



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

Romel Vázquez-Rodríguez¹; Carlos Pérez-Risquet¹ & Juan Carlos Torres-Cantero²

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

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

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

DOI: http://dx.doi.org/10.11137/2013_1_98_111

Abstract

The exploration of large GIS models containing spatio-temporal information is a challenge. In this paper we propose the integration of scientific visualization (ScVis) techniques into geographic information systems (GIS) as an alternative for the visual analysis of data. Providing GIS 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 regard, we present a new visual data mining tool for the GIS gvSIG. This tool has been implemented as a gvSIG module and contains several ScVis techniques for multiparameter data with a wide range of possibilities to explore interactively the data. The developed module is a powerful visual data mining and data visualization tool to obtain knowledge from multiple datasets in time and space. A real case study with meteorological data from Villa Clara province (Cuba) is presented, where the implemented visualization techniques were used to analyze the available datasets. Although it is tested with meteorological data, the developed module is of general application in the sense that it can be used in multiple application fields related with Earth Sciences.

Keywords: Visual data mining; scientific visualization; information visualization; GIS

Resumo

A exploração de modelos SIG de grande dimensão contendo informação espaço-temporal é um desafio. Neste artigo são apresentadas técnicas de visualização científica (ScVis) em sistemas de informação geográfica (GIS) como uma alternativa para a análise visual dos dados. Fornecer SIG com tais ferramentas melhora a análise e compreensão de conjuntos de dados com muito baixa densidade espacial e permite ainda encontrar correlações entre variáveis no tempo e no espaço. Neste âmbito, é proposta uma ferramenta visual de mineração de dados para a aplicação de GIS gvSIG. Esta ferramenta foi implementada como um módulo gvSIG e inclui várias técnicas ScVis para dados com múltiplos parâmetros, disponibilizando uma ampla gama de possibilidades de exploração interativa desses mesmos dados. O módulo desenvolvido é uma poderosa ferramenta de visualização e de mineração de dados que pode ser utilizado para extração de informação com origem em múltiplos conjuntos de dados espaço-temporais. Um estudo de caso real com dados meteorológicos da província de Villa Clara (Cuba) é apresentado. Nesta aplicação prática as técnicas de visualização desenvolvidas foram utilizadas para analisar conjuntos de dados disponíveis. Embora seja testado com dados meteorológicos, o módulo desenvolvido é suficientemente generalista para poder ser aplicado em múltiplos domínios temáticos no âmbito das Ciências da Terra.

Palavras-chave: Mineração de dados visual; visualização científica; visualização de informações; GIS

1 Introduction

Due to the fast development of Geographic Information Systems (GIS) caused in part by the revolution of the new information and communication technologies, GIS has become an essential tool for analyzing and solving several problems in business, industry, as well as social and governmental institutions (Burrough & McDonnell, 1998). Their versatility has allowed their use in almost all activities that have a spatial component. Thus, GIS have become an essential tool in many areas related to strategic management, including the processes of demographic analysis, environmental protection and planning and resource management applications. Overall, we find GIS applications in tasks such as:

- Location: asking for the features of a specific place.
- Condition: compliance or not, with a condition imposed to the system.
- Trend: comparison between different temporal or spatial situations of a given feature.
- Routes: calculation of optimal routes between points.
- Guidelines: detecting spatial patterns.
- Models: generation of models from simulated events or actions.

Scientific Visualization (ScVis) has become a research area of increasing interest in recent years. This is mainly motivated by the constant increase of the data volumes generated in multiple application fields and by the continuous increase in the power of modern graphics interfaces, which allows the generation of sophisticated images. In the past two decades, significant progress has been made in setting up the visualization as a data exploration tool flexible and easy to use. Visual data analysis is a new approach that integrates both human perception and automatic computational methods, enabling a better understanding and analysis of large and complex datasets (Keim *et al.*, 2008).

One special data class is composed of multiparameter data (multidimensional data; multivariate data). Multiparameter data consists of m -dimensional data of scalar type distributed over points in the observation space. Many visualization techniques have been developed for multiparameter

data, the most relevant of them are presented in section 2. The flexibility and possibilities of interaction of these techniques can replace the shortcomings of the interpolation methods used by current GIS when working with very low spatial resolution data.

The integration of ScVis techniques into GIS is an innovative idea that combines the advantages and strengths for data analysis of both approaches (Rhyne & MacEachren, 2004). Section 3 discusses the alternatives for the integration of ScVis techniques into GIS.

Section 4 presents the module developed for the integration of scientific visualization techniques for multiparameter data in the gvSIG GIS. This module was used with meteorological data from the province of Villa Clara, Cuba; thus showing the effectiveness of this type of analysis when the spatial resolution of the data is poor and the period of data acquisition is long.

Weather variables measured from 1977 to 2007 in Villa Clara's weather stations are used to demonstrate the system's usefulness. Although this time series is fairly complete, it only contains information from four weather stations. This information is not sufficient for the spatial interpolation techniques used by most current GIS. By using the multiparameter visualization techniques developed in this work it is possible to make an analysis of these data. The conclusions are presented in the last section of this paper.

2 Multiparameter Visualization Techniques

There is a large number of problems where each data point contains more than one attribute. These attributes can be dates, locations, prices or descriptive values, and may have or not a spatial reference. This type of data is called multiparameter, and it is usually found in applications of data mining, statistics and artificial intelligence (Keim, 2002). A more precise definition of multiparameter data can be found in Bergeron & Grinstein (1989), Brodlie *et al.* (1992), and Schumann & Müller (2000).

The main goal of multiparameter data visualization methods is to make representations or reveal correlations between attributes (Eick, 2000; Keim, 2002; Theisel, 2000). There is a wide range of ScVis techniques, for which several improvements have also been created. The techniques can be classified in geometric, icon-based and pixel-based (Keim, 2000).

2.1 Geometric Techniques

Geometric techniques are those that use elements (such as lines, points or curves) as visual properties to represent data. There is a great number of techniques, such as projection views (Bergeron & Grinstein, 1989; Brodlié *et al.*, 1992; Schumann & Müller, 2000), hyper slices (Wijk & Liere, 1993), parahistograms (Ong & Lee, 1996) and star coordinates (Kandogan, 2000).

However, three techniques stand out because of their wide use. These are the scatter plots matrices (Cleveland, 1993; Cui *et al.*, 2005; Hansen & Johnson, 2005; Keim, 2002; Theisel, 2000; Inselberg & Dimsdale, 1990), parallel coordinates (Inselberg & Dimsdale, 1990; Keim, 2002; Theisel, 2000) and Andrews graphics (Andrews, 1972). The parallel coordinates technique is especially useful for displaying patterns in data or for perceiving relationships between attributes, since the result it is very intuitive. The fundamental strength of Andrews graphics is that, it may represent datasets of a relatively large amount of records and with a higher

number of dimensions. The geometric techniques are very good for perceiving correlations between two variables. Graphical examples of these techniques can be seen in Figure 1.

2.2 Icon-Based Techniques

Icon-based techniques consisted of creating an image from a dataset. The result is a set of figures with different visual features. Icon-based techniques have two defining parameters: the first is the kind of figure that will represent each observation, i.e., the shape of the icon, and the second is how the position of each icon is defined (Theisel, 2000).

Among the methods to create icons we have Chernoff faces (Chernoff, 1973), Shape coding (Beddow, 1990), profile glyphs (Chen *et al.*, 2008) and starfield (Andrews, 1972; Eick, 2000; Theisel, 2000; Ward, 2002; Xie *et al.*, 2006). Moreover, it is common to create icon editors for specific applications (Andrews, 1972; Theisel, 2000).

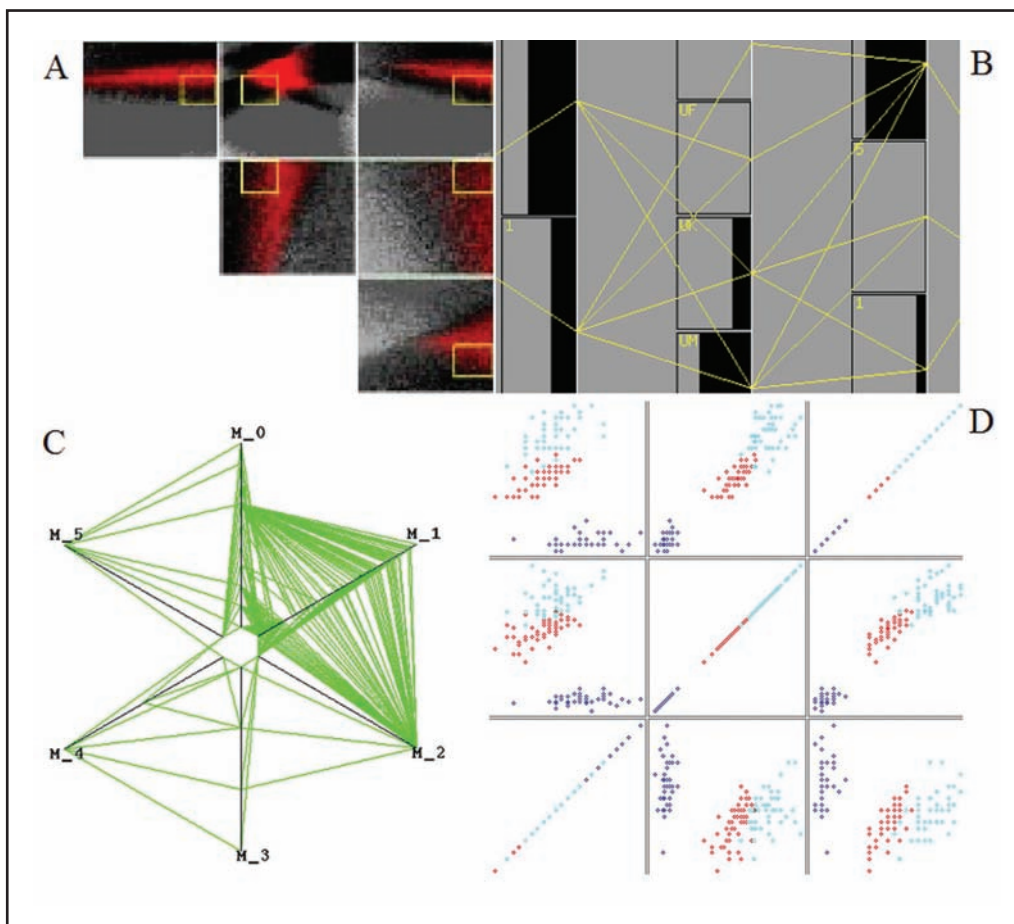


Figure 1 Representation of some geometric techniques.
A) projection views,
B) parahistograms,
C) star coordinates,
D) scatter plots matrices.

Icon-based techniques are recommended when the number of variables is between 10 and 15 and the number of data is high. These techniques can be used with a spatial reference. Graphical examples of these techniques can be seen in Figure 2.

2.3 Pixel-Based Techniques

Pixel-based techniques use a pixel to represent each attribute of an observation. The procedure in these techniques is to relate each value of a dimension to color and group the pixels from each dimension in adjacent areas. Because this method uses a single pixel per data value, this technique allows to display up to more than one million values (Keim, 2000).

Pixel-based techniques are most efficient when the number of dimensions is large and when the number of records grows. This is because a pixel usually represents each attribute of an observation. Key challenges in these methods are the choice of the color for each element and the positioning mode of the pixels (Keim, 2000; Keim, 2002).

Positioning the pixels in the appropriate way allows to see information about correlations, dependencies and transcendental regions. Two modes of pixels positioning are the recursive patterns (Keim, 2000; Keim, 2002; Keim *et al.*,

1995; Theisel, 2000) and circle segments (Hansen & Johnson, 2005; Keim, 2000; Keim, 2002). In order to have a better visual perception of these techniques see examples depicted in Figures 5 and 7.

3 SCVIS – GIS Integration

The integration of ScVis techniques into GIS is an innovative alternative for the visualization of multiparameter data. Providing GIS 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.

The integration of ScVis and GIS, according to (Rhyne, 1997), is based on 4 methods:

- Rudimentary: Minimal data sharing between GIS and visualization systems.
- Operational: Consistency of geospatial data.
- Functional: Transparent communication between GIS and visualization systems.
- Merged: Comprehensive toolkit. Tools that widely involve the 2 disciplines.

The rudimentary method uses minimal data integration and exchange between both technologies.

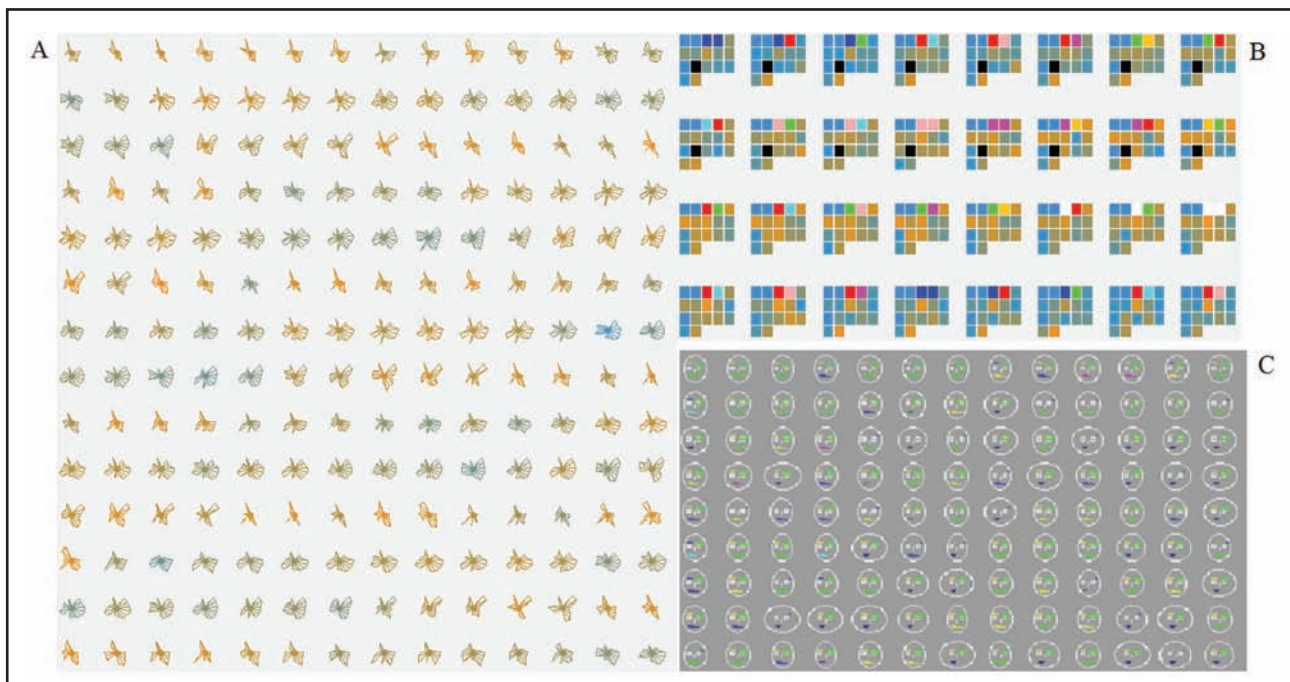


Figure 2 Representation of some icon-based techniques. A) starfield, B) shape coding, C) Chernoff faces. All of them use parts of the icon to codify the variables.

The operational method is intended to provide data consistency, while deleting redundancies between the two technologies (Cook *et al.*, 1997). The functional method provides transparent communication between the software environments (Mitas *et al.*, 1997). The merged approach is concerned with the development of systems where the concepts of cartography, GIS and ScVis are merged in a comprehensive toolkit.

In the late 1990s and continuing into the 2000s, strides were made to integrate GIS and Visualization tools. GIS developers have explored how to incorporate three-dimensional and time series animation capabilities into their software. For example, in the late 1990s, ESRI introduced ArcView 3D Analyst to allow for visualizing surface data and three dimensional modeling. Currently, ArcGIS 3D Analyst is integrated into the ArcGIS suite of tools (Rhyne & MacEachren, 2004).

Meanwhile, SciVis and InfoVis programmers built data readers that support geospatial data formats. As an example, Advanced Visual Systems (AVS), in the late 1990s, developed an AVS-Arc data reader that allowed for direct operational import of ESRI's Arc-Info data into the AVS visualization environment. Today, GIS import and database interaction are incorporated into AVS/Express Professional Edition (Rhyne, 1997). Another example in this case is the Scientific Visualization System OpenDX that had the capacity of read ESRI shape files and other geographic formats with *dxconverter* tool (Treinish, 1992).

Some early attempts of merged integration approach are materialized through tools like GeoVista Studio (Amant & Cohen, 1998; Gahegan *et al.*, 2002; MacEachren *et al.*, 2003; Takatsuka & Gahegan, 2002) and Snap-Together Visualization (North & Shneiderman, 2000). Both tools are based on networks of components that allow the construction of complex workflows linking components for managing geographic data formats and different ScVis techniques.

Snap-Together Visualization was developed by North & Shneiderman (2000) as an innovative approach to creating flexible coordinated views onto datasets. Multiple coordinated views can enable information visualization users to rapidly explore complex data and uncover patterns and relationships. Building such coordinated views when needed is a challenging task for most users. Snap-Together Visualization allows users to construct customized, coordinate multiple views that fit their needs. Users of Snap query their relational database and

load results into desired views (e.g., maps, graphs, tables). Then they specify the kinds of coordination supported between views (for selecting, navigating, or re-querying).

GeoVISTA Studio is an open source, component-based software development environment distributed through SourceForge. Studio provides a visual programming interface through which users can quickly build applications from available JavaBeans -that are distributed with Studio or acquired from other sources- (Luo, 2007). Studio's visual programming environment allows an analyst to package assembled functionality into a working program (in the form of a cross-platform, JavaBeans component, an applet, or an application). Studio supports development of non-geographic as well as geographic applications, e.g., InfoVis tools.

To support the OpenGIS data interoperability needs mentioned above, Studio developers have begun to adapt and extend the GeoTools geospatial data access and visualization methods for use with other Studio components. GeoTools is a complementary Open Source Java library for developing OpenGIS solutions to geospatial data access, analysis, and presentation tasks.

GeoVISTA includes multiple scientific visualization techniques that can be used with some maps. The Exploratory Spatio-Temporal Analysis Toolkit ESTAT was included to GeoVista Studio and constitutes a good tool to analyze Spatio-Temporal data associated to choropleth maps as area ESRI shape files.

4 GVSIG SCVIS Module

The main objective of this work was to build a multiparameter visualization data module that helps to knowledge acquisition from data measured in the same time interval with a large amount of records, but spatially scarce. The solution consisted of integrating some multiparameter visualization techniques into a GIS, in such a way that data could be analyzed by having a geographical perception of their origin, as can be seen in the Figures 4-7 and Figure 10.

A merged integration by adding a scientific-information visualization module to a GIS was implemented allowing the user to prepare projects to be used in a coordinated visualization. A coordinated visualization means visualizing, by

using scientific visualization techniques, multiple datasets corresponding to multiple data points. It is coordinated because all visualizations are normalized according to the global extreme values of every variable of all data points.

These projects can be customized to be visualized over a vector map of points (point shape file given by the user) or over a vector map of area (area shape file used as background of the visualization); in the latter, the graphics are displayed over the centroid of the corresponding area. The techniques that can not be visualized as a graphic over the map due to space restrictions are visualized in independent panels, where the geographical perception is still present.

The GIS selected to be extended with the scientific visualization module was gvSIG, a powerful open source GIS from Generalitat Valenciana. It is a java-based GIS which is very easy to extend and have very good documentation for developers. gvSIG have a well-structured class hierarchy and supports the reading of several geographical and non geographical formats like tables. The developed module in this paper for gvSIG GIS, is registered with the name extSV 1.0 and is freely available from request to the first author.

In this section the main features of the developed module that was incorporated to gvSIG GIS are explained. This section has 3 subdivisions: the first one explains the structure of the data to be visualized by the gvSIG extension, and a case study with meteorological data is presented. The second section deals with the visualization tool and the developed techniques. The third section shows some examples that illustrate the suitability of using this kind of techniques in GIS to visually extract information and knowledge from the data.

4.1 Preparing the Data to be Visualized

In order to visualize using a coordinated visualization in gvSIG, 2 types of data are necessary: geographical data and multiparameter data. The geographical data can include a background map and a point map, or only a background map, both as ESRI shape files. Each data point should match a multiparameter data file. The developed module allows reading multiparameter data from the *DBF* file associated with the shapefile and *ARFF* file format, that specify metadata like names and types

for every variable of the table. The data in each data point have to match the number of variables and records. Missing values are allowed.

Some application areas like meteorology, geology, demography, etc, generally have three components: space, time and information associated with both components. In our case study, these are the meteorological variables taken into account in the spatio-temporal analysis. The events in a given application area can be analyzed spatially to identify what is happening in different regions. The temporal component, in turn, allows to analyze these events over time and to understand the changes of the variables at different times.

The meteorological data collected by the Meteorological Institute of Villa Clara represent a time series of each of the variables collected from 1977 to 2007. From each of the 4 weather stations in the province, there is a time series of 13 variables. The meteorological variables are collected over a ten-day period (dekadal), the variables are: dekadal average mean (xy1), maximum (xy2) and minimum (xy3) temperature, dekadal average of maximum (xy4) and minimum (xy5) relative humidity, saturation deficit (xy6), dekadal total rainfall (xy7), dekadal insolation (xy8) (average daylight hours), dekadal water vapor pressure (xy9), dekadal average of mean relative humidity (xy10), cloudiness (xy11) (dekadal average in 1/8 cover sky), dekadal mean wind speed (xy12) and dekadal average atmospheric pressure (xy13).

Table A1 shows the way in which this data is stored. This table stores the average in ten days for each of the meteorological variables. Each station has a link with a table similar to this one.

Users can prepare their own multiparameter dataset for each data point and use a wizard that helps to customize datasets to be visualized. It creates a configuration file that is used by the coordinated visualization module.

4.2 Visualization

There are several ways to visualize data with multiparameter ScVis techniques. Some of them allow visualizations over maps, where the association of the data with the geographical area is perceived. In particular, ScVis techniques for multiparameter data do not necessary have to be associated with a map, as shown in Figure 3. However, these techniques can be

Year	Month	Year – dek	dek	xy1	xy2	xy3	xy4	xy5	xy6	xy7	xy8	xy9	xy10	xy11	xy12	xy13
1977	1	197701	01	24.6	27.3	19.3	91	62	6.5	8.9	8	17.1	74	3	6.5	1016.1
1977	1	197702	02	24.3	27.6	19.5	90	65	6.7	8.7	9	17.5	73	4	6.4	1005.6
1977	1	197703	03	24.3	27.6	19.5	90	65	6.7	8.7	9	17.5	73	4	6.4	1005.6
1977	2	197704	04	24.3	27.6	19.5	90	65	6.7	8.7	9	17.5	73	4	6.4	1005.6
1977	12	197736	36	21.5	24.9	16.9	94	65	4	2.1	6.2	15.3	79	4	7.9	1018.9
1978	1	197801	01	24.6	27.3	19.3	91	62	6.5	8.9	8	17.1	74	3	6.5	1016.1
1978	1	197802	02	24.3	27.6	19.5	90	65	6.7	8.7	9	17.5	73	4	6.4	1005.6
1978	1	197803	03	24.3	27.6	19.5	90	65	6.7	8.7	9	17.5	73	4	6.4	1005.6
1978	2	197804	04	24.3	27.6	19.5	90	65	6.7	8.7	9	17.5	73	4	6.4	1005.6
1978	12	197836	36	24.3	27.6	19.5	90	65	6.7	8.7	9	17.5	73	4	6.4	1005.6
2007	1	200701	01													
2007	1	200702	02													
2007	1	200703	03
2007	2	200704	04													
2007	12	200736	36	26.2	30.3	23.4	95	65	3.9	57.1	8	21.8	86	5	4.5	1006.4

Table A1 Single station meteorological data set needed as input information for the developed module^a.
^a A representative sample of variable values is shown. The meaning of variables are: Year (1977-2007), Month (1-12), Year-dekadal, dekadal (01-36), average mean (xy1), maximum (xy2) and minimum (xy3) temperature, dekadal average of maximum (xy4) and minimum (xy5) relative humidity, saturation deficit (xy6), dekadal total rainfall (xy7), dekadal insolation (xy8) (average daylight hours), dekadal water vapor pressure (xy9), dekadal average of mean relative humidity (xy10), cloudiness (xy11) (dekadal average in 1/8 cover sky), dekadal mean wind speed (xy12) and dekadal average atmospheric pressure (xy13).

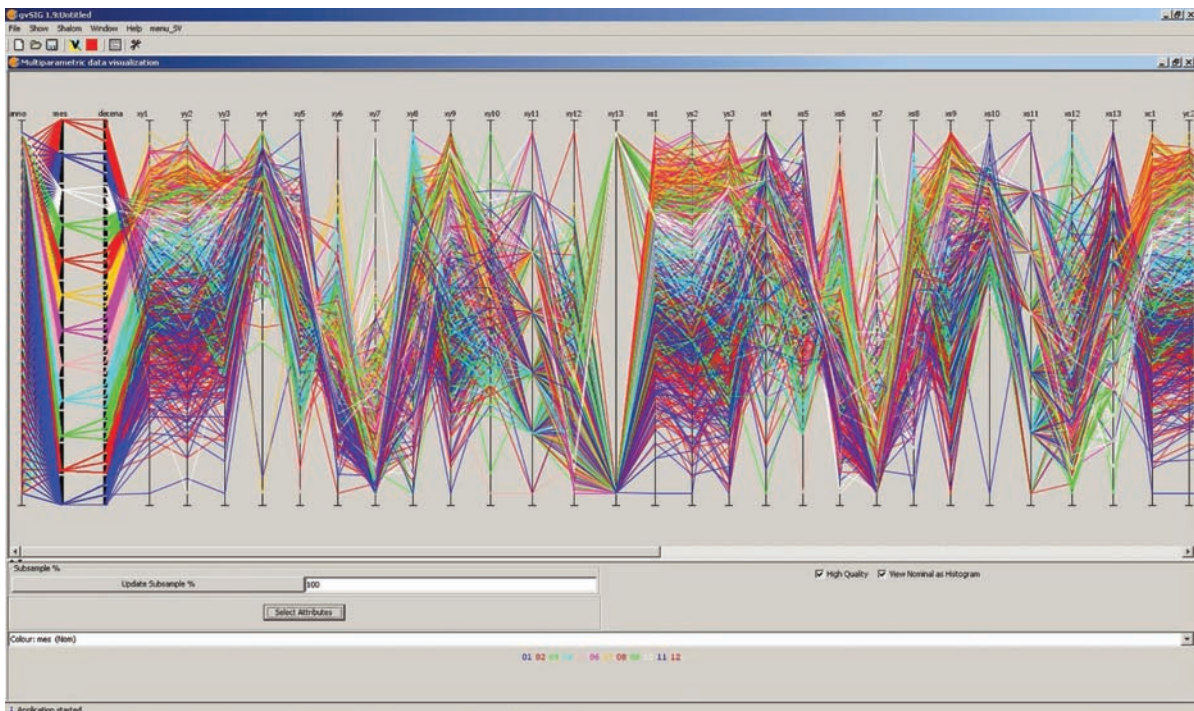


Figure 3 Independent visualization of all variables using the parallel coordinates technique. The order of the variables is: year, month, dekadal, xy1-xy13 (weather variables of the station, see the order and the respective name in Section 4.1), xs1-xs13 (weather variables of another station), similar for the other two stations. The variables month is used as class variable with a different color for each of the 12 month.

applied to all variables in each data point to obtain correlations between certain variables of several points. It is possible to use for example, parallel coordinates to display the 13x4 (52) variables corresponding to the 4 meteorological stations.

Another implementation method is to display data from each point separately, and show a map where it is evident to which region each graphic is related to. This is achieved with an arrow that connects the region in the map with the visualization panel and naming each panel with the same name of the data point file visualized on the map (see Figure 4).

Some ScVis techniques for multiparameter data can be displayed directly over the map. Circle segments, recursive pattern, and icon-based techniques are examples of such techniques (see Figures 5 and 6). With these techniques is possible to show an icon for each data source (a weather station in our case), which represents the set of variables for an observation at a given time. Some widgets (sliders, scroll bars) are used to scroll over the time, changing the icon on the map accordingly (see Figure 6). This permits the user to study the evolution of the data over time. The same approach can be used to scroll the data with respect to other variables.

All these methods were implemented in the developed module for gvSIG. For example, for non coordinated visualizations, that is, when the data do not necessarily have to be associated with a map, the system allows the user to load a data file to be analyzed using any of the following techniques:

- Parallel coordinates
- Andrews graphics
- Starfield
- Shape coding
- Profile Glyphs
- Circle segments
- Recursive patterns

Coordinated visualizations with maps include the following techniques:

- Parallel coordinates (in independent panels)
- Andrews graphics (in independent panels)
- Starfield (a record at a time)
- Shape Coding (a record at a time)

- Profile Glyphs (a record at a time)
- Circle segments (all selected records)
- Recursive patterns (all selected records)

In our case study we obtained the best results with the recursive patterns technique. It seems to be a very good technique to carry out spatio-temporal analysis in GIS. The case study of section 4.3 illustrates the effectiveness of recursive patterns for analysing meteorological data.

The developed module allows some functionalities that are available for all the techniques:

- Selecting attributes (only the selected attributes are shown in the graphic)
- Selecting a percentage of the records
- Selecting according to an attribute a range of values (makes a subsample of all data in a range of values selected for a given attribute)
- Showing legends (shows the global color range for each variable, it takes minimum and maximum values of all datasets, nominal values are assigned a different color for each value, some techniques like Profile glyphs show in the legend a different color for each attribute)
- Reorganizing attributes (a new order of attributes is given)

The *ordering by* functionality orders all datasets according to a given attribute. Pixel-based techniques and icon-based techniques use this functionality to order all the values using this attribute. By using the time as the attribute the user can analyze data over time.

Each technique has its own configuration panel and features that can be configured separately. In addition to that, there is a group of overall aspects related to technique types. For example, for all icon-based techniques visualized over the map the user can change the size of the icon to obtain a better visualization, and a sliding bar is used to move through records using a selected attribute, i.e. a time attribute.

All developed pixel-based techniques also allow to modify the size of the graphics. These techniques can be visualized in independent panels as well as over the map. The developed geometric techniques were designed to be visualized in

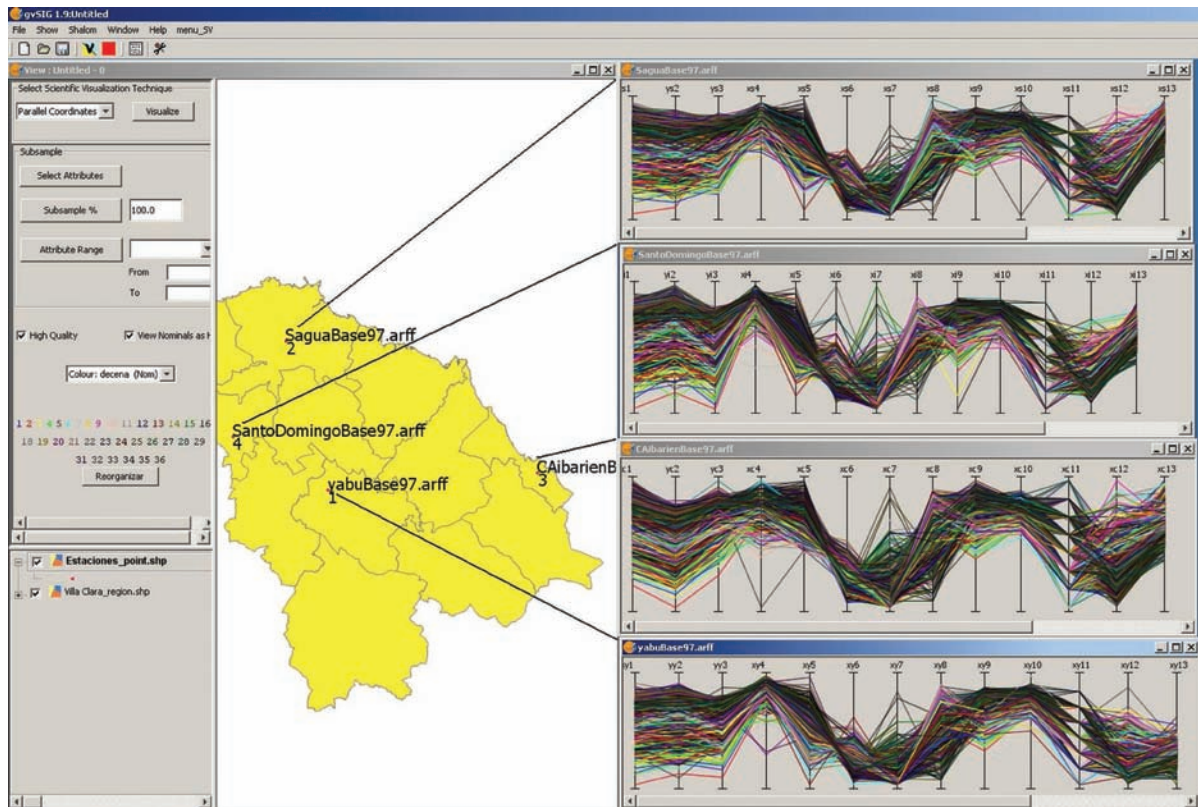


Figure 4. Coordinated visualization with parallel coordinates. Each station show their 13 weather variables, dekadal variables is used as class, there are a different color for each of the 36 dekadal. The order of the variables is the order provided in Section 4.1.

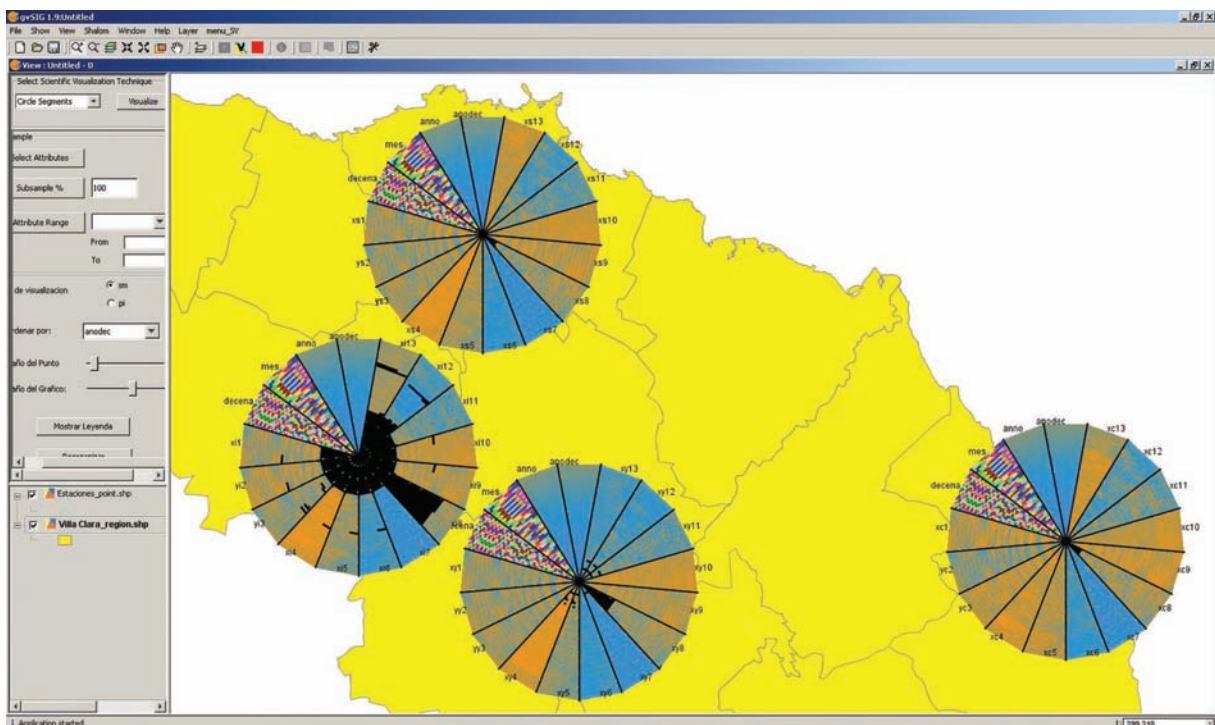


Figure 5 Coordinated visualization with circle segments over the map. Each circle segment is a variable of each station. Weather variables are stored as continues variables, the color of each continues variables range from a dark blue (low values) to dark orange (high values). Discrete variables use a different color for each value. The order of the variables is the same provided in Section 4.1.

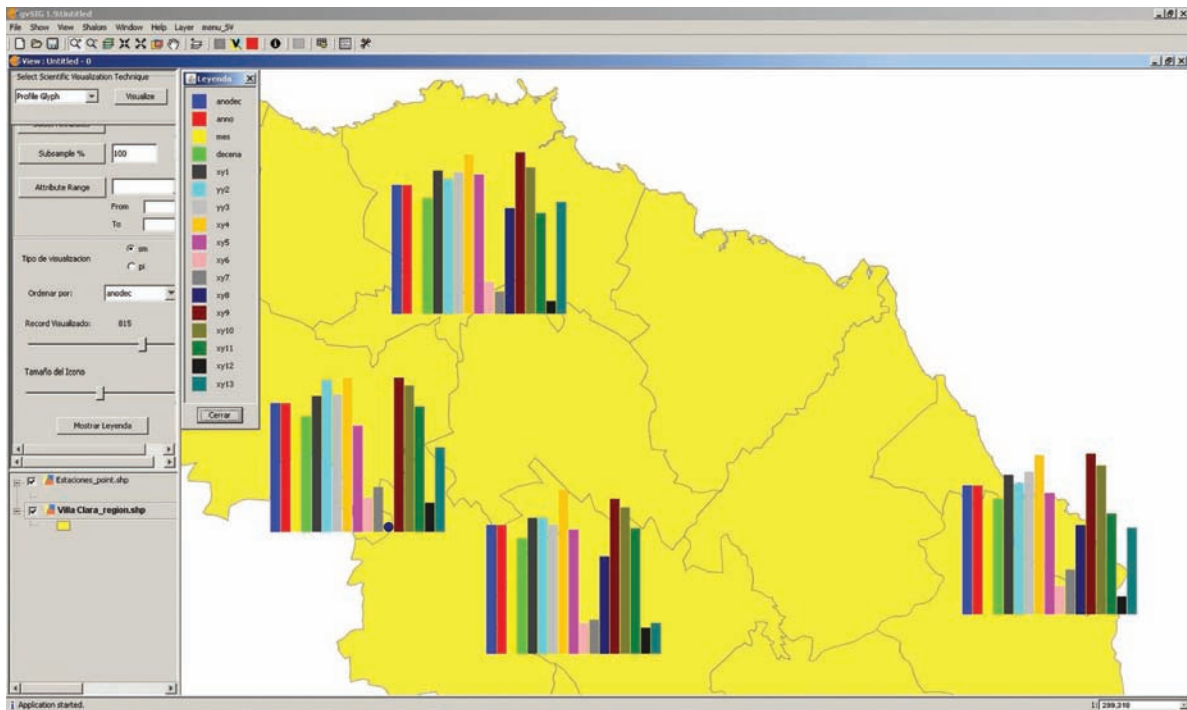


Figure 6 Coordinated visualization with profile glyphs over the map. Shows only one register simultaneously, a different color is used for each variable. The order of the variables is the same provided in Section 4.1. A sliding bar can be used to change for new registers.

independent panels, one for each data point. They are also coordinated by the main configuration panel. The coloring of the records is showed with a color that is defined for a given variable.

4.3 Case Study: Visual Analysis of Meteorological Data

The ScVis module for gvSIG allows the user to obtain visual information about the variables. In this case study, meteorological data from the Villa Clara's four weather stations is analyzed.

The main advantages of using ScVis techniques with multiparameter data in GIS are the possibilities of making comparisons, identifying patterns & correlations, detecting anomalies, variability, and extracting the different patterns that may occur in the variables, in such a way that space and time information can be obtained. Once detected, a graphical relationship or a hypothesis between variables or a pattern can be proved by mathematical methods.

Another advantage provided by the use of such techniques is the possibility offered by the developed module for data interaction, for example, changing the order of the attributes showed in the techniques, and thus having a better look at relationships between certain variables, or including

or removing certain attributes. This interaction is present in all the techniques described in this paper.

The recursive pattern, pixel-based technique, has a special way of interacting with data, which allows the definition of different recursive levels. This gives the possibility of arranging the data in different forms, even if the data has a natural ordering according to one dimension (e.g. time series data). One straightforward possibility is to arrange data items from left to right in a line-by-line fashion. Another possibility is to arrange the pixels in a column-by-column fashion. One possible way of improving the visualizations is to organize the pixels into small groups and arrange the groups to form some global pattern. This strategy corresponds to a two-step approach with a first-order pattern formed by grouping the pixels and a second-order pattern formed by the global arrangement. By taking the results of the second order structure as the basic building element for a third-level structure, a third-order pattern may be introduced. This process may be iterated up to an arbitrary level forming a general recursive scheme (Keim *et al.*, 1995).

This technique seems to be the most useful for meteorologists, due to the large amount of data that can be shown for each meteorological station and because data is time-dependent.

Figure 7 shows the coordinated visualization using recursive pattern. Each station displays its 17 attributes, that is, time attributes like year, month, dekadal, year-dekadal, and the 13 weather variables. The recursive pattern used in the picture is (36,1) (1,30) which shows a year per row from left to right (the 36 dekadal average of a year), and 30 years from top to bottom, 1080 values are shown for each attribute (small square) and for each station. It could be clearly seen the behaviour of the medium (the small square at the upper right corner), maximum (the first small square at the left of the row 2) and minimum (the next small square of the row 2) temperatures. As expected, lower values can be identified at the beginning and end of the square from left to right (blue color), and higher values can be identified in the middle (orange color) which represent the hotter months. Missing values are represented in black color, that is, the moments when weather stations did not take measures. This kind of anomalies can be detected using this module.

Some regularities stand out after using the visualization module with the meteorological data. The 1996-1999 period was analyzed in order to identify the effects of El Niño in Cuba. The recursive pattern technique was used with the pattern (36,1) (1,4). This pattern shows the four-year period.

Figure 8 shows the dekadal minimum, mean and maximum temperature using recursive pattern (36,1)(1,4) from 1996 to 1999 in Santo Domingo weather station. Each square represent a variable where the minimum value of the variable is codified in blue color and the maximum with an intense orange color. The internal rectangles represents the years 1996-1999 from top to down. Note that for the variables minimum, mean and maximum temperatures, the rectangles whose borders are highlighted in red represent the years 1997 and 1998 respectively. In the other hand the rectangles for the years 1996 and 1999 are highlighted in blue for the same variables. That figure shows that:

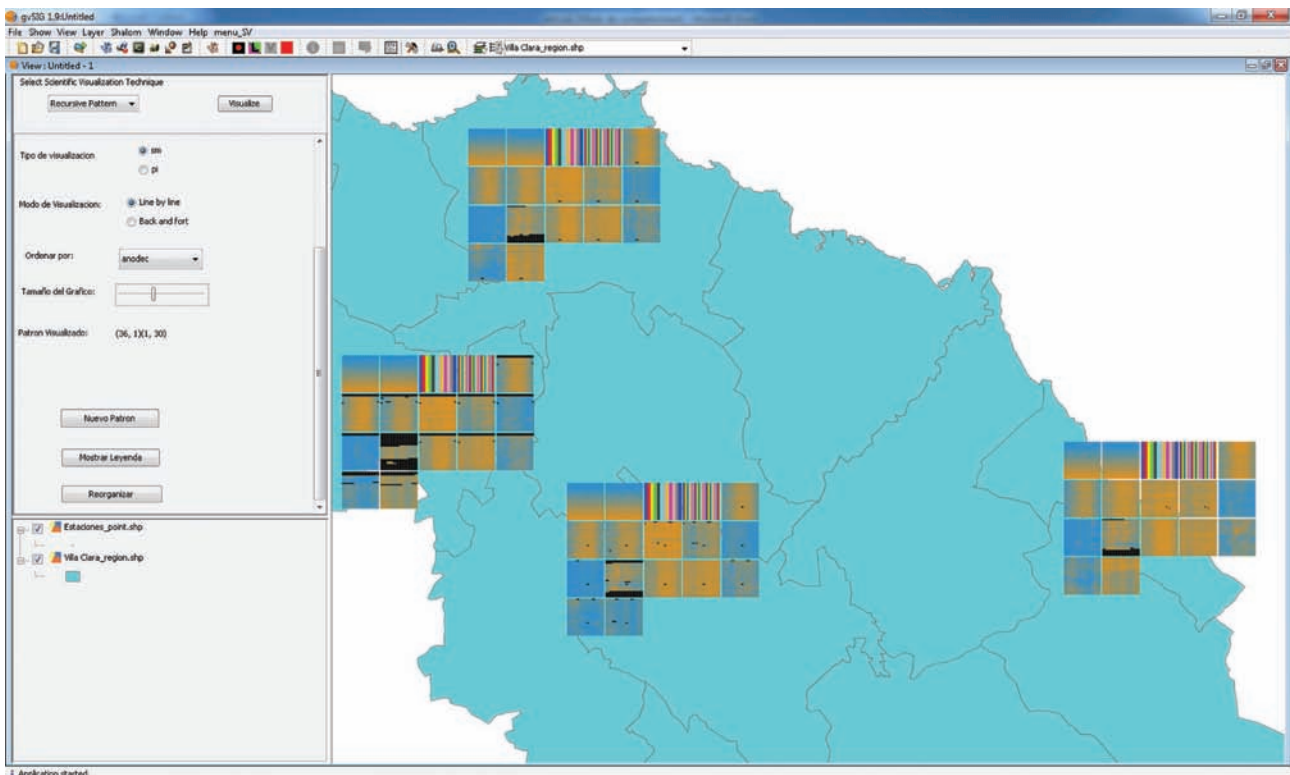


Figure 7 Coordinated visualization with recursive pattern over the map, 1080 values for each attribute of the four weather stations of Villa Clara. Each small square represents a variable, showing their values using the specified pattern. The recursive pattern used in the picture is (36,1)(1,30) which shows a year per row from left to right (the 36 dekadal average of a year), and 30 years from top to bottom. Weather variables are stored as continues variables, the color of each continues variables range from a dark blue (low values) to dark orange (high values). Discrete variables use a different color for each value. The rainbow patterns correspond to the variables month and dekadal. The order of the variables is represented from left to right and from top to bottom, following the order: natural order, year, month, dekadal, and the 13 weather variables, the variable natural order representing the first square is year-dekadal, a variable that show the sequential order of the register.

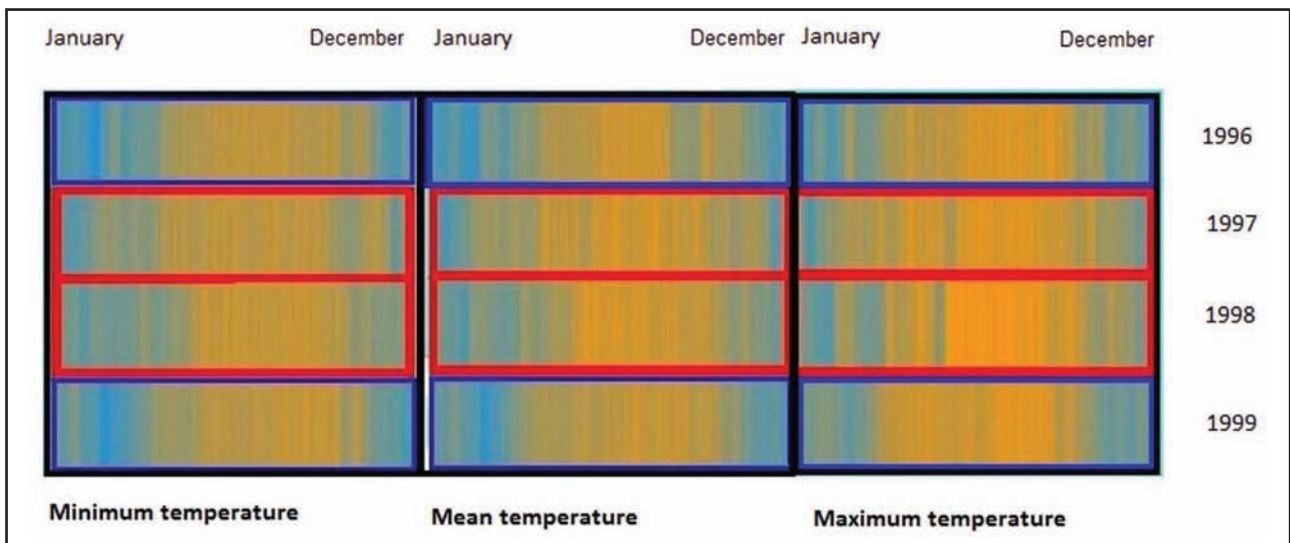


Figure 8 Dekadal minimum, mean and maximum temperature using recursive pattern (36,1)(1,4) from 1996 to 1999 in Santo Domingo weather station. The years 1997 and 1998 are highlighted in red. The color of each variable range from a dark blue (low values) to dark orange (high values).

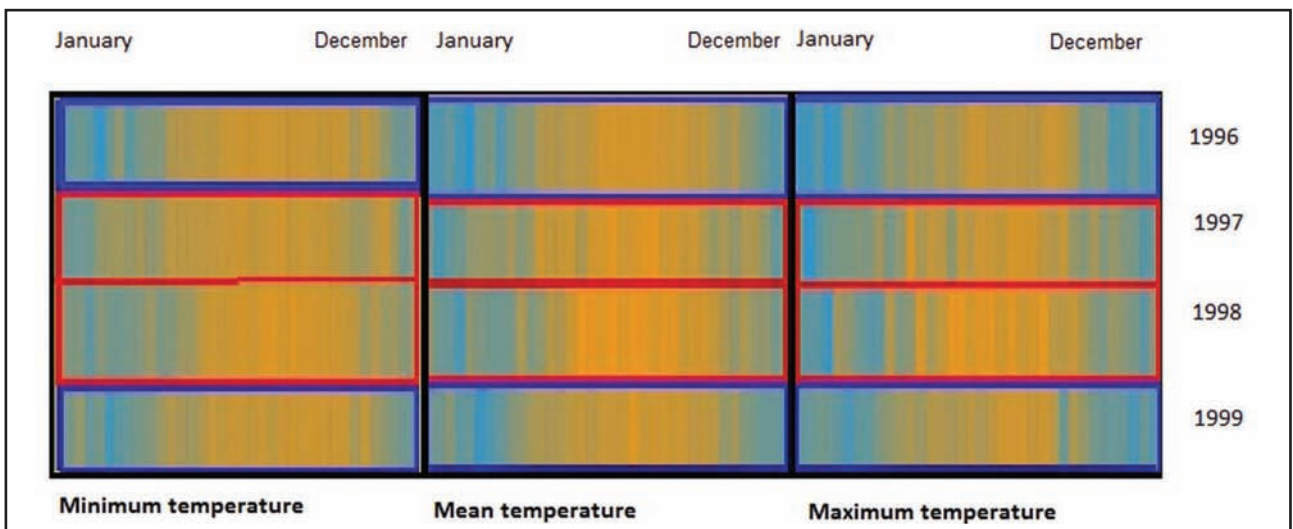


Figure 9. Dekadal minimum, mean and maximum temperature using recursive pattern (36,1)(1,4) from 1996 to 1999 in Sagua weather station. The years 1997 and 1998 are highlighted in red. The color of each continues variables range from a dark blue (low values) to dark orange (high values).

- The maximum and mean temperature are higher in 1997 and 1998.
- The maximum and mean temperature in 1997 were higher early in the year.
- In 1998 temperatures were higher during the warmer months.

Figure 9 shows the dekadal minimum, mean and maximum temperature using recursive pattern (36,1)(1,4) from 1996 to 1999 in Sagua weather station. Each square represents a variable where the

minimum value of the variable is codified in blue color and the maximum with an intense orange color. The internal rectangles represent the same data than in the previous figure but for Sagua weather station. So the red and blue rectangles also represent the same years. That figure shows that in Sagua station:

- The maximum and minimum temperatures are higher in 1997 and 1998
- Temperatures in 1997 were higher early in the year



Figure 10 Dekadal minimum temperature using recursive pattern (36,1)(1,30) from 1977 to 2007. Caibarién is the weather station at the right side. See the orange color more intense in the selected months in this station than in the others.

The minimum temperature in the first three months of the year is usually higher in Caibarién than in all other stations. This is because this place is a coastal area. (See Figure 10).

5 Conclusions

In this paper the basic ideas about the ScVis-GIS integration are presented, emphasizing on multiparameter data visualization techniques. A real case study with meteorological data from Villa Clara was presented, where several visualization techniques were implemented, and some implementation alternatives were discussed. A ScVis module was developed and incorporated into a GIS (gvSIG) by implementing the merged approach of integration between these disciplines. The developed module includes several visualization techniques, and offers a wide range of possibilities for interaction with the data.

It should be noted that, although this study was carried out to solve a real problem of visual analysis of meteorological data in Villa Clara province, the developed tools are general. They allow the reading, viewing and analysis of any kind of tabular data stored in *ARFF* and *DBF* format. Future work will

focus on extending the developed module so as to allow the reading of data in other formats. Also, new multiparameter data visualization techniques will be added to the module.

6 References

- Amant, R.S. & Cohen, P.R. 1998. Intelligent support for exploratory data analysis. *Computational and Graphical Statistics*, 7(4): 545-558.
- Andrews, D.F. 1972. Plots of high dimensional data. *Biometric*, 28: 125-137.
- Beddow, J. 1990. Shape coding of multidimensional data on a microcomputer display. In: 1990 IEEE CONFERENCE ON VISUALIZATION, 1990 Los Alamitos, CA.: IEEE Computer Society Press, 238-246.
- Bergeron, R.D. & Grinstein, G. 1989. A Reference Model for the Visualization of Multidimensional Data. In: PROCEEDINGS EUROGRAPHICS '89, 1989 Hamburg, p.393-399.
- Brodie, K.W.; Carpenter, L.A.; Gallop, J.R.; Hubbard, R.J.; Mumford, A.M.; Osland, C.D. & Quarendon, P. 1992. Scientific Visualization. *Springer-Verlag*. 284 p.
- Burrough, P.A. & McDonnell, R. 1998. *Principles of geographical information systems*, Oxford University Press Oxford. 936 p.
- Chen, C.-H.; Härdle, W. & Unwin, A. (eds.). 2008. *Handbook of Data Visualization*, Berlin: Springer-Verlag. 936 p.
- Chernoff, H. 1973. The use of faces to represent points in k-dimensional space graphically. *Journal of the American Statistical Association*, 68: 361-368.

- Cleveland, W.S. 1993. *Visualizing Data*. Hobart Press, 1993 Summit New Jersey. 360 p.
- 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.
- Cui, Q.; Ward, M.O. & Rundensteiner, E.A. 2005. Enhancing Scatterplot Matrices for Data with Ordering or Spatial Attributes. Available: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.107.3157&rep=rep1&type=pdf>.
- Eick, S.G. 2000. Visualizing Multi-Dimensional Data. *ACM SIGGRAPH*, 34(1):61-67.
- 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. *Computers, Environment and Urban Systems*, 26(4): 267-292.
- Hansen, C.D. & Johnson, C.R. 2005. *The visualization handbook*, Elsevier. 962 p.
- Inselberg, A. & Dimsdale, B. 1990. Parallel coordinates: A tool for visualizing multi-dimensional geometry. In: VISUALIZATION 90, 1990 San Francisco, p. 361-370.
- Kandogan, E. 2000. Star coordinates: A multi-dimensional visualization technique with uniform treatment of dimensions. In: PROCEEDINGS OF THE IEEE INFORMATION VISUALIZATION SYMPOSIUM, 2000, p. 9-12.
- Keim, D.; Mansmann, F.; Schneidewind, J.; Thomas, J. & Ziegler, H. 2008. Visual analytics: Scope and challenges. *Visual Data Mining*, p.76-90.
- Keim, D.A. 2000. Designing Pixel-Oriented Visualization Techniques: Theory and Applications. In: IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS, 6, p. 59-78.
- Keim, D.A. 2002. Information Visualization and Visual Data Mining. In: IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS, 7(1): 1-8.
- Keim, D.A.; Kriegel, H.-P. & Ankerst, M. 1995. Recursive Pattern: A Technique for Visualizing Very Large Amounts of Data. In: VISUALIZATION '95, 1995 Atlanta, GA, p. 279 - 286.
- Luo, J. 2007. *The semantic geospatial problem solving environment: an enabling technology for geographical problem solving under open, heterogeneous environments*. Doctor of Philosophy. The Pennsylvania State University. 208 p.
- MacEachren, A.; Dai, X.; Hardisty, F.; Guo, D. & Lengerich, G. 2003. Exploring High-D Spaces with Multiform Matrices and Small Multiples. In: INTERNATIONAL SYMPOSIUM ON INFORMATION VISUALIZATION, Oct. 19-21 2003, *Proceedings*, Seattle, WA, p. 31-38.
- 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
- North, C. & Shneiderman, B. 2000. Snap-together visualization: can users construct and operate coordinated visualizations? . *International Journal of Human-Computer Studies*, 53: 715-739.
- Ong, H.L. & Lee, H.Y. 1996. Software report WInViz -a visual data analysis tool. *Computation & Graphics*, 20(1): 83-84.
- Rhyné, T.M. 1997. Going Virtual with Geographic Information and Scientific Visualization Available: <http://www1.elsevier.com/homepage/sad/cageo/cgviz/rhyné/rhyné.htm>.
- Rhyné, T.M. & MacEachren, A. 2004. *Visualizing Geospatial Data*. ACM SIGGRAPH 2004 Course #30. 141 p.
- Schumann, H. & Müller, W. 2000. *Visualisierung Grundlagen und allgemeine Konzepte*, Springer Verlag, Heidelberg. 370 p.
- 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.
- Treinish, L. A. 1992. Visualization Techniques for Correlative Data Analysis in the Earth Sciences. In: IEEE. COMPUTER GRAPHICS AND APPLICATIONS, p.10-12.
- Ward, M.O. 2002. A Taxonomy of Glyph Placement Strategies for Multidimensional Data Visualization. Worcester Polytechnic Institute. *Journal Information Visualization*, 1(3/4): 194 - 210.
- Wijk, J.J.V. & Liere, R.D.V. 1993. Hyperslice. In: *IEEE Visualization '93*, 1993 Los Alamitos. IEEE Computer Society Press, p. 119-125.
- Xie, Z.; Huang, S.; Ward, M.O. & Rundensteiner, E. A. 2006. Exploratory Visualization of Multivariate Data with Variable Quality. In: IEEE SYMPOSIUM ON VISUAL ANALYTICS SCIENCE AND TECHNOLOGY 2006, p. 183-190.