

**Universidad Central “Marta Abreu” de Las Villas.**

**Facultad Matemática, Física y Computación**

**Licenciatura en Ciencia de la Computación**



## *Trabajo de Diploma*

*Visualización de datos multiparamétricos en Sistemas de Información*

*Geográfica*

*Autor: Alexis Fajardo Moya.*

*Tutor: MSc. Romel Vázquez Rodríguez.*

*Santa Clara*

*2010*



Hago constar que el presente trabajo fue realizado en la Universidad Central “Marta Abreu” de Las Villas como parte de la culminación de los estudios de la especialidad de Ciencia de la Computación, autorizando a que el mismo sea utilizado por la institución, para los fines que estime conveniente, tanto de forma parcial como total y que además no podrá ser presentado en eventos ni publicado sin la autorización de la Universidad.

---

Firma del autor

Los abajo firmantes, certificamos que el presente trabajo ha sido realizado según acuerdos de la dirección de nuestro centro y el mismo cumple con los requisitos que debe tener un trabajo de esta envergadura referido a la temática señalada.

---

Firma del tutor

---

Firma del jefe del  
Laboratorio

*A veces sentimos que lo que hacemos es tan solo una gota en el mar, pero el  
mar sería menos si le faltara una gota.*

*Madre Teresa de Calcuta.*

## *Dedicatoria*

*A mi mamá y a mi papá.*

*A mi hermano.*

*A Lianaday.*

*A todos los que contribuyeron en mi formación personal y profesional.*

## *Agradecimientos*

*A mi mamá y a mi papá por su amor, dedicación y ejemplo, por la educación que me brindaron y por su esfuerzo y sacrificio en pos de mi formación profesional.*

*A Lianaday por su amor y apoyo espiritual en los momentos de trabajo más intensos.*

*A mi hermano Alejandro, que me quiere como nadie.*

*A mi tutor Romel Vázquez por su valiosa ayuda y preocupación durante el desarrollo de este trabajo y por tratarme como un verdadero amigo.*

*A Inti por la certera ayuda ante cualquier interrogante.*

*A mi tía María, porque siempre me tiene presente y por quererme muchísimo.*

*A mi tía Alina por tener tan buen gusto en la selección del vestuario.*

*A Odalis y Rigoberto por el inmenso cariño que me han dado y por tratarme como un hijo.*

*A mi primo Axel por el dispositivo de almacenamiento.*

*A toda mi familia por estar siempre al tanto de lo que acontece en mi vida y por el cariño y afecto que siempre me han tenido.*

*A mis amigos Lisdany y Leduany por ser como mis hermanos.*

*A mis amigos de la carrera por el apoyo y los momentos divertidos que hemos pasado juntos.*

*A todos mis compañeros de estudio, a los que respeto y admiro.*

*A los profesores que me han aportado los más disímiles valores durante todos mis años de estudios.*

*Alexis Fajardo Moya.*

## **RESUMEN**

La integración de técnicas de Visualización Científica en Sistemas de Información Geográfica es una alternativa novedosa para el análisis visual de datos científicos. Dotar a los Sistemas de Información Geográfica con tales herramientas contribuye a la comprensión y análisis de grandes conjuntos de datos con una baja densidad espacial y permite encontrar correlaciones entre las variables en el tiempo y el espacio. En esta investigación se presenta una herramienta de minería visual de datos para el Sistema de Información Geográfica gvSIG. Esta herramienta es implementada como un módulo de gvSIG y contiene varias técnicas de visualización de datos multiparamétricos con un amplio rango de posibilidades para interactuar con los datos. En el trabajo se presentan las principales características de las disciplinas involucradas como la Visualización Científica y los Sistemas de Información Geográfica, así como su integración. Los principales aspectos relacionados con el diseño e implementación de la herramienta son explicados, quedando propuesta una metodología para la incorporación de nuevas técnicas de visualización de datos multiparamétricos al sistema desarrollado. Se desarrolla un manual de usuario que muestra el uso del sistema y se presenta un caso de estudio real con datos meteorológicos de la provincia de Villa Clara. Aunque el módulo desarrollado es probado con datos meteorológicos, las herramientas implementadas son de propósito general y permiten su utilización en múltiples campos de aplicación.

## **ABSTRACT**

The integration of Scientific Visualization techniques into Geographic Information Systems is an innovative alternative for the visual analysis of scientific data. Providing Geographic Information Systems with such tools improves the analysis and understanding of datasets with very low spatial density and allows to find correlations between variables in time and space. In this research a visual data mining tool for gvSIG Geographic Information Systems is presented. This tool is implemented as a gvSIG module and contains several Scientific Visualizations techniques for multiparameter data with a wide range of possibilities for interaction with the data. This work presents the principal features of the involved disciplines like Scientific Visualization and Geographic Information Systems, as well as their integration. The main topics related with the design and implementation are explained and a methodology for the integration of new multiparameter visualization techniques to the developed system is proposed. A user manual is developed that show the use of the system and a real case study with meteorological data from Villa Clara province is presented. Although it is tested with meteorological data, the developed module is general purpose and can be used in multiple application fields.

## ÍNDICE

<b>INTRODUCCIÓN.....</b>	<b>1</b>
--------------------------	----------

<b>CAPÍTULO 1. VISUALIZACIÓN CIENTÍFICA Y SISTEMAS DE INFORMACIÓN GEOGRÁFICA... 8</b>	
---	--

<b>1.1 Visualización Científica .....</b>	<b>8</b>
1.1.1 Técnicas de Visualización Científica .....	9
1.1.2 Técnicas de visualización para datos multiparamétricos.....	10
1.1.2.1 Técnicas geométricas .....	11
1.1.2.2 Técnicas basadas en iconos.....	14
1.1.2.3 Técnicas orientadas a píxel.....	18
<b>1.2 Sistemas de Información Geográfica.....</b>	<b>20</b>
1.2.1 GRASS.....	22
1.2.2 Quantum GIS.....	22
1.2.3 GVSIG .....	23
<b>1.3 Integración de VisCi con SIG .....</b>	<b>25</b>
1.3.1 Snap-Together Visualization .....	26
1.3.2 GeoVista Studio.....	27
<b>1.4 Conclusiones parciales .....</b>	<b>28</b>

<b>CAPÍTULO 2. DISEÑO E IMPLEMENTACIÓN DE LA EXTENSIÓN .....</b>	<b>30</b>
--	-----------

<b>2.1 Selección de las técnicas a incluir en la extensión .....</b>	<b>30</b>
<b>2.2 Diseño de la extensión realizada a gvSIG .....</b>	<b>32</b>
2.2.1 Análisis de actores y casos de uso .....	32
2.2.2 Diagramas de transición de estados .....	34
2.2.3 Diagramas de clases.....	37
2.2.3.1 Paquete para la visualización independiente.....	37
2.2.3.2 Paquete para la visualización coordinada .....	40
2.2.3.3 Jerarquía de clases para la lectura de datos.....	44
<b>2.3 Implementación de la extensión.....</b>	<b>47</b>



2.3.1	Selección de las tecnologías necesarias para la implementación .....	47
2.3.2	Consideraciones sobre las técnicas implementadas.....	49
<b>2.4</b>	<b>Metodología para la adición de nuevas técnicas al módulo de visualización de gvSIG51</b>	
2.4.1	Metodología para la adición de nuevas técnicas al paquete de visualización independiente .	51
2.4.2	Metodología para la adición de nuevas técnicas al paquete de visualización coordinada.....	53
<b>2.5</b>	<b>Conclusiones parciales .....</b>	<b>54</b>
<b>CAPÍTULO 3. PRESENTACIÓN DE LOS RESULTADOS.....</b>		<b>55</b>
<b>3.1</b>	<b>Manual de usuario .....</b>	<b>55</b>
3.1.1	Visualización de un conjunto de datos .....	56
3.1.2	Visualización coordinada de varios conjuntos de datos .....	64
3.1.2.1	Configuración de un proyecto de visualización coordinada.....	64
3.1.2.2	Utilización de un proyecto de visualización coordinada .....	65
<b>3.2</b>	<b>Caso de estudio: Análisis visual de los datos meteorológicos de la provincia Villa Clara</b>	
	.....	<b>75</b>
3.2.1	Datos meteorológicos de la provincia Villa Clara.....	76
3.2.2	Visualización de los datos meteorológicos .....	78
<b>3.3</b>	<b>Conclusiones parciales .....</b>	<b>84</b>
<b>CONCLUSIONES.....</b>		<b>85</b>
<b>RECOMENDACIONES.....</b>		<b>86</b>
<b>REFERENCIAS BIBLIOGRÁFICAS.....</b>		<b>87</b>
<b>ANEXOS .....</b>		<b>90</b>

## **INTRODUCCIÓN**

El uso de Sistemas de Información Geográfica (SIG) ha llegado a ser esencial en todos los campos relacionados con la geografía y el medio ambiente, pasando por las aplicaciones más clásicas de cartografía, urbanismo y gestión de recursos. Actualmente se utilizan los SIG para resolver problemas tan diversos como la planificación de la extinción de incendios, el análisis de riesgos ambientales o la propagación de contaminantes.

Según (Bolstad, 2005) un SIG es una integración organizada de hardware, software y datos geográficos diseñado para capturar, almacenar, manipular, analizar y desplegar en todas sus formas la información geográficamente referenciada, con el fin de resolver problemas complejos de planificación y gestión. En un sentido más genérico, los SIG son herramientas que permiten a los usuarios crear consultas interactivas, analizar la información espacial, editar datos, mapas y presentar los resultados de todas estas operaciones.

Por otra parte, la Visualización Científica (VisCi) se ha convertido en un área de investigación de interés creciente en los últimos años, motivado fundamentalmente por el incremento constante de los volúmenes de datos generados en muchos campos de aplicación, así como por el aumento sostenido de la potencia de las interfaces gráficas modernas, que permiten generar imágenes cada vez más sofisticadas.

En las últimas dos décadas se han realizado avances significativos en el establecimiento de la visualización como una herramienta de exploración de datos flexible y fácil de usar. 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.

La VisCi según (Morell and Pérez, 2006) significa encontrar una representación visual apropiada para un conjunto de datos que permita mayor efectividad en el análisis y evaluación de los mismos. Según (Rhyne, 1997) permite la transformación de los datos numéricos o simbólicos y la información en imágenes geométricas generadas por computadora. Es una metodología para interpretar, a través de una imagen en la computadora, tanto datos de mediciones como los generados por modelos computacionales. La investigación y el desarrollo de la VisCi se han centrado en cuestiones relacionadas con el renderizado de gráficos en tres dimensiones, animaciones de series temporales y visualización interactiva en tiempo real.

Una clase especial de datos son los datos multiparamétricos (datos multidimensionales o datos multivariados). Los datos multiparamétricos son los que poseen  $m$  variables o dimensiones de datos escalares distribuidos sobre puntos en el espacio de observación. Estas variables pueden ser cuantitativas o cualitativas y a su vez ordinales o nominales (Theisel, 2000, Hansen and Johnson, 2005, Pérez and Ortega, 2005). La VisCi ofrece grandes ventajas sobre otros métodos de análisis de datos cuando los mismos poseen más de dos dimensiones o variables, puesto que permite la visualización de todas las variables que se deseen al mismo tiempo. A este tipo de visualización se le conoce como visualización de datos multiparamétricos.

La integración de la VisCi y los SIG se ha ido desarrollando desde principio de la década del noventa, principalmente con incorporaciones de componentes de VisCi a SIG y viceversa. Los primeros intentos de este tipo de integración fueron desarrollados utilizando software propietario que limitaban el uso de las herramientas y el desarrollo de la unión entre estas dos disciplinas. La presente investigación se enmarca en la integración de la VisCi y los SIG aprovechando las ventajas y fortalezas de ambos enfoques para el análisis de datos espacio-temporales. Las soluciones están dadas en el área del software libre facilitando la reutilización de las aplicaciones desarrolladas. El acceso a los algoritmos y sus implementaciones les permitirá a los usuarios entender qué están haciendo al ejecutar un análisis particular y promoverá investigaciones reproducibles mediante el acceso a herramientas de código abierto.

El mundo real no es estático, tanto componentes espaciales como no espaciales tienen que ver con el tiempo. Especialistas de muchas áreas de la ciencia visualizan, consultan y analizan la información recopilada a partir del mundo real para ayudar a la toma de decisiones. Esta información existe en el dominio espacial, el dominio temporal y en el dominio temático (Hogeweg, 2000). El dominio espacial es el encargado de analizar qué se mide o se encuentra en algún lugar. El dominio temporal se encarga de analizar qué se produce en algún momento o qué existe durante un cierto tiempo. El dominio temático es el que tiene que ver con el área de aplicación que se va a analizar en el espacio y el tiempo. Una de las ciencias que mejor se ajusta a la integración de los dominios espacial, temporal y temático es la meteorología. El análisis espacio-temporal en esta área se ha realizado históricamente mediante el uso de SIG, herramientas para el análisis de series temporales, paquetes estadísticos y geoestadísticos y la animación en sistemas de visualización.

Como parte de un proyecto territorial entre el laboratorio de Computación Gráfica del Centro de Estudios de Informática de la Universidad Central “Marta Abreu” de Las Villas y el Instituto de Meteorología de la Provincia de Villa Clara, se cuenta con los datos meteorológicos de un reducido número de estaciones de esta provincia.

### **Planteamiento del problema**

Los datos mencionados no pueden ser analizados suficientemente bien en forma numérica, por lo que surge la necesidad de visualizarlos con el objetivo de realizar comparaciones, identificar patrones, encontrar correlaciones, detectar anomalías, variabilidad y las diferentes tendencias que se pueden presentar en las variables meteorológicas a lo largo del tiempo, teniendo en cuenta su ubicación espacial.

A pesar de contar con una serie temporal bastante completa de las variables meteorológicas la información espacial es muy pobre, debido a que se cuenta con los datos de sólo cuatro estaciones.

Existen múltiples áreas de aplicación en las que está presente esta problemática de la baja densidad espacial. Debido a las deficiencias de los métodos de interpolación que emplean los SIG actuales cuando los datos están poco muestreados espacialmente y las ventajas que brindan las técnicas de VisCi para datos multiparamétricos, se desarrolló la presente investigación.

Dada la cantidad de técnicas de VisCi existentes y las particularidades de los SIG, se formularon varias **Preguntas de Investigación**:

- ¿Cuáles son las técnicas de VisCi para datos multiparamétricos más adecuadas en el análisis visual de datos espacio-temporales?
- ¿Cuál de los SIG de software libre es el más adecuado para la incorporación de un módulo de VisCi?
- ¿Cuáles resultados se obtendrían al combinar técnicas de visualización de datos multiparamétricos con SIG para el análisis visual de datos meteorológicos poco muestreados espacialmente?

Teniendo en cuenta la problemática planteada y las preguntas de investigación se plantearon los siguientes objetivos:

### **Objetivo general**

Implementar una aplicación que mediante un enfoque de integración entre la VisCi y los SIG permita visualizar en un SIG datos multiparamétricos poco muestreados espacialmente y amplios en el tiempo, como es el caso de los datos meteorológicos de la provincia de Villa Clara.

### **Objetivos específicos**

- Seleccionar las técnicas de visualización más adecuadas para el análisis visual de datos multiparamétricos.

- Seleccionar un SIG adecuado para la incorporación de técnicas de VisCi de datos multiparamétricos.
- Implementar las técnicas de VisCi de datos multiparamétricos en el SIG seleccionado.
- Desarrollar una metodología para la incorporación de nuevas técnicas de VisCi al módulo desarrollado.
- Ilustrar mediante algunos ejemplos la idoneidad del uso de las técnicas desarrolladas para el análisis visual de datos meteorológicos poco muestreados espacialmente.

### **Justificación de la investigación**

Existen numerosas áreas de aplicación como la Geofísica, Geología, Medio Ambiente y la Meteorología que están generando datos producto de estaciones de medición. Muchos de los equipos necesarios para realizar estas mediciones son de tecnología muy costosa y difícil de obtener por países subdesarrollados, lo que provoca que la distribución espacial de este tipo de equipos sea pobre. Los países subdesarrollados generalmente cuentan con una infraestructura meteorológica pobre, es decir, pocas estaciones meteorológicas que se distribuyen demasiado espaciadas. Esta diferencia de espacio, junto a la poca cantidad de estaciones impide que se puedan crear mediante interpolación espacial mapas realistas de las variables registradas.

Sin embargo la información que contiene cada estación es amplia en el tiempo y la constituyen varias variables. El espacio existente entre las diferentes estaciones se puede utilizar para mostrar gráficos que contengan toda o parte de la información temporal de varias variables a la vez. La adición de un módulo de VisCi a un SIG aportaría valiosas herramientas a los especialistas, permitiéndoles visualizar los datos de cada estación mediante gráficos generados por las técnicas. La representación visual de los datos permitiría a los usuarios conocer con facilidad características de los mismos, que podrían ser desde informaciones estadísticas básicas hasta conclusiones importantes.

Las técnicas que por sus características no puedan ser visualizadas en lugares puntuales podrán ser mostradas de manera independiente, siempre y cuando se sepa a qué localización corresponden.

Estas ideas podrán ser aplicadas en cualquier tipo de problemas que contengan información multivariada de datos puntuales en el espacio.

### **Viabilidad de la investigación**

El estado actual de las técnicas de VisCi y el desarrollo de los SIG ofrece una amplia gama de ideas a desarrollar en este trabajo. Para el desarrollo de esta investigación se cuenta con los recursos necesarios para acometer las tareas propuestas, que incluyen los datos meteorológicos, los códigos fuentes de los SIG, la información cartográfica y la capacidad de procesamiento de datos y de gráficos. Por otro lado se cuenta con el apoyo del grupo de investigación Computación Gráfica del Centro de Estudios de Informática, que tiene suficiente experiencia en el área.

Después de haber elaborado el marco teórico se formuló la **Hipótesis general** de investigación siguiente:

La integración de técnicas de VisCi con SIG en una misma aplicación permite visualizar una gran cantidad de información multivariada y amplia en el tiempo e interactuar con la misma a través de las diferentes técnicas, asociándolas a lugares puntuales en el espacio. Este método de análisis espacio-temporal facilita la interpretación y comprensión de los datos multiparamétricos asociados a pocos lugares puntuales del espacio, como es el caso de los datos meteorológicos de la provincia de Villa Clara.

El trabajo se estructura en tres capítulos. En el primero de ellos se abordan los aspectos teóricos relacionados con la VisCi, se describen los tipos de datos que utilizan las técnicas de visualización y en base a ese criterio se efectúa una clasificación de las mismas, abordándose con mayor profundidad las técnicas de visualización de datos

multiparamétricos. Se exponen las características generales de los SIG y se profundiza en algunos SIG del mundo del software libre. Además se analizan las variantes de integración entre la VisCi y los SIG mostrando ejemplos de aplicaciones que implementan el enfoque mezclado de integración. En el capítulo dos se exponen los aspectos relacionados con el diseño y la implementación de un módulo de visualización de datos multiparamétricos incorporado al SIG gvSIG. Se mencionan algunos aspectos sobre la implementación de las técnicas, el modo de integración con gvSIG y la metodología para extender el módulo. En el capítulo tres se muestra el manual de usuario de la extensión y un caso de estudio sobre el análisis visual de los datos meteorológicos de la provincia Villa Clara. Finalmente se formulan las conclusiones y recomendaciones, se relaciona la bibliografía y se añaden algunos anexos.



## **CAPÍTULO 1. VISUALIZACIÓN CIENTÍFICA Y SISTEMAS DE INFORMACIÓN GEOGRÁFICA**

En el presente capítulo se tratan las generalidades de la VisCi y específicamente la visualización de datos multiparamétricos. Se abordan las principales características de los SIG así como una descripción de los más utilizados en el mundo del software libre. Además se brindan elementos que evidencian la integración de la VisCi con los SIG.

### **1.1 Visualización Científica**

La ciencia ha desarrollado diversos métodos para la obtención de información, y uno de ellos se basa en la creación de imágenes a partir de los datos. Este método, conocido como visualización, ha sido utilizado como vía natural para mostrar información (Hansen and Johnson, 2005). Recientes investigaciones han impulsado en gran medida este campo mediante el uso de la computación, motivado fundamentalmente por el incremento constante de los volúmenes de datos generados en muchos campos de aplicación, así como por el aumento sostenido de la potencia de las interfaces gráficas modernas, que permiten generar imágenes cada vez más sofisticadas.

La visualización de datos permite alcanzar diferentes metas. La naturaleza del objetivo que se desee está en relación directa al conocimiento que se tenga sobre los datos iniciales. Los objetivos pueden ser los siguientes (Theisel, 2000):

- Análisis exploratorio.
- Análisis confirmativo.
- Presentación de información.

El *Análisis exploratorio* en la visualización parte de un conjunto de datos sobre el que no se tiene hipótesis y a partir de un proceso interactivo, usando una búsqueda no dirigida, se llega a la obtención de una imagen que aporta una hipótesis sobre los datos.

El *Análisis confirmativo* comienza con datos sobre los que se tiene a priori una hipótesis; el proceso consiste en la búsqueda de la confirmación de la hipótesis.

La *Presentación de información* parte de hechos que son fijos a priori y que se desean enfatizar y mostrar con extrema calidad.

El análisis visual de datos es un nuevo enfoque que se puede utilizar con cualquiera de los objetivos anteriores. Como se ha mencionado anteriormente, 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.

### **1.1.1 Técnicas de Visualización Científica**

Diversos enfoques se han empleado para agrupar y clasificar las diversas técnicas de VisCi existentes. Un enfoque establecido para clasificar las técnicas es a través del tipo de dato sobre los que operan. Por el tipo de dato se refiere al tipo al que pertenecen los atributos o variables. Atendiendo a este criterio se encuentran las siguientes categorías (Hansen and Johnson, 2005, Theisel, 2000):

- Técnicas de visualización para datos volumétricos.
- Técnicas de visualización para fluidos.
- Técnicas de visualización para datos multiparamétricos.
- Técnicas de visualización de la información.

Existen diversos enfoques para especificar los datos (Pérez and Ortega, 2005). Estos enfoques permiten definir una serie de características de los datos como son la

dimensionalidad, la estructura y el nivel de medición. En este trabajo se utiliza un enfoque sencillo para definir los datos.

Los datos volumétricos representan una malla de tres dimensiones donde cada punto tiene asociado un valor. En general los datos se definen como un conjunto  $S$  de muestras, donde cada elemento  $s \in S$  es un vector de la forma  $(x, y, z, v)$  que contiene las coordenadas espaciales y un elemento que es un escalar (Theisel, 2000, Hansen and Johnson, 2005).

Los campos vectoriales representan una malla de dimensión menor o igual que tres donde cada punto está relacionado con un vector. Una de las áreas de mayor uso de los campos vectoriales es para representar datos de fluidos (Hansen and Johnson, 2005).

Los datos multiparamétricos son aquellos en que el número de variables relacionadas con cada observación es mayor o igual que dos. Estas variables pueden ser cuantitativas o cualitativas y a su vez ordinales o nominales (Hansen and Johnson, 2005, Pérez and Ortega, 2005).

En algunas aplicaciones los datos presentan una estructura que no concuerda con ninguna de las anteriores o que sencillamente no puede ser definida con exactitud. A estos datos se les suele llamar información y entre las principales se identifican estructuras como árboles, grafos e hipertexto (Hansen and Johnson, 2005, Theisel, 2000, Keim, 2002).

Este trabajo se centró en el desarrollo de técnicas de visualización de datos multiparamétricos. Es por ello que se hace mayor énfasis en este tipo de técnicas, las cuales son descritas a continuación.

### **1.1.2 Técnicas de visualización para datos multiparamétricos**

Existen una serie de problemas en que cada punto de dato contiene más de un atributo, estos atributos pueden ser fechas, precios o valores descriptivos. A este tipo de datos se les

llama multiparamétricos y se encuentran generalmente en aplicaciones de minería de datos, estadísticas e inteligencia artificial (Keim, 2002). Los datos multiparamétricos, también llamados multidimensionales o datos n-dimensionales, consisten en un número de n registros donde cada uno está definido por un vector de d valores. 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 (Ward, 2008).

El objetivo fundamental de los métodos de visualización para datos multiparamétricos es lograr que las representaciones revelen correlaciones o patrones entre los atributos (Theisel, 2000, Keim, 2002, Eick, 2000). Con este fin existe actualmente una amplia gama de técnicas de visualización, para las cuales se han creado además diversas mejoras. Las técnicas pueden ser clasificadas en geométricas, basadas en iconos, basadas en píxel y proyecciones (Keim, 2002, Theisel, 2000).

#### **1.1.2.1 Técnicas geométricas**

Las técnicas geométricas son aquellas que utilizan elementos como puntos, líneas o curvas como propiedades visuales para representar los datos (Keim, 2002, Theisel, 2000). Existe un gran número de ellas, entre las que se encuentran *prosections views* (Furnas and Buja, 1994), *hyper slices* (van Wijk and van Liere, 1993) y *parahistogramas* (Ong and Lee, 1996), pero hay tres que sobresalen por su generalidad y gran uso, estas son los Diagramas de Dispersión, Coordenadas Paralelas y Gráfico de Andrews .

#### **Diagramas de Dispersión**

El Diagrama de Dispersión es una técnica sencilla muy utilizada. Su forma más simple se manifiesta cuando los datos poseen solo dos dimensiones. Con dos dimensiones la técnica consiste en trazar dos ejes de coordenadas y utilizar los valores de las dimensiones como puntos (x, y) de  $R^2$ , resultando un gráfico donde se encuentran dispersos los puntos de datos. Para visualizar datos de más de dos dimensiones pueden utilizarse proyecciones, que

provocan pérdida de información debido a la reducción de la dimensión (Hansen and Johnson, 2005, Theisel, 2000, Keim, 2002).

Para datos multiparamétricos es muy frecuente utilizar matrices de diagramas de dispersión. Las matrices resultantes son cuadradas y el elemento (i, j) de la matriz es un diagrama de dispersión de la dimensión i y la j. El diseño evita la pérdida de información pero en cambio los análisis complejos son engorrosos. Una deficiencia adicional es que la diagonal principal de la matriz es subutilizada. Algunos trabajos actuales están encaminados a aprovechar mejor esta región de la representación (Cui et al., 2006). Ver Figura 1.

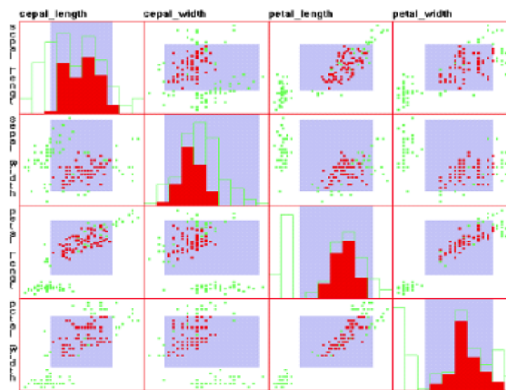


Figura 1: Matriz de diagramas de dispersión con la diagonal principal con histogramas.

## Coordenadas Paralelas

La técnica Coordenadas Paralelas es un esquema simple de gran generalidad que permite visualizar conjuntos de datos multidimensionales. Esta técnica geométrica es una de las más utilizadas producto de lo fácil que resulta su implementación y los buenos resultados que se obtienen al aplicarla (Keim, 2002). Ver Figura 2.

Esta técnica utiliza un sistema de coordenadas como base y consiste en crear un eje de coordenadas para cada atributo colocándolos paralelamente, uno a continuación del otro. El valor de cada dimensión en un determinado punto de datos es marcado en el eje

correspondiente. La representación final para un objeto es una línea que recorre las posiciones marcadas en cada dimensión (Theisel, 2000, Keim, 2002). Resulta especialmente útil para mostrar patrones en los datos o para percibir relaciones entre los atributos, pues el resultado es muy intuitivo (Theisel, 2000).

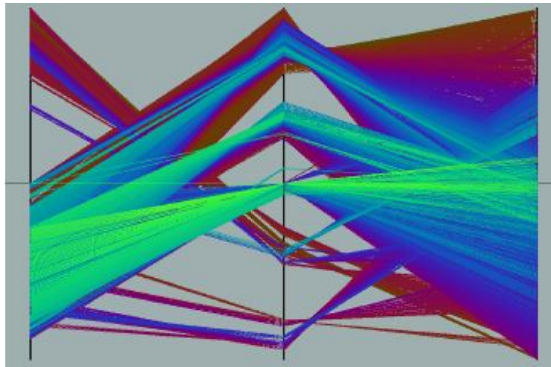
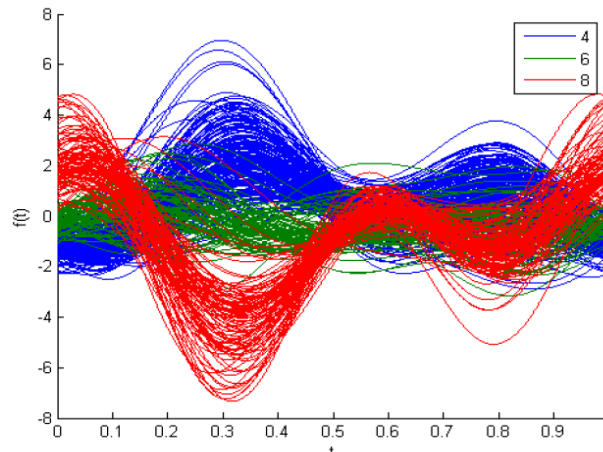


Figura 2: Coordenadas Paralelas. Se muestra la visualización de un conjunto de datos con tres variables.

El color de las líneas que representan los objetos puede ser elegido por varios criterios. El más simple es utilizar un color constante para todos los objetos. Un criterio que maximiza la calidad de la imagen es asignar el color de los objetos de acuerdo a una dimensión (Theisel, 2000, Keim, 2002).

### Gráfico de Andrews

Una idea similar para representar datos multiparamétricos es el Gráfico de Andrews (Obsérvese la Figura 3). En esta técnica cada observación es representada por una función  $f(t)$  que se evalúa en el intervalo  $[0,1]$ . Cada función es una serie de Fourier cuyos coeficientes se igualan a los valores de las dimensiones para cada observación (Matlab, 2004, Andrews, 1972).



**Figura 3: Gráfico de Andrews.**

Esta técnica permite identificar con facilidad diferencias entre grupos de observaciones, ya que por lo general observaciones pertenecientes a un mismo grupo presentan una forma de la función similar. Los análisis sobre variables individuales resultan en cambio mucho más engorrosos (Matlab, 2004). La virtud fundamental de la técnica es que puede representar conjuntos de datos de un tamaño relativamente grande y además con un número de dimensiones elevado.

### ***1.1.2.2 Técnicas basadas en iconos***

Las técnicas basadas en iconos tienen dos parámetros que la caracterizan. El primero es el tipo de figura que representará cada observación, o sea, la forma del icono; el segundo parámetro es la forma en que se definirá la posición de cada icono en la imagen (Theisel, 2000, Ward, 2002).

Estas técnicas no sufren de pérdida de información. Se logra evitar la pérdida de información al realizar una proyección de las dimensiones a los diferentes rasgos del icono (Theisel, 2000).

Entre los métodos para crear iconos están los rostros de Chernoff (Chernoff, 1973) y los campos de estrellas. Además suelen crearse editores de iconos para aplicaciones específicas (Theisel, 2000, Salgado, 2003). Por otro lado la solución más popular para la ubicación de los iconos en la imagen está basada en el uso de proyecciones (Ward, 2002).

Las técnicas basadas en iconos son recomendadas cuando el número de dimensiones oscila entre diez y quince y el número de mediciones de las mismas es alto. Estas técnicas se pueden utilizar con una referencia espacial.

### **Campo de Estrellas (en inglés *Starfield*)**

La técnica de campo de estrellas utiliza un método procedural para representar cada icono y permite elegir entre varias opciones para colocar los mismos. Tiene el potencial para mostrar datos de múltiples dimensiones sin pérdida de información y es habitualmente usada por los especialistas (Theisel, 2000, Andrews, 2005).

El campo de estrellas utiliza un algoritmo para componer los iconos, lo que le confiere cierta generalidad. En la forma básica el método utiliza dos dimensiones como coordenadas de posición en un eje imaginario. El resto de las dimensiones son normalizadas al intervalo  $[0,1]$ . Estas coordenadas son el punto de inicio en el dibujado del icono. Las dimensiones restantes se expresan a partir de líneas que parten del punto inicial y cuya longitud está determinada por el valor del atributo. Estas líneas o rayos que representan las diferentes dimensiones están dispuestos entre sí con igual distancia angular, lo que genera una figura de estrella. Frecuentemente los extremos de las líneas son conectados entre sí. Esta variación elimina la silueta de estrella y crea una figura cerrada que suele presentar más claramente las características del objeto (Eick, 2000). En la Figura 4 puede observarse un ejemplo.



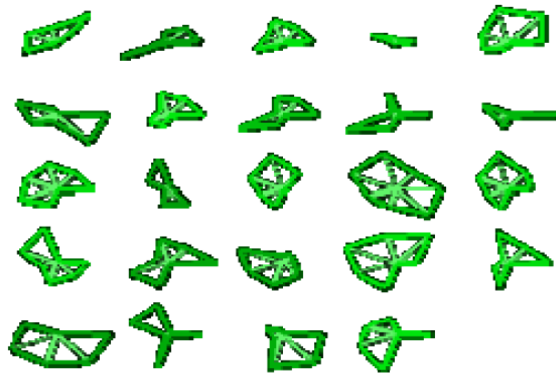


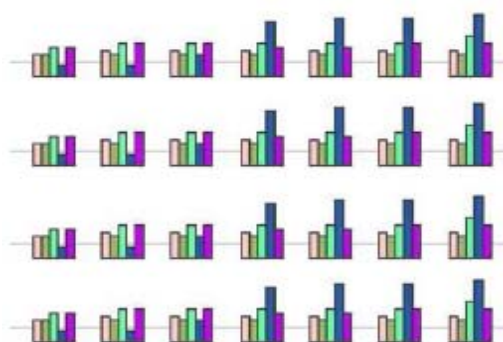
Figura 4: Campo de Estrellas.

Todos los puntos del conjunto de datos pueden mostrarse con el mismo color, pero resulta muy conveniente utilizar esta característica para codificar algún atributo de interés. Igualmente pueden usarse otros rasgos de la figura para codificar otras informaciones, como por ejemplo la calidad de los datos (Xie et al., 2006).

Una cuestión de particular importancia en la técnica es la estrategia de posicionamiento del icono. En el proceso de posicionamiento de los iconos pueden usarse los datos de ciertas dimensiones, que en el caso más simple utiliza dos dimensiones y en caso de un número mayor de dimensiones requiere el uso de proyecciones (Ward, 2002).

### **Icono en forma de barras (en inglés *Profile Glyph*)**

Las características de este icono están dadas por la longitud y el color de las barras que lo forman. Las dimensiones son representadas por barras ubicadas secuencialmente cuyas longitudes están determinadas por el valor de los atributos. El color de cada barra se selecciona de acuerdo a la dimensión visualizada, de manera tal que permita identificar qué dimensión está codificada en los atributos del icono (Toit et al., 1986). Ver la Figura 5.



**Figura 5:** Icono en forma de barras. Se muestra la visualización de 28 observaciones de cuatro variables.

Este icono permite reconocer con facilidad las relaciones existentes entre dimensiones adyacentes. La comparación de las longitudes de las barras con una línea base común es más fácil de percibir que las longitudes de los rayos de un icono en forma de estrella (Ward, 2008).

Las relaciones entre dimensiones que están más separadas son difíciles de determinar. Este problema se puede resolver reordenando las dimensiones, tratando de que queden adyacentes aquellas que se deseen correlacionar o reordenarlas de manera aleatoria hasta obtener la visualización deseada. Existen estrategias de reordenamiento de las dimensiones que aumentan las posibilidades de obtener información más valiosa que utilizando reordenamientos aleatorios (Ward, 2008).

### Shape Coding

Este icono es un rectángulo que contiene una malla regular. Cada celda de la malla es usada para una dimensión cuyo color depende del valor de la dimensión en el punto de dato visualizado. Obsérvese la Figura 6. Los iconos rectangulares usualmente son localizados de forma linear, aunque pueden ser considerados otros criterios de ordenamiento como pueden ser: correlaciones espaciales o temporales o un orden de acuerdo a valores especiales de los datos (Beddow, 1990).

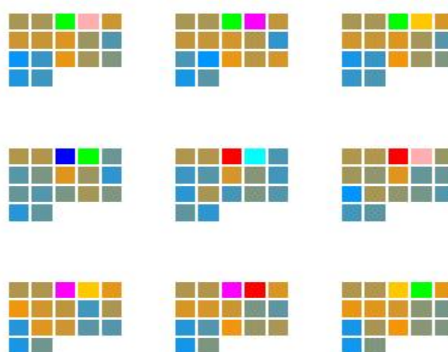


Figura 6: Shape Coding. Se muestra la visualización de nueve observaciones de diecisiete variables.

### 1.1.2.3 Técnicas orientadas a píxel

La visualización de un conjunto de datos de gran tamaño resulta un reto para técnicas geométricas y basadas en iconos. Al graficarlos suele surgir desorden en la imagen, que está originado por el tamaño de la figura que representa una observación simple. Partiendo de esta idea resulta lógico concluir que minimizando el espacio que ocupa un solo punto de dato en la imagen se mejoraría la percepción visual (Hansen and Johnson, 2005, Andrews, 2005).

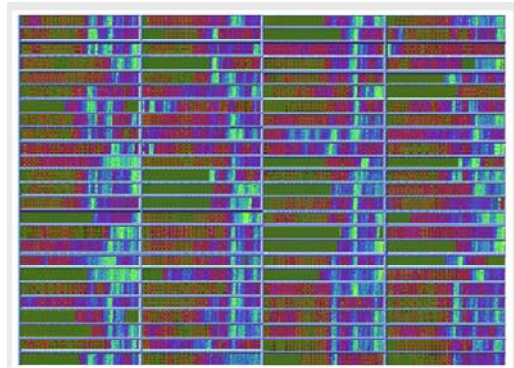
Las técnicas basadas en píxel son las más eficientes cuando el número de dimensiones es grande y cuando crece el número de registros. Esto se debe a que utilizan un píxel para representar cada atributo de una observación. Los retos fundamentales en estos métodos son la elección del color para cada elemento y el modo de posicionamiento de los píxeles (Keim, 2002, Keim, 2000).

El procedimiento en las técnicas basadas en píxel consiste en relacionar cada valor de una dimensión a un color y agrupar los píxeles de cada dimensión en áreas adyacentes. Puesto que este método utiliza un píxel simple por cada valor de dato, la técnica permite mostrar hasta más de un millón de valores (Keim, 2002).

Este tipo de técnicas utilizan diferentes modos de posicionamiento de los píxeles para lograr diferentes objetivos. Colocar los píxeles en la forma apropiada ofrece la posibilidad de observar información sobre correlaciones, dependencias y regiones trascendentales. Dos de los modos de posicionamiento de los píxeles son los patrones recursivos y los segmentos de círculo (Hansen and Johnson, 2005).

### **Patrones Recursivos**

La técnica orientada a píxel patrones recursivos tiene una manera especial de interactuar con los datos, permitiendo la definición de diferentes niveles de recursividad. Está particularmente dirigida a representar un conjunto de datos con un orden natural de acuerdo a un atributo, propiedad que la convierte en una opción para problemas de series de tiempo. Ver la Figura 7.



**Figura 7: Patrones Recursivos.**

Una posibilidad simple que provee la técnica es la de organizar los puntos de datos de izquierda a derecha, línea por línea o columna por columna. Una vía posible de mejorar la visualización es la organización de los píxeles en pequeños grupos y organizar los grupos para formar un patrón global. Esta estrategia corresponde a un planteamiento en dos fases con un patrón de primer orden formado por la agrupación de los píxeles y un patrón de segundo orden formado por el orden global. Al tomar los resultados de la estructura de segundo orden como el elemento básico de construcción de una estructura de tercer nivel, puede realizarse la introducción de un tercer patrón. Este proceso puede ser repetido hasta

un nivel arbitrario formando un esquema general recursivo. Este esquema general recursivo puede ser distribuido de dos formas para cada nivel de recursividad: línea a línea (*line by line*), donde las estructuras de cada orden se posicionan en la imagen de izquierda a derecha o intercalando el sentido (*back and forth*), las estructuras de cada orden se posicionan intercalando el sentido, es decir para una línea se posicionan de izquierda a derecha y en la siguiente línea de derecha a izquierda (Keim et al., 1995).

### Segmentos de Círculo

Los segmentos de círculo utilizan como imagen base un círculo que es dividido en segmentos iguales a partir del origen. Cada segmento corresponde a un atributo del conjunto de datos. Dentro de cada segmento el valor del atributo para cada registro de datos se representa con un píxel simple. La colocación de los píxeles comienza en el centro de la circunferencia y continúa hacia fuera dibujando sobre una línea ortogonal al segmento (Ankerst et al., 1996). Obsérvese la Figura 8.

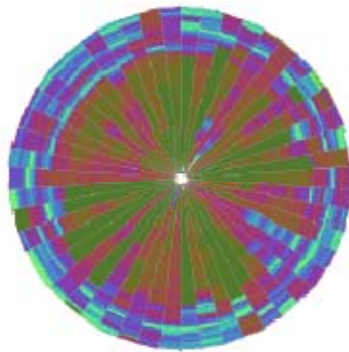


Figura 8: Segmentos de Círculo.

## 1.2 Sistemas de Información Geográfica

El acelerado desarrollo de los SIG, provocado en parte por la revolución de las nuevas tecnologías, los ha convertido en una herramienta de trabajo esencial para el análisis y resolución de diversos problemas que se presentan en empresas, industrias e instituciones

sociales y gubernamentales. Su versatilidad ha permitido que puedan ser empleados en casi todas las actividades que poseen una componente espacial, convirtiéndose en una herramienta esencial en muchas áreas relacionadas con la gestión estratégica, incluyendo también los procesos de análisis demográfico, protección del medio ambiente, y aplicaciones de urbanismo y gestión de recursos. En general, encontramos aplicaciones de los SIG en tareas como:

- *Localización*: Preguntar por las características de un lugar concreto.
- *Condición*: El cumplimiento o no de una de las condiciones impuestas al sistema.
- *Tendencia*: Comparación entre situaciones temporales o espaciales distintas de alguna característica.
- *Rutas*: Cálculo de rutas óptimas entre dos o más puntos.
- *Pautas*: Detección de pautas espaciales.
- *Modelos*: Generación de modelos a partir de fenómenos o actuaciones simuladas.

Un SIG se define a menudo como una combinación de un sistema de gestión de base de datos, un conjunto de operaciones para examinar los datos y un sistema de visualización gráfica que están vinculados al proceso de análisis espacial. Los SIG son herramientas cartográficas que permiten la creación de mapas en tiempo real y el análisis de los impactos de los cambios en el mapa de forma interactiva (Rhyne, 1997). Otras definiciones hacen énfasis en los SIG como disciplina o ciencia aplicada que incluyen en su formulación no solo al software, sino al hardware, al equipo técnico y la filosofía de trabajo, integrándolo todo de forma global.

Según (Great Britain. Committee of Enquiry into the Handling of Geographic Information, 1987) es un sistema para capturar, almacenar, comprobar, manipular y visualizar datos que estén espacialmente referenciados a la tierra. Según (Clarke, 1990) es un sistema automatizado para la captura, almacenamiento, composición, análisis y visualización de datos espaciales. Otra definición es en la que se considera como un sistema de hardware, software y procesamiento diseñado para la captura, gestión, manipulación, análisis,

modelado y visualización de datos espacialmente referenciados para resolver problemas complejos de planeamiento y gestión (Cowen, 1989).

Son muchas las ventajas que se pueden obtener de los SIG por las múltiples aplicaciones en que se pueden emplear. A continuación se abordan un conjunto de SIG del mundo del software libre, que fueron analizados con vistas a la posible incorporación de un módulo de visualización de datos multiparamétricos.

### **1.2.1 GRASS**

GRASS (de las siglas en inglés Geographic Resources Analysis Support System) fue inicialmente concebido y desarrollado en 1982 por el laboratorio de investigación del cuerpo de ingenieros del ejército de los Estados Unidos (USA-CERL) para la gestión del territorio y la gestión medioambiental (GRASS, 2008).

GRASS dispone de un gran número de herramientas y utilidades. Originalmente estuvo más orientado al aspecto matricial (raster) de la información, aunque contaba con un potente editor de topología vectorial. Sin embargo en las últimas versiones el aspecto vectorial, y sobre todo la conexión a bases de datos externas, se han potenciado. En cuanto a la interfaz de usuario también ha experimentado una gran evolución, teniendo en cuenta que en las primeras versiones todo el control se hacía por medio de comandos tipo UNIX. Otro gran avance ha sido la herramienta de visualización 3D (NVIZ), que se destaca por su potencia gráfica y las opciones de generación de salidas gráficas que permite.

### **1.2.2 Quantum GIS**

Quantum GIS es un SIG programado en C++. Posee una interfaz visual que facilita una rápida familiarización a los nuevos usuarios.

Tiene soporte directo para edición en PostGIS<sup>1</sup> y conexión con GRASS para tareas como edición de topología. Soporta formatos tanto vectoriales como raster (QGIS, 2009).

Quantum GIS está desarrollado a partir de una filosofía modular. A partir de la versión 0.9.x permite la adición de extensiones. Actualmente se pueden encontrar un buen número de extensiones para tareas como la conversión de archivos *shape* de *ESRI* (Environmental Systems Research Institute) a PostGIS. Permite la creación de extensiones en los lenguajes Python y C++.

Quantum GIS por sí solo no contiene técnicas de visualización e interpolación adecuadas para el tratamiento de datos; no obstante estas funcionalidades se pueden incorporar haciendo uso del módulo de conexión con GRASS y utilizar los módulos de interpolación que este tiene implementados.

### 1.2.3 GVSIG

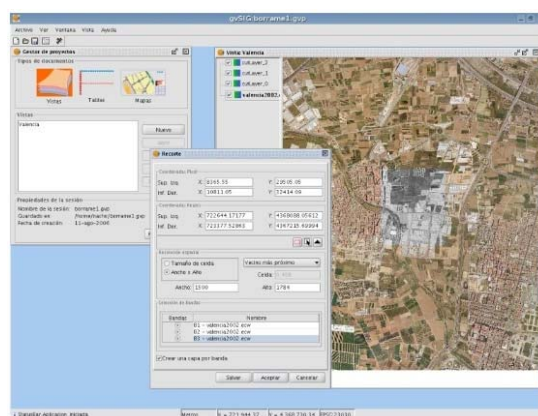
GvSIG (Generalitat Valenciana SIG) surge como un proyecto amparado por la Generalitat Valenciana de España que a finales de 2003 promocionó un concurso para el desarrollo de un SIG con una serie de características propias como: multiplataforma, de código abierto, modular, sin licencias, interoperable con formatos de otros programas (Autocad, Microstation, Arcview), y sujeto a estándares de la OGC (Open Geospatial Consortium) (Anguix and Carrión, 2005). El resultado ha sido una aplicación que ya tiene disponibles varias versiones al público y gran parte de las funcionalidades propias de los SIG cubiertas, aunque se desarrolla constantemente. Las funciones básicas que cualquier usuario desearía como diseño de impresión o soporte de formatos de imagen típicos están incorporadas sin necesidad de ningún módulo adicional.

---

<sup>1</sup> **PostGIS** es un módulo que añade soporte de objetos geográficos a la base de datos objeto-relacional PostgreSQL, convirtiéndola en una base de datos espacial para su utilización en Sistema de Información Geográfica. Se publica bajo la licencia pública general de GNU.



GvSIG posee una jerarquía de clases bien estructurada para la incorporación de nuevas funcionalidades. Permite la lectura de varios formatos de datos geográficos y no geográficos en forma de tablas así como la conexión con varias bases de datos. En la Figura 9 se muestra la interfaz visual de gvSIG.



**Figura 9: Interfaz visual de gvSIG.**

Este SIG posee las aplicaciones traducidas a veinte idiomas; toda la documentación está disponible en 5 idiomas, incluyendo español e inglés, por lo que se ha convertido en un SIG muy popular en el mundo hispano (Anguix, 2009). Se ha reportado su utilización en varios países europeos como Francia, Italia, Suiza, Austria, Reino Unido y Alemania, donde se encuentra la mayor comunidad de usuarios de gvSIG no hispano hablantes. Varias instituciones y universidades prestigiosas han utilizado esta aplicación, tal es el caso de la Agencia Espacial Europea y Oxford Archaeology. Varios países africanos han realizado trabajos con gvSIG pero su mayor uso se ha reportado en Iberoamérica.

La simplicidad para la incorporación de nuevas funcionalidades, la disponibilidad de toda su documentación, el soporte técnico suministrado por la lista de distribución de sus desarrolladores y el apoyo institucional del proyecto hace que se considere el SIG más adecuado para la incorporación de un módulo de visualización científica de datos multiparamétricos.

### 1.3 Integración de VisCi con SIG

A finales de los años ochenta y principio de los noventa la VisCi y los SIG se desarrollaron en paralelo y de forma independiente (Rhyne, 1997). Los esfuerzos para desarrollar estándares de datos espaciales rara vez consideraron la forma en que estos se visualizaban. Las bibliotecas gráficas y los estándares evolucionaron independientemente de los modelos de datos. Como resultado de esto se evidenciaron muchas ineficiencias asociadas con la visualización de datos geográficos. Entre ellas se incluyen dificultades con el registro de los datos espaciales dentro de Sistemas de Visualización Científica, engorrosas producciones de secuencias de animaciones en SIG y quizás la más importante, la falta de conexión entre bases de datos y los ambientes de visualización que soportaban la visualización de datos espaciales (Hearnshaw and Unwin, 1994).

Los desarrolladores de herramientas SIG y de VisCi hicieron esfuerzos para ampliar e integrar sus sistemas (Rhyne et al., 1994). Los desarrolladores de SIG estudiaron la forma de incorporar las capacidades de la animación de series de tiempo en tres dimensiones en su software. Los desarrolladores de herramientas de VisCi comenzaron la construcción de lectores de datos que soportaban los formatos de datos espaciales como modelos digitales de elevación, así como formatos de SIG comerciales.

Al examinar estos esfuerzos fueron definidos cuatro niveles de métodos de integración entre SIG y VisCi: *rudimentario*, *operacional*, *funcional* y *mezclado* (Rhyne, 1997).

El enfoque *rudimentario* utiliza una mínima integración de datos e intercambio entre las dos tecnologías. El nivel *operacional* proporciona coherencia entre los datos mientras se eliminan las redundancias entre las dos tecnologías (Cook et al., 1997). La forma *funcional* intenta proporcionar una comunicación transparente entre los entornos de software correspondientes (Mitas et al., 1997). El enfoque *mezclado* se refiere al desarrollo de sistemas donde los conceptos de cartografía, SIG y VisCi se funden en una única herramienta.

El nivel rudimentario de intercambiar datos en formatos de SIG hacia las herramientas de VisCi se ha logrado en casos puntuales. Algunos ambientes de VisCi se han aproximado en el nivel operacional permitiendo accesos directos a bases de datos de SIG; sin embargo esto suele ser en un solo sentido, una vez que la herramienta de VisCi genera la imagen tridimensional o la animación, generalmente no es posible activar las funciones de consulta de los SIG desde la pantalla de visualización.

Lograr la integración funcional de SIG y herramientas de VisCi requiere de estándares abiertos de datos de SIG, enlaces a programas que permitan que las herramientas de VisCi realicen análisis de datos espaciales y funciones de extracción de información. También es factible el uso de sistemas expertos o una arquitectura basada en reglas con agentes inteligentes para facilitar la comunicación transparente entre SIG y herramientas de VisCi (Treinish and Rogowitz, 1993).

Los sistemas que implementan el enfoque mezclado replantean el proceso de desarrollo de herramientas SIG-VisCi. En este sentido la cartografía está bien posicionada como puente entre ambas tecnologías.

Algunos de los primeros intentos del enfoque mezclado entre los SIG y la VisCi se han materializado a través de herramientas como GeoVista Studio y Snap-Together Visualization. Ambas herramientas están basadas en redes de componentes que permiten la construcción de complejos flujos de trabajos que vinculan componentes para el manejo de formatos de datos geográficos y de diferentes técnicas de VisCi.

### **1.3.1 Snap-Together Visualization**

*Snap-Together Visualization* es una herramienta Web que permite que los datos de los usuarios sean mezclados y correlacionados de forma dinámica en visualizaciones coordinadas para la construcción personalizada de interfaces de exploración sin necesidad de programación. El modelo conceptual de *Snap-Together* se basa en un modelo de base de

datos relacional. Este modelo permite que las relaciones sean cargadas durante la visualización y se coordinen basándose en las características que las unen. Los usuarios pueden crear diferentes tipos de coordinaciones tales como: barridos, vistas de detalles, vistas globales y desplazamientos sincronizados.

Los desarrolladores de visualización pueden integrar al sistema sus visualizaciones independientes con un API simple. La evaluación de esta herramienta reveló beneficios en cuanto a aspectos cognitivos y su usabilidad, mejorando el rendimiento de los usuarios entre un 30 y un 80 por ciento, dependiendo de la tarea realizada (North and Shneiderman, 2000).

Algunas de las técnicas de visualización que están incluidas en esa herramienta son los Diagramas de Dispersión y los TreeMap (North and Shneiderman, 2000).

### **1.3.2 GeoVista Studio**

*GeoVista Studio* es una herramienta de código abierto que implementa un ambiente de desarrollo basado en componentes. Suministra una interfaz de programación visual como muchos sistemas de VisCi, a través de la cual los usuarios pueden construir aplicaciones de forma rápida utilizando JavaBeans<sup>2</sup>. El ambiente de programación visual permite a los analistas empaquetar funcionalidades dentro de un programa de trabajo. GeoVista soporta el desarrollo de aplicaciones geográficas y no geográficas.

Para soportar la interoperabilidad de datos OpenGIS, los desarrolladores de GeoVista han comenzado a adaptar y extender la biblioteca GeoTools en cuanto al acceso a datos y los métodos de visualización. GeoTools es una biblioteca complementaria de código abierto

---

<sup>2</sup> Los **JavaBeans** son un modelo de componentes creado por Sun Microsystems para la construcción de aplicaciones en Java. La especificación de JavaBeans de Sun Microsystems los define como "componentes de software reutilizables que se puedan manipular visualmente en una herramienta de construcción".

desarrollada en Java para el desarrollo de soluciones OpenGIS que permite el acceso a datos geoespaciales, el análisis y la representación de tareas.

El principal objetivo de GeoVista Studio es soportar la fusión de diversas capacidades visuales y analíticas en una herramienta de análisis que posibilite la multiperspectiva. GeoVista incluye además un conjunto de técnicas de visualización clásicas como Diagramas de Dispersión, Coordenadas Paralelas y Self Organizing Map (SOM) (Gahegan et al., 2002, Takatsuka and Gahegan, 2002).

La herramienta Exploratory Spatio-Temporal Analysis Toolkit (ESTAT) fue incluida al GeoVista Studio y constituye una buena herramienta para analizar datos espacio-temporales asociados a mapas vectoriales de áreas en formato *ESRI shape*.

## **1.4 Conclusiones parciales**

En este capítulo se presentaron los principales elementos teóricos sobre la VisCi, profundizando en las técnicas de visualización de datos multiparamétricos. Además se analizaron características generales de los SIG y se analizaron algunos con el objetivo de seleccionar uno para la incorporación de un módulo de visualización de datos multiparamétricos. Se analizaron los esfuerzos de integración entre estas dos disciplinas y se mostraron ejemplos de integración mediante el enfoque mezclado.

Se concluye que las técnicas de visualización de datos multiparamétricos constituyen valiosas herramientas para el análisis visual de datos, por lo que se escogerán algunas de ellas para ser implementadas en un módulo de visualización e incorporarlas a un SIG.

Se decidió incorporar el módulo de visualización de datos multiparamétricos al SIG gvSIG basándose en el enfoque mezclado de integración.

La integración de este tipo de técnicas en gvSIG constituye un esfuerzo más en el acercamiento entre estas dos disciplinas. En el siguiente capítulo se presentan los aspectos de diseño e implementación relacionados con la integración de las técnicas de visualización de datos multiparamétricos en gvSIG.

## **CAPÍTULO 2. DISEÑO E IMPLEMENTACIÓN DE LA EXTENSIÓN**

En el presente capítulo se definen los principales aspectos de la extensión desarrollada para gvSIG y las técnicas que serán implementadas en el mismo. Además se exponen aspectos fundamentales en cuanto al diseño e implementación del módulo.

### **2.1 Selección de las técnicas a incluir en la extensión**

El proceso de desarrollo de técnicas de visualización para algún sistema tiene como punto de partida natural la estructura y tipos de datos que serán utilizados. Esto es producto del modelo de clasificación de las técnicas que se describió en el capítulo anterior. La correcta descripción de los datos a utilizar permite identificar las técnicas que pueden aportar más al módulo.

#### **Descripción de los datos**

La fuente de datos multiparamétricos que utiliza el módulo consiste en un conjunto  $C$  de cardinalidad  $m$  en que cada elemento  $O_i = \langle V_1, \dots, V_n \rangle$ ,  $i=1 \dots m$  es una  $n$ -upla. A cada elemento de  $C$  se le conoce como instancia u observación y cada componente de una observación es una dimensión o variable (Theisel, 2000, Pérez and Ortega, 2005). En estos datos, cada variable puede tener un nivel de medición continuo o nominal.

Los conjuntos de datos multiparamétricos pueden tener una variable cantidad de dimensiones y observaciones. Esta variabilidad depende de la problemática cuyas variables hayan sido identificadas y la cantidad de observaciones que se realicen de las mismas.

Otros tipos de datos que se tienen en cuenta son los datos geográficos. Los datos geográficos consisten en mapas vectoriales en formato *ESRI shape* que se pueden utilizar como fondo para realizar un análisis teniendo en cuenta varias regiones a la vez. También

pueden incluir las diferentes ubicaciones a las que pertenecen determinados conjuntos de datos multiparamétricos. Estos mapas de puntos se suministran en forma de mapa vectorial en formato *ESRI shape*.

### **Elección de las técnicas de visualización**

Al realizar el estudio de los objetivos del proyecto, así como las funcionalidades que se desean aportar al SIG gvSIG se decidió desarrollar dos paquetes: el primero tiene como objetivo la visualización de manera independiente de un único conjunto de datos multiparamétricos y el segundo la visualización de manera coordinada de varios conjuntos de datos multiparamétricos, teniendo una percepción espacial de la localización de cada conjunto de datos sobre un mapa.

El objetivo de ambos paquetes es la visualización de datos multiparamétricos, por lo tanto las técnicas a implementar serán las mismas para cada caso. Sólo se diferencian en algunos aspectos de implementación que serán explicados en este capítulo.

Las técnicas que se decidieron incluir en el módulo de visualización para datos multiparamétricos son:

- Coordenadas Paralelas.
- Gráfico de Andrews.
- Técnicas basadas en iconos.
  - Icono en forma de estrella.
  - Shape coding.
  - Icono en forma de barras.
- Segmentos de Círculo.
- Patrones Recursivos.



## 2.2 Diseño de la extensión realizada a gvSIG

El lenguaje UML (Unified Modeling Language) (Rumbaugh and Booch, 2000) fue el utilizado para realizar el diseño de la herramienta de VisCi, que tiene como objetivos principales la especificación, visualización, construcción y documentación de los productos de un sistema de software. Este lenguaje es usado por el RUP (Rational Unified Process) como lenguaje de modelado para lo cual se basa en todos sus tipos de diagramas, que constituyen diferentes vistas del modelo del producto.

Los diagramas UML empleados en el diseño del módulo fueron: diagramas de casos de uso, diagramas de transición de estados y diagramas de clases.

La herramienta empleada para el modelado de los casos de uso, los diagramas de transición de estados y los diagramas de clases fue *Visual Paradigm para UML versión 6.0*.

### 2.2.1 Análisis de actores y casos de uso

Los modelos de casos de uso proporcionan un medio sistemático e intuitivo de capturar requisitos funcionales del sistema basándose en los requerimientos de los usuarios. Este modelo se realiza identificando cada actor del sistema como los posibles usuarios para los cuales está realizado el mismo.

La extensión implementada está destinada a un solo tipo de actor que es el especialista o investigador de cualquier rama que desee explorar sus datos de manera visual y que sean capaces de interpretar de forma correcta los resultados que brinda la herramienta. Dicho especialista se ha nombrado en el diagrama de la Figura 10 como Especialista.

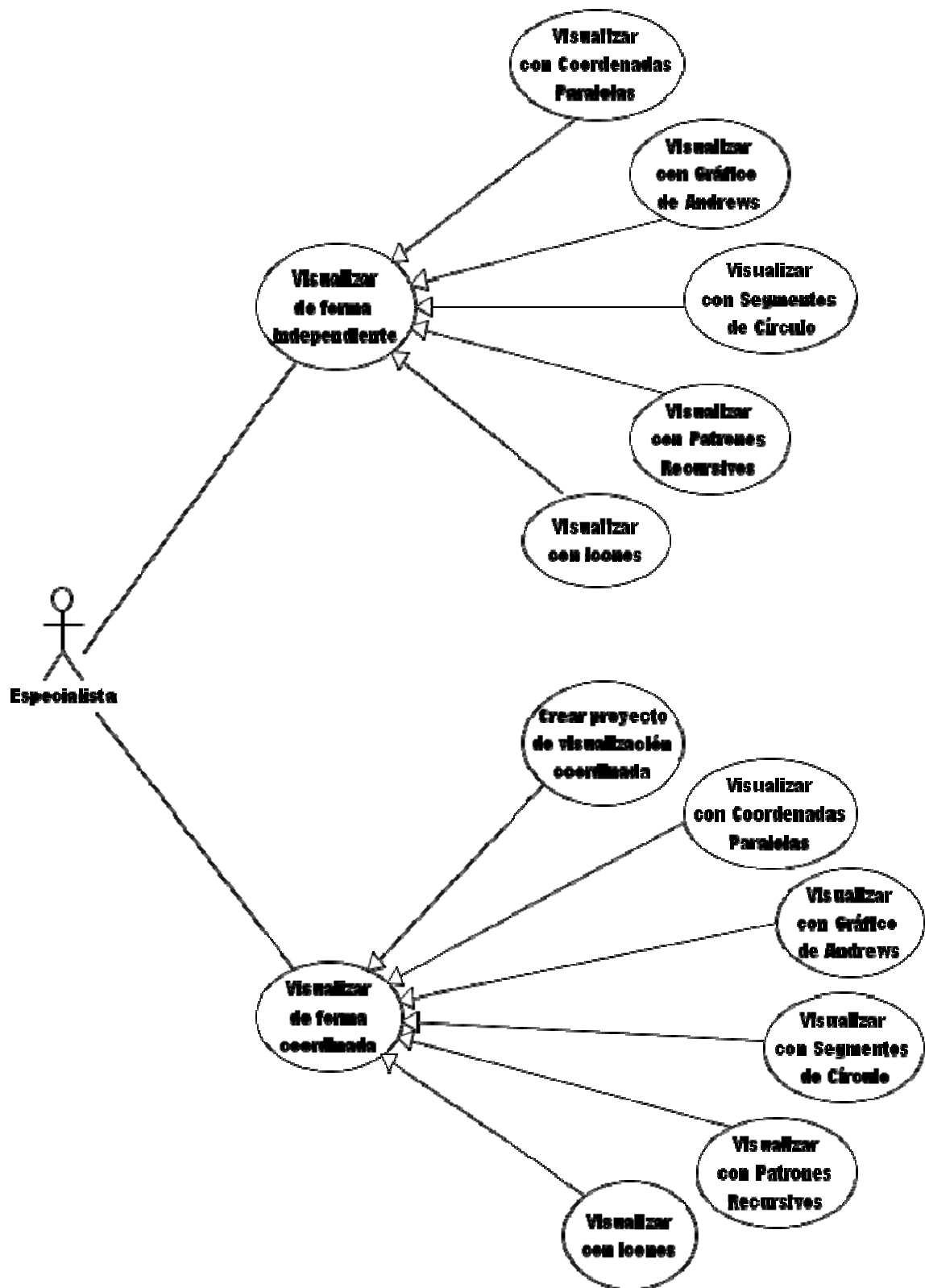


Figura 10: Diagrama de Casos de uso.

Los casos de uso principales que se consideraron fueron: visualizar de forma independiente y visualizar de forma coordinada.

La visualización independiente significa visualizar, mediante técnicas de visualización científica, un único conjunto de datos. Este conjunto de datos puede corresponder a una única localización o incluir variables de diferentes localizaciones. En este caso la visualización se realiza normalizando los valores de las variables con respecto a sus extremos en el conjunto de datos. Como instancias de este caso de uso tenemos la visualización utilizando Coordenadas Paralelas, Gráfico de Andrews, Segmentos de Círculo, Patrones Recursivos y técnicas basadas en iconos, permitiéndole al Especialista visualizar los datos con las técnicas que sean de su interés.

La visualización coordinada significa visualizar, mediante técnicas de visualización científica, múltiples conjuntos de datos correspondientes a varios puntos en el espacio. Es coordinada porque todas las visualizaciones se realizan normalizando los valores de las variables teniendo en cuenta los mínimos y máximos en todos los conjuntos de datos para cada variable. La visualización coordinada permite obtener información sobre el origen de los datos visualizados, referenciando los gráficos generados por las técnicas a la localización de los datos en un mapa; esta referencia puede realizarse emplazando los gráficos sobre el punto de origen de los datos o visualizando los gráficos en ventanas independientes y relacionarlas con sus respectivas localizaciones en el mapa. Como instancias de este caso de uso el Especialista puede crear un proyecto de visualización coordinada y visualizar los datos con Coordenadas Paralelas, Gráfico de Andrews, Segmentos de Círculo, Patrones Recursivos y técnicas basadas en iconos.

### **2.2.2 Diagramas de transición de estados**

El sistema pasa por una serie de estados durante el proceso de visualizar un único conjunto de datos; dichos procesos se muestran en el diagrama de la Figura 11.

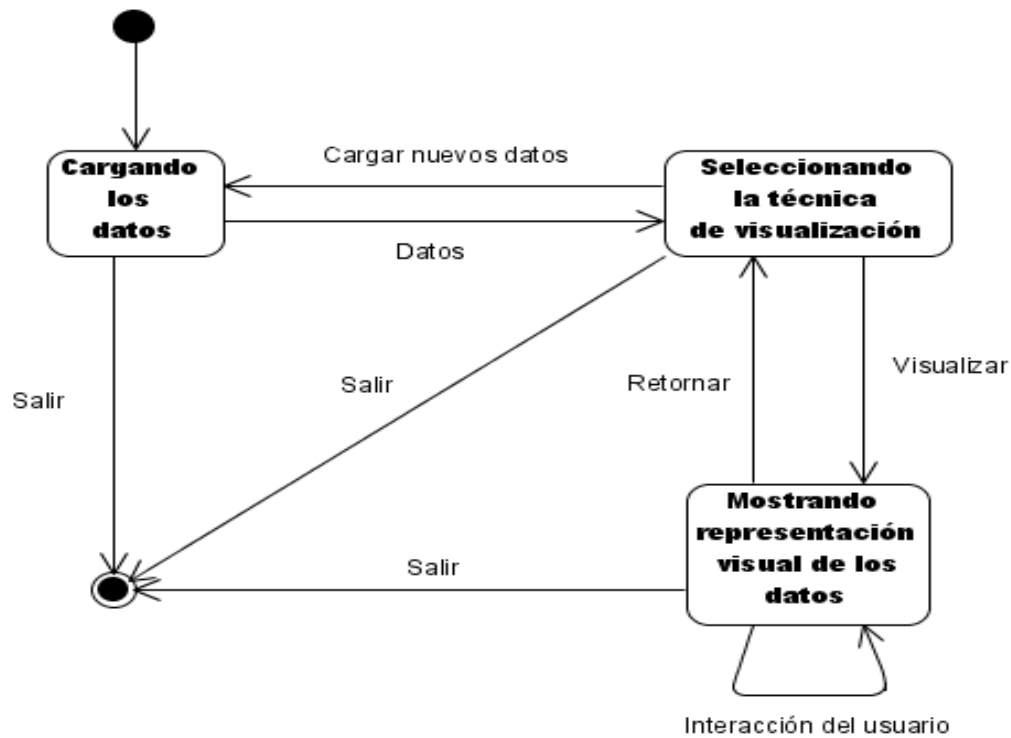


Figura 11: Diagrama de transición de estado para el caso de uso Visualizar de forma independiente.

El estado *Cargando los datos* representa la acción del usuario de seleccionar un conjunto de datos multiparamétricos, para luego visualizarlos mediante alguna de las técnicas de visualización de datos multiparamétricos. Una vez cargados los datos el especialista puede seleccionar una de las técnicas de visualización disponibles, este estado está representado por *Seleccionando la técnica de visualización*. Al visualizar los datos el sistema pasa al estado *Mostrando representación visual de los datos*, donde el especialista puede interactuar con las técnicas y cambiar los parámetros de la visualización; el sistema permite pasar al estado anterior para la selección de otra técnica de visualización y a su vez cargar otro conjunto de datos.

Los estados por los que pasa el sistema durante el proceso de visualización de manera coordinada se muestran en el diagrama de la Figura 12.

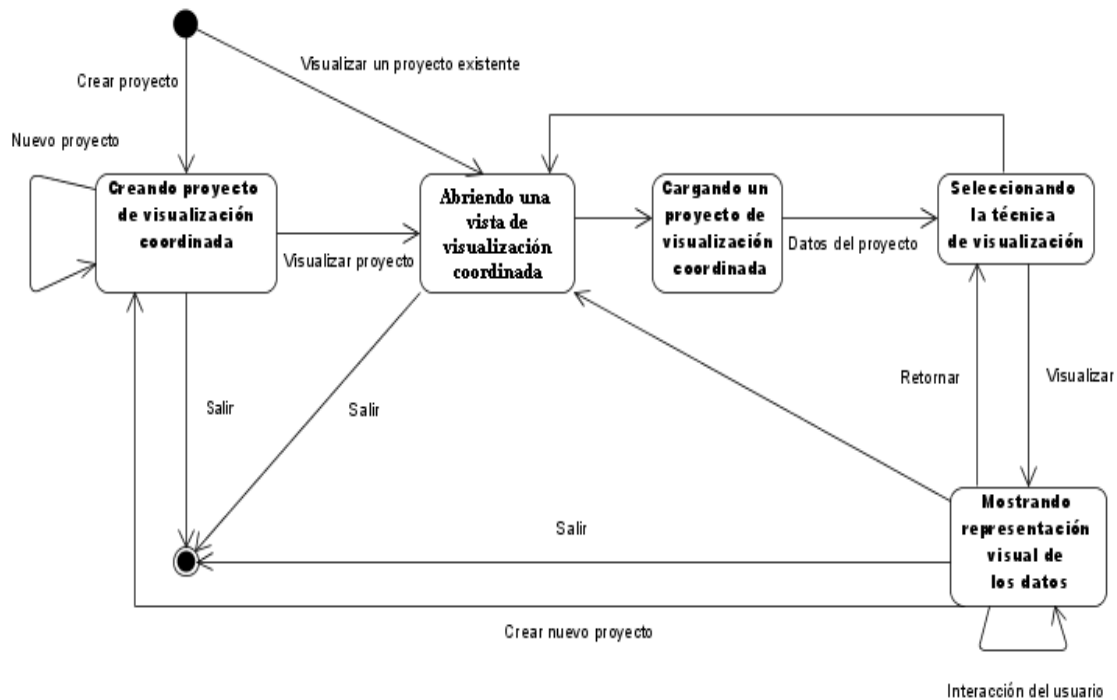


Figura 12: Diagrama de transición de estados para el caso de uso visualizar de forma coordinada.

Inicialmente el sistema puede pasar por dos estados: *Creando proyecto de visualización coordinada*, estado en que el usuario crea y configura un proyecto de visualización coordinada, y *Abriendo una vista de visualización coordinada*, estado en que el usuario crea y abre una vista de visualización para el análisis visual de varios conjuntos de datos de manera coordinada, en la cual se cargará un proyecto de este tipo. De este último el sistema pasa al estado *Cargando un proyecto de visualización coordinada*, en el que se añade un proyecto a la vista previamente abierta. Al cargarse los datos en la vista el sistema pasa al estado *Seleccionando la técnica de visualización científica*, en que el usuario selecciona la técnica de visualización que utilizará para mostrar la representación visual de los datos. Al visualizar los datos con alguna técnica el sistema pasa al estado *Mostrando representación visual de los datos*, en el que el usuario puede interactuar con las técnicas y cambiar los parámetros de la visualización.

### 2.2.3 Diagramas de clases

Como se mencionó anteriormente el módulo de visualización está dividido en dos paquetes principales: el de visualización independiente y el de visualización coordinada. En este epígrafe se muestra la estructura de clases de ambos paquetes y su integración a gvSIG.

#### 2.2.3.1 Paquete para la visualización independiente

El objetivo del paquete para la visualización independiente es permitir la visualización de un único conjunto de datos. Las clases para la visualización que conforman este paquete se pueden apreciar en la Figura 13.

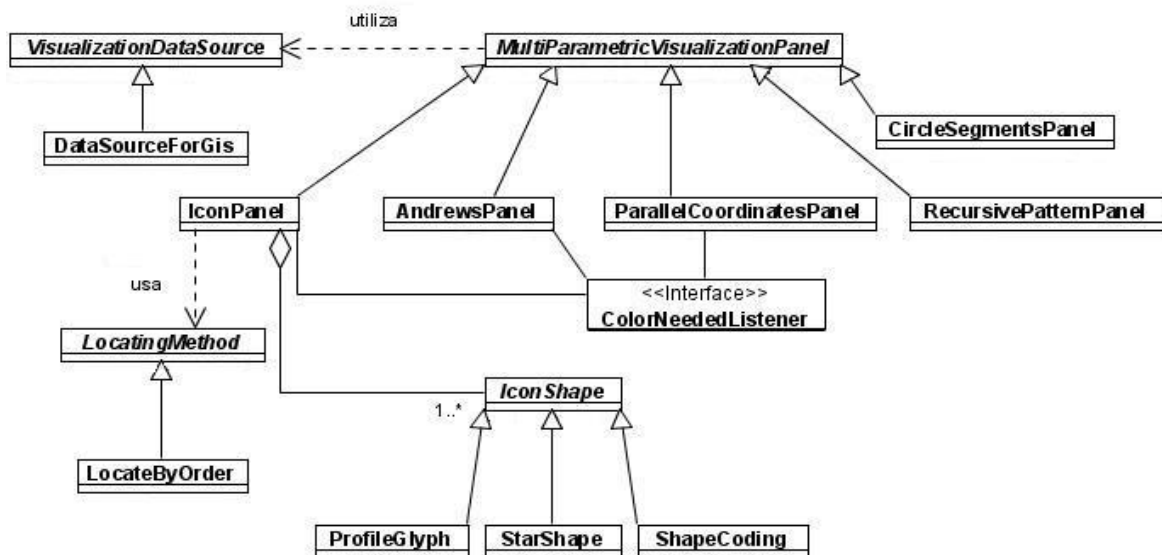


Figura 13: Diagrama de clases del paquete de visualización independiente.

Las clases *IconPanel*, *AndrewsPanel*, *ParallelCoordinatesPanel*, *RecursivePatternPanel* y *CircleSegmentsPanel* son paneles donde son visualizadas cada una de las técnicas de visualización de datos multiparamétricos. Obsérvese además que todas esas clases heredan de *MultiParametricVisualizationPanel* que es una clase abstracta que utiliza a la clase *VisualizationDataSource*.

El siguiente diagrama de clases muestra la relación entre los diferentes paneles de configuración de las técnicas (Figura 14).

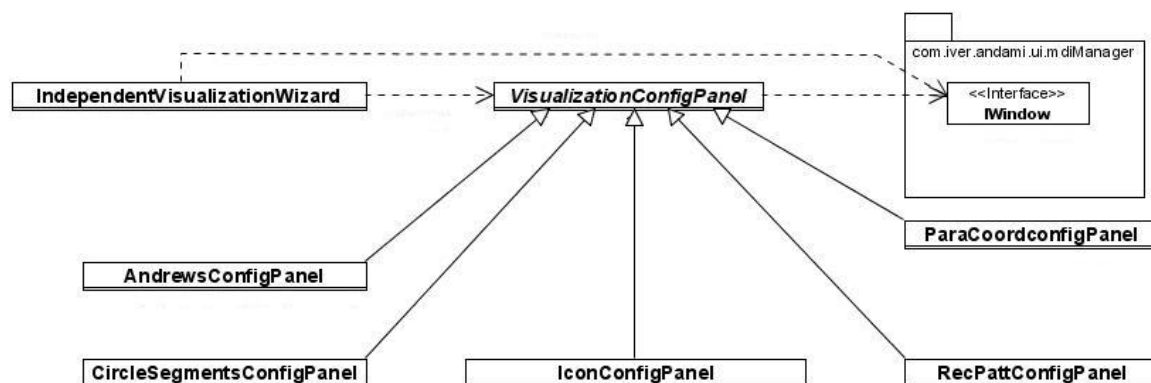


Figura 14: Diagrama de clases de configuración de las técnicas.

La clase *IndependentVisualizationWizard* permite cargar los datos y visualizarlos utilizando los paneles de configuración de las técnicas. Para poder mostrar las ventanas gvSIG se vale de la interfaz *IWindow* del paquete *com.iver.andami.ui.mdiManager*. Las clases que implementen esta interfaz deberán implementar el método **getWindowInfo()**, que devuelve un objeto de tipo *WindowInfo* que contiene la información de la ventana y las operaciones que se pueden realizar sobre la misma, como cambiar el tamaño, minimizarla, etc. Además tiene que implementar el método **getWindowProfile()** que devuelve un objeto con la descripción de la ventana.

La Figura 15 muestra la relación entre las clases de configuración y de visualización de las técnicas.

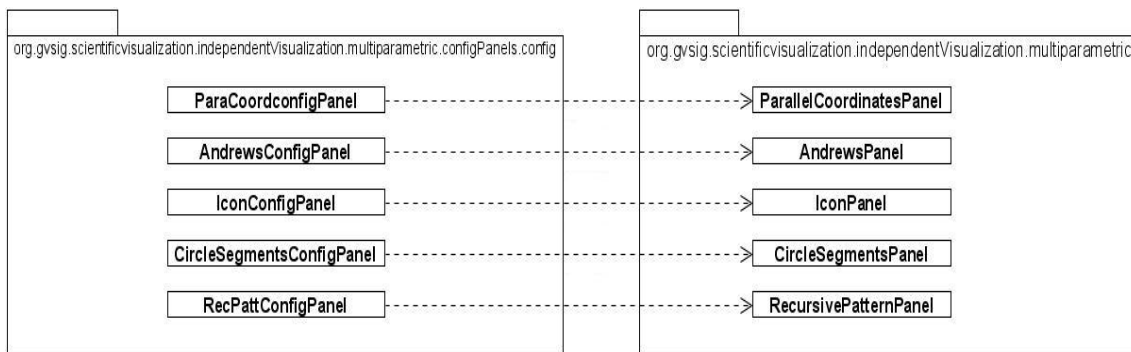


Figura 15: Diagrama de relación entre las clases de configuración y de visualización de las técnicas.

La integración de las funcionalidades de este paquete a gvSIG se muestra en la Figura 16.

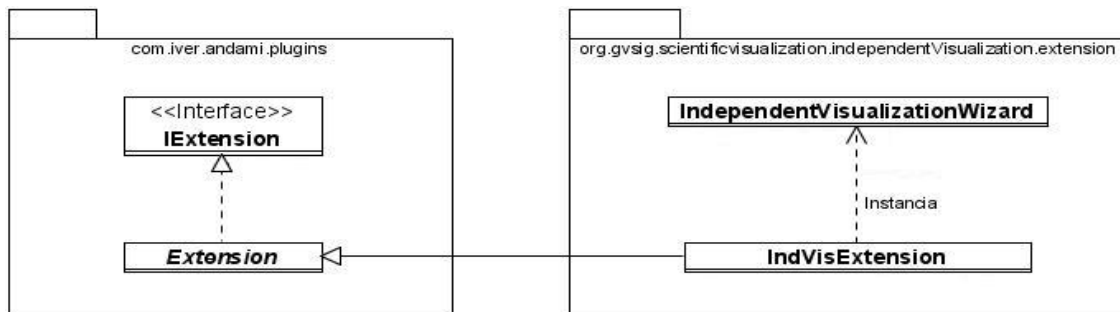


Figura 16: Diagrama de clases. Integración a gvSIG.

La Figura 16 muestra la relación de la clase *IndVisExtension* del paquete *org.gvsig.scientificvisualization.independentVisualization.extension* con la clase *Extension* del paquete *com.iver.andami.plugins*. En este caso *IndVisExtension* agrega una opción de menú y un botón a la barra de herramientas de gvSIG. Al ser pulsada alguna de estas opciones se crea una instancia de la clase *IndependentVisualizationWizard*, que es la encargada de gestionar la selección de un conjunto de datos y visualizarlos con alguna de las técnicas.

La metodología a seguir para agregar nuevas extensiones a gvSIG se explica a continuación. Las extensiones que se deseen añadir a gvSIG deben heredar de *Extension*, que implementa la interfaz *IEExtension* y deben implementar los siguientes métodos:



- **public void initialize():** Una vez que se leen y se ordenan todas las extensiones presentes en los módulos, se llama a este método para cada una de las extensiones, de forma secuencial, siguiendo el orden definido en la prioridad de las mismas.
- **public boolean isVisible():** Este método permite definir la visibilidad de la extensión haciendo chequeos de las condiciones bajo las cuales se desea que la extensión esté activa o visible.
- **public boolean isEnabled():** Estando la extensión activa puede a su vez estar habilitada o no. Mediante este método se pueden definir las condiciones bajo las cuales se le permite al usuario hacer uso de la extensión.
- **public void execute(String actionCommand):** En el *config.xml* de cada módulo añadido como una extensión a gvSIG se definen elementos de la interfaz de usuario (botones, menús, etc.). Cuando el usuario pulsa en uno de estos elementos de la interfaz, se busca en la extensión asociada y se llama a este método. Los elementos de la interfaz pueden tener asociado una cadena que actúa como identificador de comando (*actionCommand*). Cuando se llama al método *execute*, se pasa como parámetro el *actionCommand* asociado al botón o menú que fue pulsado. De esta forma, si se tienen varios botones o menús en una extensión, se puede decidir cual acción ejecutar comprobando el *actionCommand* generado.

### 2.2.3.2 Paquete para la visualización coordinada

Uno de los objetivos del paquete para la visualización coordinada de varios conjuntos de datos es poder obtener una percepción sobre la localización de los datos visualizados. Para ello se hace necesario contar con una ventana o vista donde se puedan cargar los mapas para la localización de los gráficos y permitir la visualización de las técnicas.

La siguiente figura muestra un diagrama de clases simplificado donde se observa la representación interna de la vista añadida a gvSIG para realizar la visualización coordinada. A partir del modelo extensible de documentos definido en gvSIG se ha definido un nuevo tipo de documento llamado *ProjectSV* a través de una clase base

*ProjectViewBase*. El nuevo tipo de documento tiene su propia factoría *ProjectSVFactory*, que se registra como factoría de documentos de gvSIG.

*ProjectSV* hereda de la clase base *ProjectViewBase* que contiene un *MapContext* con una instancia de *FLayers*, quien a su vez contiene implementaciones alternativas de algunos métodos (por ejemplo para añadir o borrar una capa).

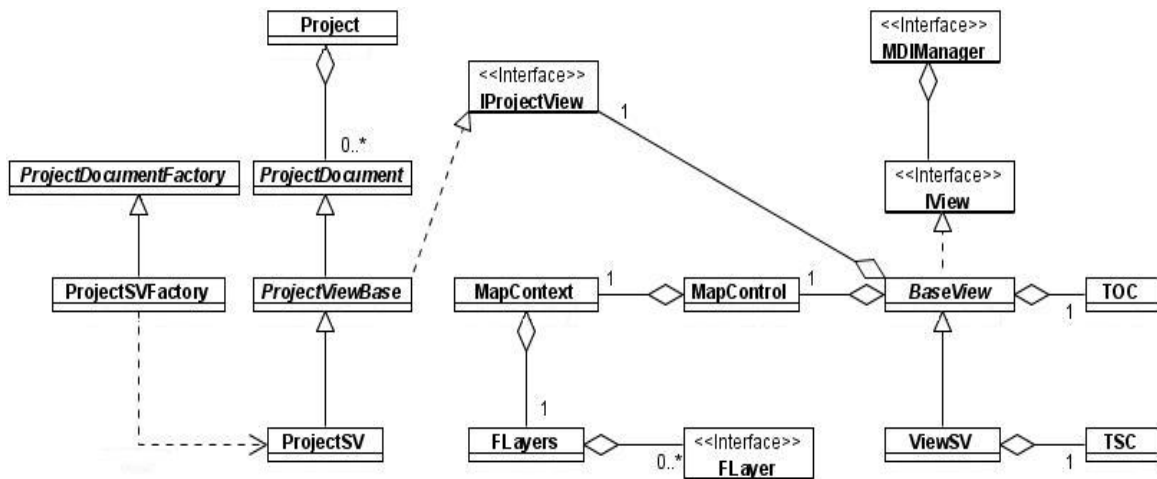


Figura 17: Diagrama de clases de la vista para la visualización coordinada.

Utilizando la metodología para la adición de extensiones a gvSIG, el nuevo tipo de documento creado puede ser registrado como documento de gvSIG. La clase *ExtensionSV*, que hereda de *Extension*, se encarga de registrar el nuevo tipo de documento llamando al método **register()** de la clase *ProjectSVFactory*. El diagrama de la Figura 18 muestra esta relación.

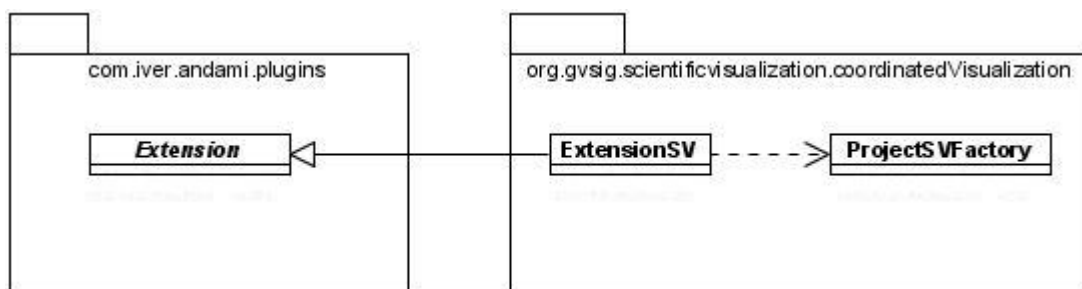


Figura 18: Integración del paquete de visualización coordinada como una extensión de gvSIG.

El método **createWindow()** de la clase *ProjectSV* devuelve en cada llamada una nueva instancia de la clase *ViewSV*, que es la vista en sí. Cada instancia de *ViewSV* contiene una instancia de la clase *TSC* (Techniques Selection and Configuration) cuya función es permitir la utilización de las técnicas de visualización. La Figura 19 muestra la relación de esta clase con los paneles de configuración de las técnicas.

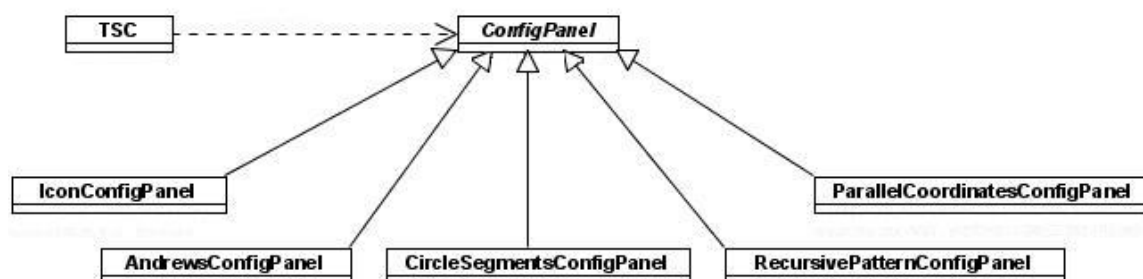


Figura 19: Diagrama de las clases de configuración de las técnicas.

En principio todas las técnicas pueden ser visualizadas en paneles independientes, teniendo una percepción del origen de los datos que están visualizando cada uno de los paneles en el mapa.

Dadas las características propias de cada técnica (tamaño de la imagen generada, etc.) se analizaron cuáles eran más efectivas para ser visualizadas sobre el mapa. Las técnicas seleccionadas fueron Segmentos de Círculo, Patrones Recursivos y las basadas en iconos. Para poder realizar la visualización de estas técnicas sobre el mapa era necesaria la utilización de una clase que implementara algún tipo de objeto gráfico que pudiese ser añadido a la capa de gráficos en gvSIG. Para esto se utilizó la capa gráfica (*GraphicLayer*) del contexto del mapa de la vista (*MapContext*). Mediante el método **addGraphic(FGraphic g)** de la clase *GraphicLayer* se pueden añadir objetos de tipo *FGraphic* que pueden ser visualizados sobre un mapa en gvSIG. Las relaciones de estas clases del paquete *com.iver.cit.gvsig.fmap* se muestra en la Figura 20.

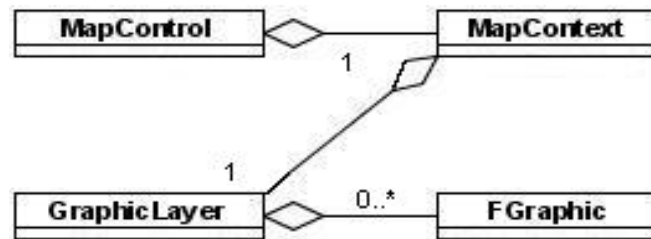


Figura 20: Diagrama de clases.

El diagrama de clases para la visualización tanto sobre el mapa como en paneles se muestra en la Figura 21.

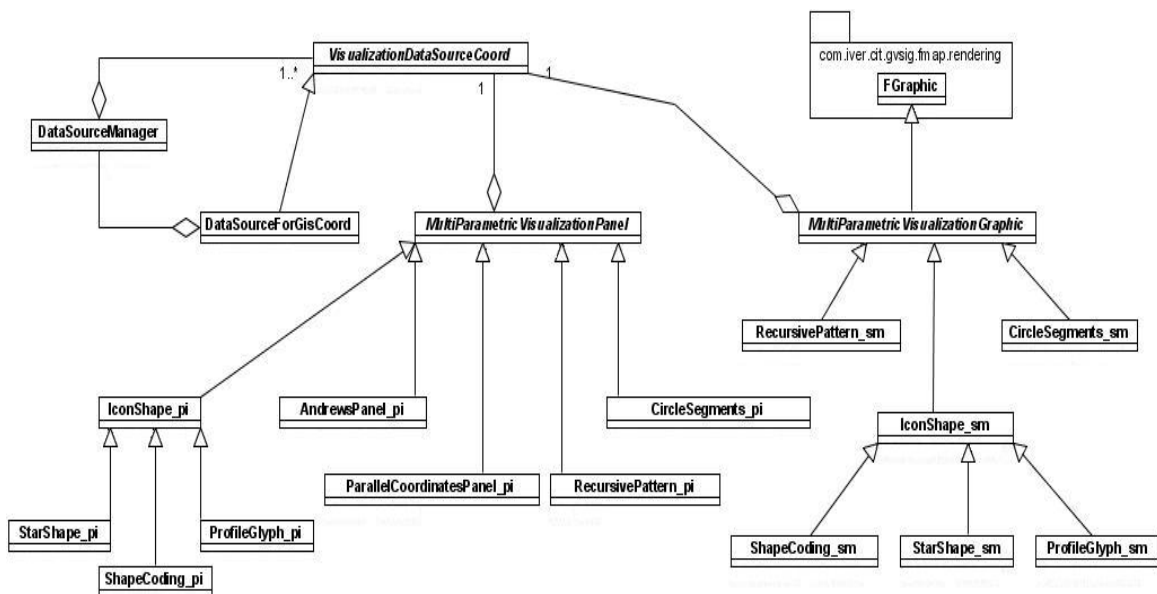


Figura 21: Diagrama de clases de las técnicas de visualización para la visualización coordinada.

En la Figura 21 se puede observar las clases que heredan de *MultiParametricVisualizationPanel*, que son las técnicas que se visualizan coordinadamente en paneles independientes. Se escogió como nomenclatura de las clases asignarles *\_pi* al final, para enfatizar que se trata de paneles independientes. Obsérvese que Coordenadas Paralelas y Gráfico de Andrews solo están en esa jerarquía porque los gráficos generados por estas técnicas ocuparían un gran espacio sobre el mapa y se dificultaría su análisis.

Obsérvese que *MultiparametricVisualizationGraphic* hereda de *FGraphic*, con el objetivo de poder integrar los gráficos de las técnicas a los mapas sobre los que se realizará la visualización. Las clases que se pueden mostrar sobre el mapa y en paneles independientes son *RecursivePattern\_sm*, *CircleSegments\_sm* y las implementaciones de iconos que heredan de *IconShape\_sm*. Note la nomenclatura finalizada en *\_sm* para enfatizar que se visualizan sobre el mapa.

El siguiente diagrama muestra la relación entre los paneles de configuración y las clases de visualización de las técnicas.

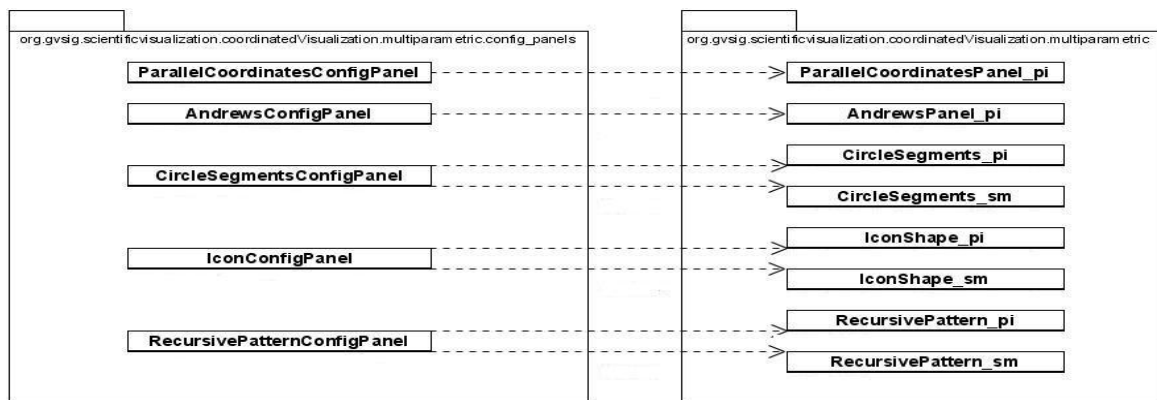


Figura 22: Relación de los paneles de configuración y los paneles de las técnicas.

Los paneles de configuración de las técnicas son los encargados de gestionar los gráficos que son visualizados sobre el mapa y en paneles independientes.

### 2.2.3.3 Jerarquía de clases para la lectura de datos

En los subepígrafes anteriores se expuso la estructura de clases para la visualización independiente y coordinada. El objetivo del presente subepígrafe es mostrar cómo ambos paquetes realizan la lectura de los datos para su posterior visualización.

Los datos meteorológicos con que se cuentan están disponibles en dos formatos de archivos: *arff* (*attribute-relation file format*) y *dbf* (*data base format*). Los ficheros *arff* son

archivos de texto donde se definen los atributos, los tipos de los atributos y la matriz de datos, cuyas columnas representan los atributos y cada fila representa las mediciones de las variables (punto de dato). El anexo 1 muestra con mayor claridad la estructura de estos ficheros. Los ficheros *dbf* son ficheros en formato de tabla y son utilizados en disímiles aplicaciones. Las columnas de la tabla de este fichero representan los atributos y cada fila representa un punto de dato. Para su utilización son necesarios lectores específicos que realicen la lectura de los valores de cada campo de la tabla.

La Figura 13 del subepígrafe 2.2.3.1 muestra la relación existente entre las clases *MultiparametricVisualizationPanel* y *VisualizationDataSource* del paquete para la visualización independiente. La clase abstracta *VisualizationDataSource* define un conjunto de métodos abstractos que serán usados por *MultiparametricVisualizationPanel* para realizar la visualización de los datos. Esta forma de diseño permite que el manejo de los datos a través de los métodos abstractos de la clase *VisualizationDataSource* sea homogéneo, independientemente de la fuente de datos que se utilice.

Teniendo en cuenta que los datos se encuentran en archivos, es necesario tener clases que realicen la lectura de cada tipo de fichero y una clase que herede de *VisualizationDataSource* para el manejo de los datos que sean extraídos. Ver la Figura 23.

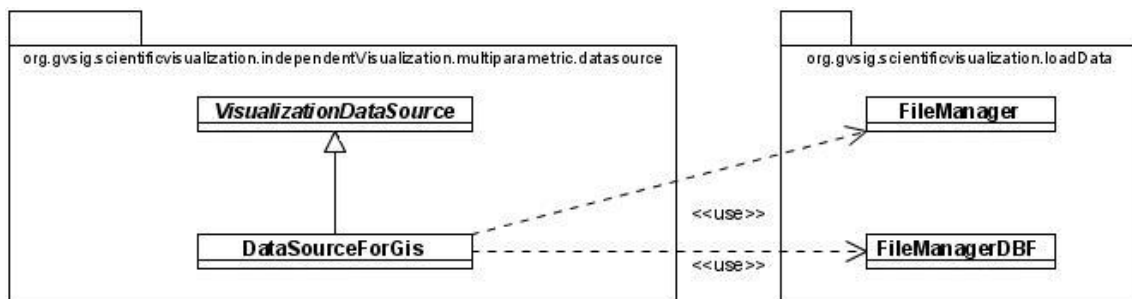


Figura 23: Clases que intervienen en la lectura de los ficheros y el manejo de los datos.

La figura anterior muestra la clase *DataSourceForGis* que hereda de *VisualizationDataSource* y que usa las clases *FileManager* para la lectura y obtención de los datos de un fichero en formato *arff* y *FileManagerDBF* para la lectura y obtención de

los datos de un fichero en formato *dbf*. La clase *FileManagerDBF* emplea para la lectura del fichero *dbf* la clase *DBFDriver* del paquete *libFMap* de gvSIG.

La Figura 21 del subepígrafe 2.2.3.2 muestra la relación existente entre las clases *VisualizationDataSourceCoord* y las clases para la visualización de las técnicas en paneles independientes y sobre el mapa, *MultiparametricVisualizationPanel* y *MultiparametricVisualizationGraphic* respectivamente, del paquete para la visualización coordinada. Como se ha explicado en epígrafes anteriores, la visualización coordinada implica visualizar todos los datos de manera tal que todos los gráficos generados por las técnicas muestren la variación de las variables para todos los conjuntos de datos, de ahí que sea necesaria una estructura donde se controlen todos los valores de las variables de manera global. Para ello se creó la clase *DataSourceManager* que contiene varios *VisualizationDataSourceCoord* y se encarga de controlar la homogeneidad de cuantos conjuntos de datos sean insertados. Ver la Figura 24.

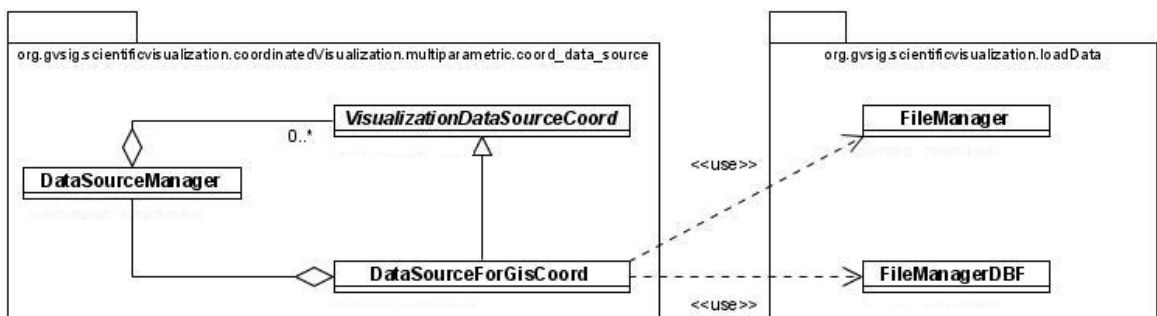


Figura 24: Clases que intervienen en la lectura de los ficheros y el manejo de los datos para la visualización coordinada.

La figura anterior muestra la clase *DataSourceForGisCoord* que hereda de *VisualizationDataSourceCoord* y usa las clases *FileManager* y *FileManagerDBF* para la lectura de los ficheros de datos. La clase *VisualizationDataSourceCoord* se diferencia a *VisualizationDataSource* en que contiene otros métodos abstractos como `getGlobalMin()` y `getGlobalMax()`, que deben devolver los valores extremos de las variables con respecto a todos los conjuntos de datos. Esto se logra asociándole a cada *DataSourceForGisCoord* el *DataSourceManager* en el cual está incluido.

## **2.3 Implementación de la extensión**

En este epígrafe se exponen los temas referentes a la implementación de la extensión teniendo en cuenta las tecnologías utilizadas para la implementación y algunas consideraciones sobre las técnicas implementadas.

### **2.3.1 Selección de las tecnologías necesarias para la implementación**

Durante el desarrollo o extensión de un sistema deben tomarse decisiones de implementación como la plataforma objetivo y el lenguaje de desarrollo. En cada proyecto hay que puntualizar en determinadas áreas específicas de la aplicación. En el caso particular de este proyecto debe fijarse la API (Application Programming Interface) gráfica a utilizar.

#### **Elección del lenguaje de desarrollo**

El objetivo de este proyecto es adicionar técnicas de visualización para datos multiparamétricos a gvSIG. Esta herramienta está construida con el lenguaje Java. El proceso de integración de las técnicas de visualización con el sistema actual inevitablemente tendrá que interactuar y utilizar el código fuente de gvSIG. Esto hace al lenguaje Java la mejor elección para mantener la mayor compatibilidad posible con el sistema actual.

A las razones descritas anteriormente se puede agregar que este proyecto necesita de un lenguaje potente en la implementación de interfaz gráfica, lo que se puede conseguir con la plataforma Java. Por otra parte la implementación del módulo de visualización científica en este lenguaje permite la fácil reutilización y adaptación de varias de sus componentes, facilitando su integración con otros sistemas, producto de la flexibilidad y potencialidades de Java.



### **Elección de la API gráfica**

Las técnicas que se decidieron implementar no requieren de gran complejidad en la API gráfica. En particular no necesitan de una API 3D. Por esto se decidió utilizar la tecnología Java2D que brinda un excelente desempeño y potencialidad para la programación gráfica en dos dimensiones. Esta decisión facilita a los usuarios la utilización del sistema ya que no necesitan paquetes adicionales a los de la plataforma Java.

La API Java 2D está constituida por un conjunto de clases que pueden ser usadas para crear gráficos de elevada calidad. Esto incluye técnicas como: transformaciones geométricas, disminución de curvas o líneas causadas por una baja resolución del gráfico, dibujado de mapas de píxeles, procesamiento de imágenes, diseño bidireccional de texto, además de excelentes vías para el uso del color, solo por mencionar algunos. Utilizamos esta API para crear las representaciones visuales de las técnicas de visualización.

### **Elección del Entorno de Desarrollo Integrado**

Existe un conjunto de Entornos de Desarrollo Integrado (IDE, de sus siglas en inglés) que permiten el desarrollo de proyectos en Java. El IDE Eclipse versión 3.4 fue el seleccionado como ambiente de programación de esta extensión, debido a que resulta cómodo y fácil de usar para depurar programas. Además la estructura de proyectos de gvSIG está diseñada para ser compilado con mayor facilidad utilizando este IDE.

### **Elección de la versión de gvSIG a utilizar como sistema base**

El progresivo desarrollo de gvSIG por parte de una comunidad de desarrolladores hace posible que con cierta regularidad se actualice su versión. Actualmente la versión estable de gvSIG es la 1.9 por lo tanto fue la elegida para integrarle el módulo de visualización para datos multiparamétricos.

### 2.3.2 Consideraciones sobre las técnicas implementadas

Las técnicas seleccionadas presentan una estructura general, la cual contiene parámetros que pueden variar de una implementación a otra. A continuación se establecen las consideraciones realizadas en cada método para la implementación realizada.

#### Técnicas de visualización basadas en iconos

En el caso de la visualización de manera independiente la técnica se implementó en la clase *IconPanel* realizando la localización de los iconos de acuerdo al orden natural de los datos. Se incluyó en el paquete tres tipos de iconos que están implementados en las clases *ProfileGlyph*, *ShapeCoding* y *StarShape*. Este último es un tipo de icono procedural que se basa en la técnica *Campo de Estrellas*.

En el caso de la visualización de forma coordinada la implementación varía con lo antes expuesto. Para este caso se necesitaba realizar la visualización de un icono a la vez. Para ello se crearon las clases *IconShape\_pi* e *IconShape\_sm* para la visualización coordinada, en paneles independientes y sobre el mapa respectivamente, que solo visualizan un registro y por lo tanto se genera un icono a la vez. Los restantes registros del volumen total de datos podrán ser visualizados cambiando solamente el índice que referencia al punto de dato que está siendo visualizado. Las subclases que implementan un tipo de icono específico son *StarShape\_pi*, *ProfileGlyph\_pi* y *ShapeCoding\_pi*, que heredan de *IconShape\_pi* y *StarShape\_sm*, *ProfileGlyph\_sm* y *ShapeCoding\_sm*, que heredan de *IconShape\_sm*.

#### Gráfico de Andrews

Como se analizó en el capítulo anterior, la técnica Gráfico de Andrews utiliza una función para representar cada observación. La implementación realizada utiliza una serie de Fourier de la forma:

$$f_i(x) = X_{i,1} + \sum_{n=1}^D (X_{i,2n} * \sin(2n * \pi * t) + X_{i,2n+1} * \cos(2n * \pi * t))$$

Esta función es evaluada para cada instancia  $i$  con  $t \in [0,1]$ .  $X$  es una matriz que contiene los datos sobre las observaciones. Cada fila de la matriz corresponde a una observación y cada columna es una variable o dimensión de la observación. El valor  $D$  que se utiliza en la sumatoria se obtiene a partir de la expresión  $d/2 - 1$ , donde  $d$  es el número de dimensiones que poseen las instancias.

Sin embargo en conjuntos de datos con una cantidad de dimensiones extremadamente elevada las funciones pueden perder su suavidad, notándose los segmentos que conforman la curva. La técnica está implementada en las clases *AndrewsPanel* del paquete para la visualización independiente y en *AndrewsPanel\_pi* del paquete para la visualización coordinada. Soporta características como reducción del conjunto de los datos de trabajo y mapa de colores para las observaciones.

### Coordenadas Paralelas

La técnica Coordenadas Paralelas fue implementada en la clase *ParallelCoordinatesPanel* del paquete para la visualización independiente y en la clase *ParallelCoordinatesPanel\_pi* del paquete para la visualización coordinada. La implementación admite el uso de variables nominales que son representadas con histogramas. Se puede además elegir subconjuntos de trabajo y mapa de colores para las observaciones.

### Segmentos de Círculo

La técnica Segmentos de Círculo fue implementada en las clases *CircleSegmentsPanel* del paquete de visualización independiente, en *CircleSegmentsPanel\_pi* para la visualización coordinada en paneles independientes y en *CircleSegments\_sm* para la visualización coordinada sobre el mapa.

La implementación difiere en cierta medida de la idea expuesta en el primer capítulo. La mayor diferencia con la idea original es que se ha utilizado un esquema en que cada par observación-dimensión ya no es representada por un simple píxel, sino por un punto o círculo cuyo tamaño puede ser variado interactivamente. Esta pequeña modificación resulta particularmente útil cuando el conjunto de datos es más bien pequeño. Puede notarse que la idea original de utilizar un solo píxel es un caso particular de la variante aquí utilizada.

### **Patrones Recursivos**

La técnica Patrones Recursivos fue implementada en las clases *RecursivePatternPanel* del paquete de visualización independiente, en *RecursivePatternPanel\_pi* para la visualización coordinada en paneles independientes y *RecursivePattern\_sm* para la visualización coordinada sobre el mapa. La técnica se implementó siguiendo la teoría planteada en el primer capítulo. La implementación permite introducir nuevos niveles de recursividad, posibilitando que el usuario personalice la visualización de los datos de acuerdo al patrón que sea insertado.

## **2.4 Metodología para la adición de nuevas técnicas al módulo de visualización de gvSIG**

El objetivo del presente epígrafe es mostrar una metodología para la adición de nuevas técnicas al módulo de visualización para datos multiparamétricos de gvSIG. Se dividirá en dos subepígrafes para tratar por separado las dos formas de realizar la visualización que se han tenido en cuenta.

### **2.4.1 Metodología para la adición de nuevas técnicas al paquete de visualización independiente**

La adición de una nueva técnica de visualización al paquete para la visualización independiente consta de tres pasos fundamentales. El primero es la creación de una

subclase de *MultiparametricVisualizationPanel*, que debe localizarse en el paquete *org.gvsig.scientificvisualization.independentVisualization.multiparametric* y es la encargada de implementar la técnica de visualización. El siguiente paso es la creación de una subclase de *VisualizationConfigPanel*, que debe localizarse en el paquete *org.gvsig.scientificvisualization.independentVisualization.multiparametric.configPanels.config*, para el manejo de la interfaz de usuario que permite la interacción con los parámetros de configuración de la técnica de visualización. El último paso es la adición de una opción de selección de la técnica a la clase *IndependentVisualizationWizard*.

La clase *MultiParametricVisualizationPanel* que hereda de la clase *javax.swing.JPanel* tiene las operaciones necesarias para acceder a la fuente de datos. Las responsabilidades de cualquier subclase de ella son:

- Realizar la representación visual deseada.
- Manejar la interacción con el usuario a través de los dispositivos de entrada.
- Ofrecer un conjunto de operaciones que permita cambiar los parámetros de la visualización. Estos parámetros pueden ser el color, la geometría y cualquier otra propiedad que afecte la representación.

La clase *VisualizationConfigPanel* hereda de *javax.swing.JPanel*. Cada subclase de *VisualizationConfigPanel* está asociada con una subclase de *MultiParametricVisualizationPanel*. La responsabilidad fundamental de todo *VisualizationConfigPanel* es crear y mantener una técnica de visualización ofreciendo una interfaz para acceder a todos los parámetros de la misma. Otras responsabilidades son:

- Ofrecer operaciones que permitan establecer los parámetros iniciales de la técnica de visualización.
- Transformar la fuente de datos de los archivos a la fuente de datos del paquete de visualización.

La clase *IndependentVisualizationWizard* constituye la interfaz de usuario para la selección de las técnicas con que se deseen visualizar los datos que sean previamente cargados. Para

añadirle la nueva técnica se le agrega una opción a la interfaz, que debe ser un *radio button*, cuya selección muestre la visualización de los datos con la nueva técnica.

### 2.4.2 Metodología para la adición de nuevas técnicas al paquete de visualización coordinada

La adición de una nueva técnica al paquete de visualización coordinada consta de tres pasos. El primero es la creación de las subclases de *MultiparametricVisualizationPanel* y *MultiparametricVisualizationGraphic*, que deben localizarse en el paquete *org.gvsig.scientificvisualization.coordinatedVisualization.multiparametric*, que son las encargadas de implementar las técnicas de visualización, para visualizar de manera coordinada en paneles independientes y sobre el mapa, respectivamente. El segundo paso es crear una subclase de *ConfigPanel*, que debe localizarse en el paquete *org.gvsig.scientificvisualization.coordinatedVisualization.multiparametric.config\_panels*, para el manejo de los parámetros de las técnicas que sean visualizadas sobre el mapa y en paneles independientes. El último paso es la adición de una opción de selección de la técnica a la clase *TSC*.

Las clases *MultiparametricVisualizationPanel*, que hereda de *java.swing.JPanel*, y *MultiparametricVisualizationGraphic*, que hereda de la clase *FGraphic* del paquete *com.iver.cit.gvsig.fmap.rendering*, tienen las operaciones necesarias para acceder a la fuente de datos. Deben permitir realizar la representación visual deseada, manejar las interacciones del usuario y ofrecer un conjunto de operaciones que permita cambiar los parámetros de la visualización como selección de atributos, conjunto de trabajo, entre otros.

La clase *ConfigPanel* hereda de *java.swing.JPanel*. Cada subclase de *ConfigPanel* está asociada con varias subclases de *MultiparametricVisualizationPanel* y *MultiparametricVisualizationGraphic*, en dependencia de la selección realizada por el usuario; las responsabilidades de cualquier subclase de *ConfigPanel* deben ser: ofrecer operaciones que permitan establecer los parámetros iniciales de la técnica de visualización

y transformar la fuente de datos de los archivos a la fuente de datos del paquete de visualización.

La clase *TSC* está integrada a la clase *ViewSV*, que representa una vista de visualización. Permite la selección y visualización de las técnicas implementadas. Para añadirle la nueva técnica se le agrega un nuevo campo al *combobox* donde están listadas las técnicas, cuya selección cargue el panel de configuración y muestre la técnica según el modo de visualización seleccionado.

## 2.5 Conclusiones parciales

La integración de técnicas de VisCi para datos multiparamétricos en el SIG gvSIG, siguiendo un enfoque mezclado entre estas dos disciplinas, le permite al usuario visualizar uno o varios conjuntos de datos, de forma independiente o coordinada respectivamente, obteniendo en el último caso una percepción espacial del origen de los datos visualizados.

Se seleccionaron un conjunto de técnicas de visualización de datos multiparamétricos que fueron incorporadas a gvSIG. Las técnicas incluidas en el módulo de visualización ofrecen un amplio número de beneficios. Se incluyeron algunas de las técnicas tradicionales, con ciertas mejoras, lo que permite mostrar conjuntos de datos con disímiles características. Además el proceso de diseño del nuevo módulo admite la realización de futuras extensiones y modificaciones, quedando propuesta una metodología para la adición de nuevas técnicas.

## **CAPÍTULO 3. PRESENTACIÓN DE LOS RESULTADOS**

En el presente capítulo se realiza una presentación al usuario de las facilidades y funcionalidades del módulo de visualización integrado a gvSIG. Se efectúa un análisis detallado de las opciones y modo de uso de cada una de las formas de visualización implementadas, que constituye una guía para explotar los beneficios de la extensión realizada. También se presenta como caso de estudio el análisis visual de los datos meteorológicos de la provincia Villa Clara con la herramienta desarrollada.

### **3.1 Manual de usuario**

La extensión realizada a gvSIG no trajo como consecuencias nuevos requerimientos del sistema para su ejecución. Para utilizar la extensión basta copiar el módulo compilado para el directorio de extensiones de gvSIG. Se recomienda utilizar la versión 1.9 de gvSIG, ya sea la versión portable o instalable.

El análisis visual de datos multiparamétricos utilizando las técnicas de visualización con que cuenta la extensión implementada, requiere un conocimiento previo de cómo aplicarlas de forma adecuada para lograr una correcta comprensión de los resultados que se obtengan.

La extensión realizada a gvSIG permite la exploración y análisis visual de datos multiparamétricos mediante las técnicas: Coordenadas Paralelas, Gráfico de Andrews, técnicas basadas en iconos, Segmentos de Círculo y Patrones Recursivos. Permite visualizar un único conjunto de datos donde los parámetros de la visualización dependen solamente de los valores de las variables del propio conjunto y visualizar coordinadamente varios conjuntos de datos donde los atributos de la visualización representan el comportamiento global de las variables para la totalidad de los datos.



Para la correcta utilización de la herramienta de visualización integrada a gvSIG se requiere el desarrollo de un manual de usuario. Este epígrafe muestra el uso de cada una de las técnicas y para esto se dividió el contenido en dos subepígrafes, donde se tratan por separado las dos opciones de visualización que se han tenido en cuenta en la implementación de la extensión.

### 3.1.1 Visualización de un conjunto de datos

El especialista puede realizar la visualización de un único conjunto de datos que estén previamente configurados en los formatos de archivos de datos *arff* y *dbf*. Para realizar este análisis se puede acceder a una opción del menú *Visualization/Open and visualize data* o a un botón de la barra de herramientas de gvSIG. En la Figura 25 se pueden observar resaltadas en rojo las respectivas opciones.



Figura 25: Opciones para realizar la visualización de un conjunto de datos.

Al seleccionar algunas de estas opciones se muestra una vista para la selección del archivo de datos, como la mostrada en la Figura 26. Resaltado en rojo con el número uno está el botón que muestra un diálogo de selección de ficheros que permite abrir solo los formatos de archivos permitidos. Luego de ser seleccionado el archivo de datos se muestra su dirección en el campo de texto resaltado con el número dos y se habilita el botón *Next* resaltado con el número tres.

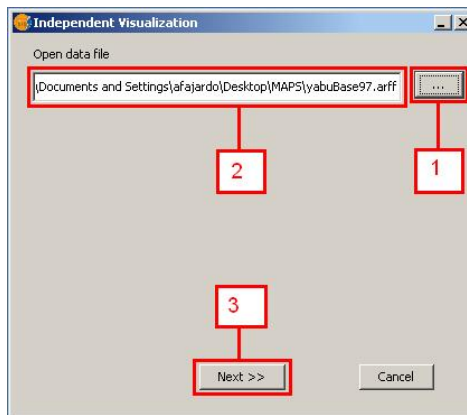


Figura 26: Vista para la selección del fichero de datos.

Al pulsar el botón *Next* se muestra la vista para la selección de las técnicas de visualización, ver la Figura 27.

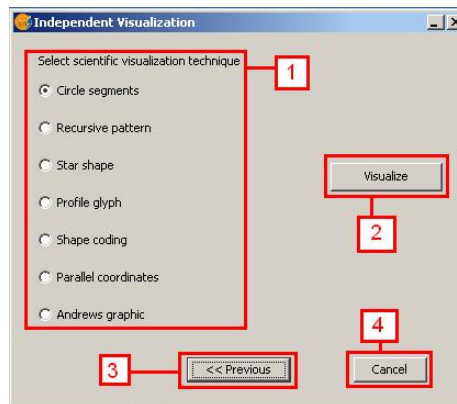


Figura 27: Vista para la selección y visualización de las técnicas.

La región identificada con el número uno muestra las técnicas con las que puede ser visualizado el conjunto de datos seleccionado. Una vez seleccionada alguna de las opciones se procede a pulsar el botón *Visualize*, número dos, que muestra la representación visual de los datos con la técnica de visualización escogida. De esta manera se puede visualizar el mismo conjunto de datos con todas las técnicas que desee el usuario. En caso de que se desee visualizar otro conjunto de datos distinto del seleccionado se pulsa al botón *Previous*, número tres, retornando a la vista de selección del archivo de datos de la Figura 26. El botón *Cancel*, número cuatro, cierra la ventana.

## Coordenadas Paralelas

En la siguiente figura se puede observar un ejemplo de la aplicación de la técnica Coordenadas Paralelas, que será utilizada para explicar las opciones de configuración que brinda la técnica.

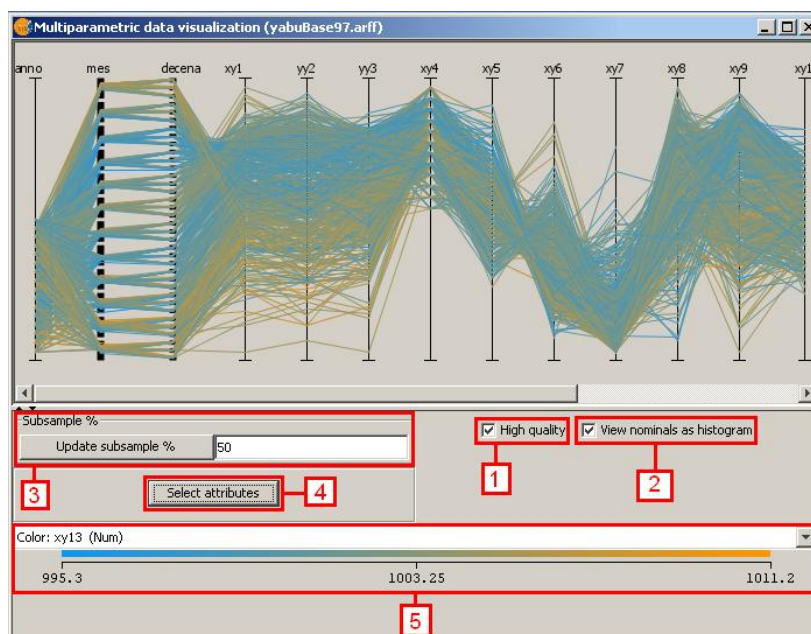


Figura 28: Visualización de un conjunto de datos utilizando la técnica Coordenadas Paralelas.

La implementación de esta técnica permite intercambiar los ejes que representan los atributos, posibilitando que puedan ser reorganizados de acuerdo a las necesidades del usuario. El procedimiento consiste en posicionar el puntero del “ratón” encima de un eje coordinado, hacer un clic izquierdo, trasladarse a otro eje y seleccionarlo mediante un clic; luego de realizar esta operación las barras se intercambian y con ello cambia el orden de los atributos en la imagen. A continuación se exponen las opciones enumeradas en la figura:

1. *High quality*: Habilita trazos mejor definidos de las líneas aumentando la calidad de la imagen.
2. *View nominals as histogram*: Las variables cuyo dominio de definición es un conjunto finito de valores pueden provocar superposiciones en la imagen de la visualización. Seleccionando esta opción el eje se divide en tantas barras como

valores tenga la variable que representa, haciendo corresponder cada barra a un valor específico del atributo.

3. *Update subsample %*: Esta opción permite obtener una muestra del conjunto de datos dado el porcentaje insertado en el campo de texto.
4. *Select attributes*: Al elegir esta opción se muestra el diálogo de la Figura 29 que permite la selección de los atributos que se deseen visualizar.
5. *Selección del color de las observaciones*: Esta opción permite asignarle el color a las observaciones con respecto a los valores de un atributo. Para ello se debe seleccionar un atributo y automáticamente se muestra su escala de colores, redibujándose las observaciones de acuerdo al valor que posee la dimensión en dicha observación.

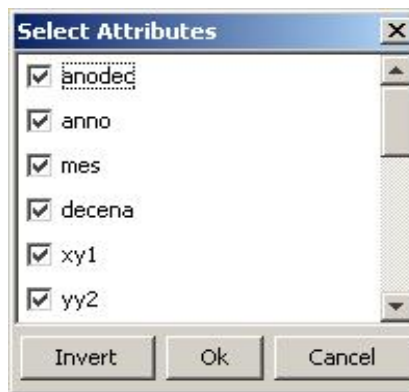


Figura 29: Diálogo de selección de atributos.

### Gráfico de Andrews

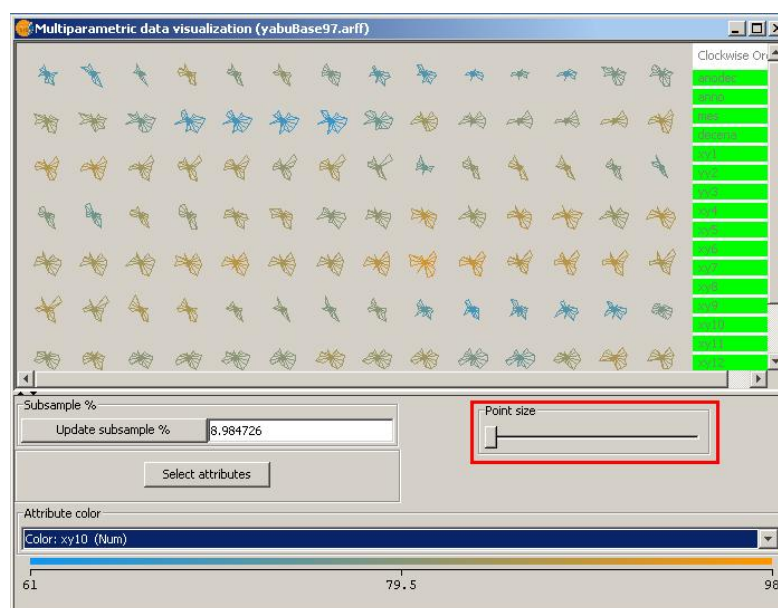
La técnica Gráfico de Andrews presenta, al igual que la anterior, las opciones de *Update Subsample %* y *Select Attributes*. Estas opciones pueden resultar en extremo útiles si el número de observaciones es grande o si existen demasiados atributos que no resultan de interés, ya que disminuye el desorden en la imagen.

La técnica permite definir el color de cada observación. El esquema es similar al utilizado en la técnica Coordenadas Paralelas anteriormente descrita, que se basa en mantener un

atributo para asignar el color de las observaciones mediante una escala de colores. La utilización de esta técnica se puede apreciar en el anexo 2.

### Técnicas basadas en iconos

La técnica basada en iconos implementada sigue el modelo de la técnica Campo de Estrellas. Se trata de visualizar los iconos en un espacio siguiendo un método de localización. En la actual implementación la localización de los iconos se realiza atendiendo al orden natural de los datos. La siguiente figura muestra la utilización de la técnica con iconos en forma de estrella.



**Figura 30: Icono en forma de estrella.**

De manera análoga a las técnicas anteriores, el usuario puede visualizar muestras de observaciones dado un porcentaje y seleccionar atributos. Permite aumentar el tamaño de los iconos mediante la barra de desplazamiento resaltada en color rojo. En el caso de los iconos en forma de estrella (Figura 30) e iconos en forma de barras (ver el anexo 3) los colores de los iconos, que representan las observaciones, pueden ser visualizados con un color de acuerdo a los valores de un atributo, de manera similar a la explicación brindada

para Coordenadas Paralelas. En el caso de shape coding (ver el anexo 4), esta operación no es posible puesto que los colores de cada celda dependen de los valores del atributo que representa; la opción *Attribute color* puede ser utilizada a manera de leyenda, se puede apreciar el rango de valores y la codificación de esos valores en una escala de colores para cada atributo.

### Segmentos de Círculo

La utilización de esta técnica puede apreciarse en la Figura 31. Esta visualización también brinda las opciones de seleccionar los atributos de interés para el usuario y la obtención de muestras del total de observaciones dado un porcentaje. Aunque esta técnica está especialmente diseñada para mostrar un gran volumen de datos, en caso de que se tenga un conjunto extremadamente grande de los mismos, sí se debe considerar la reducción de las observaciones. La cantidad de atributos suele influir mucho más en la expresividad de la representación, pues un número elevado de estos disminuye en gran medida el espacio de representación de cada atributo, en cuyo caso lo mejor es realizar una selección de los mismos.

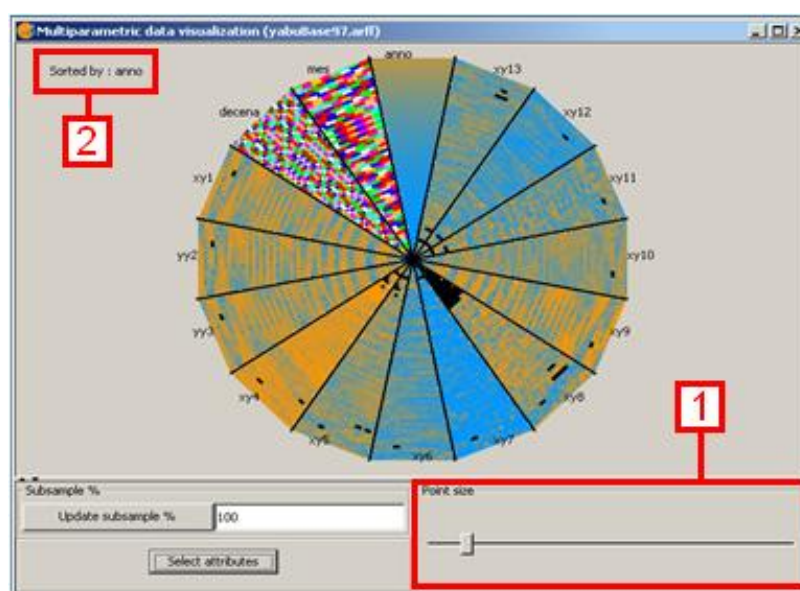


Figura 31: Técnica Segmentos de Círculo.

Otra posible interacción que se brinda es la elección del atributo mediante el cual se establece el orden de las observaciones. Por defecto las mismas aparecen ordenadas por el primer atributo que aparece en los datos, pero el usuario puede elegir cualquier otro con solo seleccionar el atributo deseado haciendo clic en la etiqueta correspondiente sobre la imagen, al realizar esta operación se visualiza el atributo seleccionado en la región resaltada en color rojo con numeración dos. Otro de los parámetros de interacción que se brinda es la barra de desplazamiento situada en la esquina inferior derecha con numeración uno, la cual permite al usuario controlar el tamaño de los puntos que representan los valores de cada variable en una observación.

### Patrones Recursivos

La utilización de la técnica Patrones Recursivos puede ser apreciada en la Figura 32. Esta técnica permite realizar las operaciones *Update Subsample %* y *Select Attributes* explicadas anteriormente.

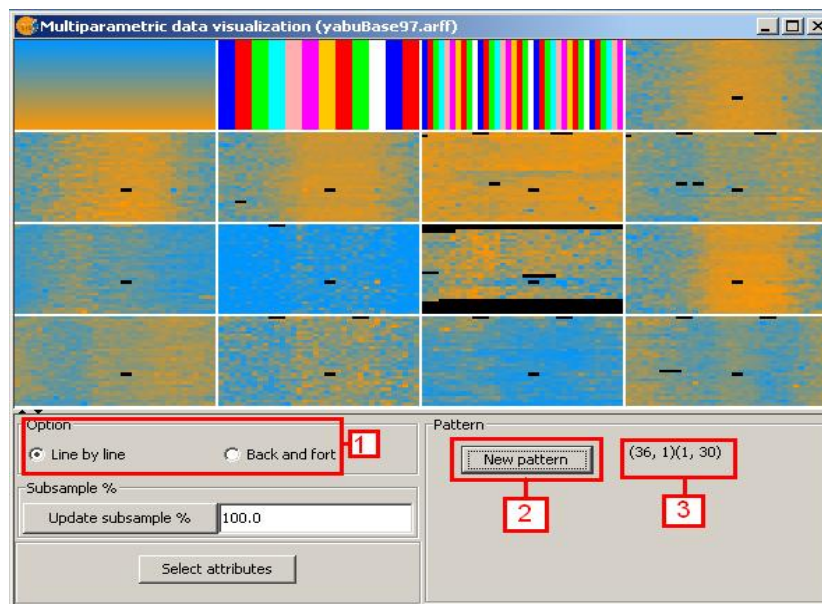


Figura 32: Visualización de un conjunto de datos mediante la técnica Patrones Recursivos.



La opción identificada con el número uno permite personalizar la forma de posicionar los píxeles en la región particular de cada atributo. Seleccionando *Line by line* los píxeles son posicionados línea a línea, de izquierda a derecha para cada nivel de recursividad; si se selecciona *Back and forth* los píxeles son posicionados de izquierda a derecha en una línea y en la siguiente de derecha a izquierda, repitiendo el procedimiento para cada nivel de recursividad. La opción dos es el botón *New pattern*, que al ser pulsado muestra el diálogo de la Figura 33 para la edición de los niveles de recursividad que conforman un nuevo patrón.

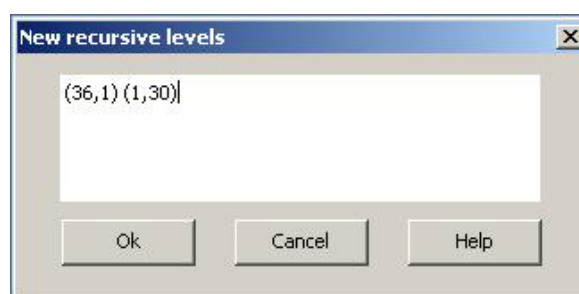


Figura 33: Diálogo para la edición de un nuevo patrón.

La forma de editar un nuevo patrón es definiendo los niveles de recursividad del mismo. Los niveles de recursividad se representan por pares de números enteros entre paréntesis y separados por coma; los niveles de recursividad se separan por un espacio entre los pares de valores.

Atendiendo a esta estructura suponga un ejemplo en que se tienen 120 observaciones de varias variables obtenidas mensualmente durante diez años; en este caso el orden de los datos lo proporcionaría la variable Mes o el orden natural de los datos si las observaciones poseen un orden cronológico. Se puede obtener el comportamiento anual, durante los diez años para todas las variables insertando el patrón (12,1) (1,10), donde el nivel de recursividad (12,1) corresponde a posicionar 12 mediciones de las variables en una línea y el nivel (1,10) repetir el procedimiento anterior 10 veces de arriba hacia abajo, por lo tanto se estarían representando las  $12 \times 10 = 120$  observaciones de las variables. Es responsabilidad del usuario decidir que patrón utilizar, para ello debe considerar la cantidad de



observaciones que se deseen visualizar, el orden natural de los datos aunque se puede utilizar una variable que represente un orden temporal y todo el conocimiento que pueda aportar el usuario sobre los datos.

El patrón que está siendo utilizado por la visualización se muestra en el campo de texto señalado con el número tres en la Figura 32.

### 3.1.2 Visualización coordinada de varios conjuntos de datos

El usuario del sistema tiene la posibilidad de realizar el análisis visual de varios conjuntos de datos de forma coordinada, teniendo una percepción del origen de los datos que están siendo visualizados referenciando las técnicas a un mapa. Para realizar esta operación es necesario tener configurado un proyecto de visualización de este tipo que contiene: uno o dos mapas en formato *shape* (*shp*) de *ESRI* y los archivos de datos asociados a las localizaciones puntuales en el mapa. Este tipo de proyectos se almacenan en archivos con extensión *syp* y su configuración se realiza utilizando una herramienta de esta misma extensión.

#### 3.1.2.1 Configuración de un proyecto de visualización coordinada

La configuración de un proyecto de visualización científica puede ser realizada accediendo a las opciones que se muestran en la figura siguiente.

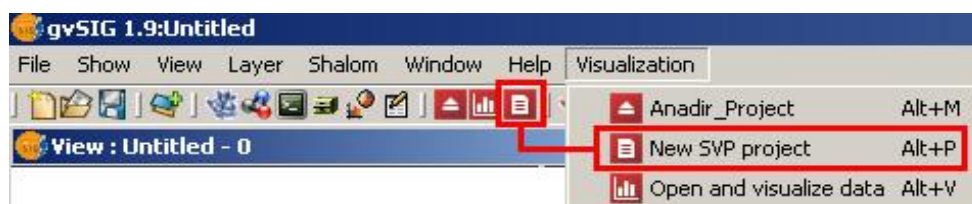


Figura 34: Opciones para la configuración de un nuevo proyecto de visualización coordinada.

Esta operación se puede realizar a través de un botón en la barra de herramientas o accediendo a la opción de menú *Visualization/New SVP project*, como se muestra en la Figura 34. Al pulsar una de las opciones antes descritas se muestra un asistente para la configuración del proyecto. El asistente permite realizar la descripción del proyecto que se desea crear, para conocimiento del usuario en caso de que desee utilizar en otro momento un proyecto creado con anterioridad. Una vez realizada la descripción del proyecto (opcional) el usuario puede seleccionar el directorio donde guardar el proyecto así como el nombre del archivo con extensión *svp* donde se guardará la configuración del mismo. A continuación se selecciona el mapa que servirá como base o fondo a la visualización y se procede a seleccionar el mapa que será utilizado para las localizaciones, que puede ser el mapa base o un mapa con las localizaciones exactas del origen de los datos que tenga total correspondencia con el mapa base. Realizada la anterior operación, el usuario define el campo de la tabla asociada al mapa de las localizaciones por el cual referenciará los archivos con los datos; a continuación el usuario puede seleccionar la localización en el mapa y el archivo de datos que desea asociar a la localización, realizando tantas inserciones como localizaciones tenga el mapa.

Cada archivo de datos asociado a una localización será visualizado en el centroide de la geometría correspondiente. El asistente controla además que los ficheros con los datos no sean utilizados en más de una localización. En el anexo 5 se muestran ejemplos de ficheros de configuración *svp*.

Una explicación más detallada del proceso de configuración de un proyecto de visualización coordinada se puede apreciar en el anexo 6.

### **3.1.2.2 Utilización de un proyecto de visualización coordinada**

Cuando se dispone de un proyecto previamente creado el primer paso para realizar la visualización lo constituye crear una vista de visualización.

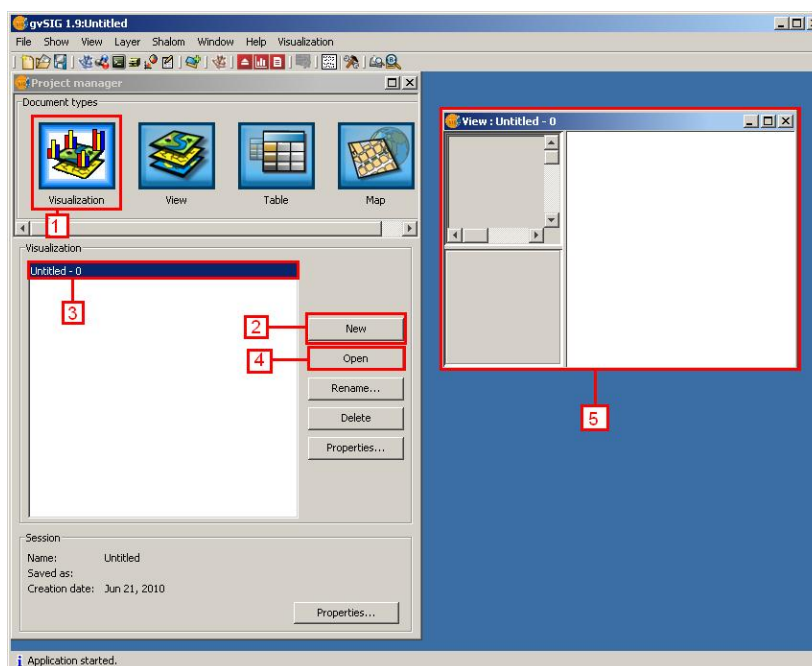


Figura 35: Creación de una vista de visualización coordinada.

La Figura 35 muestra los pasos para la creación de una vista de visualización coordinada, los cuales son: seleccionar el tipo de documento de visualización, identificado en la figura con el número uno, crear un nuevo documento de visualización oprimiendo el botón *New* (número dos), seleccionar el documento creado (número tres) y abrirlo oprimiendo el botón *Open* con numeración cuatro. Realizados estos pasos se muestra una vista como la resaltada con el número cinco.

Al estar activa una vista de visualización se habilita una opción de menú *Visualization/Load SVP project* y un botón en la barra de herramientas, como lo muestra la Figura 36.

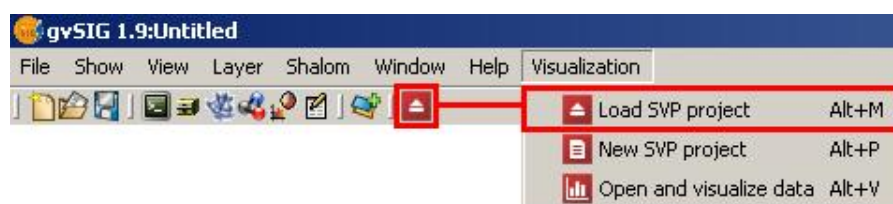
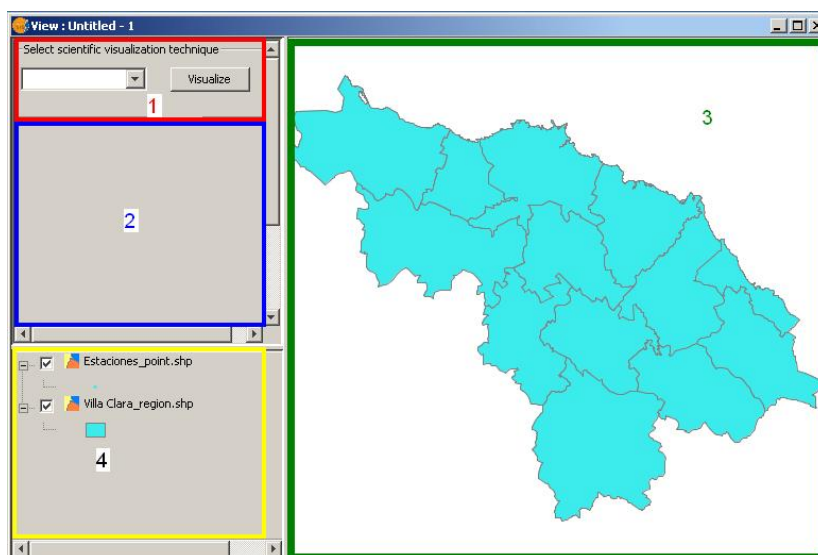


Figura 36: Opciones para la adición de un proyecto de visualización coordinada a una vista de visualización.

Cuando se selecciona una de las opciones anteriores, se muestra un diálogo para la selección del fichero *svp*, que constituye el proyecto de visualización coordinada con el que se desea trabajar. Una vez seleccionado el archivo, se carga el proyecto en la vista de visualización activa y automáticamente se cargan los mapas que estén en el archivo de configuración.

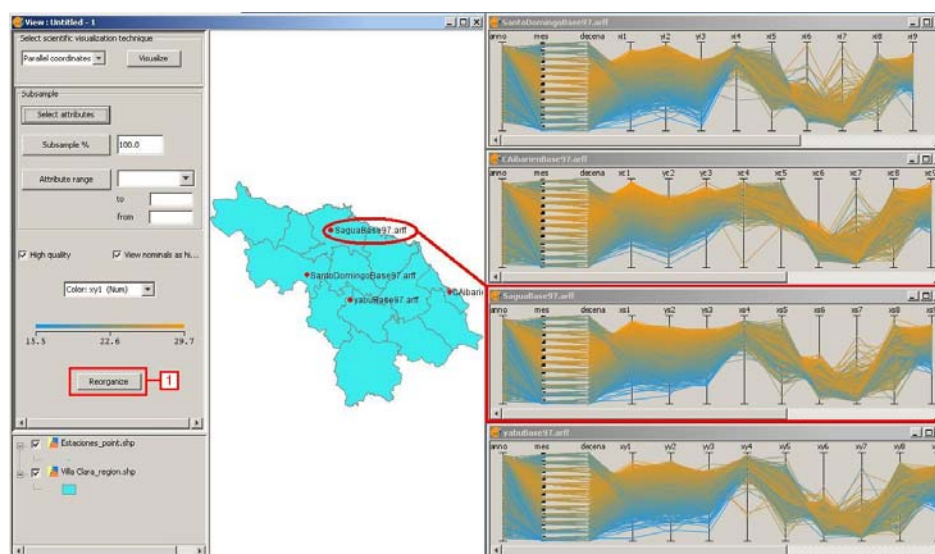


**Figura 37: Vista de visualización.**

La Figura 37 muestra los diferentes componentes que conforman una vista de visualización. Resaltado en color rojo se muestra la opción de selección de la técnica de visualización a utilizar, esta opción consta de dos componentes: un *combobox* para la selección de las técnicas de visualización y el botón *Visualize* para realizar la visualización de la técnica seleccionada. Al oprimir el botón *Visualize* se muestra en el área resaltada con color azul el panel de configuración de la técnica con que se está realizando la visualización de los datos. La región resaltada en color verde constituye el área de visualización de los mapas y de las técnicas que pueden ser visualizadas sobre los mapas. El componente resaltado con color amarillo muestra los mapas con los que cuenta el proyecto, además permite algunas configuraciones como son cambiar el color de fondo y la simbología de los mapas.

## Coordenadas Paralelas

En el capítulo anterior se mencionó la posibilidad de que algunas técnicas fueran visualizadas sobre el mapa y otras no, teniendo en cuenta las características particulares de cada una. La técnica Coordenadas Paralelas, accesible desde la opción de selección de las técnicas, no es visualizada sobre el mapa. La visualización de cada conjunto de datos se realiza en paneles independientes y se agrega el nombre del archivo de datos a la localización que le corresponde en el mapa. La siguiente figura muestra un ejemplo de lo anteriormente expuesto.



**Figura 38: Visualización coordinada utilizando Coordenadas Paralelas.**

Cada panel muestra la representación visual coordinada de un conjunto de datos. El nombre del archivo de datos representado se refleja en la barra de título del panel y a su vez en la localización que le corresponde en el mapa.

Un elemento común en todas las técnicas, independientemente de que la visualización sea en paneles independientes o sobre el mapa, es la opción de obtener muestras del conjunto de datos. Ver la Figura 39.

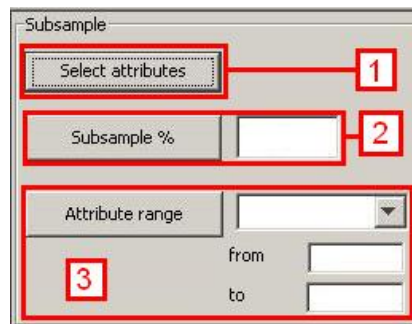


Figura 39: Obtención de muestras del conjunto de datos.

Mediante esta opción se pueden realizar las siguientes opciones:

1. Botón *Select attributes*: Selección de los atributos de interés para la visualización.
2. Botón *Subsample %*: Obtención de una muestra para cada conjunto de datos dado el porcentaje insertado en el campo de texto.
3. Botón *Attribute range*: Obtención de una muestra para cada conjunto de datos dado el rango de un atributo; el atributo se selecciona en el *combobox* y el rango de valores se inserta en los campos etiquetados *from* y *to*.

La técnica Coordenadas Paralelas en este tipo de visualización posee además las opciones *View nominals as histogram* y *High quality* descritas en el subepígrafe 3.1.1. También permite asignar los colores a las observaciones según los valores de un atributo, esto se logra mediante un mapa de colores, opción también descrita en el subepígrafe antes mencionado. En la Figura 38, con numeración uno, se muestra la opción *Reorganize*, que muestra en diálogo donde el usuario define el orden en que desea que los atributos sean visualizados (ver la Figura 40).

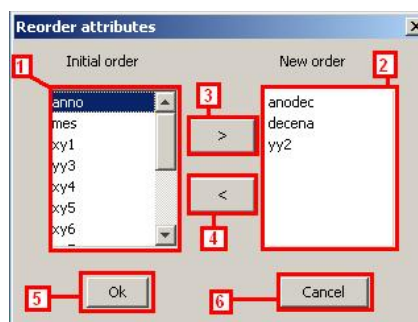


Figura 40: Diálogo para establecer un nuevo orden de visualización de los atributos.

El diálogo para establecer el orden de los atributos cuenta con una lista con el orden inicial de los atributos, identificada con el número uno; una lista con el orden final deseado, identificada con el número dos; botones para la inserción de los atributos de una lista hacia otra (tres y cuatro); botón *Ok* identificado con el número cinco para aceptar el nuevo orden establecido y el botón *Cancel* para cancelar las operaciones realizadas en esa ventana. El nuevo orden debe estar definido completamente para poder aceptar la selección, por lo tanto todos los atributos listados en la lista uno deben ser trasladados a la lista dos. Todas las operaciones realizadas en el panel de configuración afecta la visualización de todos los conjuntos de datos simultáneamente.

### Gráfico de Andrews

La técnica Gráfico de Andrews (ver la Figura 41) puede ser utilizada accediendo a la opción de selección de las técnicas. Al igual que en Coordenadas Paralelas, las visualizaciones son realizadas en paneles independientes. Permite la obtención de muestras de los conjuntos de datos con las operaciones *Select attributes*, *Subsample %* y *Attribute range*, reorganizar los atributos que serán visualizados y asignar colores a las observaciones.

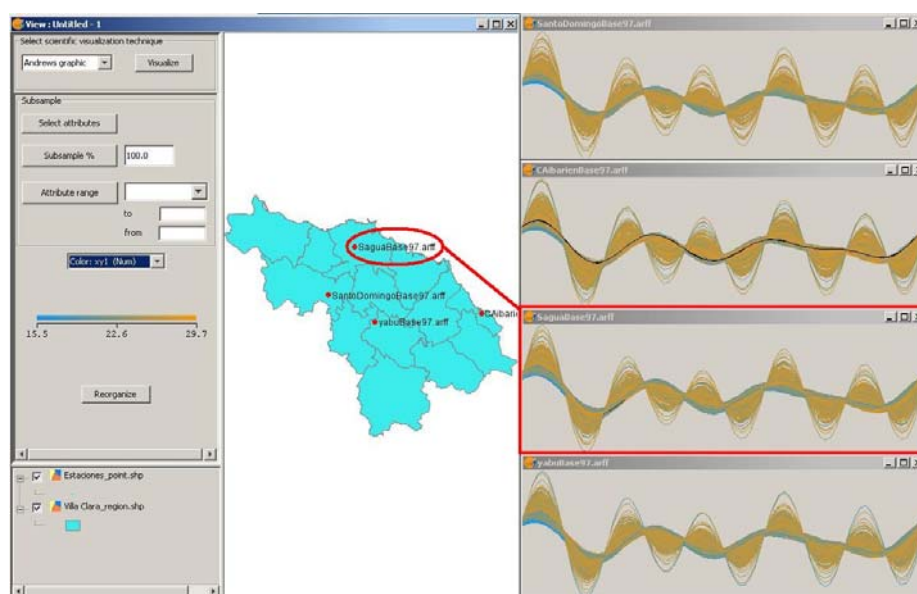
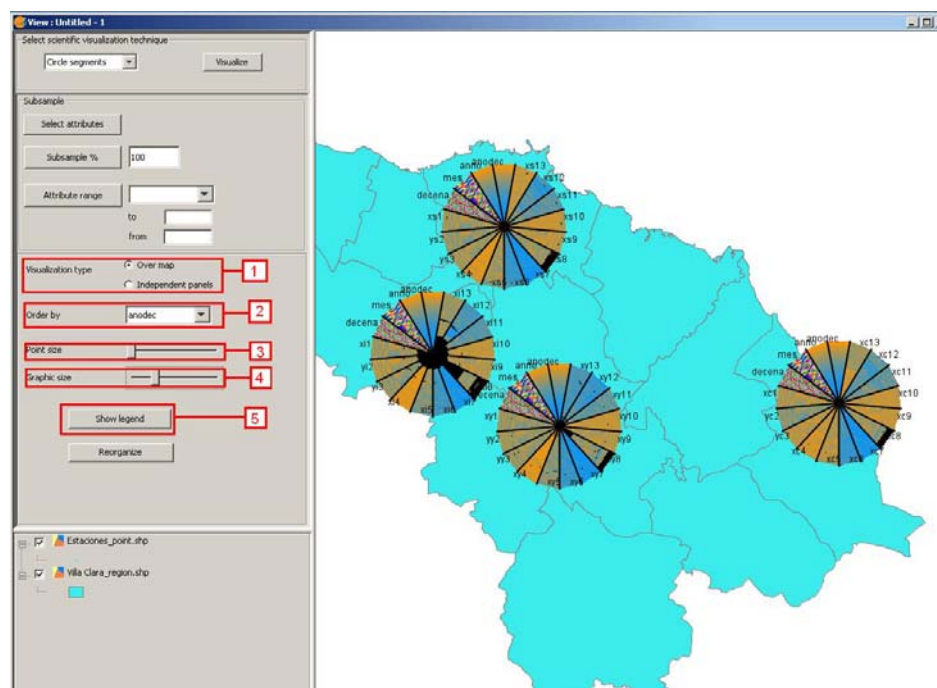


Figura 41: Visualización coordinada utilizando Gráfico de Andrews.

## Segmentos de Círculo

La técnica Segmentos de Círculo se encuentra disponible en la opción de selección de las técnicas de visualización. Al igual que las técnicas anteriormente expuestas permite obtener muestras de los conjuntos de datos y la reorganización de los atributos. La visualización de esta técnica se puede realizar sobre el mapa, obsérvese la Figura 42.



**Figura 42: Visualización coordinada utilizando la técnica Segmentos de Círculo.**

A continuación se enumeran las opciones disponibles en el panel de configuración de la técnica:

1. *Visualization type*: Permite seleccionar cómo se realizará la visualización, es decir si es en paneles independientes o sobre el mapa. La opción seleccionada por defecto es sobre el mapa.
2. *Order by*: Esta opción permite que las observaciones de los conjuntos de datos sean ordenadas por los valores de un atributo que sea seleccionado.



3. *Point size*: Aumentar o disminuir el tamaño de los puntos que representan los valores de las variables para cada observación. Esta opción resulta útil cuando la cantidad de observaciones es pequeña.
4. *Graphic size*: Permite controlar el tamaño del gráfico con el objetivo de reducir el posible solapamiento de las imágenes sobre el mapa.
5. *Show legend*: Muestra un diálogo para la leyenda. La leyenda muestra la escala en colores de los valores de los atributos. Ver la Figura 43.



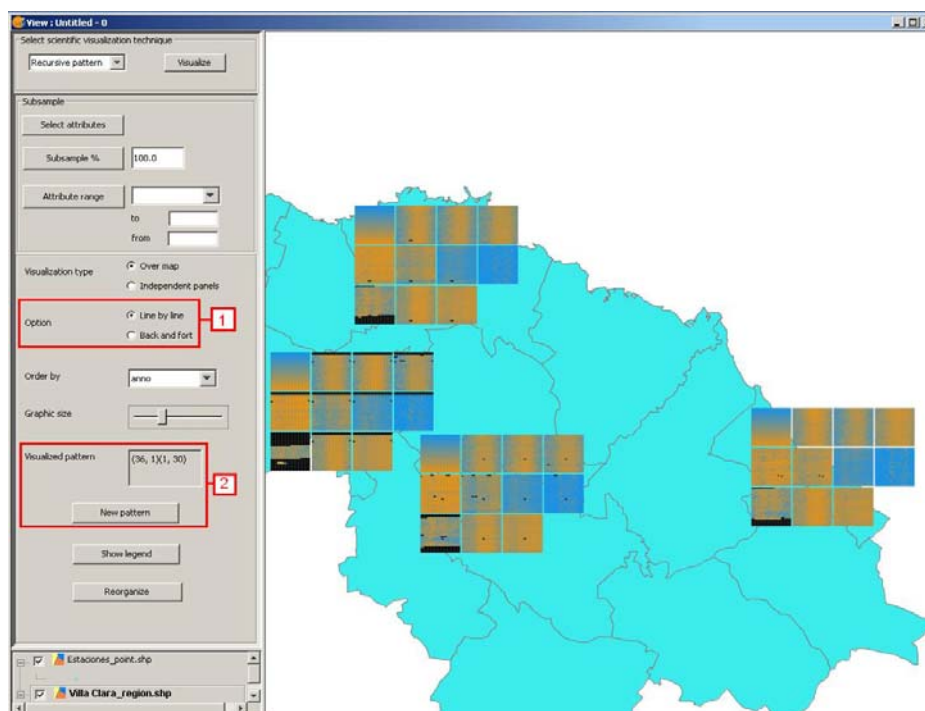
Figura 43: Leyenda de los valores de los atributos.

## Patrones Recursivos

Otra de las técnicas disponibles en la opción de selección de las técnicas de visualización es Patrones Recursivos. Esta técnica incluye las funcionalidades: visualización de muestras de observaciones de los conjuntos de datos, reorganización de atributos, aumentar y disminuir el tamaño de los gráficos generados por la técnica, ordenar las observaciones según los valores de un atributo e intercambiar entre la visualización en paneles independientes y la visualización sobre al mapa. Además permite mostrar la leyenda de los valores de las variables con el mapa de colores. Las descripciones de estas funcionalidades son similares a las explicadas en las técnicas anteriores. La utilización de esta técnica se muestra en la Figura 44, que será utilizada para exponer las opciones de interacción.

La opción identificada con el número uno permite personalizar la forma de posicionar los píxeles en la región particular de cada atributo, las posibles opciones son *Line by line* o

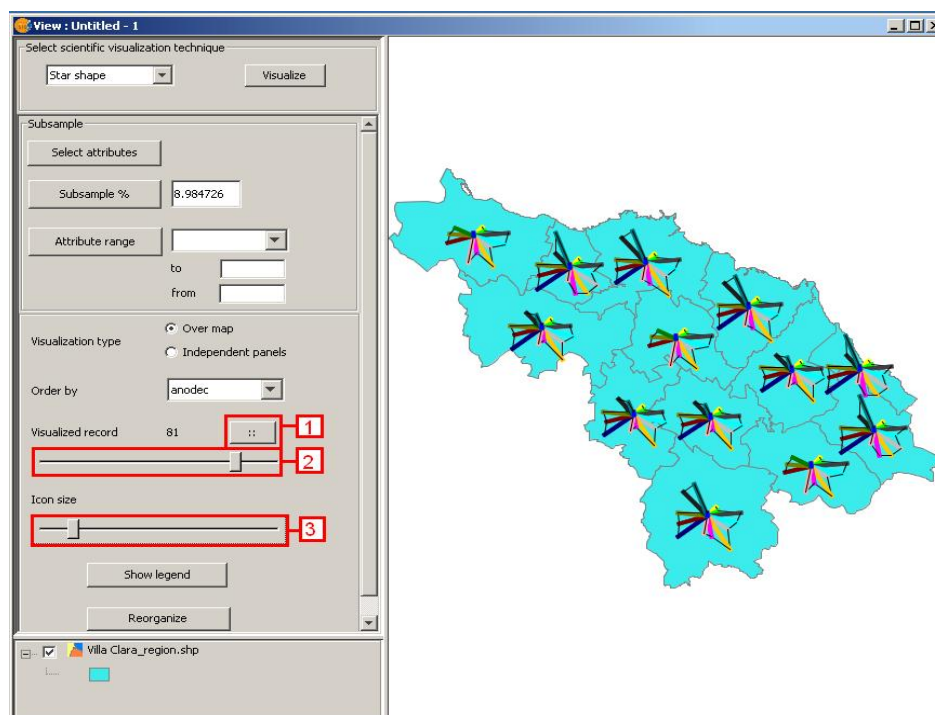
*Back and forth.* Además el usuario puede insertar nuevos patrones mediante la opción dos de la figura. Al pulsar el botón *New pattern* se muestra el diálogo de la Figura 33 para la edición de los niveles de recursividad que conforman el nuevo patrón.



**Figura 44:** Visualización coordinada utilizando la técnica patrones recursivos.

## Técnicas basadas en iconos

Las técnicas basadas en iconos también están disponibles en la opción de selección para la visualización coordinada. Estas técnicas muestran un icono a la vez, según la observación que esté siendo visualizada. Ordenando las observaciones por una variable que denote tiempo o un orden cronológico determinado en las mediciones, el usuario puede realizar un análisis temporal del comportamiento de las variables. Los iconos implementados para este propósito son: icono en forma de estrella, icono en forma de barras y shape coding. La figura siguiente muestra la visualización utilizando el icono en forma de estrella, que será utilizada para explicar las opciones específicas que brinda la técnica. Un ejemplo de visualización empleando los iconos en forma de barras y shape coding se muestra en los anexos 7 y 8, respectivamente.



**Figura 45: Visualización coordinada mediante iconos en forma de estrella. El ejemplo muestra la visualización de los iconos localizados en los centroides de los polígonos que representan a los municipios de Villa Clara.**

Estas técnicas poseen el mismo panel de configuración por lo tanto la explicación es válida para los restantes iconos. Al igual que en las técnicas anteriores el usuario puede obtener muestras del conjunto de datos, ya sea mediante un porcentaje, el rango de valores de un atributo o la selección de atributos. Puede intercambiar entre la visualización en paneles independientes y la visualización sobre el mapa, ordenar las observaciones según los valores de un atributo y reordenar los atributos.

La opción de la Figura 45 con numeración uno permite animar el proceso de visualización de los registros mediante un temporizador. Pulsando este botón se muestra un diálogo como el de la Figura 46 que permite iniciar, detener, acelerar y desacelerar el proceso de visualización de los registros.

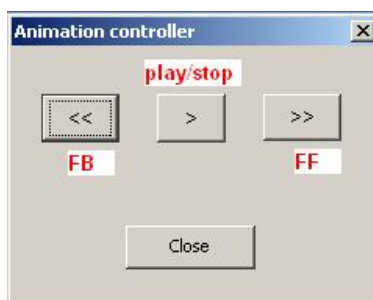


Figura 46: Diálogo para el control de la animación.

La opción dos es una barra deslizante para la selección del registro que se desee visualizar, en principio realiza la misma operación que la animación pero de forma manual. La opción tres permite controlar el tamaño del icono para evitar el solapamiento de los gráficos en la imagen.

La leyenda para el caso de los iconos en forma de estrella y en forma de barras muestra la codificación en colores de los atributos, esto se utiliza para identificar cada atributo según el color que se muestra en la leyenda (ver la Figura 47).



Figura 47: Leyenda.

### 3.2 Caso de estudio: Análisis visual de los datos meteorológicos de la provincia Villa Clara

El presente epígrafe tiene como objetivo presentar un caso de estudio donde se realiza el análisis visual de los datos meteorológicos de la provincia Villa Clara. Para ello se ha

dividido en dos subepígrafes. En el primero se describen los datos meteorológicos utilizados para el caso de estudio y en el segundo se muestran algunos resultados obtenidos al aplicar las técnicas de visualización a los datos meteorológicos.

### 3.2.1 Datos meteorológicos de la provincia Villa Clara

Los datos meteorológicos recopilados por el Instituto de Meteorología de la provincia Villa Clara representan una serie temporal de trece variables, medidas desde 1977 hasta la actualidad. Las variables meteorológicas recopiladas son: temperatura media, mínima y máxima promedio en la decena, humedad relativa media, mínima, y máxima promedio en la decena, déficit de saturación (promedio en grados en la decena), nubosidad (promedio en octavos en la decena), velocidad media del viento (promedio en la decena), lluvia total en la decena, insolación (promedio de horas luz en la decena), tensión de vapor de agua (promedio en la decena) y presión atmosférica (promedio en la decena). Para cada una de las cuatro estaciones meteorológicas de la provincia se cuenta con una serie temporal de las trece variables mencionadas. Las estaciones meteorológicas están localizadas en Yabú, Santo Domingo, Sagua la Grande y Caibarién.

Año	Mes	Año - Decena	Temp	Hum Rel	...
1977	1	197701	20	73	...
1977	1	197702	18	77	...
1977	2	197705	23	83	...
1978	...	...	...	...	...

**Tabla I: Estructura general de los datos meteorológicos.**

Para el almacenamiento de los datos meteorológicos se han utilizado varios formatos como bases de datos, archivos en formato de texto (*arff*) y en formato de tablas (*dbf*). La Tabla I muestra un fragmento de la forma en que se almacenan estos datos. Cada registro representa el promedio de los valores en diez días para cada una de las variables meteorológicas. Estos registros se tienen para cada una de las estaciones. Los datos

originales fueron procesados mediante métodos estadísticos y de limpieza de datos para corregirlos y eliminarles ruido.

Como se puede observar existen varios rasgos relacionados con el tiempo (año, mes, decena y año-decena) por lo que se puede lograr diferentes niveles de granularidad respecto al componente temporal.

Conjuntamente se posee la información cartográfica que incluye un mapa con la división político-administrativa en municipios de la provincia Villa Clara y un mapa de puntos con las coordenadas de las localizaciones de las estaciones meteorológicas. La Tabla II muestra las coordenadas geográficas de las estaciones.

No.	Estación	Latitud	Longitud	Altura
1	Yabú 343	79.991	22.461	116.44
2	Sagua 338	80.092	22.806	12,06
3	Caibarién 348	79.471	22.497	46,27
4	Sto Domingo 326	80.226	22.586	45,35

Tabla II: Coordenadas geográficas de las estaciones meteorológicas.

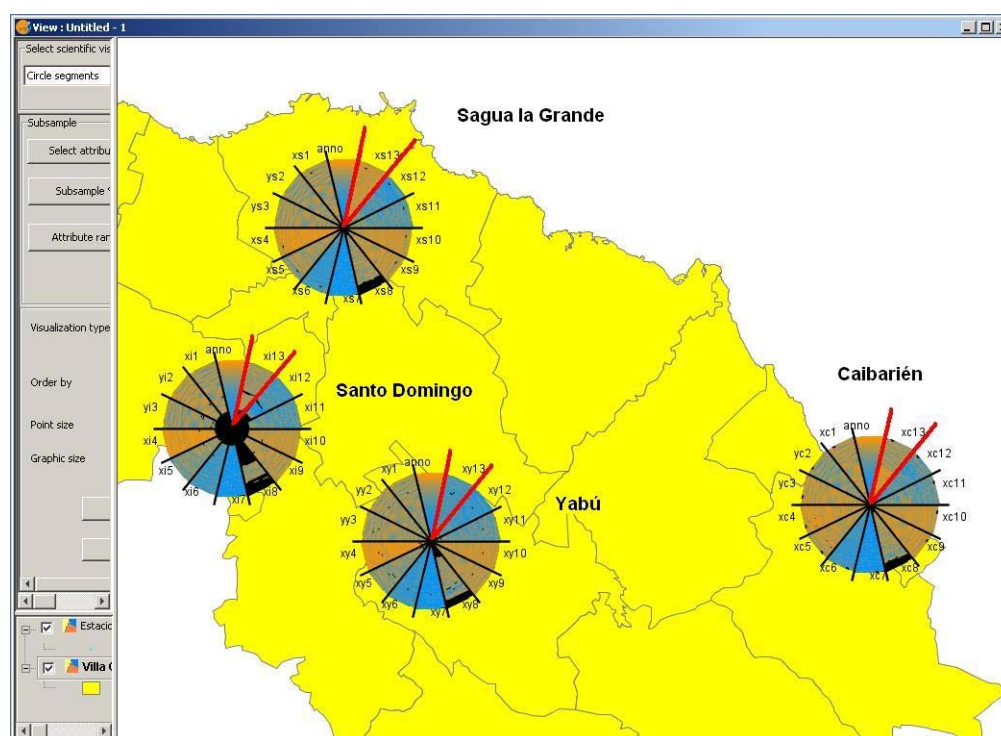
Mediante el asistente de configuración de proyectos de visualización coordinada se configuró un proyecto en el cual se tomó como mapa base el mapa de la provincia Villa Clara y para las localizaciones se empleó un mapa de puntos con las localizaciones de las estaciones meteorológicas, haciendo corresponder los archivos de datos multiparamétricos a las localizaciones.

Otra vía sería realizar la visualización de cada conjunto de datos por separado mediante la extensión para la visualización independiente. En ese caso el análisis sería localizado para cada estación meteorológica pues la visualización depende únicamente de los valores del conjunto de datos que utiliza.

### 3.2.2 Visualización de los datos meteorológicos

En el primer capítulo se mostró que el análisis visual de datos sigue varios objetivos. En el caso del análisis de los datos meteorológicos de la provincia Villa Clara se persiguen dos objetivos: explorar visualmente los datos y arribar a alguna hipótesis o confirmar alguna que se tenga del comportamiento de las variables en algún período de tiempo.

La siguiente figura muestra la visualización de los datos mediante la técnica de visualización Segmentos de Círculo.

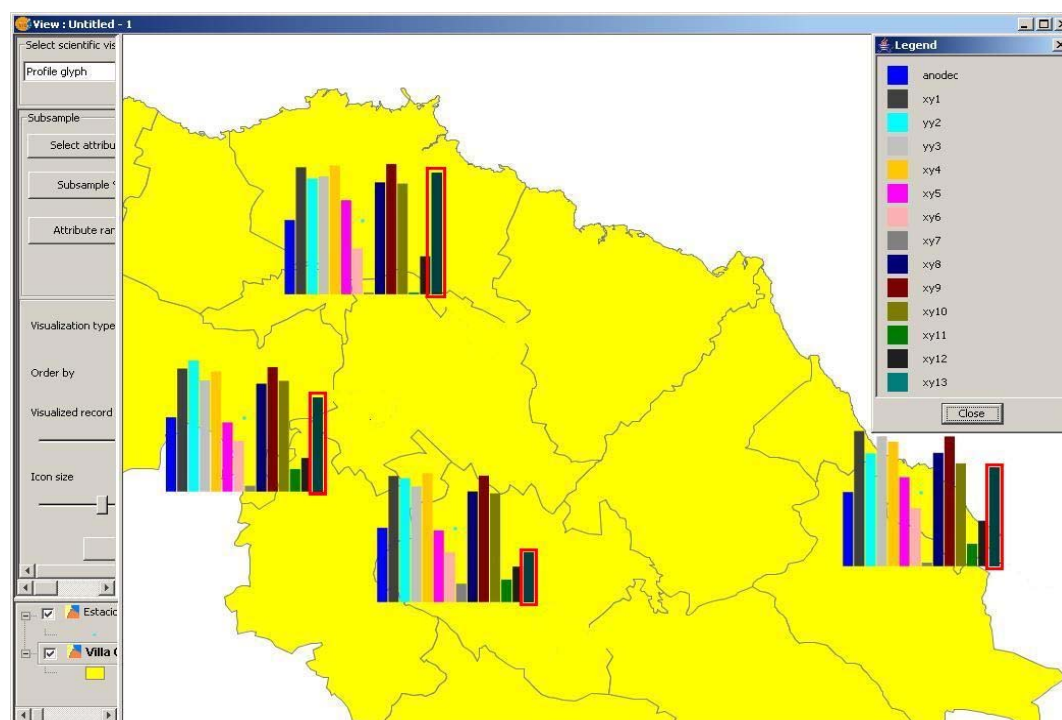


**Figura 48: Visualización coordinada mediante la técnica Segmentos de Círculo. Se muestra la representación visual de los 1113 registros de las variables meteorológicas.**

La figura muestra la visualización de los 1113 registros que conforman los datos de las estaciones meteorológicas. Los datos se muestran ordenados por la variable año, desde el centro de la circunferencia hacia afuera. Los valores en negro representan los valores perdidos de las variables. El color azul representa los valores mínimos y el anaranjado intenso los valores máximos para cada una de las variables.

Resaltado en color rojo se muestra el sector de la circunferencia correspondiente a la variable presión atmosférica. Se pueden apreciar tonalidades azules en la mayor parte de los valores de esta variable en la estación del Yabú, por lo tanto la presión atmosférica siempre es más baja en esta estación y sus valores son mínimos con respecto a los valores de esa variable para las restantes estaciones, donde se pueden apreciar tonalidades naranjas. De igual manera se pueden percibir tonalidades naranjas en esa variable para la mayoría de los datos en la estación de Sagua la Grande. En el caso de la estación de Caibarién, la variable presión atmosférica presenta un cambio de lo que se pudiera estimar que representan los valores luego de transcurridos los primeros diez años; se puede apreciar un cambio de tonalidad naranja intenso a tonalidades azules y naranja claro lo cual evidencia cambios en los valores de esa variable en cierto período de tiempo.

En la siguiente figura se evidencia este comportamiento de la variable presión atmosférica empleando iconos en forma de barras.



**Figura 49: Visualización de un registro de los datos de las cuatro estaciones meteorológicas.**



La variable está codificada en la barra resaltada con color rojo. Se puede apreciar las diferencias en el tamaño de las barras. El comportamiento cambiante de las variables al utilizar esta técnica puede ser percibido con mayor claridad si se emplea la animación.

La Figura 50 muestra la representación visual de todo el volumen de datos de las cuatro estaciones mediante la técnica Coordenadas Paralelas. El orden de arriba hacia abajo de las representaciones visuales corresponden a las estaciones de Santo Domingo, Caibarién, Sagua la Grande y Yabú.

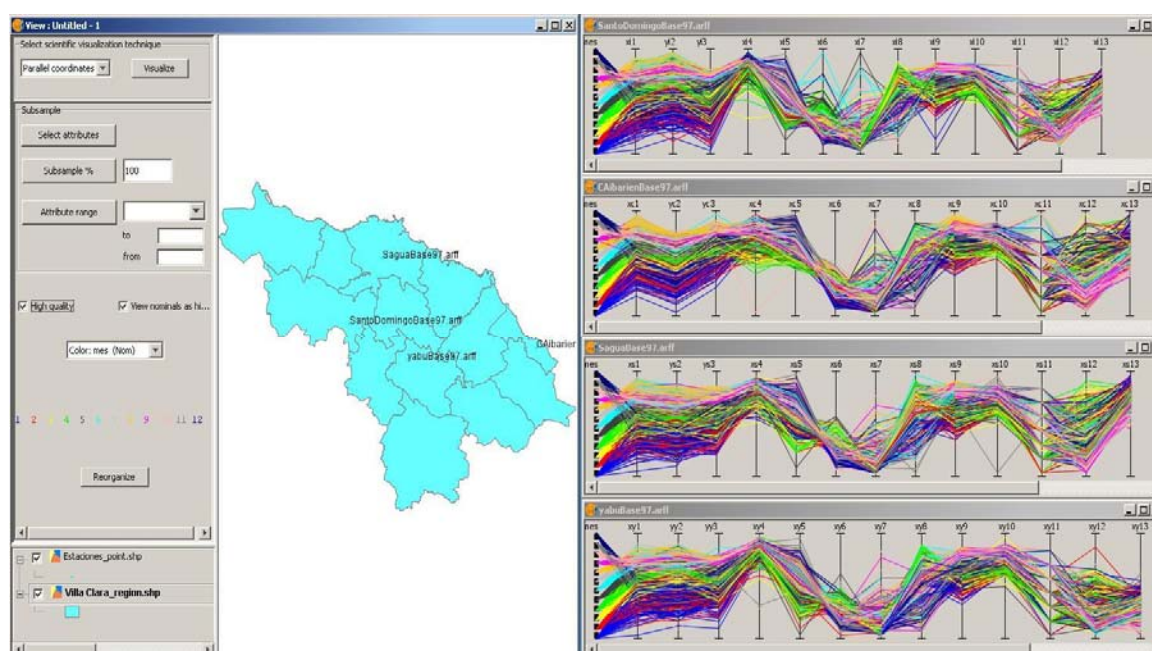
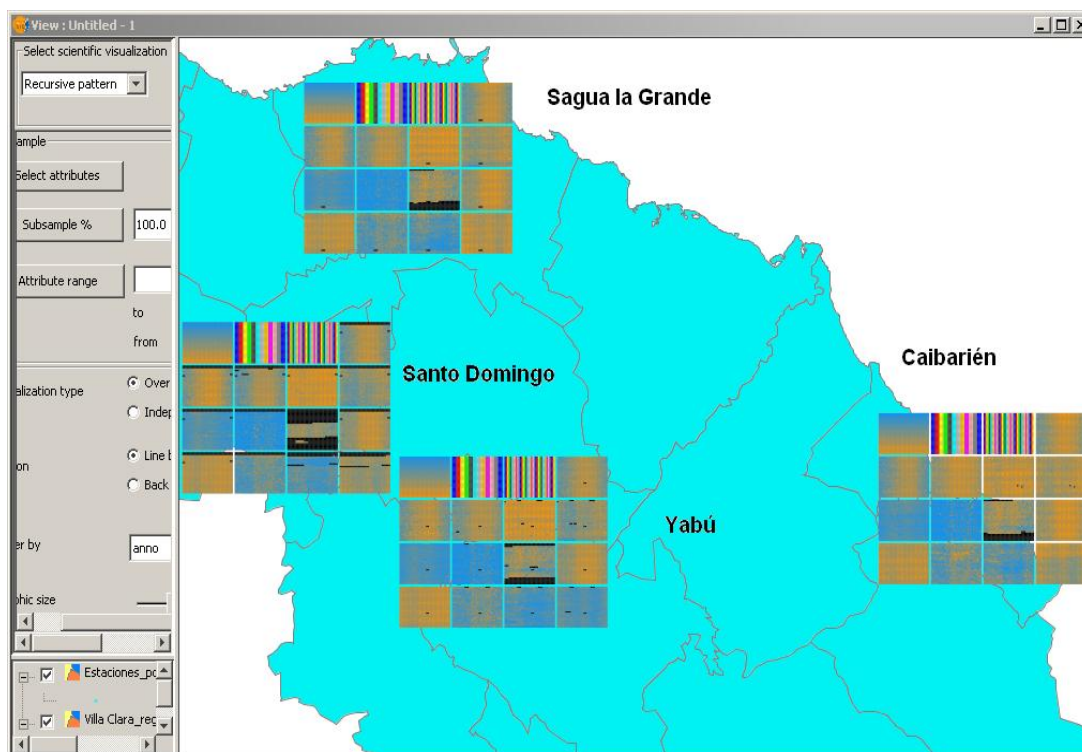


Figura 50: Visualización coordinada empleando la técnica Coordenadas Paralelas.

El color de las observaciones está asignado con respecto a la variable mes. Mediante este procedimiento se puede apreciar el comportamiento de las variables asociado a los valores del mes. Se puede apreciar cómo las observaciones con tonalidad azul oscuro, correspondiente al mes de Enero, se concentran en la parte inferior de la segunda, la tercera y cuarta barra de izquierda a derecha (temperatura media, máxima y mínima en la decena) que corresponden a los valores mínimos de esas variables, conclusión que está acorde con que los valores mínimos de temperatura en nuestro país corresponden a los primeros y últimos meses del año.

La siguiente figura muestra el análisis visual de los datos meteorológicos empleando la técnica Patrones Recursivos.



**Figura 51:** Visualización de las 1080 observaciones de las variables correspondientes a las cuatro estaciones meteorológicas de Villa Clara con la técnica de visualización Patrones Recursivos.

El patrón utilizado en la visualización es (36,1) (1,30). Las mediciones de las variables constituyen los valores promedio en un período de 10 días llamado decena, por lo tanto un mes posee tres decenas y un año 36. El nivel de recursividad (36,1) del patrón empleado representa un año de observaciones de las variables, significa tomar 36 valores de una variable y posicionar los píxeles que representa a cada uno en una línea de izquierda a derecha. El nivel de recursividad (1,30) significa reproducir 30 veces el nivel anterior, de arriba hacia abajo, por lo tanto se están representando 30 años de comportamiento de las variables, visualizándose los  $36 \times 30 = 1080$  registros de cada variable meteorológica. En este caso se está siguiendo como método de posicionamiento de los píxeles en la imagen la opción línea a línea, en inglés *Line by line*.

Se puede apreciar ciertos comportamientos en la temperatura media, máxima y mínima (cuarto, quinto y sexto cuadrados, de izquierda a derecha, de arriba hacia abajo); las tonalidades menos naranjas y más azuladas, de los extremos izquierdo y derecho de los cuadrados, evidencian los valores mínimos de estas variables que corresponden a los meses fríos del año; las tonalidades naranjas más intensas hacia el centro representan los valores máximos, que corresponden a los meses calurosos del año. Los valores perdidos se pueden apreciar en color negro, que corresponden a mediciones no realizadas en las estaciones meteorológicas.

Fue analizado el período desde 1996 hasta 1999 con el objetivo de identificar los efectos del fenómeno climatológico “El Niño”, cuya mayor afectación tuvo lugar en este rango de tiempo. Para lograr la representación de este período se utilizó el patrón (36,1) (1,4). La Figura 52 muestra los valores la temperatura mínima, media y máxima de los años analizados pertenecientes a la estación meteorológica de Santo Domingo.

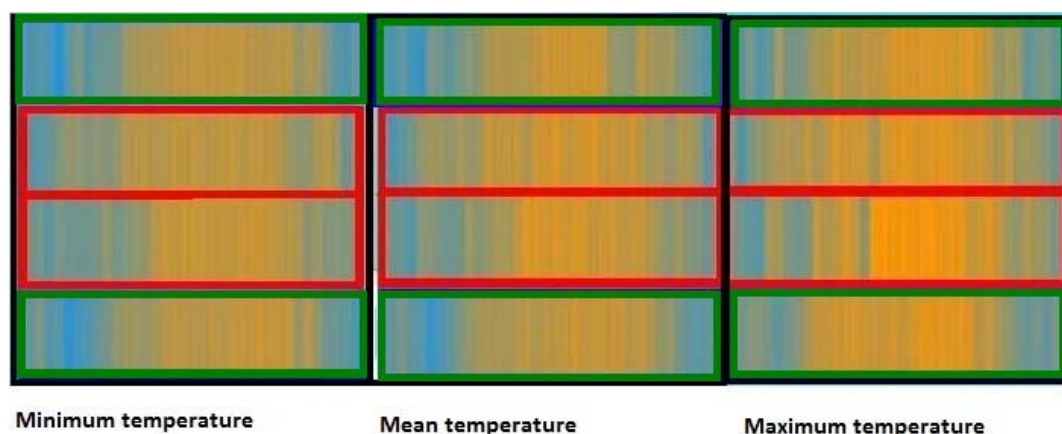


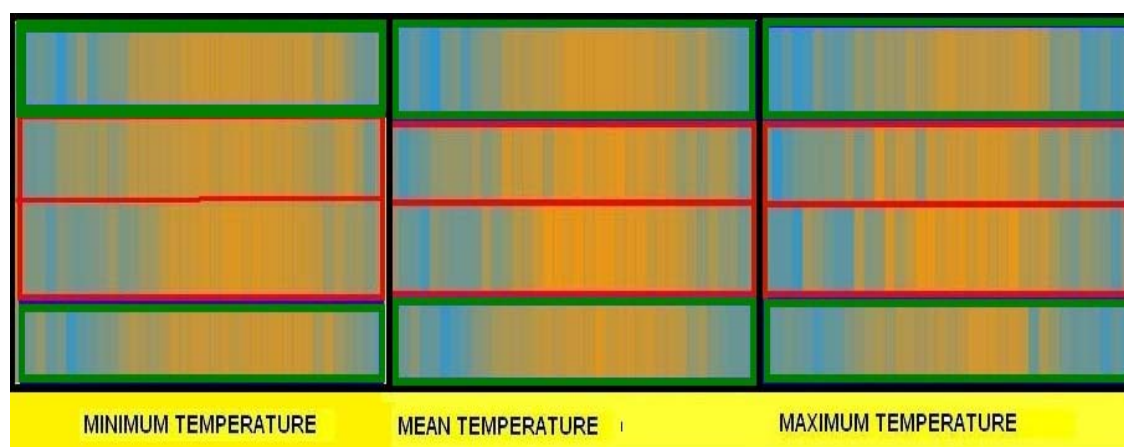
Figura 52: Visualización de la temperatura mínima, media y máxima desde 1996 hasta 1999, utilizando el patrón (36,1) (1,4). La visualización corresponde a la estación meteorológica de Santo Domingo.

Cada rectángulo representa un atributo cuyos valores mínimos están codificados en color azul y los valores máximos en color naranja intenso. Los rectángulos internos resaltados en color verde y rojo representan el período de años 1996-1999, de arriba hacia abajo. Los que están resaltados en color rojo representan los años 1997 y 1998, siguiendo el orden antes descrito.

De la figura se puede deducir que:

- La temperatura media y máxima es mayor en los años 1997 y 1998, pues se puede apreciar una coloración naranja más intensa.
- La temperatura mínima, media y máxima aumentó en los primeros meses del año 1997.
- En el año 1998 la temperatura media y máxima fue mayor durante los meses cálidos con respecto a los restantes años.

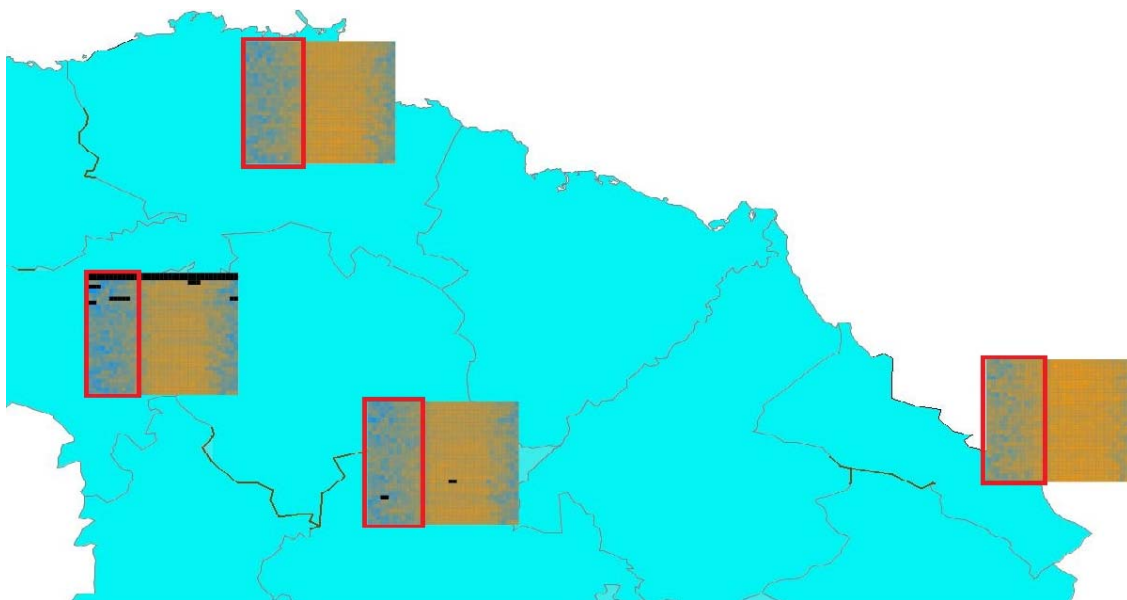
La Figura 53 muestra los mismos parámetros de la visualización tratados en el ejemplo anterior para la estación de Sagua la Grande.



**Figura 53:** Visualización de la temperatura mínima, media y máxima desde 1996 hasta 1999, utilizando el patrón (36,1) (1,4). La visualización corresponde a la estación meteorológica de Sagua la Grande.

Las características de la imagen siguen las especificaciones expuestas en el ejemplo anterior. Se puede apreciar que las variables se comportan de manera similar a la estación de Santo Domingo, con el aumento de los valores de la temperatura mínima, media y máxima hacia los meses fríos del año 1997 y mayores valores de las variables temperatura media y máxima hacia los meses cálidos del año 1998, con respecto al comportamiento en los otros años.

La Figura 54 muestra la representación visual de todos los valores de la variable temperatura mínima utilizando el patrón (36,1) (1,30) explicado anteriormente.



**Figura 54:** Visualización de todos los valores de la temperatura mínima de las cuatro estaciones meteorológicas de Villa Clara utilizando el patrón (36,1) (1,30).

Resaltado en color rojo se puede apreciar el comportamiento de la variable temperatura mínima para los primeros cuatro meses de todos los años. Obsérvese que para la estación de Caibarién las tonalidades son más naranjas (valores mayores) que para las restantes estaciones. Esto se debe a la cercanía de la estación meteorológica a la zona costera al norte de la provincia Villa Clara.

### 3.3 Conclusiones parciales

En el presente capítulo se expusieron los aspectos fundamentales para el uso del módulo de visualización de datos multiparamétricos incorporado a gvSIG. Esto constituye una importante guía para el uso eficiente de la herramienta por parte de los usuarios.

Se mostró un caso de estudio donde se evidencia la factibilidad del uso del módulo para el análisis visual de los datos meteorológicos de la provincia Villa Clara. En este se pudo constatar las ventajas que brinda la herramienta para explorar datos y obtener información espacio-temporal relevante sobre los mismos.

## **CONCLUSIONES**

Como resultado de esta investigación se desarrolló un módulo de visualización de datos multiparamétricos, que mediante un enfoque de integración entre la VisCi y los SIG, permite visualizar en un SIG datos multiparamétricos poco muestreados espacialmente y amplios en el tiempo, cumpliéndose el objetivo general planteado. Además se presentaron los siguientes aspectos:

- Se seleccionaron un conjunto de técnicas para la visualización de datos multiparamétricos, que por su generalidad y amplio uso facilitan la exploración de manera visual de uno o varios conjuntos de datos, de manera independiente o coordinada respectivamente, obteniendo en el último caso una percepción del origen de los datos que están siendo visualizados.
- Se seleccionó el SIG gvSIG para la incorporación del módulo de visualización de datos multiparamétricos implementado mediante el enfoque mezclado de integración entre las disciplinas VisCi y SIG.
- Se desarrolló una metodología para la futura adición de nuevas técnicas de visualización de datos multiparamétricos al módulo de visualización de gvSIG.
- Se presentó un caso de estudio donde se realiza el análisis visual de los datos meteorológicos de la provincia Villa Clara, que evidencia la idoneidad del uso del módulo de visualización desarrollado para el análisis visual de datos meteorológicos poco muestreados espacialmente.

Aunque el módulo desarrollado es probado con datos meteorológicos las herramientas implementadas son de propósito general, permitiendo su utilización en múltiples campos de aplicación donde la densidad espacial de los datos sea pobre y la cantidad de observaciones de cierto número de variables sea grande.

## **RECOMENDACIONES**

- Implementar nuevas técnicas de visualización y adicionarlas al módulo de visualización de gvSIG.
- Perfeccionar el sistema de gestión de consultas para una mejor selección de los datos.
- Probar la efectividad del análisis visual de datos multiparamétricos mediante la herramienta desarrollada en otros dominios de aplicación.



## REFERENCIAS BIBLIOGRÁFICAS

- ANDREWS, D. F. 1972. Plots of high dimensional data. *Biometric*, Vol.28, No.1, 12.
- ANDREWS, K. 2005. Information Visualisation. Available: <http://courses.iicm.edu/ivis/>.
- ANGUIX, A. 2009. gvSIG: un proyecto global Casos de éxito. *Primera jornada de Latinoamérica y el caribe de usuarios de gvSIG* [Online].
- ANGUIX, A. & CARRIÓN, G. gvSIG: Soluciones Open Source en las tecnologías espaciales. In: GISPLANET 2005, 2005.
- ANKERST, M., KEIM, D. A. & KRIEGEL, H. P. 'Circle Segments': A Technique for Visually Exploring Large Multidimensional DataSet. In: Visualization, 1996 San Francisco.
- BEDDOW, J. 1990. Shape coding of multidimensional data on a microcomputer display. *Proceedings of the 1st conference on Visualization '90*. San Francisco, California: IEEE Computer Society Press.
- BOLSTAD, P. 2005. *GIS Fundamentals: A first text on Geographic Information Systems*, White Bear Lake, MN, Eider Press.
- CHERNOFF, H. 1973. The Use of Faces to Represent Points in K-Dimensional Space Graphically. *Journal of the American Statistical Association*, 68, 361-368.
- CLARKE, K. 1990. *Analytical and Computer Cartography*, Prentice Hall Professional Technical Reference.
- COOK, D., SYMANZIK, J., MAJUREA, J. J. & CRESSIEB, N. 1997. Dynamic graphics in a GIS: more examples using linked software. *Computers & Geosciences*, 23, 371-385.
- COWEN, D. 1989. NCGIA lecture.
- CUI, Q., WARD, M. & RUNDENSTEINER, E. 2006. Enhancing scatterplot matrices for data with ordering or spatial attributes. *Visualization and Data Analysis, Part of IS\&T/SPIE Symposium on Electronic Imaging*. San Jose, California, USA.
- EICK, S. G. 2000. Visualizing Multi-Dimensional Data. *ACM SIGGRAPH*, 34 No.1.
- FURNAS, G. W. & BUJA, A. 1994. Prosection Views: Dimensional Inference through Sections and Projections *Journal of Computational and Graphical Statistics*, 3, 323-353.
- GAHEGAN, M., TAKATSUKA, M., WHEELER, M. & HARDISTY, F. 2002. *Introducing GeoVISTA Studio: an integrated suite of visualization and computational methods for exploration and knowledge construction in geography*. , Oxford, ROYAUME-UNI, Elsevier.
- GRASS. 2008. *Geographic Resources Analysis Support System* [Online]. Available: <http://grass.itc.it> [Accessed 15 de Febrero de 2010].
- GREAT BRITAIN. COMMITTEE OF ENQUIRY INTO THE HANDLING OF GEOGRAPHIC INFORMATION 1987. *Handling geographic information: report to the Secretary of State for the Environment of the Committee of Enquiry into the Handling of Geographic Information / Chairman, Lord Chorley*, London, H.M.S.O.
- HANSEN, C. & JOHNSON, C. R. 2005. *The Visualization Handbook*, Elsevier



- HEARNSHAW, H. M. & UNWIN, D. J. 1994. *Visualization in geographical information systems*, Chichester.
- HOGEWEG, M. 2000. *Spatio-temporal visualisation and analysis* Master of Science in Geographical Information Systems University of Salford
- KEIM, D. A. 2000. Designing Pixel-Oriented Visualization Techniques: Theory and Applications. *IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS*, 6, 59-78.
- KEIM, D. A. 2002. Information Visualization and Visual Data Mining. *IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS*, 8, 1-8.
- KEIM, D. A., ANKERST, M. & KRIEGEL, H.-P. 1995. Recursive Pattern: A Technique for Visualizing Very Large Amounts of Data. *Proceedings of the 6th conference on Visualization '95*. IEEE Computer Society.
- MATLAB. 2004. Ayuda del Matlab 7.0.
- MITAS, L., BROWN, W. M. & MITASOVA, H. 1997. Role of dynamic cartography in simulations of landscape processes based on multi-variate fields. *Computers & Geosciences*, 23, 437- 446
- MORELL, A. & PÉREZ, C. 2006. Biblioteca de módulos de visualización de fluidos para OpenDX.
- NORTH, C. & SHNEIDERMAN, B. 2000. Snap-Together Visualization: A User Interface for Coordinating Visualizations via Relational Schemata.
- ONG, H.-L. & LEE, H.-Y. 1996. Software report: winviz -a visual data analysis tool. *Computers & Graphics*, 20, 83-84.
- PÉREZ, C. & ORTEGA, J. C. Modelación de datos para la visualización científica. In: *Compumat 2005*, 2005 La Habana.
- QGIS. 2009. *Quantum GIS Official Web Site* [Online]. Available: <http://www.qgis.org/> [Accessed 22 de Febrero de 2010].
- RHYNE, T. M. 1997. Going virtual with Geographic Information and Scientific Visualization *Computers & Geosciences*, 23, 489 - 491.
- RHYNE, T. M., IVEY, W., KNAPP, L., KOCHEVAR, P. & MACE, T. 1994. Visualization and Geographic Information System integration: What are the needs and requirements, if any ? (Panel). IEEE Computer Society.
- RUMBAUGH, J. & BOOCH, G. 2000. *El Lenguaje Unificado de Modelado. Manual de Referencia*, Madrid, Addison Wesley.
- SALGADO, E. 2003. *Visualization Techniques* [Online]. Rostock, Germany. [Accessed].
- TAKATSUKA, M. & GAHEGAN, M. 2002. GeoVISTA Studio: a codeless visual programming environment for geoscientific data analysis and visualization. *Computers & Geosciences*, 28(10), 1131-1144.
- THEISEL, H. 2000. Scientific Visualization. Available: [http://www.aimatshape.net/s05\\_sciviz.zip](http://www.aimatshape.net/s05_sciviz.zip).
- TOIT, S. H. C. D., STEYN, A. G. W. & STUMPF, R. H. 1986. *Graphical exploratory data analysis*, Springer-Verlag New York, Inc.
- TREINISH, L. A. & ROGOWITZ, B. E. 1993. An Architecture for Perceptual Rule-Based Visualization.
- VAN WIJK, J. J. & VAN LIERE, R. 1993. HyperSlice - Visualization of Scalar Functions of Many Variables. IEEE Computer Society.

- WARD, M. O. 2002. A taxonomy of glyph placement strategies for multidimensional data visualization. *Information Visualization*, 1, 194-210.
- WARD, M. O. 2008. Multivariate Data Glyphs: Principles and Practice. *In*: CHEN, C.-H., HÄRDLE, W. & UNWIN, A. (eds.) *Handbook of Data Visualization*. Springer-Verlag Berlin Heidelberg.
- XIE, Z., HUANG, S., WARD, M. O. & RUNDENSTEINER, E. A. 2006. Exploratory Visualization of Multivariate Data with Variable Quality. *IEEE*.

## ANEXOS

### Anexo 1. Estructura de los ficheros en formato *arff*

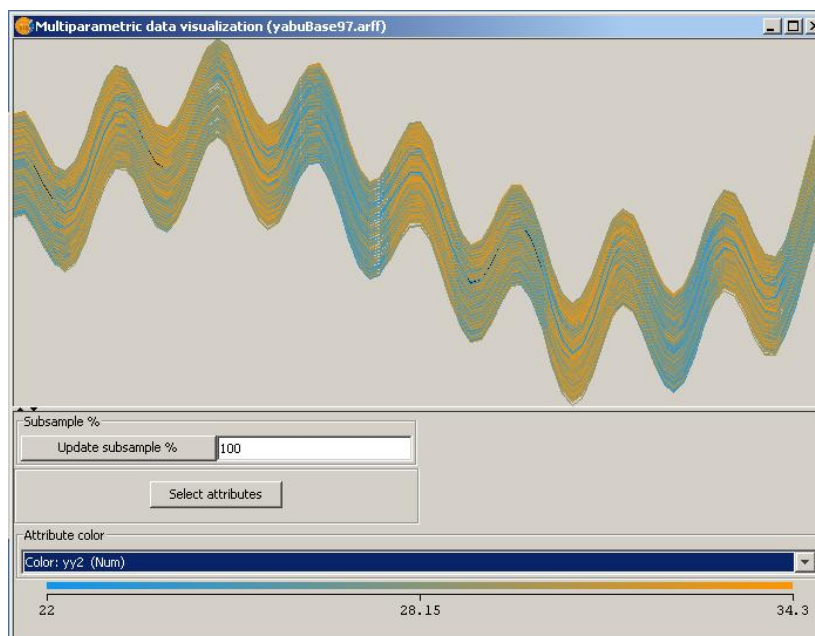
```

1 @relation weather
2
3 @attribute outlook {sunny, overcast, rainy}
4 @attribute temperature real
5 @attribute humidity real
6 @attribute windy {TRUE, FALSE}
7 @attribute play {yes, no}
8
9 @data
10 sunny,85,85,FALSE,no
11 sunny,80,90,TRUE,no
12 overcast,83,86,FALSE,yes
13 rainy,70,96,FALSE,yes

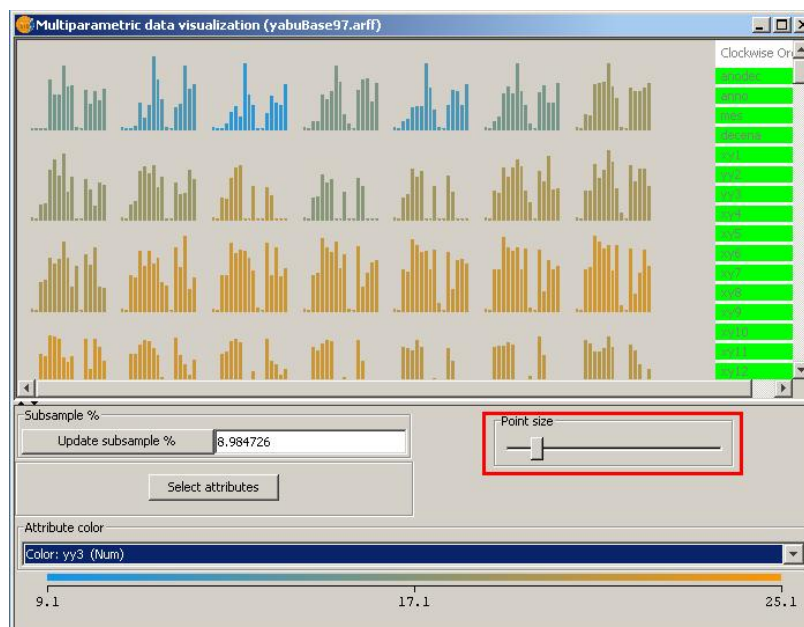
```

- **@relation** <relation-name> (línea 1): Indica el nombre de la relación.
- **@attribute** <attribute-name> <datatype> (líneas de la 3 a la 7): Sección para la declaración de cada atributo del conjunto de datos, indicando su nombre y el tipo de dato. En <attribute-name> se indica el nombre del atributo, que debe comenzar por una letra y entre comillas si presenta espacios. En <datatype> se indica el tipo de dato del atributo que puede ser:
  - **NUMERIC** (numérico).
  - **STRING** (texto).
  - **DATE** [<date-format>] (fecha). En <date-format> se indica el formato de la fecha, que será del tipo "yyyy-MM-dd'T'HH:mm:ss".
  - **<nominal-specification>**. Estos son tipos de datos definidos por el especialista y que pueden tomar valores de un conjunto finito (línea 3).
- **@data** (a partir de la línea 9): Sección donde se incluyen los datos correspondientes a las observaciones de las variables. Cada columna se separa por comas y todas las filas deben tener el mismo número de columnas, que debe coincidir con el número de atributos declarados. Los valores perdidos se representan con el signo de interrogación.

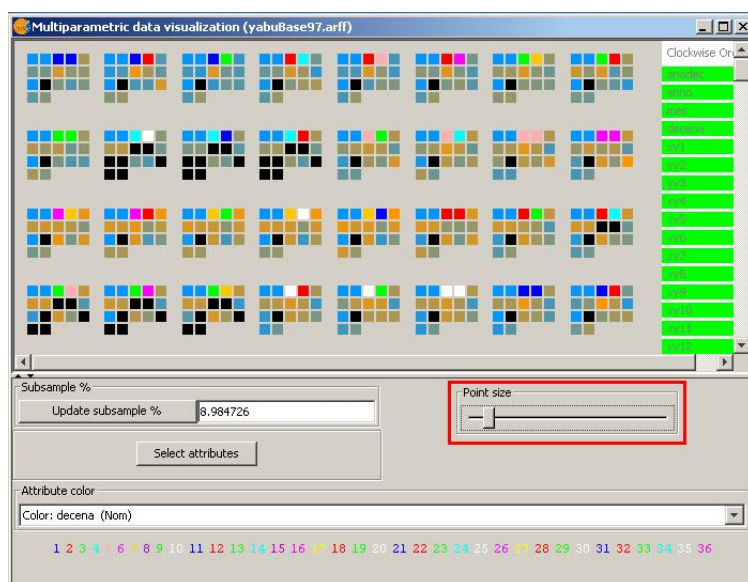
## Anexo 2. Visualización de un conjunto de datos empleando la técnica gráfico de Andrews



### Anexo 3. Visualización de un conjunto de datos empleando iconos en forma de estrellas



#### Anexo 4. Visualización de un conjunto de datos empleando shape coding



## Anexo 5. Ejemplos de ficheros de configuración de proyectos de visualización coordinada

El siguiente ejemplo muestra el fichero de configuración utilizado para la visualización de los datos meteorológicos de la provincia Villa Clara. En este caso se utiliza el mapa VillaClara\_region.shp como mapa base y Estaciones\_point.shp para las localizaciones. El campo @data muestra los archivos de datos asociados a los índices que identifican a las cuatro estaciones meteorológicas.

```
%Estaciones meteorológicas de Villa Clara%
```

```
@map VillaClara_region.shp
```

```
@location another Estaciones_point.shp
```

```
@data
```

```
0,yabuBase97.arff
```

```
1,SaguaBase97.arff
```

```
2,CAibarienBase97.arff
```

```
3,SantoDomingoBase97.arff
```

En el ejemplo que se muestra a continuación las localizaciones son realizadas en el mapa base. Se utiliza el campo MUNICIPIO del mapa VillaClara\_region.shp para asociar a cada municipio de la provincia un fichero de datos.

```
%Visualización sobre áreas %
```

```
@map Villa Clara_region.shp
```

```
@location in_map
```

```
@data
```

```
0,Corralillo.arff
```

```
1,QuemadoGuines.arff
```

```
2,Placetar.arff
```

```
3,SantoDomingo.arff
```

```
4,SagualaGrande.arff
```

- 5,Remedios.arff
- 6,Ranchuelo.arff
- 7,Manicaragua.arff
- 8,Encrucijada.arff
- 9,SantaClara.arff
- 10,Caibarien.arff
- 11,Cifuentes.arff
- 12,Camajuani.arff



## Anexo 6. Pasos para la configuración de un proyecto de visualización coordinada

Al seleccionar una de las opciones para la configuración de un proyecto de visualización coordinada se muestra un asistente cuya ventana inicial se puede observar en la Figura 55.

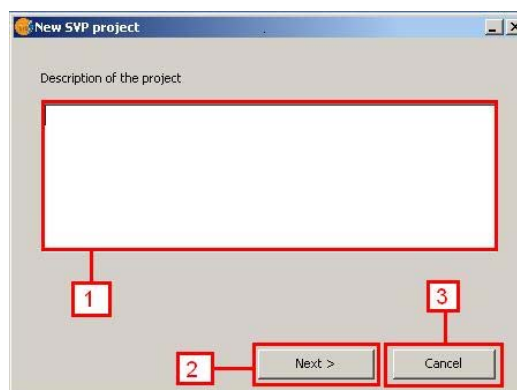


Figura 55: Vista para añadir la descripción del proyecto.

La descripción de la vista es como sigue a continuación, según la numeración en la figura:

1. Campo de texto que permite insertar una descripción del proyecto. Esta operación es opcional.
2. Botón *Next*: carga la siguiente vista que se muestra en la Figura 56.
3. Botón *Cancel*: cancela la creación del proyecto.

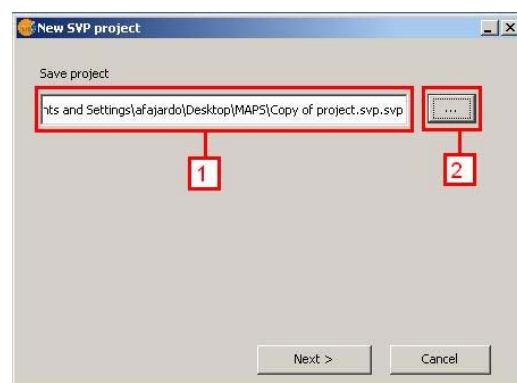
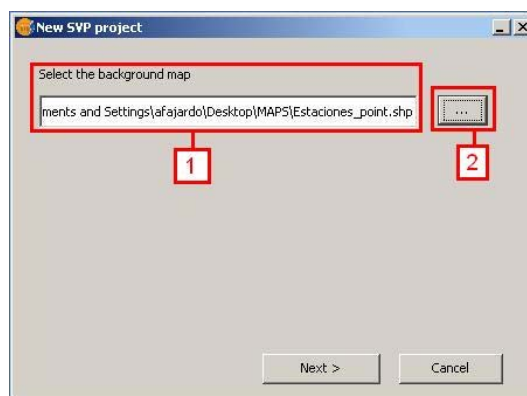


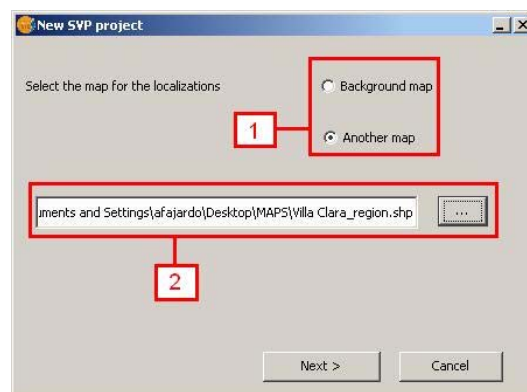
Figura 56: Selección del directorio del proyecto de visualización coordinada.

Esta vista permite salvar el fichero de configuración del proyecto con un nombre y en el directorio especificado por el usuario. Al pulsar el botón dos se muestra un diálogo para salvar el fichero, una vez aceptada la selección se muestra en el campo de texto uno la dirección donde se salvará el fichero. Al pulsar el botón *Next* se muestra la vista de la Figura 57.



**Figura 57: Selección del mapa base.**

Esta vista permite la selección, mediante el botón dos, del mapa base para la visualización de las técnicas. La dirección y nombre del archivo se muestran en uno, una vez realizada la selección. Se pulsa el botón *Next* y se muestra la vista de la Figura 58 para la selección del mapa que será utilizado para las localizaciones.

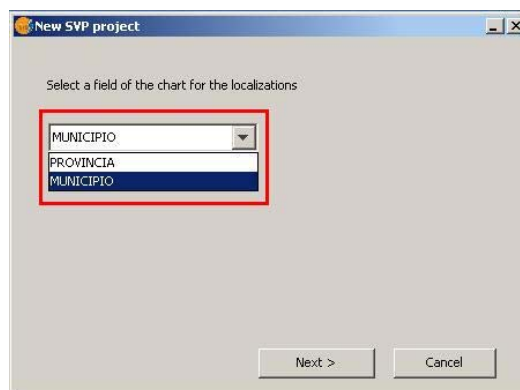


**Figura 58: Selección del mapa para las localizaciones.**

Las opciones enumeradas se explican a continuación:

1. Selección del mapa para las localizaciones: esta opción le permite al usuario definir el mapa que será utilizado para realizar las localizaciones de los archivos de datos. Por defecto la opción activada es *Background map*, que significa que la localización será realizada sobre el mapa base seleccionado en la vista anterior. Por el contrario, si elige la opción *Another map*, se habilitan el campo de texto y el botón con numeración dos.
2. Selección de otro mapa para las localizaciones: Al ser seleccionada la opción *Another map* el usuario tiene la posibilidad de seleccionar un mapa vectorial de puntos que contenga las localizaciones. Este nuevo mapa debe tener una correspondencia con el mapa base; principalmente se utiliza esta opción cuando se dispone de un mapa con las localizaciones exactas en la región del origen de los datos que se deseen visualizar.

Al pulsar el botón *Next* se muestra la vista de selección del atributo del mapa de localizaciones que el usuario utilizará para localizar los archivos de datos. Ver Figura 59.



**Figura 59:** Selección del campo para las localizaciones.

Todo archivo en formato *ESRI shape* tiene asociado un archivo en formato de tabla que contiene información asociada a los objetos geométricos, como pueden ser los nombres de las regiones, coordenadas de las localizaciones, etc. Esta tabla tiene tantas filas como geometrías tenga definido el mapa. Las geometrías a su vez pueden ser líneas, puntos o

polígonos. Es posible asociarle un archivo de datos multiparamétricos a una geometría que representa una localización. La correcta selección del atributo de la tabla que servirá como guía para localizar los datos es fundamental, pues permite conocer con mayor certeza a que región o punto del mapa se está asociando el archivo con los datos multiparamétricos. Una vez seleccionado el atributo se pulsa el botón *Next* y se muestra la vista de la Figura 60.

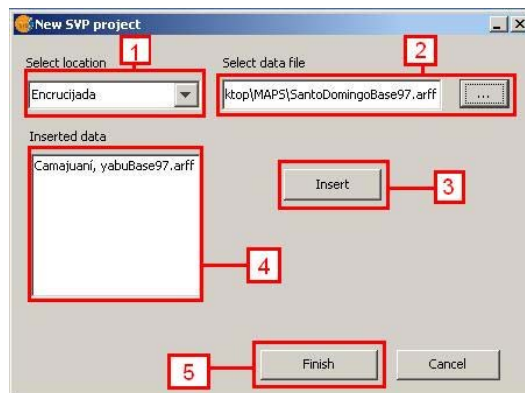
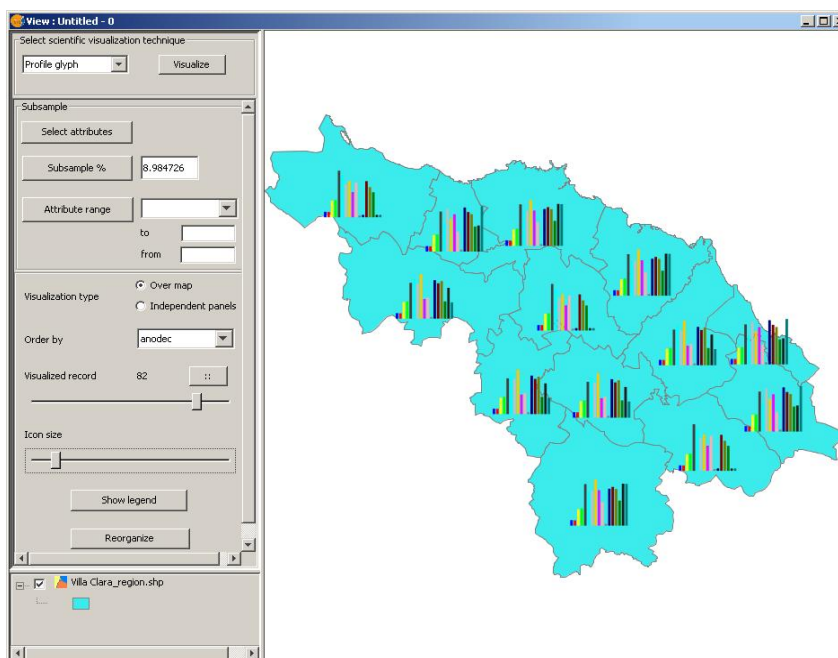


Figura 60: Inserción de los archivos de datos en las localizaciones.

Esta vista permite asociar un archivo de datos multiparamétricos a una localización del mapa. A continuación se brinda la descripción de las componentes que conforman la vista:

1. Selección de la localización a la que se va a asociar un archivo de datos. Se muestran los valores según el campo seleccionado en el paso anterior.
2. Opción que muestra un diálogo para la selección del fichero de datos asociado a la localización seleccionada en uno.
3. Botón *Insert*: inserta la localización y el nombre del fichero de datos en el archivo de configuración del proyecto.
4. Campo de texto que muestra las inserciones realizadas.
5. Botón *Finish*: termina la configuración del proyecto de visualización coordinada y cierra la vista.

## Anexo 7. Visualización coordinada utilizando iconos en forma de barras



## Anexo 8. Visualización coordinada utilizando shape coding

