

Exploratory data analysis through the integration of visualization techniques in Geographical Information Systems

**Romel Vázquez-Rodríguez¹, Carlos Pérez-Risquet¹,
Juan Carlos Torres²**

¹Centro de Estudios de Informática, Universidad Central "Marta Abreu" de Las Villas.
Carretera a Camajuaní, Km 5 ½. Santa Clara, Villa Clara, Cuba. CP: 54830.
Teléfono 281515. romel@uclv.edu.cu, cperez@uclv.edu.cu

²Laboratorio de Realidad Virtual, Universidad de Granada. jctorres@ugr.es

Abstract

Exploratory data analysis constitutes a challenge for scientists and analysts of multiple disciplines. Most current methods and tools do not allow for analyzing effectively several variables simultaneously. This paper addresses the problem of visual analysis of temporal sequences of multiple variables using multiparameter visualization techniques in geographical information systems. The concepts and methods developed here were implemented as an extension to gvSIG GIS, thus demonstrating the feasibility of this approach for extracting information and knowledge from the original data. A case study with global climate data containing 9 variables with 1224 monthly records from 1901 to 2002 is exposed. Information is given for the 67,420 cells that correspond to the Earth's land surface, where each cell occupies an area of 25km by 25km. Also, a module for handling scientific data formats was developed for gvSIG that allow storing all information in compact form for effective reading.

Keywords: Visual data mining; information visualization; geographical information systems.

Análisis exploratorio de datos mediante la integración de técnicas de visualización en sistemas de información geográfica

Resumen

El análisis exploratorio de datos constituye un reto para científicos y analistas de diferentes disciplinas. Muchos de los métodos y herramientas existentes en la actualidad no permiten realizar un análisis visual efectivo de múltiples variables a la vez. En este trabajo se aborda la problemática del análisis visual de secuencias temporales de múltiples variables, utilizando la combinación de técnicas de visualización de datos multiparamétricos en Sistemas de Información Geográfica. Los conceptos y métodos elaborados en este trabajo se implementaron como una extensión al Sistema de Información Geográfica gvSIG, demostrándose así la viabilidad de este enfoque para la extracción de información y conocimiento a partir de visualizaciones de los datos originales. Se muestra un caso de estudio con datos climáticos mundiales que contienen 9 variables con 1224 registros mensuales de 1901 a 2002. La información está dada para las 67420 celdas que corresponden a la superficie terrestre del planeta, donde cada celda ocupa un área de 25km por 25km. Se desarrolló además un módulo para la manipulación de formatos de

datos científicos en gvSIG, que permitió el almacenamiento de toda la información de forma compactada para una lectura efectiva.

Palabras clave: minería visual de datos, visualización de información, sistemas de información geográfica.

1. Introducción

La geovisualización es un campo emergente, que está basado en la integración de muchas disciplinas como la cartografía, la visualización científica (VisCi), el análisis de imágenes, la visualización de información, el Análisis Exploratorio de Datos (EDA) y la ciencia de los Sistemas de Información Geográfica (SIG) para proporcionar la teoría, los métodos y las herramientas para la exploración visual, análisis, síntesis y presentación de datos que contengan información geográfica [1-3]. El análisis visual de datos es un nuevo enfoque, que integra tanto la percepción humana como los métodos computacionales automáticos, lo que permite una mejor comprensión y análisis de grandes y complejos conjuntos de datos [4].

La integración de técnicas de visualización científica en SIG es una idea innovadora que combina las ventajas y fortalezas para el análisis de datos de los dos enfoques. Esta es un área interesante de investigación dentro de la geovisualización [5].

Cuando se analizan grandes volúmenes de datos en series temporales, es común encontrarse con varios problemas: las secuencias no pueden ser analizadas con simples imágenes de una variable, debido a que se requiere tener en cuenta el tiempo y el solapamiento de todos los mapas provoca que se tengan que analizar en secuencias de imágenes o animaciones. Los métodos y herramientas actuales no permiten el análisis espacio-temporal de múltiples variables a la vez.

En esta investigación se realiza la visualización de series temporales de múltiples variables mediante la utilización de técnicas de visualización de datos multiparamétricos (TVDM) integradas dentro de un SIG. Los datos fueron suministrados por la unidad de investigaciones climáticas de la Universidad de East Anglia, en el Reino Unido [6]. El uso de las herramientas desarrolladas en esta investigación posibilitó presentar un caso de estudio con estos datos donde se demuestra la efectividad de las técnicas para extraer tendencias, correlaciones, y patrones.

2. Materiales y métodos

2.1. Antecedentes de la integración VisCi-SIG

Desde la década del 90 y aún en la del 2000 los desarrolladores de herramientas de VisCi y SIG han realizado grandes esfuerzos para extender e integrar sus sistemas [7]. Por una parte, los programadores de sistemas de VisCi y visualización de información (VI) han construido lectores de datos que soportan los formatos de datos espaciales como los modelos digitales de elevación (MDE) y los formatos de SIG comerciales [7]. De igual manera, se han desarrollado intentos de integrar algunos de los formatos de datos científicos más comúnmente utilizados en la VisCi, –como *Common Data Format* (CDF), *Network Common Data Format* (NetCDF), *Hierarchical Data Format* (HDF), y *Flexible ImageTransportSystem* (FITS)–, con SIG; por ejemplo actualmente la suite de ArcGIS permite la manipulación de algunos de estos formatos, como NetCDF y HDF [8].

Algunos de los primeros intentos de integración entre los SIG y la VisCi se han materializado a través de herramientas como *GeoVista Studio* [9-11] y *Snap-TogetherVisualization* [12]. La mayoría de los paquetes se dedican a la visualización de datos espacio-temporales se enfocan más hacia las visualizaciones que hacia la integración con el SIG. Utilizan los SIG de manera separada para preparar datos, exportar e importar datos y finalmente hacer las visualizaciones con una herramienta determinada. *VIS-STAMP* [13] y *GAV Flash* [14] son ejemplos de integración más recientes. Otro de los intentos de integración entre estas disciplinas se puede ver en [15], donde se presenta una herramienta que integra un grupo de TVDM en un SIG para el análisis de datos con baja densidad espacial. Las técnicas implementadas en este trabajo fueron coordenadas paralelas, gráficos de Andrews, *starfield*, *shapecoding*, segmentos de círculo y patrones recursivos. Además estas técnicas fueron utilizadas para el análisis de

variables meteorológicas asociadas a pocas estaciones de medición.

En el presente trabajo se implementan otras TVDM, que son integradas a un nuevo módulo de un SIG. Con el objetivo de validar el método propuesto para el análisis visual de múltiples variables a la vez, almacenadas en series temporales con una alta densidad espacial, se implementaron las técnicas: series temporales, matrices de diagramas de dispersión, iconos en forma de barras, rueda de tiempo, parcelas múltiples y mapas auto-organizados de conjunto con coordenadas paralelas. Otro aspecto novedoso del nuevo módulo es la posibilidad de manipular grandes volúmenes de datos almacenados en formatos de datos científicos.

2.2. Técnicas de visualización de datos multiparamétricos

Existen una serie de problemas en que cada punto de dato contiene más de un atributo. A este tipo de datos se les llama multiparamétricos y se encuentran generalmente en aplicaciones de minería de datos, estadística y la inteligencia artificial [16]. Los datos multiparamétricos, también llamados multidimensionales o datos n-dimensionales, estos datos pueden ser vistos como una matriz de $n \times d$, donde cada fila representa un registro y cada columna representa una observación, variable o dimensión [17]. Las TVDM pueden ser clasificadas en geométricas, basadas en iconos, basadas en píxel y proyecciones [16, 18]. Una descripción más detallada de estas técnicas de visualización se pueden observar en [15].

Existen otras formas de visualizar datos que permitan extraer información relevante de estos. Ejemplo de estas son las técnicas basadas en ejes como la Rueda de Tiempo (*Time Wheel*) [19, 20] (obsérvese Figura 1A): su idea

básica es presentar el eje de referencia (tiempo en este caso) en el centro de la pantalla, y circularmente organizar las demás variables como ejes alrededor del eje temporal trazando líneas de los valores correspondientes para una variable hacia el valor correspondiente en el eje temporal. Parcelas Múltiples (*MultiComb*) [21] (obsérvese Figura 1B): el objetivo básico de esta técnica es alinear los datos por parcelas de manera circular como si fueran gráficos de tiempo sobre cada eje. Otro ejemplo es la visualización de Río Temático (obsérvese Figura 1C) que es de gran utilidad para ver la variación de datos en una gran colección de información temporal. Los cambios son mostrados en el contexto de una línea de tiempo, permitiéndole al usuario discernir patrones más fácilmente y analizar la relación entre los datos [22].

Los mapas auto-organizados (obsérvese figura 1D) o redes neuronales de Kohonen (SOM por sus siglas en inglés, Self-Organizing Map) fueron introducidos por el profesor finlandés Teuvo Kohonen [23-25]. Un Mapa Auto-Organizado es una herramienta que analiza datos de muchas dimensiones con relaciones complejas entre ellas y los presenta en una visualización sencilla en sólo dos dimensiones [25]. La propiedad más importante de SOM es que preserva las propiedades topológicas de los datos, es decir, que datos parecidos aparecen cercanos en la visualización. Este tipo de red posee un aprendizaje no supervisado competitivo [24]. La red auto-organizada debe descubrir patrones regulares, elementos comunes, interrelaciones o categorías en los datos de entrada y adicionarlos a su estructura interna de conexiones [26]. Martín-Del-Brío y Sanz-Molina en [26] dicen que “en este caso que las neuronas deben autoorganizarse en función de estímulos (señales o datos) procedentes del exterior”.

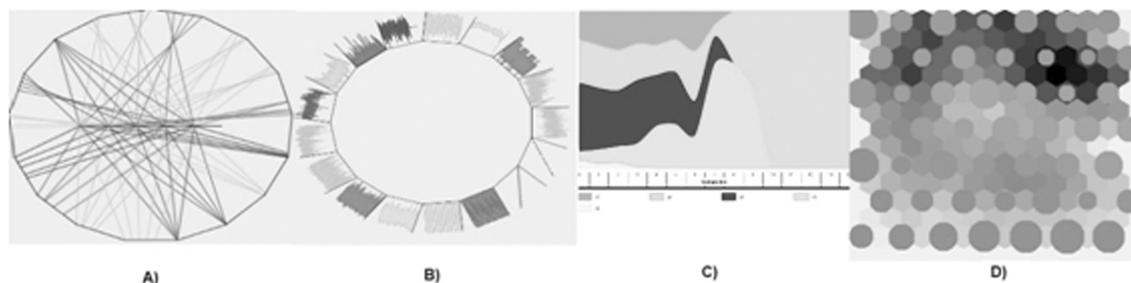


Figura 1. Otras técnicas de visualización de datos multiparamétricos.

3. Solución propuesta

En este artículo se propone un nuevo método para analizar de forma interactiva, mediante visualizaciones de múltiples variables a la vez, datos espaciales almacenados en series temporales. Esto se lleva a cabo usando algunas TVDM como las mencionadas anteriormente: estas se utilizan sobre una región para analizar datos que están almacenados en las series temporales. Los SIG no están diseñados para trabajar de forma óptima con grandes conjuntos de información, por lo tanto se quiere modificar la forma en que se almacenan y utilizan los datos para poder visualizarlos.

3.1. Arquitectura general

La Figura 2A muestra la arquitectura general propuesta en este trabajo. Se propone la implementación de dos nuevos módulos a un SIG de código abierto: un módulo de visualización científica con todas las posibilidades de mostrar diferentes técnicas e interactuar con ellas, y un módulo para la lectura de grandes volúmenes de datos en formatos de datos científicos, como puede ser el caso de HDF. El usuario tiene la posibilidad de seleccionar las técnicas de visualización que desee usar, para luego mostrar los datos asociados a las regiones del mapa que sean de su interés, permite escoger cuál de las técnicas seleccionadas será mostrada sobre el mapa de fondo y cuáles estarán en paneles independientes. Esta arquitectura permite al usuario mostrar con las mismas técnicas de visualización los datos asociados a distintas localidades del mapa, y así poder analizar los re-

sultados para obtener conclusiones, o corroborar hipótesis.

3.2. Interacción del usuario con el sistema

La aplicación funciona de la siguiente forma: El usuario puede seleccionar la técnica o las técnicas que desee visualizar a la vez y se le permite escoger cuál de estas desea mostrar sobre el mapa, las demás técnicas seleccionadas son mostradas con los mismos datos pero en paneles independientes para que no ocurra solapamiento de imágenes sobre el mapa de fondo. La Figura 2B muestra el diagrama de interacción entre el usuario y el software en una sesión de trabajo.

Una vez seleccionadas y preparadas las técnicas de visualización, el usuario selecciona sobre un mapa de fondo las localizaciones que desea analizar; en este punto se cargan desde el formato de datos científicos los datos correspondientes a las selecciones consultadas, los que se visualizan utilizando las técnicas de visualización previamente seleccionadas por el usuario. Este mecanismo se puede considerar una excelente herramienta de exploración de datos, ya que permite analizarlos con múltiples técnicas de forma interactiva, permitiendo la exploración visual de varias variables en todo el espacio de información.

La interfaz de la selección de las técnicas de visualización científica se muestra en la Figura 3A. Se pueden seleccionar varias técnicas a la vez, de las cuales una se puede visualizar sobre el mapa de fondo. La selección de las localizaciones

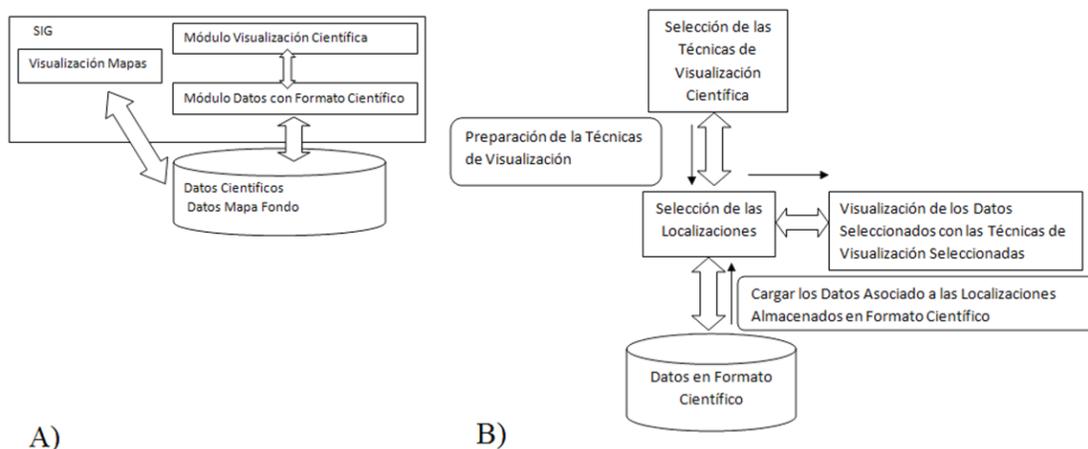


Figura 2. A) Arquitectura general. B) Esquema general de interacción del usuario con el sistema.

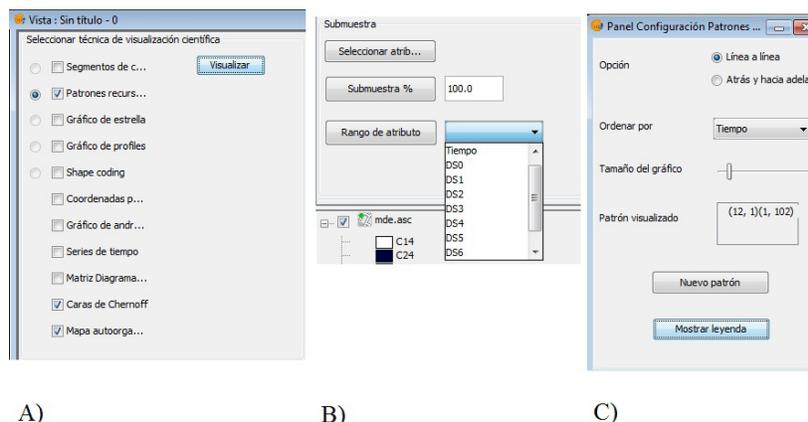


Figura 3. A) Selección de las técnicas de visualización científica. B) Panel de configuración general de todas las técnicas. C) Panel de configuración particular de la técnica Patrón Recursivo.

se realiza con el ratón de forma interactiva sobre una región del mapa de fondo, luego de seleccionar las técnicas y oprimir el botón visualizar de la Figura 3A). Cada técnica de visualización posee un panel de configuración general (Figura 3B), que permite seleccionar los variables involucradas en el análisis. Cada una de las técnicas tiene su configuración particular. La Figura 3C muestra la configuración para la técnica patrones recursivos. La operación de cargar datos se hace de manera automática, una vez seleccionadas las regiones de interés para el análisis sobre el mapa de fondo. Ejemplos de las interfaces de las operaciones de visualización se pueden observar en las Figuras 4 y 5.

3.3. Implementación

El SIG de escritorio seleccionado para ser extendido con el nuevo módulo de VisCi fue gvSIG y el formato de dato científico utilizado para manipular la información fue HDF. La versión actual del sistema incluye las técnicas: Coordenadas paralelas, gráficos de Andrews, series temporales y matrices de diagramas de dispersión (todas ellas en paneles independientes); *starfield*, *shapecoding* y (todos los registros juntos en paneles independientes); iconos en forma de barras (un registro a la vez); segmentos de círculos, patrones recursivos, rueda de tiempo y parcelas múltiples (todos los registros seleccionados sobre el mapa y en paneles independientes); y mapas auto-organizados de conjunto con coordenadas paralelas (en paneles independientes).

En esta investigación todas las visualizaciones son normalizadas teniendo en cuenta los valores globales extremos de cada variable para todos los puntos de datos, las variables son transformadas a valores dentro del intervalo (0,1) con el objetivo de que puedan ser comparables. Existen infinitas funciones que permiten realizar esto, sin embargo algunas pueden favorecer a determinados valores. Debido a que las variables fueron registradas en iguales intervalos de tiempo, lo más razonable es utilizar la función de normalización lineal $(val-min)/(max-min)$; así, variables que tengan dominios numéricos diferentes, pueden ser analizadas conjuntamente.

4. Caso de estudio

Como se mencionó anteriormente, la solución propuesta para el problema planteado fue evaluada mediante un caso de estudio que involucra los datos que fueron utilizados y las herramientas desarrolladas para visualizar los datos.

4.1. Datos climáticos

En esta investigación se trabajó con el conjunto de datos CRU (*Climatic Research Unit*) TS (*Time Series*) 2,1 [6]. Este comprende 1224 registros mensuales de 9 variables climáticas en el período 1901-2002 y cubre la superficie mundial con una resolución de 0,5 grados.

Como uno de los resultados de esta investigación quedó creado un fichero HDF, llamando CRU2,1.h5 que almacena toda la información

descrita anteriormente. De esta forma se logró compactar la información de 3,5GB a 910MB. La estructura del archivo es la siguiente: 9 conjuntos de datos, uno para cada una de las variables, cada conjunto de datos es una matriz de 67420 filas por 1224 columnas, representando las filas, a cada una de las celdas de que se dispone información y las columnas los 1224 valores mensuales de las variables en cada celda desde 1901 hasta el 2002. Cada uno de estos conjuntos de datos se comprimió con el algoritmo *gzip* utilizando valor de 9, el máximo valor de compresión y los trozos de escritura que se utilizaron fueron de 16855 y 306 para cada dimensión respectivamente. Además se creó un conjunto de datos de 720 por 360 que coincide con las celdas del MDE utilizado como fondo y que contiene en cada escaque el número de la celda que le corresponde de donde hay información. De esta forma esta matriz se puede leer y tenerse un acceso directo a los datos de cada una de las celdas.

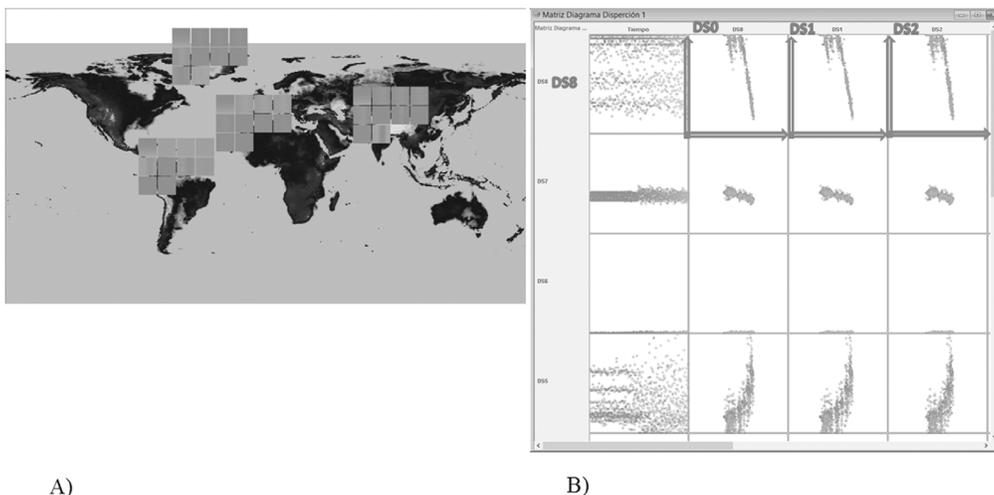
El proceso de lectura de los datos dada una petición en una celda realizada por un usuario es el siguiente: abrir el conjunto de datos que almacena los índices (Matriz de 720 por 360) y cargarlo en memoria. Al acceder a esta matriz en (fila, columna), se devuelve el número de la celda si contiene información, si no contiene información el valor en el escaque dado es 0. Luego se abre cada uno de los conjuntos de datos que almacenan las variables climáticas y se lee sólo los 1224 valores de cada conjunto de datos, accediendo por

las filas al número de celda devuelto restándole 1 y leyendo en cada conjunto de datos los 1224 valores de la fila correspondiente, de esta forma se seleccionan 9×1224 valores que luego son utilizados para crear el *VisualizationDataSource* (objeto que contiene la fuente de datos para la visualización) que es pasado a cada técnica que se visualice sobre la celda correspondiente en el mapa de fondo.

Los conjuntos de datos nombrados con DS# (obsérvese Figura 4B), contienen los valores de las variables para cada una de las celdas que corresponden a las zonas terrestres del planeta. La correspondencia de las 9 variables climáticas con el nombre de cada conjunto de datos es: el promedio mensual de la temperatura diaria media (DS0), mínima (DS1) y máxima (DS2), nubosidad (DS3), precipitación (DS4), promedio mensual de la frecuencia de humedad diaria (DS5), presión de vapor (DS6), rango de temperatura diurna (DS7) y promedio de la escarcha diaria (DS8).

4.2. Validación del sistema. Detectar lo esperado

Con el objetivo de ejemplificar el uso del sistema se seleccionaron como técnicas a visualizar: matrices de diagramas de dispersión y patrón recursivo. Se tomaron cuatro puntos contrastantes: uno sobre el Himalaya, otro sobre el desierto del Sahara, y otros dos sobre la cuenca del Amazonas y Groenlandia respectivamente. Mediante una



A) Visualización de diferentes puntos utilizando Patrón Recursivo.
B) Gráfico de matrices de diagramas de dispersión asociado al Himalaya.

herramienta que permite modificar el tamaño del gráfico se aumentaron todos los gráficos para analizar mejor la información de los cuatro puntos en el espacio y el tiempo. En la Figura 4A) se muestran los resultados. Patrón recursivo se utilizó con el patrón (12,1) (1,102) que dispone los datos de los 12 meses de izquierda a derecha y cada año de arriba hacia abajo. La variable de la parte superior izquierda es una variable adicional que representa el orden temporal; se puede observar que en los cuatro gráficos la variable es igual. Se utiliza el color anaranjado fuerte para representar valores altos de la variable (en escala de grises corresponde con tonos claros) y azul oscuro para representar valores bajos (en escala de grises corresponde con tonos oscuros), el resto de los valores se representa con un gradiente entre azul y anaranjado (en escala de grises corresponde con tonos intermedios).

Realizando un análisis visual con Patrón Recursivo: se puede ver claramente cómo el gráfico sobre el Himalaya y el de Groenlandia poseen valores más bajos durante todo el año para las variables, temperatura diaria media (DS0), mínima (DS1) y máxima (DS2) que el gráfico sobre el desierto de Sahara y sobre el Amazonas, sin embargo en el gráfico sobre el Amazonas, la humedad (DS5), y la presión de vapor de agua (DS6) posee valores más altos de esas variables que todas las demás regiones analizadas. Se puede observar además cómo el promedio de escarcha diaria (DS8), la última de las variables, es más intensa en Groenlandia y en el Himalaya que los otros dos lugares, notándose una disminución considerable de los valores de esta variable en el Himalaya en los meses más calurosos del año para la zona norte del planeta, este patrón no ocurre tan pronunciado en Groenlandia.

Por otra parte utilizando matrices de diagramas de dispersión para mostrar relaciones entre las variables que pertenecen a una misma localización, por ejemplo con los datos del punto seleccionado sobre el Himalaya, se pueden observar claramente como hay una correspondencia entre los valores de temperatura (ya sean medios, mínimos o máximos) y el promedio de escarcha diaria (Obsérvese Figura 4B). Centrándonos en la celda donde se grafica DS8 x DS0, se observa que a medida que aumenta la temperatura por el eje x (de izquierda a derecha como muestra la flecha),

disminuye la escarcha por el eje y. Este es por supuesto es un resultado esperado, ya que a menor temperatura mayor será el nivel de congelación.

4.3. Validación del sistema. Descubrir lo inesperado

El sistema fue utilizado para analizar los datos correspondientes a Cuba, para eso se focalizó en el mapa de fondo la zona de la isla de Cuba. Se seleccionaron como técnicas a visualizar patrón recursivo y mapas auto-organizados. En este caso se utilizó patrón recursivo con el mismo patrón anterior (12,1) (1,102) que dispone los datos de los 12 meses de izquierda a derecha y cada año de arriba hacia abajo. Se escogieron tres puntos representativos del territorio, uno en oriente (zona alta), otro en el centro (llanura) y el último en occidente (zona baja) (Figura 5A).

La variable promedio de escarcha diaria (DS8) tiene un comportamiento interesante, porque en Cuba nunca ha nevado. Se puede ver en los gráficos de patrón recursivo asociados a las regiones de oriente y centro no hay valores significativos de escarcha (última variable) en estas regiones, pero en la región de occidente se observa un pequeño incremento de los valores de esta variable en el mes de enero. Como los valores de escarcha en esa localización son tan bajos comparados con el máximo global no se puede realizar un buen análisis con esta técnica, por esta razón se utilizó mapas auto-organizados, para intentar detectar más claramente el incremento que apenas se distingue con patrón recursivo en escarcha diaria (DS8) del punto seleccionado en occidente.

La técnica SOM en el sistema opera enlazada de conjunto con coordenadas paralelas de manera tal que los grupos detectados por la red neuronal pueden ser seleccionados de manera interactiva sobre el gráfico y los elementos pertenecientes a los grupos son resaltados en la técnica de coordenadas paralelas. Con la técnica SOM se le asignó un mayor peso (>2) a la variable escarcha diaria (DS8) y se le quitó peso (<0.05) a la variable tiempo para que el orden temporal no tuviera tanta influencia en la formación de los grupos con la red neuronal, esto es debido a que ese atributo sólo se utiliza para dar un orden lógico al resto de las variables. El resto de las variables les mantuvo el peso por defecto (1). El objetivo de este

análisis es tratar de ver mejor las diferencias que no son tan notables con las otras técnicas. Con el resultado de ejecutar las técnicas de mapas auto-organizados sobre los 3 puntos y seleccionando la variable escarcha (DS8) para colorear los grupos, se generaron los gráficos de la Figura 5B.

Como se puede observar en la Figura 5B los gráficos identificados por los número uno y dos presentan el mismo color en sus grupos (note la tonalidad de gris en los SOM de la parte superior); si se observa en la técnica de coordenadas paralelas asociados a cada uno de los gráficos anteriores, se puede ver que la variación de la variable escarcha es de 0-1, esto significa que no hubo ninguna ocurrencia de valores de escarcha en esos puntos (Oriente y Centro). En cambio en la localización tres existen grupos de coloración diferente (círculos más claros en el SOM de la parte inferior) lo cual significa que poseen valores más altos.

Como se muestra en la Figura 6, los elementos que pertenecen a los grupos seleccionados tienen valores altos de escarcha. Además se observa en la técnica SOM el tono oscuro de fondo que separa los grupos seleccionados de los demás, lo cual expresa la gran diferencia que existe entre los grupos seleccionados y el resto. La variable de cobertura de nubes (DS3 correspondiente al quinto eje vertical de izquierda a derecha en coordenadas paralelas) posee un patrón que llama la atención. Posee valores altos en los meses de invierno en la zona occidental como se muestra en la Figura 6B, algo que no se nota en las regiones del centro y oriente. Este aspecto tiene relación con la entrada de los frentes fríos a Cuba en época de invierno, estos frentes fríos por lo general llegan a la isla por occidente acompañados de mucha nubosidad que luego se va disipando a su paso de Occidente a Oriente.

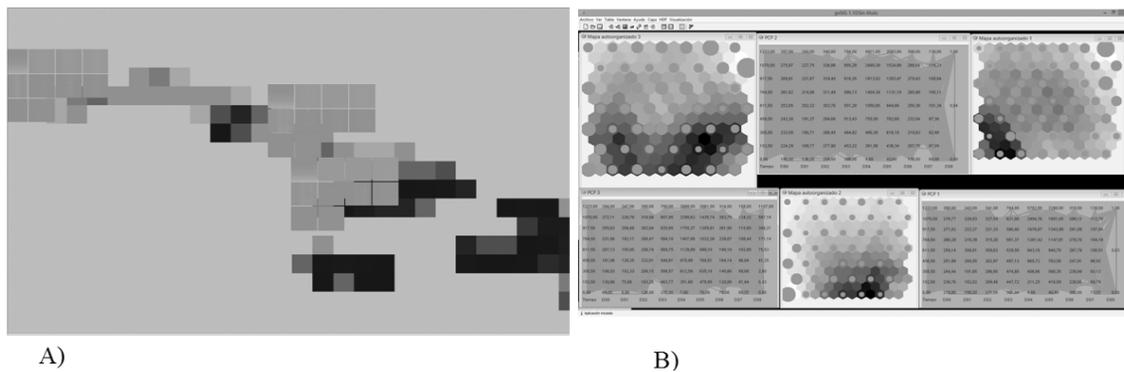


Figura 5. A) Visualización de diferentes puntos del territorio de Cuba utilizando Patrón Recursivo. B) Visualización con SOM y coordenadas paralelas asociados a los mismos puntos del territorio de Cuba.

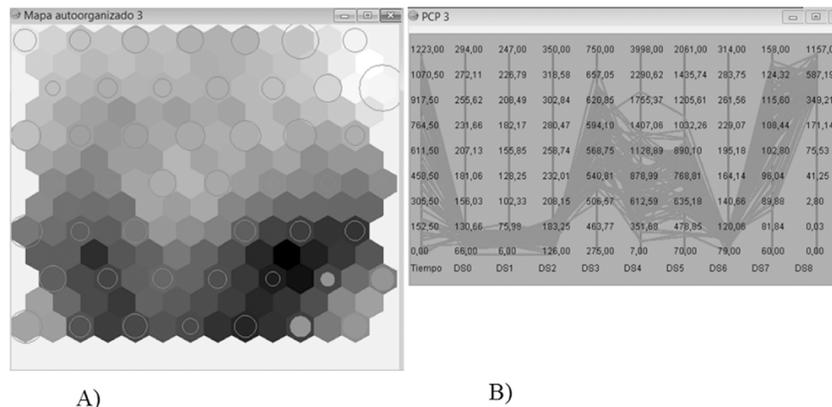


Figura 6. A) Visualización de SOM y B) Coordenadas Paralelas asociadas a un punto del Occidente de Cuba.

5. Conclusiones

En este trabajo se abordó la problemática del análisis visual de secuencias temporales de múltiples variables, utilizando la combinación de TVDM en SIG. Las herramientas fueron implementadas como una extensión al Sistema de Información Geográfica gvSIG, demostrándose así la viabilidad de este enfoque para la extracción de información y conocimiento a partir de visualizaciones. Se mostró un caso de estudio con datos climáticos mundiales, que contienen 9 variables con 1224 registros mensuales de 1901 a 2002, información dada para las 67420 celdas que corresponden a la superficie terrestre del planeta. Se desarrolló además un módulo para la manipulación de formatos de datos científicos en gvSIG, que permitió el almacenamiento de toda la información de forma compactada para una lectura efectiva.

Referencias bibliográficas

1. Dykes, J., A. MacEachren, and M. Kraak, Exploring Geovisualization, Chapter 1, in Exploring geovisualization, Elsevier: Amsterdam, 2005.
2. Kraak, M.: "Visualization viewpoints: beyond geovisualization". IEEE Comput Graph, Vol. 26, No. 4 (2006) 6-9.
3. MacEachren, A. and M. Kraak: "Research challenges in geovisualization". Cartography and Geographic Information Science, Vol. 28, No. 1 (2001) 3-12.
4. Keim, D., *et al.*: "Visual analytics: Scope and challenges". Visual Data Mining, Vol. No. (2008) 76-90.
5. Andrienko, N., G. Andrienko, and P. Gatalsky: "Exploratory spatio-temporal visualization: an analytical review". J Visual Lang Comput, Vol. 14, No. 6 (2003) 503-541.
6. Mitchell, T.D. and P.D. Jones: "An improved method of constructing a database of monthly climate observations and associated high resolution grids". Int J Climatol, Vol. 25, No. 6 (2005) 693-712.
7. Rhyne, T.: "Going virtual with geographic information and scientific visualization". Comput Geosci, Vol. 23, No. 4 (1997) 489-491.
8. Zhao, S., *et al.*: "GDAL-based extend ArcGIS Engine's support for HDF file format". Memorias de: Geoinformatics, 2010 18th International Conference, 2010.
9. Gahegan, M., *et al.*: "Introducing GeoVISTA Studio: an integrated suite of visualization and computational methods for exploration and knowledge construction in geography". Comput Environ Urban, Vol. 26, No. 4 (2002) 267-292.
10. MacEachren, A., *et al.*: "Exploring High-D Spaces with Multiform Matrices and Small Multiples". Memorias de: International Symposium on Information Visualization, Seattle, WA, 2003.
11. Takatsuka, M. and M. Gahegan: "GeoVISTA Studio: a codeless visual programming environment for geoscientific data analysis and visualization". Comput Geosci, Vol. 28, No. 10 (2002) 1131-1144.
12. North, C. and B. Shneiderman: "Snap-together visualization: can users construct and operate coordinated visualizations?". Int J Hum-Comput St, Vol. 53, No. 5 (2000) 715-739.
13. Guo, D., *et al.*: "A visualization system for space-time and multivariate patterns (vistamp)". IEEE T Vis Comput Gr, Vol. 12, No. 6 (2006) 1461-1474.
14. Ho, Q., *et al.*: "A web-enabled visualization toolkit for geovisual analytics". Memorias de: SPIE, the International Society for Optical Engineering: SPIE: Electronic Imaging Science and Technology, Visualization and Data Analysis, 2011.
15. Vázquez-Rodríguez, R., C. Pérez-Risquet, and J.C. Torres-Cantero: "A Novel Visual Data Mining Module for the Geographical Information System gvSIG". Anuário do Instituto de Geociências, Vol. 36, No. 1 (2013) 98-111.
16. Keim, D.: "Information visualization and visual data mining". IEEE T Vis Comput Gr, Vol. 8, No. 1 (2002) 1-8.
17. Ward, M.O.: "Multivariate data glyphs: Principles and practice". Handbook of data visualization, Vol. No. (2008) 179-198.
18. Theisel, H.: "Scientific Visualization". 2000.
19. Tominski, C., J. Abello, and H. Schumann: "Axes-based visualizations with radial lay-

- outs". Memorias de: 2004 ACM symposium on Applied computing, 2004.
20. Aigner, W., *et al.*: "Visual methods for analyzing time-oriented data". Visualization and Computer Graphics, IEEE Transactions, Vol. 14, No. 1 (2008) 47-60.
21. Abello, J. and J. Korn, Mgv: A system to visualize massive multi-digraphs. IEEE Transactions on Visualization and Computer Graphics, Vol. 8 No. 1 (2002).
22. Havre, S., *et al.*, Theme river: Visualizing thematic changes in large document collections. Visualization and Computer Graphics, IEEE Transactions on, 2002. 8(1): p. 9-20.
23. Kohonen, T.: "Self-organized formation of topologically correct feature maps". Biological cybernetics, Vol. 43, No. 1 (1982) 59-69.
24. Kohonen, T.: "The self-organizing map". Proceedings of the IEEE, Vol. 78, No. 9 (1990) 1464-1480.
25. Kaski, S., J. Nikkilä, and T. Kohonen: "Methods for interpreting a self-organized map in data analysis". Memorias de: 6th European Symposium on Artificial Neural Networks (ESANN98). D-Facto, Brugfes, 1998.
26. Martín-del-Brío, B. and A. Sanz-Molina, Redes neuronales y sistemas borrosos. 2001.

Recibido el 7 de Enero de 2014

En forma revisada el 2 de Febrero de 2015