosociological data into non-exclusive groups with a conjecture about determining the optimum number of groups in a classification. *Vegetatio*, 24: 1-12.
- YEE, T. W. & MITCHELL, N. D. 1971. Generalized additive model in plant ecology. *J. Veg. Sci.*, 2: 587-602.
- ZHANG J. & OXLEY, E. R. B. 1994. A comparison of three methods of multivariate analysis of upland grasslands in North Wales. *J. Veg. Sci.* 5(1): 71-76.
- ZIMMERMAN, H. J. 1984. *Fuzzy set theory and its applications*. Kluwer-Nijhoff Publishing. Boston. Dordrecht.

Recibido 14 de febrero de 1994

Aceptado 28 de abril de 1994.