

Módulo de Análisis basados en métodos de interpolación para GeneSIG.

Ing. Arletty Silvera Boffill

GEySED, Universidad de las Ciencias Informáticas, Carretera a San Antonio de los Baños, km 2 ½, Boyeros, La Habana, Cuba.

asilvera@uci.cu

Ing. Celia Torres Reyes

GEySED, Universidad de las Ciencias Informáticas, Carretera a San Antonio de los Baños, km 2 ½, Boyeros, La Habana, Cuba.

ctreyes@uci.cu

RESUMEN

La plataforma GeneSIG es diseñada para capturar, almacenar, manipular, analizar y desplegar en todas sus formas la información geográficamente referenciada en Sistemas de Información Geográfica (SIG). Cuenta actualmente con tres formas de realizar tematizaciones para el análisis de la información que se representa en el mapa: coropletas, símbolos proporcionales y gráficas dinámicas. Aun cuando estas son ampliamente utilizadas, algunas presentan problemas que condicionan los resultados en dependencia del criterio de análisis sobre el que se esté trabajando.

Mediante la presente investigación se implementó un módulo de análisis basado en métodos de interpolación el cual garantiza mejorar la precisión del visualizado de los datos espaciales en las tematizaciones sobre la plataforma, generando resultados precisos y confiables; logrando un aporte significativo en el proceso de toma de decisiones a través de la utilización de personalizaciones basadas en GeneSIG. La implementación del sistema se llevó a cabo en un ambiente web, utilizando tecnologías libres.

Palabras claves: Interpolación, Plataforma GeneSIG, Sistemas de Información Geográfica y Tematizaciones.

INTRODUCCIÓN

La necesidad de representar y gestionar el gran volumen de datos que se genera diariamente en las empresas, ha posibilitado que se desarrollen los Sistemas de Información (SI) y las tecnologías computacionales, cambiando progresivamente la forma de procesar los resultados obtenidos. Una de las áreas de mayor aplicación de estos sistemas son los Sistemas de Información Geográfica (SIG), los cuales constituyen una herramienta indispensable para el estudio, gestión y análisis territorial.

El desarrollo de los Sistemas de Información Geográfica ha tenido una importante representación en el área del análisis espacial, nutriéndose de investigaciones del campo de las matemáticas y la estadística. Algunos de sus aportes en dicha área son utilizados para realizar exploraciones de los datos geográficos posibilitando extraer información adicional que podría no ser evidente observando el mapa simplemente.

La evolución de las Tecnologías de la Información y las Comunicaciones (TIC) ha propiciado el perfeccionamiento del análisis espacial de los datos; conllevando al hombre a realizar disímiles estudios encaminados a mejorar temas relativos a la representación de la información geográfica. La necesidad de potencializar el uso de las TIC en la sociedad cubana, posibilitó que fueran informatizadas varias de sus esferas, como es el caso de las ciencias geográficas. Un papel protagónico en esta área lo desempeña la Universidad de las Ciencias Informáticas (UCI), la que cuenta con el centro de desarrollo Geoinformática y Señales Digitales (GEySED). Dentro de la amplia gama de productos que se llevan a cabo en el centro, muchos son desarrollados sobre la plataforma GeneSIG, realizada íntegramente con tecnologías libres, en cuyo desarrollo participan actualmente tres entidades: grupo empresarial GEOCUBA (encargándose del suministro cartográfico), las Fuerzas Armadas Revolucionarias (FAR) y la UCI. Esta plataforma fue creada para el desarrollo de aplicaciones SIG en entornos web y cuenta con las funcionalidades comunes de estos sistemas, pudiendo ser personalizada como aplicación a la medida, incluyéndosele nuevas funcionalidades en caso de ser necesario.

GeneSIG cuenta actualmente con tres formas de realizar tematizaciones para el análisis de la información que se representa en el mapa: coropletas, símbolos proporcionales y gráficas dinámicas. Aun cuando estas son ampliamente utilizadas, algunas presentan

problemas que condicionan los resultados en dependencia del criterio de análisis sobre el que se esté trabajando.

Por ejemplo, las tematizaciones por coropletas posibilitan a los especialistas obtener una representación detallada mediante el uso de colores, permitiendo de este modo definir las distintas áreas visualizadas en el mapa, a pesar de esto presenta las siguientes inconvenientes:

- Se limita a colorear polígonos (ver Fig.1), esto significa que siempre la tematización se realizará coincidiendo con los límites definidos en el mapa según las características del mismo. Por lo que no se podría dibujar una de estas áreas formando diferentes zonas de colores en función de los datos socioeconómicos que se tengan, que facilitarían el análisis de indicadores tales como la distribución poblacional.

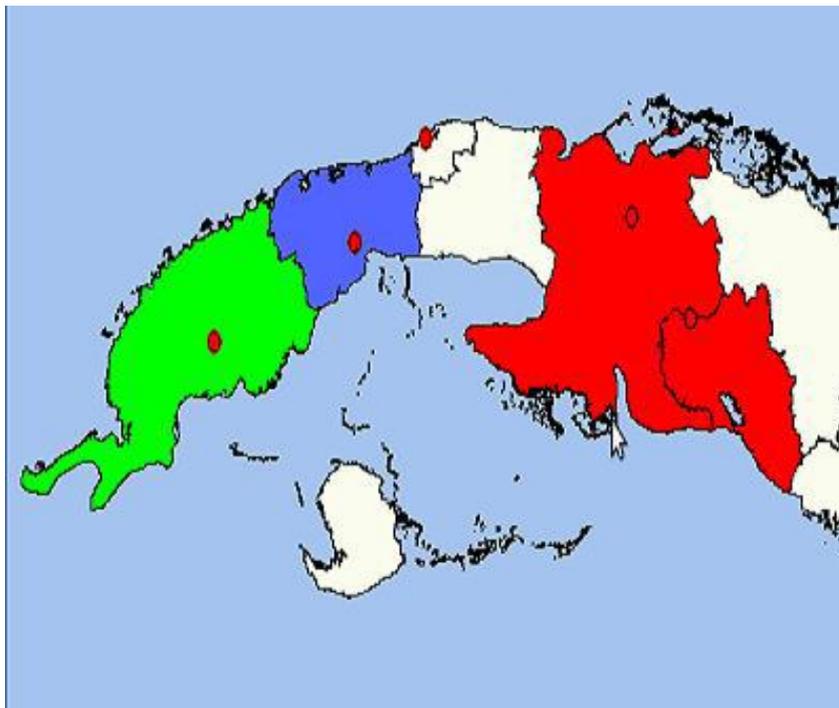


Fig. 1 Resultados sobre la disponibilidad de los almacenes de LABIOFAM arrojados mediante la tematización por coropletas en la plataforma GeneSIG.

- No presenta el mismo nivel de detalle de la información en áreas grandes que en áreas pequeñas, lo que significa que cuando se realiza la representación de un elemento en el mapa, por ejemplo un municipio con gran cantidad de población, se muestra abundante información sobre los barrios y consejos populares. Sin embargo en los municipios poco poblados, habitualmente es atípico encontrar mapas con la representación de la información sobre sus barrios y consejos populares.

Los símbolos proporcionales a diferencias de las demás tematizaciones realizan su representación mediante símbolos significativos (círculos, cuadrado, triángulos o la combinación entre ellos), permitiendo así una vista más detallada del mapa. Su uso se fundamenta en el tamaño de la variable que se va a tematizar, la que resulta ser más representativa para variables cuantitativas y para la representación de valores absolutos. En variables cuya diferencia entre el valor mínimo y máximo sea muy grande, puede provocar que algunos símbolos resulten prácticamente imperceptibles (ver Fig.2), lo que determina que las entidades con muy poca información a representar en el mapa sean casi inapreciables frente a las más populosas.

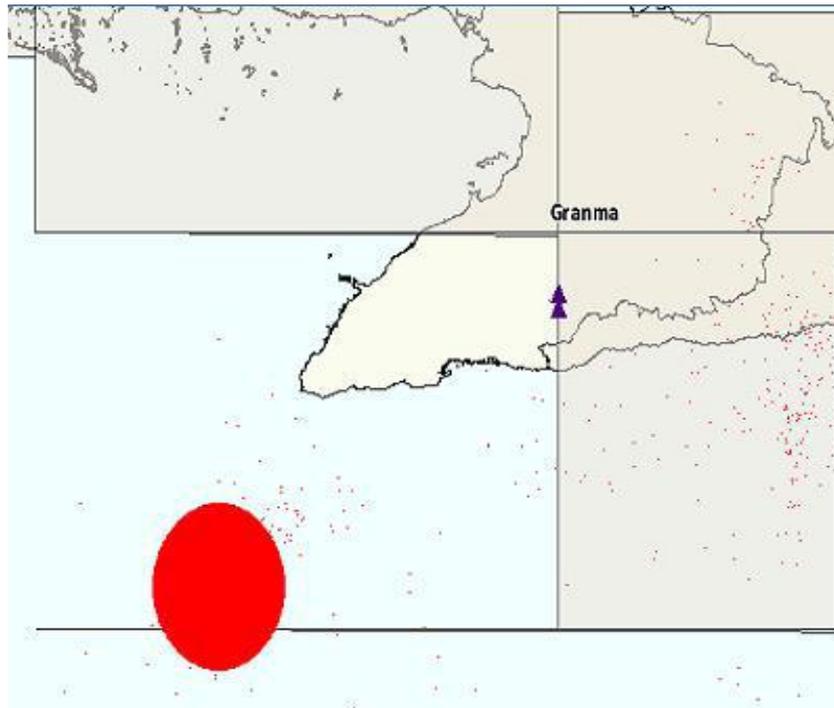


Fig. 2 Resultados de los epicentros arrojados mediante la tematización por símbolos proporcionales en la plataforma GeneSIG.

Los resultados arrojados por mapas que utilizan coropletas y símbolos proporcionales como parte de su proceso de visualización, pueden influir gravemente en la toma de decisiones y el análisis erróneo en la solución de determinada situación tendría secuelas negativas.

Las gráficas dinámicas como su nombre lo indica están destinadas a mostrar los datos obtenidos mediante gráficas de barras o pastel, logrando así que los especialistas puedan entender el resultado visualizado mediante una de las formas más utilizadas actualmente como método representativo. Estas no evidencian ningún inconveniente debido a que no se representan sobre el mapa y no afectan la correcta visualización de los datos en él.

MATERIALES Y MÉTODOS

Análisis espacial basado en métodos de interpolación

Uno de los objetivos que persigue el análisis espacial de los datos es obtener información temática cartográfica, representada a través de mapas que reflejen de la forma más precisa el área de estudio.

En esencia el análisis espacial de los datos posibilita identificar los componentes del espacio analizado (datos); luego se realiza un procedimiento o conjunto de estos (métodos de análisis numérico) que permitan comprender la funcionalidad de los componentes obtenidos. Es por ello que el resultado alcanzado depende de la selección del método utilizado.

El empleo de métodos de análisis numérico provocado por el creciente avance de las TIC ha posibilitado que las ciencias sean usadas en el desarrollo de la computación. La interpolación espacial es uno de los procesos que se desarrolla sobre bases matemáticas, lo que posibilita exactitud en los resultados.

Cuando se refiere a interpolación espacial se puede definir como la estimación de los valores que alcanza una variable Z en un conjunto de puntos definidos por un par de coordenadas (X, Y). Los SIG emplean la interpolación espacial para obtener capas que constituyen la variable a interpolar. Por lo que cada celdilla en la capa se convierte en un punto a interpolar.

Todos los métodos de interpolación se basan en la semejanza que tomarán los valores de cualquier variable cuantitativa que se midan en dos puntos, mientras más cercanos estén estos. No se puede afirmar que un método es superior a otro, ya que cada uno cumple una función determinada. El empleo de un método en específico está en dependencia de la situación espacial existente.

Descripción del proceso de interpolación

Para la realización de una correcta interpolación espacial se debe primeramente construir la grilla o malla que permite la regularización de los valores de Z, de forma que la estimación para el punto de grilla sea lo más cercano posible al valor esperado de Z.

Al construir la malla se forma una cuadrícula, la que se realiza mediante la utilización de un método de interpolación que en el caso de esta investigación se realiza mediante la función lineal o el método Spline Cúbico.

Cuando los valores están regularizados se realiza la interpolación aplicando un método que puede ser el mismo que se empleó para construir la grilla u otro de los mencionados

anteriormente. Este método utiliza funciones matemáticas que permiten establecer la estimación de los valores que alcanza una variable Z (valores a interpolar) en un conjunto de puntos definidos por un par de coordenadas (X, Y); obteniendo como resultado final una función que pase lo más cerca posible por los valores resultantes del proceso. Una vez formada la función que recorre todos los puntos se realiza el suavizado de contornos para mejorar la visualización de los resultados.

Para llevar a cabo el proceso de interpolación se hizo uso de la biblioteca para python SciPy la cual está organizada en sub-paquetes que cubren diferentes dominios de la computación científica como son: funciones especiales, integración, optimización, interpolación, transformada de Fourier, procesamiento de señales, álgebra lineal, estadística y procesamiento de imágenes (SciPy, 2013).

Ventajas del uso de SciPy:

- Posee módulos para optimización de funciones.
- Cubre gran parte de las necesidades básicas de un software científico.
- Es ampliamente utilizada.
- Emplea disímiles funciones de Matlab que pueden ser utilizadas en el procedimiento de interpolación, logrando obtener exactitud en los resultados.

La biblioteca seleccionada permite realizar el proceso mediante los métodos de Interpolación lineal, Spline Cúbico, Funciones de base radial (RBF), El Vecino más Cercano (nearest).

De estos métodos fueron seleccionados la Interpolación lineal y Spline Cúbico como candidatos debido a sus resultados sobre la variable a medir.

Interpolación lineal

La interpolación lineal es un método relativamente sencillo que se basa en la suposición de que la tasa de cambio entre los valores conocidos es constante y puede calcularse a partir de estos valores utilizando la fórmula de la pendiente simple. Entonces, un valor desconocido entre los dos puntos conocidos se puede calcular usando uno de los puntos y la tasa de cambio. Proporciona resultados de forma rápida y sencilla, debido a que solo realiza el cálculo de trayectoria de dos puntos. (IRI/LDEO, 2006).

Spline Cúbico

Estima valores analizados usando una función matemática que minimiza la curvatura general de la superficie, dando como resultado una superficie lisa que

pasa exactamente a través de los puntos de entrada. Este método es el mejor para suavizar superficies distintas, como la elevación, la altura de la capa freática, o las concentraciones de contaminación (GIS_Resources, 2013).

Ofrece un resultado continuo, suave y adaptable, si se ajustan adecuadamente sus parámetros se suele llegar a una interpolación de mayor precisión (PESQUER, y otros, 2007).

Lenguajes, Tecnologías y Herramientas utilizadas

Preprocesador de Hipertexto (PHP) 5

Lenguaje de programación código abierto que se emplea para programar del lado del servidor. Además de que GeneSIG emplea PHP en el desarrollo de sus componentes y que el módulo a desarrollar deberá integrarse a la misma, PHP permite lograr un sistema lo más libre de licencias posible cubriendo todas las necesidades.

Python 2.7

Python es un lenguaje orientado a objetos e interpretado, lo que significa que no se necesita compilar el código fuente para poder ejecutarlo, esto ofrece ventajas como la rapidez de desarrollo. Integra gran cantidad de bibliotecas que contienen tipos de datos y funciones incorporadas en el propio lenguaje. La selección del lenguaje se debe a su simplicidad lo que permitió integrar fácilmente los módulos psycopg2, SciPy, numpy y pyplot.

Se empleó psycopg2 para la conexión a la base de datos de PostgreSQL. SciPy permitió la utilización de la función interpolate que contiene funciones Spline, interpolación de Lagrange así como interpoladores polinómicos, los que fueron aplicados para realizar la interpolación. El módulo numpy permitió operar con matrices; este constituye una biblioteca de funciones matemáticas de alto nivel. Por último pyplot de la biblioteca matplotlib garantizó la realización de contornos para obtener los polígonos generados en la interpolación y la visualización de los resultados generados a través de su interfaz gráfica.

JavaScript

Es un lenguaje de programación sencillo y ligero del lado del cliente; no es conveniente usarlo como un lenguaje independiente, más bien está diseñado para una fácil incrustación en otros productos y aplicaciones, tales como los navegadores web (Network, 2005-2014.).

Framework de JavaScript ExtJS 3.2

Es una biblioteca de JavaScript para el desarrollo de aplicaciones web interactivas. Permite flexibilizar el manejo de componentes de la página y permite crear interfaces de usuario bastante funcionales. Además posibilita crear aplicaciones complejas utilizando componentes predefinidos (Ecured, 2012).

Editor de Cartografía Quantum GIS (QGIS) 1.7.5

(Ferri-Benedetti, 2012) define a QGIS como “un sistema de información geográfica que permite visualizar, editar e imprimir mapas, así como analizar todo tipo de datos demográficos y económicos usando la geografía como referencia.

Esta herramienta se utilizó para insertar en la cartografía los datos de la distribución poblacional de los consejos populares de cada municipio de la provincia Ciudad de La Habana, empleados para realizar la interpolación. Estos datos fueron almacenados simultáneamente en el gestor de base de datos PostgreSQL.

Sistemas Gestores de Bases de Datos (SGBD) PostgreSQL9.1

PostgreSQL es un sistema de gestión de bases de datos objeto-relacional, utilizado por ser un sistema de código abierto de gran robustez y por soportar datos espaciales mediante la extensión PostGIS (Martínez, 2010). Además permite almacenar gran cantidad de información y es empleado por la plataforma GeneSIG para gestionar los datos.

PostGIS 1.5.8

Es una extensión que convierte a PostgreSQL en una base de datos espacial, por lo que contiene cientos de funciones espaciales. Realiza la administración de la base de datos a través de pgAdmin. Posee soporte de datos ráster y vectoriales. Es compatible con el servidor de mapas MapServer (Morales, 2012).

Servidor Web Apache 2.2

Permite personalizar la respuesta ante los posibles errores que se puedan dar en el servidor y posibilita configurarlo para que ejecute un determinado conjunto de instrucciones cuando ocurra un error en concreto (CIBERAULA, 2012).

Servidor de Mapas MapServer 6.0

Es una plataforma de código abierto para la publicación de los datos espaciales y aplicaciones de mapas interactivos para la web. Es multiplataforma, soporta

formatos de datos ráster y vectorial, presenta una producción cartográfica avanzada y posibilita la automatización del mapa (barra de escala, mapa de referencia, y la leyenda) (MapServer, 2013-2014).

Para el desarrollo de este software se seleccionaron tales tecnologías debido a que son compatibles y posibilitan una buena unificación del módulo con GeneSIG. La integración de las herramientas se ve favorecida, pues muchas están construidas sobre el mismo lenguaje. Se tuvo en cuenta los principios establecidos por la Universidad de promover el software libre, con el fin de garantizar la soberanía tecnológica de la nación.

PRINCIPALES RESULTADOS

El diseño de un sistema se basa en su arquitectura, la cual está compuesta por varios estilos o patrones que posibilitan detallar un software con suficiente claridad, lo que facilitaría llevar a cabo la realización física del mismo. El diagrama de clases del diseño (ver Fig.13) muestra la integración del módulo de análisis desarrollado para la plataforma GeneSIG, cuya estructura está basada en el framework CartoWeb el cual tiene como característica sobre MapServer, que posee una arquitectura bastante modular y escalable, lo que permite poder separar la lógica del servidor (cartoserver), de la lógica del cliente (cartoclient).

Debido al gran número de datos que genera el proceso de interpolación y lo que puede llegar a tardar el mismo, se llevó a cabo la implementación del patrón SOA. Se utilizó básicamente para asegurar la continuidad de ejecución de procesos soportados por varios sistemas, haciendo el procesamiento de la información más rápido e integrado. Un servicio o Daemon (como se conoce en UNIX) es un proceso informático no interactivo; esto significa que se ejecuta en segundo plano en vez de ser controlado directamente por el usuario.

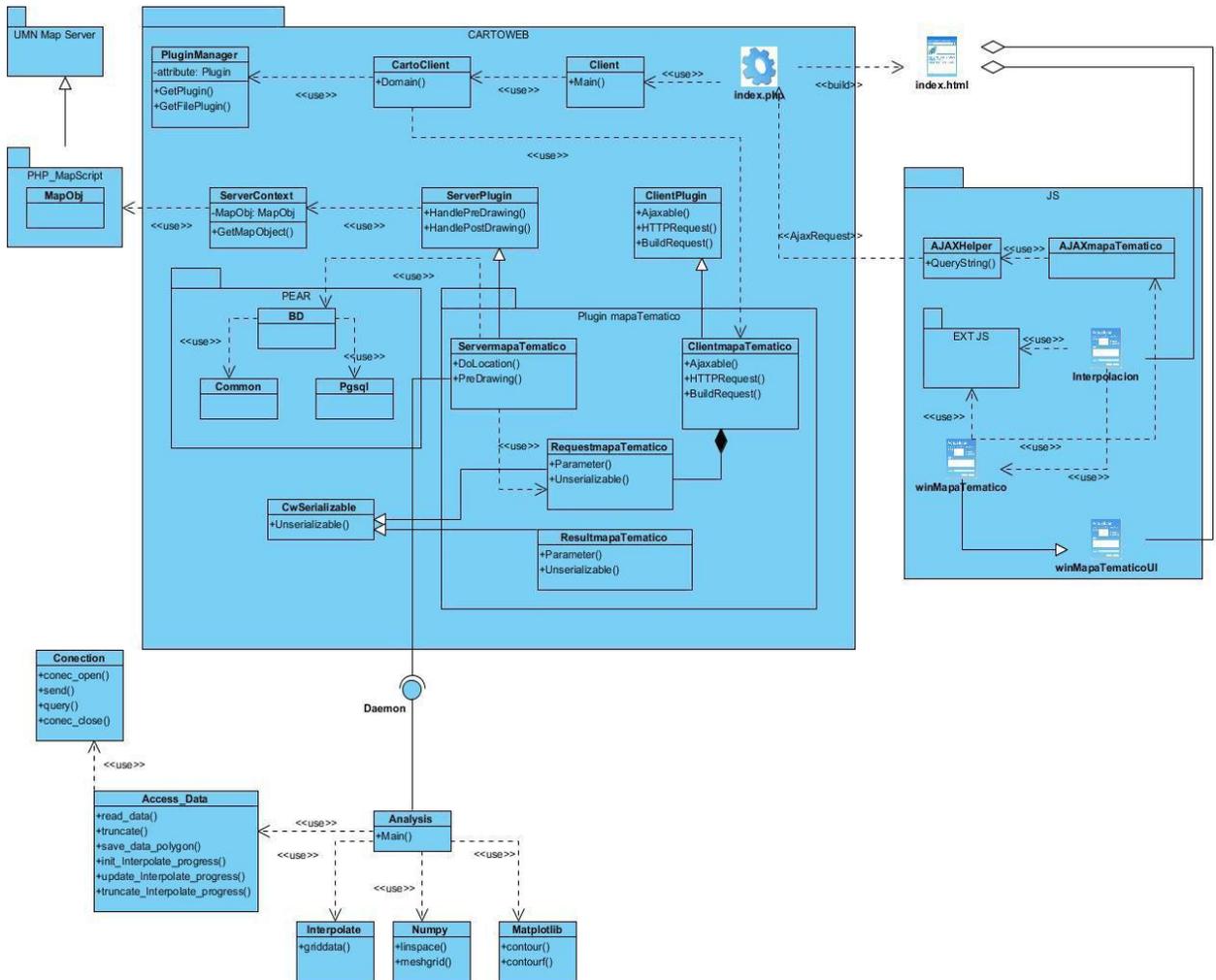
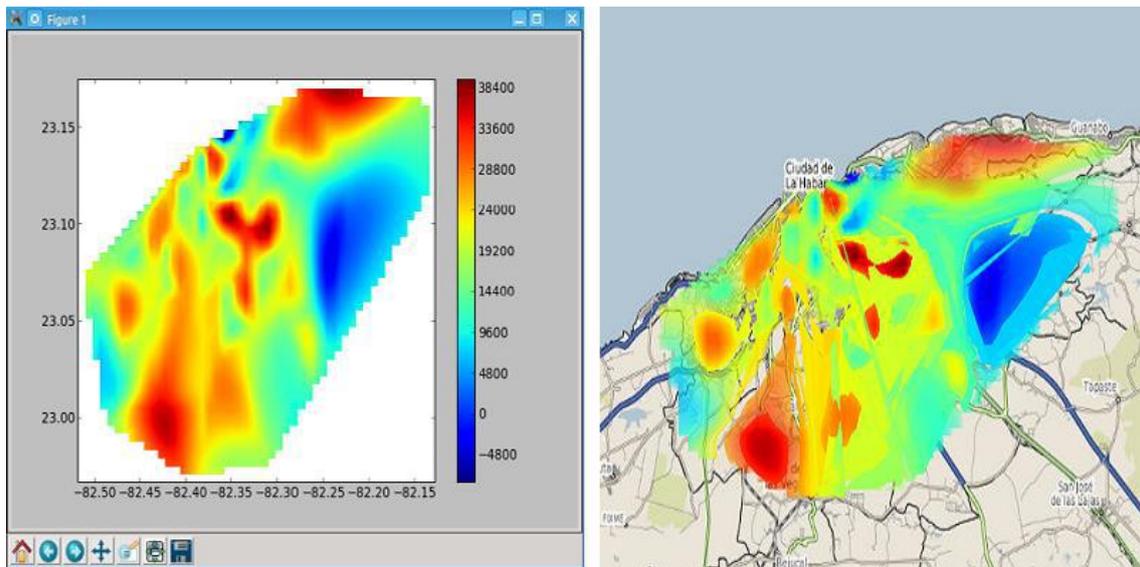


Fig. 13 Diagrama de clases del diseño

La investigación se centra en la Distribución poblacional de los consejos populares de la provincia de La Habana (variable a medir). Para extraer los datos que confirman la cantidad de personas que habitan en esta localidad se consultó como fuente a la Oficina Nacional de Estadísticas e Información (ONEI) generando para cada consejo popular la cantidad de habitantes.

Una vez terminada la fase de implementación se obtiene como resultado un módulo capaz de realizar la tematización mediante el proceso de interpolación, por los métodos Spline Cúbico y Lineal, con niveles especificados por el usuario. Actualmente solo representa polígonos gestionados por MapServer; mientras que la biblioteca de visualización gráfica pyplot, la cual constituye la principal referencia en cuanto a visualización de la tematización, permite la graficación de rutas (paths) compuestas por líneas, curvas y saltos.

Al representar los resultados obtenidos de ejecutar el proceso de interpolación, para 950 niveles y con el método “Spline Cúbico”, sobre la distribución poblacional de los consejos populares de la provincia La Habana, mediante la librería de visualización pyplot del módulo matplotlib de python (A) y el servidor de mapas MapServer (B), se detecta dificultades en el proceso de visualización en B respecto a; siendo A la representación exacta de los resultados de la interpolación.



(A) Interfaz de la biblioteca gráfica pyplot

(B) Plataforma GeneSIG

Fig. 18 Tematización de la provincia La Habana por métodos de interpolación

La inexactitud en la visualización se debe a que, como PostGIS en su versión 1.5 no tiene concebido la forma de almacenar objetos geográficos de tipo rutas (compuestas por líneas, curvas y saltos) y que el servidor de mapas MapServer solo permite definir color de fondo a los objetos geográficos cerrados, se procedió a transformar las rutas obtenidas mediante el proceso de interpolación a polígonos compuestos por solo rectas; lo cual condiciona pérdida en la precisión de la información generada por el proceso de interpolación desarrollado. Estos polígonos son los que se almacenan en el gestor de base de datos, para posteriormente ser dibujados por el servidor de mapas.

CONCLUSIONES

La solución propuesta garantiza la obtención de resultados precisos y confiables sobre los datos analizados en las tematizaciones realizadas en la plataforma Ge-

neSIG, lo que posibilita un aporte significativo para el proceso de toma de decisiones a través de la utilización de personalizaciones basadas en esta plataforma.

El “Módulo de análisis basado en métodos de interpolación para GeneSIG” brinda a la sociedad un sistema completamente libre, que posibilita la obtención de resultados precisos sobre los datos analizados, teniendo en cuenta la variable a medir: distribución poblacional, temperatura, suelos, zonas de altas y bajas presiones, etc. Los resultados generados por el software permiten detectar la existencia de áreas propensas a formación de ciclones tropicales u otros fenómenos naturales.

REFERENCIAS

SciPy. 2013. SciPy Tutorial. [En línea] 21 de Octubre de 2013. [Citado el: 01 de Marzo de 2014.] <http://docs.scipy.org/doc/scipy/reference/>.

IRI/LDEO. 2006. IRI/LDEO Climate Data Library. Interpolation Techniques. [En línea] 2006. [Citado el: 08 de Diciembre de 2013.] <http://iridl.ldeo.columbia.edu/dochelp/StatTutorial/Interpolation/>.

GIS_Resources. 2013. GIS Resources. Types of Interpolation Methods. [En línea] 2013. http://www.gisresources.com/types-interpolation-methods_3/.

PESQUER , L., MASÓ , J. y PONS , X. 2007. INTEGRACIÓN EN UN SIG DE MÉTODOS ESTADÍSTICOS, INTERPOLACIÓN E IMÁGENES DE TELEDETECCIÓN PARA LA OBTENCIÓN MENSUAL DE DATOS METEOROLÓGICOS VALIDADOS. Barcelona : s.n., 2007.

Network, Mozilla Developer. 2005-2014.. Mozilla Developer Network. . JavaScript. . [En línea] 2005-2014. [Citado el: 24 de Marzo de 2014.] <https://developer.mozilla.org/en-US/docs/Web/JavaScript..>

Ecured. 2012-2013. Ecured. Sencha Ext JS. [En línea] 2012-2013. [Citado el: 11 de Diciembre de 2013.]

http://www.ecured.cu/index.php/Sencha_Ext_JS .

Ferri-Benedetti, Fabrizio. 2012. Quantum GIS. Datos y mapas, una gran unión. [En línea] 24 de Abril de 2012. [Citado el: 22 de Diciembre de 2013.] <http://quantum-gis.softonic.com>.

Martínez, Rafael. 2010. PostgreSQL. Portal en español sobre PostgreSQL. [En línea] 02 de Octubre de 2010. [Citado el: 24 de Diciembre de 2013.] http://www.postgresql.org.es/sobre_postgresql.

Morales, Aurelio. 2012. MappingGIS. Por qué utilizar PostGIS. [En línea] 19 de Septiembre de 2012. [Citado el: 14 de Enero de 2014.] <http://mappinggis.com/2012/09/por-que-utilizar-postgis/>.

CIBERAULA. 2012. Ciberaula. Apache. [En línea] 2012. [Citado el: 20 de Marzo de 2014.] http://linux.ciberaula.com/articulo/linux_apache_intro/.

MapServer. 2013-2014. MapServer . MapServer open source web mapping. [En línea] 2013-2014. [Citado el: 12 de Enero de 2014.] <http://mapserver.org/about.html#about> .



www.sociedadelainformacion.com

Edita:



Director: José Ángel Ruiz Felipe
Jefe de publicaciones: Antero Soria Luján
D.L.: AB 293-2001
ISSN: 1578-326x