

Minería y Visualización de Paleodatos^{*}

Roberto Therón y María Vaquero

Departamento de Informática y Automática
Universidad de Salamanca
Plaza de la Merced S/N, Salamanca, 37008
theron@usal.es, mvaquero@usal.es

Resumen El interés en la comprensión de la dinámica climática del pasado puede beneficiarse de la oportunidad de introducir métodos de análisis procedentes de la Minería de Datos. Con el mismo propósito de proporcionar a los paleontólogos formas alternativas de mirar a los datos (y de sacar a la luz conocimiento oculto), se han desarrollado herramientas de visualización de información *ad hoc*; es decir, herramientas pensadas para ofrecer el máximo de información paleoceanográfica y paleoclimática que permita realizar reconstrucciones fiables, proponer modelos climáticos del pasado y, en último término, realizar predicciones. En este trabajo se presentan algunos de las herramientas desarrolladas y los resultados obtenidos con ellas.

1. Introduction

Este trabajo es un esfuerzo por acercar el enorme potencial de las técnicas procedentes de la Minería de Datos y de la Visualización de la Información a los investigadores de la climatología del pasado. Se trata de una ambiciosa intención que está plagada de numerosas dificultades.

La primera de ellas es la elevada complejidad de la paleoclimatología. Es muy difícil para un experto en las técnicas convencionales de Minería de Datos, su aplicación directa a los paleodatos. Estos son producidos en colaboración multidisciplinar entre diversos grupos, laboratorios e instituciones, que a menudo involucran a varios países. Por otro lado, la interpretación de los resultados para los expertos geólogos no es trivial, incluso con la ayuda de técnicas de visualización, ya que requiere un entrenamiento por parte de los usuarios para entender la información que se les está proporcionando [1].

Así, para poder obtener resultados fiables, se requiere un conocimiento profundo de todos los procesos necesarios: desde la toma de datos (en los buques

^{*} Los resultados presentados en este trabajo son fruto de la colaboración de los autores con el Grupo de Micropaleontología Oceánica de la Universidad de Salamanca y han sido parcialmente subvencionados a través de la Acción Especial del Programa Nacional Antártico, del Ministerio de Ciencia y Tecnología, de referencia REN2002-1112-E/ANT.

oceanográficos, por ejemplo) hasta la generación de modelos climáticos del pasado (realizando modelos de edad, análisis espectrales, reconstrucciones de paleovariables, etc.).

Lógicamente, al tratarse de un estudio del clima del pasado, los principales objetos de análisis serán series temporales.

En décadas pasadas se ha desarrollado una gran cantidad de trabajos de investigación relacionados con las series temporales y la Minería de Datos ha encontrado una nueva oportunidad de aplicar sus métodos y algoritmos. El estudio del cambio climático no ha sido ajeno a esta tendencia y, por ejemplo, en [2] se han ocupado de series temporales oceánicas. Sin embargo, no hemos encontrado en la bibliografía algún caso de aplicación de estas técnicas al análisis de paleoregistros, el cuál se ha realizado típicamente a través de técnicas estadísticas (análisis factorial) o análisis espectral [3].

Por otro lado, es importante proporcionar herramientas de visualización potentes que permitan a los paleontólogos acceder a información que de otro modo quedaría oculta. En este sentido se han realizado muy interesantes avances en lo referente a sistemas de información geográfica (para datos oceánicos PANGAEA proporciona la herramienta PanMap[4]). Sin embargo, herramientas especiales son necesarias para la visualización de algunos de los resultados que proporcionan las técnicas de Minería de datos, o para el siempre problemático análisis de datos multivariados. En este sentido se han hecho contribuciones fundamentales como es el caso de Inselberg, que propuso las denominadas Coordenadas Paralelas [5].

Un aspecto fundamental es la posibilidad de dotar a estas herramientas de interactividad. Por un lado, los geólogos quieren participar en el proceso de análisis. De hecho, es numeroso el número de investigadores que rechazan interesantes y potentes técnicas porque no se fían de herramientas (o algoritmos) *caja-negra* a las que introducen datos, ejecutan, y obtienen resultados, sin tener la más mínima noción de la filosofía subyacente. Por otro lado, la reorganización visual y la interactividad pueden llevar a un sorprendente grado de penetración en los datos [6].

En el panorama descrito en los párrafos anteriores surge en la Universidad de Salamanca la iniciativa PaleoSoftTools¹. Este grupo tiene como objetivo la aplicación de técnicas de Minería de Datos y de Visualización de la Información adaptadas a la Paleoclimatología, haciendo especial énfasis en la interactividad. Las herramientas desarrolladas gozan de una elevada aceptación internacional y son recomendadas por los dos centros de datos paleoclimáticos más importantes a nivel mundial: el *World Data Center for Paleoclimatology*², de la americana *National Oceanic and Atmospheric Administration*, y PANGAEA³, red financiada por la Comisión Europea.

El resto del capítulo se organiza como sigue: en primer lugar se expone de forma sucinta, aunque suficientemente explicativa, en qué consiste el área de la

¹ <http://carpe.usal.es/paleosofttools>

² <http://www.ngdc.noaa.gov/paleo/paleo.html>

³ <http://www.pangaea.de>

paleoclimatología, qué tipo de datos son los que proporciona y cómo son susceptibles de análisis; la sección tercera está dedicada a las técnicas de Minería de Datos que el grupo investigador ha aplicado a estos datos, así como a presentar algunas de las herramientas que se han desarrollado específicamente para analizar paleodatos; en la siguiente sección se hace lo propio para el caso de la Visualización de la Información, la cuál es especialmente relevante para los usuarios finales de estas herramientas: los geólogos. Finalmente se esbozan las principales conclusiones a las que se ha llegado a través de las diferentes experiencias de colaboración con micropaleontólogos. Asimismo, se indican las principales líneas de trabajo futuro del grupo.

2. Series Temporales Paleoclimáticas

De entre las geosferas más características de la Tierra, el océano representa, volumétricamente, la porción más importante de la hidrósfera. Ésta interactúa con otra geosfera: la atmósfera. Entre ambas y la litosfera se produce un intercambio de energía y materia que genera una serie de ciclos que gobiernan la dinámica externa del Planeta. Instaurada en estas geosferas, la biosfera, quizás la más peculiar de todas en el contexto planetario en lo que a organización y administración energética se refiere, añade un componente que en buena medida complica y hace más dependiente cada uno de los sistemas referidos.

Una de las consecuencias de la interacción entre las geosferas —con la adición de factores externos al sistema, como es el caso de las radiaciones solares—, es el clima.

La dinámica climática ha variado considerablemente a lo largo de la historia del Planeta y sus modificaciones, son por lo general, procesos que tienen lugar a lo largo de lapsos largos. No obstante, existen casos llamativos de modificaciones rápidas en la temperatura media, alteraciones de la dinámica hídrica, cambios en el nivel del mar, etc. que desde una óptica antrópica, se califican como desastres (un célebre ejemplo, efecto directo de las modificaciones de diversos parámetros atmosféricos y oceánicos, es el fenómeno conocido como El Niño).

Estas modificaciones rápidas y radicales en el entorno ambiental no se considerarían importantes si no llegasen a afectar a la sociedad. En muchos casos, se observa una cierta periodicidad en estos procesos, de orden bianual o decadal. Pero, ¿a qué obedece este fenómeno? ¿Son procesos naturales o están afectados por el hombre? ¿Pueden anunciarse para tratar de paliar en lo posible sus efectos?

La prevención, por lo tanto, ha de considerarse como la meta de cualquiera de los científicos que se dediquen al tema. Sin embargo, previo a cualquier proceso de aplicación, es preciso generar datos que nos permitan conocer los procesos.

Una posibilidad para modelar, y consiguientemente de predecir, es trabajar con modelos matemáticos. Estos no siempre cuentan con la suficiente información o tienen en cuenta sólo alguno de los factores que puedan afectar. La aplicación de un método empírico en esos mismos términos de predecibilidad, es, cuando menos, compleja.

Existe, sin embargo, otra posibilidad de estimar los efectos: disponer de datos que permitan estimar la variación en el tiempo de ciertos parámetros que han operado en el clima. Esta metodología la constituye el análisis de series temporales, y la disciplina que las acoge se denomina, de forma genérica, Paleoclimatología.

Sirviéndonos de la interacción de las geosferas, en concreto al ciclo sedimentario, una posibilidad de hallar información almacenada durante largos periodos de tiempo se encuentra precisamente en los sedimentos. Si, además, consideramos que el océano, las cuencas que en él se definen, son los puntos con menor energía potencial, esto es, las regiones donde mayores posibilidades existen de encontrar registro sedimentario con cierto grado de continuidad, los fondos del mismo, sus secuencias sedimentarias, constituyen el mayor archivo potencial. Y así es, la práctica totalidad del fondo oceánico está constituido por sedimentos producto de la acumulación durante largos lapsos de material proveniente de los continentes (vertido de ríos y transporte eólico), y en mucha menor medida, material extraterrestre.

Pero curiosamente, el componente cuantitativamente mayoritario de estos sedimentos, en especial a cierta distancia de la costa, lo constituyen elementos orgánicos: una cantidad inestimable de microesqueletos de carbonato cálcico y sílice (esencialmente) se producen en la parte superior de la columna de agua, y una vez mueren, se precipitan hacia el fondo. Estas partículas son conocidas como microfósiles, y tanto ellos, como datos obtenidos por técnicas indirectas, van a proporcionar información acerca de las condiciones que originalmente operaron en el océano.

Una de las técnicas más empleadas para reconstruir las características del océano en el pasado es la micropaleontológica. La infinidad de pequeños organismos con caparazón mineral que viven en el océano pasan a formar parte de los sedimentos una vez mueren. Variaciones en la temperatura, en la salinidad, en el contenido en nutrientes, etc., darán lugar a que los organismos que viven en esas condiciones varíen, cualitativa o cuantitativamente, en función de los cambios temporales de esos parámetros. Además de la información recogida directamente de organismos, en los sedimentos puede quedar registrada (fosilizada) otra información relativa a la composición química de los organismos que se desarrollaron allí. Una de las técnicas habitualmente empleadas es el análisis de isótopos estables. Otras técnicas auxiliares, que pueden incluso proporcionar valores absolutos acerca de la temperatura de las masas de agua, parte del análisis de biomarcadores.

Registros de similares características se estudian en cuencas de todo el océano con el objeto de hacer una reconstrucción de las condiciones que se daban en distintos puntos.

La reconstrucción de ambientes del pasado requiere, no sólo una aproximación cualitativa que nos aporte información acerca de cómo ha variado el escenario climático terrestre (global y/o regional), sino que ha de tratar de materializar sus propuestas en términos cuantitativos, comparables con parámetros que pudieran obtenerse en las diversas masas de agua actuales.

El análisis visual de estas series por parte de los micropaleontólogos se limita a la confrontación visual de representaciones 2D (figura 1).

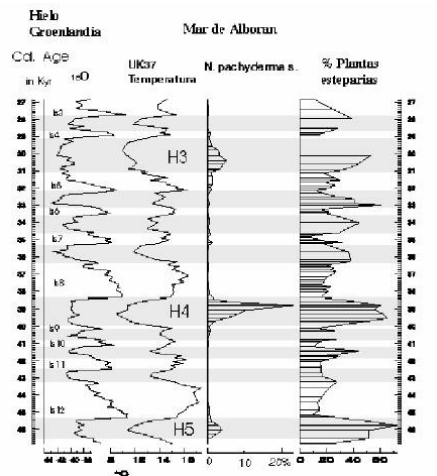


Figura 1. Confrontación de diferentes series temporales

Ésta es precisamente la propuesta de nuestro trabajo: establecer herramientas que, a partir de datos obtenidos en series pretéritas, los compare con modelos del presente y concrete en valores numéricos manejables esas variaciones observadas, además de esclarecer relaciones no visibles a priori. La acción conjunta de estas dos operaciones permitirá la reconstrucción fiable de las condiciones climáticas del pasado.

3. Minería de Paleodatos

En esta sección se van a presentar dos casos en los que la Minería de Datos puede ser de gran ayuda para comprender el clima del pasado. El primer caso se corresponde a la aplicación de diversas técnicas mediante el uso de un paquete comercial de propósito general (Mineset⁴). El segundo caso se corresponde a una herramienta desarrollada específicamente para reconstruir variables paleoclimáticas.

⁴ Mineset era un producto de Silicon Graphics en el momento en que se realizó este trabajo. Actualmente es propiedad de Purple Insight.

3.1. Reconstrucción de la Dinámica Oceánica durante los Últimos 130000 Años

El objetivo era la reconstrucción de la dinámica de las aguas superficiales en el Mar de la China durante los últimos 130000 años [7]. Para poder realizar este trabajo fue necesaria la construcción de un conjunto de datos obtenidos a partir de un testigo extraído del fondo oceánico ($8^{\circ}20.4'N$, $112^{\circ}19.9'E$; 1968 m de profundidad, 992 cm de longitud del testigo) utilizando diversas técnicas como: análisis cuantitativo de coccolitofóridos, técnica U_k^{37} para la temperatura superficial del mar (SST), y ratio $\delta^{18}O/\delta^{16}O$ para la salinidad superficial del mar.

Se trataba de comparar tres variables a través del tiempo: la temperatura y la salinidad estaban disponibles, pero era necesario encontrar algún medio de establecer la variación de la estratificación en la columna de agua.

Para ello fue necesario establecer qué especies de coccolitofóridos contribuyeron más a la señal biológica de estratificación de la columna de agua. Esto se hizo en dos pasos: 1) mediante *clustering* se encontraron los grupos de especies que se comportaron de forma similar a través del tiempo; y 2) mediante análisis de árboles de decisión (figura 2) se permitieron encontrar las reglas que establecían cómo las diferentes especies eran afectadas en mayor o menor medida bajo diferentes condiciones de temperatura y salinidad durante los diferentes estados isotópicos marinos, MIS (tiempo).

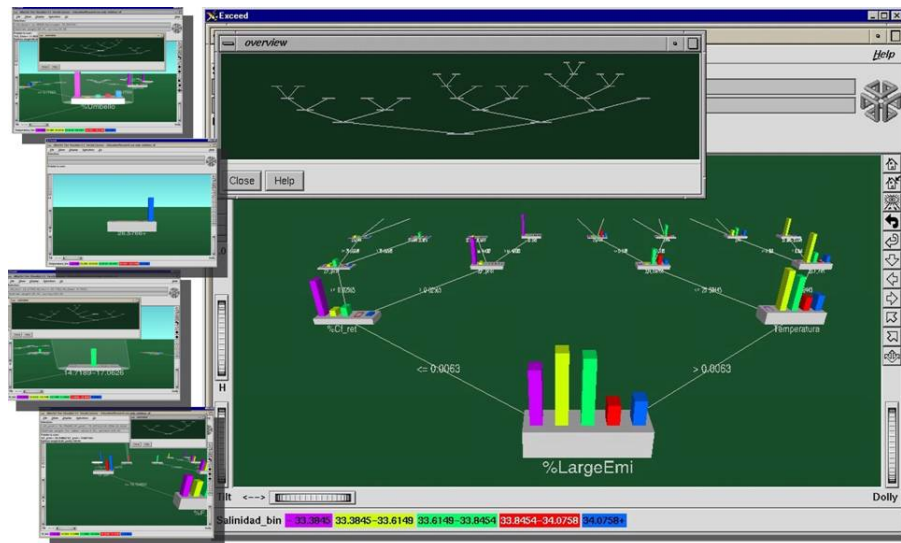


Figura 2. Uso de árboles de decisión para encontrar un índice de estratificación de las aguas

En la ventana más grande de la figura 2 se ve este proceso para el caso de la salinidad. Previamente fue necesario establecer 5 niveles de salinidad (muy baja, baja, normal, alta y muy alta). Entonces se aplicó el algoritmo clasificador (eliminando la variable temperatura) para cada MIS. En la figura se puede ver cómo una proporción pequeña de *Large Emiliana* (primera rama del árbol a la izquierda), para el MIS 2, es una señal de baja salinidad (sólo está presente la primera columna a la izquierda en cada nodo).

Similarmente se realizó este proceso para el caso de la temperatura (eliminando la salinidad). A la vista de las reglas proporcionadas por los árboles, se pudo establecer el índice N de estratificación de la columna de agua (1).

$$N = \frac{\sum(\text{placolitos} < 3\mu\text{m})}{F.\text{profunda} + \sum(\text{placolitos} < 3\mu\text{m})} \quad (1)$$

Finalmente se repitió el procedimiento para analizar cómo eran afectadas las especies por el índice N (eliminando temperatura y salinidad). Estos resultados, junto con los procedentes de las técnicas de visualización aplicadas (se comentan en la sección siguiente), permitieron a los micropaleontólogos reconstruir la dinámica oceánica del pasado.

3.2. Reconstrucción de Paleovariables

A continuación se presenta un ejemplo de reconstrucción de temperaturas (SST) del pasado de un sitio en particular, a partir de muestras modernas tomadas en diferentes puntos geográficos bajo distintas condiciones medioambientales.

El primer intento de encontrar una relación empírica entre el clima del pasado y componentes de un sedimento fue realizado por Imbrie y Kipp [8]. Este trabajo se basaba en la asunción de que la distribución de las especies de foraminíferos planctónicos dependía de los valores de la SST. Posteriormente, Hutson en [9] utilizó aprendizaje basado en casos (más concretamente *K-nearest neighbors*) para encontrar el conjunto de los K puntos geográficos que más se parecen (tienen distribuciones de especies similares) al sitio que se está estudiando en un momento del pasado (un estrato dentro del sedimento), y se obtiene el valor promediado de la temperatura medida en esos K sitios. Hutson denominó a este método *Modern Analog Technique*, MAT.

En el trabajo de Hutson se utilizó la distancia coseno-theta como medida de disimilaridad. Posteriormente, Overpeck [10] realizó un estudio con 8 diferentes medidas de disimilaridad, concluyendo que la medida de distancia *Squared chord* era la mejor elección. Desde entonces se ha convertido en una técnica estándar en Paleoclimatología. En los últimos años se han desarrollado programas como ANALOGS [11], y modificaciones a la técnica como SIMMAX [12], RAM [13] y, más recientemente, se han utilizado redes neuronales [14] para estimar la SST. Ninguna de las citadas ha superado la popularidad de MAT.

PaleoAnalog [15] es una aplicación multiplataforma que implementa MAT. El objetivo de esta herramienta era proporcionar a los geólogos la técnica estándar

para la reconstrucción cuantitativa de paleovariables, haciendo especial énfasis en la facilidad de uso a la vez que intentaba involucrar al usuario en la reconstrucción, intentando eliminar el efecto *caja-negra*, además de permitir validar o afinar las reconstrucciones.

El proceso de reconstrucción comienza mediante la selección de los ficheros que contienen los datos modernos (*database*) y los del testigo que se quiere reconstruir (*core*). Generalmente, estos dos ficheros contienen diferentes taxones (figura 3.a y 3.b), debido a que proceden de diferentes investigadores y/o laboratorios que utilizan diferentes categorías (especies y subespecies), nombres y abreviaturas.

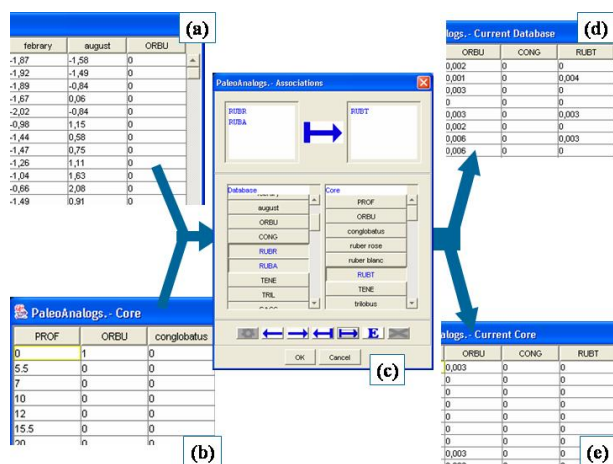


Figura 3. Asistente para la asociación taxonómica

Para poder calcular las distancias, es necesario que las variables en ambos conjuntos de datos sean comparables. Con la ayuda del asistente de asociaciones taxonómicas (3.c) se realiza esta tarea fácilmente. Además, se calculan las proporciones de cada especie y se determina el conjunto de paleovariables a reconstruir. Una vez que ambos conjuntos de datos son comparables (tienen taxones equivalentes, 3.d y 3.e) se selecciona una de las medidas de disimilitud estudiadas en [10] y finalmente se realiza la reconstrucción en función de los K análogos (K vecinos).

Es en este punto en donde PaleoAnalog ofrece su ampliación al método MAT y característica diferenciadora. Es posible realizar una análisis interactivo (parte del cual se comenta en la siguiente sección) por parte del usuario, de forma que puede conocer exactamente qué muestras modernas se han utilizado para reconstruir las variables de cada muestra del testigo. De esta manera, a partir de ahora el proceso de reconstrucción está dirigido por el experto.

Se debe tener en cuenta que la técnica MAT clásica sólo permitía elegir el número K de análogos (vecinos), normalmente 10, que se promediaban para reconstruir una variable. Pero podía darse el caso de que sólo 3 de estos 10 análogos modernos se parecieran realmente a la muestra que se estaba calculando, mientras que los 7 restantes fueran solamente los siguientes más parecidos, y, por tanto, sería un error utilizarlos para la reconstrucción desde el punto de vista geológico.

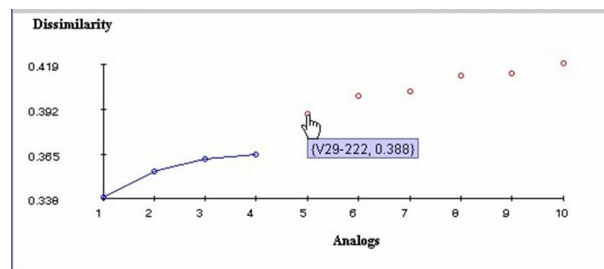


Figura 4. Validación/Afinación de la reconstrucción

La figura 4 muestra la herramienta de análisis interactivo de PaleoAnalog, que permite para cada muestra (por ejemplo, la correspondiente al centímetro 7 del testigo que, según la tasa de sedimentación corresponderá a una edad concreta), observar cuáles han sido los sitios más cercanos (en la figura, el quinto análogo se corresponde al sitio V29-222, con un valor de disimilitud de 0.388). En este ejemplo se ha considerado que el salto entre el cuarto y el quinto análogo no es aceptable; interactivamente, pulsando en cada punto de la gráfica se selecciona el análogo, de forma que se recalcula el promedio, sin tener en cuenta los casos desestimados por el usuario.

El método MAT clásico inherentemente produce reconstrucciones cuya precisión es difícil de estimar[16]. Gracias al método de validación interactiva, al introducir la experiencia del usuario, gran parte de esta incertidumbre puede ser despejada.

4. Visualización de Paleodatos

A continuación se exponen brevemente las herramientas de visualización utilizadas en los dos casos contemplados en la sección anterior.

4.1. Visualización de la Dinámica Oceánica

La herramienta Mineset permite realizar representaciones tipo *Splat*, una agregación espacial que muestra la distribución de las muestras. Ésta puede

incluir además de las tres dimensiones espaciales, otras codificaciones de variables como la opacidad o la animación. Para el problema de la dinámica oceánica se eligió representar (figura 5) las tres variables temperatura-salinidad-N en los tres ejes, utilizando la animación para ver la evolución temporal (a lo largo de los diferentes estados isotópicos marinos, MIS). Esto permitió encontrar relaciones temporales que facilitaron a los geólogos la elaboración de un modelo de la dinámica oceánica del Mar de la China. Sin embargo, es conveniente señalar que a estos les resultaba muy dificultoso interpretar la animación, por lo que se troceó en las instantáneas que se muestran en la figura.

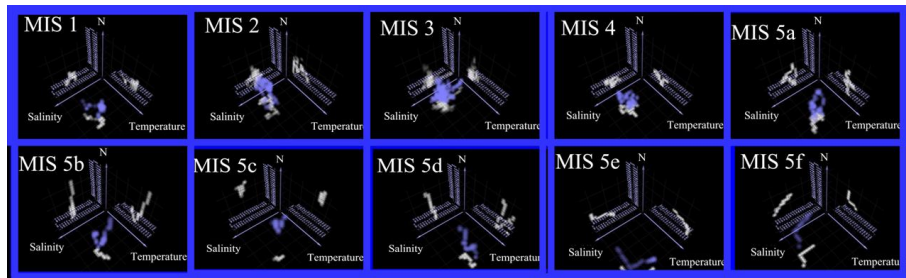


Figura 5. Temperatura-Salinidad-N a lo largo de los MIS

4.2. Visualización de Análogos

Como se ha comentado antes, la herramienta PaleoAnalog se diseñó con la intención de involucrar al usuario de la herramienta en las reconstrucciones. Una de las características más destacables de esta aplicación es la posibilidad de realizar un análisis interactivo de la distribución de los análogos para cada muestra estudiada del testigo (figura 6).

El análisis es muy versátil de forma que se pueden representar tres dimensiones: eje X, eje Y, y el color para la disimilaridad (se ha elegido apropiadamente [6] una gama azul-rojo, de menor a mayor disimilaridad). Es posible elegir cualquier variable reconstruida en cualquiera de los dos ejes; asimismo, se puede representar cualquier especie, también en cualquiera de los dos ejes. El número de análogos representados es seleccionable por el usuario, así como la etiqueta del sitio de cada análogo. La combinación de todos estos elementos permiten adquirir mucha información sobre el testigo que se está estudiando.

Por ejemplo, en la figura 6, se ha elegido mostrar los análogos para la muestra centímetro 20 del testigo. Se está representando latitud frente a longitud, junto con la disimilaridad de los análogos. De esta forma, el experto comprueba que para esa muestra, hace x años (20cm corresponderá a x años según la tasa de sedimentación del lugar de donde se extrajo el testigo), en ese lugar la distribución de especies era muy similar a la de los sitios polares de la actualidad (en la

figura, el círculo punteado engloba los análogos de color azul, es decir, de menor disimilitud; estos coinciden en una zona de latitudes polares).

Mediante un estudio similar combinando las diferentes posibilidades: cada especie frente a la temperatura, temperatura frente a latitud, reduciendo la representación al número de análogos que se van a utilizar para reconstruir (K), etc., y junto con el ajuste posterior de la reconstrucción, se obtienen rápidamente valores de paleovariabes para todo el testigo que son, a la vez, más precisos y más fiables desde el punto de vista geológico.

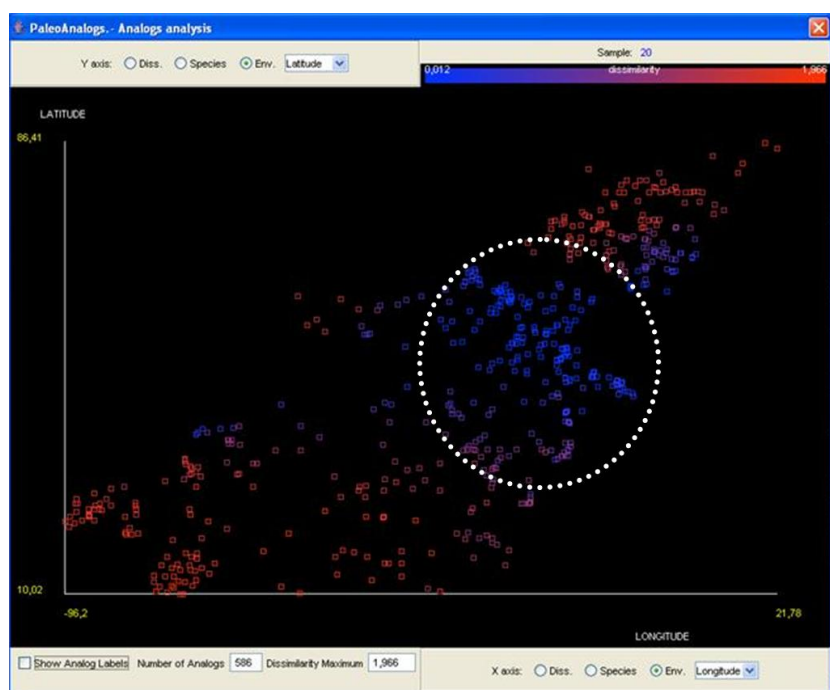


Figura 6. Temperatura-Salinidad-N a lo largo de los MIS

5. Conclusiones y Trabajos Futuros

Este trabajo es un ejemplo de cómo la unión de la Minería de Datos y de la Visualización interactiva pueden ayudar en el descubrimiento de conocimiento para el campo de la Paleoclimatología. Los resultados obtenidos por el grupo de investigación han demostrado cómo la involucración del usuario en los procesos de análisis favorecen la adopción de estas técnicas que en ocasiones son de difícil aceptación por comunidades inicialmente reticentes. Actualmente estamos trabajando en la incorporación de visualizaciones más avanzadas como es el análisis

temporal y la representación geográfica de análogos, así como otras técnicas en principio más complejas, como es el caso de las Coordenadas Paralelas.

Referencias

1. Two Crows, Introduction to data mining and knowledge discovery, 3rd edition, Two Crows Corporation, 1999.
2. M. Steinbach, P. N. Tan, V. Kumar, S. Klooster, C. Potter. Discovery of Climate Indices using Clustering, in Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining , (2003) 446-455
3. R. Vautard and M. Ghil, Singular spectrum analysis in nonlinear dynamics, with applications to paleoclimatic time series, in *Physica*, **35** (1989) 395-424
4. M. Diepenbroek, H. Grobe, M. Reinke, U. Schindler, R. Schlitzer, R. Sieger, G. Wefer, PANGAEA-an information system for environmental sciences, in *Computers and Geosciences*, **28** (2002) 1201-1210
5. Inselberg, A.: Visualization and knowledge discovery for high dimensional data Proceedings. Second International Workshop on User Interfaces to Data Intensive Systems. UIDIS 2001 (2001) 5-24
6. Spence, R. Information Visualization. ACM Press, Addison-Wesley, (2001)
7. Therón R., Flores, J. A., Sierro, F. J., Pelejero, C., Grimalt, J. and Vaquero: M. Using Data Mining and Visualization Techniques for the Reconstruction of Ocean Paleodynamics. Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, **IV** (2002) 2382-2384
8. Imbrie, J. and Kipp, N.G.: A new micropaleontological method for paleoclimatology: Application to a Late Pleistocene Caribbean core. *The Late Cenozoic Glacial Ages*. New Haven, Yale University Press, (1971) 71-181
9. Hutson, W. H.: The Agulhas Current during the Late Pleistocene: Analysis of modern faunal analogs. *Science*, **207** (1980) 64-66
10. Overpeck, J.T., Webb, T. and Prentice, I.C.: Quantitative interpretation of fossil pollen spectra: Dissimilarity Coefficients and the method of modern analogs. *Quaternary Research*, **23** (1985) 87-108
11. Schweitzer, P. N.: ANALOG: A program for estimating paleoclimate parameters using the method of modern analogs. U. S. Geological Survey Open-File Report **94-645** (1994)
12. Pflaumann, U., Duprat, J., Pujol, C. and Labeyrie, L.: SIMMAX: A modern analog technique to deduce Atlantic sea surface temperatures from planktonic foraminifera in deep-sea sediments. *Paleoceanography* **11** (1996) 15-35
13. Waelbroeck, C., Labeyrie, L., Deplessy, J.-C., Guoit, J., Labracherie, M., Leclaire, H., and Duprat, J. Improving past sea surface temperature estimates based on planktonic faunas. *Paleoceanography*, **13**(1998) 272-283
14. Malmgren, B. A., Kucera, M., Nyber, J. And Waelbroeck, C.: Comparison of statistical and artificial neural network techniques for estimating past sea surface temperatures from planktonic foraminifer census data. *Paleoceanography* **16** (5) (2001) 520-530
15. Therón, R., Paillard, D., Cortijo, E., Flores, J.A., Vaquero, M. and Sierro, F. J., Waelbroeck, C. Data-mining the past environment, in proceedings of the IEEE International Geoscience and Remote Sensing Symposium, **6** (2003) 3688-3690
16. Mannila, H., Toivonen, H. Korhola, A. and Olander, H. Learning, mining or modeling? A case study from paleoecology. *Discovery Science*, (1998) 12-24