

# Análisis geo-visual de datos con gvSIG

Autor: Romel Vázquez Rodríguez

Departamento de Ciencias de la Computación  
Universidad Central "Marta Abreu" de Las Villas

Valencia, 2015



# Índice

- 1 Motivación e introducción
- 2 Visualización científica
- 3 Sistemas de información geográfica
- 4 Análisis exploratorio de datos
- 5 Integración de visualización científica con sistemas de información geográfica para el análisis exploratorio de datos
- 6 Análisis exploratorio de datos con baja densidad espacial
- 7 Análisis exploratorio de datos con alta densidad espacial
- 8 Herramientas para el soporte de archivos de formatos de datos científicos en sistemas de información geográfica
- 9 Conclusiones y trabajos futuros

## Motivación e introducción

- El análisis de secuencias temporales de datos en forma de mallas regulares es un reto.

## Motivación e introducción

- El análisis de secuencias temporales de datos en forma de mallas regulares es un reto.
- Estos datos comúnmente contienen información en el dominio espacial y representan mediciones de múltiples variables.

## Motivación e introducción

- El análisis de secuencias temporales de datos en forma de mallas regulares es un reto.
- Estos datos comúnmente contienen información en el dominio espacial y representan mediciones de múltiples variables.
- Varios problemas deben ser resueltos para analizar grandes conjuntos de datos con toda esta información de una manera consistente y efectiva.

## Motivación e introducción

- El análisis de secuencias temporales de datos en forma de mallas regulares es un reto.
- Estos datos comúnmente contienen información en el dominio espacial y representan mediciones de múltiples variables.
- Varios problemas deben ser resueltos para analizar grandes conjuntos de datos con toda esta información de una manera consistente y efectiva.
- La mayoría de los métodos y herramientas actuales no permiten un análisis espacio-temporal efectivo de múltiples variables simultáneamente.

## Motivación e introducción

- El análisis de secuencias temporales de datos en forma de mallas regulares es un reto.
- Estos datos comúnmente contienen información en el dominio espacial y representan mediciones de múltiples variables.
- Varios problemas deben ser resueltos para analizar grandes conjuntos de datos con toda esta información de una manera consistente y efectiva.
- La mayoría de los métodos y herramientas actuales no permiten un análisis espacio-temporal efectivo de múltiples variables simultáneamente.
- Este trabajo trata el problema del análisis visual de secuencias temporales de múltiples variables con baja y alta densidad espacial.

## Motivación e introducción

- El análisis de secuencias temporales de datos en forma de mallas regulares es un reto.
- Estos datos comúnmente contienen información en el dominio espacial y representan mediciones de múltiples variables.
- Varios problemas deben ser resueltos para analizar grandes conjuntos de datos con toda esta información de una manera consistente y efectiva.
- La mayoría de los métodos y herramientas actuales no permiten un análisis espacio-temporal efectivo de múltiples variables simultáneamente.
- Este trabajo trata el problema del análisis visual de secuencias temporales de múltiples variables con baja y alta densidad espacial.
- Se presentan casos de estudio que demuestran la efectividad de las herramientas y métodos desarrollados.

## El problema

- El principal problema es que el análisis de series temporales distribuidas espacialmente se dificulta cuando se analizan múltiples variables simultáneamente.

## El problema

- El principal problema es que el análisis de series temporales distribuidas espacialmente se dificulta cuando se analizan múltiples variables simultáneamente.
- La integración de técnicas de visualización científica con sistemas de información geográfica facilita la solución de este problema.

## El problema

- El principal problema es que el análisis de series temporales distribuidas espacialmente se dificulta cuando se analizan múltiples variables simultáneamente.
- La integración de técnicas de visualización científica con sistemas de información geográfica facilita la solución de este problema.
- La principal ventaja de usar técnicas de visualización científica en SIG consiste en la posibilidad de:

## El problema

- El principal problema es que el análisis de series temporales distribuidas espacialmente se dificulta cuando se analizan múltiples variables simultáneamente.
- La integración de técnicas de visualización científica con sistemas de información geográfica facilita la solución de este problema.
- La principal ventaja de usar técnicas de visualización científica en SIG consiste en la posibilidad de:
- Hacer comparaciones, identificar patrones, correlaciones, detectar anomalías, variabilidad.

## Concepto

La visualización científica transforma los datos científicos y abstractos en imágenes.



## Datos multiparamétricos

- Consisten en un número de  $m$  registros donde cada uno está definido por un vector de  $n$  valores.

## Datos multiparamétricos

- Consisten en un número de  $m$  registros donde cada uno está definido por un vector de  $n$  valores.
- Estos datos pueden ser vistos como una matriz de  $m * n$ , donde cada fila representa un registro y cada columna representa una variable o dimensión.

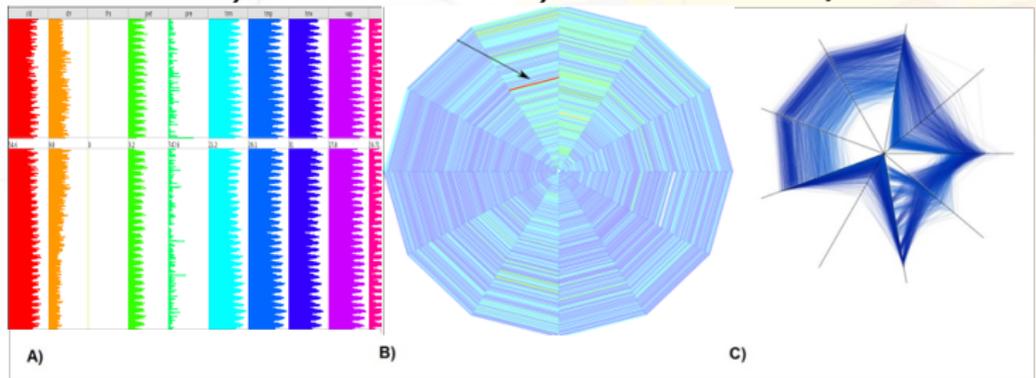
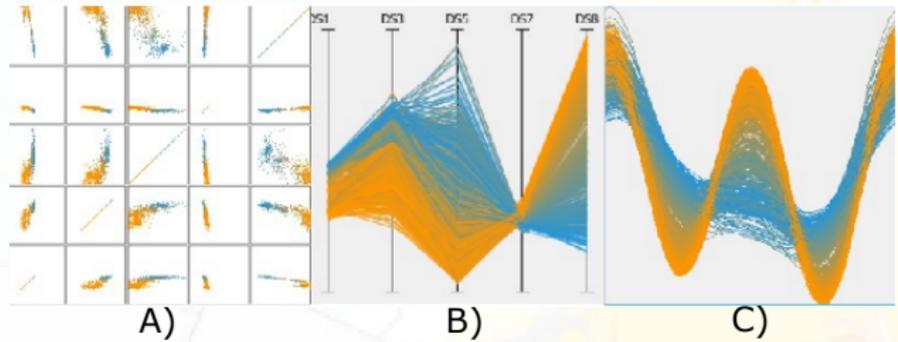
## Datos multiparamétricos

- Consisten en un número de  $m$  registros donde cada uno está definido por un vector de  $n$  valores.
- Estos datos pueden ser vistos como una matriz de  $m * n$ , donde cada fila representa un registro y cada columna representa una variable o dimensión.

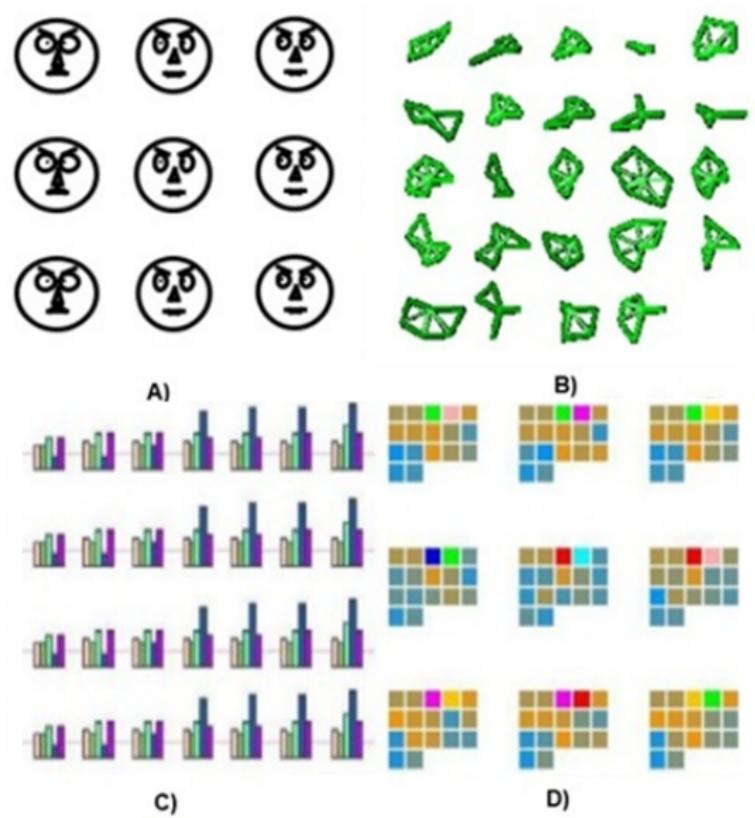
## Las técnicas de visualización de datos multiparamétricos

- 1 Geométricas
- 2 Basadas en íconos.
- 3 Basadas en píxel.
- 4 Otras

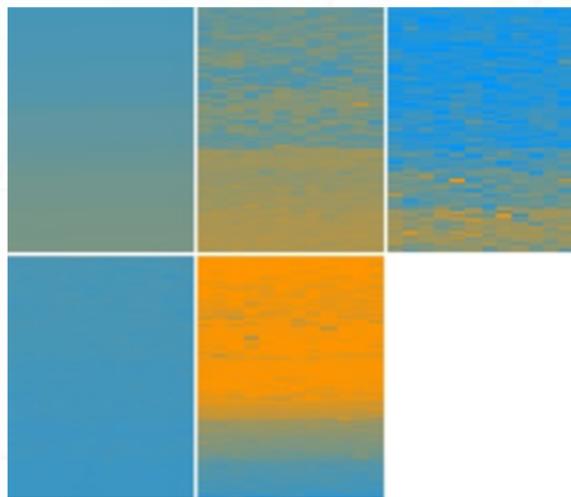
# Técnicas geométricas



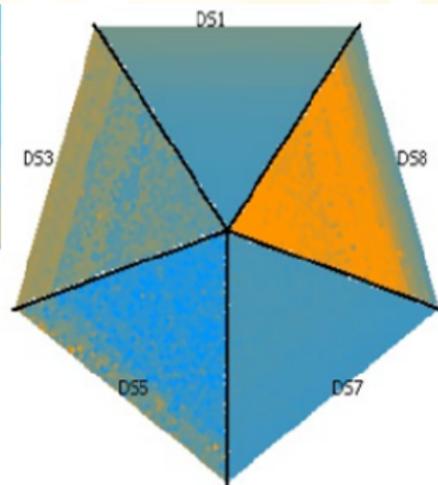
# Técnicas basadas en íconos



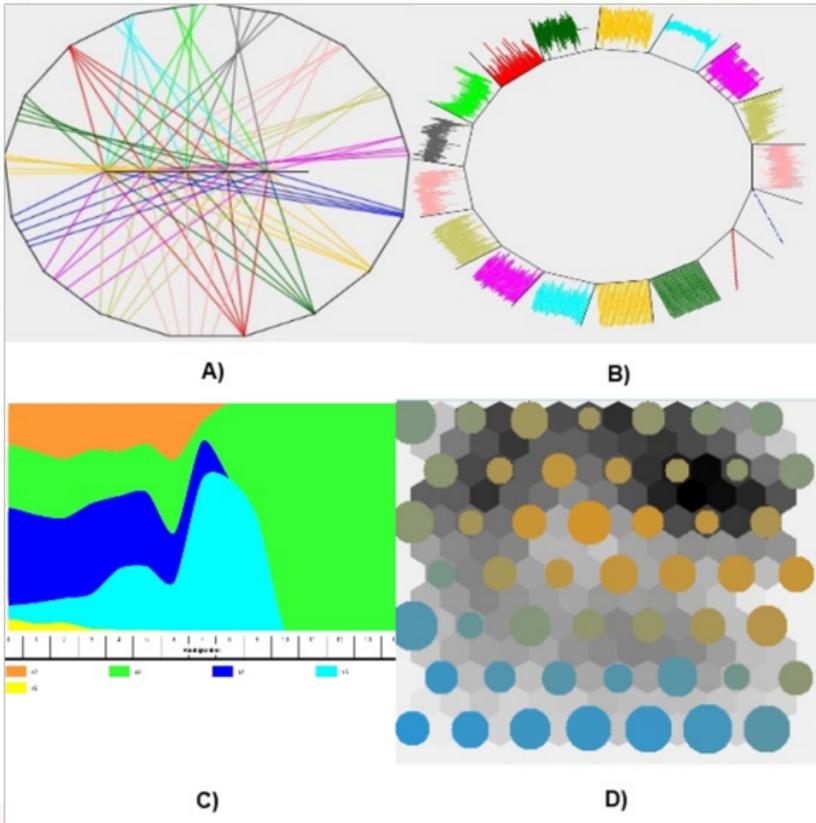
## Técnicas basadas en píxel

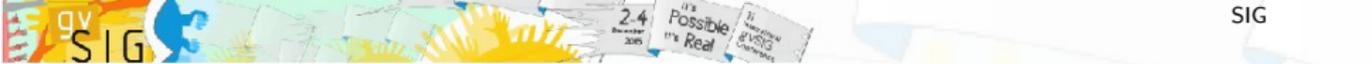


A)



B)





# Sistemas de información geográfica

## SIG estudiados

- ArcGIS.
- GRASS GIS
- Quantum GIS
- gvSIG
- Open JUMP
- uDIG
- Sextante

# Análisis exploratorio de datos. Análisis visual

## Introducción al AED

- El AED tiene sus orígenes en la estadística
- Fue el regreso a los objetivos originales de la estadística, por ejemplo detectar y describir patrones, tendencias, y relaciones entre los datos (Tukey, 1977).
- El AED está más relacionado con la generación de hipótesis que con la prueba de hipótesis (Andrienko y Andrienko, 2006)
- Está muy relacionado con el Análisis visual de datos.

Se puede estudiar desde la perspectiva de:

- Datos
- Tareas
- Herramientas
- Principios

## Integración entre SIG y VisCi

La VisCi y los SIG se desarrollaron en paralelo y de forma independiente

- Los sistemas de visualización tenían dificultades para incorporar formatos de datos geográficos. Eran fuertes en la manipulación de formatos de datos científicos.
- Los SIG tenían pobres herramientas de visualización. Incorporaban múltiples funciones de análisis geográfico.

Niveles de integración entre ambas disciplinas Rhyne (1997)

- Rudimentario
- Operacional
- Funcional
- Mezclado

# Integración. Herramientas

## Intentos de integración

- Snap-Together Visualization
- Geovista Studio
- VIS-STAMP
- GAV Flash tools
- ArcView-xGobi

## Desventajas de estas herramientas

- No incorporan formatos de datos científicos
- Algunos no incorporan grandes cantidades de técnicas de VisCi
- Algunos no permiten la manipulación completa de sus componentes
- Son pobres en el análisis de series temporales mediante visualizaciones
- No soportan secuencias temporales en mapas raster

# Integración. Formatos de datos científicos

## Principales formatos de datos científicos

- HDF Hierarchical Data Format
- HDF-EOS Hierarchical Data Format - Earth Observing System
- CDF Common Data Format
- NetCDF Network Common Data Format
- FITS Flexible Image Transport System

## Ventajas formatos de datos científicos

- Modelos simples
- Metadatos
- Manipulación de grandes volúmenes de datos
- Portabilidad

# AED para la baja densidad espacial

## Datos espacio-temporales con baja densidad espacial y amplios en el tiempo

- datos históricos sobre estaciones meteorológicas donde se miden múltiples variables
- datos de censos tomados en regiones delimitadas por la distribución político-administrativa (municipios, provincias, países, etc.)
- pocos conjuntos de datos multiparamétricos que pueden contener muchos registros temporales.



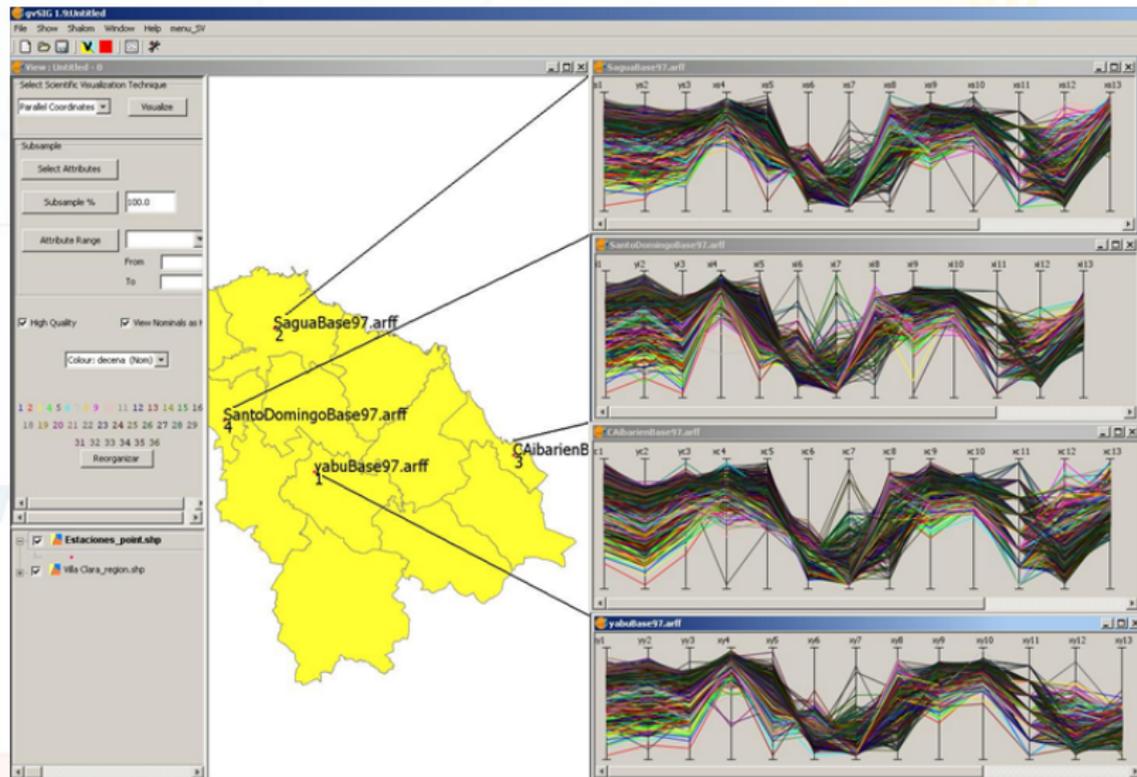


# Enfoque general

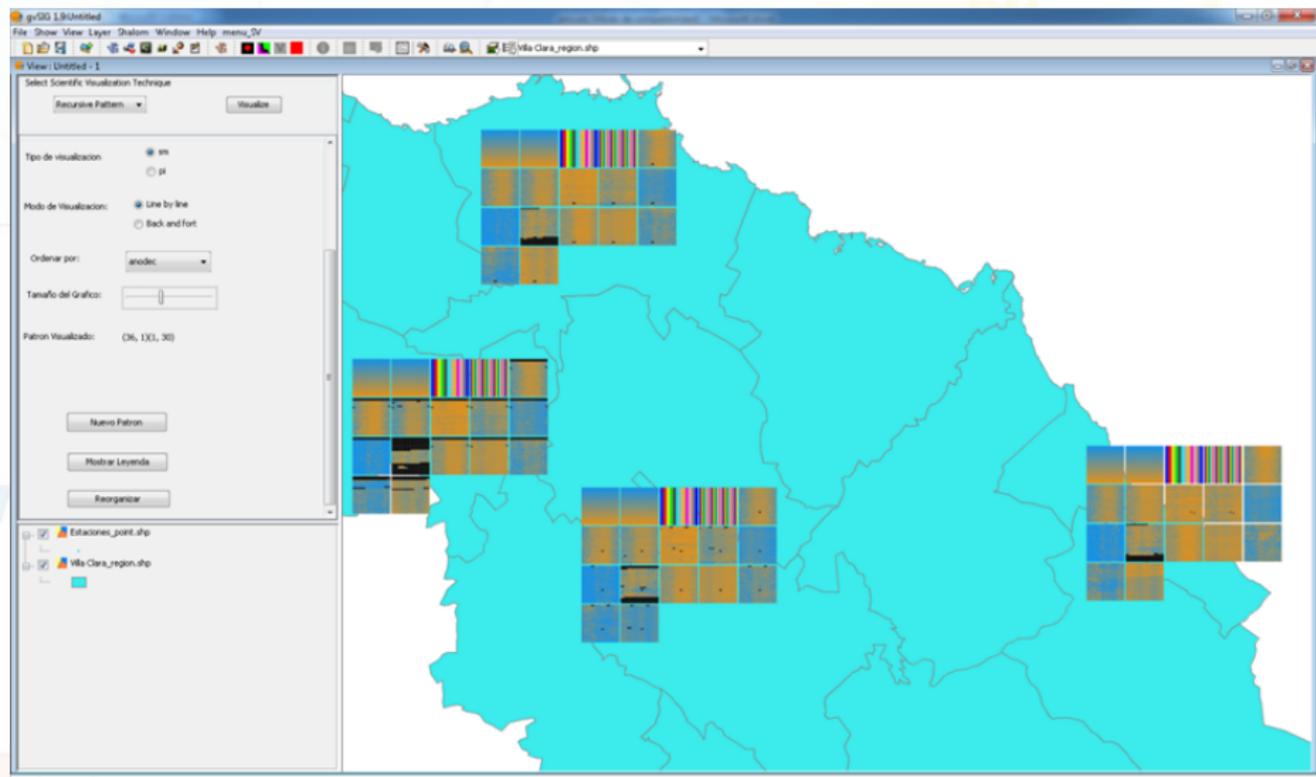
## Características

- gvSIG SIG Español de código abierto, libre, modular y fácil de extender.
- El módulo de visualización científica fue agregado con las técnicas: Coordenadas Paralelas. Gráfico de Andrews. Icono en forma de estrella. Shapecoding. Icono en forma de barras. Segmentos de Círculo. Patrones Recursivos.
- Datos multiparamétricos en formatos planos arff y dbf. Capas vectoriales en formato shp.

# Caso de estudio. Datos meteorológicos de la provincia de Villa Clara



# Caso de estudio. Datos meteorológicos de provincia la de Villa Clara



# Caso de estudio. Datos meteorológicos de la provincia de Villa Clara

Los principales resultados fueron publicados en la revista brasileña *Anuario do Instituto de Geociências* Vázquez-Rodríguez *et al.* (2013b)



Anuário do Instituto de Geociências - UFRJ  
www.anuario.igeo.ufrj.br

**A Novel Visual Data Mining Module for the Geographical Information System gvSIG**  
Um Novo Módulo de Mineração VISUAL de Dados para o Sistema de Informação Geográfica gvSIG

Romel Vázquez-Rodríguez<sup>1</sup>; Carlos Pérez-Risquet<sup>1</sup> & Juan Carlos Torres-Cantero<sup>2</sup>

<sup>1</sup>Center of Studies on Informatic, Central University of Las Villas, Santa Clara, Cuba.  
Carretera a Camajuaní km 5 ½, Santa Clara, Villa Clara. CP 54830 Cuba

<sup>2</sup>Virtual Reality Laboratory, Granada University, Granada, Spain  
E-mails: romel@uclv.edu.cu; cperez@uclv.edu.cu; jctorres@ugr.es

Recibido em: 12/09/2012 Aprovado em: 18/07/2013

DOI: [http://dx.doi.org/10.11137/2013\\_1\\_98\\_111](http://dx.doi.org/10.11137/2013_1_98_111)

# Caso de estudio. Datos meteorológicos de la provincia de Villa Clara

Presentados en la Conferencia Internacional de Descubrimiento de Conocimiento y Recuperación de Información (KDIR) Vázquez-Rodríguez *et al.* (2010a).



# AED para la alta densidad espacial

Datos espacio-temporales con alta densidad espacial y amplios en el tiempo

- El desarrollo de la aeronáutica ha permitido desplegar sensores remotos que recogen información sobre la superficie terrestre.

# AED para la alta densidad espacial

Datos espacio-temporales con alta densidad espacial y amplios en el tiempo

- El desarrollo de la aeronáutica ha permitido desplegar sensores remotos que recogen información sobre la superficie terrestre.
- Se han creado secuencias temporales de múltiples variables para una alta cantidad de lugares del espacio terrestre.

## AED para la alta densidad espacial

Datos espacio-temporales con alta densidad espacial y amplios en el tiempo

- El desarrollo de la aeronáutica ha permitido desplegar sensores remotos que recogen información sobre la superficie terrestre.
- Se han creado secuencias temporales de múltiples variables para una alta cantidad de lugares del espacio terrestre.

Problemas del análisis de grandes volúmenes de datos espacio-temporales

- Las secuencias no pueden ser analizadas con simples imágenes de una variable, hay que tener en cuenta el tiempo.

## AED para la alta densidad espacial

Datos espacio-temporales con alta densidad espacial y amplios en el tiempo

- El desarrollo de la aeronáutica ha permitido desplegar sensores remotos que recogen información sobre la superficie terrestre.
- Se han creado secuencias temporales de múltiples variables para una alta cantidad de lugares del espacio terrestre.

### Problemas del análisis de grandes volúmenes de datos espacio-temporales

- Las secuencias no pueden ser analizadas con simples imágenes de una variable, hay que tener en cuenta el tiempo.
- El solapamiento de todos los mapas provoca que se tengan que analizar en secuencias de imágenes o animaciones.



## AED para la alta densidad espacial

Datos espacio-temporales con alta densidad espacial y amplios en el tiempo

- El desarrollo de la aeronáutica ha permitido desplegar sensores remotos que recogen información sobre la superficie terrestre.
- Se han creado secuencias temporales de múltiples variables para una alta cantidad de lugares del espacio terrestre.

### Problemas del análisis de grandes volúmenes de datos espacio-temporales

- Las secuencias no pueden ser analizadas con simples imágenes de una variable, hay que tener en cuenta el tiempo.
- 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 visual espacio-temporal de múltiples variables a la vez.

# AED para la alta densidad espacial

La solución a los anteriores problemas ha dado origen a:

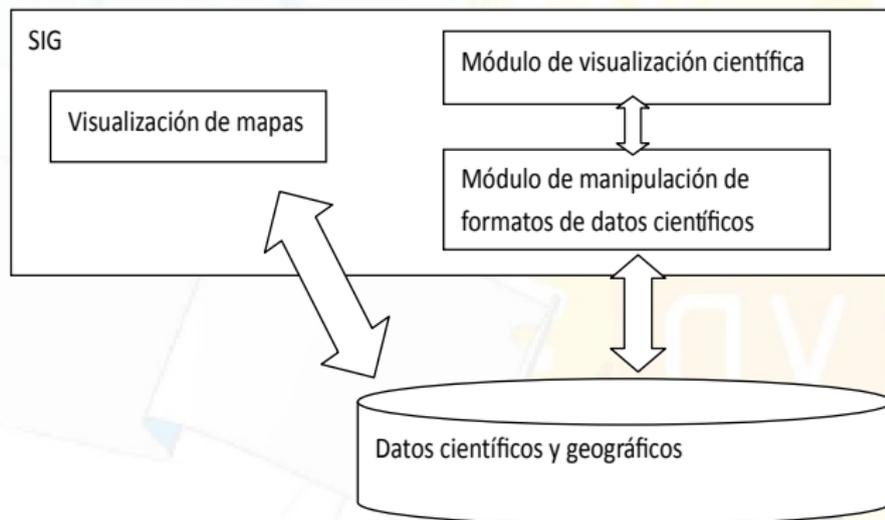
- Un nuevo método para la visualización de series temporales de múltiples variables, mediante la utilización de técnicas de visualización de datos multiparamétricos integradas en sistema de información geográfica.
- Casos de estudio donde se demuestra la efectividad de las técnicas para extraer tendencias, correlaciones y patrones.

Ejemplo de datos espacio-temporales con alta densidad espacial

Dataset	Espacio	Tiempo	Variedad	Variables
CRU TS3.21	0.5° global	1901- 2012	Series temporales de observación	pre, tmp, tmx, tmn, dtr, vap, cld, wet, frs, pet

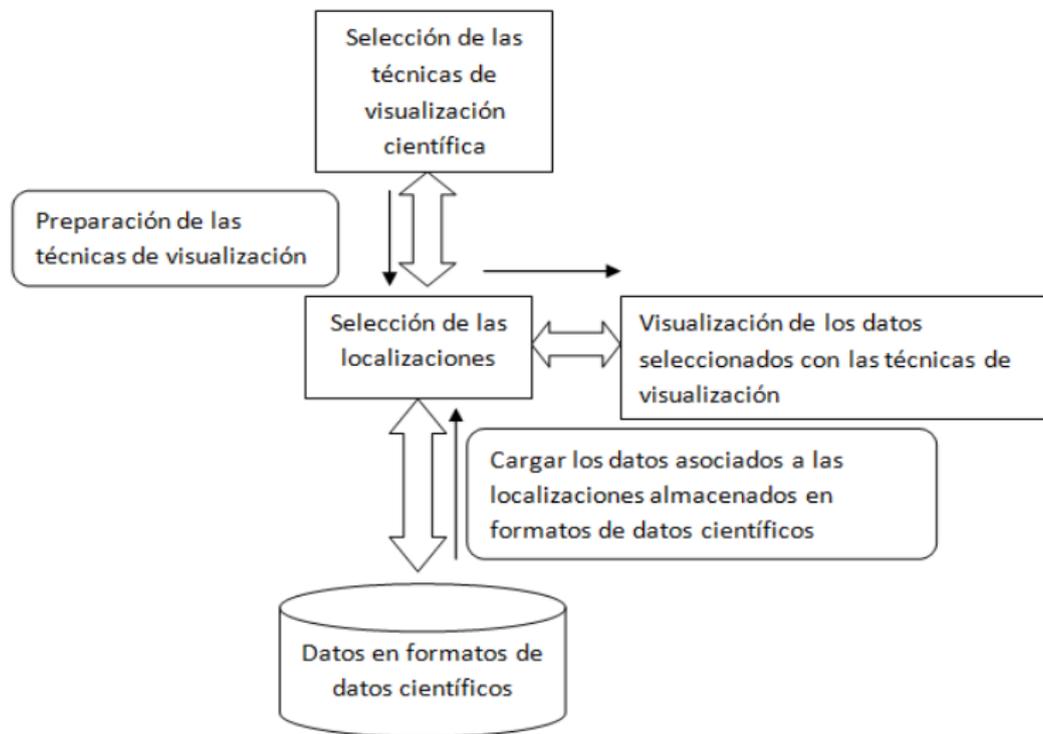
# AED para la alta densidad espacial

## Arquitectura general para el AED con alta densidad espacial



# AED para la alta densidad espacial

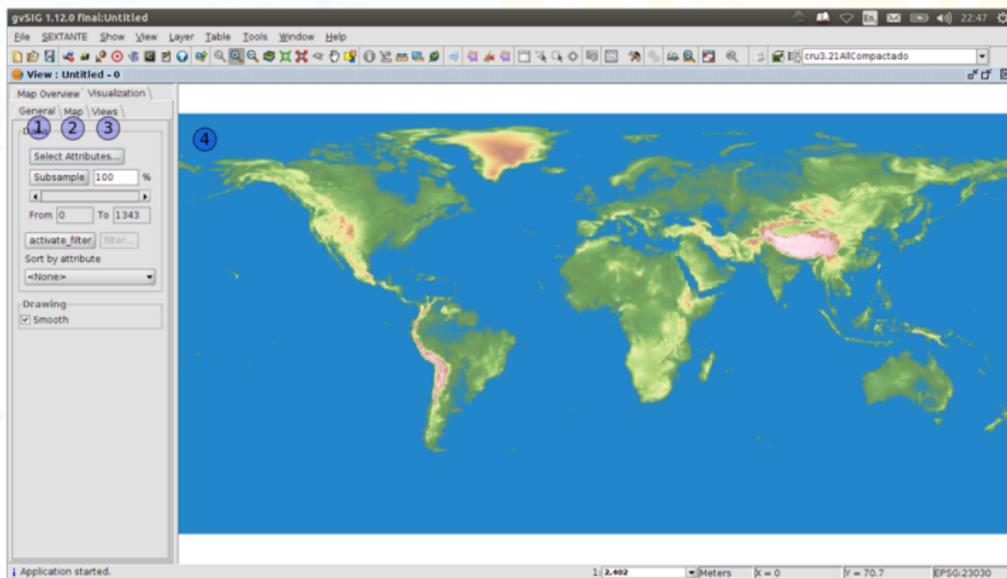
## Interacción





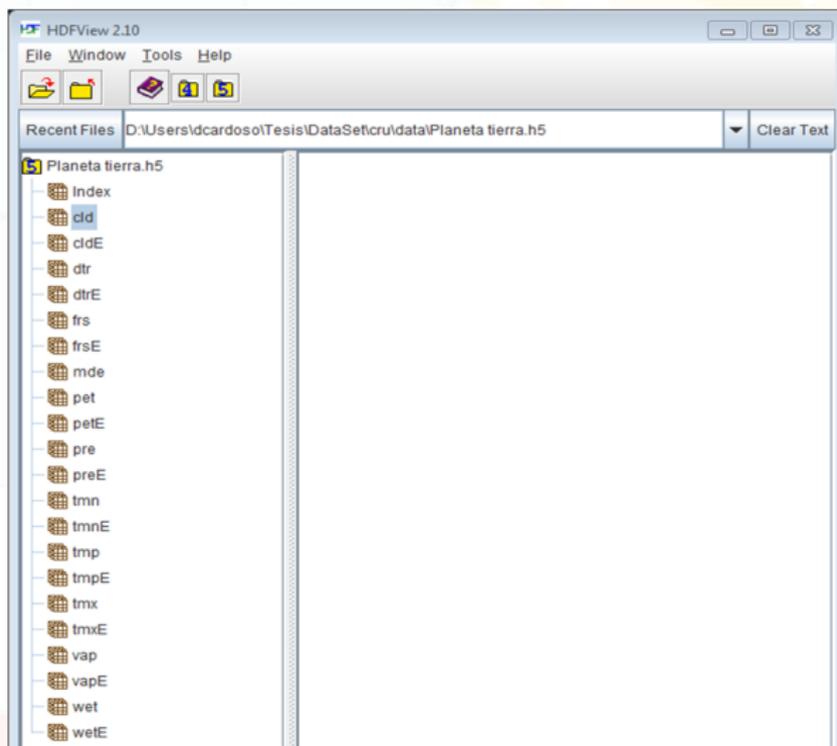
# AED para la alta densidad espacial

## Visualización del mapa de fondo



# AED para la alta densidad espacial

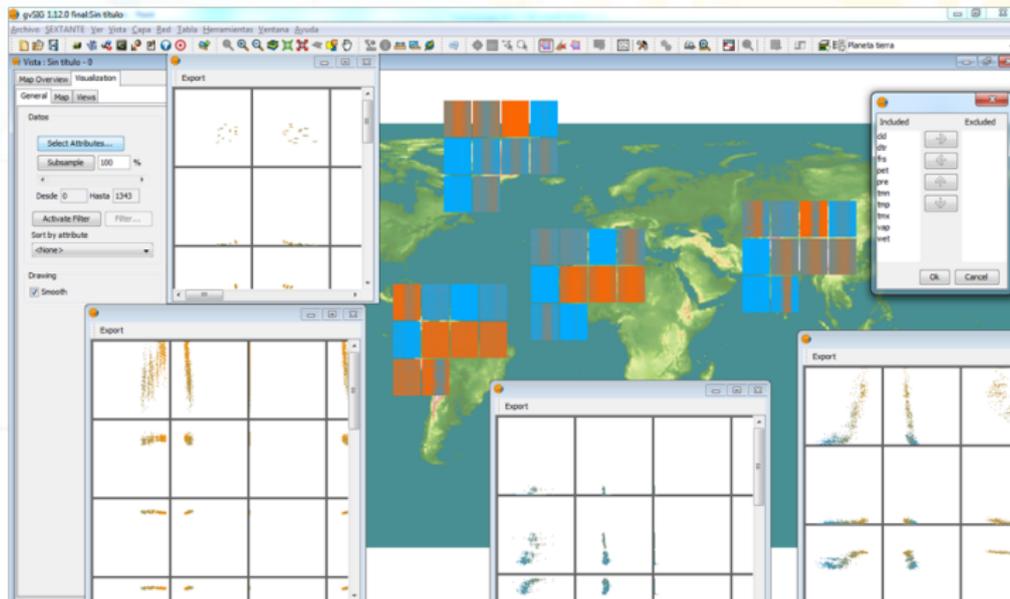
Estructura de un conjunto de datos HDF con alta densidad espacial





# AED para la alta densidad espacial

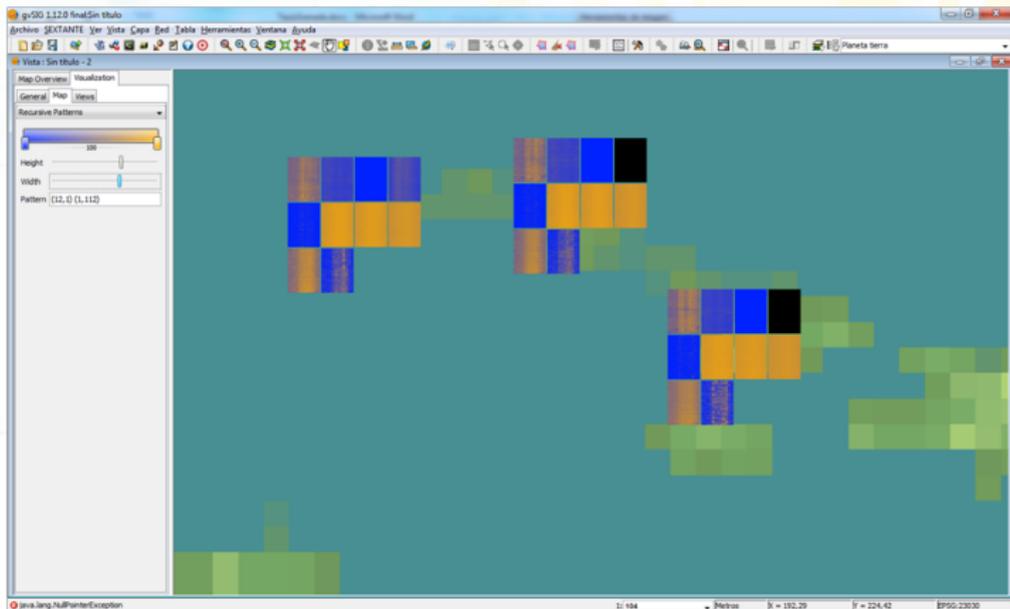
## Caso de estudio. Datos mundiales





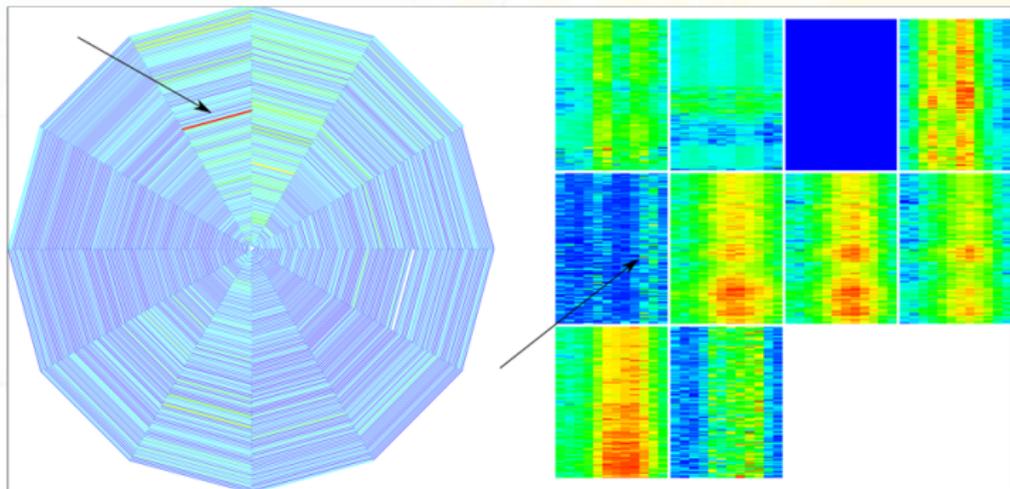
# AED para la alta densidad espacial

## Caso de estudio. Enfocado en Cuba



# AED para la alta densidad espacial

Caso de estudio. Enfocado en el Oriente de Cuba





# Resultados

Los resultados fueron publicados en la *Revista Técnica de la Facultad de Ingeniería de la Universidad del Zulia* Vázquez-Rodríguez et al. (2015)

ISSN 0354-0770  
Depósito legal pp 1978022038

## REVISTA TECNICA

DE LA FACULTAD DE INGENIERÍA  
UNIVERSIDAD DEL ZULIA  
MARACAIBO - VENEZUELA



Una Revista Internacional Arbitrada  
que está indexada en las publicaciones  
de referencia y comentaristas:

- Science Citation Index (SCIExpanded)
- Compendex
- Chemical Abstracts
- Metal Abstracts
- World Aluminium Abstracts
- Mathematical Reviews
- Petroleum Abstracts
- Zentralblatt für Mathematik
- Current Mathematical Publications
- MathSci (online database)
- Revescylt
- Materials Information
- Periódica
- Actualidad Iberoamericana

VOL. 38                      ABRIL 2015                      No. 1

Rev. Téc. Ing. Univ. Zulia. Vol. 38, N° 1, 73 - 82, 2015

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

Romel Vázquez-Rodríguez<sup>1</sup>, Carlos Pérez-Risquet<sup>1</sup>,  
Juan Carlos Torres<sup>2</sup>

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

<sup>2</sup>Laboratorio de Realidad Virtual, Universidad de Granada. jctorres@ugr.es

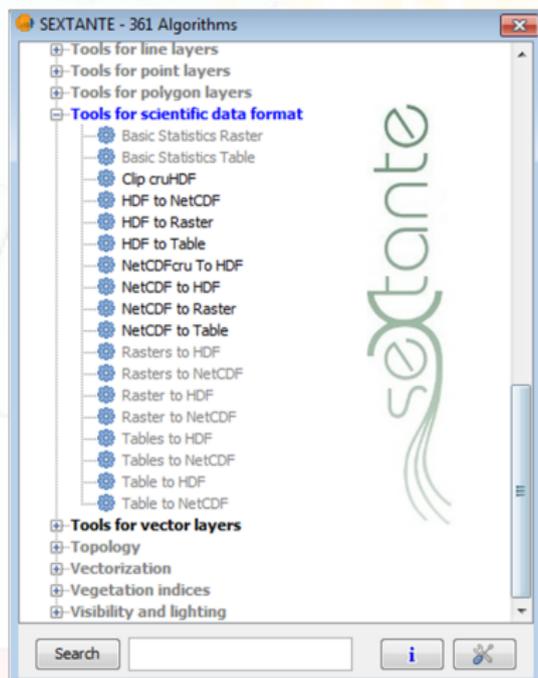
# Herramientas para la manipulación de formatos de datos científicos

## Selección de la tecnología

- HDF y NetCDF fueron los formatos seleccionados para la incorporación en SIG.
- Los SIG seleccionados para la incorporación de FDC fueron gvSIG y la biblioteca Sextante.
- se utilizaron las bibliotecas HDF-Java y NetCDF 4.2.

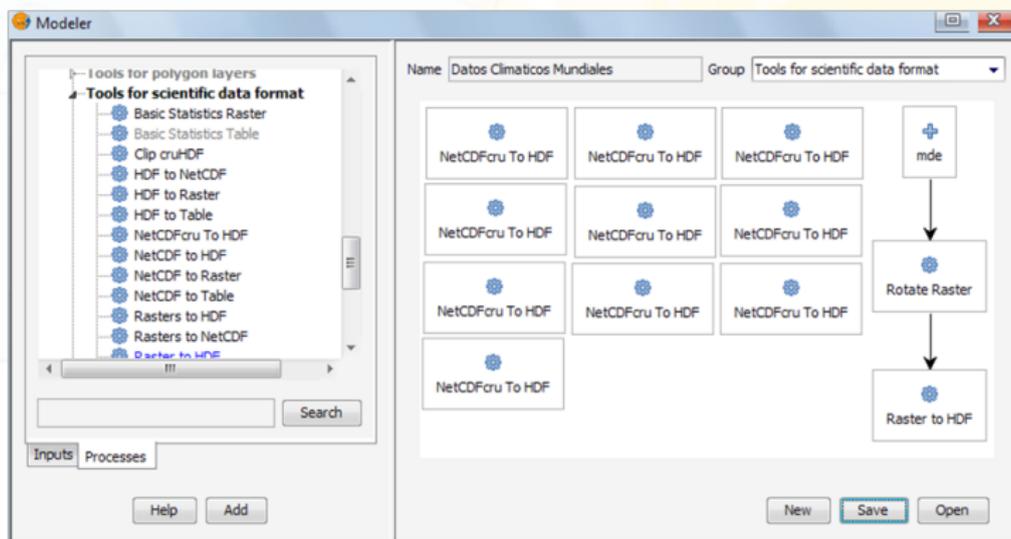
# Herramientas para la manipulación de formatos de datos científicos

Algoritmos incorporados a Sextante



# Herramientas para la manipulación de formatos de datos científicos

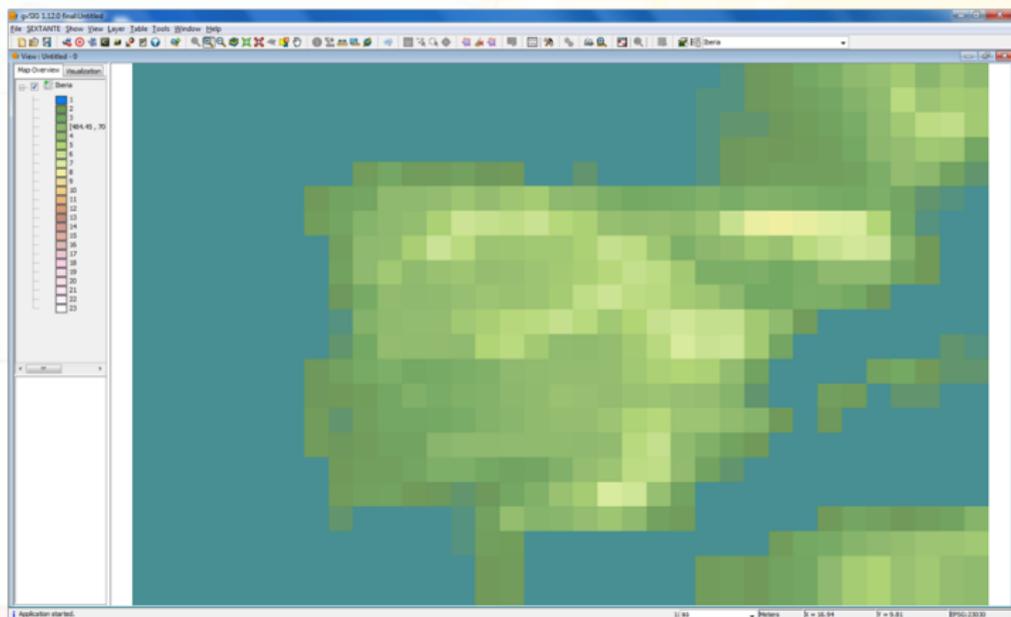
Uso del modelizador gráfico de Sextante. Caso de estudio.





# Herramientas para la manipulación de formatos de datos científicos

Caso de estudio construcción de un dataset HDF con los datos climáticos de la península ibérica



# Conclusiones y trabajos futuros

## Conclusiones

- Se crearon herramientas en gvSIG que permiten el descubrimiento de patrones, tendencias, anomalías y relaciones entre los datos. Las herramientas incluyen módulos de visualización científica y módulos para la manipulación de formatos de datos científicos como HDF y NetCDF.
- Se comprobó mediante casos de estudio la viabilidad de la utilización de las herramientas para la extracción de conocimiento, reconocimiento de patrones, tendencias, anomalías y relaciones entre múltiples variables.

## Referencias I

- Bolstad, P. 2005. *GIS fundamentals: A first text on geographic information systems*, Eider Pr.
- Clarke, Keith. 1990. *Analytical and Computer Cartography*, Prentice Hall Professional Technical Reference, first edn.
- Star, Jeffrey y John Estes. 1990. *Geographic information systems*, prentice-Hall Englewood Cliffs.
- Tukey, J.W. 1977. *Exploratory Data Analysis*, Reading, MA: Addison-Wesley.
- Andrienko, Natalia y Gennady Andrienko. 2006. *Exploratory analysis of spatial and temporal data*, Berlin, Germany: Springer-Verlag.
- Yi, Ji Soo, Youn ah Kang, John T Stasko, y Julie A Jacko. 2007. Toward a deeper understanding of the role of interaction in information visualization, *Visualization and Computer Graphics, IEEE Transactions on*, 13(6), 1224–1231.

## Referencias II

- Shneiderman, B. 1996. The eyes have it: A task by data type taxonomy for information visualizations, en *IEEE Symposium on Visual Languages*, IEEE, 336–343.
- Rhyne, Theresa Marie. 1997. Going virtual with geographic information and scientific visualization, *Computers & Geosciences*, 23(4), 489 – 491.
- Valle-Lima, A. D. 2012. *La investigación pedagógica. Otra mirada*, Pueblo y Educación.
- Vázquez-Rodríguez, Romel, Carlos Pérez-Risquet, y Juan Carlos Torres-Cantero. 2013b. A novel visual data mining module for the geographical information system gvSIG, *Anuário do Instituto de Geociências*, 36(1), 98–111.
- Vázquez-Rodríguez, R., C. Pérez-Risquet, I. Y. Gonzalez-Herrera, A. Fajardo-Moya, y J. C. Torres-Cantero. 2010a. A new visual data mining tool for gvSIG gis, en *KDIR 2010 - Proceedings of the International Conference on Knowledge Discovery and Information Retrieval*, 428–431.



## Referencias III

Vázquez-Rodríguez, Romel, Carlos Pérez-Risquet, y Juan Carlos Torres. 2015. Exploratory data analysis through the integration of visualization techniques in geographical information systems, *Revista Técnica de la Facultad de Ingeniería de la Universidad del Zulia*, 38(1), 73–82.