Visual Exploration of Spatio-Temporal Databases

MILTON HIROKAZU SHIMABUKURO[1,2]
VINÍCIUS MARQUES ALVES BRANCO[1]
MARIA CRISTINA FERREIRA DE OLIVEIRA[1]
EDILSON FERREIRA FLORES[2]

[1]ICMC - USP, Av. Trabalhador São-Carlense, 400, C.P. 668, 13560-970 São Carlos, SP, Brasil
[2]FCT - Unesp, Rua Roberto Simonsen, 305, 19060-900 Presidente Prudente, SP, Brasil
miltonhs@prudente.unesp.br
vinicius@icmc.usp.br
cristina@icmc.usp.br
efflores@prudente.unesp.br

Abstract. Visualization research deals with the use of graphical models to represent data, coupled with suitable interaction operations that support an active user exploration of the data representations. Visualization techniques can greatly enhance knowledge discovery processes involving geo-referenced data, and the study of visual displays to assist users of geographic information, which typically includes spatial and temporal attributes, motivates developments in the new field known as Geovisualization. Users of such data are typically interested in spatial dynamics, or changes that occur over time, and the capability of simultaneously depicting the temporal component of data and the spatial information, in an integrated manner, is necessary to support real data analysis tasks. In this paper we introduce several strategies to support visual exploration of spatio-temporal databases, focusing on assisting end users on the initial stages of a knowledge discovery process. The visual approaches proposed are illustrated with an application that demands the exploration of a pluviometric database.

1 Introduction

Recent developments in Scientific and Information Visualization [21] are greatly impacting the geographic information sciences. Visualization research deals with the use of graphical models to represent data, coupled with suitable interaction operations that support an active user exploration of the data representations [5,13]. A new field of Geographic Visualization (*GVis*), or Geovisualization, has emerged [21] which spans both Scientific and Information Visualization, and deals with the effective use of visual displays to assist users of geographic information. Although the early focus of geographical visualizations was mainly on dynamic maps and cartographic representations, additional visual representations are necessary to enhance and complement the use of maps. Georeferenced data, such as those contained in environmental data sets, usually include both spatial and temporal attributes. Users of geographic information are typically interested in spatial dynamics, or changes that occur over time. Depicting the temporal component of data simultaneously with the spatial information, in an integrated manner, is necessary to support real data analysis tasks. Data selection, filtering and/or quality assessment are necessary activities prior to the application of 'conventional' statistical methods or mining techniques as part of a KDD (*Knowledge Discovery in Databases*) process [7,8]. Because raw data pre-processing is a critical activity when building knowledge from data, offering visual support to typical user tasks in this context may greatly enhance user effectiveness and productivity. Visualization techniques have a role to play in a knowledge discovery process, as they may be tailored to support either data pre-processing or the mining stages, making user participation easier, simplifying interpretation of the results and rendering the process more understandable [2,20].

In this paper we focus on visual support to the initial stages of a knowledge discovery process or a data analysis process, and introduce several strategies to support visual exploration of databases characterized by including attributes of spatial and temporal nature. We assume that many user exploration tasks are driven by these attributes, and thus focus on providing visual support to these tasks. The proposed strategies are illustrated with an application involving a pluviometric database provided by the DAEE – *Departamento de Águas e Energia Elétrica do Estado de São Paul*o (*Water and Energy Department*). This database contains rain precipitation measures collected in different points of the state of São Paulo, and one of its uses is in deriving climate features classification models for the state of São Paulo. This work is part of a research project funded by FAPESP, named *InfoVis*, whose goal is to investigate strategies for making Information Visualization techniques more effective and provide an integrated and accessible tool to support visual data exploration of large datasets [19,4,22]. The motivation for proposing the strategies introduced in this paper was the identification of many difficulties associated with processing the DAEE database for input in further steps of the climate modeling process.

This paper is organized as follows. In Section 2 we review related work dealing with the visualization of data with spatial and temporal attributes. In Section 3 we identify two typical tasks faced by a user of a geographical database containing data measured at geographical locations over time, and describe in general terms the visual strategies conceived to support such tasks. These strategies integrate conventional spatial visualizations based on maps with dynamic visual queries and innovative temporal visualizations. Section 4 actually introduce these strategies and describe how they were applied in the visual exploration of the DAEE database, simplifying the execution of multiple activities requiring information on the spatial location and temporal distribution of measurements collected in multiple collection stations. Conclusions and further work are in Section 5.

2 Visualization of Spatio-Temporal Data

In this section we discuss several research initiatives that investigate and implement visualization techniques to support analysis of data with spatial and temporal attributes. The goal is not to provide an exhaustive list of techniques or systems, but rather to identify different approaches to assist domain specialists in processing and analyzing spatio-temporal data, their advantages and limitations.

Visualization of spatial data is widely employed in Cartography. From 1993 the International Cartographic Association started paying special attention to establishing connections with related disciplines, Scientific Visualization in particular. A Commission on Visualization has been created that evolved, in 1999, into a Commission on Visualization and Virtual Environments[1] A list of research themes elaborated by this commission includes as a major topic the integration of visual and computational approaches into tools to support knowledge acquisition from spatial data, with emphasis on exploratory tools that do not require pre-defined hypotheses [16]. MacEachren et al. [15] believe that visualization has a role to play in all stages of the KDD process, and discuss the integration of Geographic Visualization and KDD into visual exploration tools to assist knowledge building from great volumes of spatio-temporal data. Geographically referenced data with spatial attributes may come, for example, from earth monitoring systems and environmental studies. Repeated observations of the environmental processes produce the temporal attribute. These authors argue that the difference between GVis and KDD lies in the emphasis placed by the former on the human visual system, versus the emphasis on the automated computational approaches by the latter.

Research at the GeoVISTA Center (Geographic Visualization Science, Technology, and Applications Center[2]), from the Pennsylvania State University, is focused on developing user-centered methods and technologies to solve problems involving geo-spatial data using visual exploration and analysis. GeoVista Studio is an environment that allows end users to create visualization applications with a visual programming scheme: users define applications by connecting available JavaBeans components. Dynamically linked visual representations such as maps, scatterplot matrices and parallel coordinate visualizations may be used for exploration and analysis. Component based programming allows incorporating JavaBeans components developed by third parties [17].

SPIN!, Spatial Mining for Data of Public Interest[3] is a project at the Fraunhofer Institute for Autonomous Intelligent Systems[4] targeted at proposing innovative approaches for analyzing geo-referenced data. Their solution integrates capabilities of spatial data mining techniques with those of a Geographical Information System (GIS) for interactive visual exploration. Implementation also uses JavaBeans and a client-server architecture, thus meeting requirements of scalability and platform independence. This project uses the Descartes system[5], which automatically generates thematic maps from heuristics provided for interactive visualization [18]. Integrating visualization with GIS's also offers a solution for exploratory analysis of spatially referenced data, an approach exemplified by the integrated use of XGobi, a system for multivariate data visualization, with ArcView, a geographical information system [25,26]. Using a different approach, GeoMiner, a prototype at the Simon Fraser University[6], employs visualizations in the form of maps, diagrams and tables to show the results of mining processes [10].

DEVise (Data Exploration and Visualization)[7], offers configuration mechanisms and creates visualizations. A mapping model defines how to transform data into visual

[1] www.geovista.psu.edu/sites/icavis/

[2] www.geovista.psu.edu

[3] www.ccg.leeds.ac.uk/spin/overview.html

[4] ais.gmd.de/index.en.html

[5] allanon.gmd.de/and/descartes.html

[6] db.cs.sfu.ca

[7] www.cs.wisc.edu/~devise

representations, and an end user may generate interactive coordinated visualizations. Supported visual representations include images, bar diagrams and scatter plots. Applets coupled to the environment allow Web-based interaction with the visualizations produced, making them accessible from different computational platforms [14,27].

The visual strategies proposed in this paper are targeted specifically at assisting the execution of initial exploration tasks on raw data containing attributes of spatial and temporal nature. They are being integrated into a multi-platform general-purpose visualization system that offers specific support to the visual exploration of this type of data [22,23,4], in addition to other techniques targeted at high-dimensional data. This system is being implemented in the Java Platform, ensuring features such as platform independence, easy customization to different domains and users, and extensibility by the gradual incorporation of components.

3 Integrated Strategies for Visual Exploration of Spatio-Temporal Data

The strategies proposed are targeted at the visualization of spatio-temporal databases, being suitable for situations in which a single spatial localization is associated with a range of temporal value(s) of interest. A typical example is a database produced by an earth monitoring system, containing measurements collected over time at the same set of locations (spatial positions). Our visual strategies combine well-known general-purpose interaction and visualization techniques in a manner to ensure effective user-driven exploration in this context. Their novelty is on the ability of assisting the integrated analysis of both the spatial and temporal data components simultaneously.

To illustrate the fact that such an integrated analysis is required in many situations, let us consider two general tasks typically associated with the exploration of this kind of data. The first one (Task 1) involves checking spatial positioning of elements of interest, in order to verify spatial proximity amongst different elements and verify their spatial density. The second one (Task 2) involves obtaining an overview of how a target value measured at one particular spatial location, or at various neighboring locations, varies over time, maybe with the goal of comparing this variation across multiple spatial locations. Note that both tasks are likely to be executed simultaneously in an integrated manner. A user might, for example, want to analyze and compare data quality across multiple nearby locations sharing common characteristics. Execution of such tasks is greatly improved by visual representations capable of depicting (1) the relative spatial positioning of data elements, coupled with (2) an overview of the temporal distribution of target values of interest associated with each spatial location or with a set of neighboring locations.

Visual strategies are thus introduced to support the integrated execution of the above tasks. To provide the spatial localization required in (1), graphical markers (e.g., dots, triangles, circles, or any other) representing the different spatial locations are positioned and displayed over a background visual spatial reference, usually a map of the relevant region. As important as displaying spatial positioning is to give users the ability to select the data elements at the spatial locations in which s/he is interested, focusing on regions of interest. As focusing may be based in the spatial location and other (maybe non-spatial) attributes, users should have freedom to select elements of interest based on the values of multiple attributes. Dynamic Queries (DQ) [24], also known as Range Queries, offer a powerful filtering mechanism in which visual controls are associated with attributes of the database, allowing the visual specification of queries over the database. Associated with map visualizations, they are naturally useful for querying geographical data to convey a direct observation of the spatial relationships amongst elements [1,24]. This allows a user to select attributes and attribute ranges dynamically for visualization, and then activate other visual strategies to inspect temporal information associated with one or multiple spatial locations.

To provide the temporal information required in (2), we introduce various 'pixel-based' temporal visual representations. Pixel-based visualizations map an attribute value to a pixel in the screen that is colored according to a pre-defined color map. Pixels are arranged in the screen according to a certain criteria, and multiple attributes may be displayed in multiple sub-windows of rectangular or circular shape [11,12,3]. Pixel-based visualizations are particularly suitable to convey temporal information if the arrangement of the pixels in a rectangular window is properly chosen. We suggest three visualizations that use different pixel arrangement criteria to convey temporal distribution of a data attribute of interest: time interval status, multiscale temporal behavior and yearly temporal behavior. In these representations the data attribute of interest associated to the target data values may be measurement values, data quality, or other, and the one chosen attribute is mapped to the pixel color, whereas pixel positioning conveys the temporal variation. The visualization user (the domain specialist) must have the freedom to define a suitable color map. For example, a discrete attribute assuming *n* different values may be mapped to a discrete color map with *n* color hues, whereas a continuous attribute may be mapped to varying intensities of the same hue or to different intensity ranges of multiple hues.

The temporal and spatial visual strategies outlined in this section are presented in Section 4, where their use is illustrated with a case study. They were implemented as components of a Java visualization environment that allows the integrated and simultaneous use of multiple visualizations. Thus, users can display temporal information of interest in relation with spatial location, with ´linked´ interaction: user actions in one visual representation, e.g., data selection, are reflected on the remaining ones, a strategy known as linking [6].

4 A Case Study with the BcDAEE Database

In this Section, the application of proposed visual strategies in a problem of processing a pluviometric database is described.

Problem Description

The proposed strategies have been applied in a real problem, in which data from the BcDAEE relational database is used to derive a climate model for the western region of the state of São Paulo [9]. BcDAEE 1.0 – The Pluviometric Database of the State of São Paulo, version 1.0 – contains daily precipitation measurements collected in 1.660 stations located in the state of São Paulo from 1888 to 1997. The time interval for which data is available is not the same for all stations, as not all of them started operating at the same time, some became inactive at some point, and others had data collection temporarily interrupted over different periods. In addition to missing (non-collected) data, stations may have non-consistent data, which have not been checked for validity.

Climate modeling is a complex process that comprises many steps and tasks, and makes extensive use of multivariate statistical data analysis techniques. Flores (2000) used GIS and geostatistical krigging techniques to explain spatial and temporal variations of rain precipitation patterns on the western region of the state of São Paulo. An initial step of the modeling process requires grouping years into representative raining patterns, namely dry, rainy, or normal. This has been done from the BcDAEE data, and involved three initial tasks: (a) selecting relevant collection stations, i.e., those that contain enough reliable data over a suitable time interval; (b) treating missing and extraneous data within stations; and (c) applying a clustering algorithm to identify groups of representative years. A brief description of these tasks follows:

a) Selecting collection stations. A huge table has been manually constructed from the BcDAEE data with table lines describing a station and table columns describing years. Each table cell thus describes a pair (station,year), and was labeled with either an "X" for consistent data, an "*" for non-consistent data, or a blank for missing data. The climatologist then inspected the table searching for stations labeled "X" or "*" and containing measurements over a common interval of years. Initially 200 stations have been selected in this process, closer inspection reduced the number to 150;

b) Handling Missing and Extraneous Data. Two different approaches were adopted to handle a station with missing precipitation values: (1) a missing data is filled in with the precipitation value collected in the corresponding month of another year with similar precipitation behavior, within the same station; or (2) a missing data is filled in with a value interpolated from the corresponding month in neighboring stations. In the last approach information on the spatial position and altitude of stations is relevant for a more informed decision on which stations to use in the interpolation. Extraneous data are handled similarly;

c) Identifying representative years (dry, rainy, normal). A hierarchical clustering technique was employed, for each one of the selected stations, to group the years based on the total yearly precipitation value. From these groupings, one representative year for each class (rainy, dry or normal) is selected for use in the modeling process. Several levels of clustering are possible, depending on the loss of information admitted in the process.

Data preparation prior to input into the GIS software is an important procedure in the modeling process, and it is extremely time consuming and painful, as the climatologist has to handle a huge volume of data represented in conventional data tables. The amount of data renders conventional approaches extremely ineffective. To make the problem more manageable one approach is to consider data at monthly, rather than daily levels. However, the ability to handle data at the daily level greatly impacts the quality of the resulting model.

Visual Strategies

The visual representations outlined in Section 3 were employed to assist the climatologist in the procedures involved in the initial treatment of the database. Tasks (a) to (c) previously described were redone by the climatologist using the interactive visual representations. In this particular case, visualizations were tailored to depict (1) the relative spatial positioning and altitude of the stations; and (2) an overview of the temporal precipitation data with respect to data quality, missing data and interval of years for which data is available, both for a particular station and for a set of neighboring stations. In the following we describe the major steps executed by the climatologist – the domain specialist – when exploring the BcDAEE database with support of the visual strategies provided. The domain specialist received a short training on the visual strategies in order to understand them and to learn the basic interaction actions. His interaction with the visualization tools when accomplishing the tasks was monitored by the visualization specialist, in order to identify difficulties and to validate the visual strategies and their implementation.

Due to access limitations of the original database, data were exported into a set of files, each one containing complementary information about each station and the daily measurements collected at the station. Stations have the following attributes: *Prefix*, *Name*, *City*, *Hydrographic Basin*, *Altitude*, *Latitude*, *Longitude*, *Initial Year* (of the collection), *Final Year* and *Interval* (range of years with consistent data). The need of this initial data preparation step illustrates the difficulties in offering a generic, domain independent solution for data pre-processing, as the syntax and semantics of the source databases vary considerably.

Once the data is available, the first user task was to select collection stations in the region of interest to the study, based on an assessment of the quality of their data and on the range of years with available measurements. The initial action is to use the spatial visualization that plots stations over the map, using the DQ interface to dynamically manipulate the map visualization based on the values of the spatial attributes *Latitude* and *Longitude* and the attribute *Interval* (range of years with consistent data), as illustrated in Figure 1[8]. By inspecting these attributes, the analyst verifies that only a few stations in the region of interest – delimited by *Latitude* and *Longitude* – have a collection interval greater than 30 years (this is the minimum interval length recommended by the *World Meteorological Organization*). The bigger map shows all the stations on the region of interest, and the small rectangle at the bottom depicts those stations with collection intervals within 30 and 51 years.

He thus considers the possibility of selecting stations based on the total year collection interval disregarding whether data is consistent or not, and treating extraneous and missing values later – these may be replaced with plausible values collected at similar time and/or similar spatial positions and which are known to be consistent. Figure 2 shows visual representations used to investigate this alternative approach. One visual representation is still the map, used to select stations according to their positions (the inner rectangle in the map delimits selected stations, shown highlighted; in the color view[8] they are shown with a thin vertical mark in cyan). Each point or station on the map is colored according to the value of a chosen attribute – in this case the attribute is *Altitude* and thus pixel color represents station altitude.

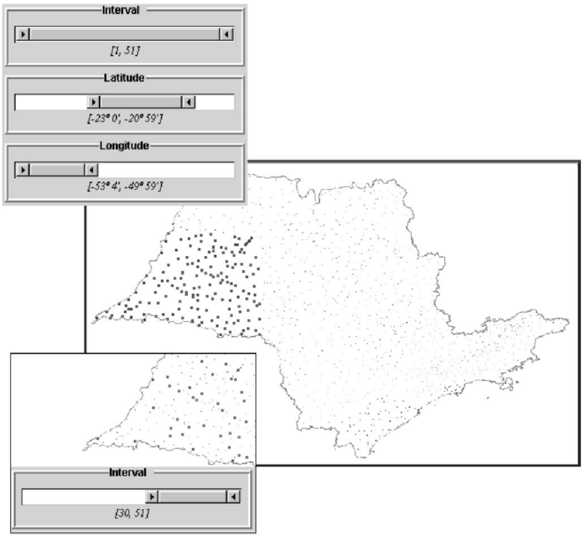

Figure 1 Examining interval length with consistent data for selecting stations

The other visual representation shown exhibits information about the collection interval for each station along with station data quality. A horizontal line is associated with each station, and the set of lines (stations) is displayed in multiple columns, organized from top to bottom, left to right, ordered by station's prefix. Each point (pixel) on a line represents a specific year – the leftmost pixel corresponds to the first collection year, the rightmost pixel corresponds to the last one. Note that the initial collection year varies across stations, which is the reason for lines having different starting (i.e., leftmost end) points. The length of each line, defined by the *Initial* and *Final Year* attributes, maps the length of the overall collection interval, and each line is initially painted with a user selected hue (cyan in the color views, light gray in the black and white view). Within each station, years with consistent data – as given by the *Interval* data attribute – are mapped to a different pixel color (dark blue in the color views, darkest gray in the black and white view). Lines spanning the whole column width, prolonged further to the left, indicate user selected stations – here they have been selected in the map visualization, and hence are automatically highlighted in both visualizations

[8] Color figures with higher definition are available at www.nec.prudente.unesp.br. Although it is not necessary to interpret Figure 1, color is required to interpret Figures 2 to 4.

(highlighted lines shown in magenta in the color view, an in mid-intensity gray in the black and white view). A thin vertical line (yellow in the color figure) may be positioned in the columns to provide a visual reference for a given year (the box at bottom left in the figure indicates that the vertical lines mark year 1971). One may observe in the figure that, from that year on, most selected stations have a common interval of years with consistent rain measurements, indicating that precipitation data collected within this year interval is a good choice of input to the next climate modeling steps. Reliability of further steps in the process is highly dependent on the quality of the input rain measurement data.

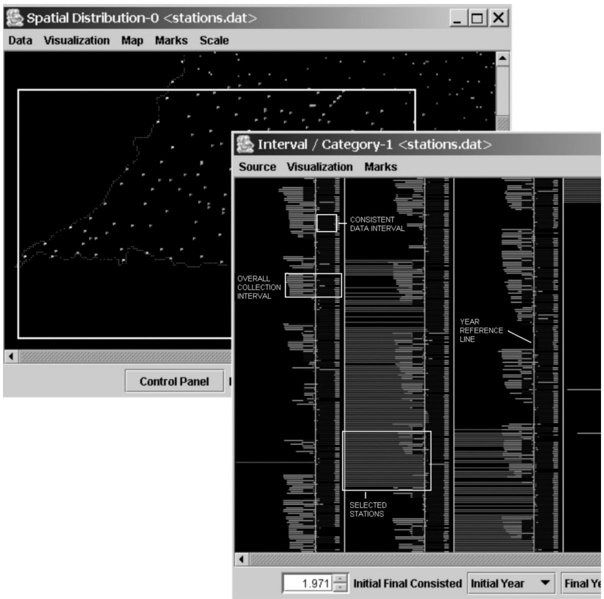

Figure 2 Analyzing which stations within the region of interest have good collection periods

A pair of maps depicting stations spatial localization, one mapping station's altitude to color and the other mapping the total length of the data collection interval to color, can help the analyst to refine the initial station selection process, reducing the number of stations for further processing. A simple rule adopted by the domain specialist is: "if two or more neighboring stations are within a certain altitude range, the one with greater amount of collected data may be chosen as representative of the area". This is not a final decision, as further steps can lead to different choices.

The next step is to treat missing and extraneous values within stations, replacing such values with reasonable approximations. This requires inspecting data collected within a station and in its neighboring stations for plausible replacement values. Visual support to this task is offered by the visual representation shown in Figure 3. Precipitation values measured over all the collection period at one particular station are visually displayed simultaneously in three scales: daily, monthly and yearly. In the figure each column represents one specific year, and yearly precipitation measures are organized from top to bottom in a single column as follows: the first twelve rectangles render daily precipitation values (color maps precipitation intensity), with values positioned within the rectangle so as to mimic a calendar; the following twelve rectangles indicate the total monthly precipitation values (total precipitation intensity mapped to color); and the last rectangle displays the accumulated precipitation value for that year. Each scale uses a different, user defined, color table. Missing values in the raw data are mapped to an arbitrary color (black in this case). As the goal is to identify plausible values for missing or extraneous data inspecting values collected in the same station or in neighboring ones, multiple instances of this visual representation may be created for simultaneous inspection by the user. Visualizations corresponding to different stations are positioned on the screen so as to reflect their relative geographical positioning – these visualizations may also be inspected jointly with an underlying map visualization depicting station spatial positioning.

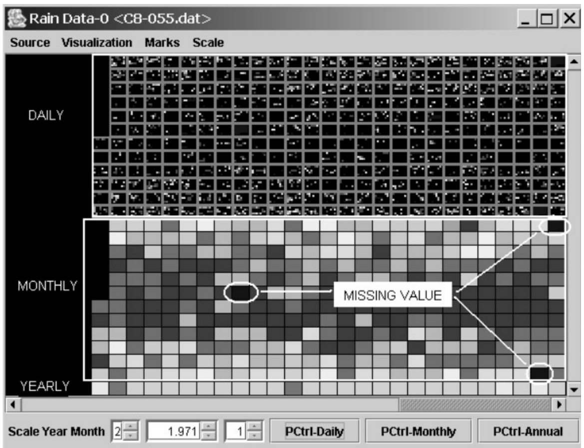

Figure 3 Multiscale visual representation for handling missing and extraneous values

After selecting stations and correcting missing and extraneous values, the user proceeds to determining representative years. To support this latter task, the visual representation defined is a generalization of the one used for displaying data collection intervals (depicted in Figure 2), with a different arrangement. Now, each horizontal line is associated with a specific year, and each point (pixel) in the line depicts a station. Thus, rather than representing data quality status for one particular measurement, as in Figure 2, each pixel on the line represents the accumulated yearly precipitation value, for that particular year, at one station. Though not implemented in our visualization environment, this representation, illustrated in Figure 4, was tested by adapting a visualization available in the *GGobi*[9] software. A suitable color mapping can visually uncover years that are representative of rainy, dry and normal conditions. The color mapping chosen by the analyst in this case is shown in Figure 4. One observes, in the color version, that the predominance of: (1) redish to orange-ish hues indicates a predominantly dry year (such as 1985), (2) orange-ish to yellowish indicate normal (1984), and (3) yellowish to greenish indicate a rainy year (1983) (the years within the rectangle). The representative years identified visually match those previously obtained with a clustering algorithm [9]. Furthermore, from the visual representations the analyst realized that other choices of representative years could have been made that would be equally suitable.

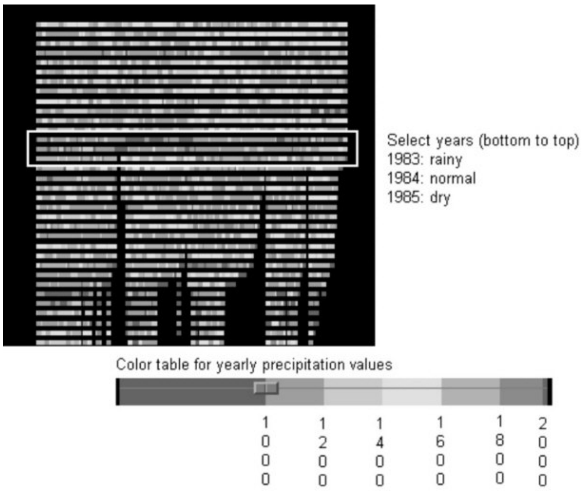

Figure 4 Visual representation to identify representative rainy, dry and normal years

The visual exploration approaches allowed the user to handle a greater amount of data and made it easier to incorporate in the analysis multiple features such as spatial distribution, range of years with consistent data, altimetry and rain behavior. The ability of inspecting multiple data features simultaneously allowed a broader view of data from the initial stages, leading to a better understanding. This contributes to improve the modeling data to be input into GIS software for the next analysis stages. Although familiar with handling graphics, the domain specialist took a few training sessions before being able to correctly interpret the visual representations. Training time was mostly spent in demonstrating interaction capabilities, with emphasis on integrated and coordinated use of the visualizations.

5 Conclusions

Visualization is useful to assist domain specialists in exploring information in large databases, because it has the capability of successfully integrating the huge processing power of computers with the human ability of recognizing visual patterns. Even when conventional statistics and automated mining techniques will be applied in further stages, many specialists still approach a raw data exploration process using conventional table-driven approaches. The size and other characteristics of ´real´ databases do not encourage data analysts to apply alternative approaches, and one reason for this is that there are few robust (and affordable) general-purpose tools that effectively support data pre-processing tasks. We are currently working on the implementation of such an environment, adaptable to user and task requirements, and in this paper we introduce some strategies included to assist users in data pre-processing of large databases containing spatial and temporal attributes. The proposed strategies combine well-known interaction and visualization techniques to support exploration tasks driven by the spatial and temporal data attributes. Coupling multiple pixel-based visualizations with specific interaction operations to display large volumes of temporal data proved particularly effective to support comparison tasks. Interaction with domain specialists was valuable to tailor the visualizations to user tasks and preferences and to validate the strategies. Using them the specialist can handle more data (e.g., at a daily, rather than a monthly level) more effectively. These techniques are now being applied to other spatio-temporal databases, in order to identify their advantages and limitations for exploration tasks in this context.

Acknowledgments

The authors acknowledge the financial support of FAPESP (Proc. 01/07566-2), CNPq (Proc. 521931/97-5) and CAPES (V.M.A. Branco´s M.Sc. dissertation program).

References

[1] C. Ahlberg, B. Shneiderman, “Visual Information Seeking: Tight Coupling of Dynamic Query Filters with Starfield Displays”, *Human Factors in Computing Systems: Proc. CHI’94*, New York, NY, 1994, 313--317.

[2] M. Ankerst, “Visual Data Mining”, *Ph.D. Thesis*, University of Münich, Germany, November 2000, 216p.

[3] M. Ankerst, D.A. Keim, H.P. Kriegel, “Circle Segments: A Technique for Visually Exploring Large

[9] www.ggobi.org

Multidimensional Data Sets". *Proc. IEEE Visualization'96*, Hot Topic Session, San Francisco, CA, 1996.

[4] V.M.A. Branco, "Visualization to Support Exploration of a Pluviometric Database", *M.Sc. Dissertation*, ICMC-USP, São Carlos, April 2003. (In Portuguese)

[5] S.K. Card, J.D. Mackinlay, B. Shneiderman (eds.), *Readings in Information Visualization – Using Vision to Think*, Morgan Kaufmann, San Francisco, CA, 1999.

[6] W.S. Cleveland, *Visualizing Data*, Hobart Press, Summit, NJ, 1993.

[7] U.M. Fayyad, G. Piatetsky-Shapiro, P. Smith, "From Data Mining to Knowledge Discovery: An Overview", U.M. Fayyad et al. (eds.) *Advances in Knowledge Discovery and Data Mining*, AAAI Press and MIT Press, 1996, 1--34.

[8] U.M. Fayyad, "Mining Databases: Towards Algorithms for Knowledge Discovery", *Data Engineering Bulletin*, 21(1) (1998), 39--48.

[9] E.F. Flores, "Modeling in Geographic Climatology: A Methodological Essay Applied to the Western Region of São Paulo", *Ph.D. Thesis*, Instituto de Geociências e Ciências Exatas, Unesp, Rio Claro, March 2000. (In Portuguese)

[10] J. Han, K. Koperski, N. Stefanovic, "GeoMiner: A System Prototype for Spatial Data Mining", *Proc. 1997 ACM-SIGMOD Int. Conf. on Management of Data*, Tucson, Arizona, May 1997. (System prototype demonstration).

[11] D.A. Keim, "Visual Database Exploration Techniques", Tutorial KDD'97 *Int. Conf. on Knowledge Discovery and Data Mining*, Newport Beach, CA, 1997.

[12] D.A. Keim, "Designing Pixel-Oriented Visualization Techniques: Theory and Applications", *IEEE Trans. on Visualization & Computer Graphics*, 6(1) (2000), 59--78.

[13] D.A. Keim, "Information Visualization and Visual Data Mining", *IEEE Trans. on Visualization & Computer Graphics*, 8(1) (2002), 1--8.

[14] M. Livny, R. Ramakrishnan, K. Beyer, G. Chen, D. Donjerkovic, S. Lawande, J. Myllymaki. K. Wenger, "DEVise: Integrated Querying and Visual Exploration of Large Datasets", *Proc. ACM-SIGMOD Int. Conf. on Management of Data,* May 1997.

[15] A.M. MacEachren, M. Wachowicz, R. Edsall, D. Haug, R. Masters, "Constructing Knowledge from Multivariate Spatiotemporal Data: Integrating Geographic Visualization (GVis) with Knowledge Discovery in Database (KDD) Methods", *Int. J. of Geographic Information Science*, 13(4)(1999), 311--334.

[16] A.M. MacEachren, M-J. Kraak, "Research Challenges in Geovisualization", *Cartography & Geographic Information Science*, 28(1) (2001), 3--12.

[17] A.M. MacEachren, F. Hardisty, M. Cahegan, M. Wheeler, X. Dai, D. Guo, M. Takatsuka, "Supporting Visual Integration and Analysis of Geospatially-Referenced Data Through Web-Deployable, Cross--Platform Tools". *Proc. National Conf. for Digital Government Research*, Los Angeles, CA, May 21-23, 2001.

[18] M. May, A. Savinov, "An integrated platform for spatial data mining and interactive visual analysis", Proceedings of Data Mining 2002, *3rd Int. Conf. on Data Mining Methods and Databases for Engineering, Finance and Other Fields*, Bologna, Italy, September, 2002.

[19] M.C.F. de Oliveira, "Projeto InfoVis: Visualização de Informação Aplicada a Dados de Comércio Eletrônico e Climatologia", FAPESP 01/07566-2, *Scientific Report nº 01*, September 2002. (in Portuguese)

[20] M.C.F. de Oliveira, H. Levkowitz, "From Visualization to Visual Data Mining: A Survey", *IEEE Trans. on Visualization & Computer Graphics*, 9(3) (July/September 2003), 37--394.

[21] T.M. Rhyne, "Does the Difference between Information and Scientific Visualization Really Matter?", *IEEE Computer Graphics & Applications*, 23(3) (May/June 2003), 6--8.

[22] M.H. Shimabukuro, "Visual Approaches to Support Exploration and Analysis of Spatial Data", *Ph.D. Proposal*, ICMC-USP, São Carlos, September 2001. (In Portuguese).

[23] M.H. Shimabukuro, V.M.A. Branco, M.C.F. de Oliveira, "InfoVis – Description of the Architecture and Implementation Aspects", *Technical Report No. 178*, ICMC-USP, November 2002, 24p. (In Portuguese)

[24] B. Shneiderman, "Dynamic Queries for Visual Information Seeking", *IEEE Software*, 11(6) (1994), 7--77.

[25] D.F. Swayne, D. Cook, A. Buja, "XGobi: Interactive Dynamic Data Visualization in the X Window System", *J. Computational & Graphical Statistics*, 7(1) (1998).

[26] J. Symanzik, D. Cook, N. Lewin, J.J. Majure, I. Megretskaia, "Linking ArcView 3.0 and Xgobi: Insight Behind the Front End", *J. Computational & Graphical Statistics*, 9(3) (2000), 470--490.

[27] H. Yao, K. Wenger, M. Livny, "DEVise and JavaScreen: Visualization on the Web", *Proc. SPIE Conf. on Visual Data Exploration & Analysis*, January 2000, 375--384.