

BANCO DE DATOS OCEANOGRÁFICOS PARA COMPUTADORA INSTALADA A BORDO (*)

por

Antonio Cruzado y Mario Manríquez (**)

INTRODUCCION

En todo trabajo científico de observación y experimentación y especialmente en las investigaciones oceanográficas, existe un importante desfase entre la recogida de muestras y la edición de los datos de forma más o menos elaborada, sobre todo en aquellos estudios que deben ser desarrollados en contacto directo con la naturaleza. No obstante, es totalmente indispensable el conocimiento y estudio de dichos datos, para poder planificar de una manera científica nuevas observaciones y experimentos para acelerar la resolución del problema en estudio.

La utilización de instrumentos de medida y de análisis capaces de tener una salida analógica o digital ha reducido considerablemente los tiempos muertos, ya que permite obtener las listas de datos pocos minutos (o segundos) después de efectuada la toma de muestras. Al mismo tiempo, la frecuencia de muestreo ha sido incrementada en forma tal, que la acumulación de números está desbordando la capacidad de lectura no automática. La Oceanografía no ha quedado ajena a esta tendencia y son hoy muchos los grupos de trabajo que utilizan modernas técnicas instrumentales (BALLESTER et al. 1972) en el estudio de los fenómenos físicos, químicos y biológicos que se producen en el mar.

Existen ya algunos grupos investigadores que planean sus campañas oceanográficas con la flexibilidad suficiente para poder establecer, sobre la marcha, el plan de acción inmediato, contando

(*)

(**) Inst. Invest. Pesqueras. Paseo Nacional Barcelona-3 España

para ello con tiempos muertos del orden de las 24 horas entre el muestreo y la aparición de los datos elaborados con un cierto grado de complejidad. Esto sólo puede lograrse por el empleo de computadoras digitales que, de forma automática, realicen todas las funciones numéricas, lógicas y de edición en un tiempo mínimo y con una intervención humana también mínima.

Es por ello absolutamente indispensable la ordenación de los datos para su utilización por el científico en el curso de la investigación, ya sea a bordo o luego en tierra, así como la disponibilidad de toda la información pertinente en un momento dado y en una forma fácilmente interpretable, que es generalmente gráfica. Para que esto pueda tener lugar es preciso: primero, tener dicha información estructurada en una forma conveniente y almacenada en unidades de acceso relativamente rápido; segundo, un conjunto de instrucciones escritas en lenguaje simple y mnemónico que permitan al científico no informático dar las órdenes oportunas al sistema para recuperar la información que en aquel momento le interesa y, tercero, un conjunto de programas que soporten, junto con los periféricos adecuados, la edición alfanumérica o gráfica solicitada y en la forma más idónea.

El presente trabajo se está llevando a cabo en parte dentro del proyecto CISERMAR y como alternativa a otro intento que se está llevando a cabo dentro del mismo proyecto (GARBAYO y BALLESTER 1974), por considerar que la estructura paralela propuesta por estos autores, aunque adecuada al lenguaje utilizado por ellos el APLSV/360, no responde a las características de la información oceanográfica, según se indicó en un memorandum presentado por uno de los autores (CRUZADO, 1973). Además el banco de datos descrito en el presente trabajo, que tiene cabida en un sistema de computación de capacidad más bien modesta (IBM 1130), resultaría óptimo e idóneo, con las oportunas modificaciones de lenguaje, para un sistema de computación de mayor envergadura. De esto se deduce que actualmente adolece de las limitaciones del lenguaje utilizado (FORTRAN/1130), aunque se beneficia de la capacidad de presentación gráfica con la presen-

cia de un Plotter Calcomp y de una pantalla Tektronix 611 ambos conectados en línea con el Sistema 1130 por medio de una interfase WDV.

ESTRUCTURA DEL BANCO

La estructura del banco de datos ha sido elegida de acuerdo con las características de la información oceanográfica recogida, tomando como base en principio los resultados de las observaciones y experimentos realizados durante las campañas Atlor-II y Atlor-I en la región del NO de Africa, así como información procedente de otros centros interesados en la zona. Aún cuando esta información puede ser de una gran variedad, es posible establecer los tipos de observaciones siguientes:

a) Perfiles verticales

Estación fija
(Euleriana)

Estación a la deriva
(Lagrangiana)

b) Barridos horizontales

Casi sinópticos

Intervalos de tiempo considerables

Si bien los perfiles verticales son considerados como instantáneos y en un lugar geográfico definido, en las estaciones a la deriva se repiten dichos perfiles siguiendo el movimiento del agua de superficie generalmente.

Por lo tanto la distinción entre estación fija o a la deriva no afecta las características de los datos, aunque de hecho esta última contiene una información implícita que es la del movimiento de la masa de agua y que puede ser solicitada al banco de datos.

Los barridos horizontales contienen observaciones a lo largo del recorrido del barco y se diferencian básicamente de los perfiles verticales, en que las profundidades de toma de muestra son muy escasas y por lo general limitadas a la superficie. La distinción entre sinopticidad y temporalidad es importante en cuanto a la presentación de los datos ya que las variaciones temporales pueden ser y son, en muchos casos, tan importantes o más que las variaciones espaciales.

Existe también en ambas categorías la distinción entre muestreo continuo y discontinuo, perteneciendo un grupo de observaciones a una u otra categoría según su frecuencia. Un tercer grupo de observaciones son las procedentes de los ecosondadores que participan a la vez de a y de b, ya que producen información continua sobre toda la columna de agua en un lugar fijo mientras el barco se desplaza horizontalmente.

Número de variables

Generalmente el número de variables medidas no es muy grande ya que las técnicas de medida y análisis, así como la instrumentación y el personal que las pone en práctica son limitados. Sin embargo, el número varía constantemente ya que de una manera ininterrumpida se van incorporando nuevas líneas de trabajo a las existentes. Por esta razón, el banco de datos debe estar estructurado de tal manera que permita esta continua expansión. Como existen diferencias notables entre variables que justifican una subdivisión jerárquica, proponemos una ordenación de esta naturaleza basada en grupos que denominamos "Niveles". En particular el Nivel 0 (cero) se ha considerado primordial ya que contiene la información de espacio y tiempo que caracteriza una observación así como las referencias a todos aquellos niveles que contienen el resto de datos pertenecientes a aquella misma observación. El Nivel 1 (uno) tiene toda aquella información que interesa a un punto geográfico independientemente de la profundidad. El Nivel 2 (dos) tiene todas aquellas variables que se refieren a un punto concreto de la columna de agua. Aún cuando en el presente estado no han sido utilizados más que estos niveles, la inclusión de la información proceden-

te del trabajo biológico deberá ser hecha en otros Niveles dadas sus especiales características (Ecosondad, Fitoplancton, Zooplancton, etc.).

--La distribución de variables en los distintos niveles es:

- NIVEL 0 Año/Mes/Día/Hora/Minuto/Longitud/Latitud/Referencias Niveles.
- NIVEL 1 Refª Nivel 0/String Composición/Nº Estación/Profundidad Fondo/Disco Secchi/P. Atmosf/Temp. aire/Hum. Rcl./Viento/Olas
- NIVEL 2 Refª Nivel 0/String Composición/Prof/Temp/Salin/Densi/Oxig/pH/NO₂/NO₃/PO₄/SiO₄/NH₄/Fluores/Alcalinidad

La inclusión de nuevas variables puede hacerse por niveles enteros lo que evita la reestructuración de los ficheros ya existentes.

Número de observaciones

El número de observaciones, en principio, no debería tener límites puesto que las investigaciones siguen y seguirán efectuándose y con este criterio ha sido estructurado el banco de datos cuya entrada principal es el tiempo. Las únicas limitaciones vienen impuestas por la configuración del sistema y aun sólo por el número de discos magnéticos (o cintas) que pueden ser accedidos al tiempo. Existe sin embargo la posibilidad de utilizar una unidad de discos adicional donde se vaya almacenando secuencialmente la información recogida bastando que, una vez completada su capacidad, el operador efectúe cambio del disco. No existe limitación en el número de observaciones impuesta por la capacidad de la memoria central, ya que las funciones operan directamente sobre los ficheros de datos y traen a ésta tan sólo aquellos que cumplen las condiciones establecidas. Esto va, naturalmente, en detrimento de los tiempos de proceso pero es obligado teniendo en cuenta la frecuencia con que son solicitados conjuntos de datos que exceden dicha capacidad.

Datos inexistentes

Una de las diferencias básicas de este sistema con el descrito por GARBAYO y BALLESTER consiste en el trato dado a los datos no existentes. Mientras en aquél no se tienen en cuenta en la propia estructura, y sólo por el contenido de la información (que está fuera de las escalas normales) se les puede detectar, en el propuesto por nosotros, la presencia o ausencia de un dato queda registrada en un string de composición que encabeza los registros de todos los Niveles excepto en el Nivel 0. En dicho string de 15 bits, cada uno se refiere a una variable y vale 1 si existe dato o 0 (cero) si no existe. A continuación los datos quedan compactados por lo que, si se dispone de ficheros con tamaño de registro variable, se consigue un ahorro de espacio considerable. Aunque la presencia de dicho string representa la pérdida de 1/30 del espacio dedicado a los datos, la proporción de datos inexistentes puede llegar a ser la mitad del espacio total, según se deduce del Test de Compactación de Datos de la campaña SAHARA-I efectuado por el grupo de CIBEMAR-UAM/IBM.

Referencias cruzadas

La búsqueda de datos puede ser realizada según cualquiera de las variables presentes. Todos los Niveles están referenciados en el Nivel 0 y éste a su vez está referenciado en cada uno de los demás. Ello quiere decir que si, por ejemplo, nos interesan todos los datos de aquellas observaciones cuya temperatura es inferior a 17.5° C, la búsqueda se realizará en el Nivel 2 y cuando sea localizada una temperatura que cumple tal condición, se hace referencia al Nivel 0 donde se encuentra tiempo y lugar en que tuvo lugar dicha observación y la referencia del registro donde se encuentra la información en el Nivel 1 si éste existe, ya que los datos correspondientes a los barridos horizontales carecen, en principio, de tal Nivel.

LIMITACIONES Y OPCIONES

En el estado actual se ha dado mayor énfasis a la estruc-

turación y a las funciones de búsqueda que al procesado de la información, ya que existe en nuestro poder un considerable número de programas capaces de producir gráficos y cálculos estadísticos y cuya adaptación al sistema se puede considerar totalmente trivial.

Por este motivo se han incluido en el programa de búsqueda además de las limitaciones de tiempo, situación y valor de las variables, las opciones de reticulado geográfico y cálculo de cocientes entre ellas.

Función búsqueda

La función de búsqueda es efectuada básicamente en cinco pasos sucesivos:

- 1º Búsqueda del registro de Nivel 0 dentro de los límites de tiempo fijados. Esta operación es totalmente secuencial ya que las observaciones se encuentran ordenadas según el tiempo.
- 2º Comparación con los límites de situación.
- 3º Búsqueda del registro de Nivel 1 si existe. La posición de este registro ha sido dada por el Nivel 0.
- 4º Comparación del string de composición de Nivel 1 con un string que contiene las variables cuyos límites se van a comparar.
- 5º Comparación de los valores de las variables con los límites establecidos en el caso de que en dicho registro existan las variables limitadoras.

.....

Estas operaciones se repiten para cuantos niveles se deseen pero se saltan si una de las variables comparadas queda fuera de límites o no existe.

.....

Ejecución de las opciones seleccionadas.

Función reticulado

Es esta una función del banco de datos que establece un retículo geográfico según unos límites y un tamaño de malla fijados por el usuario y ejecuta todas las operaciones de búsqueda citadas más arriba.

Funciones estadísticas

Todas aquellas funciones estadísticas que suponen acumulación de sumas o sumas de productos (Medias, Desviaciones standard, Coeficientes de correlación y de regresión, etc.) pueden ser hechas simultáneamente a las operaciones de búsqueda, aunque también podrían establecerse independientemente como se verá en el próximo apartado.

Funciones de salida

En la actualidad la única función de salida es por la impresora de 120 caracteres e imprime los datos y los cocientes deseados dejando en blanco aquellos valores que no existen o aquellos cocientes que no han podido ser calculados por ser el denominador cero o uno de los términos inexistente. En la cabecera se incluyen todos los datos que definen las limitaciones impuestas.

Las funciones gráficas no han sido incorporadas al banco de datos por no poder disponer de los periféricos adecuados por el momento, pero existen para efectuar gráficos temporales, planos de fase y mapas de curvas de nivel.

Existe también, aunque no ha sido todavía incorporada al banco, una función de salida a un fichero en disco que permite, mediante la operación CALL LINK enlazar con programas de proceso complejo y que utilizan los datos recuperados pero sin poder ser su ejecución simultánea a la búsqueda.

CONCLUSIONES

- 1º En lo que respecta a la estructura propiamente dicha, puesto que refleja fielmente las características de la información tratada y su situación en el contexto espacio-temporal, pocos comentarios caben hacer.
- 2º La disposición de las variables por Niveles cumple la doble misión de agrupar los datos con características semejantes permitiendo la inclusión de datos tan dispares como intensidades de eco cada 30 cm. desde la superficie hasta el fondo y frecuencias de aparición de especies en una pesca vertical de zooplankton, y la de permitir la ampliación del banco con nuevas variables sin tener que modificar los ficheros existentes.
- 3º El sistema de referencias cruzadas entre registros correspondientes de Niveles distintos permite entrar al banco por cualquiera de los Niveles sin necesidad de pasar forzosamente por el Nivel 0, si solamente interesa la información contenida en aquél.
- 4º El sistema de string de composición permite al mismo tiempo ahorrar una considerable cantidad de espacio en disco y efectuar una rápida decisión cuando se busca una determinada variable, ya que la comparación entre bits debería ser bastante más rápida que la comparación entre variables reales.
- 5º La información procedente de perfiles o de barridos recibe exactamente el mismo tratamiento y sólo se diferencia por su contenido.
- 6º El desarrollo futuro evidentemente debe prever el aumento de variables y de niveles y, por otro lado, la implementación de las opciones de salida en forma gráfica y de cálculo de parámetros estadísticos.

- 7º Su utilización en Sistemas Operativos que permitan el trabajo con registros de longitud variable y optimización de las subrutinas de utilidad mediante su escritura en Assembler en lugar de FORTRAN en que se hallan ahora, dará lugar a un sistema de banco de datos considerablemente ágil y flexible.

- 8º Por último, por haber sido programadas todas las instrucciones en un lenguaje simple e intuitivo permite un uso fácil por científicos no familiarizados con las técnicas de programación.

BIBLIOGRAFIA

BALLESTER A., A. CRUZADO, A. JULIA, M. MANRIQUEZ y J. SALAT 1972
Análisis automático y continuo de las características
físicas, químicas y biológicas del mar. Publ. Tec. Patr.
J. Cierva, 1:1-71

CRUZADO A. 1973 Posible clasificación de los datos oceanográficos
para su estructuración en un banco de datos. Informe
Técnico.

GARBAYO E. y A. BALLESTER 1974. Banco de Datos de Cibermar. Cen-
tro Investig. UAM-IBM.

ANONIMO 1973. Test de compactación de datos de Sahara-I. Centro
Investig. UAM-IBM.

Apéndice I: Estructura del banco de datos.

Apéndice II: Ejemplos de utilización.

Apéndice III: Listado de programas