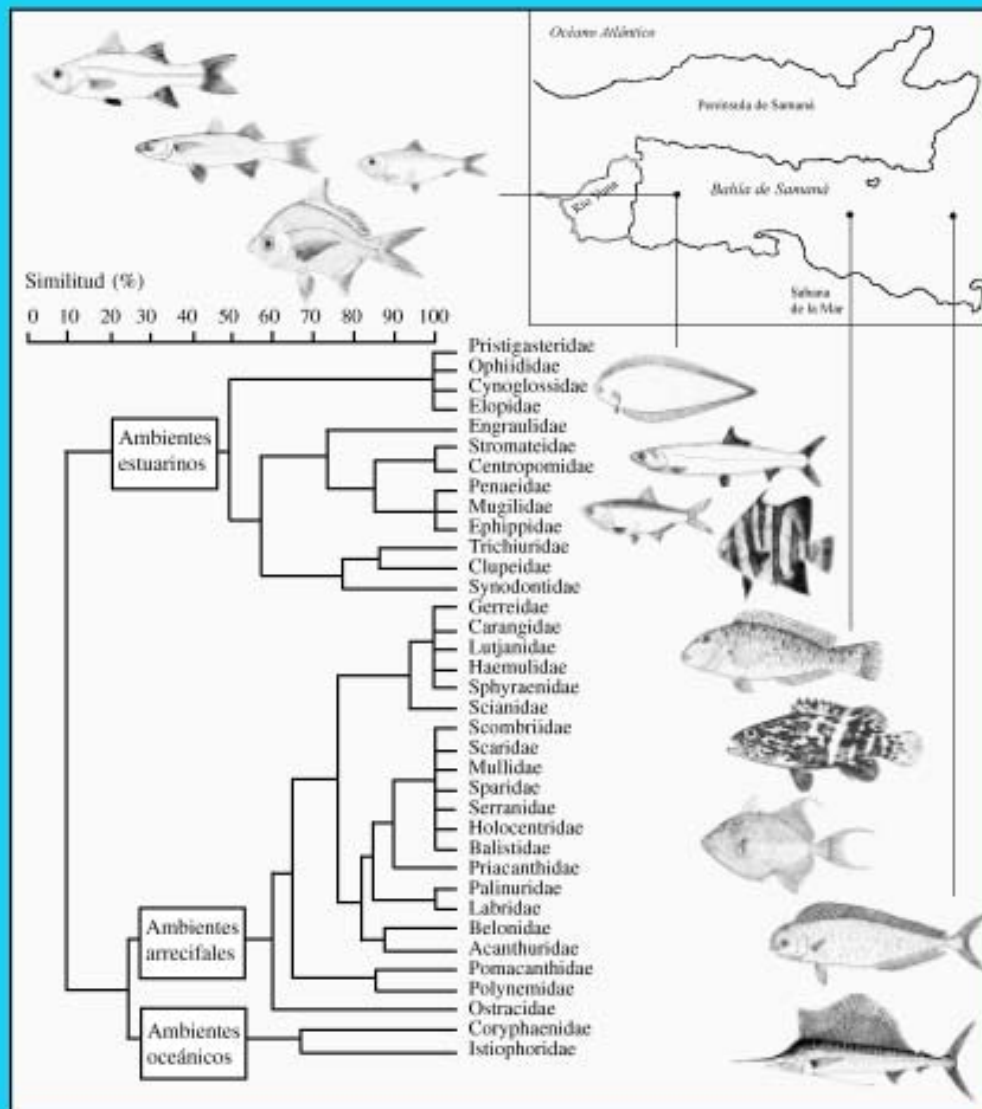


# LA CLASIFICACIÓN NUMÉRICA Y SU APLICACIÓN EN LA ECOLOGÍA

Alejandro Herrera Moreno



INSTITUTO TECNOLÓGICO DE SANTO DOMINGO  
Santo Domingo, República Dominicana



**LA CLASIFICACIÓN  
NUMÉRICA  
Y SU APLICACIÓN  
EN LA ECOLOGÍA**

ISBN 99934-25-12-5

Primera edición 2000

© Alejandro Herrera Moreno

Reservados todos los derechos. Se prohíbe toda reproducción -parcial o total- de la presente obra, sin la correspondiente autorización expresa del autor.

Edición y diseño

Alejandro Herrera Moreno y Liliana betancourt Fernández

Editora Sammenycar C. x A.

Santo Domingo República Dominicana

# **LA CLASIFICACIÓN NUMÉRICA Y SU APLICACIÓN EN LA ECOLOGÍA**

**Alejandro Herrera Moreno**  
Programa EcoMar, Inc.

**Publicación auspiciada por:**



**INSTITUTO TECNOLÓGICO DE SANTO DOMINGO**  
**Santo Domingo, R. D.**  
**2000**



*A mi adorada esposa Lily, quien “clasificó”  
definitivamente mi vida del lado de la felicidad.*





Estimado lector:

No es pretensión de este libro -al margen de aportes personales- abordar originalmente un campo en el cual existen obras clásicas y autores consagrados, pero un propósito sí me anima: allanar a otros un camino arduo que he debido transitar. Mucho se ha escrito sobre clasificación, pero dispersos los trabajos en tiempo y espacio, escritos en su mayoría en un lenguaje esencialmente matemático, en ocasiones con un tratamiento parcial del tema, con una extensa terminología técnica no siempre bien explicada, y en idiomas diferentes del español; no es fácil para el que comienza, orientarse y avanzar. La experiencia de numerosos cursos por el mundo me ha mostrado la necesidad de una guía que recoja lo esencial de los métodos y permita al que comienza, particularmente estudiantes e investigadores jóvenes de Nuestra América, orientarse en este campo no siempre bien aplicado de la clasificación y hacerle más accesible el estudio de los textos especializados. Siguiendo en lo fundamental las ideas de los clásicos, particularmente a Boesch (1977), que es a mi juicio lo más didácticamente escrito sobre la materia; con aspectos nuevos de todos los trabajos a los cuales he podido acceder y con la experiencia de mi trabajo práctico se ha confeccionado esta guía, que es además un reclamo de mis alumnos y colegas.

En el logro de este empeño no son pocos a los que debo gratitud por su ayuda desinteresada. Agradezco en primer lugar a la MSc. Rosa del Valle que me enseñó mi primer «cluster». al Dr. Jesús Sánchez del Instituto de Matemática y Cibernética Aplicada de Cuba, que como maestro me acercó por primera vez a estos métodos. Al Dr. Pedro Alcolado del Instituto de Oceanología de Cuba que revisó con su acostumbrada meticulosidad la primera versión de este libro (sin que ello le haga responsable de ninguno de sus errores) llenándolo de acertadas sugerencias. También debo agradecer a profesores de algunas Universidades que me facilitaron la presentación de mis cursos y a través de ellos a los alumnos que nutrieron esta obra de nuevas experiencias. En Brasil, agradezco a la Dra. Erika Schlenz de la Universidad de Sao Paulo y a la Dra. Maria Julia da Costa Belem de la Universidad de Río de Janeiro. En Venezuela, a la Dra. Elisabeth Méndez de Elguezábal de la Universidad de Oriente, en Cumaná. En México agradezco a los colegas del Centro de Investigaciones de Quintana Roo en Chetumal, el Instituto Tecnológico Agropecuario de Oaxaca y el Centro de Investigaciones Avanzadas de Mérida. En Cuba a los colegas del Instituto de Oceanología y del Instituto de Ecología y Sistemática, donde hallé entusiastas seguidores. En República Dominicana, tuve la oportunidad de exponer estos temas por primera vez gracias a la amabilidad del Dr. José Contreras, Decano del Área de Ciencias Básicas y Ambientales, y la Profesora Lic. Ana Mercedes Henríquez, ambos de la Universidad INTEC; la misma Universidad que hoy publica este libro bajo la dirección del Lic. Antonio Fernández, del Departamento de Investigaciones y Publicaciones Científicas, a quien extiendo un agradecimiento especial. Finalmente, deseo dar gracias al Sr. José A. Mari Mutt, Editor del Caribbean Journal of Science, por su gentileza en el envío electrónico del trabajo que sirve de base al último ejemplo de nuestro capítulo sobre la práctica de la clasificación.

Como pienso con José Martí, que «cada alumno que progresa es un maestro que nace», he intentado con este libro desbrozar de malezas un camino. Si en algunas partes éste es aún escabroso, si la hierba aquí o allá no ha sido bien cortada, no ha sido por dejadez ni apuro. Asumo la responsabilidad, sabiendo que los que vienen lo harán mejor y cumplirán su parte como he tratado yo, sencillamente- de cumplir la mía.

El Autor.



## CONTENIDO

---

1. INTRODUCCIÓN	1
2. PASOS GENERALES DE LA CLASIFICACIÓN	6
3. TIPOS DE DATOS Y SU ANÁLISIS	9
Datos cualitativos	10
Datos de doble estado: presencia-ausencia	10
Datos de multiestado ordenado	14
Datos de rango	15
Datos cuantitativos	17
Reducción de los datos	20
Transformación de los datos	21
Tratamiento de datos mezclados	23
4. MEDIDAS DE AFINIDAD	26
Medidas de afinidad cualitativas	27
Índice de similitud de Sorensen	30
Medidas de afinidad cuantitativas	31
Medidas de distancia	32
Distancia euclidiana	32
Medidas de similitud o disimilitud	36
Índice de Bray Curtis	37
Índice de Sanders	37
Índice de Canberra	38
Medidas de correlación	40
Correlación lineal	40
Correlación de Spearman	41
Alternativas de empleo de la matriz de afinidad	42
Diagrama de Trellis	42
Proyección de similaridad cenoclínica	42
5. MÉTODOS DE AGRUPAMIENTO	44
¿Cómo operan los métodos aglomerativos combinatorios?	44
Ligamiento simple o vecino más cercano	46
Ligamiento completo o vecino más alejado	47
Promedio simple	47
Promedio de grupos	47

---

---

Estrategia flexible	48
¿Cómo se hace un agrupamiento «a mano»?	49
Una alternativa de agrupamiento de datos porcentuales	54
<b>6. INTERPRETACIÓN DE LAS CLASIFICACIONES</b>	<b>58</b>
Medidas de la bondad del ajuste	58
Reglas de decisión	60
Reasignación	64
Comparación interclasificatoria	65
Evaluación de las diferencias entre grupos	66
Relación de las clasificaciones con factores externos	67
Análisis nodal	68
<b>7. LA CLASIFICACIÓN EN LA PRÁCTICA</b>	<b>72</b>
• Ejemplo 1. <i>Estructura ecológica de las comunidades de gorgonáceos en un gradiente de contaminación en los arrecifes coralinos del litoral de La Habana, Cuba.</i>	72
• Ejemplo 2. <i>Estudio de la comunidad de corales escleractíneos en los arrecifes coralinos del borde de la plataforma Suroccidental de Cuba.</i>	74
• Ejemplo 3. <i>Tipificación de biotopos en la Bahía de Cárdenas en Cuba, a través de la estructura ecológica de sus comunidades de bivalvos</i>	77
• Ejemplo 4. <i>Reclasificación de la biodiversidad coralina caribeña incluyendo los datos de la Hispaniola en la matriz de Chiappone et al. (1996).</i>	77
• Ejemplo 5. <i>Resultados de la clasificación numérica de los datos de las pesquerías de Samaná bajo el concepto de los complejos ecológicos de pesca.</i>	82
<b>8. A MODO DE CONCLUSIÓN</b>	<b>85</b>
<b>9. REFERENCIAS</b>	<b>86</b>

“La estadística es una forma de control social sobre  
la conducta profesional de los investigadores”  
Richard J. Harris

## 1. INTRODUCCIÓN

Aunque la interacción de la matemática y la estadística con la ecología no es algo nuevo, el extraordinario desarrollo de la informática en los últimos tiempos ha contribuido a popularizar el empleo de nuevos y más complejos métodos de análisis e interpretación, que han venido a dar respuesta a las necesidades de esta materia cuyo carácter multidisciplinario hace del análisis de numerosas variables bióticas y abióticas y sus interrelaciones, un objetivo esencial.

La diversidad de métodos existentes, cada uno con particularidades metodológicas y prácticas<sup>1</sup>, y ajustados a los más diversos intereses, es algo con lo cual el ecólogo debe familiarizarse -al menos en las intenciones básicas de cada uno de ellos- a fin de orientarse debidamente tanto en la obtención como en la interpretación de sus datos. En este sentido, no son una excepción los métodos de clasificación numérica.

Definida en su forma más general la *clasificación* no es más que el ordenamiento de las entidades en grupos sobre la base de las relaciones entre sus atributos (Boesch, 1977); o en palabras más sencillas: el agrupamiento de cosas similares en clases (Pielou, 1984), con lo cual se distingue de la *identificación* que pretende encontrar la clase en la cual ha de ubicarse un nuevo individuo, entre clases ya establecidas (Orlóci, 1978). Esta subdivisión parece correcta para definir dos acciones relacionadas pero diferentes: la clasificación, en su acepción común de “ordenar o disponer por clases” y con ordenación, separación, distribución u organización como sinónimos; y la identificación en el sentido de “reconocer si una cosa es la misma que se supone o se busca” con sinónimos como filiación, identidad, reconocimiento o unificación.

Sin embargo, existe ambigüedad en el uso de estos términos en la literatura. En numerosos textos (Anderson, 1984; Morrison, 1990; Johnson y Wichern, 1992; Esbensen *et al.*, 1994; Rencher, 1995) el término clasificación se emplea para denominar la ubicación de individuos en categorías pre-establecidas como parte de lo que Rencher (1995) llama el aspecto predictivo del análisis discriminante, lo cual usualmente recibe el nombre de asignación, diagnosis o como habíamos dicho, identificación, quedando el vocablo clasificación para los métodos de construir grupos (Chatfield y Collins, 1992).

Al margen de estas definiciones técnicas, la clasificación es algo inherente a la vida humana pues en este proceso aumentamos nuestro conocimiento, hacemos un uso más eficiente de la información,

---

<sup>1</sup>Existen campos bien definidos de interacción de la matemática y la estadística en la bioecología. Son algunos de ellos, además de la llamada estadística clásica (paramétrica y no paramétrica), diseño de muestreos, diseño experimental, distribuciones espaciales, índices ecológicos, métodos de clasificación, de ordenamiento y el modelaje matemático.

logramos acceder a ella más fácilmente cuando lo requerimos y además valoramos las cosas que nos rodean. El niño clasifica sus juguetes: algunos privilegiados, generalmente los más simples y desdeña otros, casi siempre los más caros. El adolescente clasifica asignaturas y amores en difíciles y fáciles. El adulto clasifica opiniones, amistades, oficios y no pocas veces en demasía los objetos del mundo material olvidando que su sentido clasificatorio infantil le acercaba más a la felicidad. Tal vez por ello Stuessy (1990), que trata en amplitud la historia de la clasificación vegetal, dice que el ser humano tiene compulsión por el orden.

En el pensamiento científico hallamos a los biólogos clasificando taxones; los geógrafos, regiones; los médicos, enfermedades; los químicos, compuestos; los historiadores, épocas; los arqueólogos, hallazgos; y los astrónomos, estrellas. En cualquier estudio, para poder interpretar la realidad circundante, el investigador clasifica y ordena los componentes de su sistema de una forma intuitiva buscando una estructura natural entre sus observaciones basada en su perfil multivariado (Hair *et al.*, 1995).

Desde Aristóteles hasta Linneo, la teoría y práctica de la clasificación aplicada al estudio de los animales y plantas ha jugado un papel importante en la biología, que tuvo su momento cumbre en las teorías de Darwin. La clasificación correcta de los organismos es tan importante en la biología, que se ha desarrollado una disciplina particular: la taxonomía, para tratar la teoría y práctica de la clasificación biológica (Fielding, 1999). La clasificación de los elementos de la Tabla Periódica por Mendeleiev ha tenido un profundo impacto en las concepciones sobre la estructura atómica. La clasificación de las estrellas en “gigantes” y “enanas” sobre la base de su temperatura y luminosidad ha afectado fuertemente la teoría de la evolución estelar (Everitt, 1993).

En el campo de la ecología estructural los biólogos marinos daneses en la primera década del siglo clasificaron con acierto las comunidades bentónicas. A partir de especies comunes de moluscos y equinodermos, junto a otros taxones marinos, establecieron varias comunidades en relación con el tipo de fondo y la profundidad, cada una de las cuales se definía por la proporción de sus grupos componentes y era denominada acorde a las dos especies dominantes. Estos trabajos marcaron el punto de partida de los actuales estudios cuantitativos del bentos. Pero sin duda alguna, corresponde a los botánicos en este campo un mérito especial, pues ya desde principios de siglo la escuela fitosociológica de Braun-Blanquet (1979) estableció bases muy claras para la clasificación de las comunidades vegetales. Sus principios de ordenación espacial y cartografía, basados en inventarios florísticos y factores climáticos y edáficos, son un ejemplo de combinación de parámetros ecológicos para lograr una subdivisión coherente de la vegetación.

Estas tempranas clasificaciones, que se han dado en llamar “subjetivas”, fueron la base de la actual ecología vegetal cuantitativa que nos presenta Greig-Smith (1983), donde la botánica entra de lleno en el campo de la clasificación numérica, a la cual se le atribuyen ventajas por encima de cualquier otro tipo de clasificación subjetiva, pues además de permitir el análisis de un número mayor de atributos de lo que es capaz de considerar la mente humana, tiene la propiedad de ser repetible una vez determinados los criterios clasificatorios.

Tales características están implícitas en su definición que establece que la *clasificación numérica* o el análisis de grupos, conjuntos, cúmulos o conglomerados (“cluster analysis”) comprende una amplia variedad de técnicas para ordenar entidades en grupos sobre la base de ciertos criterios formales objetivamente establecidos (Boesch, 1977). Kaufman y Rousseeuw (1990) la definen como el arte de encontrar grupos en los datos; y más recientemente Krzanowski y Marriott (1966a) reiteran que con esta denominación los estadísticos distinguen las técnicas de análisis que dividen los datos en grupos. Chatfield y Collins (1992) además de considerar su objetivo de simplificación mediante la agrupación, señalan su valor para la exploración de los datos, generar hipótesis acerca de la estructura de la población o derivar predicciones a partir de las tendencias de agrupamiento.

Para Sharma (1996) el análisis de grupos es una técnica que combina observaciones en conjuntos de modo que, sobre la base de determinadas características, éstos sean homogéneos o compactos internamente y a su vez bien diferentes de otros. Este proceso brinda un resumen conveniente de los datos multivariados en los cuales se basa pues tiene el efecto de reducir la dimensionalidad de la tabla de datos (Fielding, 1999), pero generalmente rinde mucho más, pues ayuda a memorizar y entender mejor nuestros datos, facilita la comunicación entre especialistas y puede tener importantes implicaciones teóricas y prácticas (Everitt y Dunn, 1991).

Actualmente, estos métodos se aplican no solo en ecología sino en los más variados dominios (ver Statistica, 2000) como la inteligencia artificial, los patrones de reconocimiento, química, biología, economía, geociencias, estudios de mercado, medicina, ciencias políticas, sicología y otros. Aunque en esta diversidad de materias se le ha conocido con diferentes nombres: taxonomía cuantitativa o matemática, sistemática numérica, morfometría multivariada, taxometría (Sneath y Sokal, 1973), cladística numérica (Neff y Marcus, 1980), clasificación automática, análisis Q, clasificación jerárquica, botriología, análisis tipológico o taxonomía numérica (Kaufman y Rousseeuw, 1990), las denominaciones generales de clasificación numérica o análisis de grupos se han ido imponiendo.

En este contexto el término “cluster” es frecuentemente empleado -como si no fuera bien rico nuestro idioma- y empleado además ambiguamente para denominar cualquier paso del proceso clasificatorio, incluidos los diagramas arbóreos que como resultado final de éste se derivan, cuya estructura no siempre son grupos en sentido estricto. Siempre que deseemos discriminar las tendencias de relación grupal de nuestros datos podemos emplear métodos de clasificación; agruparemos, si entre las posibilidades de este método usamos determinadas técnicas; que nos darán siempre una subdivisión de nuestros datos originales pero que podrán ser grupos o no, según determinadas reglas.

Los métodos de clasificación no deben ser confundidos con los de otro campo relativamente cercano: los métodos de ordenamiento, cuyo objetivo no es establecer grupos ni delimitar clases sino expresar las relaciones entre entidades en modelos espaciales simplificados de varias dimensiones. Son ejemplo de ellos: el análisis de componentes y coordenadas principales, el factorial de correspondencias, la correlación canónica, el ordenamiento Gaussiano y el escalado multidimensional, entre otros (ver Bakus, 1990; Fielding, 1999).

La literatura recoge cierto debate acerca de cuál es el tipo de análisis multivariado más apropiado para los datos ecológicos, y distintas teorías favorecen los de clasificación o los de ordenamiento. En la ecología estructural estas discrepancias surgen al considerar la variación de las comunidades como un continuo pues con este concepto, los modelos de ordenamiento espacial se ajustarían mejor a la realidad, mientras que la partición en grupos que propone la clasificación puede parecer inapropiada.

Esta diferencia de enfoques es artificial (Digby y Kempton, 1991) pues el empleo de métodos de clasificación no implica que se desconozca la naturaleza continua de la variación ecológica y de hecho, en un gradiente de cambios muchas clasificaciones son capaces de identificar grupos precisos que identifican a zonas estructuralmente estables y grupos de “ruido” que son zonas de cambio. La clave, como en todo análisis ecológico no radica tanto en los métodos de interpretación sino en su base: el diseño del muestreo, especialmente en lo concerniente a qué puntos del gradiente se eligen y cuán representativos son los parámetros seleccionados y la calidad del muestreo.

Desde el punto de vista práctico, Boesch (1977) comenta que esta discusión no tiene sentido. Ambos tipos de métodos son herramientas interpretativas útiles, que incluso se complementan, si bien algunos pueden ser más relevantes en determinadas circunstancias. La clasificación es más útil para simplificar conjuntos complejos de datos y se le concede actualmente un alto valor exploratorio y descriptivo, aunque sus propiedades estadísticas no han sido totalmente desarrolladas (Morrison, 1990). El ordenamiento es más recomendable para el análisis de conjuntos de datos más pequeños y homogéneos donde sea de interés interpretar en detalle las relaciones entre entidades, y sus técnicas están mucho más elaboradas estadísticamente. De cualquier forma, todas estas técnicas multivariadas pueden ser utilizadas para analizar los mismos datos y ayudar en la búsqueda y comprobación de los fenómenos latentes en ellos, muchas veces con resultados muy similares (Popper y Heymann, 1996).

Hasta aquí hemos hablado de métodos, pero hay otros aspectos que sería útil comentar. Es natural que tras el muestreo se busque la mejor forma de analizar, interpretar y expresar los resultados, pero: ¿estamos seguros de la calidad de nuestros datos? No puede levantarse un edificio sobre cimientos débiles y no pocas veces se ven grandes intenciones interpretativas sobre datos pobres.

Por otra parte, si aún el muestreo y los datos que de él se derivan son óptimos, todas las variantes de análisis directo de la matriz de datos deben ser agotadas antes de pensar en métodos matemáticos o estadísticos complejos. No hay método matemático ni estadístico que supere la intuición de un investigador que conoce a fondo el material con que trabaja. Los gráficos de relación entre variables bióticas y abióticas, de variaciones estacionales o locales, la mapeación de variables, la inspección de la tendencia de variación entre estaciones y en los patrones de distribución de las especies y el empleo de métodos sencillos de correlación o pruebas de la estadística clásica -tanto paramétricas como no paramétricas- son algo insustituible que proporcionan la clave para orientarse hacia métodos más complejos, si son realmente necesarios.



Algunas importantes revisiones sobre análisis multivariado (Digby y Kempton, 1991; Everitt, 1993; Hair *et al.*, 1995) incluyen de hecho desde su inicio, resúmenes de métodos para el examen inicial de los datos. Krzanowski y Marriott (1966a) dicen que el primer paso de cualquier análisis estadístico es mirar a los datos e identificar sus peculiaridades fundamentales; simples plots de los datos pueden revelar fácilmente algunas facetas como las tendencias grupales de los individuos, las relaciones entre variables o la presencia de datos aberrantes, además de que pueden destacar aspectos relevantes, brindar puntos de análisis y hasta generar hipótesis para futuras investigaciones.

En relación con el campo que nos ocupa no es poco común que el interesado en clasificar sus datos aplique el primer índice de afinidad que tenga a mano; o éste u otro método de agrupamiento, sin conocer en esencia los mecanismos que le conducirán a subdividir sus datos en ciertos conjuntos. El resultado final, generalmente un árbol de clasificación, es considerado erróneamente como algo estático a lo cual debe buscarse necesariamente una explicación desconociendo que la estructura de grupos obtenida es simplemente el producto de todo un proceso matemático que de conocerse a cabalidad puede ser usado de manera objetiva en favor de una mejor interpretación ecológica.