

# *Palaeoceanographer*, una herramienta informática para estimar características paleoceanográficas de las aguas marinas superficiales del Cuaternario y Neógeno superior

*Palaeoceanographer, a computer tool estimating paleoceanographic characteristics of the Quaternary and late Neogene sea surface waters*

F. Serrano Rebollo <sup>(1),(2)</sup> y F. Serrano <sup>(3)</sup>

<sup>(1)</sup> Servicio de Informática. Diputación de Málaga. C/ Pacífico 54, 29004 Málaga. f.serrano@correos.centrafusion.net

<sup>(2)</sup> Novasoft. C/ Marie Curie, nº 14. Parque Tecnológico de Andalucía (PTA). 29590 Campanillas (Málaga)

<sup>(3)</sup> Departamento de Ecología y Geología. Universidad de Málaga. Campus de Teatinos, 29071 Málaga. f.serrano@uma.es

Correspondencia: Francisco Serrano. Departamento de Ecología y Geología. Universidad de Málaga. Campus de Teatinos, 29071 Málaga f.serrano@uma.es

## ABSTRACT

*This work presents Palaeoceanographer as a new computer tool that we designed for palaeoceanographic studies. However, the software utility can be also extended to other scientific areas in which the aim is to carry out a similarity multivariate analysis between cases, or the assignment of characteristics to samples on the basis of a calibration set. The computer application analyzes the proximity between samples by means of distance algorithms. Different options allow to select among the utilities offering the program, the mathematical operators of similarity and the number of cases to consider.*

**Key words:** *palaeoceanography, software, similarity analysis, planktonic foraminifera, modern analog technique (MAT)*

*Geogaceta, 44 (2008), 219-222  
ISSN: 0213683X*

## Introducción

Una de las aplicaciones clásicas de los fósiles, junto a las paleobiológicas y bio-cronoestratigráficas, es su utilidad para la determinación de características paleoambientales. En las últimas décadas hemos asistido a un avance significativo en este campo, motivado principalmente por dos causas:

Por una parte, el desarrollo de proyectos internacionales de investigación oceanográfica (DSDP; ODP; CLIMAP; etc.) y el acceso a bases de datos de organismos administrativos de los países más desarrollados (v.g. NOAA, <http://www.noaa.gov/>) permite disponer de conjuntos de datos representativos a escala global.

Paralelamente, el desarrollo de la informática que ha permitido que análisis estadísticos complejos por el enorme número de operaciones matemáticas implicadas se puedan abordar actualmente en tiempos muy breves.

La conjunción de ambos progresos ha dado la posibilidad de aplicar nuevas metodologías para conocer características paleoambientales con resultados más fiables.

En este trabajo se presenta *Palaeoceanographer* (Fig. 1) una nueva aplicación informática diseñada especialmente para

estudios paleoclimatológicos mediante la estimación de parámetros paleoceanográficos, tales como la temperatura, la estacionalidad y la salinidad de las aguas superficiales marinas a partir de análisis de similitud de asociaciones de foraminíferos planctónicos. La metodología del programa es aplicable al Pleistoceno y, con algunas consideraciones, su uso también puede extenderse al Neógeno superior (Serrano *et al.*, 2007). La aplicación está basada en la técnica de análogos modernos (MAT) de Hutson (1980), pero la herramienta se puede usar directamente en otros campos de investigación en los que se pretendan comparar, caso a caso, la similitud de dos conjuntos de datos o asignar cualidades a muestras tomando como referencia un conjunto de calibración, siempre que se siga un procedimiento simple. Aunque el programa no ofrece todas las utilidades de los grandes paquetes estadísticos disponibles en el software de mercado (*SPSS*, *Mathematica*, etc.), su principal ventaja es la sencillez de uso para los fines a los que se ha diseñado.

## Metodología

Los paleoclimatólogos han utilizado diferentes técnicas de estimación de tem-

peratura de las aguas superficiales marinas (SST). Imbrie y Kipp (1971) usaron una función de transferencia obtenida mediante análisis de regresión de componentes principales a partir de un conjunto de muestras de fondos oceánicos actuales. La técnica de la función de transferencia ha sido aplicada posteriormente por numerosos autores, bien tal como fue descrita inicialmente, bien obtenida mediante otros tratamientos matemáticos: regresión múltiple utilizando las frecuencias relativas de las especies; regresión lineal obtenida a partir de la media de los valores óptimos de la variable ambiental para cada especie, ponderados de acuerdo con sus abundancias relativas, etc.

Hutson (1980) presentó como alternativa a las funciones de transferencia mediante análisis de regresión, la técnica de análogos modernos (MAT), que está basada en el análisis de la similitud de asociaciones fósiles con las de un conjunto de calibración actual.

Diversos autores han evaluado diferentes técnicas propuestas con el fin de determinar la fiabilidad de las diferentes metodologías. Aunque los resultados no han sido unánimes, Dowsett y Robinson (1997) y González Donoso y Linares (1998) concluyeron que los mejores re-

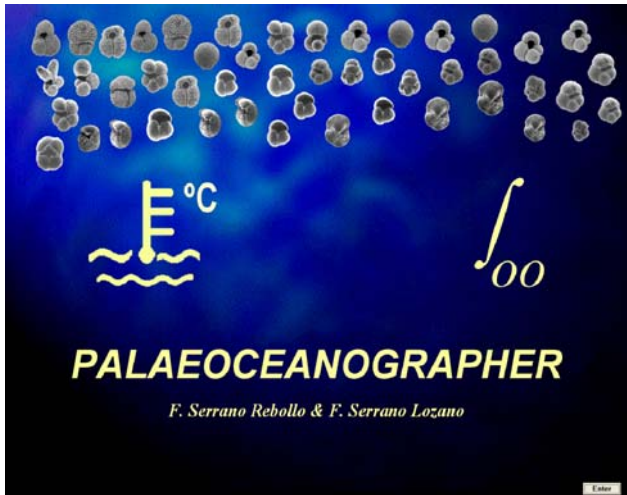


Fig. 1.- Pantalla de acceso al programa.

Fig. 1.- Screen of access to the program.

sultados se obtienen con MAT, usando la distancia de la cuerda al cuadrado como medida de proximidad.

De acuerdo con los resultados anteriores, *Palaeoceanographer* contempla algoritmos de distancia como operadores de similitud entre las asociaciones que componen la base de datos de calibración y las asociaciones fósiles que contienen las muestras que se desean analizar. El programa ofrece la posibilidad de elegir entre cuatro de las medidas de distancia más usuales:

1. Distancia euclídea:

$$d_{ij} = (S(p_{ik} - p_{jk})^2)^{1/2}$$

2. Distancia de Manhatan:

$$d_{ij} = S(p_{ik} - p_{jk})$$

3. Distancia de la cuerda al cuadrado:

$$d_{ij} = S(p_{ik}^{1/2} - p_{jk}^{1/2})^2$$

4. Distancia euclídea al cuadrado:

$$d_{ij} = (S(p_{ik} - p_{jk})^2)$$

Donde:  $d_{ij}$  es la distancia entre dos muestras  $i$  y  $j$ ;  $p_{ik}$  es la frecuencia de la especie  $k$  en la muestra  $i$ ;  $p_{jk}$  es la frecuencia de la especie  $k$  en la muestra  $j$ .  $S$  es el sumatorio para las distintas especies;

Una vez que el programa ha determinado la distancia de cada muestra analizada con cada una de las del conjunto de calibración, selecciona las más próximas de este último y valora cuantitativamente las características paleoceanográficas de la muestra en función de la media ponderada inversamente por su distancia de los valores de sus análogas.

### Funciones del programa

El programa permite realizar tres tipos de operaciones básicas (Fig. 2): autoevaluación de la base de datos de calibración (*Autoevaluation*), estimación de valores de variables paleoceanográficas

(*Parameters*) y localización de análogos (*Analogs*). Para desarrollar cualquiera de las tres operaciones se puede optar entre los algoritmos de distancia disponibles.

Mediante *Autoevaluation* el programa calcula los valores de las variables oceanográficas para cada muestra de las que componen la base de datos de calibración, de modo que se puedan comparar con los valores que tienen asignados en la propia base de datos.

La metodología que sigue el programa es extraer de la base de calibración la muestra de la que se van a estimar los valores de las variables oceanográficas (leave-one-out cross-validation procedure). Para efectuar la autoevaluación, se introduce como *sampling dataset* una copia de la base de datos de calibración en la que se suprimen las columnas de las variables estimativas. El programa ofrece una tabla con los resultados obtenidos para cada muestra de la base de calibración.

Es muy recomendable que, una vez que se disponga de una base de calibración, se realice una autoevaluación para cada uno de los algoritmos de distancia contemplados en el programa, con el fin de determinar cual presenta mayor precisión de estimación. Nuestra experiencia con las bases de datos disponibles es que la distancia de la cuerda al cuadrado es la que ofrece mejores resultados. No obstante, los estadísticos que permitan valorar los resultados de la autoevaluación (v.g., error medio y error absoluto medio) deben ser obtenidos fuera del programa.

*Parameters* estima valores de variables paleoceanográficas para muestras en las que hemos determinado la composición cuantitativa de sus asociaciones fósiles.

Por defecto, para el cálculo de los valores, el programa selecciona de acuerdo con el operador elegido, las 10 muestras del conjunto de calibración más próximas a la muestra fósil, pero ofrece la posibilidad de seleccionar otro número de muestras más próximas de entre las que componen la base de calibración que se desean tener en cuenta.

Alternativamente, *Analogs* ofrece la posibilidad de seleccionar los análogos modernos de la base de calibración más próximos a la muestra fósil analizada. En este caso, las variables oceanográficas en nuestra base de calibración son: latitud, longitud y el valor de la latitud de Mercator, útil para situar puntos sobre mapas geográficos representados en coordenadas Mercator. De esta forma, el programa indica el punto o los puntos del océano actual donde las condiciones de las aguas superficiales son más parecidas a las que existían en la cuenca en el momento en que se depositó la muestra fósil. Como resultados, el programa ofrece para cada muestra tantas muestras análogas como se hayan elegido, ordenadas por su mayor similitud. Por defecto, el programa selecciona las 3 más análogas, pero como en los casos anteriores, el número de muestras análogas puede ser modificado.

El conjunto de resultados es una tabla de cuatro columnas; la primera identifica la muestra en el mismo orden en el que se han incluido en la base de datos, y las tres restantes con las estimaciones de las variables oceanográficas. Finalmente, para la lectura de los resultados, el programa ofrece la posibilidad de que aparezcan en una ventana emergente en forma de cuadro de texto o crear un nuevo fichero del que indicaremos su ubicación (el progra-

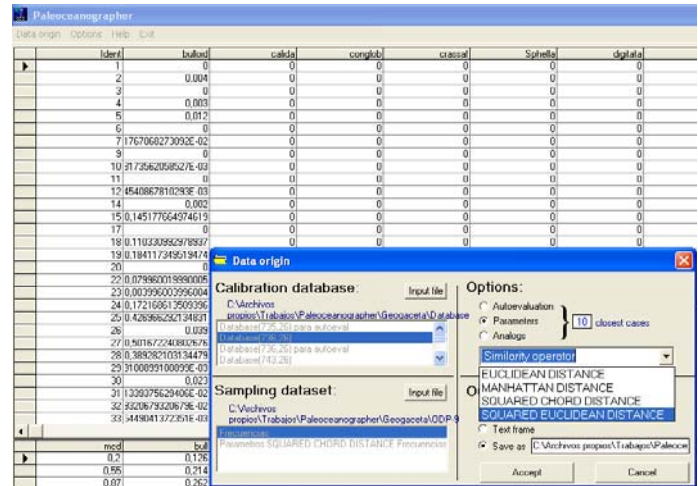


Fig. 2.- Pantalla de inserción de datos y opciones de programa.

Fig. 2.- Screen for the input of database files and the program options.

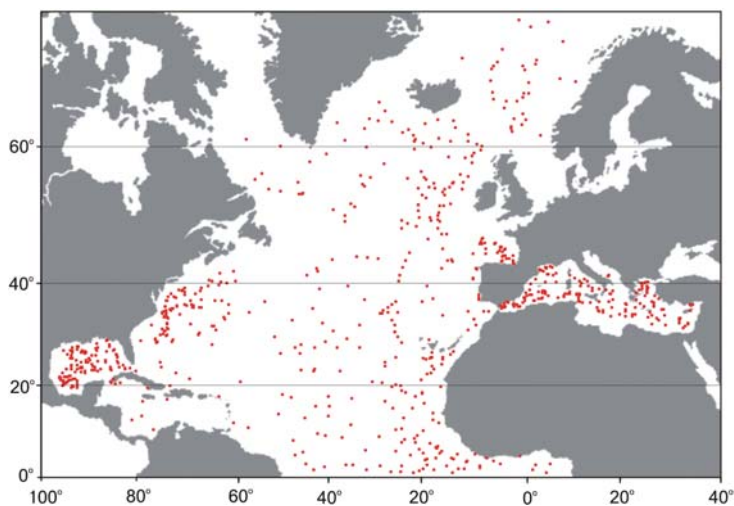


Fig. 3.- Localización de las muestras de cimbras de sondeos que componen la base de datos de calibración usada en este trabajo

Fig. 3.- Location of the core-tops composing the calibration dataset used in this work.

ma le asigna un nombre indicativo de las opciones elegidas).

### Preparación de las bases de datos

Las bases de datos deben ser matrices rectangulares formateadas en ficheros de Microsoft Access o Excel.

En la base de datos de calibración, la primera columna está reservada a la identificación de las muestras; a continuación una serie de columnas correspondientes a las variables taxonómicas utilizadas en las que los valores representen las frecuencias de los taxones (especies o morfotipos) en la asociación de cada muestra. Finalmente, seis columnas reservadas a las variables estimativas. De éstas, las tres primeras columnas corresponden a las variables oceanográficas que se desean estimar cuando se utiliza *Autoevaluation* o *Parameters*, y las tres últimas corresponden a las coordenadas geográficas cuando se opera con *Analogs*. Estas últimas seis columnas deben estar siempre contempladas, incluso en el caso de que el número de variables oceanográficas que se desean estimar sea menor (v.g., completándolas con ceros o repitiendo la columna de alguna variable).

La base de datos de las muestras a analizar debe contemplar también una primera columna con las etiquetas y la serie de columnas relativas a las variables taxonómicas, siendo imprescindible que estén en el mismo orden que en la base de datos de calibración. En este caso, lógicamente faltan las seis columnas relativas a las variables oceanográficas que se desean estimar.

### Rutina

La rutina del programa es extraordinariamente sencilla. Una vez activado el ejecutable, aparece la pantalla principal que da acceso a una segunda pantalla desde donde se realiza todo el procedimiento (Fig. 2). La parte izquierda de la ventana activa está dedicada a la introducción de las bases de datos (*Data origin*). En primer lugar se introduce el fichero conteniendo la base de calibración (*Calibration database*) y a continuación la base de datos de las muestras que se desean analizar (*Sampling dataset*). En la parte derecha de la pantalla podemos seleccionar la función que vamos a realizar, el número de muestras más similares que se quieren tener en cuenta y el operador matemático con el que queremos establecer la similitud. Finalmente, seleccionamos como queremos ver los resultados (cuadro de texto o guardar en fichero).

Una vez que empieza a correr el programa, nos indica el tiempo estimado que tardará en mostrarnos los resultados.

### Comprobación del funcionamiento del programa

Para mostrar la precisión de la metodología y el propio funcionamiento de la aplicación hemos hecho correr el programa con la opción de *Autoevaluation* sobre una base de datos de calibración (Fig.

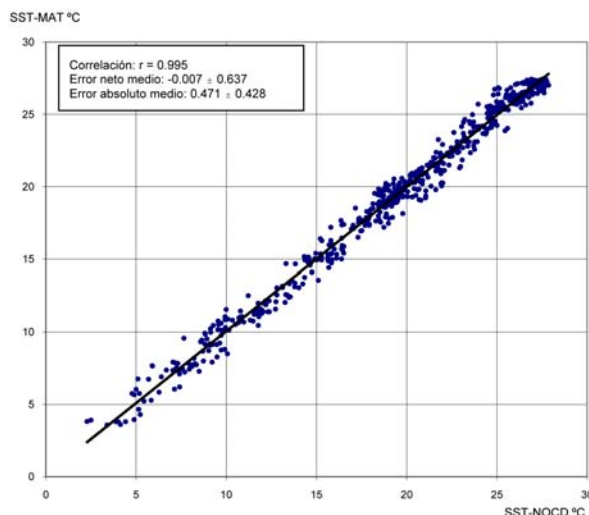


Fig. 4.- Temperaturas medias anuales estimadas por MAT (SST-MAT) frente a sus correspondientes valores del conjunto de calibración (SST-NOCD) para cada muestra de la base de datos y la línea de regresión resultante. En la gráfica se indica el coeficiente de correlación de Pearson ( $r$ ), y el resumen del análisis de errores.

Fig. 4.- Annual mean SST estimates using MAT vs. reciprocal SST values from the calibration database and the respective regression line. The value of the Pearson correlation coefficient and the summary of the error analysis are shown.

3) compuesta por 735 muestras de cimbras de sondeos de fondos oceánicos del Atlántico Norte recopilados de la base de datos de CLIMAP (1984) complementada con muestras de cimbras de sondeos de fondos marinos del Mediterráneo y del Golfo de Méjico, a partir de datos de Thunel (1978), Brunner (1979) y Loubere (1981). Las asociaciones de foraminíferos planctónicos están definidas por las abundancias relativas de 26 especies de foraminíferos planctónicos que constituyen casi la variabilidad taxonómica más significativa de este grupo de organismos en el universo oceanográfico actual. Las variables oceanográficas asociadas a cada muestra (temperaturas, estacionalidades y salinidades medias anuales de las aguas superficiales) fueron obtenidas de los archivos del National Oceanographic Data Center (Conkright *et al.*, 2002), aunque el proceso de autoevaluación que mostramos corresponde sólo al de las temperaturas (SST).

Entre las diferentes opciones de análisis de similitud, presentamos los resultados obtenidos con el algoritmo de la distancia de la cuerda al cuadrado que, en un ensayo previo fue el que, ligeramente, ofreció mejores resultados. Se pidió al programa que tomara las 10 muestras más próximas (más similares) de la base de datos de calibración, puesto que, al estar la media ponderada con la distancia, el peso de las muestras en la estimación baja

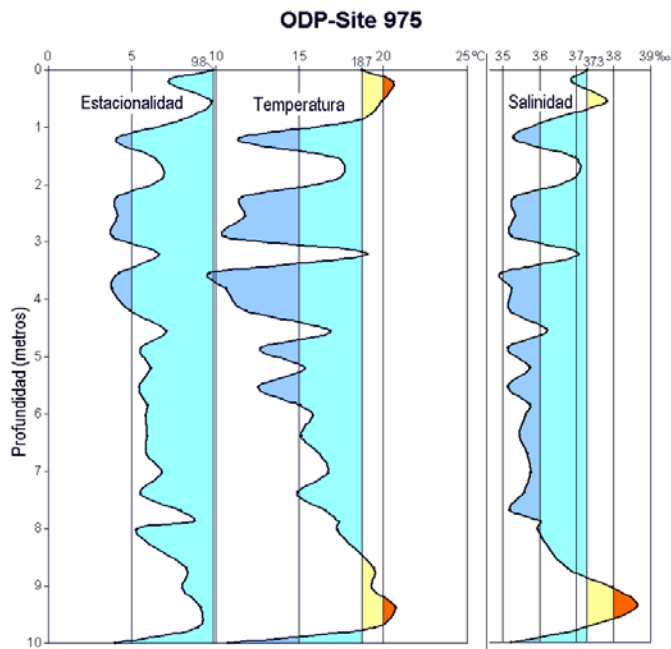


Fig. 5.- Estimaciones de los valores medios anuales de temperatura, estacionalidad y salinidad de las aguas superficiales marinas en Baleares para aproximadamente los últimos 150.000 años, deducidas a partir del análisis de muestras de los 10 m superiores de sedimentos perforados en el ODP-Site 975. En la gráfica se han incluido como referencias, los valores actuales de las variables oceanográficas (Temp: 18.7 °, Est.: 9.85 °C; Sal.: 37.3 ‰).

Fig. 5.- Estimates of the annual means of temperature, seasonality and salinity of sea surface in Balearics for approximately the last 150,000 years, deduced from the sampling analysis of the uppermost 10 m of sediments of the ODP-Site 975. As references, the present values of the oceanographic variables (SST: 18.7 °C, Seasonality: 9.85 °C; Salinity: 37.3 ‰) are represented.

con la distancia y un número más alto no ofrece mayor precisión.

Para evaluar la precisión de los resultados que ofrece el programa, se han calculado dos estadísticos:

- (1) El error medio de las SST estimadas por la autoevaluación del programa (en adelante SST-MAT), calculado como la media de la diferencia entre las temperaturas obtenidas y las SST de la base de datos de calibración (en adelante SST-NOCD).
- (2) El error absoluto medio de las estimaciones de SST, calculado como la media de los valores absolutos de las diferencias entre SST-MAT y SST-NOCD.

El error medio es indicativo del sesgo de las estimaciones. Esto permite detectar si las estimaciones de SST obtenidas muestran una sistemática desviación no aleatoria significativa en un determinado sentido, en comparación con los valores de la base de datos de calibración. Si el

error medio resultante es próximo a cero, significará que el procedimiento no produce una sistemática supervaloración o infravaloración de las estimaciones. Sin embargo, el error medio no ofrece información sobre la fiabilidad de la metodología para suministrar estimaciones precisas de SST porque compensa las sobreestimaciones con las subestimaciones. La precisión del método es medido mediante el error absoluto medio de las estimaciones.

La comparación de las temperaturas estimadas por MAT con las temperaturas incluidas en la base de datos para las 735 muestras del conjunto de calibración ofrece una concordancia muy buena. La correlación es muy alta (coeficiente de Pearson  $r = 0,995$ ; con probabilidad de hipótesis nula  $p < < 0,001$ ) y la recta de regresión prácticamente cruza por valores homogéneos (Fig. 4). La pequeña desviación que muestra la media de las diferencias entre SST-MAT y SST-NOCD (sesgo

= -0,007 °C; desviación típica  $s = \pm 0,64$  °C) indica que, en conjunto, MAT no sobrevalora o infravalora las estimaciones SST. Por otra parte, los valores absolutos de las diferencias entre SST-MAT y SST-NOCD muestra que el error de estimación medio de MAT es de 0.47 °C ( $s = \pm 0,43$  °C).

Finalmente, como ejemplo de la aplicación, en la figura 5 presentamos los resultados obtenidos con el programa en la estimación de las temperaturas, las estacionalidades y las salinidades media anuales para muestras tomadas en los 10 m superiores de sedimentos (aproximadamente 0-150.000 años) cortados por el sondeo Site 975 del Ocean Drilling Program, Leg 161 (Comas, Zahn, Klaus *et al.*, 1996) realizado en el área de Baleares (38° 53.7' N, 4° 30.6' E).

## Referencias

- Brunner, Ch. A. (1979). *Micropaleontology*, 25, 325-335.
- CLIMAP (1984). *Quaternary Research*, 21, 123-224.
- Comas, M.C., Zahn, R., Klaus, A., *et al.* (1996). *Proc. ODP, Init. Repts.*, 161: College Station, TX (Ocean Drilling Program)
- Conkright, M.E., Locarnini R.A., García H.E., O'Brien T.D., Boyer T.P., Stephens C., y Antonov J.I. (2002). *World Ocean Atlas 2001: Objective Analyses, Data Statistics, and Figures, CD-ROM Documentation*. National Oceanographic Data Center, Silver Spring, MD.
- Dowsett, H.J. y Robinson, M.M. (1997). *Palaeontology Electronica*, 1 ([http://palaeo-electronica.org/1998\\_1/dowsett/issue1.htm](http://palaeo-electronica.org/1998_1/dowsett/issue1.htm))
- González Donoso, J.M. y Linares, D. (1998). *Revista Española de Paleontología*, 13, 107-129.
- Hutson, W.H. (1980). *Science*, 207, 64-66.
- Imbrie, J. y Kipp, N.G. (1971). En: *The Late Cenozoic glacial ages* (Turekian, K.K. Ed.). Yale Univ. Press, New Haven, 71-131.
- Loubere, P. (1981). *Journal Foraminiferal Research*, 11, 137-158.
- Serrano, F., González Donoso, J.M., Palmqvist, P., Guerra-Merchán, A., Linares, D. y Pérez-Claros, J.A. (2007). *Palaeogeography, Palaeoclimatology, Palaeoecology*, 243, 174-188.
- Thunell, R.C. (1978). *Marine Micropaleontology*, 3, 147-173.