

Un sistema de procesamiento de datos ecológicos

ARMANDO GUERRA-TREJO

Departamento de Biofísica
Escuela Nacional de Ciencias Biológicas, IPN
Prol. de Carpio y Plan de Ayala
Apartado Postal 42-186
11340 México, D. F.

ERNESTO A. CHAVEZ

Centro de Investigación y de Estudios Avanzados, IPN
Unidad Mérida
Apartado Postal 73
Cordemex, 97310
Mérida, Yucatán, México

GUERRA TREJO, ARMANDO y E. A., CHÁVEZ, 1986. Un sistema de procesamiento de datos ecológicos. *An. Esc. nac. Cienc. biol., Méx.* 30: 211-220.

RESUMEN: La necesidad de evaluar en corto tiempo la magnitud de los cambios espacio-temporales de conglomerados faunísticos en un ecosistema bajo tensión inducida por la contaminación, fue el motivo para el desarrollo de un sistema analítico e interactivo de colectivos taxonómicos. Los datos se almacenan en archivos creados en un disco y son procesados sucesivamente por una serie de programas organizados modularmente. Las posibilidades de análisis incluyen: estadísticas básicas (total de individuos, abundancia media y desviación estándar de la abundancia de cada taxón por estación); estimación de los parámetros que describen la abundancia relativa de las especies en la comunidad (serie logarítmica y serie log-normal); índice de diversidad (Shanon) y equidad (J); análisis comparativo de la diversidad y de su variación a través de la *t* de Student; coeficiente de correlación entre estaciones; una transformación logarítmica de los datos de abundancia es opcional; análisis de conglomerados ordenados jerárquicamente (dendrogramas), análisis de componentes principales (con opción para comparar taxa o estaciones); matriz de similitud; de esta matriz, el sistema identifica los valores de asociación más altos y los incorpora a un análisis de correlación múltiple. Este procedimiento analítico aplicado a muestreos realizados a través del tiempo, al hacer la evaluación de parámetros y estimar la relación entre las variables constituye una herramienta muy eficaz para el conocimiento de la estructura de las comunidades y para la evaluación del impacto inducido por las actividades humanas en los ecosistemas.

INTRODUCCIÓN

El desarrollo de la ecología moderna impone la necesidad de realizar investigación interdisciplinaria como única vía para el entendimiento de la naturaleza y de los procesos que en ella ocurren en su interacción con los organismos a través de los flujos de energía y materia de los ecosistemas. Una consecuencia de esta actividad desarrollada colectivamente es la obtención de gran cantidad de in-

formación cuyo análisis e interpretación resultan lentos y tediosos si se trata de llegar a ellos a base de procedimientos tradicionales, que al ser tan limitados, resultan poco prácticos y no muy recomendables. La ciencia actual obliga al uso de herramientas puestas a su servicio por la tecnología como parte del proceso interactivo de ambas disciplinas y esto ha caracterizado el progreso de la humanidad durante el presente siglo. La Ecología, como una de las ciencias más recientes, no podía estar al margen de esta tendencia y ha hecho de los computadores una de las herramientas de uso cotidiano. En estas circunstancias, la necesidad de analizar el posible impacto de los hidrocarburos sobre el zooplankton de la Sonda de Campeche, constituyó el estímulo para hacer uso de los principios del análisis de sistemas y aplicarlo al estudio de ese ecosistema. Debe señalarse que el sistema, objeto de este trabajo, fue creado con el propósito de someter los datos de la comunidad a una serie de procedimientos analíticos que suelen utilizarse con cierta frecuencia en ecología, cuya particularidad, en este caso, estriba en que el sistema permite realizar en cadena toda la serie de análisis para un mismo conjunto de datos. Esto significa un ahorro de tiempo considerable, pues si bien algunos de los procedimientos aquí utilizados suelen estar disponibles en las bibliotecas de programas en las centrales de computación, como los paquetes de estadística que, por ser de uso general, imponen la necesidad de tediosas modificaciones en los formatos de entrada de datos o en la estructura misma de los programas. Esto, que comúnmente ocurre, limita su uso por parte de los biólogos poco familiarizados con los lenguajes de las computadoras, quienes, a su vez, con frecuencia encuentran dificultad para que los programadores interpreten sus necesidades. Gran parte de los procedimientos incorporados en este sistema se basaron en los trabajos de Davies (1971) y de Poole (1974).

EL SISTEMA OPERATIVO

El sistema (Fig. 1) es un conjunto de programas o módulos escritos en basic, desarrollados o adaptados originalmente para el sistema Wang 2200-T con 16 K-bytes de memoria y 2 unidades de diskette. Es pertinente señalar que este sis-

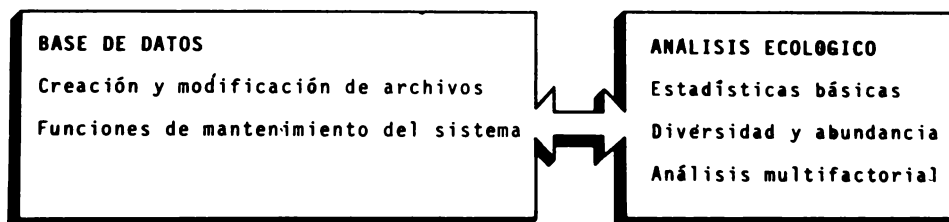


FIG. 1. Estructura del sistema operativo.

tema ha sido adaptado al equipo *Atari* por Gustavo de la Cruz; esta adaptación incluye mayor número de opciones que las del sistema original, así como mayor capacidad en el tamaño de la matriz de datos que maneja. Este sistema permite el análisis cuantitativo de datos ecológicos almacenados en un disco como un archivo en el cual se almacenan las variables. La información contenida en ellas es: número de archivos (especies) en el disco, número de parámetros, fecha de creación del disco, nombre del archivo 1, número de estaciones activas, nombre del parámetro correspondiente, datos para hacer una regresión lineal múltiple con parámetros, número de variables independientes y última estación analizada. Cada archivo de datos, de los cuales se pueden tener hasta 54 por disco, contiene las variables con la siguiente información: género-especie/nombre del archivo, última fecha de acceso y el número de parámetros en los archivos. También indica si hay o no datos en cada uno de ellos, e identifica la variable dependiente o independiente para hacer una regresión múltiple con especies.

Los programas que constituyen el sistema (Fig. 2) permiten realizar las siguientes actividades: iniciar un disco, crear un archivo, modificar los datos que éste tenga, escribir el disco tal y como está, escribir el directorio del disco de datos; obtener las estadísticas básicas de los archivos para los cuales se tengan datos, consultar datos (ya sea por especie o por estación), determinar la diversidad muestral mediante el índice de Shanon, evaluar las diferencias de la diversidad observada entre distintas muestras con una estimación de su variancia y luego con una prueba de *t* de Student (Poole, 1974); posteriormente el sistema determina el espectro de la diversidad mediante la adición sucesiva y aleatoria de

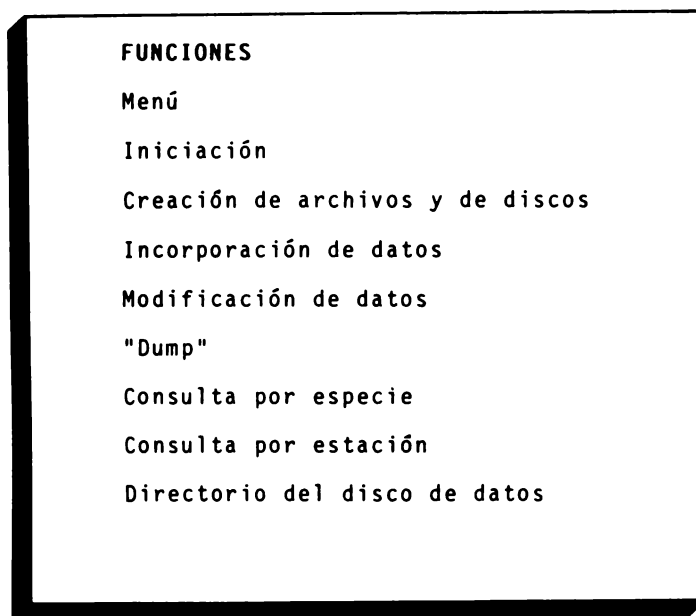


FIG. 2. Estructura y opciones de la base de datos.

muestras (Margalef, 1969); hace el ajuste a la serie log-normal a la serie logarítmica; análisis de componentes principales, análisis de regresión múltiple (ya sea por parámetros o por especies) y realiza actividades de mantenimiento del sistema que opera por medio de un menú que se exhibe en la pantalla y, una vez que se selecciona la opción deseada, el módulo irá dando las indicaciones adecuadas. Desde cualquier módulo se puede volver a llamar al menú del sistema.

Menú. Contiene la primera parte del sistema; una vez iniciada la ejecución, buscará el disco de datos y si lo encuentra leerá el número de estaciones que se están manejando; si no lo encuentra pondrá en la pantalla el mensaje correspondiente y repetirá el ciclo; una vez que tiene determinado el número de estaciones esperará a que el usuario le dé la indicación correspondiente para continuar. También ejecuta algunas tareas sencillas como son: definir el número de estaciones que se desea trabajar (esta información la usa todo el sistema) y controlar las instrucciones de mantenimiento del mismo.

Iniciación. Cada disco de datos se identifica con un archivo para ser reconocido por el sistema, este módulo es el que se encarga de esta operación, para lo cual se coloca un disco nuevo y con formato adecuado; si el disco que se puso tiene archivos o cualquier otra información el sistema lo rechazará. Si el disco es aceptado por el sistema se preguntará la fecha y el nombre de los parámetros que se van a utilizar, verificará que el archivo temporal esté en el disco del sistema y si no está lo creará.

Creación de archivos y de discos. Mediante dos módulos, por una parte, se crea el archivo de cada una de las especies que se van a almacenar en el disco de datos, para lo cual se preguntará el nombre de la especie; la respuesta deberá ser de la forma género-especie, el símbolo "-" que separa los dos nombres es necesario, ya que el sistema tomará los tres primeros caracteres de cada uno para formar el nombre del archivo. El sistema verificará que en el disco de datos no exista otro archivo con el mismo nombre; si existe, el sistema procederá a poner dos caracteres numéricos a continuación de los seis anteriores y los irá incrementando de uno en uno hasta encontrar un nombre único. Una vez que se ha hecho lo anterior el sistema graba el archivo.

Por otra parte, para crear un disco, en el caso de tener varias muestras en las que haya una repetición significativa en las especies capturadas, se creó un módulo que permite reproducir en un nuevo disco el de datos que se tenga en otro disco creado previamente, copiando solamente los datos generales del disco y de los archivos y poniendo ceros en todos los registros que se tengan.

Consulta por especie. Este módulo y el siguiente permiten utilizar los archivos como un banco de datos; en este caso, el programa preguntará por el nombre del archivo correspondiente. Una vez que se confirme su presencia preguntará si se quieren los datos de todas las estaciones o de una en particular, si se quieren los datos en la pantalla o en la impresora y procederá a escribir los valores de los parámetros correspondientes.

Consulta por estación. Es semejante al anterior, sólo que en este caso se puede revisar la misma estación en todas las especies (archivos) que se tengan.

Incorporación de datos. Para la incorporación de los datos en los archivos, al iniciarse la ejecución del programa, el sistema preguntará la fecha (ésta será la última fecha de acceso al archivo) y luego especie por especie irá preguntando si hay o no datos o si se desea terminar la sesión. Si hay datos se preguntará estación por estación y dentro de cada estación preguntará los parámetros correspondientes; se tienen opciones para indicar que ya no hay datos para más parámetros o que ya no hay datos para más estaciones.

Modificación de datos. Este programa permite la modificación de los datos que ya se tienen en el archivo. El programa irá poniendo en la pantalla los nombres de las especies y archivos correspondientes y preguntará si hay o no modificaciones que hacer; si se va a modificar, entonces se preguntará por la estación y poniendo parámetro por parámetro preguntará si es o no el que se desea modificar; una vez modificado el parámetro se graba la modificación y el programa preguntará si hay más modificaciones, si las hay se repetirá el ciclo.

Recuperación. Permite escribir todos los archivos que se tengan en el disco tal y como están grabados.

Directorio del disco de datos. Con él es posible escribir el directorio del disco de datos poniendo en orden alfabético los nombres de los archivos y el número correspondiente, este listado se puede obtener en la pantalla o en la impresora.

ANÁLISIS ECOLÓGICO

La base de datos es interactiva con un conjunto de programas que hacen un análisis de tipo esencialmente estadístico (Fig. 3) que constituyen el objetivo del sistema y son los siguientes:

Estadísticas básicas. Obtienen las estadísticas básicas de todos los archivos: medias y varianzas de cada uno de los parámetros sobre todas las estaciones, esto se hace para cada especie; también escribe el total de individuos en el archivo y los totales por estación.

Análisis de diversidad. Lleva a cabo el análisis de Shanon-Wiener y calcula por cada estación el valor de H' , posteriormente calcula la estadística de prueba para comparar la diversidad de cada estación con la de cada una de las demás; para ello hace una estimación previa del valor esperado de H' , su variancia y luego somete las estimaciones a una prueba de t de Student y evalúa la significancia de las diferencias. Para este procedimiento se ha seguido a Bowman *et al.* (según Poole, 1974); por lo tanto, las unidades de diversidad empleadas son beles naturales, aunque también calcula la diversidad en bits.

Espectro de diversidad. Este módulo calcula el espectro de la diversidad: para ello se calcula el índice de Shanon-Wiener de una estación seleccionada al azar y a continuación se van incorporando nuevas estaciones, también seleccionadas al azar y calcula el índice para cada nuevo agregado hasta que se tengan todas las estaciones. Para este procedimiento se ha seguido a Margalef (1969).

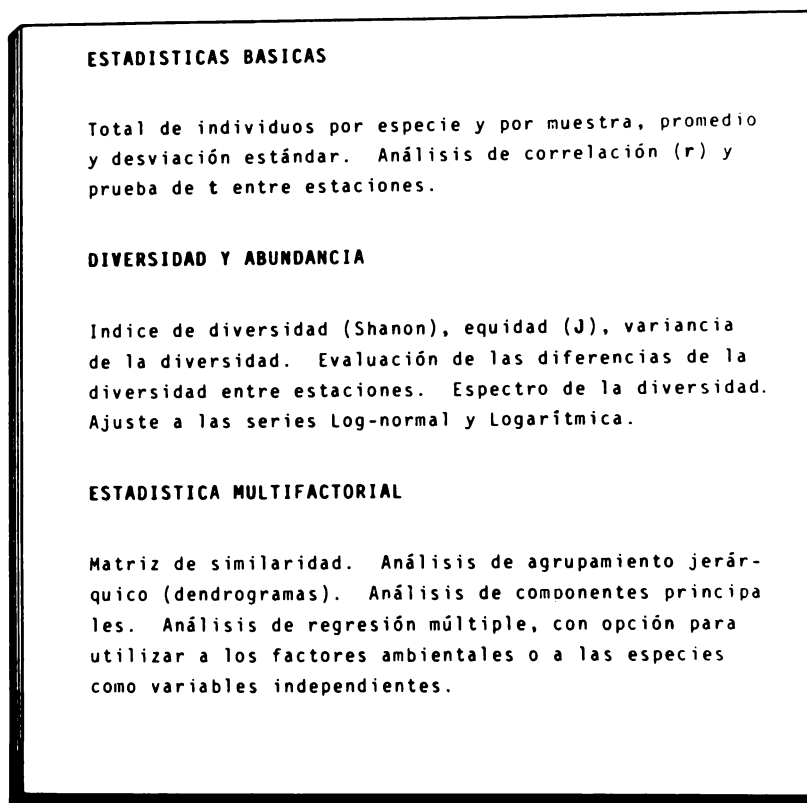


FIG. 3. Estructura y opciones del análisis ecológico del sistema.

Serie logarítmica. Con este módulo se calculan los términos de la serie logarítmica para cada especie en cada estación; primero el programa lee todos los archivos para obtener totales por estación y luego los vuelve a leer (lo mismo sucede en el módulo anterior) para calcular el parámetro alfa; como este es un proceso iterativo se limitó a 100 el número máximo de iteraciones; si al llegar a este número el valor no converge el proceso aborta automáticamente.

Transformación logarítmica. El sistema ofrece la opción de hacer una transformación logarítmica de los datos de abundancia de cada especie. Esta transformación tiene el propósito de minimizar las diferencias entre las especies dominantes y las raras, y es un tratamiento previo que con frecuencia se aplica a los datos antes de ser sometidos a los análisis de estadística multifactorial, que en este sistema son la clasificación por dendrogramas, componentes principales y la regresión múltiple.

Matriz de similitud y análisis de agrupamiento jerárquico (dendrogramas). Consiste de dos partes, la primera se encarga de crear la matriz de datos, calcular y escribir la semimatriz de similitud. La segunda parte se encarga de hacer un

análisis de agrupamiento, empezando con un nivel de similitud mayor que el máximo encontrado en la matriz de similitud; a continuación prosigue incorporando nuevas estaciones cuya similitud sea progresivamente más baja. El nivel de similitud aquí utilizado para discriminar el orden de agrupamiento es de 0.01. Es pertinente señalar que podría haberse utilizado un nivel de precisión más alto; sin embargo, en vista de que por cada orden de magnitud que se bajara significaba un tiempo de análisis considerablemente más largo, se optó por tomar el valor señalado como base, en vista de que daba un nivel de discriminación suficiente para extraer la información ecológica fundamental.

Serie-log-normal. El programa escribe para cada estación el intervalo, en base 2 del número de especies (octavas); calcula el total de individuos, el índice de diversidad en bits/individuo, la equidad (J) y los parámetros de la serie (Preston, 1948).

Componentes principales. Este módulo consta de tres partes, la primera se encarga de construir la matriz de datos y de calcular la matriz de correlación, la segunda se encarga de calcular el determinante de la matriz, los autovalores y los autovectores, y la tercera parte se encarga de calcular los componentes principales y escribir resultados.

Análisis de regresión múltiple (parámetros). Este módulo también consta de tres partes, la primera se encarga de formar la matriz de datos, la segunda invierte la matriz y la tercera escribe los resultados. Para poder construir la matriz de datos se deberá especificar el número de variables independientes (parámetros); la variable dependiente será el número de individuos.

Análisis de regresión múltiple (especies). Este módulo consta de tres partes, la primera se encarga de formar la matriz de datos, la segunda es la misma que en el módulo anterior y la tercera (datos/regresión múltiple especies) se encarga de marcar los archivos para que se tomen como variables dependiente o independientes, para esto, irá poniendo en la pantalla el nombre de la especie y del archivo correspondiente y preguntará cómo se le quiere marcar, la única manera de quitar la marca a una especie es oprimiendo la opción 3, si no se hace así y se oprime la opción 4 todas las marcas que no se hayan revisado permanecerán iguales; este módulo hace uso del disco de matrices.

DISCUSIÓN

La variabilidad intrínseca de los ecosistemas es una de las causas sobre las que se explica su estabilidad, según ha sido señalado por Holling *et al.* (1978) y por esos atributos es muy difícil definir un procedimiento rígido que permita conocer con precisión no sólo los cambios espacio temporales que ocurren en ellos, sino también los patrones estructurales que intervienen. Resulta aún más difícil poder separar claramente los cambios naturales de los inducidos por las actividades humanas, especialmente si se tiene en cuenta el señalamiento hecho por Lamont y Grant (1979) quienes opinan que al usar organismos como indicadores ecológicos, si estos son capaces de tolerar amplia variación ambiental,

entonces no serán buenos discriminadores del ambiente; pero si son estenotópicos, entonces será necesario que haya una reducción muy notable en el número de especies para poner de manifiesto un cambio ambiental significativo. Estas características hacen resaltar la importancia de los métodos de procesamiento de datos, como los aquí expuestos, como valiosos auxiliares en el análisis de grandes conglomerados de datos que a su vez facilitan la interpretación de los procesos que caracterizan a los ecosistemas y ayudan a entender algunos de los mecanismos responsables de la estructura de los mismos.

Por otra parte, es necesario resaltar el hecho de que las investigaciones hechas en áreas relativamente grandes como la expuesta a los efectos de la contaminación generalizada, deberán hacer enfoques globales sobre toda la región, ya que debido a la amplitud temporal e intensidad de los focos de impacto, la limitación espacial en el muestreo hace muy remota la posibilidad de identificar áreas libres de contaminación sobre las proximidades. En todo caso, cualquier intento de separación que se hiciera dentro de las áreas de impacto en áreas contaminadas y no contaminadas, en tanto no esté verificado por análisis químicos, la delimitación de estas áreas sería arbitraria en gran medida y tendría un carácter poco objetivo. Por la misma razón, así como en función de la heterogeneidad espacial de los ecosistemas y sin un gradiente ambiental o ecológico claramente identificable, los análisis orientados a la detección de los cambios en la estructura de las comunidades deberán ser objeto de un enfoque global y de manera análoga al estudio de las poblaciones; tales cambios no serán las características de muestras consideradas aisladamente, sino el resultado de atributos estadísticos del conjunto de muestras que se obtengan en toda el área expuesta. Es en este sentido en el que se ha interpretado el análisis del coeficiente de correlación entre muestras, así como el análisis de la diversidad en la región que sirvió de base para el estudio de este problema (De la Cruz *et al.*, en prensa; De la Campa *et al.*, en prensa, y Guzmán del Pro. *et al.*, en prensa).

Por su parte, el uso de los índices de diversidad para el estudio de las comunidades ha sido objeto de controversia, tanto porque se ha aplicado tautológicamente como porque a veces se ha sobrevalorado en sus alcances. Como consecuencia de ello, Green (1979) ha objetado su validez como una medida útil para la evaluación del impacto ambiental. A pesar de ello, estamos convencidos de que una selección atinada de los datos a evaluar a través de esta vía puede ser muy informativa para este propósito, especialmente si se hace una selección previa de los datos y se evalúa la diversidad de niveles tróficos, o bien, de gremios ecológicos previamente definidos (Pielou, 1975). En algunos casos, cuando la depauperación es muy severa, la aplicación del índice de diversidad a muestras no seleccionadas de la comunidad ha resultado ser un buen indicador de la magnitud del impacto de las actividades humanas sobre los ecosistemas (De Lachica, 1980). Y más aún el contenido de información en una comunidad, medido a través de este índice puede resultar informativo como un útil auxiliar en el proceso para la evaluación del impacto sobre un ecosistema si se evalúa no sólo sobre especies, como sería lo más recomendable, sino sobre entidades taxonó-

micas de rango supra específico, tal como se hizo con el zooplacton de la Sonda de Campeche (De la Cruz *et al.*, en prensa; Guzman del Proo *et al.*, en prensa). En ese análisis, a pesar de que mucha información quedó enmascarada, cualquier cambio detectable a este nivel indudablemente tiene mayor significación en la estructura de la comunidad que cualquier cambio similar en el valor de ese índice si fuese aplicado a colectivos sistemáticos identificados a nivel de especie.

Los tres procedimientos de estadística multifactorial que se incorporan en este sistema son, el análisis para la formación de dendrogramas, el de componentes principales y el de regresión múltiple, pretenden identificar patrones generales en la constitución de los ecosistemas. Debe puntualizarse que estas tres técnicas constituyen, en su conjunto, una secuela de etapas de la investigación que corresponden al método inductivo, pues las dos primeras constituyen poderosas herramientas para la formulación de hipótesis y la última evalúa la correlación entre variables; es decir, una variable dependiente, en este caso la densidad de una población, es una función de las densidades de las poblaciones de varias otras especies que concurren en las mismas áreas y que bajo estas circunstancias operan como variables independientes.

Finalmente, debe señalarse como corolario que la aplicación mecánica de todo este conjunto de procedimientos analíticos resultará estéril en tanto no se haga intervenir el espíritu crítico del investigador que logre escudriñar y extraer de los conglomerados de datos la información que tenga un sentido ecológico trascendental y le permita identificar los patrones de cambio en la estructura y organización de las comunidades. En otras palabras, es el investigador el que en última instancia dirá la última palabra en la interpretación de los datos ecológicos, y en la medida en que esta situación se cumpla, la aplicación de los más elaborados procedimientos analíticos los convertirán en grandes auxiliares del investigador científico y poderosas herramientas de cálculo, pero bajo ninguna circunstancia deberán considerarse como sustitutos de los ecólogos.

El uso de este sistema como herramienta de análisis ecológico ha sido aplicado para comparar la composición ictiofaunística de las lagunas costeras del Golfo de México, analizar los cambios estacionales de la macrofauna del fondo en una región costera de Veracruz, así como en el análisis de la variabilidad espacio temporal en la composición de la meiofauna de una laguna costera; desafortunadamente estos estudios aún no se publican.

SUMMARY

The necessity of evaluating in a short time the magnitude of the space-time changes of the fauna in an ecosystem stressed by pollution was the aim for the development of an interactive analytical system.

Data are stored in several files created in a diskette and then are processed by computer programs structured modularly. The possible analysis are: basic statistics, estimation of parameters that describe the relative abundance of species

in the community; diversity index and equity; comparative analysis of the diversity and its variation; correlation coefficient between different points of observation (optionally a logarithmic transform of abundance is available; dendrograms) principal component analysis (with the options of analyze taxa or stations); similarity matrix; from this matrix the system identifies the highest association values and these are included in a multiple correlation analysis.

This analytical procedure is a very efficient tool for understanding the structure of the community and it also may evaluate the human impact upon the ecosystem.

BIBLIOGRAFÍA

- DAVIES, R. G., 1971. *Computer programming in quantitative biology*. Academic Press.
- DE LA CAMPA, S., R. GUADARRAMA y S. MILLE. En prensa. Análisis comparativo de la comunidad zooplanctónica de la Sonda de Campeche antes y después de la derrama del pozo Ixtoc-I. Simposio Internacional Ixtoc-I, 2-5 de junio, 1982.
- DE LACHICA, F., 1980. Diversidad espacial y temporal de la bentocenosis del estuario del río Coatzacoalcos, Veracruz, México. *Memorias del Seminario sobre el Estudio Científico e Impacto Humano en el Ecosistema de Manglares*, pp. 121-134. UNESCO.
- DE LA CRUZ, G., L. GÓMEZ y D. TORRUCO. En prensa. Estructura y diversidad de la comunidad planctónica de la Sonda de Campeche. El caso del Ixtoc-I. Simposio Internacional Ixtoc-I, 2-5 de junio, 1982.
- GREEN, R. H., 1979. *Sampling design and statistical methods for environmental biologist*. Wiley.
- GUZMÁN DEL PROO, S. A., G. DE LA CRUZ, E. A. CHÁVEZ, S. DE LA CAMPA, L. GÓMEZ, R. GUADARRAMA, S. MILLE, D. TORRUCO y M. F. ALATRISTE. En prensa. Impacto de la derrama del pozo Ixtoc-I sobre el zooplancton de la Sonda de Campeche. Simposio Internacional Ixtoc-I, 2-5 de junio, 1982.
- HOLLING, C. S., 1978. *Adaptive environmental assessment and management*. Wiley.
- LAMONT, B. B. y K. J. GRANT, 1979. A comparison of twenty-one measures of site dissimilarity. In: L. Orloci, C. R. Rao y W. M. Stiteler, (eds.). *Multivariate Methods in Ecological Work*, pp. 101-126. Int. Co. Pub. House.
- MARGALEF, R., 1969. Estudios sobre la distribución a pequeña escala del fitoplancton marino. *Mem. Real Acad. Cienc. Artes Barcelona*, 40 (1): 3-22.
- PIELOU, E. C., 1975. *Ecological diversity*. Wiley-Interscience.
- POOLE, R., 1974. *An introduction to quantitative ecology*. MacGraw Hill.
- PRESTON, E. W., 1948. The commonness and rarity of species. *Ecology*, 29: 254-283.