



INTEGRACIÓN DE FORMATOS DE DATOS CIENTÍFICOS EN SISTEMAS DE INFORMACIÓN GEOGRÁFICA

Romel Vázquez Rodríguez⁽¹⁾, Carlos Pérez Risquet⁽¹⁾, Juan Carlos Torres Cantero⁽²⁾, Alexis Gallardo Segura⁽¹⁾

1 Centro de Estudios de Informática, UCLV, Cuba, Carretera a Camajuaní km 5 ½, Santa Clara, Villa Clara, Cuba. E-mail: {romel, cperez, agsegura}@uclv.edu.cu.

2 Departamento de Lenguajes y Sistemas Informáticos, Universidad de Granada, España.

RESUMEN

La integración de formatos de datos científicos en Sistemas de Información Geográfica es una alternativa novedosa para el almacenamiento de grandes volúmenes de datos. Dotar a los Sistemas de Información Geográfica con tales formatos de datos contribuye a un efectivo almacenamiento y gestión de múltiples variables en el tiempo y el espacio. En este trabajo se presentan los principales formatos de datos científicos más utilizados en la visualización científica y cómo se han ido integrando en los SIG. En particular se presenta la integración del formato HDF en el SIG gvSIG y se muestra un caso de estudio con datos climáticos a escala global.

ABSTRACT

The integration of scientific data formats in Geographic Information Systems is an innovative alternative for storing large volumes of data. Equipping GIS with such data formats contributes to effective storage and management of multiple variables in time and space. This paper presents the main scientific data formats used in scientific visualization and how they have been integrated into GIS. In particular presents HDF format integration in GIS gvSIG and shows a case study with global climate data.

INTRODUCCIÓN

El gran volumen de datos espacio-temporales que se genera a diario ha ocasionado que el almacenamiento de dicha información sea un poco costosa, y manipularla también es difícil por lo que es posible mejorar el trabajo de los sistemas SIG con la introducción en estos de nuevos tipos de datos científicos como pueden ser HDF (Hierarchical Data Format), este formato es desarrollado por el Centro Nacional de Aplicaciones de Supercómputo (National Center for Supercomputing Applications, NCSA) en la Universidad de Illinois. Es una biblioteca y un formato de fichero multi-objeto para la transferencia de datos gráficos y numéricos entre máquinas (Fortner, 1998). Otro de los formatos que pueden ser incluido es el CDF (Common Data Format), es un formato de datos auto-descriptivo para el almacenamiento y manipulación de datos escalares y multidimensionales en una forma independiente de la plataforma y el campo de especialidad. A pesar de que CDF tiene su propio formato interno auto-descriptivo, éste consiste en algo más que sólo un formato de datos. CDF es un paquete de administración de datos científicos (conocido como la "Biblioteca CDF") la cual permite a los programadores y desarrolladores de aplicaciones administrar y manipular arreglos de datos escalares, vectoriales y multidimensionales (Atkinson and Hankin, 1995). Y por ultimo The Network Common Data Form o netCDF, es una interfaz a una biblioteca de funciones de acceso a datos diseñada para el almacenamiento y recuperación de datos en forma de arreglos, donde cada arreglo es una



estructura rectangular n-dimensional (con $n = 0,1,2,\dots$) en la que todos sus elementos son del mismo tipo de dato (ejemplo: caracteres de 8-bit, entero de 32-bit) y un escalar (un solo valor) equivale a un arreglo 0-dimensional (Rew and Davis, 1990).

Con estos formatos de datos científicos se puede hacer más potentes los Sistemas de Información Geográfica, con lo cual es posible mejorar el sistema de almacenamiento de los datos de diferentes variables temporales para luego a estas aplicarles diferentes técnicas de visualización científica para datos multiparamétricos. Utilizando estos formatos de datos científicos se puede reducir el espacio destinado al almacenamiento de dichos datos y una mayor organización de estos.

Para lograr esto se introdujo una nueva biblioteca en gvSIG que permite manipular los datos que están en este formato, para de manera independiente del modulo de visualización se pueda hacer uso de este formato.

1. Formatos de datos científicos

A continuación se describen los principales formatos de datos científicos, los cuales cuentan con diferentes niveles de complejidad. Se describen las características de estos tipos de formatos de datos científicos. Se muestran los que se consideran los más populares como son el Hierarchical Data Format, más conocido por HDF, el *Common Data Format* (CDF) y NetCDF, estos serán descritos con más detalles en los epígrafes siguientes (Cohen et al., 2006).

1.1 Hierarchical data format (hdf)

Hierarchical Data Format, más conocido por HDF, es desarrollado por el Centro Nacional de Aplicaciones de Supercómputo (National Center for Supercomputing Applications, NCSA) en la Universidad de Illinois. Es una biblioteca y un formato de fichero multi-objeto para la transferencia de datos gráficos y numéricos entre máquinas (Fortner, 1998).

HDF es un proyecto de datos científicos de gestión que proporciona herramientas básicas para la gestión de datos científicos, y apoyo a los científicos que usan software HDF. Los productos incluyen dos formatos de datos distintos, HDF4 y HDF5, bibliotecas de soporte para leer y escribir datos HDF y utilidades para trabajar con HDF. Disponible a partir de 1988, HDF se ha convertido en un estándar para los datos científicos archivar y compartir datos. La gama de usuarios de HDF se ilustra en forma tabular. HDF es el formato de base de la norma HDF-EOS y por consiguiente es el formato de destino para una gran cantidad de datos de teledetección. HDF5 es una versión completamente nueva de HDF que ofrece un formato mejorado y admite el almacenamiento eficiente de los conjuntos de datos muy grandes, y cuenta con el apoyo de una biblioteca de entrada salida que se ejecuta tanto en serie como en paralelo (Folk et al., 1999).

HDF se encuentra disponible de forma gratis. La distribución consiste en la biblioteca HDF, las utilidades de línea de comando HDF y una *suite* de prueba (código fuente solamente). HDF presenta 4 niveles de interacción. En su nivel más bajo es un formato de fichero físico para el almacenamiento de datos científicos. En su nivel más alto es una colección de utilidades y aplicaciones para manipular, ver y analizar datos en los ficheros HDF, y entre esos niveles HDF, una biblioteca de programas que provee APIs de alto nivel y una interfaz



de datos de bajo nivel como se puede ver en la 1 (Group, 2000) .



Figura 1: Niveles de interacción de HDF

Aplicaciones generales

En el nivel más alto hay utilidades de línea de comandos de HDF, aplicaciones de NCSA que soportan visualización de datos y análisis, y variedad de aplicaciones de terceros desarrolladores.

Existen utilidades de línea de comandos de HDF para:

- Convertir de un formato a otro (por ejemplo, desde y hacia JPEG/HDF)
- Analizar y ver ficheros HDF (siendo *hdp* una de las herramientas más útiles)
- Manipular los ficheros HDF

De las utilidades HDF, la "*hdp*" es una de las más importantes. Provee información rápida sobre los contenidos y objetos de datos en un fichero HDF, puede listar los contenidos de los ficheros HDF en varios niveles con diferentes detalles, además puede también vaciar los datos de uno o más ficheros en el fichero, en formato binario o ASCII.

Compresión de datos

HDF 4.0 (y versiones posteriores) soportan una interfaz de compresión de bajo nivel, la cual permite que cualquier objeto de dato sea comprimido utilizando una variedad de algoritmos. Actualmente solo tres algoritmos de compresión están soportados: Run-length Encoding (RLE), Huffman adaptativo, y un codificador de diccionario LZ-77 (el algoritmo de decodificación de gzip). Planes para algoritmos futuros incluyen un codificador de diccionario Lempel/Ziv-78, un codificador aritmético y un algoritmo rápido de Huffman. HDF 4.0 (y entregas posteriores) soportan compresión de n-bit para SDS, RLE (Run- Length Encoding), IMCOMP, y compresión JPEG para imágenes de puntos.

El fichero hdf

Un fichero HDF contiene un encabezamiento de fichero, al menos un bloque descriptor de datos, y cero o más elementos de datos. El encabezamiento de fichero identifica el fichero como un fichero HDF. Un bloque descriptor de datos contiene un número de descriptores de



datos. Un descriptor de datos y un elemento de dato juntos, conforman un objeto de dato, el cual es la estructura básica de agrupamiento para el encapsulamiento de datos en un fichero HDF.

El fichero HDF posee una serie de rasgos que se mencionan a continuación:

Es versátil. HDF soporta muchos tipos diferentes de modelos de datos. Cada modelo de dato define un conjunto específico de tipo de dato y provee una API para lectura, escritura y organización de datos y metadatos del tipo correspondiente. Los modelos de datos soportados incluyen arreglos multidimensionales, imágenes de puntos, y tablas.

Es auto-descriptivo, permitiendo una aplicación para interpretar la estructura y contenidos de un fichero sin ninguna información proveniente del exterior.

Es flexible. Con HDF se pueden mezclar y asociar objetos relacionados, agrupados en un fichero y acceder a ellos como un grupo o como objetos individuales. Los usuarios pueden además crear sus propias estructuras de agrupamiento utilizando un rasgo de HDF llamado *vgroups*.

Es extensible. Puede acomodar fácilmente nuevos modelos de datos, sin tener en cuenta si fueron adicionados por el equipo de desarrolladores de HDF o por los usuarios de HDF.

Es portable. Los ficheros HDF pueden ser compartidos a través de la mayoría de las plataformas comunes, incluyendo muchas estaciones de trabajo y computadoras de alto desempeño. Un fichero HDF creado en una computadora puede ser leído en un sistema diferente sin modificación alguna.

1.2 Common data format

El *Common Data Format* (CDF) es un formato de datos auto-descriptivo para el almacenamiento y manipulación de datos escalares y multidimensionales en una forma independiente de la plataforma y el campo de especialidad. A pesar de que CDF tiene su propio formato interno auto-descriptivo, éste consiste en algo más que sólo un formato de datos. CDF es un paquete de administración de datos científicos (conocido como la "Biblioteca CDF") la cual permite a los programadores y desarrolladores de aplicaciones administrar y manipular arreglos de datos escalares, vectoriales y multidimensionales (Atkinson and Hankin, 1995).

El fichero dotCDF (Fichero de datos CDF) contiene un número mágico y dos o más registros internos (RIs) que son utilizados para organizar los contenidos de un CDF. Diferentes tipos de registros internos son utilizados para almacenar información sobre varios aspectos y/u objetos en el CDF.

Todos los ficheros dotCDF contienen un *CDF Descriptor Record* (CDR) y un *Global Descriptor Record* (GDR). Otros registros internos estarán presentes dependiendo de los contenidos del CDF. El único registro interno en una localización fija en el fichero dotCDF es el CDR. Todos los otros registros internos (incluyendo el GDR) pueden estar presentes en cualquier orden (el cual depende, generalmente, del orden en que fueron creados por la aplicación los contenidos del CDF).

Los desplazamientos dentro del fichero son utilizados para "apuntar" a otros registros



internos. Las listas enlazadas de registros internos son implementadas mediante el almacenamiento del desplazamiento dentro del fichero del primer registro interno en la lista enlazada, almacenando ese registro interno, el desplazamiento dentro del fichero del próximo registro interno y así sucesivamente.

1.3 The network common data form

NetCDF es una abstracción de datos para almacenar y recuperar datos multidimensionales. NetCDF se distribuye como una biblioteca de software que proporciona una implementación concreta de esa abstracción. La aplicación proporciona un formato independiente de la máquina para la representación de datos científicos. En conjunto, la abstracción, la biblioteca y el formato de datos es compatible con la creación, el acceso y el intercambio de información científica. NetCDF es útil para soportar objetos que contienen tipos diferentes de datos en un entorno de red heterogénea y para la escritura de software de aplicación que no depende de los formatos específicos de la aplicación. La independencia de las representaciones particulares se logra mediante el uso de un estándar común para la representación de datos externos. (Rew and Davis, 1990)

Una de las metas de netCDF es soportar acceso eficiente a pequeños subconjuntos de grandes conjuntos de datos. Para lograr esto, netCDF utiliza accesos directos en lugar de accesos secuenciales. Esto puede ser mucho más eficiente cuando el orden en que los datos son leídos es diferente del orden en que fueron escritos, o cuando debe ser leído en diferentes órdenes para distintas aplicaciones.

El costo para la representación externa portable depende de muchos factores, incluyendo el tipo de dato, el tipo de computadora, la granularidad del acceso a datos, y de cuán bien se ha ajustado la implementación a la computadora en la que está corriendo. Este costo es típicamente pequeño en comparación con todo el conjunto de recursos utilizados por una aplicación. De cualquier forma, el costo de la capa de representación externa es, usualmente, un precio razonable a pagar por acceso de datos portables.

A pesar de que la eficiencia en el acceso a datos ha sido un asunto importante en el diseño e implementación de netCDF, todavía es posible utilizar la interfaz netCDF para acceder a datos de formas ineficientes: por ejemplo, mediante la petición de una porción de datos que requiere un sólo valor de cada registro.

PARTES DE UN FICHERO netcdf

Un conjunto de datos de netCDF clásico o de desplazamiento de 64 bit es almacenado como un sólo fichero que consiste en dos partes:

Un encabezado (*header*), conteniendo toda la información sobre las dimensiones, atributos y variables, excepto para los datos de variable.

Una parte de datos, que comprende los datos de tamaño fijo (datos para variables que no poseen una dimensión ilimitada) y los datos de tamaño variable (datos para variables que presentan una dimensión ilimitada).

El encabezado, al principio del fichero, contiene información sobre las dimensiones, variables, y atributos en el fichero, incluyendo sus nombres, tipos, y otras características. La información



sobre cada variable incluye el desplazamiento al inicio de los datos de la variable para variables de tamaño fijo, o el desplazamiento relativo de otras variables dentro de un registro. El encabezado también contiene longitudes de dimensión e información necesaria para trasladar (*map*) índices multidimensionales de cada variable hacia los desplazamientos apropiados dentro del fichero.

Las partes de datos con tamaño fijo que siguen al encabezamiento contienen todos los datos de variables para las variables que no presentan una dimensión ilimitada. Los datos para cada variable son almacenados contiguamente en esta parte del fichero. Si no hay variables con dimensión ilimitada, esta es la última parte del fichero netCDF.

La parte de datos de registro que le sigue a los datos de tamaño fijo consiste en un número variable de registros de tamaño fijo, que cada uno contiene datos para todas las variables de tipo registro. Los datos registro para cada variable son almacenados contiguamente en cada registro.

El orden en que aparecen los datos de la variable en cada sección de datos, es el mismo en que fueron definidas: un orden numérico incremental por el ID de variable de netCDF. Este conocimiento puede ser utilizado a veces para mejorar el desempeño de acceso a datos, puesto que el mejor acceso a datos es actualmente logrado mediante la lectura y escritura de datos en orden secuencial.

2 Integración de los formatos de datos científicos en los sig

En investigaciones realizadas sobre la introducción de estos formatos de datos científicos en los Sistemas de Información Geográfica, se pudo constatar que hay ejemplos de algunos SIG que están logrando la introducción de dichos formatos pero aun así es una cantidad reducida para la gran cantidad de SIG que existen. También se comprobó que existe un gran interés en la introducción de estos formatos en los SIG dado la ventaja que esto representa para el almacenamiento de las variables espacio-temporales.

Las nuevas tecnologías para la computación de objetos distribuidos (DOC) desafían nuevos enfoques para integrar sistemas de información geográfica (SIG) y las herramientas de simulación para el modelado espacio-temporal. Varias iniciativas de estandarización trabajan sobre la especificación de las interfaces y la definición de arquitecturas de servicio para los componentes apropiados. En la actualidad, los intentos más prometedores están representados por el OpenGIS y la arquitectura de alto nivel. Define un marco posible para el SIG integrado de simulación de procesos espacio-temporales. Dos capas de abstracción proporcionan interfaces para diferentes kernels SIG y permitir la integración de los modelos de simulación diferentes, la implementación de sistema de archivos es basado en netCDF (Bernard and Krüger, 2000).

Otro de los proyecto de inclusión de formatos de datos científicos es el del gigante Google con su proyecto de Google Earth como plataforma SIG para aplicaciones en climas. Cuenta con el sistema WDSS-II que incluye herramientas para convertir datos de netCDF genérico con otros datos, un ejemplo es la "rotación de pistas" producto generado por el sistema WDSS-II mediante la combinación de este producto con información de SIG de alta resolución, los usuarios pueden determinar cuándo ocurre una tormenta (Smith and Lakshmanan, 2006).

La visualización y el trabajo con estos formatos de datos en ArcGIS ya es posible para los



formatos netCDF y HDF y a continuación se muestra un ejemplo de integración además con SIG basados en Web para el trabajo con datos meteorológicos.

Con el fin de obtener información precisa, específica y oportuna previsión meteorológica, el Departamento Meteorológico desea emitir el pronóstico del tiempo utilizando WebGIS y ArcGIS, el papel acabado de visualización de imágenes de los NetCDF (Formato de datos de red común) datos que WRF (Investigación y Pronóstico del tiempo) de salida, combinada con ArcGIS NetCDF mapa electrónico, lograron una fusión de la información meteorológica y predicción de la información geográfica. Por último, se ha diseñado el sistema de predicción del tiempo con información de SIG basados en Web para mejorar la distribución, la visualización y el enriquecimiento de la información del tiempo previsto. Es beneficioso para predecir los efectos sobre los factores geográficos por factores meteorológicos (Cong-cong and Li-ying, 2010). La incorporación de NetCDF en ArcGIS fue promovido por el Centro Nacional de Investigación Atmosférica (NCAR, véase: <http://www.ncar.ucar.edu>) y del Grupo de Interés Especial de la Atmósfera (ver: <http://www.gis.ucar.edu/sig>) (Groot et al., 2008).

Al SIG gvSIG se le integró el formato de dato científico HDF, con el objetivo de poder almacenar variables espacio-temporales para el módulo de Visualización Científica pueda hacer uso de grandes volúmenes de datos de estas variables con una gran calidad y mejorando el aprovechamiento del espacio dado que el HDF cuenta con mecanismos de compresión de datos.

3. Módulo de manipulación de archivos hdf en gvSIG

A continuación se muestran los principales aspectos de la extensión implementada para gvSIG, así como las funcionalidades del trabajo con datos de formatos HDF que fueron implementadas para lograr la manipulación de este tipo de formatos dentro del SIG.

El objetivo del paquete de visualización de datos de formatos científicos es permitir la manipulación y visualización de conjuntos de datos de tipo HDF. Este paquete está formado a la vez por varios paquetes, estos se interrelacionan entre sí como un todo a la hora de lograr la visualización.

EL paquete *ObjetchDF* de Java que implementa objetos de datos HDF4 y HDF5. Este paquete provee APIs comunes de Java para la manipulación de archivos de ambos tipos de formatos. Este paquete no es una envoltura (*wrapper*) natural para las bibliotecas HDF. Las envolturas (*wrappers*) para HDF4 y HDF5 están implementadas por separado.

El *ObjectHDF* implementa APIs de alto nivel y encapsula las llamadas a la biblioteca HDF en forma orientada a objetos, para facilitar el acceso a los archivos. Este paquete *ncsa.hdf.object*, provee clases que reflejan conceptos fundamentales para el diseño de objetos HDF. Los objetos de HDF5 (grupos y conjuntos de datos) y HDF4 (grupo, arreglos multidimensionales, imágenes raster, vdata y anotaciones) son presentados como clases Java. El *ObjectHDF* tiene dos cometidos principales: primero, que simplifica el proceso de información de lectura o escritura de datos para un archivo porque los detalles de ganar acceso a la biblioteca HDF son encapsulados en clases respectivas; en segundo lugar, los objetos HDF4 y HDF5 heredan del mismo objeto y de la misma interfaz. Las aplicaciones pueden acceder a los objetos a través de esta clase sin tener que ir a las bibliotecas.

En los paquetes *HDFLib* y *HDF5Lib* se implementan las bibliotecas para el trabajo específico



en cada uno de los tipos de objetos. En el caso del paquete HDFLib, se define una serie de clases Java que posibilita el trabajo con archivos de datos tipo HDF en particular. En el caso de HDF5lib se definen funciones específicas para el trabajo con archivos de datos HDF5 que, por ser una versión más reciente, incorpora funcionalidades nuevas que se irán describiendo en el transcurso de este trabajo.

A continuación se exponen los temas referentes al uso de la extensión, teniendo en cuenta las tecnologías utilizadas y algunas de las funcionalidades del trabajo con archivos de tipos de datos científicos importantes a la hora de realizar este proyecto.

La biblioteca HDF-JAVA nos brinda la posibilidad de manipular tipos de datos científicos con la cual se creó un archivo con datos climáticos a escala global para su posterior uso con la extensión de gvSIG de Visualización Científica, ya que estas variables tienen asociadas a ellas grandes volúmenes de datos almacenados por lo que hace el trabajo muy engorroso (McGrath et al., 1997).

Con la creación de programas sencillos y usando las funciones de la biblioteca de HDF-JAVA se construyó un archivo en el cual se almacenaron los datos con variables climáticas a escala global, esto redujo significativamente el espacio que ocupaban dichos datos y con una mejor organización.

HDFView, que es el software que con más frecuencia es utilizado para ver HDF5, se procesó el archivo HDF creado con las variables climáticas para agregarles otros conjuntos de datos. HDFView es basado en Java, con el cual es posible navegar y editar archivos HDF4 y HDF5. HDFView permite a un usuario descender a través de la jerarquía de archivos y navegar entre los objetos de datos del fichero y con el mismo se muestra cómo quedó la estructura del archivo creado (Yang et al.).

Los datos meteorológicos tomados como caso de estudio pertenecen a una recopilación de los cinco continentes durante un siglo. El archivo origen de los datos consta de diecisiete registros. Los primeros ocho registros están destinados a los datos. Estos están separados en matrices de acuerdo al momento al que pertenecen y la localización en que fueron tomados. Los ocho siguientes campos son dependientes de los anteriores ya que están destinados a llevar el registro de los valores mínimo y máximo globales para cada variable. Por ejemplo DS0 tiene a DS0E que tiene dos campos donde el superior es el valor mínimo y el inferior el máximo que pueden encontrarse en DS0.

Cada DS# está referenciado a una variable con un valor a una matriz, en este caso de 720X360 que es la amplitud del mapa de fondo. Los nombres de las variables son: DS0, temperatura; DS1, temperatura mínima; DS2, temperatura máxima; DS3, nubosidad; DS4, presión; DS5, humedad; DS6, vapores; DS7, precipitaciones y DS8, escarcha.

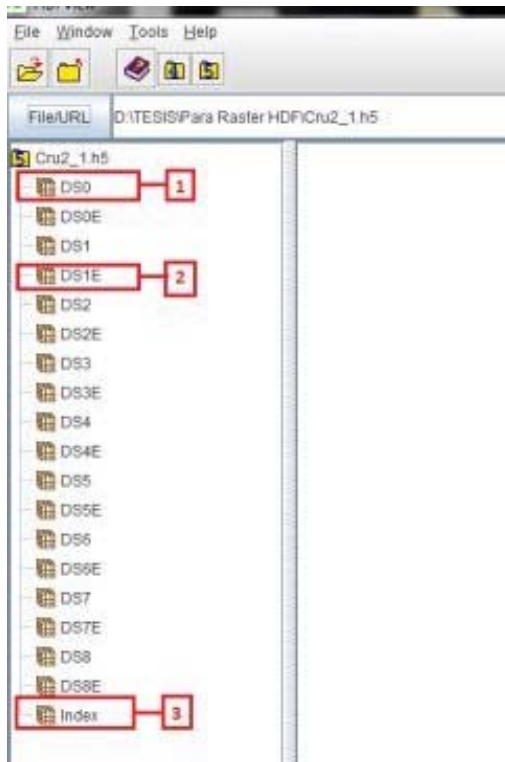


Figura 2 Lista de los Datos en HDFview

En la Figura 2 se puede apreciar la distribución de la información en el fichero.

EL último *dataset* es el llamado Índex, señalado por el número tres. Este es el que nos permite dibujar una imagen para referenciar cada variable. Este registro es una matriz donde el tamaño de sus filas y columnas es consecuente con las dimensiones del mapa al cual hacen referencia los datos

En el caso del HDFview, permite generar una imagen del mapa tomando como referencia este registro. El Índex además de generar la imagen base de los datos permite dado un punto acceder a la información referente a la guardada en cualquiera de los dieciséis registros restantes.



	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
0	270	274	269	265	259	253	246	242	248	252	261	268	270	274	269	265	259	253
1	-187	-212	-195	-137	-30	47	89	81	38	-56	-132	-189	-197	-213	-195	-137	-30	47
2	-195	-221	-203	-145	-35	43	85	75	28	-61	-138	-193	-199	-221	-203	-145	-35	43
3	-208	-235	-222	-165	-53	27	68	58	13	-78	-154	-207	-208	-235	-222	-165	-53	27
4	-215	-245	-230	-174	-59	22	61	52	7	-82	-160	-214	-216	-245	-230	-174	-59	22
5	-228	-257	-245	-180	-73	9	47	38	4	-83	-172	-225	-228	-257	-245	-180	-73	9
6	-233	-261	-247	-190	-72	11	48	37	5	-84	-172	-225	-233	-261	-247	-190	-72	11
7	-244	-273	-262	-206	-85	-1	34	23	-17	-105	-183	-239	-244	-273	-262	-206	-85	-1
8	-244	-270	-263	-194	-72	14	45	34	-6	-86	-173	-236	-244	-270	-263	-194	-72	14
9	-248	-275	-258	-202	-81	1	27	19	-20	-103	-174	-237	-248	-275	-258	-202	-81	1
10	-252	-280	-267	-209	-86	-1	25	16	-23	-108	-183	-242	-252	-280	-267	-209	-86	-1
11	-253	-229	-213	-157	-47	21	71	62	18	-59	-145	-205	-203	-229	-213	-157	-47	21
12	-245	-244	-232	-177	-65	15	54	45	2	-84	-160	-213	-215	-244	-232	-177	-65	15
13	-202	-240	-238	-180	-68	15	43	43	1	-88	-182	-218	-222	-240	-238	-180	-68	15
14	-230	-258	-244	-188	-71	10	46	36	5	-91	-187	-225	-230	-258	-244	-188	-71	10
15	-236	-264	-249	-193	-74	9	42	32	-8	-84	-170	-230	-236	-264	-249	-193	-74	9
16	-238	-263	-244	-195	-64	19	50	38	0	-89	-163	-229	-238	-263	-244	-195	-64	19
17	-251	-278	-262	-205	-84	-1	25	15	-22	-105	-176	-240	-251	-278	-262	-205	-84	-1
18	-249	-279	-269	-214	-93	-11	14	7	-30	-111	-184	-240	-249	-279	-269	-214	-93	-11
19	-253	-280	-264	-205	-81	5	30	21	-18	-103	-176	-241	-253	-280	-264	-205	-81	5
20	-255	-275	-265	-203	-78	23	28	22	-23	-108	-183	-242	-255	-275	-265	-203	-78	23
21	-272	-317	-272	-265	-250	-242	-238	-244	-249	-260	-269	-272	-277	-272	-265	-250	-242	-238
22	-173	-185	-178	-131	-34	32	61	57	22	-68	-108	-164	-173	-185	-178	-131	-34	32
23	-188	-224	-207	-152	-43	33	72	63	20	-65	-138	-194	-198	-224	-207	-152	-43	33
24	-211	-239	-227	-173	-62	15	54	46	4	-80	-155	-208	-211	-239	-227	-173	-62	15
25	-215	-242	-227	-171	-58	20	56	48	6	-79	-153	-210	-215	-242	-227	-171	-58	20
26	-223	-250	-236	-180	-65	14	49	40	0	-85	-159	-217	-223	-250	-236	-180	-65	14
27	-207	-253	-236	-180	-63	17	49	40	1	-84	-158	-219	-207	-253	-236	-180	-63	17
28	-217	-241	-224	-171	-59	17	45	38	1	-77	-145	-208	-217	-241	-224	-171	-59	17
29	-252	-279	-263	-205	-84	-1	25	16	-22	-105	-177	-241	-252	-279	-263	-205	-84	-1
30	-241	-268	-259	-204	-87	-7	17	11	-25	-103	-174	-230	-241	-268	-259	-204	-87	-7
31	-272	-318	-273	-271	-268	-264	-268	-265	-269	-263	-269	-273	-273	-272	-268	-264	-268	-265
32	-170	-192	-175	-129	-35	29	59	56	21	-48	-108	-161	-170	-192	-175	-129	-35	29
33	-187	-213	-195	-143	-38	36	75	66	25	-57	-129	-183	-187	-213	-195	-143	-38	36
34	-194	-221	-204	-150	-43	32	69	62	20	-63	-135	-190	-194	-221	-204	-150	-43	32
35	-207	-235	-222	-170	-60	15	63	46	5	-77	-150	-203	-207	-235	-222	-170	-60	15
36	-212	-239	-225	-172	-61	16	62	44	4	-78	-151	-207	-212	-239	-225	-172	-61	16
37	-218	-245	-230	-176	-63	15	49	41	1	-81	-154	-211	-218	-245	-230	-176	-63	15
38	-221	-247	-230	-174	-60	18	50	42	3	-80	-152	-213	-221	-247	-230	-174	-60	18
39	-253	-280	-264	-209	-84	-1	25	16	-22	-105	-177	-241	-253	-280	-264	-209	-84	-1
40	-254	-281	-269	-210	-85	0	23	15	-19	-106	-178	-242	-254	-281	-269	-210	-85	0

Figura 3 Vista de la información de DS0

En la Figura 3 se muestra una tabla que contiene los datos de una de las variables, DS0. Como se puede apreciar, en las columnas resaltadas se encuentran los valores referentes a la posición y tiempo.

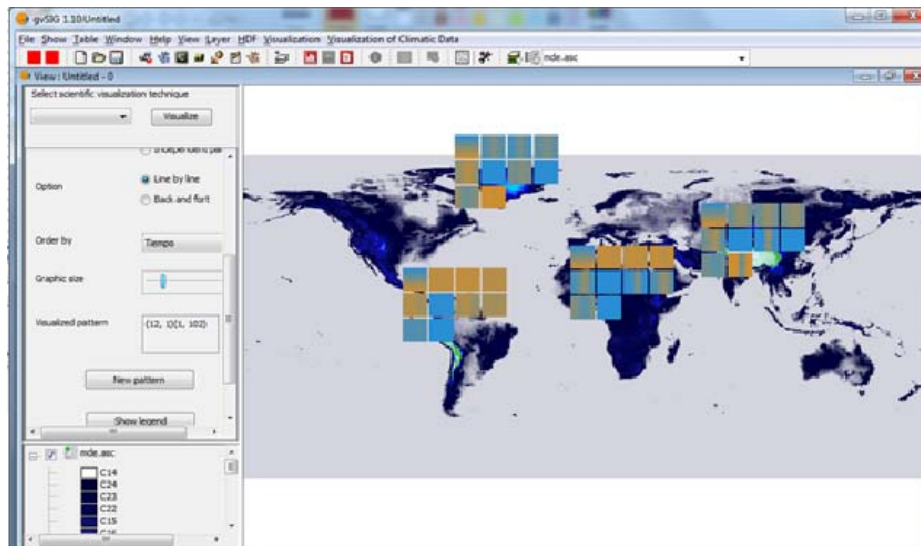


Figura 4. Integración de HDF con gvSIG.

En la Figura 4 se muestra cómo hace uso de las técnicas de visualización científica, las cuales están actuando sobre los datos espacio-temporales de variables como temperatura, nubosidad, presión, humedad, vapores, precipitación y escarcha a escala mundial. Todos los



datos están almacenados en el archivo que se expuso anteriormente y gvSIG al tener el módulo de HDF se hace posible la lectura del archivo HDF. También es posible hacer la visualización de varias variables a la vez y sin la complicación de que se tenga que hacer acceso a otros archivos ya que todos los datos están almacenados en *dataset* diferentes, sin la utilización de este formato se hubiese tenido que abrir un archivo para cada variable con las complejidades a que esto conlleva.

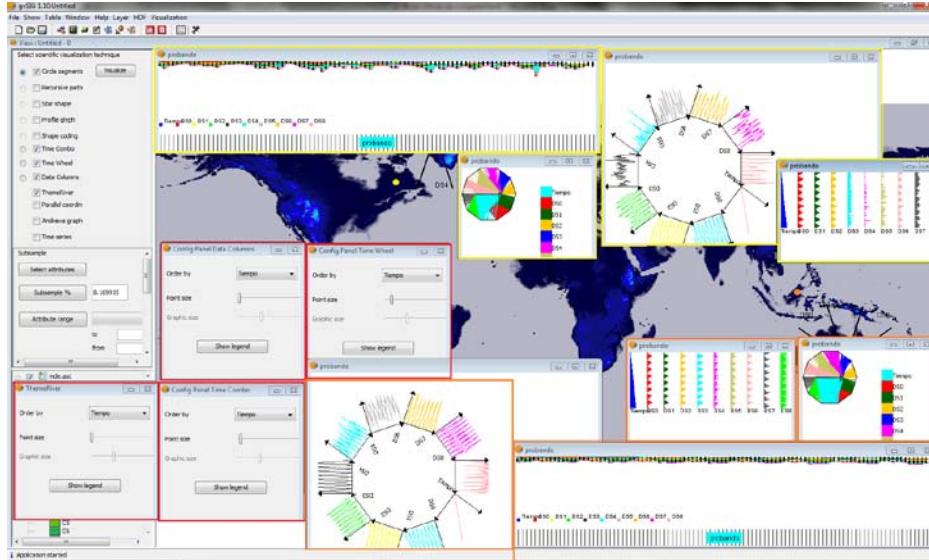


Figura 5 Visualización de múltiples técnicas

CONCLUSIONES

En este trabajo se presentaron los principales formatos de datos científicos más utilizados en la visualización científica. En particular se presentó la integración del formato HDF en el SIG gvSIG mediante la incorporación de un módulo basado en java para permitir el soporte de estos formatos de datos en el SIG.

BIBLIOGRAFÍA

- ATKINSON, L. & HANKIN, S. 1995. Common Data Formats. *OCEANOGRAPHY-WASHINGTON DC-OCEANOGRAPHY SOCIETY*-, 8, 100-101.
- BERNARD, L. & KRÜGER, T. 2000. Integration of GIS and Spatio-temporal Simulation Models: Interoperable Components for Different Simulation Strategies. *Transactions in GIS*, 4, 197-215.
- COHEN, S., HURLEY, P., SCHULZ, K. W., BARTH, W. L. & BENTON, B. 2006. Scientific formats for object-relational database systems: a study of suitability and performance. *ACM SIGMOD Record*, 35, 10-15.
- CONG-CONG, X. & LI-YING, W. Year. Study of Image Display with NetCDF Data on WebGIS. *In: Information Technology and Computer Science (ITCS), 2010 Second International Conference on*, 2010. IEEE, 368-371.
- FOLK, M., MCGRATH, R. E. & YEAGER, N. Year. HDF: an update and future directions. *In: Geoscience and Remote Sensing Symposium, 1999. IGARSS'99 Proceedings. IEEE 1999 International*, 1999. IEEE, 273-275.
- FORTNER, B. 1998. HDF: The hierarchical data format. *DR DOBB'S J SOFTWARE TOOLS PROF*



- PROGRAM, 23, 42.
- GROOT, N., VEGTE, J., SOM DE CERFF, W., OORD, G., SLUITER, R., NEUT, I., PLIEGER, M., HEES, R., JEU, R. & SCHAEPMAN, M. 2008. Atmospheric data access for the geospatial user community (ADAGUC).
- GROUP, H. 2000. Hierarchical Data Format, Version 5. The National Center for Supercomputing Applications, <http://hdf.ncsa.uiuc.edu/HDF5>.
- MCGRATH, R. E., LU, X. & FOLK, M. 1997. Java (TM) applications using NCSA HDF files. *Concurrency Practice and Experience*, 9, 1113-1125.
- REW, R. & DAVIS, G. 1990. NetCDF: an interface for scientific data access. *Computer Graphics and Applications, IEEE*, 10, 76-82.
- SMITH, T. M. & LAKSHMANAN, V. Year. Utilizing Google Earth as a GIS platform for weather applications. *In: 22nd International Conference on Interactive Information Processing Systems for Meteorology, Oceanography, and Hydrology*, 2006.
- YANG, M. Q., MCGRATH, R. E. & FOLK, M. P2. 42 HDF5—A HIGH PERFORMANCE DATA FORMAT FOR EARTH SCIENCE.