Computational and Visual Support for Exploratory Geovisualization and Knowledge Construction

Computational and Visual Support for Exploratory Geovisualization and Knowledge Construction

Etien Luc Koua

ISBN 90-6164-229-9
ITC dissertation number 118

c/o Faculty of Geographical Sciences
Utrecht University
P.O. Box 80115
3508 TC Utrecht, The Netherlands

Cover designed by Andries Menning

Printed by ITC Printing department

Computational and Visual Support for Exploratory Geovisualization
and Knowledge Construction

Computationele en visuele ondersteuning voor exploratieve geovisualisatie
(Met een samenvatting in het Nederlands)

Dissertation

presented for the degree of doctor
at Utrecht University under the authority of the
Rector Magnificus Prof. Dr. W. H. Gispen,
to be defended in public on Friday January 14, 2005 at 10.30 o'clock,
in accordance with the decision made by the
Board for the Conferral of doctoral degrees

By

Etien Luc Koua

Born on 10 June, 1968
in Bongouanou, Côte d'Ivoire

Promotor: Prof. Dr. Menno-Jan Kraak
International Institute for Geo-information Science (ITC)

&
Faculty of Geographical Sciences, Utrecht University

Examining committee:

Prof. dr. A. M. MacEachren, Penn State University
Prof. dr. F. J. Ormeling, Utrecht University
Prof. dr. S. M. De Jong, Utrecht University
Prof. dr. P. Verhagen, Twente University
Dr. M. J. van Kreveld, Utrecht University

To my wife Rachel, my son Christian
and my mother Alika

Acknowledgements

This thesis was made possible by the help and support of a number of people. I would like to start by thanking many staff members at ITC, and particularly those in the Department of Geo-information Processing, who contributed one way or another to the results presented in this thesis.

First of all, my deepest gratitude goes to Prof. Menno-Jan Kraak, my supervisor, who has been of great help, and his scientific guidance and reviews have tremendously improved the quality of the work presented here.

All the other staff in the department helped me in many ways. My special thanks go to my good friend Wim Feringa, who was very supportive and also designed the nice cover of my thesis. I want to thank all the other staff members of the department: François de Sousa, Jeroen van den Worm, Wim Bakker, Rolf de By, Rob Lemmens, Blanca Perez Lapena, Willy Kock, Richard Knippers, Ton Mank, Ellen-Wien Augustijn, Marijke Smit, Connie Blok, Corné van Elzakker, Andreas Wytzisk, Yuxian Sun, Jantien Stoter and Martin Ellis; and of course my fellow PhD students in the department: Trias Aditya, Ulanbek Turdukulov and Arko Lucieer.

The support of the ITC library staff was very important for gaining access to scientific resources inside and outside ITC. I would like to thank Carla Gerritsen, Marga Koelen and Petry Maas–Prijs for their help and support. The same gratitude goes to Ard Blenke, who was always available to provide support regarding tools and equipment.

I would like to acknowledge the help from Prof. Skidmore of the Department of Natural Resources and Prof. John van Genderen of the Department of Earth Observation Science, who were always available to discuss many aspects of my work.

The financial support I received from the ITC/DGIS fund enabled me to conduct my research at ITC. The wonderful support of the research coordination office, especially Dr. Elisabeth Kosters, Loes Colenbrander and Prof. Martin Hale, was crucial to my completing the work and in facilitating my stay at ITC in general, as was the support of the Education Affairs Department, with special thanks to Marie-Chantal Metz, Teresa Brefeld and Bettine Geerdink.

Many people outside ITC also supported me during my research. I would like to thank Prof. Alan MacEachren of Pen State University for his rigorous reviews of some of the papers produced during my research and many parts of this thesis. I have learned a lot from his long comments and suggestions that always came in response to my many questions about my work and any ideas that I was putting in. Many thanks also go to Dr. Sara Fabrikant of the University of California at Santa Barbara, Dr. André Skupin of the University of New Orleans, Dr. Bill

Bertrand and Dr. Eamon Kelly of Tulane University, Tom Scialfa and Valerie Koscelnik, both with the CDC in Rwanda, and Prof. Verhagen, Jan Nielsen and Dionysia Loman of the University of Twente.

The support of fellow PhD students was very important. I particularly remember the fellows I found here, a nice PhD group, with whom I enjoyed coffee breaks and discussions. I think about Dr. Jose Campos dos Santos, Dr. Patrick Ogao, Dr. Liu Yaolin, Dr. Javier Morales, Dr. Yvan Bacic and Dr. Citlalli Lopez. My current PhD fellows at ITC made my life at ITC enjoyable: Richard Onchaga, Arta Dilo, Alfred Duker, Martin Yemefack, Moses, Peter, Masoud, Elisabeth Toe and Carlos.

I cannot finish this acknowledgement without mentioning my wife, who agreed to support me, travelling back and forth, and who helped me by accepting many duties and responsibilities back home in my absence. And finally, my mother and many of my friends, Lon Hustinx, Lynne Sinnung, Maura Merson, Amani Boro, Lucien N'gadi and Toure Ngoma, who strongly encouraged me during my research years.

Table of content

CHAPTER 7: USABILITY TESTING AND RESULTS

CHAPTER 8: CONCLUSIONS AND RECOMMENDATIONS

List of figures

List of tables

Chapter 1

Introduction

1.1. Context

The exploration of patterns and relationships in large and complex geospatial data is a major research area in geovisualization. Traditional forms of geospatial analysis can become difficult when dealing with such large and multivariate databases. Extracting patterns and understanding the underlying processes may be difficult, as certain patterns may remain hidden when using common geospatial analysis techniques. New approaches in spatial analysis and visualization are needed to represent such data in a visual form that better stimulates pattern recognition and hypothesis generation, allows better understanding of structures and processes, and supports knowledge construction.

Information visualization techniques are increasingly used in combination with other data analysis techniques. Recent efforts in knowledge discovery in databases (KDD) have provided a window for geographic knowledge discovery. Data mining, knowledge discovery and visualization methods are often combined to try to understand structures and patterns in complex geographic data (MacEachren et al. 1999; Wachowicz 2000; Gahegan et al. 2001). One way to integrate the KDD framework in geospatial data exploration is to combine the computational analysis methods with visual analysis in a process that can support exploratory and knowledge discovery tasks.

In this research, we explore the integration of computational and visual approaches, to contribute to the analysis of complex geospatial data. Computational algorithms are used in a framework for data mining, knowledge discovery and spatial analysis, as well as for uncovering the structure, patterns, relationships and trends in the data. Graphical representations supported by information visualization techniques and cartographic methods are then used to portray derived structures and patterns in a visual form that allows better understanding of the structures and the geographic processes. The use of these graphical representations (information spaces) plays a role by offering visual representations of data that bring the properties of human perception to bear (Card et al. 1999).

1.2. Geoinformation and knowledge acquisition

To be of value, data must be organized, transformed and presented in a way that gives meaning and makes them useful. From the information science perspective, the following seven ways suggested by (Jacobson 1999) are commonly used to organize anything in general: alphabet (name), location, time, continuum (value scale and order of importance), number, category, or randomly (meaning no organization). For communicating the message, each organization of the same set of data may express different attributes and information. To enable the transfer of knowledge, the patterns and meanings of the information must be assimilated. Creative manipulation of data is therefore necessary to assist our understanding (Hodge and Janelle 2000). Cognitive research suggests that, by using an experimental component such as interaction, inspection, evaluation, contemplation or interpretation, meaning and deep understanding can be constructed. Rules are often generated from these processes to form intellectual skills used in problem solving, and are applied to achieve a solution to a novel situation, based on a combination of previously learned rules (Gagné 1977). This process of encoding and subsequently entering the encoded information into long-term memory is a central and critical event in the acquisition of knowledge. Information must therefore be organized or transformed into a form that is semantic or meaningful.

1.2.1. Complexity in geospatial data

Information derived from geospatial data can be considered as a different type of information, due to their inherent structure (location, attributes and time), the semantics, and the geographic scale used (MacEachren and Kraak 2001). These characteristics are meaningful in geographic space. Information in this context can be organized in association with geographic positions in a natural and intuitive way (maps). Complexity in geospatial data analysis arises from the large volumes of data, underlying relationships, and the nature of geographic problems (Openshaw 2000; Miller and Han 2001; Gahegan and Brodaric 2002). The process of geospatial data handling consists of methods and techniques used to collect and analyze data and explore insight related to the dataset in order to solve particular problems. Traditionally, maps are the results of this process and are used to give a visual representation of an existing phenomenon. This role of maps has changed and expanded (Kraak 2000) owing to the technological capacities for data acquisition and data processing (e.g. satellite imagery), and the sophisticated nature of new visual representation techniques (e.g. 3D representation). Today, with the huge volume of data, static non-interactive maps do not satisfy the fundamental demands of exploratory data analysis (Andrienko et al. 2000). However, the many alternative interactive forms of the map that are now available and the use of dynamic links mean it can still play a key role in exploring geospatial data (Kraak 2000).

1.2.2. Geoinformation and the decision-making process

It is recognized that the use of maps and visual representations derived from geospatial analysis operations leads to decisions (Kraak et al. 1995). Decision making is one of the main goals of geospatial data analysis. It involves moving through a series of steps in order to evaluate a number of possible alternatives and decide upon a particular course of action. Malczewski (1999) suggested three phases in the geospatial decision-making process:

- *Intelligence* (searching for conditions calling for a decision)
- *Design* (finding and evaluating decision alternatives)
- *Choice (*choosing between decision alternatives).

Decision situations can be related to different levels of tasks such as mapping, monitoring, management and specific problem solving. The last two phases of the decision-making process (design and choice) may be attributed in certain conditions to human intelligence or other support tools helping the decision-making process. This relies on the first phase of the process (intelligence), in which appropriate information must be retrieved. This phase is particularly important and requires the transformation of appropriate data into relevant information for finding, evaluating and choosing decision alternatives, in a way that effectively facilitates knowledge acquisition. Effective extraction of patterns in the data is crucial to supporting the geospatial data analysis process.

1.3. Information extraction from complex geospatial data

Extracting information from large geospatial data is a major issue in GIScience research. Many efforts across such disciplines as statistics, machine learning, and database and information visualization have emphasized different aspects of exploratory analysis and knowledge discovery (Gahegan 2001) for uncovering structure within geospatial data and producing hypotheses with which to explain the patterns (Agrawal et al. 1993; Gahegan and Takatsuka 1999; MacEachren et al. 1999). Several techniques are used, including artificial intelligence and machine learning techniques (Openshaw and Openshaw 1997; Openshaw 2000; Gahegan 2000a).

1.3.1. Artificial intelligence and geoinformation science

Artificial intelligence is a domain of computer science dealing with the automation of intelligent behaviour for solving complex tasks. It is an inclusive term for several areas of computing that attempt to mimic processes that humans carry out without much conscious thought (Mallach 1994). Major areas covered by

artificial intelligence include artificial neural networks, robotics, machine vision, speech recognition, interpreting sentences in natural languages, and expert systems. In geoinformation science, there are many complex tasks related to data processing and manipulation for which a number of applications of artificial intelligence have been used. Expert systems or rule-based knowledge engineering systems have been used to design geoinformation systems (Openshaw and Openshaw 1997) to automate either highly skilled tasks such as map generation and name placement on maps in cartography, or rules for selecting good locations in complex planning situations. Neural networks are also applied in various areas of geoinformation science, including mapping, data classification and prediction.

1.3.2. Neural networks

Artificial neural networks, or neural networks, can be defined as a trainable or learning program that can be used to induce or extract a pattern of information from a structured set of data (Mockler and Dologite 1992). They are designed to model the way in which the brain performs a particular task or function of interest (Haykin 1994). The network is made up of many simple, highly interconnected processing elements that receive input signals and, based on these inputs, either generate output signals (fire) or do not. The output signal of an individual processing element is sent to many other processing elements via the programmed interconnections between processing elements. The network is built of a specified number of neurons and a specified number of connections between them called *weights*, which have certain values. What changes during the learning process are the values of these weights. Incoming information stimulates certain neurons, which pass the information to connected neurons or prevent further transportation along the weighted connections. The value of a weight will be increased if information should be transported, and decreased if not. While learning different inputs, the weight values are changed dynamically until their values are balanced, so each input will lead to the desired output. The training of a neural network results in a matrix that holds the weight values between the neurons. Once the network has been trained correctly, it can be used to find the most appropriate output to a given input that has been learned, by using these matrix values.

Like many techniques applied to data analysis (statistical methods, rule-based systems and decision trees), neural networks are an emerging solution for data analysis and pattern recognition. In geographic problems, they are found to be suitable because of their freedom from assumptions, their inherent non-linearity, and their ability to handle noisy data in difficult non-ideal contexts, even when the available knowledge is regarded as sufficient to employ a conventional modelling or statistical approach (Openshaw and Openshaw 1997).

1.3.3. Kohonen self-organizing map (SOM)

The Kohonen Feature Map or Self-Organizing Map was introduced in 1982 (Kohonen 1989) as an unsupervised neural network to simulate the learning process of the human brain. The basis of the SOM network is the *feature map*, a neuron layer where neurons organize themselves according to certain input values. The conceptual basis is the concept of *self-organization*, a neural process that describes the way the brain operates. The neural cells organize themselves into groups according to incoming information. This incoming information is not only received by a single neural cell, but also influences other cells in its neighbourhood. This organization results in some kind of map, where neural cells with similar functions are arranged close together. During the learning process, the nodes that are close in the array to a certain distance will activate each other to learn from the same input. This effect will create an ordering in the neighbourhood and a global ordering for a long learning time.

Several information visualization applications of the SOM have been proposed, mainly for text document visualization. The ET-MAP (Chen 1999), a hierarchical set of category maps that are essentially visual directories, was created with the SOM. The ET-MAP uses a land use metaphor (Dodge and Kitchin 2001) to display over 110,000 entertainment-related web pages listed by Yahoo. Cyberspace geography visualization developed by Girardin (1995) at the Graduate Institute of International Studies in Switzerland is another example of information visualization using SOM-based analysis for web content. WebSOM (Kohonen 1997) also uses SOM analysis but, rather than mapping the web, it maps thousands of articles posted on the Usenet newsgroup. WebSOM was developed by researchers at the Neural Network Research Centre at Helsinki University of Technology.

The SOM basically produces a similarity graph of input data, and converts the non-linear relationships between high-dimensional data into simple geometric relationships of their image points, usually on a 2D grid of nodes, by combining clustering and projection operations. One of its main applications has been the description of statistical data. The mapping produced by the SOM is expected to be explainable in terms of classical concepts of statistics. Other applications have been successful in optical character and script reading, speech recognition, image analysis, robot learning strategies, and biomedical applications (Behme et al. 1993; Alhoniemi et al. 1999; Kohonen 2001). In these applications, the SOM was found to be helpful in uncovering the relationships in the data, and finding patterns and trends based on its unsupervised learning method. It has been commonly argued that SOM networks are less sensitive to limitations (speed of convergence, local minima, etc.) known from classical neural networks such as multi-layer perceptrons and radial-basis function networks (Cottrell et al. 2001).

Application of the SOM for geospatial analysis has been considered mainly for classification problems (Weijan and Fraser 1996; Gahegan and Takatsuka 1999;

Gahegan 2000a; Luo and Tseng 2000; Tso and Mather 2001). More of the potential of this algorithm for geospatial pattern extraction and visualization remains to be explored.

Pattern extraction should also be combined with appropriate representation of information in order to allow users to understand the underlying relationships in the data and build knowledge about the geographic processes.

1.4. Alternative representation and visualization of large geospatial data

A new form of visual representation for geographic data is visualization, which attempts to give a response to the increasing needs of users. There is a continuous search for better visualization tools to allow users to benefit from geospatial analysis results, and to support the decision-making process. Visualization is the use of computer-supported interactive visual representations of data to amplify the cognition (Card et al. 1999), acquisition or use of knowledge. The cognitive support underlined in this definition is provided in six ways, by:

- increasing the memory and processing resources available to the users
- reducing the search for information
- using visual representations to enhance the detection of patterns
- enabling perceptual influence operations
- using perceptual attention mechanisms for monitoring
- encoding information in a manipulable medium.

Visualization can allow the data and information to be explored graphically, in order to gain the understanding, insight or knowledge from complex multidimensional datasets (McCormick et al. 1987) that is necessary for decision making, problem solving and explanation tasks. Visualization could also enhance the understanding of complexity, retaining user participation through computational steering to enhance the overall effectiveness of data analysis.

1.4.1. Geovisualization

Visualization in scientific computing emerged from the computer science field along with fields such as scientific visualization and information visualization. Information visualization is recognized as having the potential to enable better understanding of complex systems and the discovery of information that might otherwise remain unknown, and, by so doing, to facilitate better decisions (Card et al. 1999). Cartographic research efforts in visualization have been extended to meet other research activities in information science disciplines. This recognition was advanced by the creation in 1995 of a Commission on Visualization (later to

become the Commission on Visualization and Virtual Environments) within the International Cartography Association (ICA) (MacEachren and Kraak 2001). The research objective within this commission is to cope with the increasing volume of geospatial data by developing theory and practice that facilitate knowledge construction through the visual exploration and analysis of geospatial data. In addition, emphasis is laid on the visual tools necessary to support knowledge retrieval, synthesis and use (MacEachren and Kraak 2001).

Geovisualization (visualization applied to geospatial data) can be considered as the core discipline for understanding complex phenomena and processes, and structures and relationships in complex geospatial datasets. It integrates perspectives on the representation and analysis of geospatial data with recent developments in scientific and information visualization, exploratory data analysis (EDA), GIS, cartography and image analysis (Kraak 2000). It includes the use of a number of techniques for exploring data, answering questions, generating hypotheses, developing solutions and constructing knowledge. Such visual exploration of geospatial data (Kraak 1998) can be useful for displaying patterns with interaction and dynamics in order to achieve better decision making (see figure 1.1).

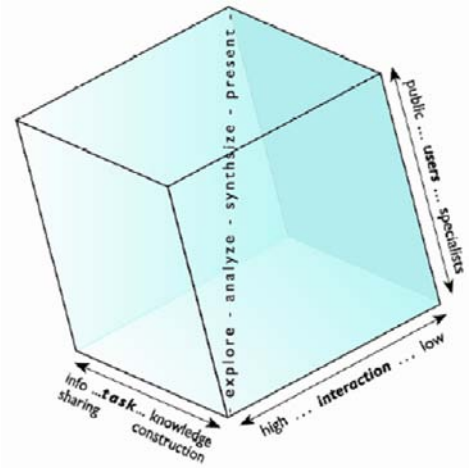


Figure 1.1. Geovisualization use space, with four dimensions: users, task, interaction and goal depicted along the central diagonal. (MacEachren 1994; MacEachren et al. 2004)

1.4.2. Information visualization and spatialization

Information visualization is a fast-growing research field with considerable potential and application in industry. Research in the field is concerned with graphically representing complex, abstract data domains in order to facilitate

knowledge extraction from very large non-spatial data. Representations used in information visualization often apply geographic metaphors to structure human-computer interaction, and are commonly referred to as *spatializations* or information spaces (Fabrikant 2001b). Most of these are generated outside the GIScience and cartography disciplines. Information visualization has two fundamentally related aspects: structural modelling and graphical representation (Chen 1999). The structural modelling intends to detect, extract and simplify underlying relationships, whereas the graphical representation is to transform an initial representation of a structure into a graphical one, so that the structure can be visually examined and interacted with.

While scientific visualization is applied to scientific data, information visualization is often applied to abstract data. Scientific data are often physically based (the human body, the earth, molecules, etc.), whereas abstract data are non-physical information and may not have obvious spatial mapping (financial data, business information, collections of documents, and abstract conceptions).

Information visualization can enable complex systems to be better understood, better decisions to be made, and information to be discovered that might otherwise remain unknown (Card et al. 1999). Users looking at large amounts of complex data can quickly find the information they need; navigate and interact with data more easily; recognize patterns and trends; discover errors in the data; easily identify minimum and maximum values, and clusters; and obtain a better understanding of the underlying structure and processes. This is possible, for example, by grouping or visually relating information in order to reduce the search for data, or by aggregating data in order that they may be revealed through clustering or common visual properties.

1.5. Problem and motivation

The growing volumes of geospatial data presents a difficult challenge in the exploration of patterns and relationships. With such large volumes of data, common geospatial analysis techniques are often limited in revealing patterns and relationships, a process necessary for understanding underlying structure and related real world processes. It seems that the geoscientist is dealing more and more with unknown data or data where he or she does not have enough knowledge of the underlying relationships (Openshaw and Openshaw 1997). In order to decide the class to which the pattern belongs, common classification or pattern recognition methods are used to compare the unknown pattern with all known reference patterns, on the basis of some criterion for the degree of similarity. These techniques are difficult to apply in the case of unknown data, as it is not obvious what mechanisms or rules are behind the data or classes of interest. New approaches in exploratory geospatial data analysis and visualization are needed to effectively extract patterns and relationships, and represent such data in a visual form that can better stimulate exploration, pattern recognition

and hypothesis generation, as well as allow better understanding of structures and processes and support knowledge construction.

1.6. Research goals

The objective of the research is to provide methods and techniques for integrating effective pattern extraction, based on computational analysis and graphical representations that can allow visual exploration of large geospatial data, and support knowledge construction. The specific objectives include providing appropriate tools that facilitate the development of problem solutions, and support hypothesis generation and testing, and the evaluation and interpretation of patterns. The ultimate goal is to support visual data mining and exploration, and gain insights into appropriate underlying distributions, patterns and trends, and therefore contribute to enhancing the understanding of geographic processes and knowledge construction.

The main research questions related are:

1. What computational and visual tools can be used to effectively extract patterns and represent information from very large geospatial data in order to allow understanding and knowledge acquisition?

2. What tools and methods can support visual data mining and knowledge discovery in large geospatial datasets?

3. How can visual and computational methods be integrated to support the design of exploratory geovisualization and knowledge discovery?

4. How can geographic and non-geographic information spaces be integrated to improve visual interaction and exploration of geospatial data?

5. How can the representation and exploration of spatio-temporal patterns in large geospatial data be supported?

6. Can a usability evaluation methodology based on an understanding of visual exploration and visualization tasks be developed to assess the exploratory geovisualization environments?

7. Can visual-computational approaches contribute to the exploratory analysis of geospatial data and knowledge construction?

1.7. Methodology

To address the research problems presented above, a methodology is proposed. It includes a theoretical basis and the subsequent experiments.

1.7.1. Theoretical basis

A theoretical basis for combining computational analysis and visual support is first proposed in a conceptual framework. This framework relates methods and techniques for spatial analysis, data mining, knowledge discovery and information visualization, and cartographic methods for guiding the design of exploratory geovisualization. This framework is intended to offer alternative and different views of the data, and as such stimulate the visual thinking process characteristic of visual exploration.

1.7.2. Experiments and validation of results

A number of experimental studies are conducted to investigate the potential of the proposed approach. These studies involve the exploration of large socio-demographic and health data in Chapter 3, and spatio-temporal data in Chapter 4, in order to provide some understanding of the complex relationships between socio-economic indicators. The experiments provide the opportunity, by way of a usability study, to assess the effectiveness of the proposed method in Chapter 6 and 7.

1.8. Structure of the thesis

To fulfil the research goals set out above, a research plan was implemented. It includes the activities presented in figure 1.2. These activities are reported in the different chapters of the thesis. The structure of the thesis is presented in figure 1.2.

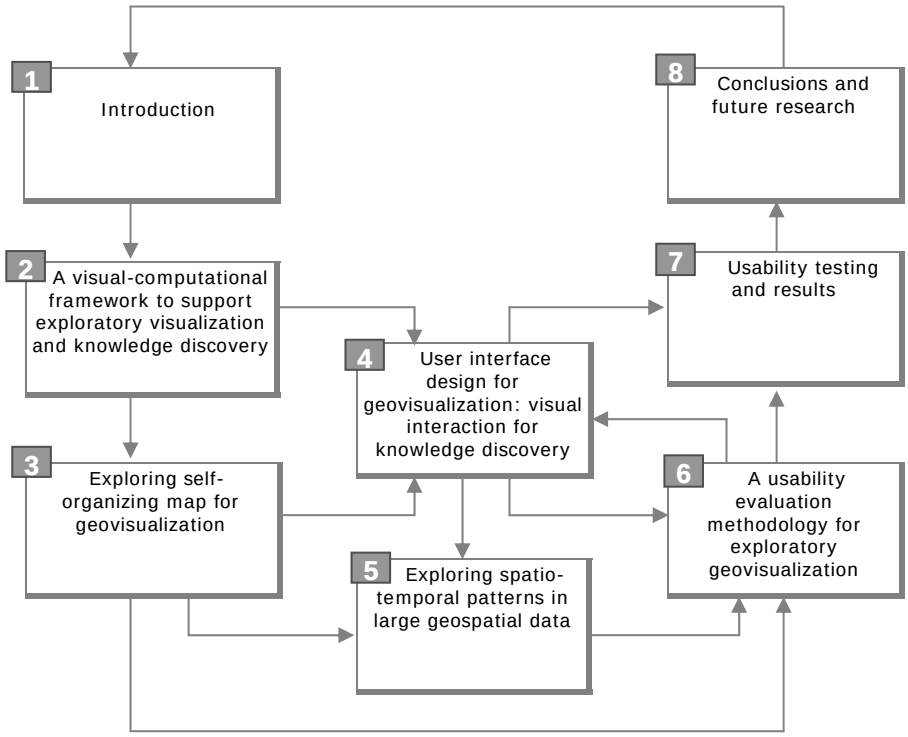


Figure 1.2. Structure of the thesis.

A brief description of the different chapters is provided below.

Chapter 1: Introduction
This chapter provides an introduction to the scope of the research, the research problem, the relevance of the topic, the context, the objectives, and the research questions.

Chapter 2: A visual-computational framework to support exploratory visualization and knowledge discovery
This chapter proposes a conceptual framework for the integration of computational and visual analysis methods for the exploratory visualization of large geospatial datasets.

Chapter 3: Exploring self-organizing map for geovisualization
This chapter explores the SOM potential for pattern extraction and investigates its application in visual exploration of large geospatial data. Two example cases are explored. At the end of this chapter, a preliminary usability feature inspection is conducted to gather users' views on the different representation forms offered.

Chapter 4: User interface design for geovisualization: visual interaction for knowledge discovery
This chapter describes the design approach to integrating visual and computational analysis for geospatial data exploration.

Chapter 5: Exploring spatio-temporal patterns in large geospatial data
This chapter presents an application of the proposed method for the representation and exploration of spatio-temporal patterns in a large dataset. Several representation techniques are proposed, based on the extraction capabilities offered by the computational analysis.

Chapter 6: A usability evaluation methodology for exploratory geovisualization
A usability evaluation methodology is proposed for assessing the usability and usefulness of the exploratory geovisualization environment. The method is based on a taxonomy of visualization tasks and operations.

Chapter 7: Usability testing and results
An empirical usability test is conducted and reported in this chapter. Some usability indicators (effectiveness, usefulness, and user reaction) are examined in relation to the use of the different graphical representations proposed, and are compared with maps and parallel coordinate plots.

Chapter 8: Findings and conclusions
This chapter presents the final findings of the research. Answers to the research questions posed in the introductory chapter are provided, as well as related future research issues.

Chapter 2

A visual-computational framework to support exploratory geovisualization and knowledge discovery

2.1. Introduction

A large proportion of geovisualization research (Dykes et al. 2004) is concentrated on the exploration of patterns and relationships in large and complex geospatial data, in order to provide alternatives representation and visualization techniques to improve geospatial data analysis. Information visualization and scientific visualization, particularly multidimensional visualization techniques (Nielson et al. 1997), are increasingly used in combination with exploratory data analysis techniques to explore the structure of large geospatial datasets. The integration of information visualization techniques with cartographic methods can benefit the geovisualization community. Examples of such integration are the dynamic and interactive maps designed in cartography (Kraak 2000). These interactive visual geospatial displays are used to explore data, generate hypotheses, develop problem solutions and construct knowledge (MacEachren 1994). This is by definition what geovisualization is all about: an active process that uses advanced user interfaces to allow users to highlight, filter and sort data as they search for patterns and relationships. To improve pattern extraction in large geospatial datasets, geographic data mining and knowledge discovery (Miller and Han 2001) have emerged from the application of data mining and knowledge discovery in databases (KDD) methods to the geographic domain. However, some difficulties remain in the application of data mining and KDD in geospatial data analysis. Besides the data volume, problems are often related to complexities caused by data gathering and the integration of different datasets and local relationships, and, more importantly, to the lack of appropriate methods and the difficulty in formulating the geographic domain (Gahegan and Brodaric 2002).

We explore a framework to integrate computational algorithms such as the self-organizing map (SOM) and visual analysis in a process that can support exploratory and knowledge discovery tasks. The SOM algorithm is used for data

---

This chapter is partly based on:

Koua E. L. and Kraak M.J. (2004). Self-organizing maps for exploratory visualization and knowledge discovery in large geopatial datasets. In: P. Agarwal and A. Skupin (Eds.) Self-organising maps: applications in geographic information sciences. New York: Wiley and Sons.

mining, knowledge discovery and spatial analysis, and for uncovering the structure, patterns, relationships and trends in the data. Some graphical representations are then used to portray extracted information in a visual form that can allow better understanding of the structures and the geographic processes. The design integrates non-geographic information spaces with maps and other graphics that allow patterns and attribute relationships to be explored, in order to facilitate knowledge construction. This allows the attribute space to be visualized while a link with the geographic space is maintained in multiple views. These graphical representations (information spaces) combine information visualization techniques and cartographic methods to improve the interaction and exploration of extracted patterns, and facilitate human perception and cognitive processes (MacEachren 1995; Card et al. 1999), by offering visualizations of the general structure of the dataset (clustering), as well as the exploration of relationships among attributes.

The chapter describes the framework in which pattern extraction with the SOM is combined with graphical representations in order to provide effective exploration of the data and support knowledge construction. The ultimate goal is to support visual data mining and exploration, and allow users to gain insights into appropriate underlying distributions, patterns and trends, and thus contribute to enhancing the understanding of geographic processes and support knowledge construction.

2.2. Geospatial data exploration and visualization

Continuous effort is put into adapting, improving and developing techniques for geospatial data exploration. We provide a description of some of the geospatial data exploration techniques in the next subsections, including the use of maps (traditional and new forms of map representations) and other visual exploration techniques.

To explore and compare the different visualization techniques, we use a dataset on geography and economic development (Gallup et al. 1999) to analyze the complex relationships between geography and macroeconomic growth. The dataset contains 48 variables on the economy, physical geography, population and health of 150 countries. This dataset will be further examined in Chapter 3.

2.2.1. Representation of geographic information: use of maps and beyond

The representation of geographic data has long been the primary role of cartography as a communication-oriented discipline with well-defined messages. A common way of representing geographic space is to use maps. Attention to maps

as spatial representation has expanded the field of cartography and makes links to a number of other disciplines, such as geographic information systems (GIS), remote sensing, cognitive science, sociology, cognitive and environmental psychology, and semiotics (MacEachren 1995). In the representation of geographic space, the map usually has a dual identity (Peuquet 1984): the map as a graphical image, which is the natural view we take of maps, and the geometric structure view, where the map is viewed as a composition of lines, points, polygons, curves, surfaces and volumes.

Although the map is a common way of representing geographic space, some geographic representation problems are more productively addressed by using other displays, as space to directly signify an attribute (MacEachren 1995). The cartogram (spatial transformations that depicts attributes of geographic objects as the object's area), a technique that can trade off shape and area adjustments, is often used to represent geographic information, as multidimensional scaling techniques are also a common alternative used in geographic problem solving. The SOM is getting a lot of attention as an alternative for representing and visualizing complex geographic data.

These new forms of representation of geospatial data utilize the geometric structure view of map representations, such as volumes, surfaces, points and lines, which encourages exploration and the subsequent discovery of novel insights into geographic databases (Peuquet and Kraak 2002). The geometry allows the exploration of relationships between items in the representation. Distance (similarity between data items), regions (aggregation of similar data items), and scale (level of detail in a database) are examples of spatial metaphors used in new representation spaces (Fabrikant 2001b). Coordinate systems allow distance and direction to be determined, from which other spatial relationships (size, shape, density, arrangement) may be derived. The scale allows exploration of the information space at multiple levels of detail, and provides the potential for the hierarchical grouping of items, and for revealing categories or classifications.

In geovisualization, all mappings of geospatial information in perceptible forms (visual, haptic or audible) are being used. The integration of senses such as hearing and touch, immersive display forms (MacEachren et al. 2003), interactions and dynamism into representations is raising new cognitive issues, important for navigating and exploring geoinformation spaces. The representations need to be appropriate to the task, and an examination of the semantics within the data is needed to maintain a connection between the real world and its iconic representation in map or more schematic forms. For example, creating a geographic analogy can help generate an information landscape based on experiential properties of the real world (Fabricant and Buttenfield 2001).

2.2.2. Multidimensional multivariate visualization and exploratory data analysis techniques

Multidimensional multivariate visualization (mdmv) has been an active research field for more than three decades, from which scientific visualization is a sub-field dealing with the analysis of data with multiple parameters and factors (Nielson et al. 1997). These techniques aim at extending the possibilities of multivariate correlation in an effort to provide correlation information among many variates simultaneously in large datasets, as a solution to limitations of statistical visualization techniques. A comprehensive classification of multivariate visualization techniques was provided by Nielson et al. (1997) and Keim et al. (2001).

Exploratory data analysis (EDA) is a well-established tradition in statistics (Tukey 1977) applied to a distinctive approach to the exploration of data, including pattern recognition and uncovering data structure (Sibley 1988). EDA methods are used to explore the data and reveal patterns or structures in the data in order to understand the underlying processes, as opposed to confirmatory methods, which are conventional statistical techniques based on classical theory and designed to test hypotheses. Projections and clustering techniques are used to emphasize the entire knowledge discovery process, and the discovery of novel patterns. Each variable can be shown either as an independent variable or in relation to other variables. However, for a large number of variables, it remains difficult to interpret and understand the underlying structure of a dataset, as the display of all data components becomes difficult and incomprehensible for large multidimensional datasets. Among the many exploratory data analaysis techniques applied in geospatial data analysis, parallel coordinates (Inselberg 1985) are particularly common. While a parallel coordinate is a useful interactive exploration technique (interactive brushing), it falls short of providing a useful overview of the full dataset (Keim et al. 2001). For example, a parallel coordinate plot of the dataset studied (150 samples and 48 variables) would be a rather noisy picture despite brushing features. The method appears satisfactory only when the number of observations is small. Slocum (1999) suggested that the maximum number of variables for interpretable display using a parallel coordinate plot should be between 10 and 20, which corresponds to a rather small dataset. Other common EDA techniques include scatter plots, scatterplot matrixes, multidimensional graphs, linear projection methods such as principal component analysis (PCA), and non-linear methods such as multidimensional scaling (MDS). Projection methods are meant to reduce the dimensionality of the dataset and represent the input data space on a lower-dimensional space, so that certain properties of the dataset structure, such as the distances between data items (variations presented in the data), are preserved as much as possible. Several non-linear projection techniques have been introduced to deal with highly asymmetric distributions, where linear projections may not be effective in visualizing the structures of the distributions or other structures. The MDS technique (Torgerson 1952) is a widely used non-linear projection method. It is a

method that represents measurements of similarity or dissimilarity among pairs of objects as distances between points of a low-dimensional space. In addition to detecting underlying structure and reducing data, MDS provides a spatial representation of data that can facilitate interpretation and reveal relationships.

Sammon's mapping (Sammon 1969) is another non-linear projection technique that is widely used and very similar to MDS methods. It tries to match the distance among pairs of data items of the low-dimensional space with their original distances. Figure 2.1 provides a representation of the dataset, using MDS (a), PCA (b), K-means clustering (c) and the SOM grid (d).

The major drawback, however, is that Sammon's mapping algorithm, like MDS methods, is a point-to-point mapping, which does not provide an explicit mapping function and cannot accommodate new data points. For any additional input data, the projection will be re-calculated completely; this makes these techniques computationally very intensive.

A number of authors have proposed using artificial neural networks as part of a strategy to improve geospatial analysis of large, complex datasets (Schaale and Furrer 1995; Openshaw and Turton 1996; Skidmore et al. 1997; Gahegan and Takatsuka 1999; Gahegan 2000a). Artificial neural networks have the ability to perform pattern recognition and classification. They are especially useful in situations where the data volumes are large and the relationships are unclear or even hidden; this is because of their ability to handle noisy data in difficult non-ideal contexts (Openshaw and Openshaw 1997). Particular attention has been directed to using the self-organizing map (SOM) neural network model as a means of organizing complex information spaces (Girardin 1995; Chen 1999; Fabricant and Buttenfield 2001). The SOM is also generally acknowledged to be a useful tool for the extraction of patterns and the creation of abstractions where conventional methods may be limited because underlying relationships are not clear or classes of interest are not obvious.

A wide range of SOM applications in geospatial analysis have been explored, including geospatial data mining and knowledge discovery (Gahegan and Brodaric 2002), map projection (Skupin 2003), and classification (Gahegan and Takatsuka 1999; Gahegan 2000a). This interest in the SOM in geographic data analysis is due partly to its multidimensional data reduction and topological mapping capabilities.

A more detailed description of the SOM is provided in the next section.

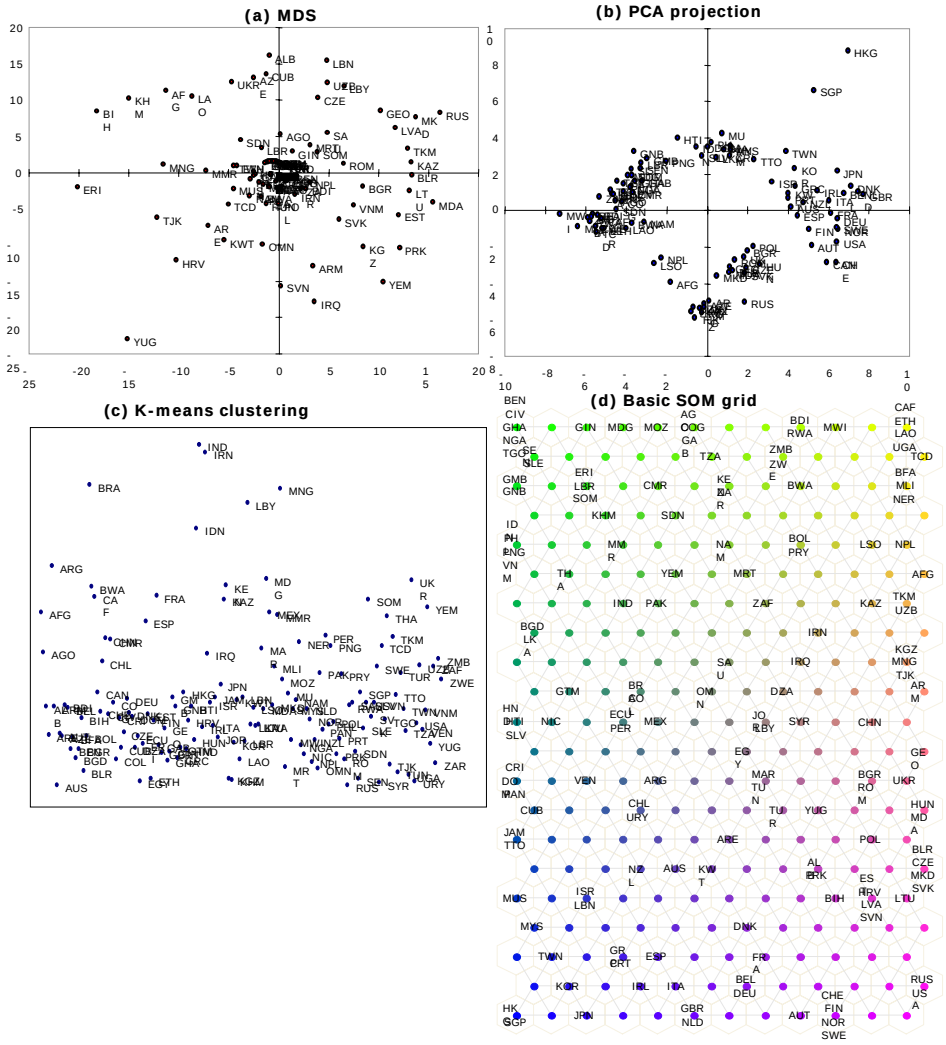


Figure 2.1. A representation of the structure of the dataset (geography and economic development) using MDS (a), PCA projection (b), and K-means clustering (c), and a basic SOM grid representation showing neuron positions and countries to which they were adapted. Nearby locations on the SOM grid represent countries with similar characteristics according to the multivariate attributes.

One fundamental difference between the SOM and other projection methods such as MDS and Sammon's mapping is that the SOM tries to preserve not the distances between original data items, but the neighbourhood relations on the map lattice (see figure 2.1d). Nearby locations on the SOM lattice represent similar data items. The second major difference is that the SOM estimates an explicit mapping function based on a dataset, which can be used for new

observations. Finally, the SOM reduces the input data to a small number of vectors, so the burden of computation is reduced.

2.3. SOM and geospatial data exploration: a framework to support exploratory visualization and knowledge discovery

Like other neural networks in general, the SOM has the ability to perform pattern recognition and classification, and to handle noisy data. This SOM property has offered an alternative to non-linear projection and multidimensional data visualization (Yin 2001) as one of the attempts to improve data analysis in general. Based on adaptive mapping methods, this neural network can learn complex non-linear relationships, often unclear or hidden, from vast numbers of variables in data.

The proposed framework explores ways to effectively extract patterns in the data by using data mining techniques, and represents the results by using graphical representations for visual exploration. This framework is based on current understanding of the effective application of visual variables (MacEachren 1995; Wilkinson 1999) for cartographic and information design, on developing theories on interface metaphors for geospatial information displays, and on previous empirical studies of map and information visualization effectiveness. In the next subsection, we outline the main components of the approach, including a detailed description of the algorithm; computational analysis steps, including data mining and knowledge discovery issues; and the representational and exploratory visualization framework.

2.3.1. The self-organizing map (SOM) algorithm

The SOM (Kohonen 1989) is an artificial neural network used to map multidimensional data onto a low-dimensional space, usually a 2D representation space (see figure 2.2). The network consists of a number of neural processing elements (units or neurons) usually arranged on a rectangular or hexagonal grid, where each neuron is connected to the input. Each of the units i is assigned an n-dimensional weight vector m_i that has the same dimensionality as the input patterns.

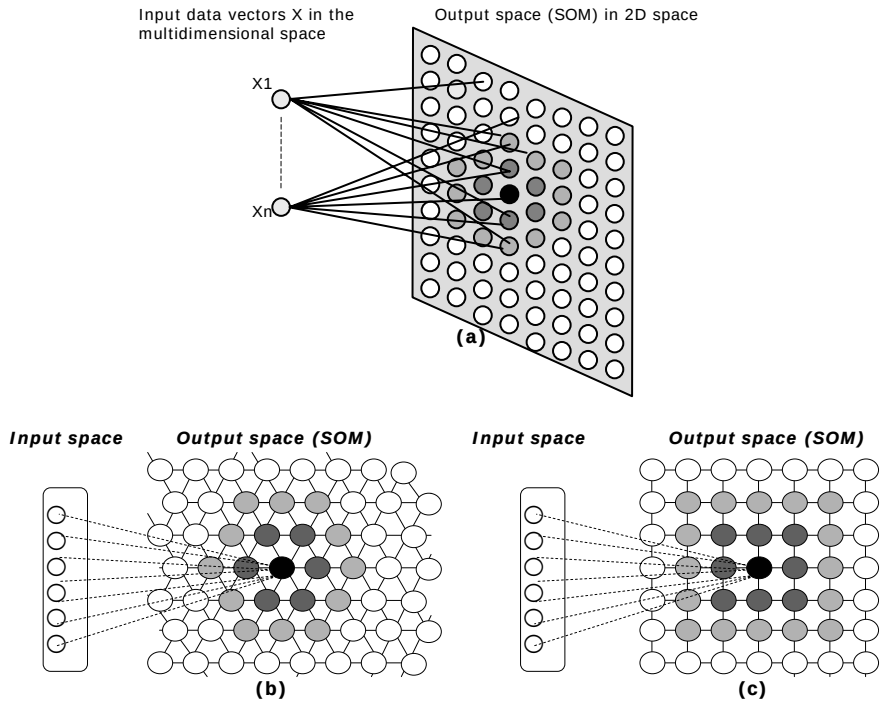


Figure 2.2. Structure of a SOM network (a) with the selection of a node and adaptation of neighbouring nodes to the input data. The SOM grid can be hexagonal (b) or rectangular (c). The black object indicates the node that was selected as the best match for the input pattern. Neighbouring nodes adapt themselves according to the similarity with input pattern. The degree of shading of neighbouring nodes corresponds to the strength of the adaptation.

What changes during the network training process are the values of these weights. Each training iteration t starts with the random selection of one input pattern $x(t)$. Using Euclidean distance between weight vector and input pattern, the activation of the units is calculated. The unit with the lowest activation is referred to as the winner, c, of the training iteration:

$$m_c(t) = \min_i \{\|x(t) - m_i(t)\|\} \qquad (1)$$

Finally the weight vector of the winner as well as the weight vectors of selected units in the neighbourhood of the winner are adapted to represent the input pattern. At each step t of the random sequence of the given $x(t)$ values, the values of m_i are gradually and adaptively changed in the following adaptation process:

$$m_i(t+1) = m_i(t) + h_{ci}(t)[x(t) - m_i(t)] \qquad (2)$$

The degree of adaptation in the neighbourhood is characterized by a neighbourhood function h, which is a decreasing function of the units from the winning unit on the map lattice until no noticeable changes are observed. As a result of a general adaptation process, a number of units in the neighbourhood of the winner lead to a spatial clustering of similar input patterns in neighbouring parts of the SOM.

The resultant maps (SOMs) are organized in such a way that similar data are mapped onto the same node or onto neighbouring nodes in the map. This arrangement of the clusters in the map reflects the attribute relationships of the clusters in the input space. For example, the size of the clusters (the number of nodes allotted to each cluster) is reflective of the frequency distribution of the patterns in the input set. Actually, the SOM uses a distribution-preserving property, which has the ability to allocate more nodes to input patterns that appear more frequently during the training phase of the network configuration. In other words, the topology of the dataset in its n-dimensional space is captured by the SOM and reflected in the ordering of its nodes. This is an important feature of the SOM and allows the data to be projected onto the lower-dimension space while roughly preserving the order of the data in its original space. Another important feature of the SOM for knowledge discovery in complex datasets is the fact that it is an unsupervised learning network, meaning that the training patterns have no category information that accompanies them. Unlike supervised methods that learn to associate a set of inputs with a set of outputs by using a training dataset for which both input and output are known, the SOM adopts a learning strategy where the similarity relationships between the data and the clusters are used to classify and categorize the data.

The SOM can be useful for knowledge discovery in database methodology as it follows the probability density function of underlying data. It also offers visual representations that enable easy data exploration.

We have integrated the SOM in a knowledge discovery framework (figure 2.5) for the exploration of complex geospatial data. The use of the SOM is intended to provide additional exploratory data analysis techniques for complex geospatial data. The strategy is to integrate the computational analysis (extraction of patterns and relationships) using the SOM algorithm with graphical representations that can stimulate pattern recognition and hypothesis generation. For the user, the main goal is the acquisition of knowledge through exploration and discovery for decision making, problem solving and explanation. These goals are targeted in the framework, using interaction and exploratory tasks for understanding the structures and processes and the knowledge construction process.

Self-organizing map quality

After the SOM has been trained, it is important to know whether it has properly adapted itself to the training data. Because it is obvious that one optimal map for

the given input data must exist, several map quality measures have been proposed (Kohonen 1995; Kohenen 2001). The maps have two primary quality properties: data representation accuracy (mapping precision or resolution) and dataset topology (attributes relationships) representation accuracy (topology preservation). A common measure that calculates the precision of the mapping is the average quantization error over the entire dataset. This mapping precision measure describes how accurately the neurons respond to the given dataset. Since the responses of the network neurons to the data samples are based on Euclidean distance, the nearest vector is the best match unit for that sample. The average quantization error measures the precision of the mapping, using the average of the Euclidean distances of each input vector x_i and its best matching reference vector m_c in the SOM, using:

$$E_q = \frac{1}{N}\sum_{i=1}^{N}\left\|x_i - m_c\right\| \qquad (3)$$

where N is the number of data vectors in the input data space. For example, if the reference vector of the best matching unit calculated for a given testing vector x is exactly the same x, the error in precision is then 0. Normally, the number of data vectors exceeds the number of neurons and the precision error is thus always different from 0.

For the topology representation accuracy (topology preservation), an error measure (percentage of data vectors for which the first and second best matching units are not adjacent units) is used. The topology preserving property comes from the fact that similar data are mapped onto the same node, or to neighbouring nodes in the map. This is described as:

$$E_t = \frac{1}{N}\sum_{k=1}^{N}u(x_k) \qquad (4)$$

where $u(x_k)$ is 1 if the first and second best matching unit of x_k are not next to each other (not adjacent units), otherwise $u(x_k)$ is 0.
An illustration of the training process is presented in figure 2.3.

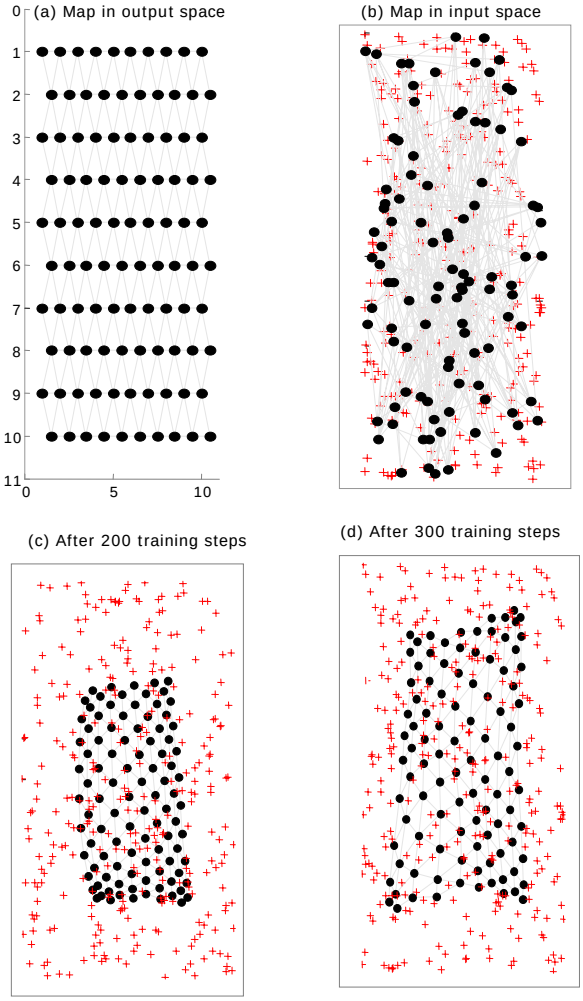


Figure 2.3. Illustration of the SOM training process with a 10 x 10 network of neurons. The black dots show the positions of the map units or neurons, and the red crosses represent 300 randomly sampled data points. The SOM grid is a 2-dimensional grid of 10 x 10 neurons (a) that show the connections between neighbouring map units in the output SOM space. The positions of the map units are disorganized together with the training data in the input space after a random initialization of the network (b). During training, the map self-organizes and folds to the training data at each learning step (c),and (d), so that the map units represent the data vectors as similar as possible.

After training, the distances between neighbouring map units can be represented in a distance map, to show the overall clustering of the data. For a given map of $[n, p]$size, the distance map is a $[k,l]$ vector $U_{ij} = [u_1, u_{12}, \ldots u_p]$.

For example, in the case of a $[1x5]$ size map, $[m_1, m_2, m_3, m_4, m_5]$ where m_i denotes one map unit. The distance map is a $[1x9]$ vector $[u_1, u_{12}, u_2, u_{23}, u_3, u_{34}, u_4, u_{45}, u_5]$ where $u_{ij} = \|m_i - m_j\|$ is the distance between map units m_i and m_j, and u_b is the mean (or minimum, maximum or median) of the surrounding values, for example $u_3 = (u_{23} + u_{34})/2$.

2.3.2. Visual data mining and knowledge discovery for understanding geographic processes

One approach to analyzing large amounts of data is to use data mining and knowledge discovery methods. In geospatial analysis, data mining tools are applied to extract patterns from large datasets and help uncover structures in complex data (Openshaw et al. 1990). The main goal of data mining is to identify valid, novel, potentially useful patterns in data, and ultimately to understand them (Fayyad et al. 1996). Generally, three general categories of data mining goals can be identified (Weldon 1996): explanatory (to explain some observed events), confirmatory (to confirm a hypothesis), and exploratory (to analyze data for new or unexpected relationships). Typical tasks for which data mining techniques are often used include clustering, classification, generalization and prediction. These techniques vary from traditional statistics to artificial intelligence and machine learning. The most popular methods include decision trees (tree induction), value prediction, and association rules often used for classification (Miller and Han 2001). Artificial neural networks are used particularly for exploratory analysis as non-linear clustering and classification techniques. For example, unsupervised neural networks such as the SOM are a type of neural clustering technique, and neural architectures using backpropagation and feedforward are neural induction methods used for classification (supervised learning). The algorithms used in data mining are often integrated into KDD, a larger framework that aims at finding new knowledge from large databases. While data mining deals with transforming data into information or facts, KDD is a higher-level process using information derived from the data mining process to turn it into knowledge or integrate it into prior knowledge (see figure 2.4).

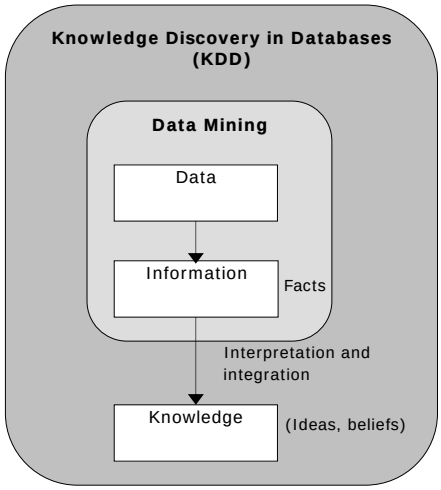


Figure 2.4. Data mining and knowledge discovery frameworks. Data mining is part of the knowledge discovery process. Data is first processed and transformed into information at the data mining stage. Information derived from the data mining process is turned into knowledge or integrated into prior knowledge through interpretation and evaluation. Visualization can support any stage in this process to enhance exploration (visual data mining).

In general, KDD stands for discovering and visualizing the regularities, structures and rules from data (Miller and Han 2001), discovering useful knowledge from data (Fayyad et al. 1996), and finding new knowledge. It consists of several generic steps, namely data pre-processing, transformation (dimension reduction, projection), data mining (structure mining) and interpretation/evaluation.

Applications of data mining and KDD methods have advanced geographic data mining and knowledge discovery, which has become an established field in geographic visualization (Sibley 1988; Weijan and Fraser 1996; MacEachren et al. 1999; Gahegan et al. 2001; Liu et al. 2001; Miller and Han 2001; Roddick and Lees 2001). This framework has been used in geospatial data exploration (Openshaw et al. 1990; MacEachren et al. 1999; Wachowicz 2000; Gahegan et al. 2001; Miller and Han 2001) to discover unexpected correlation and causal relationships, and understand structures and patterns in complex geographic data. The promises inherent in the development of data mining and knowledge discovery processes for geospatial analysis include the ability to yield unexpected correlation and causal relationships. A large proportion of these applications are directed towards spatio-temporal data mining (Roddick and Lees 2001).

The dimensionality of the dataset is very high, and searching for patterns in such high-dimensional space is often ineffective. We use the SOM algorithm as a data mining tool to project input data into an alternative measurement space, based on similarities and relationships in the input data, which can aid the search for

patterns. It becomes possible to achieve better results in this similarity space than in the original attribute space (Strehl and Ghosh 2002). As described in the previous section, the SOM adapts its internal structures to the structural properties of the multidimensional input, such as regularities, similarities and frequencies. These SOM properties can be used to search for structures in the multidimensional input. Graphical representations are then used to enable visual data exploration, allowing the user to gain insight into the data, evaluate, filter, and map outputs. This is intended to support visual data mining (Keim 2002) by allowing several variables and their interactions to be inspected simultaneously, and by receiving feedback from the knowledge discovery process by means of interaction techniques that support the process (Cabena et al. 1998).

2.3.3. Computational analysis and visualization framework

One of the advantages of the SOM is that the outcome of the computational process can easily be portrayed through visual representation. The first level of the computation provides a mechanism for extracting patterns from the data. The output of this computational process is depicted using graphical representations (information spaces) to facilitate human perception and cognitive processes (MacEachren 1995; Card et al. 1999), by offering visualizations of the general structure of the dataset (clustering), as well as the exploration of relationships among attributes. Users can perform a number of exploratory tasks not only to understand the structure of the dataset as a whole, and also to explore detailed information on individual or selected attributes of the dataset.

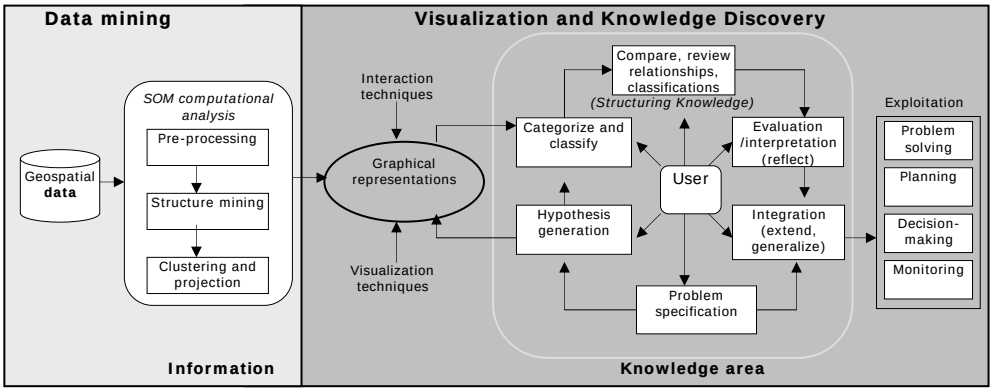


Figure 2.5. Computational analysis and exploratory visualization framework.

Like other artificial neural networks, the SOM is used as a non-linear clustering and classification technique that provides ground for extracting patterns from the data at the data mining level in the exploratory framework for visualization and knowledge discovery described below (figure 2.5).

To enhance the exploration of the graphical representations, visualization and interaction techniques such as brushing, focusing, filtering, browsing, querying, selecting and linking are used. Projection techniques such as Sammon's mapping and PCA are also used to support the representations of SOM results. As with maps, these representations use visual variables in addition to the position property of the map elements. Multiple views are used to offer alternative and different views of the data in order to stimulate the visual thinking process that is characteristic of visual exploration.

The extraction results in maps (SOMs) can then be visualized using graphical representations. We propose two levels of visualization and knowledge discovery processes closely related to the concept of abduction (Gahegan and Brodaric 2002). These are supported by a number of activities, including selection, analysis, comparison, and the relation of spatial locations or attributes, starting from the general patterns extracted and moving on to more user selection and refinement, which allow the exploration of relationships and the structure of a particular area of interest.

The first level of this framework consists of the visualization of the general structure (Shneiderman 1997) of the dataset (clustering); the second level focuses on the exploration for knowledge discovery and hypothesis generation. These two levels of the visualization process are provided with different SOM representations that can be combined with other visualization techniques. The fundamental idea is centred around four basic visualization goals, the basis for the exploratory visualization and knowledge discovery process (Weldon 1996):

- discovering patterns (through similarity representations)
- exploring correlations and relationships for hypothesis generation
- exploring the distribution of the dataset on the map
- detecting irregularities in the data.

2.4. Conclusion

The framework presented in this chapter was based on an approach to combine visual and computational analysis for the development of a visualization environment intended to contribute to the analysis of large volumes of geospatial data. This approach focuses on the effective application of computational algorithms to extract patterns and relationships in geospatial data, and the visual representation of derived information, which involves the effective use of visual variables used in such complex information spaces to facilitate knowledge construction. The main components of the approach were outlined, including a detailed description of the algorithm used; the computational analysis steps, including data mining and knowledge discovery issues; and the representational and exploratory visualization framework. The first level of the framework consists

of the visualization of the general structure of the dataset (clustering); the second level focuses on the exploration for knowledge discovery and hypothesis generation. A number of activities are supported in this framework, including selection, analysis, comparison, and the relation of spatial locations or attributes, starting from the general patterns extracted and moving on to more user selection and refinement, which allow the exploration of relationships and the structure of a particular area of interest.

Several multidimensional multivariate visualization and exploratory data analysis techniques were explored. The spatial representation of the SOM (grid) provides opportunities for exploring the attribute space in relation to the spatial locations. It can be used to search for structures in the multidimensional input as a data mining tool based on similarities and relationships in the input data. One of the advantages of the SOM is that the outcome of the computational process can easily be portrayed through visual representation. The first level of the computation provides a mechanism for extracting patterns from the data at the data mining stage in the framework. The output of this computational process is depicted using graphical representations that offer visualizations of the general structure of the dataset (clustering), as well as the exploration of relationships among attributes. These graphical representations are used to enable visual data exploration, allowing the user to gain insight into the data, evaluate, filter, map outputs, in order to understand structures and patterns in data.

The overall objective of the proposed framework is to explore ways of supporting visual exploration and knowledge construction in large geospatial data. In this respect, the SOM computational analysis can support exploratory visualization and the knowledge discovery process when integrated with appropriate visual exploration tools. The goal is to integrate the computational process and the graphical representations so that users can perform a number of exploratory tasks not only to understand the structure of the dataset as a whole, and also to explore detailed information on individual or selected attributes of the dataset. Multiple views can be used to offer alternative and different views of the data in order to stimulate the visual thinking process that is characteristic of visual exploration. Interactive manipulation (zooming, rotation, panning, filtering and brushing) of the graphical representations can enhance both user goal-specific querying and selection from the general patterns extracted, and more specific user querying and selection of attributes and spatial locations for exploration, hypothesis generation, explanation and knowledge construction. The link between the attribute space visualization based on the SOM, the geographic space with maps representing the SOM results, and other graphics such as parallel coordinate plots in multiple views can provide alternative perspectives for the better exploration, evaluation and interpretation of patterns, and ultimately supports knowledge construction. These aspects will be the focus of a subsequent design in Chapter 4 and usability test in Chapter 6 and 7. One goal will be to characterize the overall effectiveness of the representations used and how they can support exploratory geovisualization.

Chapter 3

Exploring self-organizing map for geovisualization

3.1. Introduction

With volumes of data becoming larger and data structures more complex, designing an effective visualization environment for analyzing large geospatial datasets has become one of the major concerns in the geovisualization community. In these large and rich databases, uncovering and understanding patterns or processes presents a difficult challenge as they easily overwhelm mainstream geospatial analysis techniques oriented towards the extraction of information from small and homogeneous datasets (Gahegan et al. 2001; Miller and Han 2001).

As described in the previous chapter, SOMs (and other artificial neural network methods) can be used to extract features in complex data. To interpret these (often abstract) features, appropriate visualization techniques are needed to represent extracted information in a way that allows better understanding of underlying structures and processes. The goal is to represent the data in a visual form in order to stimulate pattern recognition and hypothesis generation. The use of information spaces can play a role by offering visual representations of data that bring the properties of human perception to bear (Card et al. 1999). Spatial metaphors such as distances, regions and scale are used to facilitate the representation and understanding of information in such spaces (Fabrikant et al. 2002). An important step in the design of effective visualization tools will rely on understanding the way users interpret and build a mental model of these information spaces.

The relative effectiveness of integrating the SOM with visualization methods for exploration and knowledge discovery in complex geospatial datasets remains to be explored. In particular, we believe that the visual design of SOM graphical representations will significantly affect how successful they are for exploratory analysis purposes. This chapter discusses the potential of the SOM in an integrated visual-computational environment, presents four alternative visual renderings of the SOM that can be used to highlight different characteristics of the computational solution it produces, and proposes an evaluation strategy for

---

This chapter is partly based on:

Koua, E. L. and Kraak, M. J. (2004). Evaluating Self-organizing Maps for Geovisualization. In J. Dykes, A. M. MacEachren and M. J. Kraak (Eds.). Exploring Geovisualization. Amsterdam: Elsevier.

assessing their relative effectiveness in terms of three common visualization tasks: (1) identifying clusters, (2) relating distances (similarity), (3) relating values. Two datasets are explored. The first, a simple case on socio-economic data in a region of the Netherlands, is a known dataset used to explain the different graphical representations. The second, is a case related to geography and economy development, in which complex attributes relationships and hypotheses can be explored. The chapter concludes by discussing some next steps.

3.2. Usability framework for the design of the visualization environment

The design of the visualization environment is based on a usability framework structured to develop a tool that is useful and appropriate for the user needs and tasks. This framework not only includes the techniques, processes, methods and procedures for designing usable products and systems, it also focuses on the user's goals, needs and tasks in the design process (Rubin 1994). User characteristics, visualization tasks and operations are examined to improve user interaction and to support activities involved in the use of the visualization environment, and in related information spaces. Figure 3.1 shows the underlying design concept and usability framework. This framework is informed by current understanding of effective application of visual variables for cartographic and information design, developing theories of interface metaphors for geospatial information displays, and previous empirical studies of map and information visualization effectiveness. The framework guided the initial design decisions presented here and will be used to structure subsequent user studies (the strategy for which is introduced in section 3.5).

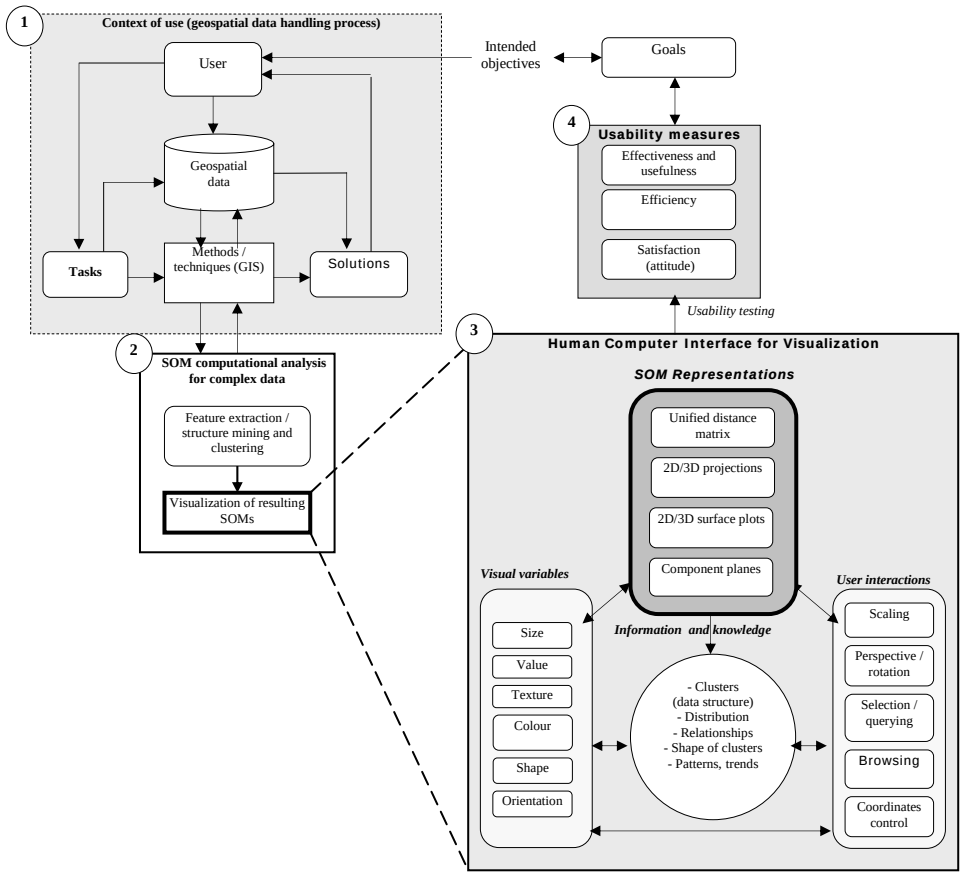


Figure 3.1. Usability framework for the design of the SOM-based visualization environment. Stage (1) describes the general geospatial data handling process; (2) represents the proposed computational analysis and visualization method based on the SOM algorithm for complex geospatial data; (3) is the design framework for the human computer interface for the visualization of the SOMs and includes the representations to evaluate; (4) shows the usability measures used to test the outcome of interaction and use of the visualization tool.

The objective of developing a SOM-based visualization environment is to contribute to the analysis and visualization of large amounts of data, as an extension of the many geospatial analysis functions available in most GIS software. The design of the tool is intended to help uncover structure and patterns that may be hidden in complex geospatial datasets, and to provide graphical representations that can support understanding and knowledge construction. The framework includes spatial analysis, data mining and knowledge discovery methods integrated into an interactive visualization system. Users can perform a number of exploratory tasks, not only to understand the structure of the dataset as a whole but also to explore detailed information on individual or

selected attributes of the dataset. In order to examine how users understand these representations, and to improve the overall effectiveness of the design, a usability assessment plan is proposed at the end of this chapter for evaluating the graphical representation forms accessible through the tool, as well as the visual variables used to depict data within each form of representation.

For the user, the main goal is the acquisition of knowledge through discovery for purposes of decision making, problem solving and explanation. The first level of the computation provides a mechanism for extracting patterns from the data. Resultant maps (SOMs) are then visualized using graphical representations. We use different visualization techniques to enhance data exploration, including brushing, multiple views and 3D views. Projection methods such as Sammon's mapping (Sammon 1969) and principal component analysis (PCA) are also used to depict the output from the SOM. Spatial metaphors are used to guide user exploration and interpretation of the resulting non-geographic representation; this is an example of spatialization, an approach discussed more generally by Fabrikant and Buttenfield (2001) and by Fabrikant and Skupin (2004). These metaphors are combined with alternative 2D and 3D forms of representation and user interaction in the information spaces.

The resulting information spaces suggest and take advantage of natural environment metaphor characteristics such as '*near=similar, far=different*' (MacEachren et al. 1999), which is epitomized by Tobler's first law of geography (Tobler 1970). Various types of map representations are used, including volumes, surfaces, points and lines. This allows exploration of multiple kinds of relationships between items. A coordinate system allows the user to determine distance and direction, from which other spatial relationships (size, shape, density, arrangement, etc.) may be derived. Multiple levels of detail allow exploration at various scales, creating the potential for hierarchical grouping of items, regionalization and other types of generalization.

3.3. The SOM graphical representations

The design of the visualization environment incorporates several graphical representations of SOM output. These include a distance matrix representation, 2D and 3D projections, 2D and 3D surfaces, and component plane visualization (in a multiple view). These representation forms are introduced briefly using the dataset described below.

3.3.1. Map

The first dataset explored to illustrate the SOM-based representations is a collection of socio-economic indicators related to municipalities in a region in the

Netherlands. It consists of 29 variables, including population and habitat distributions, urbanization indicators, income of inhabitants, family and land data, as well as industrial, commercial and non-commercial services data. The idea is to find multivariate patterns and relationships among the municipalities. This dataset was selected for the study because it is a known dataset in which we can test different hypotheses about both the geographic patterns and the representation/analysis methods investigated. The maps assist in understanding SOM representations. Unusual SOM patterns can be verified with reality. At the end, the use of SOM is applied to far larger datasets than the one used in this experiment.

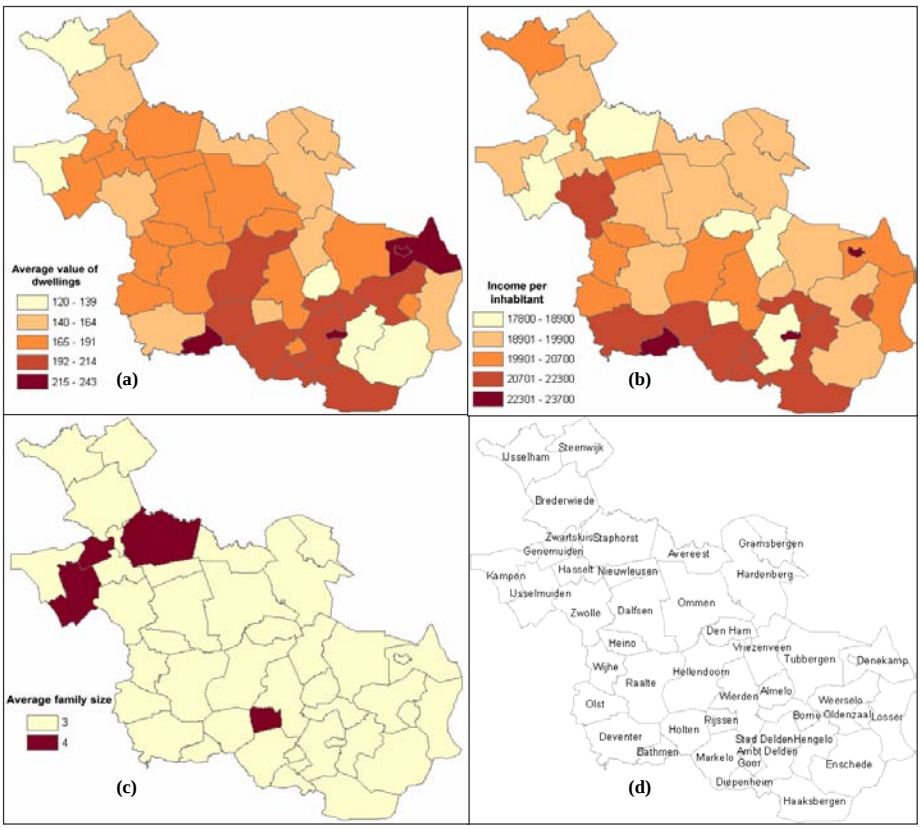


Figure 3.2. Examples of attributes of the test dataset: average family size (a), average income per inhabitant in the municipalities (b), average value of dwellings (c). (d) shows the names of municipalities.

Three attributes of the dataset (family size, income per inhabitant, and average value of dwellings), as well as a reference map with the names of the municipalities, are presented in figure 3.2. The maps show, for example, that

there are four municipalities where the average family size is four, compared with the rest of the region where the average family size is three.

With the SOM, such relationships can be easily examined in a single visual representation using the component planes. Component planes show the values of the map elements for different attributes. They show how each input vector varies over the space of the SOM units. Unlike standard choropleth maps, the position of the map units (which is the same for all displays) is determined during the training of the network, according to the characteristics of the data samples. A cell or hexagon here can represent one or several political units (municipalities), according to the similarity in the data. Two variables that are correlated will be represented by similar displays.

In the example described above, the SOM shows that there is a cluster of municipalities that have a family size of more than three (see figure 3.3a). It also shows the relationships between the municipalities for the different attributes. The two other attributes (income per inhabitant and average value of dwellings) are presented as component planes extracted from the SOM (figure 3.3b and 3c) for exploratory analysis purposes. By relating component displays we can explore the dataset, interpret patterns as indications of structure, and examine relationships that exist. For example, figures 3b and 3c indicate that the highest dwelling values correspond to municipalities (StadDelden, Bathmen, Diepenheim) where the average income per inhabitant is highest. The representations of the SOM make it possible to easily find correlations in a large volume of multivariate data. New knowledge can be unearthed through this process of exploration; this is followed by the identification of associations between attributes through using the various representations, and finally by the formulation and ultimate testing of hypotheses.

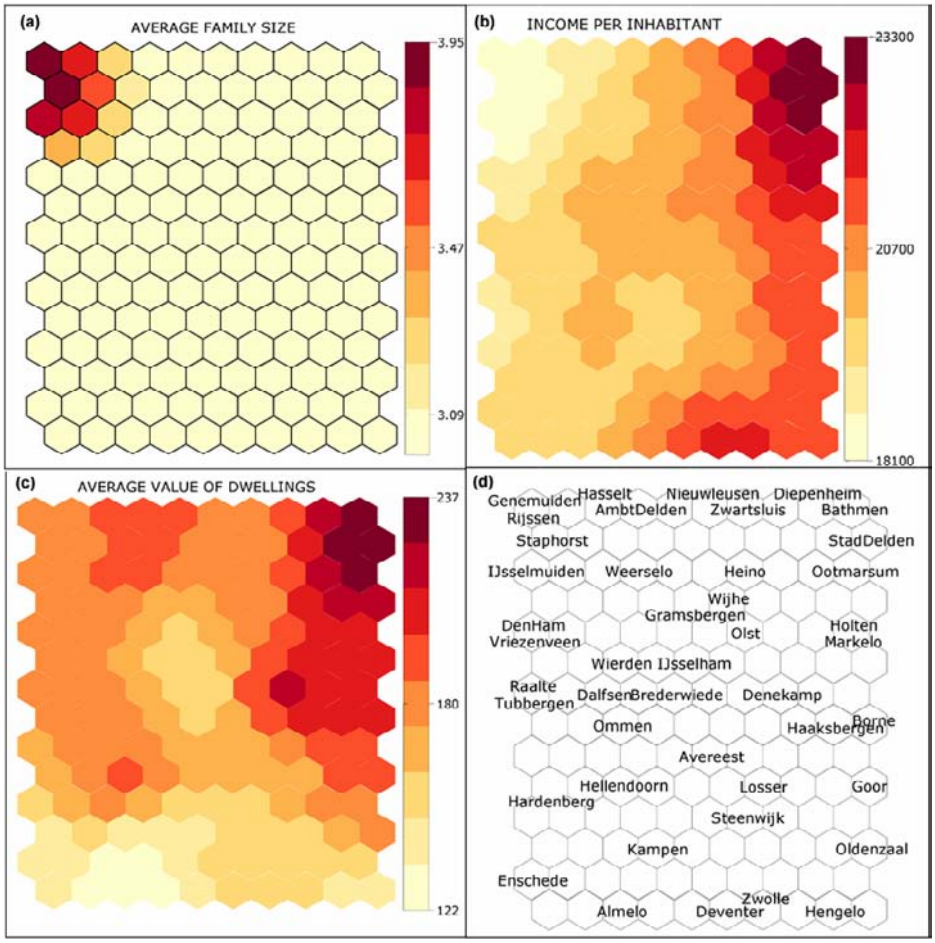


Figure 3.3. SOM component planes depicting a univariate space for selected attributes of the dataset: (a) the average family size, (b) the income per inhabitant, (c) the average value of dwellings. (d) shows the labels corresponding to the position of the map units (municipalities).

3.3.2. Unified distance matrix representation

The unified distance matrix or U-matrix (Ultsch and Siemon 1990) is a representation of the SOM that visualizes the distances between the network neurons or units. It contains the distances from each unit centre to all of its neighbours. The neurons of the SOM network are represented here by hexagonal cells (see figure 3.4). The distance between the adjacent neurons is calculated and presented with different colourings. A dark colouring between the neurons corresponds to a large distance and thus represents a gap between the values in the input space. A light colouring between the neurons signifies that the vectors

are close to each other in the input space. Light areas represent clusters and dark areas cluster separators. This representation can be used to visualize the structure of the input space and to get an impression of otherwise invisible structures in a multidimensional data space.

The U-matrix representation (figure 3.4) reveals the clustering structure of the dataset used in this experiment. Municipalities having similar characteristics are arranged close to each other and the distance between them represents the degree of similarity or dissimilarity. For example, the municipality of Enschede is well separated from the rest by the dark cells showing a long distance from the rest of the municipalities. This is expected, since Enschede is the largest and the most developed and urbanized municipality in the region. At the top left corner of the map, the municipalities Genemuiden, Rijssen, Staphorst and IJsselmuiden are clustered together. These are small localities that have common characteristics according to the data. This kind of similarity can be composed of a number of variables provided by the dataset. The U-matrix shows more hexagons than the component planes (discussed below) because it shows not only the values at map units but also the distances between map units.

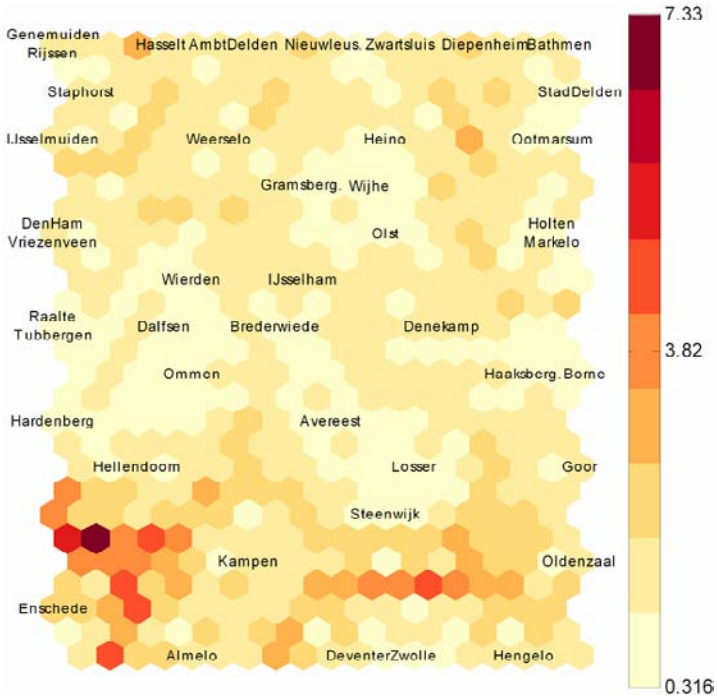


Figure 3.4. The unified distance matrix showing clustering and distances between positions on the map. Municipalities having similar characteristics are arranged close to each other and the distance between them represents the degree of similarity or dissimilarity. Light areas represent clusters and dark areas cluster separators (a gap between the values in the input space).

In contrast to other projection methods in general, the SOM does not try to preserve the distances directly but rather the relations or local structure of the input data. While the U-matrix is a good method for visualizing clusters, it does not provide a very clear picture of the overall shape of the data space because the visualization is tied to the SOM grid.

Alternative representations to the U-matrix can be used to visualize the shape of the SOM in the original input data space. Three are discussed below: 2D and 3D projections (using projection methods such as Sammon's mapping and PCA), 2D and 3D surface plots, and component planes.

3.3.3. 2D and 3D projections

The projection of the SOM offers a view of the clustering of the data with data items depicted as coloured nodes (figure 3.5). Similar data items (municipalities in this dataset) are grouped together with the same type or colour of markers. Size, position and colour of markers can be used to depict the relationships between the data items. This gives an informative picture of the global shape and the overall smoothness of the SOM in 2D or 3D space.

In 3D space, the weight of the data items according to the multivariate attributes can be represented using the third dimension to show a hierarchical order or tree structure. Exploration can be enhanced by rotation, zooming and selection in the 3D representation and by interactive manipulation of features such colour, size, and type of marker. Connecting these markers with lines can reveal the shape of clusters and the relationships among them.

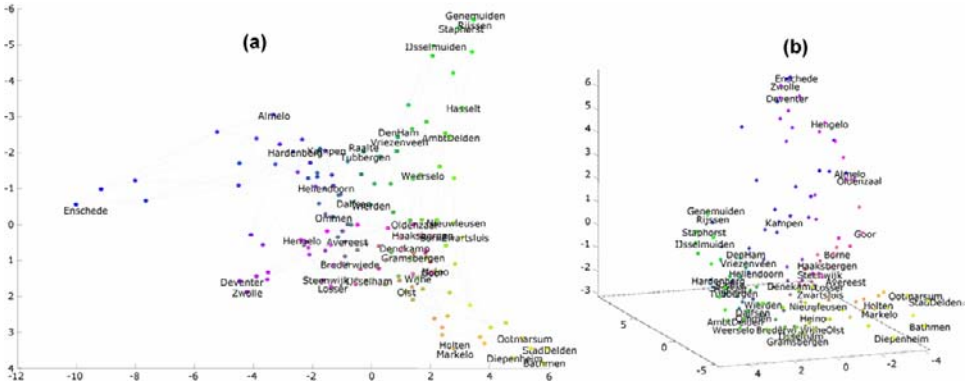


Figure 3.5. Projection of the SOM results in 2D space (a) and 3D space (b). Municipalities having similar characteristics according to the multivariate attributes in the dataset are represented using points (markers) with colour coding and connecting lines to depict relationships between them.

3.3.4. 2D and 3D surface plots

The 2D surface plot of the distance matrix (figure 3.6a) uses colour value to indicate the average distance to neighbouring map units. It is a spatialization that uses a landscape metaphor to represent the density, shape, and size or volume of clusters, and can be used for further cluster investigation in relation with the similarity representation. Unlike the projection in figure 3.5 that shows only the position and clustering of map units, areas with uniform colour are used in the surface plots to show the clustering structure and relationships among map units.

In the 3D surface (figure 3.6b), colour value and height are used to represent the regionalization of map units according to the multidimensional attributes.

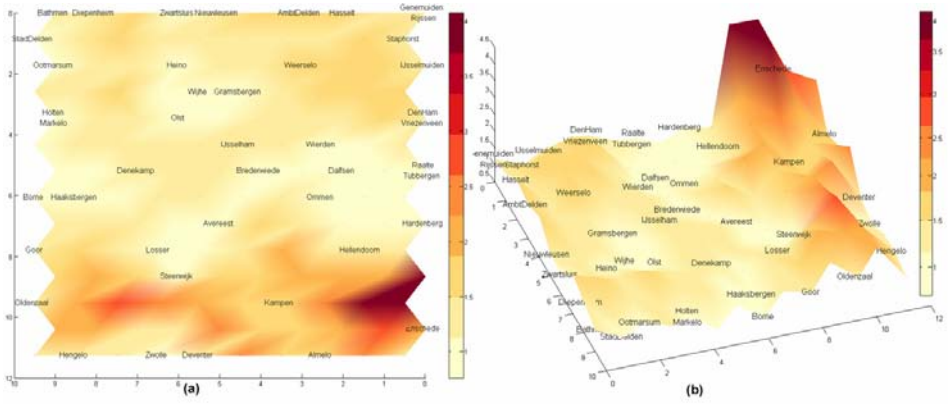


Figure 3.6. Surface plots: the density, shape, and size or volume of clusters are shown in 2D surface (a) and 3D surface (b) to depict a multivariate space. Darker colour indicates greater distance and light colour small distance.

3.3.5. Component planes

Component planes (figure 3.7) represent a multivariate visualization of the attributes of the dataset, allowing easy detection of relationships among the attributes, as described in the previous section. Each component plane shows the values of one variable for all map units, using colour coding that follows colour scheme guidelines presented by Brewer (1994). This makes it possible to visually examine every cell corresponding to each map unit or data item. By using the position and colouring, all relationships between different map units (municipalities in this dataset) can be easily explored in a single visual representation. For example, the average income per inhabitant is correlated somewhat with the number of inhabitants between the ages of 45 and 64 (INH_45-64Y) and the number of inhabitants older than 64 (INH_65_p) in municipalities such as StadDelden, Bathmen, Diepenheim, Ootmarsum, Holten and Markelo (see figure 3.3d for corresponding names of municipalities). These municipalities have the highest income of the region.

These displays can be arranged in any order (alphabetical, geographic pattern, or any order that makes it easy to see the relationships among them), in a way similar to the collection maps of Bertin (1981).

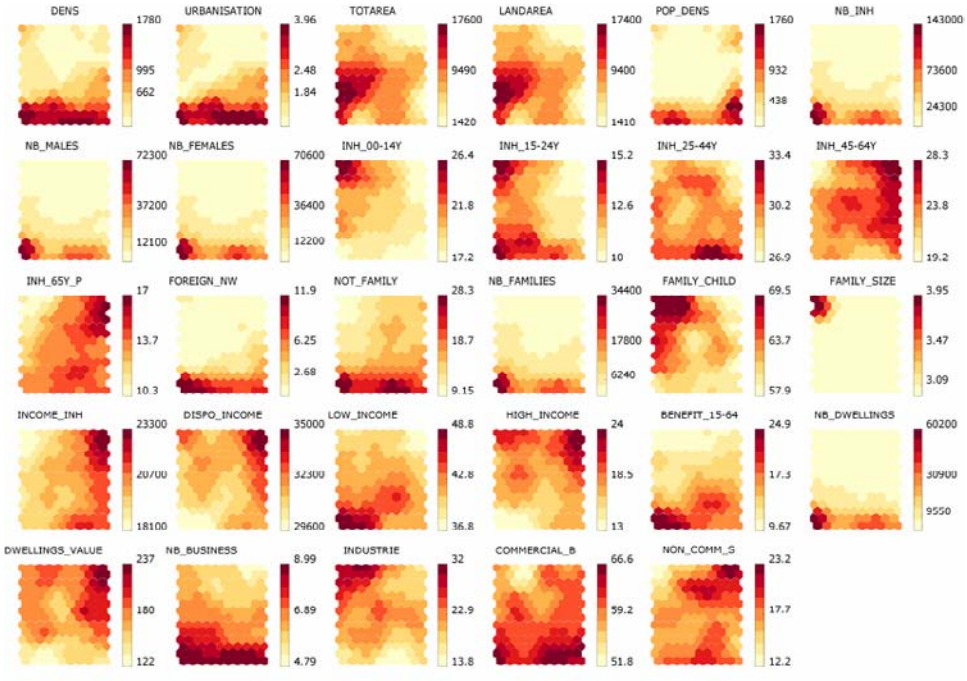


Figure 3.7. SOM component planes depicting the different attributes of the dataset and the relationships among them for all the municipalities. Relationships between different municipalities in the dataset are explored in a single visual representation.

3.4. A case of the exploration of complex relationships in geospatial data

An application of the different techniques explored in the previous section is presented for the exploration of a larger dataset with complex attributes relationships, as an example case of geospatial data exploration. Here, the four goals of the visualization described in Chapter 2 (discovering patterns through similarity representation, exploring correlations and relationships for hypothesis generation, exploring the distribution of the dataset, detecting irregularities in the data) are explored in the representation of the data described below. Similarity (patterns) is represented in the distance matrix representation and in the projections. Relationships, distribution and irregularities are viewed in fine detail with the component plane visualization.

3.4.1. The data

The example exploration uses a complex dataset on geography and economic development (Gallup et al. 1999), compiled to support analysis of the complex relationships between geography and macroeconomic growth (e.g. the ways in which geography may directly affect growth, and the effect of location and climate on income levels, income growth, transport costs, disease burdens and agricultural productivity). Additionally, the relationships between geographic regions, whether located far from the coast, and population density, population growth, economic growth and the economic policy itself are other aspects the study of this dataset intends to explore. The dataset contains 48 variables on the economy, physical geography, population and health of 150 countries. This dataset will be used in the usability test in Chapter 7.

Table 3.1 describes the variables of the dataset. Table 3.2 gives a list of the countries included in the study with their codes used in the SOM representations.

For further investigation of the SOM visualizations in the exploration of the dataset, some maps are represented in figure 3.8 for selected attributes: coastal population density, percentage population within 100 km of coast or river, GDP per capita, distance (km) to closest major port in Europe, percentage of land area in the subtropics, and percentage of land in the geographic tropics.

Table 3.1. Description of the variables of the dataset

Variable	Description	Variable	Description	Variable	Description
gdp50	gdp per capita in 1950	ciffob95	shipping cost, 1995	pop95	population in 1995
gdp90	gdp per capita in 1990	tropicar	% land in geographic tropics	zpolar	% land area in polar non-desert
gdp95	gdp per capita in 1995	troppop	%population in geographic tropics, 1994	zboreal	% land area in boreal regions
gdp65	gdp per capita in 1965	malfal66	malaria index, 1966	zdestmp	temperature desert
gdpg6590	gdp per capita growth from 1965 to 1990	maffal94	malaria index 1994	zdestrp	tropical + subtropical desert
lnd100km	% land within 100 km coast	lhcpc	log hydrocarbons per capita 1993	zdrytemp	% land area within dry temperature
pop100km	% population within 100km coast	south	southern hemisphere countries	zwettemp	% land area wet temperate
lnd100cr	% land within 100 km coast or river	landarea	land area (sq km)	zsubtrop	% land area in the subtropics
pop100cr	% population within 100km coast or river	open6590	openness, 19965-1990	ztropics	% land area in the tropics
dens65c	coastal population density, 1965	icrg82	quality of public institution,	zwater	water (lakes and ocean)
dens65i	inland population density, 1965	newstate	timing of independance	eu	western europe
dens95c	coastal population density, 1995	socialist	socialist country, 1950-1995	safri	sub-saharan africa
dens95i	inland population density, 1995	lifex65	life expectancy 1965 (UN)	sasia	south asia
landlock	landlocked	syr15651	log years secondary schooling, 1965	transit	transition countries
lnadlneu	landlocked, not west and central europe	urbpop95	% population urban, 1995 (world bank)	latam	latin america and caribbean
airdist	km to closest major port	wardum	had external war, 1960-1985	eseasia	east and southeast asia

Table 3.2. The country codes used in the training of the neural network

code	country	code	country	code	country	code	country
AFG	Afghanistan	ERI	Eritrea	LBR	Liberia	RUS	Russian Federation
AGO	Angola	ESP	Spain	LBY	Libya	RWA	Rwanda
ALB	Albania	EST	Estonia	LKA	Sri Lanka	SAU	Saudi Arabia
ARE	United Arab Emirates	ETH	Ethiopia	LSO	Lesotho	SDN	Sudan
ARG	Argentina	FIN	Finland	LTU	Lithuania	SEN	Senegal
ARM	Armenia	FRA	France	LVA	Latvia	SGP	Singapore
AUS	Australia	GAB	Gabon	MAR	Morocco	SLE	Sierra Leone
AUT	Austria	GBR	United Kingdom	MDA	Moldova	SLV	El Salvador
AZE	Azerbaijan	GEO	Georgia	MDG	Madagascar	SOM	Somalia
BDI	Burundi	GHA	Ghana	MEX	Mexico	SVK	Slovak Republic
BEL	Belgium	GIN	Guinea	MKD	Macedonia	SVN	Slovenia
BEN	Benin	GMB	Gambia	MLI	Mali	SWE	Sweden
BFA	Burkina Faso	GNB	Guinea Bissau	MMR	Myanmar	SYR	Syrian Arab Rep.
BGD	Bangladesh	GRC	Greece	MNG	Mongolia	TCD	Chad
BGR	Bulgaria	GTM	Guatemala	MOZ	Mozambique	TGO	Togo
BIH	Bosnia and Herzegovina	HKG	Hong Kong	MRT	Mauritania	THA	Thailand
BLR	Belarus	HND	Honduras	MUS	Mauritius	TJK	Tajikistan
BOL	Bolivia	HRV	Croatia	MWI	Malawi	TKM	Turkmenistan
BRA	Brazil	HTI	Haiti	MYS	Malaysia	TTO	Trinidad & Tobago
BWA	Botswana	HUN	Hungary	NAM	Namibia	TUN	Tunisia
CAF	Central African Rep.	IDN	Indonesia	NER	Niger	TUR	Turkey
CAN	Canada	IND	India	NGA	Nigeria	TWN	Taiwan
CHE	Switzerland	IRL	Ireland	NIC	Nicaragua	TZA	Tanzania
CHL	Chile	IRN	Iran	NLD	Netherlands	UGA	Uganda
CHN	China	IRQ	Iraq	NOR	Norway	UKR	Ukraine
CIV	Côte d'Ivoire	ISR	Israel	NPL	Nepal	URY	Uruguay
CMR	Cameroon	ITA	Italy	NZL	New Zealand	USA	United States
COG	Congo	JAM	Jamaica	OMN	Oman	UZB	Uzbekistan
COL	Colombia	JOR	Jordan	PAK	Pakistan	VEN	Venezuela
CRI	Costa Rica	JPN	Japan	PAN	Panama	VNM	Vietnam
CUB	Cuba	KAZ	Kazakhstan	PER	Peru	YEM	Yemen
CZE	Czech Republic	KEN	Kenya	PHL	Philippines	YUG	Yugoslavia
DEU	Germany	KGZ	Kyrgyz Republic	PNG	Papua New Guinea	ZAF	South Africa
DNK	Denmark	KHM	Cambodia	POL	Poland	ZAR	Zaire
DOM	Dominican Republic	KOR	Korea	PRK	Korea Dem.People's Rep.	ZMB	Zambia
DZA	Algeria	KWT	Kuwait	PRT	Portugal	ZWE	Zimbabwe
ECU	Ecuador	LAO	Lao PDR	PRY	Paraguay		
EGY	Egypt	LBN	Lebanon	ROM	Romania		

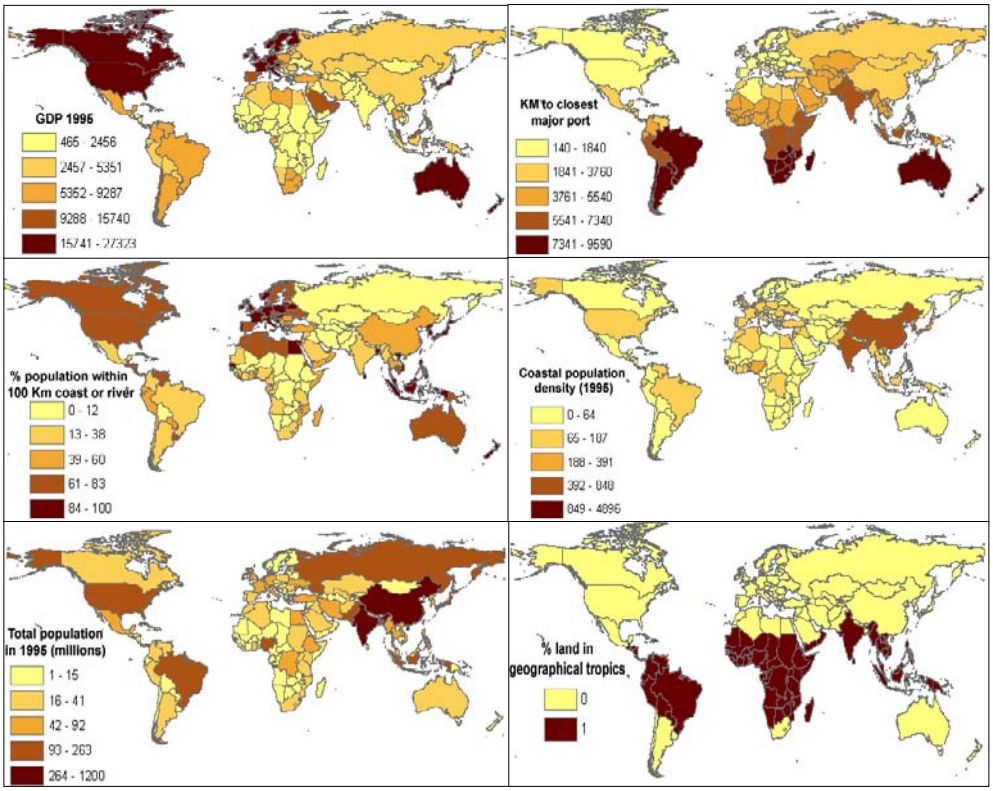


Figure 3.8. Some attributes of the dataset explored: GDP per capita, distance (km) to closest major port in Europe, percentage population within 100 km of coast or river, coastal population density, total population, and percentage of land in the geographic tropics.

3.4.2. General patterns visualization

The general patterns are represented by the unified distance matrix. The unified distance matrix is used to reveal the commonalities between the countries, based on the multivariate attributes (figure 3.9a). At the top of the map, we have the poor economies, mostly the African countries, and at the bottom the rich economies.

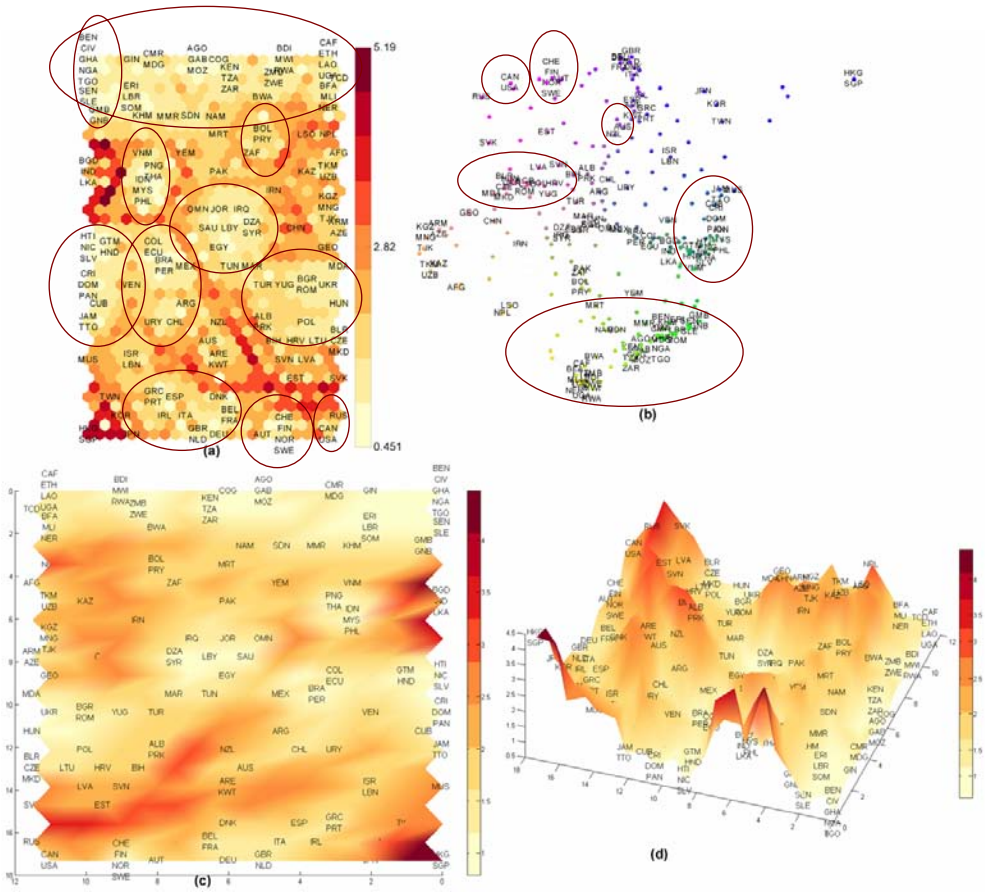


Figure 3.9. Similarity matrix representation of the dataset (a), PCA projection of SOM results (b), 2D surface plot of distance matrix (c) and 3D surface plot of distance matrix (d).The circles in (a) and (b) show the clusters also revealed in (c) and (d), and discussed in the text.

From this clustering structure, differences can be observed between countries in different parts of the world. A very striking observation is that the clustering somehow reflects the geography of the countries. This confirms the general hypothesis suggesting that there is a relationship between the geography of the countries and economic growth (Gallup et al. 1999). Even further clustering that reflects the distinct geographic regions is obtained with the similarity matrix representation: West Africa, Southern Africa, the Middle East, Europe, South America, North America (USA and Canada), and Asia. The European countries are in three different clusters next to each other; USA and Canada are clearly the richest economies.

A few cases do not reflect this geographic relationship. Laos is found in a cluster with some poor African economies (Central Republic of Africa, Ethiopia, Uganda, Chad, Burkina Faso, Mali, Niger). This may be because Laos's economic characteristics are low compared with those of the other Asian countries and it falls closer to Africa than Asia in this respect. Other countries that have no obvious characteristics in common with the others in the same geographic region include Mauritania, Yemen, Pakistan, Iran and Mauritius. South Africa has particular characteristics that position the country far away from other African countries and closer to the Middle East, Iran and Pakistan, on the one hand, and close to Bolivia and Paraguay on the other.

The same information provided in the distance matrix can be viewed using a projection (figure 3.9b), and 2D or 3D surfaces (figure 3.9c and 3.9d) in different perspectives.

3.4.3. Exploration of correlations and relationships

The exploration of correlations and relationships can be done using the component plane visualization (figure 3.10). The component planes show the values of the different attributes at different locations on the SOM grid.

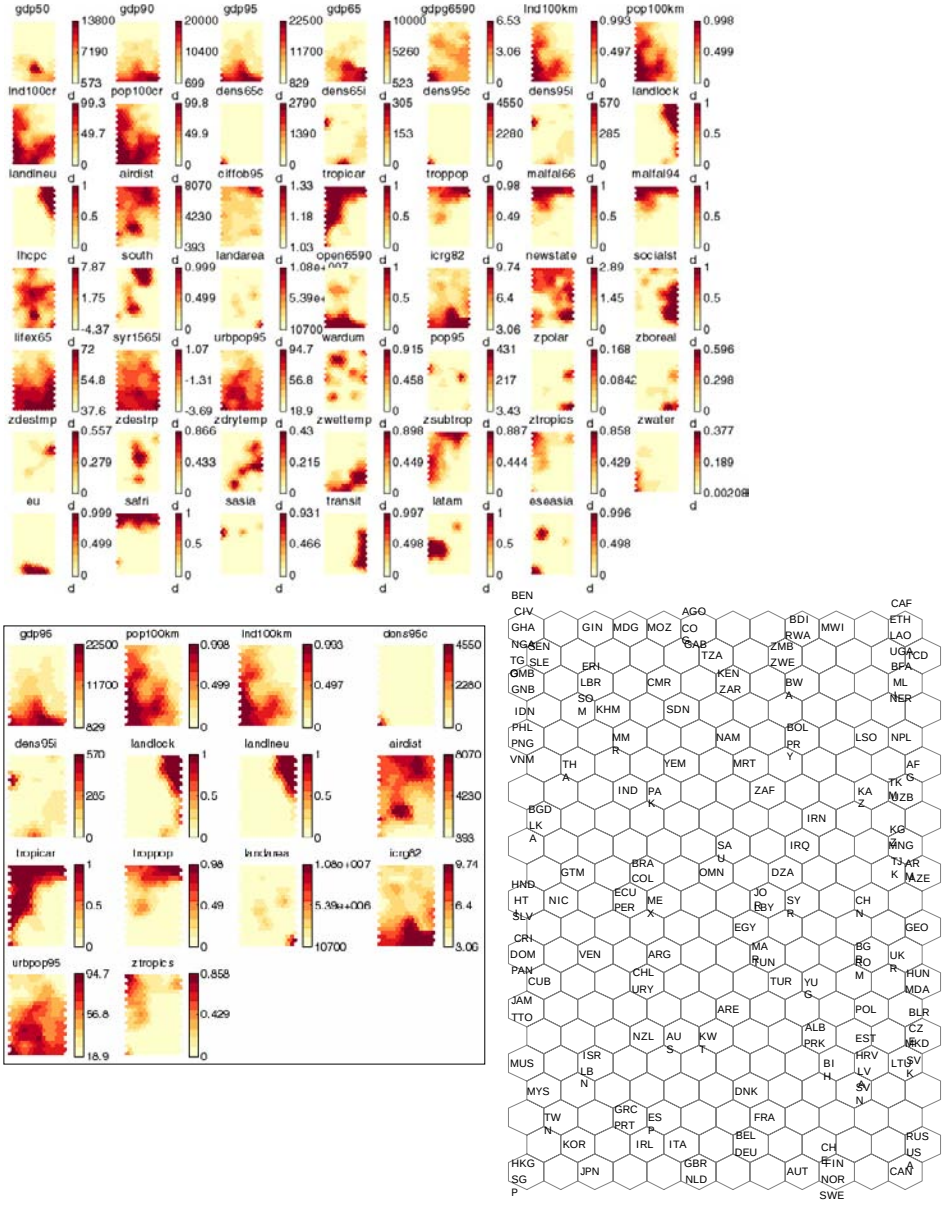


Figure 3.10. The component plane visualization: all the components at the top and selected components related to economic development and examined in the chapter are shown at the bottom. Labels of the components (countries) are shown at the bottom right.

A summary of the geographic patterns was made based on the average GDP per capita, total population and land area, and several key variables that can be related to economic development: the extent of land in the geographic tropics, the proportion of the population within 100 km of the coastline or within 100 km of the coastline or ocean-navigable river, the percentage of population that lives in landlocked countries, the average distance by air (weighted by country populations) to the closest core economic areas, the density of human settlement (population per square km) in the coastal region (within 100 km of the coastline) and the interior (beyond 100 km from the coastline). These variables are presented in the component plane visualization in figure 3.10 (bottom left). The tropical countries were defined as being those that have half or more of the land area in the geographic tropics.

From these patterns the following question can be raised: How great a role has geography played in economic growth, assuming that economic policies and institutions are well established?

The exploration here is limited to the dataset examined. Other environmental and political factors that are not included in the dataset might have some influence on economic development.

This complex linkage between geography, demography, health and economic performance requires closer examination. Using the SOM visualization, we examine two geographic correlates of economic development that were outlined by Gallup et al. (1999) and generate other possible hypotheses that the SOM technique allows in such a complex dataset. The countries in the geographic tropics are nearly all poor. Almost all high-income countries are in the mid and high latitudes. Coastal economies are generally higher-income than the landlocked economies.

From the component plane visualization in figure 3.10, a simple view of the displays allows the attributes to be visually related to the spatial locations and hypotheses to be generated based on observed correlations and relationships. The SOM component planes can be ordered so that the displays that seem to have high correlation are placed next to each other, in a way similar to the collection maps of Bertin (1981). From these displays in figure 3.10, it can be easily observed in one single view that the poorest economies (reference to the 1995 GDP from the dataset) have characteristics such as large proportion of land and population in the geographic tropics, population highly concentrated in the interior, often landlocked, small proportion of land within 100 km coast or river, located in the southern hemisphere, small proportion of land in wet temperature, and often with tropical or subtropical deserts. Most of these characteristics were identified as closely associated with low income in general (Gallup et al. 1999). Other common characteristics of these countries that can be seen as a consequence of the low income are also visualized in the component planes. The poor countries have low life expectancy, high shipping costs, and heavy disease

burdens of malaria; they are very far from the closest core markets in Europe, and many have external wars. From these observations, it can be hypothesized that various aspects of tropical geography and public health are vitally important and affect economic growth (Bloom and Sachs 1998). South Asia, Latin America, the eastern European countries and the former Soviet Union are like Sub-Saharan Africa, with more concentrated in the interior rather than at the coast. Landlocked countries may be particularly disadvantaged by their lack of access to the sea. They all have low income except those in western and central Europe (integrated into regional European market and associated low-cost trade). High population density seems to be favourable for economic development in coastal regions with good access to internal, regional and international trade. The poorest economies have low urban population density. The urban areas seem to develop more in the coastal regions.

3.5. Usability evaluation plan

After introducing the potential of the SOM and its graphical representations for visual exploration of geospatial data, the question of its effectiveness and efficiency remains to answer. This section presents a general strategy for assessing the proposed visual-computational approach (Chapter 2) and related graphical representations (Chapter 3) for exploring multivariate geospatial data. The evaluation is conducted at a later stage in Chapter 6 and 7. Here a preliminary involvement of the users is made to guide the design process in the next chapter.

3.5.1. Overview of the evaluation plan

We present a general approach useful for the empirical assessment of the intended exploratory visual-computational environment. The strategy is developed for examining the effectiveness of alternative representation forms (the SOM-based visualization environment presented above), and visual variable choices within those forms. A user-centred approach to developing integrated computational-visual analysis tools (whether based on the SOM or other inductive learning methods) should include attention to the user's understanding of the representation forms of multivariate relationships in the data. This strategy focuses on gathering information about how users interpret and understand the basic visualization features and representation forms, in order to improve their design. Specifically, we are interested in knowing whether users can actually comprehend the meaning of the proposed representations, how the different representation forms influence the effectiveness of the visualization tool in terms of analysis and exploration of data, and what type of representation is suitable for exploratory tasks.

To this end, the general evaluation strategy is to conduct a usability test comparing different options for map-based visualization of the output of SOM multivariate analysis. Since the design of the visualization environment is based on a user-centred approach (see Chapter 4), early involvement of users was necessary and took the form of a preliminary interface feature inspection in which several aspects of the representation forms, graphics and colour schemes were presented to users for analysis.

3.5.2. Usability inspection

Feature inspection is a usability inspection method (see Chapter 6 for more detail on usability evaluation methods) that involves the evaluation of functions delivered in a software tool. As part of the iterative design process (see Chapter 4), a preliminary feature inspection was conducted at this stage in the development of the visualization environment, in order to gather users' views and preferences about features of the different SOM-based graphical representations described above. An empirical user test was conducted at a later stage in the development of the prototype in Chapters 6 and 7.

The initial inspection of the prototype graphical representations described in the previous section helped identify a number of usability problems. The use of colour hue was a general problem observed by the participants with cartographic design background. They generally suggested the use of one colour scaled from light to dark. Grey scale was found to be generally difficult to use for investigating the clusters in the graphical representations. A general problem concerned the difficulty in relating different representation forms (the unified distance matrix, 2D and 3D projections, 2D and 3D surfaces). Another general observation concerned the need for interaction to support visual exploration of the patterns and relationships.

The suggestions and opinions of the participants in the inspection of the graphical representations provided a guideline for improving the design of the graphical representations. As a result a different colour scheme based on (Brewer 1994) was used in the current version of the graphical representations. The design has incorporated multiple views that relate the different representations and combine with maps to provide a geographic reference for the users during exploration activities (see the next Chapter). All graphical representations now use colour and not grey scale to represent clusters and distances. The graphical representations are provided with interaction tools for rotation, zooming, selection, and free form flying for the 3D views.

3.6. Conclusion

In this chapter we have explored some graphical representations based on the SOM, in relation to a usability framework for the design of an exploratory geovisualization environment based on the visual-computational approach presented in Chapter 2. The SOM graphical representations were examined together with maps in two example cases. The first example was basically used to illustrate the SOM graphical representations. The second example was a case of the exploration of more complex attribute relationships in a larger dataset.

With the map, and without any prior hypothesis, some individual attributes were presented. Although the patterns in the data were visible, all the attributes needed to be mapped to allow a complete visual comparison, and draw conclusions. This can be difficult for visual comparison and analysis.

The SOM offers a summary of the patterns in the data with the distance matrix representation and projections. Commonalities between the map units based on the multivariate attributes were easily viewed. From the clustering structure offered by the distance matrix representation and the projection, the different categories of the map units could be observed. The exploration of correlations and relationships was possible with the SOM component plane display, which presents in a single visual representation, the relationships between the attributes of the dataset. Two variables that are correlated are represented by similar displays, so correlations and relationships are easily detected visually. This provides ground for generating or further exploring hypotheses.

As a general strategy to assess how users understand the representations, and to improve the overall effectiveness of the design of the exploratory visual-computational environment, an assessment plan was proposed. The plan is based on a usability test involving a representation of intended users and a number of selected tasks performed in visualization environments. This assessment plan will be applied in the specific case of the design of the proposed visualization environment, in Chapter 6 and 7, in order to assess the design concepts and aspects of the implementation of the computational-visual analysis environment. This includes an assessment of the appropriateness of the representation metaphors applied to, as well as of the visual variables used in, the design of specific representations. The assessment can provide some insight into the effectiveness and usefulness of, and user reactions (users' preferences and views) to, the representations for exploratory visual analysis, interpretation and understanding of the structure in the dataset. Here a preliminary involvement of the users is made to guide the design process in the next chapter.

The next steps in developing the visualization environment discussed in this chapter will focus on the extension and improvement of the graphical representations presented, following the usability framework. One part of this work will be to integrate the representation approaches into a multiple-view

approach to visualization. By combining user interactions with these forms of representation, the visualization environment will be extended and improved to focus on the interactive manipulation of the representations to support the cognitive activities involved in the use of the visualization environment, and to provide the querying and exploration of features in a user-friendly interface. Such advances are likely to have additional impacts upon the user's preferences and responses.

Chapter 4

User interface design for geovisualization: visual interaction for knowledge discovery

4.1. Introduction

More integrated visualization tools are needed for the extraction of patterns and relationships in data. The integration of feature extraction tools with appropriate user interfaces is important to support the user's understanding of underlying structures and processes in geodata. Designing such tools is one of the major research areas in geovisualization.

An interesting development in the design of geovisualization environments is the integration of information visualization and cartographic methods for the exploration of geospatial data. Design methods for applying information visualization and scientific visualization techniques in geovisualization are, however, not clearly defined. In particular, the design of interactive representation forms still lacks a delineation of fundamental operations that users might apply to an interactive map or related graphics, as well as guidelines for their appropriate application (MacEachren 2000). Some authors have proposed a set of visualization operations (Keller and Keller 1992; Qian et al. 1997), but no comprehensive description of the operations and guidelines for design is available.

On a more conceptual level, Bertin's (1983) concept of graphical constructions can be a guide to establishing guidelines for the manipulation of graphics in today's graphical interfaces. Cartographic methods serve as the basis for most representation methods used in information visualization (Fabrikant 2001b). On the other hand, information visualization techniques are applied in cartography for the design of dynamic and interactive displays. This integration of cartographic

---

This chapter is based on:

Koua E. L. and Kraak, M. J. (2004). Geovisualization to support the exploration of large health and demographic survey data. International Journal of Health Geographics 2004, 3:12.

Koua E. L. and Kraak M. J. (2004). Integrating computational and visual analysis for the exploration of health statistics. In: SDH 2004: Proceedings of the 11th international symposium on spatial data handling : advances in spatial data handling II. : 23-25 August 2004, University of Leichester. / ed. by P.F. Fisher. - Berlin etc.: Springer, 2004. pp. 653-664.

methods with information visualization techniques can help provide ways of exploring large geospatial data, and support knowledge construction.
This chapter presents key design issues and a prototype geovisualization environment in which these design concepts are used to integrate representation forms with visualization and interaction techniques for the exploration of patterns and relationships in large geospatial datasets. An example of exploration of a dataset on health statistics on Africa is used to demonstrate the integration of the different graphical representations and the different options of the user interface.

4.2. Visual exploration support for large geospatial data

The basic idea of visual data exploration is to present the data in some visual form, allowing the human to gain insight into the data and draw conclusions (Keim 2002). Visual data mining is the use of visualization techniques to allow users to evaluate, monitor and guide the inputs and the process of data mining. Two main concepts are integrated into the visual data mining framework: feedback based on the knowledge discovery process and visualization (including mapping, filtering capabilities). Although direct manipulation techniques can facilitate interaction (Eick 1997), they can be difficult to apply to the process of data visualization in very large multidimensional datasets. The data must be processed in some way before they can be manipulated. The proposed framework explores ways of effectively extracting patterns, using data mining based on the self-organizing map (SOM), and of representing the results, using graphical representations for visual exploration. As presented in figure 4.1, the data mining stage allows a clustering (similarity matrix) of the multidimensional input space to be constructed, using the SOM training algorithm tool (SOM toolbox) and graphics processing with Matlab software. From this computational process, the global structure and patterns can be represented with graphical representations and maps (geographical view) of similarity results. Further exploration can be carried out on the relationships and correlations among the attributes. The framework includes spatial analysis, data mining and knowledge discovery methods, supported by interactive tools that allow users to perform a number of exploratory tasks in order to understand the structure of the dataset as a whole, as well as to explore detailed information on individual or selected attributes of the dataset. Different representation forms are integrated and support user interaction for exploratory tasks to facilitate the knowledge discovery process. They include some graphical representations based on the SOM, maps, and other graphics such as parallel coordinate plots. Cartographic methods support this design for the effective use of visual variables with which the visualizations are depicted. The graphical representations can be interactively manipulated in the Matlab graphical interface, using rotation, zooming, panning, and brushing.

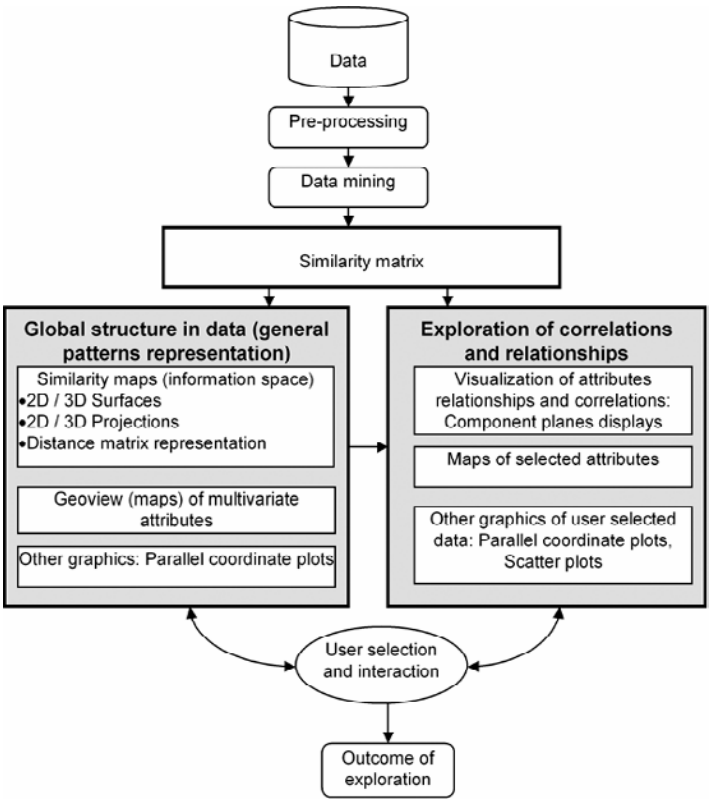


Figure 4.1. Data exploration framework: from the computational process, global structure and patterns can be visualized with graphical representations and maps of similarity results. Relationships and correlations among the attributes are presented with interactive graphical representations, maps, and other graphics such as parallel coordinate plots.

4.3. Conceptual design

Most design models focus on the available technologies for improving interactions and providing expert design guidelines on the way a user interface can be useful. To meet the goals of the geovisualization tool, we developed a user-centred approach in order to emphasis the central value of user tasks and visualization operations in the design of such environments. A focus on users can provide useful guidelines for designing an environment that better corresponds to users' analysis needs.

In the next subsections, we examine some of the design issues, including a usability specification, a conceptual approach to examining basic visualization

tasks and operations, and a model for integrating information, visualization and cartographic methods for geovisualization.

4.3.1. User-centred design

User-centred design (UCD) is an approach that views knowledge about users and their involvement in the design as a central concern. It aims at identifying the prospective users, studying their activities and how they perform them, and identifying what they need in order to perform the tasks better. Taking into account these user characteristics implies including them to some extent in the design process. UCD involves tests and evaluations with users in an interactive design process. The traditional approach for developing software advocates a number of processes that are produced in an essentially linear fashion (design, development, implementation, test). In UCD, however, a star model is used (see figure 4.2).

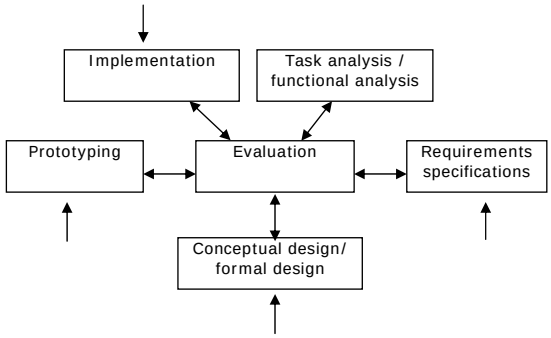


Figure 4.2. The star model (Source: Hix and Hartson 1993)

In this model, task analysis is a fundamental activity in usability specification. It allows user interaction with the system to be structured according to:

- Goals
- Methods: sequences of operators, or procedures for accomplishing a goal
- Operators: the basic actions available to the user for performing a task.

A number of these concepts, including goals, tasks, operations, plans and hierarchies, were introduced in the literature on task analysis. Task analysis is useful to the extent that it helps improve the design or implementation of the system by gathering information, representing it in an appropriate manner, and then utilizing this representation to establish the system improvement (Sheperd 1989). Task analysis models assume that a task can be broken down into a series of tasks. One of the most popular approaches is hierarchical task analysis, which can yield many practical benefits in complex situations by establishing a task description hierarchy that complies with a particular set of rules, allowing the user's goal to be described as a hierarchy of operations and plans. Based on the

task analysis, a model of user/system interaction can then be constructed, usually hierarchically composed and emphasizing the sequence of operations.

4.3.2. Usability specification for geovisualization design

The design of the visualization environment is based on a usability framework structured to develop a tool that is useful and appropriate for the user needs and tasks (see Chapter 3). This framework not only includes the techniques, processes, methods and procedures for designing usable products and systems, it also focuses on the user's goals, needs and tasks in the design process (Rubin 1994). User characteristics, visualization tasks and operations are examined in order to improve user interaction and support activities involved in the use of the visualization environment and in related information spaces.

One of the objectives of human-computer interaction (HCI) research is to achieve two goals: usefulness and usability. Usefulness refers to the achievement of the user's goals, and addresses the way in which the tool supports the user's tasks. Usability refers to the extent to which users can use the visualization environment to achieve specific goals effectively, efficiently and to their satisfaction. The usability specification is drawn up by gathering knowledge on the context of use (the characteristics of potential users, their environment, their tasks, and the main objectives) and determining user requirements for such tools, system requirements, a definition of the functionality of the system, and ultimately the design of the human-computer interface.

Specifying general user requirements for geovisualization

Specifying user requirements in software system development for general purposes and a wide target group is a difficult task. A classical study of user requirements is conducted by gathering information on potential users' preferences, objectives and views. Ideally, the system should be designed with a complete knowledge of the intended application domain and task structure. This is, however, constrained by the complex and changing nature of users' tasks, and because software typically needs to be designed to suit a wide range of users. User modelling and user profiles help achieve this goal. In spatial data analysis and geovisualization environments, it becomes even more difficult to derive user requirements as these may depend on several factors: individual user needs, tasks, knowledge domain, type of data, etc. Since the analysis needs for each dataset are often unique, some of the best visualizations are task-oriented. Nevertheless, some general requirements in spatial data analysis can be identified. We propose some requirements that can be appropriate for geovisualization, based on the analytical aspect of spatial analysis:

- Accuracy in results
- Effective extraction of patterns and relationships

- Flexibility in use (scaling, rotation, panning, brushing, browsing, focusing)
- Adaptability (appropriate to the task and applicable for different situations)
- Exploration-oriented
- Multiple (alternative) views to consolidate knowledge construction.

Other issues include the detection of irregularities (unusual and predictable behaviour), and knowledge discovery support.

Specification for the human-computer interface

Besides these analytical aspects, which are more focused on the quality of data analysis, a number of human-computer interface design issues can be added. Some key aspects of the usability of the human-computer interface, based on those proposed by Ravden and Johnson (1989), can be useful for the design of geovisualization tools. They include visual clarity, consistency, compatibility, informative feedback, explicitness, appropriate functionality, flexibility and control, error prevention and correction, and user guidance and support. These design guidelines can also serve as indicators for evaluating the human-computer interface. Cartographic visualization methods (Kraak 1998) provide other guidelines for navigation and the exploration of the data. They include controlling the user's viewpoint, a clear description of the environment, querying, browsing the database, and providing different views (maps and other graphics).

Examining visualization tasks and operations

The main goal of geospatial data analysis is to find patterns and relationships in the data that can help solve a particular geo-problem. The analysis process can be viewed as a set of tasks and operations. To understand visualization operations needed in geospatial data exploration, we can examine the actual tasks users are likely to perform in relation to finding patterns and relationships. We first look at how analysis tasks are generally performed from a cognitive perspective. In this respect, Norman and Draper (1986) provided a summary of analysis tasks in seven stages (see figure 4.3). These include establishing the goal, forming the intention, specifying the action sequence, executing the action, perceiving the system state, interpreting the state, and evaluating the system state with respect to the goals and intentions.

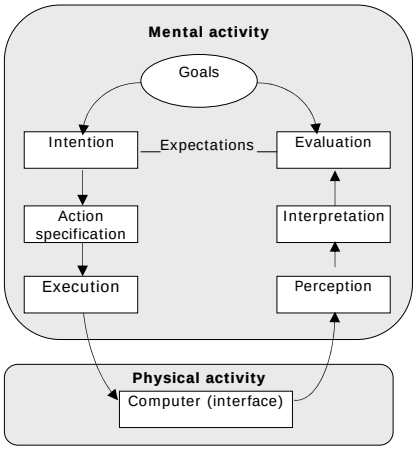


Figure 4.3. Seven stages of user activities involved in the performance of a task. Adapted from Norman and Draper (1986)

Fundamentally, two main categories of tasks are identified in figure 4.3: the mental activity and the physical activity. In general, the mental activity represents the user's psychologically expressed goals, which demands physical controls and task variables. The user starts with goals and intentions, which are psychological variables that exist in the mind of the user and are related to his or her needs and concerns. The task is performed on the physical system with physical mechanisms, resulting in changes to the physical variables and system state, which are then interpreted and evaluated in relation to the set goals.

Although in the real world, these stages may appear to be out of order (some may be skipped, some repeated), they can be a basis for examining basic visualization tasks. In exploratory data analysis, the user might not have specific goals other than the general purpose of finding patterns and relationships in the dataset. However, each step of the seven stages described above can help define specific analysis and visualization tasks. One of the few attempts to provide a description of visualization operations, and based on visualization goals, was proposed by (Keller and Keller 1992) in seven broad categories:

- Comparing: positions, datasets, subsets of data, images
- Distinguishing: importance, objects, activities, range of value
- Indicating directions: orientation, order, direction of flow
- Locating: position relative to axis, object, map
- Relating: concepts (e.g. value and direction, position and shape, temperature and velocity, object type and value)
- Representing values: numerical value of data
- Revealing objects: exposing, highlighting, bringing to the front, making visible, enhancing visibility.

We emphasize three main operations, part of which are included in the list provided above.

- Categorize and classify: identify the different categories and classifications

- Compare: review relationships, commonalties and differences, the different classifications, etc.
- Evaluate: analyze relevance of information, interpret.

4.3.3. Integrating information visualization and cartographic methods

An important issue in the design of geovisualization environments is to provide ways of representing similarity (patterns) and relationships in a way that facilitates the perceptual and cognitive processes involved (MacEachren 1995). To achieve this goal, cartographic design principles are needed to provide an effective integration of visual variables used in the representation forms, while information visualization techniques provide alternatives for the user interaction necessary to complete the tasks. Bertin's fundamental six visual variables (Bertin 1983) for graphical information processing can serve as the basis for this integration. These variables (size, value, texture/grain, colour, orientation and shape) can be used, either alone or in combination, to depict different arrangements of objects in the graphical representations. For example, size is an effective perceptual data-encoding variable and shape is useful for visual segmentation. Although direct manipulation techniques can facilitate interaction, they can be difficult to apply to the process of data visualization in very large multidimensional datasets. The data must be processed in some way before it can be manipulated. We use a SOM neural network for extracting patterns and relationships in the data, a first step in the visual-computational analysis environment. We base the design of the geovisualization environment on a model of the user's visualization tasks and operations (described above), a user perception model that contains generic interpretation capabilities of the human visual system, a user profile for the specific user preferences derived from the early involvement of the user in an inspection exercise (Chapter 3), and a model of the user's visualization goals.

4.4. Prototype exploratory geovisualization system design

Based on the conceptual design approach described above, we implemented a prototype geovisualization environment. The visualization environment is intended to contribute to the analysis and visualization of large amounts of data, as an extension of the many geospatial analysis functions available in most GIS software. The objective of the tool is to help uncover structure and patterns that may be hidden in complex geospatial datasets, and to provide graphical representations that can support understanding and knowledge construction. The design of the visualization environment incorporates several graphical representations of SOM output, including a distance matrix representation, 2D

and 3D projections, 2D and 3D surfaces, and component plane displays (described in Chapter 2).

The subsections that follow provide a description of the computational analysis and system architecture, the user interface, the functionality of the visualization environment, and some interaction design issues.

4.4.1. Structure of the integrated visual-computational analysis and visualization environment

We have extended the graphical representations of the SOM results (described in Chapter 3), to highlight different characteristics of the computational solution and integrate them with other graphics into multiple views to allow brushing and linking for exploratory analysis and knowledge discovery purposes. There are a number of researches reflecting the interest in dynamic displays on the part of experts in cartographic data presentation (Egbert and Slocum 1992; Monmonier 1992; Cook et al. 1996; Dykes 1997). Most often they suggest that brushing be applied to a map linked with one or more non-geographical presentations, showing individual values and statistics, and the visualization of neighbourhood relationships. We use multiple views to offer alternative and different views of the data in order to stimulate the visual thinking process that is characteristic of visual exploration. Cartographic methods support the design for the effective use of visual variables with which the visualization is depicted. This makes the exploratory geovisualization environment appropriate for relating the position of the map units and the value at the map units represented by colour coding, and for exploring correlations and relationships. The design incorporates several graphical representations that provide ways of representing similarity (patterns) and relationships. They are illustrated in figure 4.5, and include a distance matrix representation, 2D and 3D projections, 2D and 3D surfaces, and component plane visualization.

The tool was developed based on the integration of Matlab, the SOM toolbox and spatial analysis (Martinez and Martinez 2002). The main functionality of the visualization system includes pre-processing, the initialization and training of a SOM network, and visualization. Figure 4.4 describes the structure of the geovisualization system. The pre-processing consists of transforming primary data and converting them into an appropriate format. At this stage, input data are transformed and all components and variables of the dataset are normalized. After training the network, the visualization component provides features for visualizing the data, using different techniques.

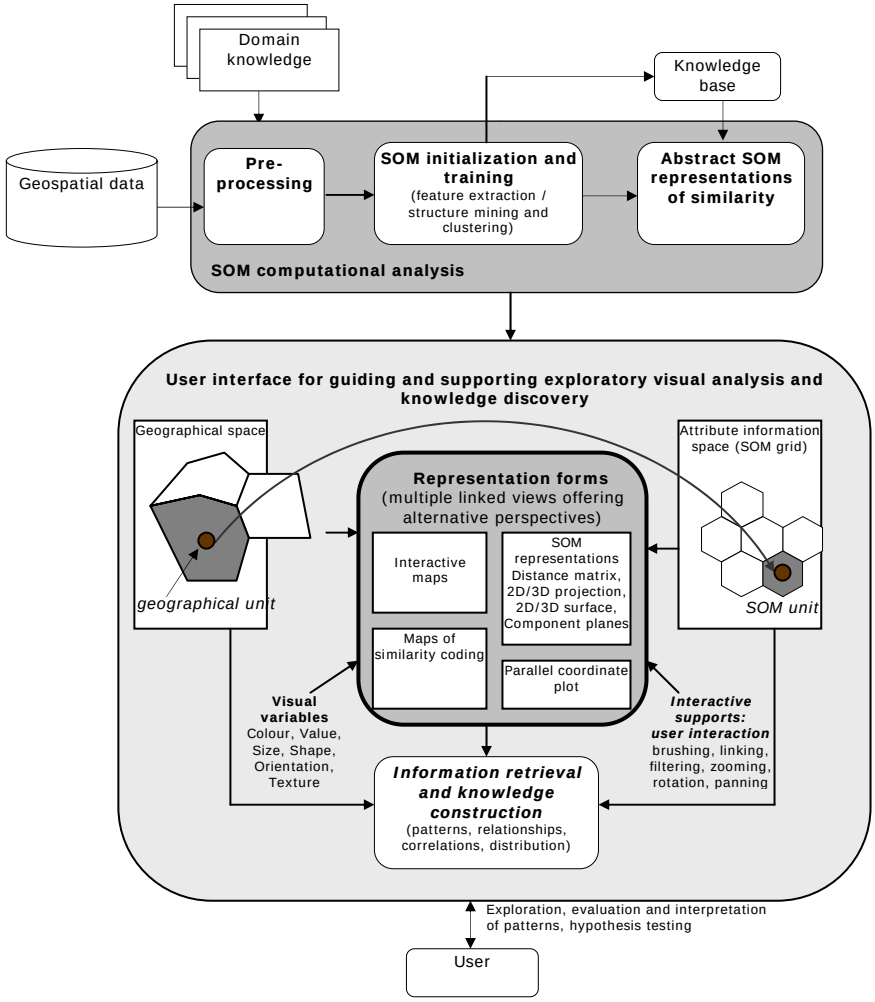


Figure 4.4. Structure of the geovisualization system.

The SOM network was trained using the SOM toolbox. In the SOM toolbox, the dataset is first put in a Matlab "struct", a data structure that contains all information related to the dataset in different fields for the numerical data (a matrix in which each row is a data sample and each column a component), strings, as well as other related information. Since the SOM algorithm uses Euclidean metric distance to measure distances between vectors, scaling of variables is needed to give equal importance to the variables. Linear scaling of all variables is used so that the variance of each is equal to 1. Other normalization methods such as logarithmic scaling and histogram equalization are offered. The original scale values can easily be returned when needed. Missing data are also handled in the SOM toolbox. The input vectors x are compared with the reference vectors m_i, using those components that are available in x.

4.4.2. User interface and interaction design

Some of the efforts to develop and apply usability engineering methods for the design and evaluation of computer interfaces have been directed towards interfaces for geospatial information representation (Cartwright et al. 2001).

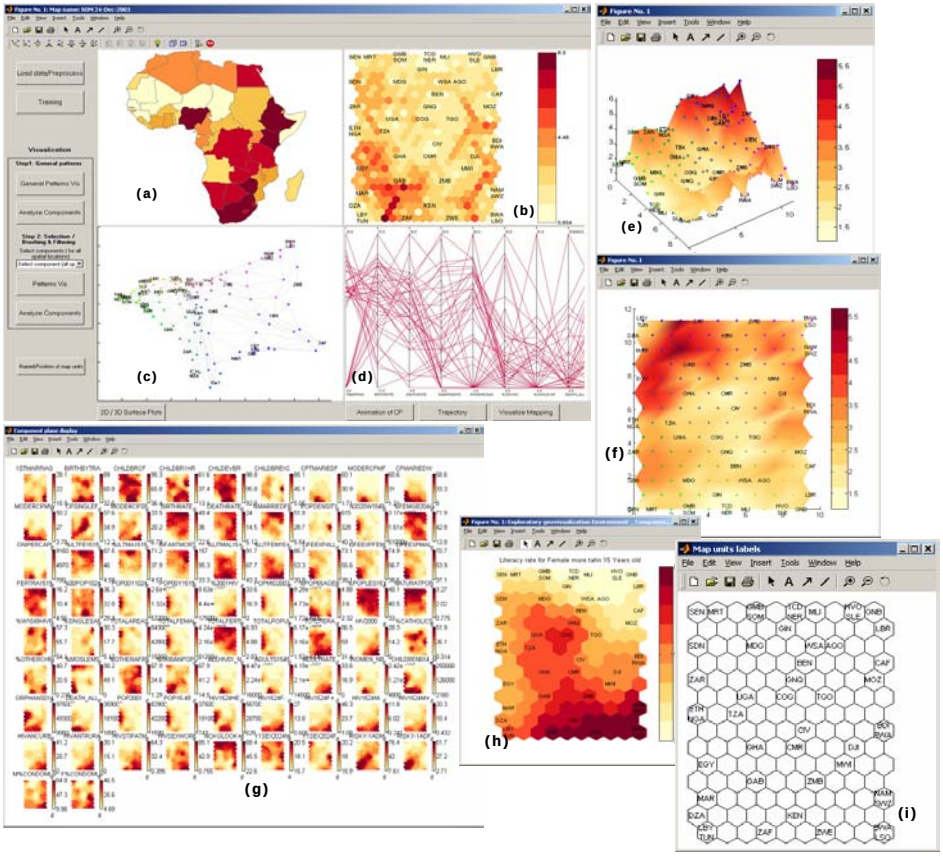


Figure 4.5. The user interface for the exploratory geovisualization environment in multiple views. The main view shows the representation of the general patterns and clustering in the input data: a map of the similarity coding (a), the unified distance matrix shows clustering and distances between positions on the map (b), the projection of the SOM results in 3D space (c), and parallel coordinate plot (d). The other windows show the alternative representations of the SOM general clustering of the data as options of the interface: 3D surface plot (e), 2D surface plot (f), and the visualization of component planes for the exploration of relationships and correlations among the attributes (g). An example of individual component is shown in (h) and the map unit labels are shown in (g).

The proposed geovisualization environment offers a similarity-based exploration that allows a distance matrix representation to be visualized in the SOM space (abstract information space). The similarity coding extracted from the SOM is used in a number of representations, including projection, surface plots,

component plane displays, and a geographical view (maps). A link between the different views is provided for the exploration of relationships (see figure 4.5).

The interface design focuses on three important aspects:
- Representation forms (map, grid, surface, projection).
- Visualization techniques (distance matrix, component planes display, 2D/3D views of surface plots and projections).
- Interaction techniques (brushing, panning, rotation, zooming).

The interface integrates the different representations into multiple views, which are used to simultaneously present interactions between several variables over the space of the SOM, maps and parallel coordinate plots, and to emphasize visual change detection and the monitoring of the variability through the attribute space. These alternative and different views of the data can help stimulate the visual thinking process that is characteristic of visual exploration. These alternative views are supported by user interaction for exploratory tasks to facilitate the knowledge discovery process. Users can perform a number of exploratory tasks to understand the structure of the dataset as a whole and to explore detailed information such as correlations and the relationships for selected attributes of the dataset. This is intended to guide them in hypothesis testing, evaluation, and the interpretation of pattern, from general patterns extracted to specific selection of attributes and spatial locations. Other supportive views are provided for further exploration of the displays, including zooming, panning, rotation and 3D views. In this integrated visual-computational environment, effective exploration of the data can be performed to support knowledge construction through user interactions.

Because the user develops a mental model of the system, it is important that the design helps construct a clear image of the system. For perceptual effectiveness (Eick 1997), the interface attempts to provide displays in a way that seems natural for interpretation: in a grid, on a map, on a surface, in a 3D space, with position showing internal relationships. Users have the possibility of visually relating information or aggregations of data to reveal the clustering structure or common visual properties. From the HCI perspective, a number of interaction strategies can help achieve the goals of visual exploration. The interface offers interactive filters for changing the relative positions of elements of the display, changing by rotation the perspective from which it is seen, and displaying detailed information to have access to actual data values on a specific data item of interest. Such transformations of views can interactively modify and augment visual structures, and support the likelihood of emergence (Peuquet and Kraak 2002). We use different interaction techniques to enhance data exploration, including brushing and linking, panning, zooming, and rotation.

Representational variables

We identified key elements of representation that may be necessary for the design of the graphical representations (see table 4.1).

Table 4.1. Representation variables.

Visuals	*Representation*
Representation form (metaphors)	Volume Cells Landscape Network or mesh Surface
Visual properties	Colour Objects identifiers (icons, markers,. etc.) Shape Lighting Position Surface reflectance Transparency Scale Legend Colour coding
Spatial arrangement	Geometry Topology Relationships Clustering structures Spatial distribution Object characteristics Similarity and dispersion Groups Organization (hierarchy, semantic generalization) Spatial relationships (taxonomic, thematic membership) Interpretation of geographical primitives (features, regions, boundaries) Formalized space structure (space type, scale)

4.5. Example exploration of geographical patterns in health statistics using the graphical user interface

The prototype is used in an example for exploring a large dataset containing heath statistics on Africa. In this section, the dataset is explored, and different visualization techniques are used to illustrate the exploration of (potential) patterns within the different options of the interface. This example is used to examine the integration of the different graphical representations in the user interface.

4.5.1. The dataset explored

The dataset consists of 74 variables, including health, demography and other socio-economic indicators, for 50 African countries. The idea is to find multivariate patterns and relationships among different attributes and countries. Maps of a few attributes of the dataset are provided in figure 4.6.

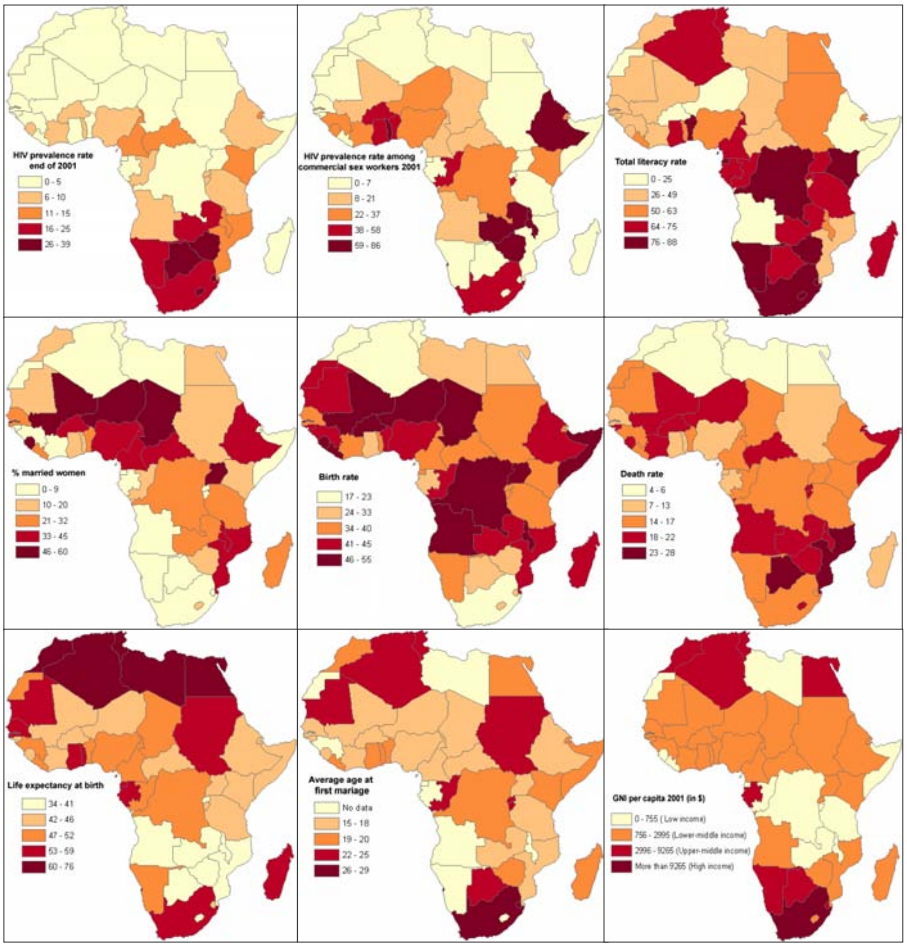


Figure 4.6. Example of attributes of the test dataset: HIV prevalence rate end of 2001, HIV rate among commercial sex workers, total literacy rate, percentage of married women, birth rate, total death rate, life expectancy at birth, average age at first marriage, and GNI per capita 2001.

4.5.2. Visual exploration support for general patterns and clustering

The default view in the interface offers after the data has been loaded and the SOM network trained is the general clustering structure of the data in different perspectives (maps, projections, unified distance matrix and parallel coordinate plot). This general view implements a number of distance matrix visualizations to explore the SOM results and show the cluster structure and similarity (patterns). An example is the unified distance matrix representation discussed in Chapter 3. In figure 4.7, the different views of the general structure of the dataset provided in the user interface are presented. In the distance matrix (figure 4.7a), countries having similar characteristics based on the multivariate attributes are positioned

close to each other and the distance between them represents the degree of similarity or dissimilarity. These common characteristics representations can be regarded as the health standard for the countries. In figure 4.7c, the projection of the SOM offers a view of the clustering of the data with data items depicted as coloured nodes (as described in Chapter 3). The clustering structure can also be viewed in the interface, as 2D or 3D surfaces representing the distance matrix (figure 4.7d), using colour value to indicate the average distance to neighbouring map units. This is a spatialization (Fabrikant and Skupin 2003) that uses a landscape metaphor to represent the density, shape, and size or volume of clusters. The landscape metaphor is a geographical analogy commonly used to facilitate the representation of information by creating an information landscape that can be easily assimilated by the viewer based on his or her experience of the real world. Such a geographical metaphor is found in most information visualization systems and has even become a design model for virtual environments (Chen 1999). The intention behind using a spatial metaphor is to create a graphical representation that is accessible to human cognition (Skupin and Buttenfield 1997) and to allow the viewer's intrinsic comfort with everyday concepts of human spatial orientation and way finding to guide their exploration and interpretation of the representation (Fabrikant 2001b).

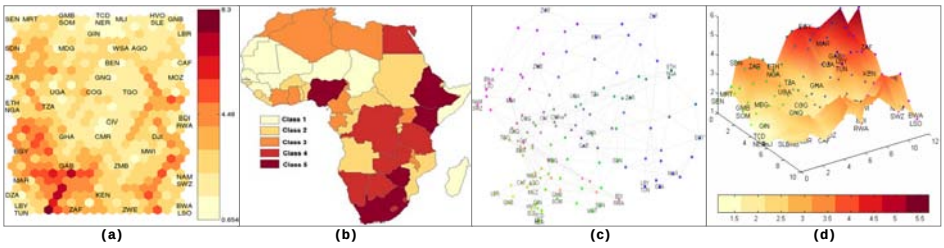


Figure 4.7. Representation of the general patterns and clustering in the input data from the interface: the unified distance matrix showing clustering and distances between positions on the map (a). Other representations of the SOM general clustering of the data are offered: a map of the similarity coding extracted from the SOM computational analysis (b), a projection of the SOM results in 3D space (c), and a 3D surface plot (d).

4.5.3. Exploration of correlations and relationships

As a second stage of the visualization process, the interface offers options to explore correlations and relationships in the input data. This is implemented by the component plane display (figure 4.8). As discussed in Chapter 3, here the component planes show the values of different attributes for the different countries. They are used to support exploratory tasks, facilitate the knowledge discovery process, and improve geospatial analysis.

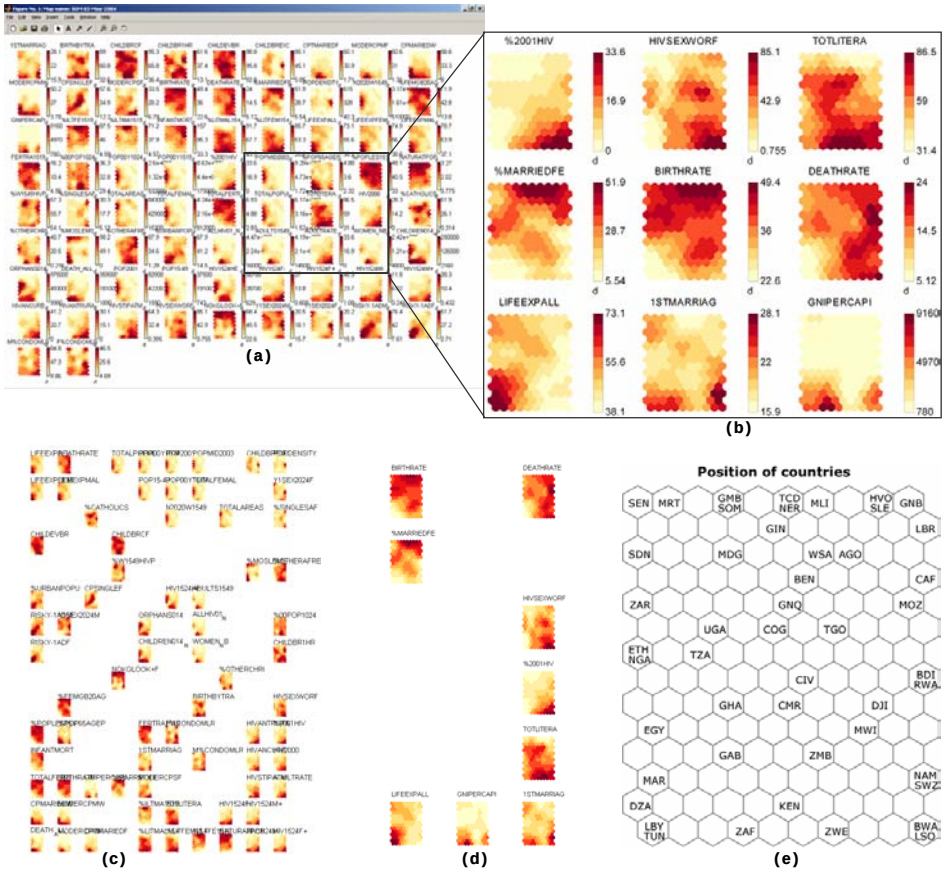


Figure 4.8. Detailed exploration of the dataset using the SOM component visualization: all the components can be displayed to reveal the relationships between the variables and the spatial locations (countries) (a). Selected components related to a specific hypothesis can be further explored (b). All the component planes can be ordered based on correlations among the variables (c). Selected components for a particular investigation (here the relationships between the HIV prevalence rate and socio-demographic variables) can be ordered to facilitate visual recognition of relationships among selected variables (d). Position of the countries on the SOM map (e).

Compared with the maps in figure 4.6, patterns and relationships among all the attributes can be easily examined in a single visual representation, using the SOM component plane visualization. Since the SOM represents the similarity clustering among the multivariate attributes, the visual representation becomes more accessible and easy to explore for exploratory analysis and knowledge discovery. The component planes can be displayed for selected attributes of the dataset. When overlaid with environmental, social, transportation and facilities data, this kind of spatial clustering makes it possible to conduct exploratory analyses to help identify the causes and correlates of health problems (Cromley and McLafferty

2002). These map overlays have been important hypothesis-generating tools in public health research and policy making (Croner et al. 1992). In figure 4.8a, all the components are displayed and a selected few are made more visible for the analysis in figure 4.8b. The kind of visual representation (imagery cues) provided in the SOM component plane visualization can facilitate visual detection, and has an impact on knowledge construction (Keller and Keller 1992). As such, the SOM can be used as an effective tool to visually detect correlations among operating variables in a large volume of multivariate data. From the exploration of global patterns, correlations and relationships in figure 4.8a, hypotheses can be made and further investigation can follow in the process of understanding the patterns. To enhance visual detection of the relationships and correlations, the components can be ordered so that variables that are correlated are displayed next to each other (see figure 4.8c and 4.8d) in a way similar to the collection maps of Bertin (1981). It becomes easy to see, for example, that the HIV prevalence rate in Africa is related to a number of other variables, including literacy rate and behaviour (characterized in the dataset as high-risk sexual behaviour and limited knowledge of risk factors), and other factors such as the high prevalence rate among prostitutes, and the high rate of infection for other sexually transmitted diseases. This exploration of the attributes of the dataset allows distinguishing clearly that as a consequence of the high prevalence rate in regions such as Southern Africa, there seem to be a low birth rate and life expectancy at birth, and a high death rate, highly impacted by the HIV infection. The birth rate in the most infected regions seems to be a consequence of the prevention measures. The increased use of condoms among a large proportion of single females, although for contraception purposes, seems to safeguard them against HIV. It is also observed through the component plane visualization that factors such as the percentage of married women, the percentage of sexually active single females, and the average age at first marriage in these countries are highly related to the prevalence rate. No significant differences are found between the prevalence rate in rural and urban areas. This may be due to the fact that over the last decades the infection, originally not known in rural areas, has gained ground in all parts of the countries. Actually the spread of HIV/AIDS follows a mixed pattern of diffusion through space and time (Gould 1995). No relationships are found between the poverty of the countries and the prevalence rate.

4.6. Conclusion

In this chapter we have presented the implementation of the proposed approach to integrate computational and visual analysis into the design of a prototype visualization environment. A user interface was developed to integrate the different graphical representations and support the exploration process by supporting a number of user activities. The interface is structured to provide global view and summary of the data as well as tools for detail exploration of relationships and correlations for exploratory analysis purposes. Interaction was needed to enhance user goal-specific querying and selection from the general

patterns extracted to more specific user selection of attributes and spatial locations for exploration, hypothesis generation, and knowledge construction. Interactive manipulation (zooming, rotation, panning, filtering and brushing) of the graphical representations was used provided to enhance user interaction, the objective being to explore ways of supporting visual exploration and knowledge construction. The link between the attribute space visualization based on the SOM, the geographical space with maps representing the SOM results, and other graphics such as parallel coordinate plots in multiple views offers alternative perspectives for better exploration, evaluation and interpretation of patterns, which ultimately supports knowledge construction.

Chapter 5

Exploring spatio-temporal patterns using self-organizing maps and cartographic animation

5.1. Introduction

Advances in data acquisition (e.g. satellite imagery, GPS) are creating new applications as well as challenges in information extraction from the large amounts of time-series data captured. The study of geographical processes has an important component about time since events or processes happen in time.

Geovisualization has been particularly addressing issues related to the analysis and exploration of patterns in spatio-temporal datasets. Visualizing the time dimension in geospatial data has long been in the centre of research in cartography. Maps are often used to depict events in snapshot views. However using maps to represent events that happen over time result in complex designs, with too many maps. Usually maps will only focus on part of such process. Animation is often used to integrate many maps and to add dynamics and interaction to the representation of time. Animation can be very useful to clarify trends and processes, as well as to provide insight into spatial relations (Kraak 2000b).

Some authors have proposed spatio-temporal modelling (Wachowicz 2000; Roddick and Lees 2001) for understanding space-time dynamics, based on modelling abstractions and concepts such as states, events and episodes as used in a specific knowledge domain. In this case, understanding the spatial, temporal and thematic aspects of the knowledge domain is crucial for the representation of variations and structure in the spatio-temporal phenomena. Much effort is needed on the representational issues of space-time dynamics. Some recent work has been based on the space-time cube concept (Hägerstrand 1970; 1982) for representing geospatial processes (Andrienko et al. 2003; Kraak 2003). However, pattern extraction issues remain limited and more exploration of techniques is needed to support visual exploration and understanding of space-time dynamics related to complex geographical phenomena.

This chapter is based on:

Koua E. L. and Kraak M. J. (2004). Alternative visualization of large geospatial data. Cartographic Journal vol 41 (3).

Koua E. L. and Kraak M. J. (In review). Exploring spatio-temporal patterns in large geospatial data using self-organizing maps and cartographic animation. Cartography and Geographic Information Science.

In this chapter, we use the SOM to process and extract patterns, relationships and trends from a large spatio-temporal dataset related to the production of food in Africa over the last 40 years. The issue of food production in relation to continued population growth, rapid urbanization, loss of forestland, land productivity, national unrest and resultant high refugee populations, droughts and floods, and more recently the HIV/AIDS pandemic has emerged as a critical development challenge (Turner and Schwarz 1980; Turner et al. 1993). Some of these conditions have resulted in recurrent food shortages in parts of Africa. The objective of exploring this dataset is to provide some understanding of the interrelationships between some of factors mentioned above, specifically changes in socio-economic factors such as population and their relation to food shortages and famine situations in parts of the continent. For representation of the space-time variations, we investigate ways of visualizing underlying dynamics in the dataset, using multiple views that simultaneously present interactions between several variables over the attribute space and time for different locations. The use of these techniques intends to allow visual change detection, support the exploration of time-related geographical trends and patterns, improve data analysis, and ultimately provide better understanding of the interrelationships between the different factors in space and time. The SOM-based representations are combined into animations using cartographic design.

5.2. Space-time representation

The proposed approach integrating computational and visual analysis for exploratory geovisualization (described in the previous chapters) is applied here for extracting and visualizing spatio-temporal patterns. Spatial analysis, data mining and knowledge discovery methods are combined for the extraction of patterns, as well as to enhance visual change detection and hypothesis generation, and therefore contribute to the understanding of spatio-temporal geographical processes. A number of representation techniques are explored. In the next subsection, we examine the different spatio-temporal representation techniques based on the SOM, including component plane displays in multiple views and the visualization of trajectories. The SOM-based techniques are then combined with cartographic animation to further explore interactions between several attributes, space and time, and to emphasize visual change detection and the monitoring of the variability through the attribute space. The goal is to provide alternative and different views on the data that can help stimulate the visual thinking process.

Spatio-temporal representations are an important aspect of research in geographical information science. Traditionally, maps have been used to represent spatio-temporal dynamics. A large part of the research effort in this area has been directed at data models and focused on two main representation models: models based on space representation and models focusing on time

representation of data. The traditional approach in GIS focuses on the spatial representation of entities based on the geometric and thematic properties. In such models, the main concept is the absolute view of space, and time is implicitly represented by changes that occur over the space. Time-based models focus on time representation as a fourth dimension or a parameter in the data in which events that occur can be located. This time structure of the representation has often been organized according to intervals between events, points of occurrence of events, or both intervals and key points (Wachowicz 2000). In temporal GIS research, data models for the representation of space-time have been proposed. A representative of this approach is Peuquet's work (1994), which proposed TEMPEST (temporal geographical information system) to integrate space and time data models in GIS. In this approach, the primary organization is based on time, representing processes by time line to show changes that occur. This approach suggests the key notions of location, time and object (where, when and what), the basic characteristics of geospatial data. On the representational level, two main models exist: models based on space representation and models focusing on time representation of data. The greater promise of spatio-temporal GIS resides ultimately in the capacity to examine causal relationships and their effects for exploration, explanation, prediction and planning (Peuquet 1994). Animation is particularly used for representing spatio-temporal patterns, to reveal the dynamics, represent processes, track changes, and attract the attention of the user (Edsall et al. 1997; Andrienko et al. 2000; Blok 2001; Harrover 2002). In general, computer-based animation (Dorling and Openshaw 1992) and visualization techniques rely on three strategies for depicting change: sequence of discrete displays or snapshots at various times, dynamically and interactively modifying display elements as time goes along, and depicting change in specific locations or over the entire region. This was explored in the triad framework (Peuquet 1994), in which information is stored relating to where (location-based view), what (object-based view) and when (time-based view).

How to design effective spatio-temporal representations will relate to the ability to handle locations, times, objects and events as primary entities, to assign attributes to any of these, and to keep track of the interdependencies among the various attributes (Galton 2001). The main views can be summarized in two ways:

- Comparing a sequence of temporal overlays to determine how factors at a set of locations change over time
- Comparing layers representing the spatial distribution of a single variable at different times instead of different variables at a single time.

In practice, the choice between the two types of temporal representation will depend on the type of data. In large spatio-temporal data, the effective representation of underlying processes will also rely on the ability to extract patterns, relationships and trends in the data.

In the next section, an application of the SOM for a large spatio-temporal dataset is explored, and the visualization techniques are used to illustrate the exploration of (potential) patterns.

5.3. Exploration of space-time patterns in a dataset on food production in Africa

An application of the method described in the previous section is applied to the annual food and agriculture statistics in Africa collected from 1961 to 2002. This dataset is provided by FAO for all African countries and consists of the production in metric tons of three main cereals (rice, maize and millet) for the last 42 years. It also includes socio-economic indicators such as the populations (total population, male and female populations, rural population, urban population, agricultural population, non-agricultural population) of 48 African countries. Data on rainfall and the vegetation index (NDVI) for the same period of time are also used for the analysis. Finding patterns and understanding the variations in the food production in such a large dataset can be very complex. For example, understanding how aspects of population growth, climate or other factors have impacted the production of cereals, or how these factors relate to famine situations in parts of the region, requires a clear depiction of the processes. The exploration of this dataset using the proposed approach and techniques intends to allow the analyst to formulate hypotheses in the process of understanding the geographical patterns of shortages, famine and socio-economic changes. Further exploration and explanation of the results might need domain expertise in an agriculture-related discipline. Additionally, factors related to governmental policies and political instability, and other environmental factors such as droughts, should be closely analyzed in relation to the patterns found in the exploration of the dataset. Some maps of the African rice, maize and millet production for selected years are presented in figure 5.1.

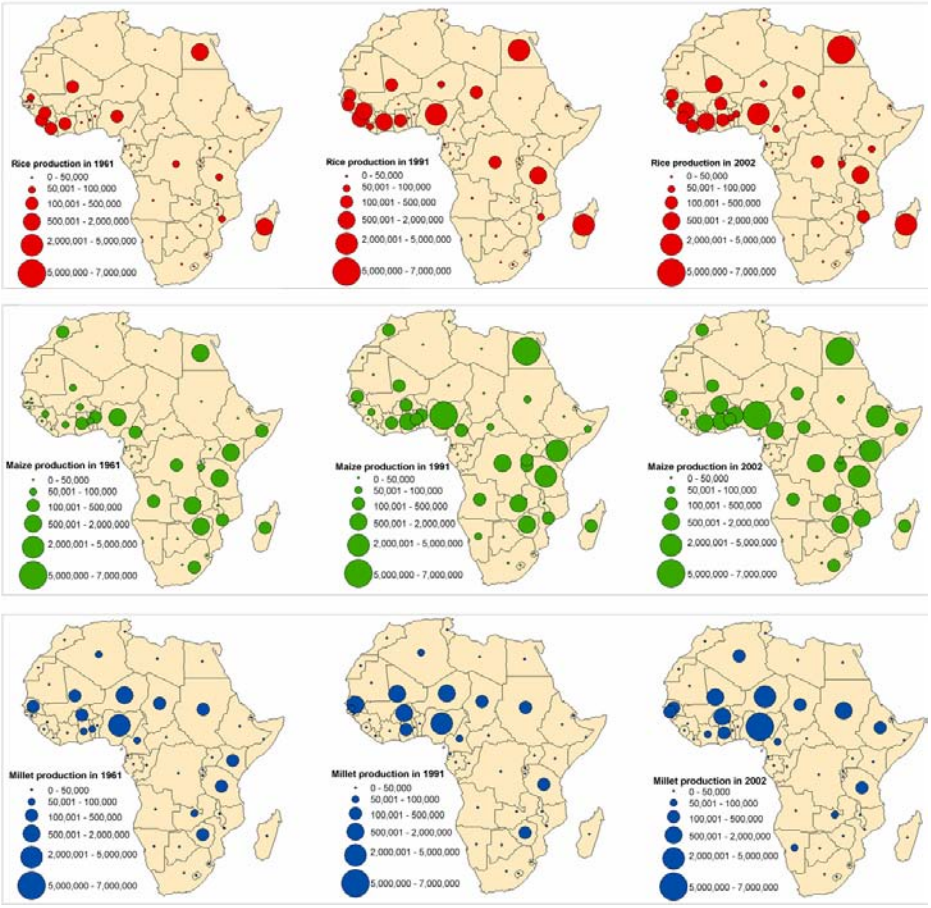


Figure 5.1. Some maps of the production of the three cereals (rice, maize and millet) for selected years (1961, 1991 and 2002). The entire period is 42 years (1961 to 2002).

5.3.1. SOM-based exploration and visualization of space-time patterns

Based on the SOM computational process, a number of visualization techniques can be explored. Non-linear dependencies between variables can be presented using three main categories of visualization and exploration technique:

- Visualization of the overall structure of the dataset introduced in Chapter 3. This refers to clustering, patterns (similarities) and irregularities (such as important gaps), and includes a similarity representation, projections in 2D or 3D space, and 2D and 3D surfaces.
- Exploration of correlations and relationships introduced in Chapter 3. This is primarily based on component plane displays in multiple views and allows the visualization of very detailed information that can support hypothesis generation.

- Visualization of temporal patterns. Examples are ordered component displays, trajectories, as well as the representation of time-related vectors in 2D or 3D projections. Examples of techniques for the representation of spatio-temporal dynamics are given in the next paragraphs.

Different graphical representations can support the exploration of space-time dynamics, offering the possibility of visually relating several variables simultaneously and thus helping in the knowledge discovery process. The SOM output provides ways of visualizing the general structure of the dataset (clustering), as well as exploring relationships among attributes. The different stages of the process of mapping the data on the SOM can be visualized as a trajectory on the SOM grid, with the display of component planes and projections, as well as animations. These can be used for each of the different representations mentioned, and combined with maps, which makes it possible to track the process dynamics and enable interpretation of the temporal relations among patterns at distinct levels (Guimaraes 2000).

5.3.2. Exploration of spatio-temporal patterns and relationships with component plane displays

The main representational technique for spatio-temporal patterns using the SOM is the visualization of component planes (see figure 5.2). The spatial and temporal attributes can be explored using the component plane visualization. The component plane display (figure 5.2) shows the values of the map elements for different attributes, and time. As with a collection of maps or processing maps (Bertin 1983) representing one attribute at the overall level, defining regions and geographical correlations, the component plane display answers the elementary question: At a given location, what is there in a given state? This results in the perception of similarity, and helps determine geographical correlations or define regions of a particular characteristic. The component planes are easy to read, provide an immediate answer to questions, and are useful for relationships involving the entire dataset. This is an exploratory process that does not need an initial hypothesis to represent patterns and facilitate visual comparison and perception, since visual recognition of the elements of the graphics and the relationships among them can be easily compared with colour over the attributes for different locations (see figure 5.2 for illustration of the reading of the component plane).

Since the component planes can be displayed and ordered in sequence to represent time-related attributes, they can be used to relate attributes to locations, times and events, which are the primary entities in spatio-temporal representation (Galton 2001). This can allow exploration of the interdependencies among the various attributes over time for different locations.

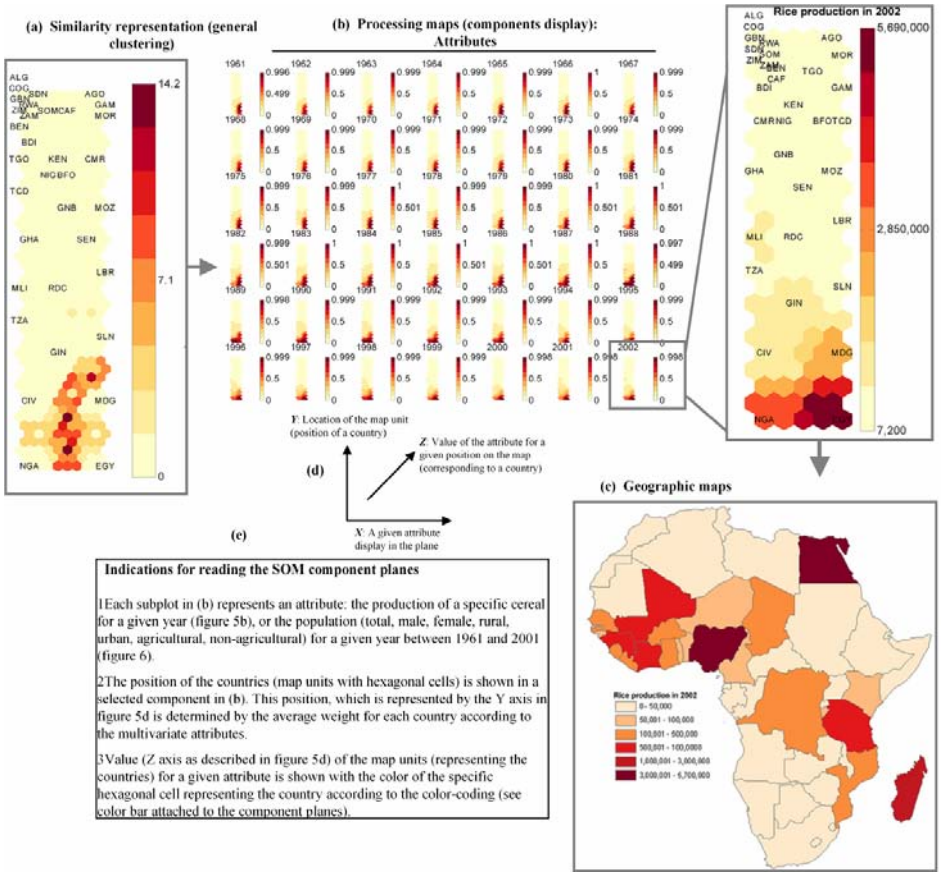


Figure 5.2. Component display and time. Detailed exploration of the dataset using the SOM component visualization: all the components can be displayed to reveal the relationships between the variables and the spatial locations (countries) (b). The variations in value (colour) indicate the relationship between the countries for the attribute represented in the plane. Selected components related to a specific hypothesis can be further explored to facilitate visual recognition of relationships among selected variables. Geographical maps of components corresponding to the hypothesis found in the exploration can be displayed (c). A description of the different axes used in the component plane display is described in (d), as well as indications for their reading in a text box (e).

In figure 5.2b, all the components are displayed and a selection of one sample attribute is made more visible for the analysis, with the name and position of the map units (countries). From the view in figure 5.2b, correlations and relationships

can be explored, and hypotheses can be made. Individual component planes are shown in subplots linked together through similar position. In each component plane, a particular map unit (hexagon) in the SOM is always in the same place and the value of one variable is shown using colour coding (see notes in figure 5.2e on how to read the component plane visualization). By using the position and colour (value), relationships between different map units can be easily explored. This can be used to visualize the variations among the attributes of the input data (figures 5.3, 5.4 and 5.5). Further analysis can be conducted by searching for correlations and interactions between different variables. This visualization reveals very detailed information. The actual values can be returned for every component (see the selected component display for agricultural population in 2001 in figure 5.5), which allows comparison between correlated attributes and places (countries). For example, if we consider attributes such as total population, rural population, urban population, agricultural population and non-agricultural population, we can easily view relationships between them (see figure 5.5). Figure 5.5 shows important changes in population patterns for Nigeria over the years. It shows the urbanization trend. Relatively more growth is observed in the urban population from the ‘80s (see green circles around Nigeria in figure 5.5). This can be partly due to a rural exodus, a persistent phenomenon in the ‘80s and ‘90s in most African countries. The agricultural population follows the reverse effect, dropping dramatically in 2001 (see figure 5.5). This may be one of the reasons for the decline in the production of rice in this country during recent years (see figure 5.5). Geographical maps can be made to represent the result of this reasoning process for better geographical exploration and comparison (figure 5.2c).

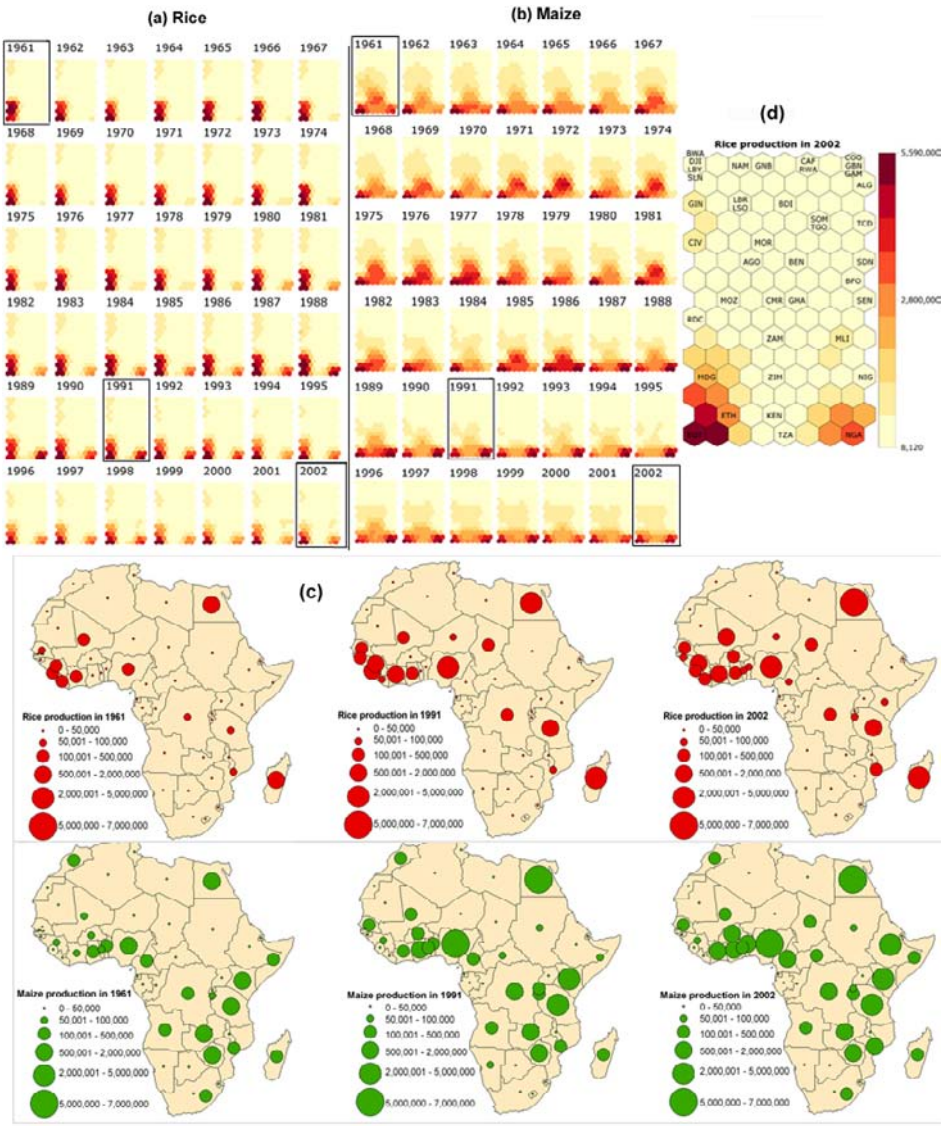


Figure 5.3. SOM component plane displays and maps for comparison. Changes are easier to detect in the component plane displays for the production of rice (a) and maize (b).

The SOM grid can be adjusted to a set of variables of interest. In figure 5.5, only the demographic variables were used to represent population patterns. The position of the map units is relative to these particular variables. Different spaces (SOM grids) can be used to explore sets of variables in each individual space. For example, the analysis of population patterns in figure 5.5 can easily be related to the production of the cereals in figure 5.4, by relating the value attached to the map units (countries) in each grid. For purposes of comparison between the different years, the normalized values of the vectors were used. The actual values of the vectors can be returned when needed for detailed analysis.

The idea is to compare a sequence of patterns for a set of locations over time in order to determine how factors affect changes. The different component displays representing the spatial distribution at different times can be compared in multiple views. For example, in figure 5.4 the urban population growth and agricultural population changes between 1961 and 2001 can be explored to observe population dynamics for Ethiopia. There has been more growth in Ethiopia's rural population over the years than in the urban population (see figure 5.5 and black circles for Ethiopia for the selected years displayed). The agricultural population for this country has experienced important growth, showing greater concentration of the agricultural population in rural areas. People in rural areas are generally poorer, and tend to be more strongly affected by crop failures and consequently more vulnerable to famine. This is a contrast to the frequent food shortages and famine in this country. This kind of exploratory analysis can be performed to include other factors that may play a role in the geographical process under study. Weather conditions, for example, are one of the factors to consider in the production of cereals. This situation in Ethiopia can be due to other environmental factors, such as droughts, rainfall and land degradation. In the next section we explore vegetation changes over the years relative to the production of cereals.

The component plane visualization is shown in figure 5.4 for the production of rice, maize and millet over the years. Information on the variations in the production of these cereals over the years can be revealed, as well as other interactions between the different attributes. For example, it is very easy to see that Nigeria started to increase its production of rice in the '80s and has suffered a significant decrease in production in recent years. Zimbabwe was one of the largest producers of maize in the '70s and '80s, but has seen a dramatic decrease in production over the last few years. In figure 5.3 the SOM component planes are displayed next to maps for selected years for comparison purposes.

This kind of visual exploration offered with the component planes can facilitate visual detection, and have an impact on knowledge construction (Keller and Keller 1992). Modelling and prediction can be made possible by using the SOM as a non-linear regression (Alhoniemi et al. 1999).

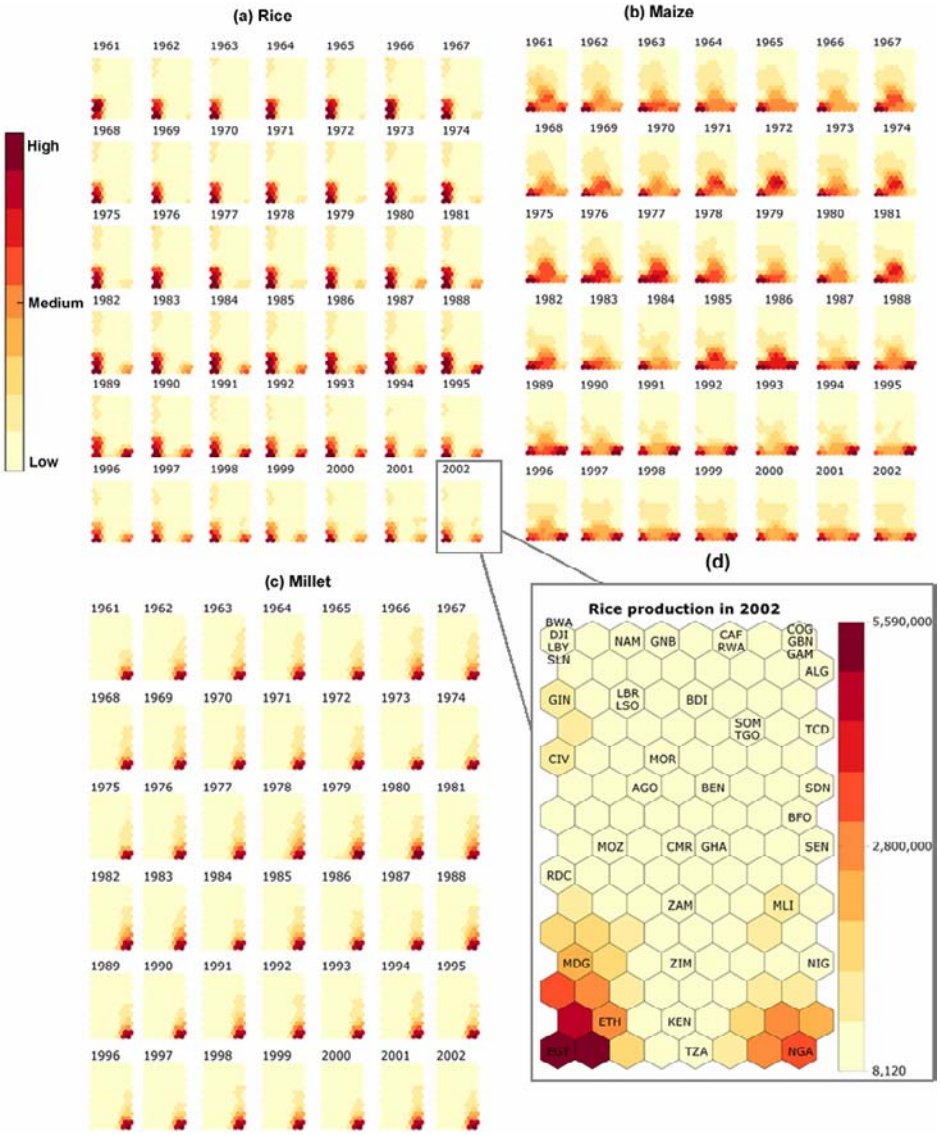


Figure 5.4. SOM component plane visualization for the production of rice (a), maize (b) and millet (c). The values of the production for the different years were normalized between 0 and 1. A detailed component showing the production of rice is provided in 2002 and a geographical index in (d).

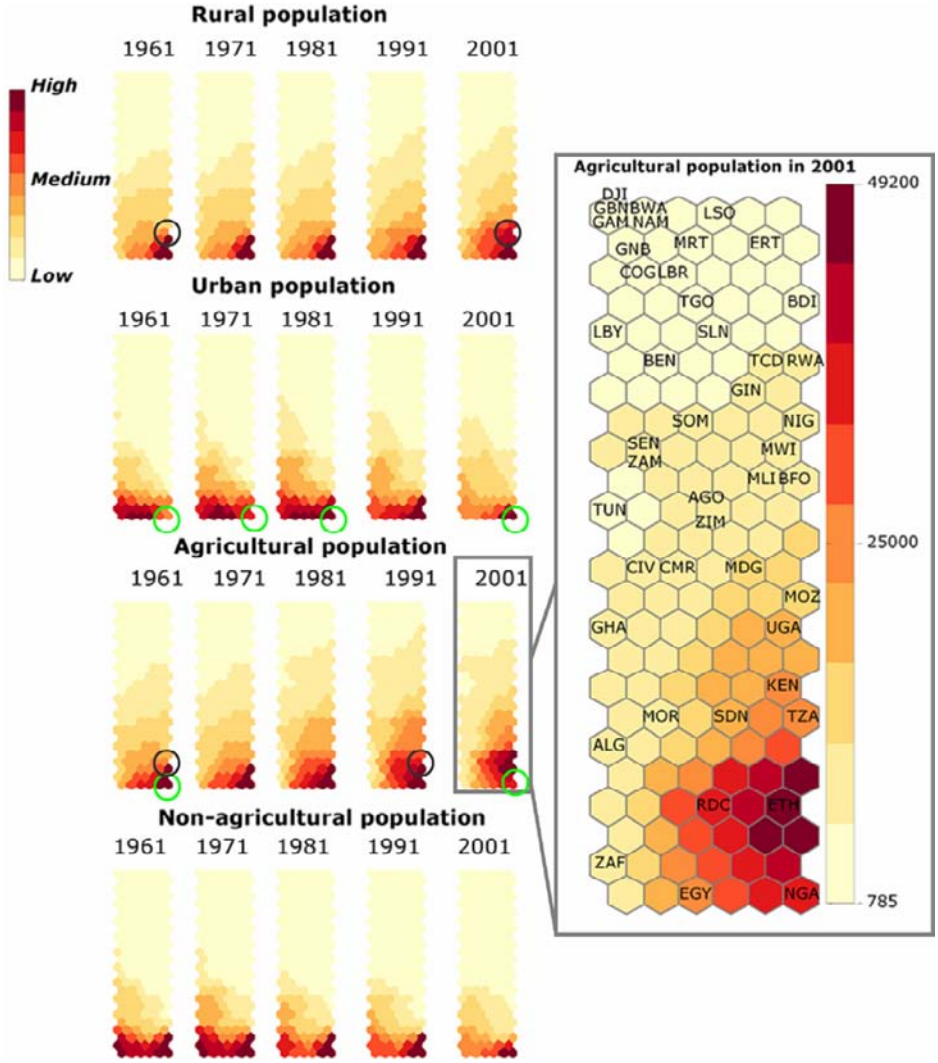


Figure 5.5. Population changes from 1961 to 2001. The figure shows a few selected years (1961, 1971, 1981, 1991, 2001) for rural, urban, agricultural and non-agricultural populations. The values were normalized between 0 and 1 for comparison. Information of the individual components can be retrieved with the legend corresponding to the actual value in the input data space (see example of display on right side for the agricultural population in 2001). The black circles show Ethiopia and the green circles Nigeria.

5.3.3. Visualization of trajectories

The different stages of the process of mapping the data on the SOM can be visualized as a trajectory on the SOM grid, which makes it possible to track the process dynamics and enable interpretation of the temporal relations among patterns at distinct levels (Guimaraes 2000). A display of the process as a trajectory linking the different moments in time can help visualize the process dynamics in the data. This visualization can be used to study the behaviour of a phenomenon over time. To illustrate the use of trajectories in analyzing the behaviour of a process, a time series extracted from the dataset explored in the experiment and related to the production of rice in Nigeria and maize in Zimbabwe over the last 40 years is considered. For clarity, a simple view of the data samples selected using scatter plots is provided (see figure 5.6a and 5.6c). A trajectory of these productions in the two countries is presented in figure 5.6b and 5.6d.

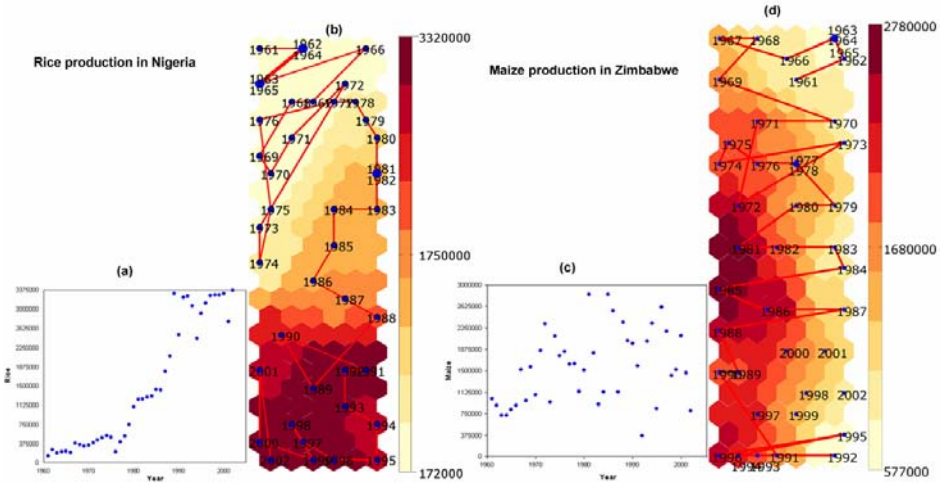


Figure 5.6. Scatter plots and trajectories of the selected data samples. Production of rice in Nigeria in scatter plot (a) and in trajectories (b). Production of maize in Zimbabwe in scatter plot (c) and in trajectories (d).

The visualization of the trajectory in figure 5.6b and 5.6d relates the different time states (years of production) on top of a SOM component display representing a clustering of the years of production. This reveals the different states of the production, and relates values and similarities between the different years of rice production in Nigeria (figure 5.6b) and maize production in Zimbabwe (figure 5.6d). In this SOM component plane, a clustering of the years of production is shown and one can easily see similarities and differences between the different years of production. This provides an easy way of visualizing the production process: gaps between years of production, changes in production levels, similarity among the years of production, and patterns of production for different countries. For example, the years where the production of rice in Nigeria was

highest can be easily seen (1989, 1991, 1992, 1993, 1996, 1997, 1998, 1999, 2000, 2002), and compared with other years in between when there was a drop (in 1990, 1994, 1995 and 2001). This can prompt the analyst to search for more patterns for these years in order to understand these changes. The production of maize in Zimbabwe has apparently had a few bad years every decade. This explains why the trajectory constantly shows a back-and-forth process from high values to low values. The path linking the years reveals the variability in the production over the years. Trajectories can be projected on top of all component planes to compare patterns in the productions for the different countries.

A number of spatio-temporal representation techniques have been based on such visualizations of paths or trajectories in recent years. Some recent work has been based on the space-time cube concept (Hägerstrand 1970; Hägerstrand 1982) for the representation of geospatial processes (Andrienko et al. 2003; Kraak 2003).

5.3.4. Projections

Projecting the SOM results offers a view of the clustering of the data and depicts the process dynamics at different times. The projection can reveal the relationships between different states. The general structure and tendency can be depicted in 2D or 3D space. The projection reveals not only the situation at each point in time, but also the similarity found between them. This gives an informative picture of the global shape and the overall smoothness of the SOM projection of the data. Exploration can be enhanced by interactive manipulation of the projection in 3D space, for example by rotating, zooming and panning. An example of a projection is shown in figure 5.7. This projection depicts the trends in the production of cereals (rice, maize and millet) in Nigeria over the years. Similarity between the years of production is shown, using colour and size of objects to further differentiate between clusters and cluster members that appear to have important gaps in the data value range. For example, the productions in the ‘60s and ‘70s are clustered together with the same colour and the same size of object. However, the productions in 1973, 1974 and 1975 have a slightly bigger size of object, showing an increase in production during these years.

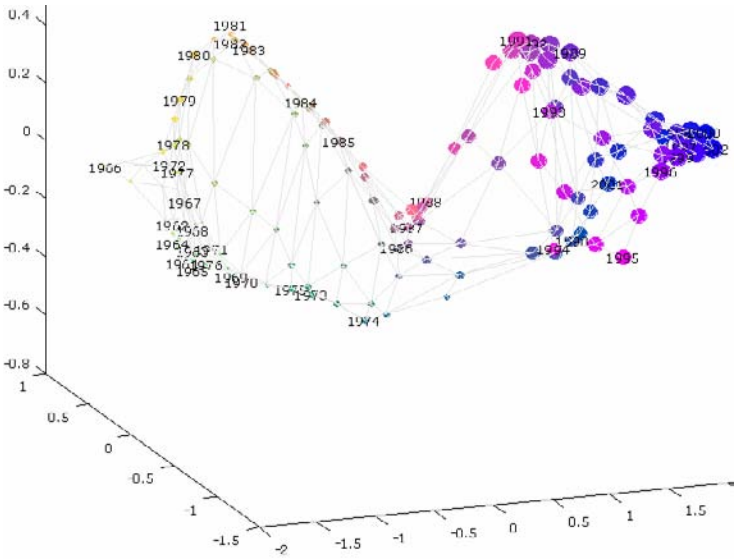
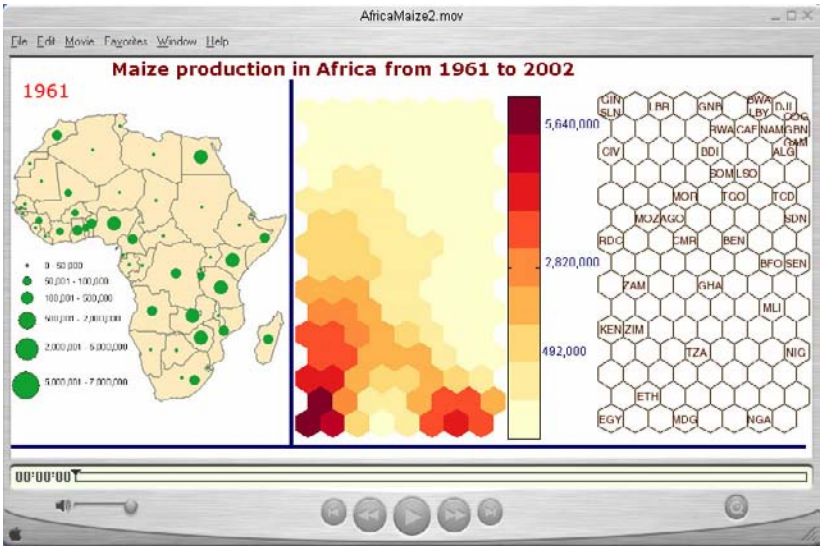


Figure 5.7. Example of projection for Nigeria's production of the three cereals together.

5.4. Integrating the SOM and cartographic animation for spatio-temporal patterns exploration

To enhance visual detection and exploration of temporal patterns, the results of the SOM computational process are integrated in a cartographic animation that provides different views, in maps and in the SOM space.

Animation as a visualization technique has been used in cartographic and geovisualization research (Monmonier 1990; Slocum and Egbert 1993; Edsall et al. 1997; Blok et al. 1999) to depict change in geographical processes and the representation of dynamic geographical phenomena. Animation can help support the understanding of geographical changes over space and time. Based on the SOM results, the animation was subsequently built using Macromedia Director 8 software, with Lingo scripting language (see figure 5.8 for the production of maize from 1961 to 2002) to relate the component plane display and the visualization of trajectories to maps.



(b)

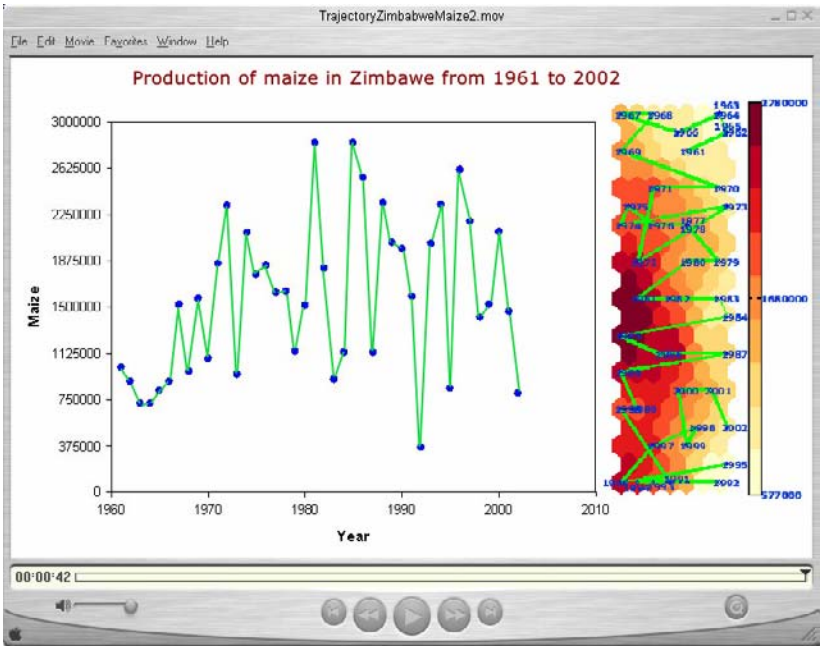


Figure 5.8. Animation of maps and component plane display (a). Scatter plots and trajectories of the selected data samples (b) for the production of maize in Zimbabwe.

The maps in figure 5.8a and scatter plot in figure 5.8b were used for comparing the animation with component planes and trajectory respectively. Changes in the production of the cereals are better revealed in the animation than in the

component planes and trajectory. Interactive features allow the user to have control over the animation (start, stop, and temporal navigation controls). The potential power of animation lies in its ability to represent change over time and thus facilitate an understanding of process rather than state (Harrover 2002).

5.5. Conclusion

In this chapter we have explored a number of computation and visual representation techniques for a large spatio-temporal dataset. The SOM algorithm was used for the extraction of patterns, relationships and trends and as the basis for visual representation of spatio-temporal processes. The techniques explored for spatio-temporal representation (component planes, trajectories and projection) provide a way of improving geospatial data analysis. Some techniques to specifically address temporal representations were explored.

We used ordered component planes to simultaneously present interactions between time vectors over the space of the SOM. This was to emphasize visual change detection and the monitoring of the variability through the attribute space. A visualization of trajectories was used to understand space-time dynamics in the data. One of the advantages of the SOM is that the algorithm is fast and effective for extracting patterns and relationships in very large datasets. Based on a similarity analysis, the algorithm was found to be effective in searching for correlations among operating variables. This can be achieved using the SOM component plane visualization, which allows the understanding of processes through visual representation and enables several variables and their interactions to be inspected simultaneously. Patterns, relationships, irregularities and distributions can be effectively visualized. This method provides opportunities to improve geographical analysis and to support exploration and knowledge discovery in the context of large geospatial datasets.

To enhance exploration and provide more flexibility and control for spatial analysis purposes, a user interface is being developed to integrate the SOM representations into a multiple-view environment linking other views, such as parallel coordinate plots, and the maps. Interaction is central to this design. A number of interaction techniques (rotation, panning, brushing, zooming and motion-related interactions) are provided in the graphics and allow different viewpoints.

Chapter 6

A usability evaluation methodology for assessing exploratory analysis and knowledge discovery tasks in geovisualization

6.1. Introduction

Usability of GIS products and specifically geovisualization tools has received considerable attention in recent years and is increasingly the focus of a number of research activities in the GIScience field. This is due partly to the fact that usability and human-computer interaction (HCI) in general are becoming important in new designs in geovisualization that integrate techniques and methods from other disciplines such as information science and computer science. There is, however, a lack of evaluation methodologies and particularly task specifications for user-based testing in exploratory geovisualization tools (Slocum et al. 2001). The need to assess the usefulness and usability of geovisualization tools is increasing as new types of interactions emerge (Muntz et al. 2003).

New developments in the design of geovisualization tools in recent years have amplified the need for usability evaluation of both these tools and the effectiveness of user interfaces. Increasing research interest in the usability of geoinformation systems has recently linked the HCI field, cognitive science and information science in a number of studies (MacEachren and Kraak 2001; Haklay and Tobon 2003; Fuhrmann et al. 2004; Koua and Kraak 2004a). There is, however, a lack of evaluation methodologies for formally assessing geovisualization tools. The map use studies (MacEachren 1995) usually conducted in the field of cartography are not necessarily fully applicable in new interactive visualizations that involve new representational spaces and advanced user interfaces.

Based on an approach to combine visual and computational methods for knowledge discovery in large geospatial data, a visualization environment has been developed. This environment integrates non-geographic information spaces with maps and other graphics that allow patterns and attribute relationships to be

---

This chapter is based on:

Koua E. L. and Kraak M.J. (2004). A Usability framework for the design and evaluation of an exploratory geovisualization environment. In: Proceedings 8th International Conference on Information Visualization. 14-16 July 2004 London. IEEE Computer Society Press, 2004. pp 153-158.

explored. The tool intends to facilitate knowledge construction, using a number of steps provided in data mining and knowledge discovery methodology. The development of the tool is based on the self-organizing map (SOM) neural network algorithm, and relates to data mining and knowledge discovery methods for the extraction of patterns. Some graphical representations are used to portray extracted patterns in a visual form in order to support the understanding of the structures and the geographic processes. In order to investigate the effectiveness of the design concept, an empirical usability test is planned to assess the tool's ability to meet user performance and satisfaction. In the test, different options of map-based and interactive visualizations of the SOM output are used to explore a socio-demographic dataset. The study emphasizes the knowledge discovery process based on exploratory tasks and visualization operations.

In this chapter, we propose the usability assessment methodology for evaluating a number of issues related to the user interface, as well as examining the support for exploratory tasks and knowledge discovery in the geovisualization environment. The methodology is based on an understanding of several knowledge discovery activities, visualization operations, and a number of steps in computational analysis used to visualize patterns in the data.

6.2. Exploration and knowledge discovery tasks in the visualization environment

One way to examine exploration and knowledge discovery support in the visualization environment is by assessing user performance for a number of defined tasks and steps. The model described in figure 6.1 emphasizes the exploratory nature of the visualization environment and the support for knowledge construction, from hypothesis formulation to the interpretation of results. This figure is an extension to figure 2.5 and focuses on the exploration steps undertaken by users in the knowledge discovery process. Some of these steps may be repeated. The next subsections provide an outline of these steps for this evaluation.

6.2.1. Defining user tasks for usability evaluation

The main goal of geospatial data analysis is to find patterns and relationships in the data that can help answer questions about a geographic phenomenon or process. The geographic analysis process can be viewed as a set of tasks and operations needed to meet the goals of the data exploration. The primary tasks in this process involve checking the spatial positioning of elements of interest in order to verify spatial proximity among different elements; verifying their spatial density; and obtaining an overview of how a target value measured at one particular spatial location, or at various neighbouring locations, varies for different

attributes. These tasks involve a number of activities and operations that users will perform:

- identification of the different clusters in the data, attributes, and relationships between elements (within clusters and between different clusters)
- comparison of values at different spatial locations, distinguishing the range of value and the order of importance of objects accordingly
- relation of value, position and shape of object type
- reflection on relevance of information extracted.

At interface level, some of the basic actions of users include selection, scaling, rotation, panning, brushing, browsing, filtering, and querying the database. For evaluation, the task of the test subjects can start with the visual exploration (Keim 2002) of the dataset, with the aim of finding patterns, correlations and relationships, by manipulating the representation forms provided, in order to gain insight into the data and draw conclusions.

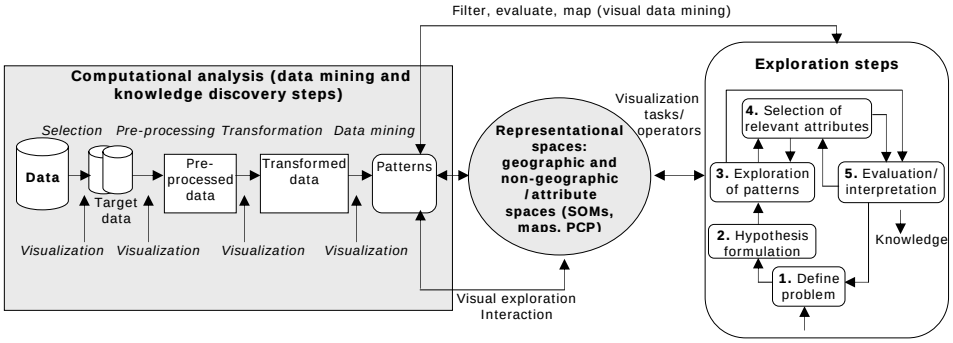


Figure 6.1. Data mining, exploratory visualization and knowledge discovery processes. The first part of this process consists of the general data mining and knowledge discovery steps (computational analysis). Each of the steps of the computational analysis can allow visualization. Patterns extracted as a result of the computational process can be explored using graphical representations (geographic and non-geographic information spaces). This exploration is guided by a number of steps to support knowledge construction.

One of the most important issues in the design of the visualization environment has been how to support the user in the different steps or in a way that best supports knowledge construction. The design provides an initial guide to take the user through the first critical steps of the exploration process. The first step (default view) of the exploration process provides the general patterns in the data in different representational spaces. The next steps of the discovery process include selecting a problem, formulating the problem, selecting data, and the relevant tasks (Michalski and Kaufman 1997). Our view of exploration is based on the fact that users primarily have access to the global views or results from the data mining process at the starting point of the visualization process. The tasks of exploration are then based on the selection of objectives, data, views and tools

for completing the required tasks for knowledge discovery. In general, the following goals will guide the exploration and knowledge discovery process in the evaluation:

- selecting data to explore
- specifying the problem or question to answer
- preparing data (selecting appropriate components of the data)
- using appropriate representations (how to find patterns)
- interpretation/evaluation (comparing views, validating patterns)
- application (how to use the knowledge).

Independent of the tool's support, users will have to go through a set of steps necessary to complete the tasks. The starting point of this process is the identification of a need or a problem. To support the exploration and knowledge discovery scenario described in figure 6.1, we provide an example of application with the exploration of a dataset to be used in the evaluation. It is a dataset on geography and economic development for 150 countries and contains 48 variables on socio-demography, economy, geography, health and other development-related indicators, explored in Chapter 2. The example problem definition and hypothesis for the exploration of this dataset can be summarized by the question: "Is economic growth related to the location of the countries?"

We propose three phases in the reasoning related to the exploration steps provided above and related to figure 6.1. This is an iterative process in which users can make new selections depending on the results of the exploration (step 3) or at the evaluation and validation step (step 4).

Phase 1: Global patterns exploration (with reference to the exploration tasks in figure 6.1)

1. Define problem (specify needs and questions to answer)
2. Formulate hypothesis
3. Explore general patterns from the data mining process (investigate clusters in different views and representational spaces as well as the relationships and correlations among attributes)
4. Evaluate and validate results or patterns (compare views).

Example of application for phase 1

From the general structure extracted in the data, the user is offered the possibility of investigating the patterns and relationships in the data (see figures 6.3 and 6.4). The user can view the clustering structure, the similarities based on the multivariate attributes (distance matrix, point projection, surface plots) and the relationships among the attributes by using the component plane visualization. Maps are also available to relate the position of the data items displayed on the grids to their geographic location.

Phase 2: Detailed exploration based on user selections (with reference to the exploration steps in figure 6.1)

In addition to the steps undertaken in phase 1 (steps 1, 2, 3 and 5) for the general patterns, users can initiate actions for more detailed exploration based on a selection of attributes and/or spatial locations:

4. Select the most relevant attributes for the hypothesis or problem, or find more attributes or data items satisfying given constraints/rules
3. Explore patterns for the selection by constructing new relevant derived visualizations of selected attributes (examine relationships and correlations among attributes)
5. Evaluate and validate results (patterns, compare views).

Example of application for phase 2

For the socio-demographic dataset explored, users can, for example, identify variables to be looked at. In this example, factors such as landlocked countries, access to ports, characteristics of countries that are in the tropics and those that are not, population density in coastal areas and population inland can be considered. From the general pattern structure, the user is offered the possibility of selecting the most relevant attributes for representation. A number of representational spaces are used for visual exploration of the data: SOM-based representations, maps, and other graphical representations such as parallel coordinate plots (PCP) and scatter plots. From these representations of selected attributes, correlations and relationships can be explored. Some answers may be found or other hypotheses generated. This iterative process can lead to the construction of necessary knowledge for the problem at hand.

Phase 3: Interpretation and evaluation of extracted knowledge (with reference to the exploration tasks in figure 6.1)

Verify relevance of classified data items. The user reflects on the results of the exploration process, makes interpretations, and evaluates the extracted knowledge for use.

The exploration steps described above with the example to illustrate the different phases of the knowledge discovery process are supported by basic visualization tasks and operators, as users manipulate the graphical representations and initiate actions at the different steps. These visualization tasks are the basis for the success of the exploration process.

6.2.2. Exploration tasks and visualization operators

To complete the tasks described above, the user will have to execute a number of visualization operations during the exploration process described in figure 6.1. Several authors have suggested taxonomies for visualization operations (Keller and Keller 1992; Qian et al. 1997; Zhou and Feiner 1998; Ogao and Kraak 2002). The most comprehensive list (Keller and Keller 1992; Wehrend and Lewis 2000) includes: identify, locate, distinguish, categorize, cluster, distribution, rank, compare, associate, and correlate.

- Identify: to establish the collective characteristics by which an object is distinctly recognizable
- Locate: to determine the absolute or relative position
- Distinguish: to recognize as different or distinct
- Categorize: to place in specifically defined divisions in a classification; this can be done by colour, position, type of object (shape)
- Cluster: to join into groups of the same, similar or related type
- Distribution: to describe the overall pattern. This is closely related to cluster in the same way that locate and identify are related. The cluster operation asks that the groups be detected, whereas the distribution operation requires a description of the overall clustering.
- Rank: to give an order or position with respect to other objects of like type
- Compare: to examine so as to notice similarities, differences, order
- Associate: to link or join in a relationship
- Correlate: to establish a direct connection (correlation).

A delineation of some of these operations for the visualization and analysis of spatial data was provided by (Qian et al. 1997), and includes selection, association, and grouping.

From these taxonomies of visualization goals described above, three key exploratory tasks for knowledge construction can be identified:

1. Categorize and classify: users must be aware of the different clusters that were found in the data. The different clusters can be viewed in different perspectives, 2D and 3D space, rotation, etc.
2. Compare: users can categorize and review relationships, and perceive commonalities and distinctions.
3. Reflect (evaluate, integrate, extend, generalize): after completing most activities, users can reflect on the patterns they observe. What general rules can be constructed?

For the dataset selected for the evaluation and described above, a first task is to analyze similarities and differences between the countries, their geography and their economic situations; and investigate correlations between economic growth and, for example, access to the sea, the relationships between geographic regions, whether located far from the coast, and population density, population growth, and economic growth. These factors are considered to be related to economic development (Gallup et al. 1999).

The model of the exploratory visualization and knowledge discovery provided in figure 6.1 is used to examine the exploration process. The steps provided in this model lead to the exploration of the global patterns (first level of clustering of all variables together) and to the examination of detailed information on individual attributes. The exploration tasks consist of finding correlations and relationships and exploring cause-and-effect scenarios in the dataset.

A set of representative tasks derived from the exploration scenarios described in figure 6.1 and key visualization operations described above are identified in visualization task scenarios for the evaluation study. This results from a decomposition of the basic visualization tasks and is presented in the next section. The rationale behind the use of evaluation scenarios is that they can represent how the system is intended to be used by end users. Task scenarios provide a task-oriented perspective on the interface and represent a structure and flow of goals and actions that participants are supposed to evaluate. Such scenarios ensure that certain interface features are evaluated.

6.3. Usability and human-computer interaction

Human-computer interaction (HCI) is a discipline concerned with improving the quality of interaction between users and the computer in information systems. In general HCI studies focus on how characteristics of machines and systems affect user performance (Shneiderman 1997). This involves the cognitive as well as the physical, and theoretical issues. HCI has a strong emphasis on user-centred design, a design approach that views knowledge about users and their involvement in the design as a central concern, and includes users in testing and evaluations in an interactive design process. The HCI approach is supported by cognitive theory to provide insights into how better to design interfaces that better correspond to the perception of users. In new designs of interfaces for geovisualization, this link between usability testing and user-centred design is becoming more prominent (Haklay and Tobon 2003; Fuhrmann et al. 2004; Koua and Kraak 2004e).

In the next subsections, we briefly describe some important issues of usability and HCI in general in the design of information systems, and outline some of the major approaches in usability evaluation.

6.3.1. Usability evaluation

Usability evaluation is central to HCI, to ensure that the design of a user interface meets the user requirements. It is a technique for ensuring that the intended users of a system can carry out the intended tasks efficiently, effectively and satisfactorily, and must be conducted at different stages in the design process. There are basically two types of usability evaluation technique: usability testing and usability inspection methods (Nielsen and Mack 1994). Usability testing involves assessing the tool's ability to meet user performance and satisfaction objectives, and is conducted based on a number of representative user tasks, for which a certain number of usability factors are measured. Usability inspection is a generic term for a range of usability engineering methods that look at problems in the design, and are generally used for evaluating early HCI design concepts.

Usability engineering is a larger usability context that helps identify usability problems either in testing or through inspections during the design process.

Both user testing and inspection methods can address different usability objectives.
Usability inspection methods include:

- Heuristic evaluation (usability specialists judge whether each dialogue element conforms to established usability principles or heuristics)
- Guideline reviews (inspections of the interface for conformance with a comprehensive list of usability guidelines)
- Pluralistic walkthroughs (users, developers and human factors experts discuss usability issues in a scenario associated with dialogue elements involved in the scenario steps)
- Consistency inspections (evaluate consistency across the family of products that has been evaluated by an inspection team)
- Standards inspection (experts on some interface standard inspect the interface for compliance)
- Cognitive walkthrough (use of a more explicit detailed procedure to simulate a user problem-solving process at each step in the human-computer dialogue)
- Formal usability inspections (participants have well-defined responsibilities in an inspection meeting where design problems can be identified)
- Feature inspections (involves the evaluation of functions delivered in the software tool).

In early design stages, inspection techniques such as cognitive walkthrough (Polson et al. 1992), and usability review or heuristic evaluation (Nielsen and Molich 1990; Nielsen 1994a) are often used for identifying usability issues, for validating design decisions, and for getting feedback on key aspects of the functionality, interface, and overall navigation. In a cognitive walkthrough, expert evaluators construct task scenarios from a specification or early prototype and then role-play as a user working with that interface. Heuristic evaluation involves a small set of expert evaluators examining the interface and judging its compliance with recognized usability principles (the "heuristics") so that they can be attended to as part of an iterative design process.

Choosing a usability evaluation method requires consideration of methodology issues and the objectives of the evaluation. For evaluating user interfaces, empirical methods are the main methods used and user testing is the most commonly used method (Nielsen and Mack 1994).

6.3.2. Approaches to usability evaluation

Based on the evaluation methods described above, usability evaluation can take several forms: user-based, expert-based or theory-based (Sweeney et al. 1993).

User-based evaluation (user testing) involves having selected users complete tasks with the system. Expert-based evaluation (inspection) involves having the evaluator(s) (e.g. human factors experts) use the system in a more or less structured way in order to determine whether the system matches the predefined design criteria. A scenario that characterizes a theory-based approach involves a designer or evaluator calculating the match between the task or user model and the system specification, using inspection methods.

Depending on the objectives of the assessment, a number of usability indicators can be examined. They generally include functionality, ease of use, learnability, effectiveness, efficiency and user satisfaction. These indicators are defined in slightly different ways in the literature. One of the most comprehensive taxonomies of usability indicators is provided by (Sweeney et al. 1993); it is described in table 6.1 below. In figure 6.2, we provide a description of the proposed usability evaluation framework. This framework shows the three stages in the development of the tool (design, development, evaluation). The proposed usability evaluation involves user-based testing on selected tasks to assess the tool's usability (effectiveness, efficiency, user satisfaction) and usefulness (compatibility, flexibility, appropriateness), as well as attitude and user reactions.

Table 6.1. Usability approaches, indicators, and measures (Sweeney et al. 1993).

Approach	***Usability indicators***	***Data gathered***	***Objective / subjective***
User-based evaluation	**Performance** (user): speed (time), accuracy scores (error)	Task time, % completed, error rate, duration of time in HELP, continuance of usage, range of function used	Objective
	Non-verbal behavior	Eye movement, orientation duration and frequency of documents access	Subjective
	Attitude (user's attitude and opinions, satisfaction and preferences)	Questionnaire and survey responses, comments from interviews and ratings, answers to comprehension questions	Subjective
	Cognition (user's understanding and knowledge of system)	Verbal protocols, post-hoc comments	Objective
	Stress	Galvanic skin response, heart rate Event-related brain potentials Electro-encephalograms Ratings or comments	Objective and subjective
	Motivation	Enthusiasm, willingness and effort	Subjective
Theory-based	**Performance** (idealized) (prediction of usage)	Predictions of: - Task performance times - Learning times - Likely ease of understanding	Objective
Expert-based	**Conformance** (level of conformance with standards, guidelines and design criteria)	Level of adherence or conformance with - Guidelines, principles and standards - Design criteria	Objective
	Attitude (expert) (professional opinion)	Comments Rating of usability properties	Subjective

From the taxonomy described in table 6.1, usability evaluation can have main orientations depending on the objectives. Usually a combination of techniques is used.

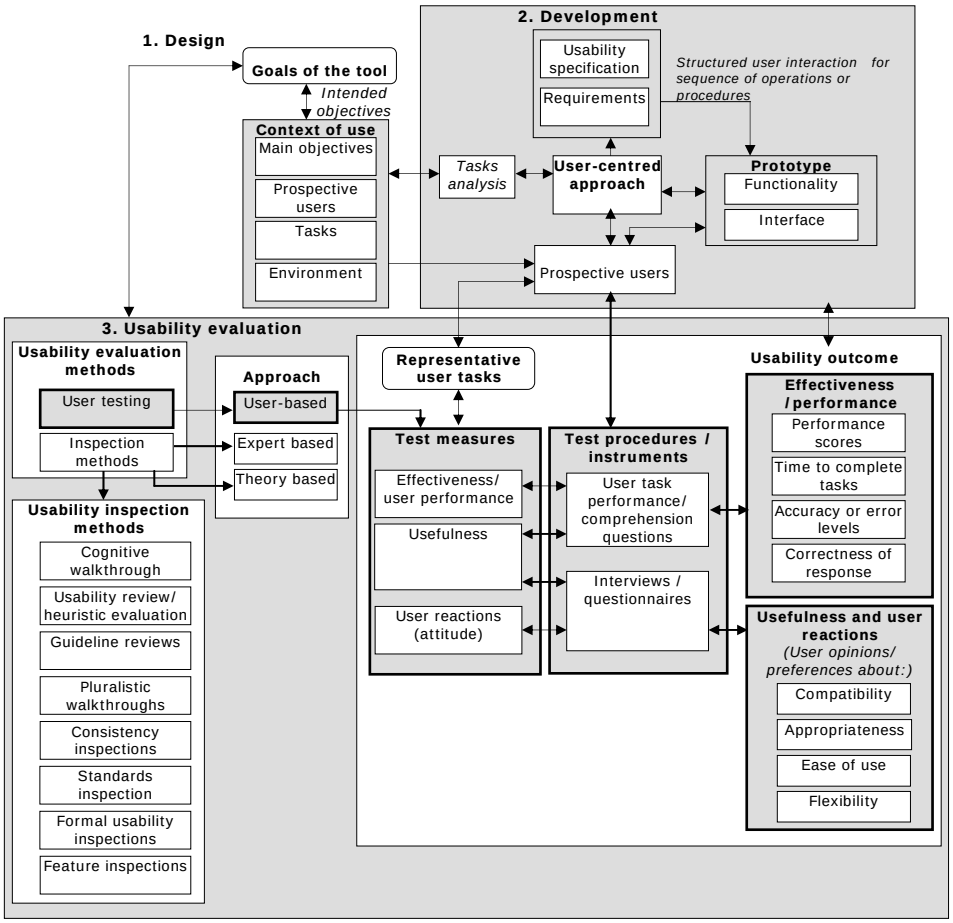


Figure 6.2. Usability evaluation framework: the framework shows the general usability and design framework, as well as the specific usability evaluation approach applied in the proposed methodology. The grey boxes indicate the applied method, and the white boxes the general usability framework. Selected user tasks are measured according to performance indicators, attitude and reactions. In this framework, the outcome of the usability evaluation is used to improve the design.

6.4. A user-based and task-based usability evaluation of exploratory geovisualization

Given the exploratory nature of geovisualization environments, particularly the visual-computational environment for which this assessment methodology is developed, a user-based evaluation (user testing) is certainly the most suitable approach to assess usability. Usability testing is more effective for evaluating the overall usability of the interface and can address a wider range of evaluation objectives than inspection methods (Nielsen and Mack 1994). Since the design of the tool is based on a user-centred approach, early involvement of users was employed in a preliminary interface feature inspection (Koua and Kraak 2004a), in which several aspects of the representation forms, graphics and colour schemes were presented to users for analysis (see Chapter 3).

Screen shots of the overall user interface and different representations are provided in figures 6.3 and 6.4 below.

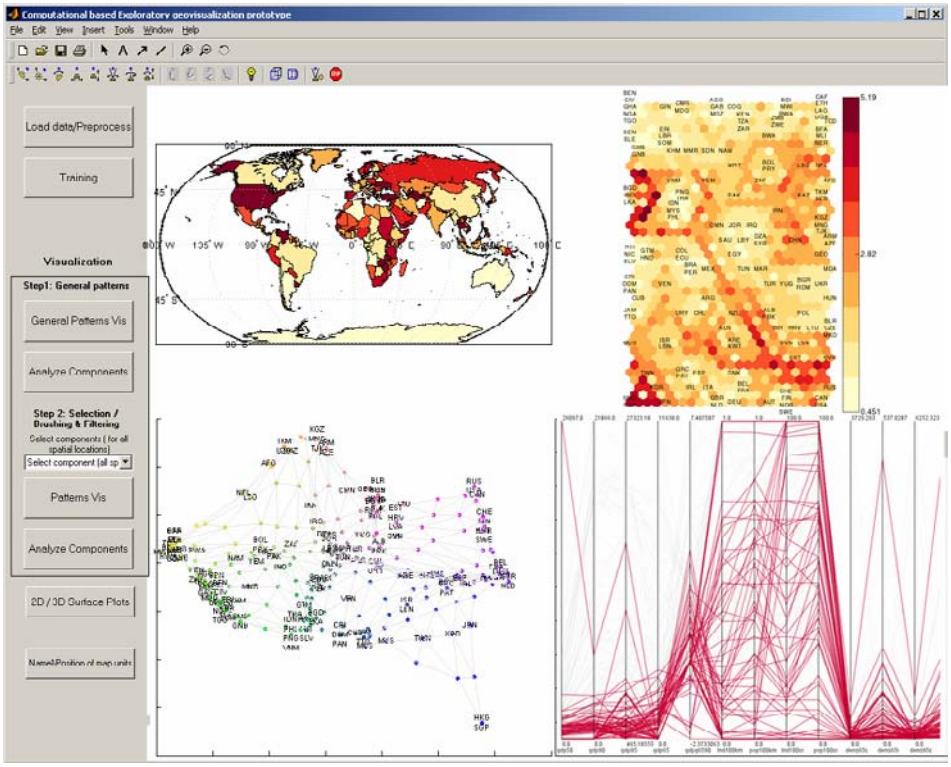


Figure 6.3. Screen shots of the different representations and the overall user interface and tools used in the evaluation.

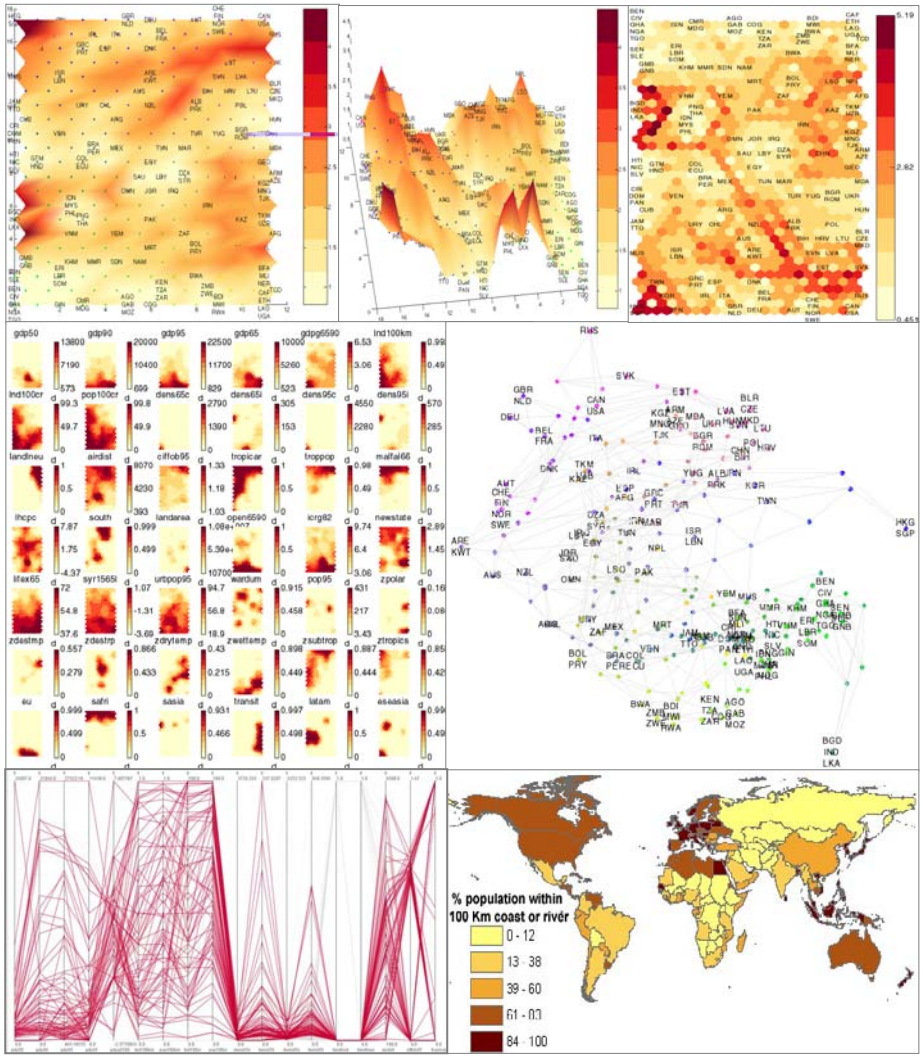


Figure 6.4. The different representations used in the evaluation: 2D surface, 3D surface, distance matrix representation, component plane visualization, 2D/3D projection, parallel coordinate plot (PCP) and maps.

The usability evaluation proposed here is goal-oriented and a number of specific test variables are defined. Appropriate usability indicators are drawn from the goals and corresponding measurements are set. The usability evaluation is designed to assess the user interactions, the functionality of the tool, the flexibility of the interface for exploratory tasks, and most importantly the ability to support the knowledge construction process. Test participants are involved in the evaluation of the different graphical representations in terms of specific end-user

tasks related to the general goal of finding patterns and relationships in the data. Participants are encouraged to interact with the interface.

6.4.1. Study design

To be able to answer the questions set for the evaluation, a clear design of the study is needed, in which each indicator is tested by multiple conditions. The objectives of the usability evaluation need to be clearly defined, as well as methods that are best suited for capturing the necessary data. There are several objectives for the proposed usability evaluation. The evaluation intends to assess the visualization tool's ability to meet user performance and satisfaction with regard to the general goal of exploring patterns and relationships in data. Examples would be the percentage of users that will be able to complete representative tasks within a certain time or without requiring assistance, or the percentage of users that will be satisfied with the usability of the tool. The assessment is based on a user test through which different options for map-based and interactive visualization of the output of SOM multivariate analysis are compared. Variants of the design (representation and display options) are presented to test subjects, and their performance scores are compared for the different representations. Interactive maps and PCP are used for comparison with the SOM-based representations (figure 6.4).

Users are asked to visually examine the representations, respond to questions, and report their preferences and viewpoints about the representation forms and the effect of the visual variables used, while completing a number of tasks defined for the evaluation (see previous section). The objective is to measure and compare task performance and the user's level of understanding for the different representation forms and tasks. The kind of evidence the evaluation intends to provide includes responses to specific questions focused on the use of features and representations to perform specific tasks, and how users interpret and understand and use the basic visualization features and representation forms. The evaluation examines the effectiveness of the particular displays. The idea is to investigate the use of SOM-based representations or a combination of them with other graphics and maps. Different performances offered by the different representations can then be compared according to the measures described in table 6.2. The evaluation study design is presented in figure 6.5 below.

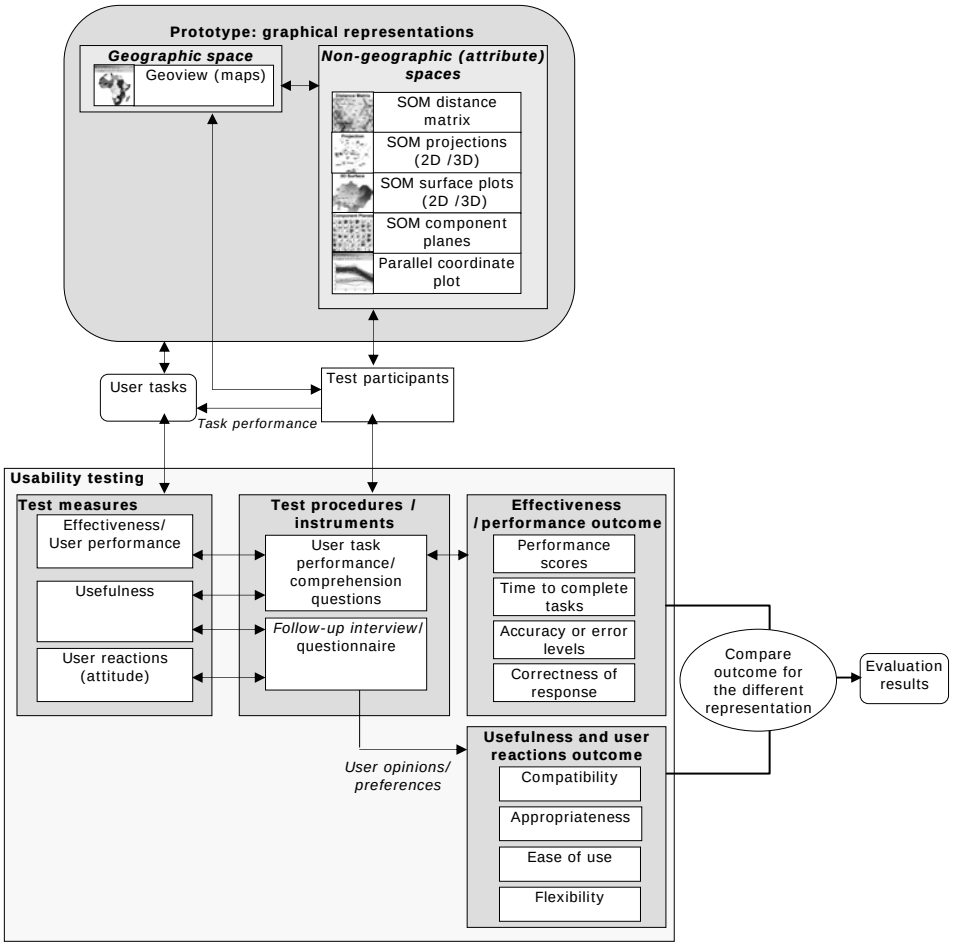


Figure 6.5. Evaluation study design: the different representations and display options are used by users to perform a number of exploratory tasks. The usability indicators examined include effectiveness/performance, usefulness and attitude/reactions. Performance/effectiveness measures used include task performance score, time to complete tasks, accuracy of user results, and correctness of user responses. Usefulness and attitude provide user opinions on the functionality, compatibility, and flexibility of the tool for user tasks.

6.4.2. Test measures

The proposed assessment methodology includes three criteria (see table 6.2): effectiveness/user performance, usefulness, and user reactions (attitude).

1. Effectiveness focuses on the tool functionality and examines the user's performance of the tasks, and how to manipulate any parameters or controls available to complete the tasks. This can be measured by the time spent on

completing tasks, the percentage of completed tasks (Sweeney et al. 1993; Rubin 1994; Fabricant 2001), the correctness of outcome of task performance and response, the success and accuracy (error rate and error types), the duration of time spent for help and questions, the range of function used and the level of success, the ease of use or level of difficulty, the frequency of documents access or request for help, and the level of user guidance and support.

2. Usefulness refers to the appropriateness of the tool's functionality and assesses whether the tool meets the needs and requirements of users when carrying out tasks, the extent to which users view the tools as supportive for their goals and tasks, and the individual user's level of understanding and interpretation of the tool's results and processes. It includes, flexibility, compatibility in relation to the user's expectations: finding patterns in data, relating different attributes, comparing values of attributes for different spatial locations. This is gathered through task performance, verbal protocols, post-hoc comments and responses on questionnaire.

3. User reactions refer to the user's attitude, opinions, subjective views, and preferences about the flexibility, compatibility (between the way the tool looks and works and the user's conventions and expectations). It can be measured using questionnaires and survey responses, and comments from interviews and ratings.

The specific usability measures and measuring methods used for the different tasks is described in table 6.2 below.

Table 6.2. Usability indicators used in the assessment.

	Usability indicators used		
	Effectiveness / user performance	**Usefulness**	**User reactions (attitude)**
Specific usability measures	- Correctness of outcome of task performance and response (success, percentage of completed tasks, accuracy or error rate) - Time to complete tasks - Time spent for help, documents access, guidance and support	- Compatibility and appropriateness in relation to user's expectations and goals - User's level of understanding and interpretation of the tool's results and processes	- Opinions, subjective views on the flexibility, compatibility (between the way the tool looks and works and the user's expectations), functionality, and appropriateness of the tool for the tasks - User preferences
Measuring method	- Examines tool functionality and the user's performance of the tasks and response to specific questions	- Task performance - Verbal protocols - Post-hoc comments - Responses on questionnaire - Answers to comprehension questions	- Questionnaires, interviews and survey responses, - Ratings

6.4.3. Evaluation tasks model

The evaluation of the graphical representations and interfaces needs to be grounded in a task model that can focus more on the user's goals and the tasks he or she needs to perform than on the interface side. The task model intends to support the development of the experimental set-up for the evaluation to cover the different levels of analysis included in the use of visualization tools. A task can be seen as a sequence of necessary steps and is comprised of objectives, the definition of the problem, and the methods necessary for the resolution of the problem, as described earlier in section 2. The conceptual goals and the different steps of the exploration and knowledge discovery process described in section 2 are used as the basis for defining low-level (operational) tasks that users need to perform to meet the conceptual goals.

Based on the three key visualization tasks and operators identified in section 2 (categorize and classify, compare, and reflect), we provide a low-level taxonomy of tasks (by decomposition of the basic visualization operators) that users might perform in a visual environment (see table 6.3). This decomposition of the basic visualization operators was obtained by analyzing task structures of real world visualization problems, representing the collection of subtasks, corresponding taxonomy or classification as well as a set of semantic relationships among these concepts, and other entities necessary to perform the task.

Table 6.3. Operational tasks derived from the visualization taxonomy.

Conceptual goals / visualization operators	**Operational visualization task**	**Task number**
Locate	Indicate data items of a certain range of value	1.1
	Indicate spatial positioning of elements of interest and spatial proximity among the different elements	1.2
Identify	Identify the different clusters or regions in the data	2.1
	Identify relationships between attributes	2.2
	Identify relationships between data items (within clusters and between different clusters)	2.3
Distinguish	Distinguish how a target value measured at one particular spatial location, or at various neighboring locations, varies for different attributes (e.g. different values of the same attribute at different spatial locations, and the value of different attributes at a specific spatial location)	3
Categorize	Define all the regions on the display, and draw boundaries	4
Cluster	Find gaps in the data on the display	5
Distribution	Describe the overall pattern (overview)	6
Rank	Indicate the best and worst cases in the display for an attribute	7
Compare	Compare values at different spatial locations, and the order of importance of objects (data items) accordingly	8
Associate	Form relationships between data items in the display	9
Correlate	Discern which data items share similar attributes	10

The full taxonomy mapped on the different representation methods used to represent each task contains a total of 49 tasks. Since each task is executed with 3, 4, 5 or 6 different representations, much time is needed to complete the test. In order to create a test that could be administered within a target duration of 1 hour and half, it was necessary to review the task structure. The process of grouping tasks was based on the conditions that the sampling of the experimental tasks is as broad as possible, and that the operational description of the selected tasks varies significantly. This was realized based on visual tasks taxonomy (Zhou and Feiner 1998) that include a set of dimensions by which the tasks can be grouped. The major dimensions of this taxonomy include visual accomplishments and visual implications. Visual accomplishments refers to the type of presentation intents that a visual might help to achieve while visual implications specify a particular type of visual action that a visual task may carry out. Three types of visual implications are proposed by (Zhou and Feiner 1998). They include: (1) visual organization, visual signaling, and visual transformations. The structure of the implication dimension of the visual taxonomy is described in table 6.4.

Table 6.4. Visual implications and related elemental tasks (Source: Zhou and Feiner 1998).

Implication	Type	Subtype	Elemental taks
Organization	Visual grouping	Proximity	Associate, cluster, locate
		Similarity	Categorize, cluster, distinguish
		Continuity	Associate, locate, reveal
		Closure	Cluster, locate, outline
	Visual attention		Cluster, distinguish, emphasize, locate
	Visual sequence		Emphasize, identify, rank
	Visual composition		Associate, correlate, identify, reveal
Signaling	Structuring		Tabulate, plot, structure, trace, map
	Encoding		Label, symbolize, portray, quantify
Transformation	Modification		Emphasize, generalize, reveal
	Transition		Switch

Based on the visual implications and related elemental tasks described in table 6.3, task 1.2 has been combined with task 4, task 2.1 with task 5, and task 2.3 with task 9.

The following experimental tasks are derived for the test (table 6.5).

Table 6.5. List of operational tasks derived from the taxonomy, and specific example tasks for the evaluation.

Conceptual goals / visualization operators	**Operational visualization task**	**Specific task explored in the study**	**Task number**
Locate	Indicate data items of a certain range of value	Indicate the poorest countries (reference to the 1995 GDP lower than 750)	1
Identify	Identify relationships between attributes	Identify possible relationships between the following attributes: population density in the coastal region and in the interior, and GDP per capita 95	2
Distinguish	Distinguish how a target value measured at one particular spatial location, or at various neighboring locations, varies for different attributes (e.g. different values of the same attribute at different spatial locations, and the value of different attributes at a specific spatial location)	How does income (GDP 1995) of the countries vary across space? Define differences and similarities between the countries	3
Categorize	Define all the regions on the display, and draw boundaries. Indicate spatial positioning of elements of interest and spatial proximity among the different elements	Define all the regions on the display, and draw boundaries. Define categories of countries such as rich, and poor countries on the display, and indicate the to which category South Africa belong. Are there African countries in this category? List the countries	4
Cluster	Find gaps in the data on the display	Find gaps in the data and indicate the different clusters	5
Distribution	Describe the overall pattern (overview)	What are the common characteristics of low-income countries (GDP lower than 750)?	6
Rank	Indicate the best and worst cases in the display for an attribute	Indicate the 5 lowest GDP countries and the 5 highest GDP	7
Compare	Compare values at different spatial locations, and the order of importance of objects (data items) accordingly	Compare population density on coastal regions (within 100 km of the coastline) and inland regions (beyond 100 km from the coastline)	8
Associate	Form relationships between data items in the display; Identify relationships between data items (within clusters and between different clusters)	Form relationships between the countries in the geographic tropics and their economic development (GDP 1995) as compared with the other countries	9
Correlate	Discern which data items share similar attributes	Examine economy development (GDP 95) across the countries: landlocked countries and countries that have access to the sea	10

The operational tasks described in table 6.5 are tested against all three usability indicators and corresponding measures described in table 6.2. Specific domain exploration tasks related to the dataset explored are used to illustrate each operational tasks (see table 6.6).

Table 6.6. Specification of user tasks and representation method used to represent task.

Conceptual goals /visualization operators	Operational visualization task	Task n°	Method used in the prototype to represent task	Representation number
Locate	Indicate data items of a certain range of value	1	Maps	1
			Parallel coordinate plot	2
			Component planes	3
Identify	Identify relationships between attributes	2	Maps	1
			Parallel coordinate plot	2
			Component planes	3
Distinguish	Distinguish how a target value measured at one particular spatial location, or at various neighboring locations, varies for different attributes (e.g. different values of the same attribute at different spatial locations, and the value of different attributes at a specific spatial location)	3	Maps	1
			Parallel coordinate plot	2
			Component planes	3
Categorize	Define all the regions on the display, and draw boundaries. Indicate spatial positioning of elements of interest and spatial proximity among the different elements	4	Unified distance matrix	4
			2D/3D projection	5
			2D/3D surface	6
Cluster	Find gaps in the data on the display	5	Unified distance matrix	4
			2D/3D projection	5
			2D/3D surface	6
			Parallel coordinate plot	2
Distribution	Describe the overall pattern (overview)	6	Map	1
			Parallel coordinate plot	2
			Component planes	3
			Unified distance matrix	4
			2D/3D projection	5
			2D/3D surface	6
Rank	Indicate the best and worst cases in the display for an attribute	7	Map	1
			Parallel coordinate plot	2
			Component planes	3
Compare	Compare values at different spatial locations, and the order of importance of objects (data items) accordingly	8	Maps	1
			Parallel coordinate plot	2
			Component planes	3
Associate	Form relationships between data items in the display; Identify relationships between data items (within clusters and between different clusters)	9	Maps	1
			Parallel coordinate plot	2
			Component planes	3
			Unified distance matrix	4
			2D/3D projection	5
			2D/3D surface	6
Correlate	Discern which data items share similar attributes	10	Maps	1
			Parallel coordinate plot	2
			Component planes	3

Each task is separately evaluated for the different test measures described in table 6.2 in a random order (see Appendices A1 and A2 for random numbers used in the test). The procedure for reporting task performance and the user's answers to questions is described below:

For ***effective / user performance,*** user performance is reported by the test administrator at the end of each task:

- completed/not completed

- time spent on completing the task
- time spent on help, documents access, guidance and support.

For ***usefulness,*** the participants are asked at the end of each task to indicate on a form if and how the different representations were supportive for the task, in terms of compatibility, flexibility, and appropriateness in relation to the user's expectations and goals and the user's level of understanding and interpretation of the tool's results and the processes.

For ***user reactions,*** participants are asked to provide their opinions and comments on the tools used for each task, as well as their preferences for the representations used for the specific task by rating them.

The forms used to record the measurements and answers are presented in Appendixes B1 and B2.

6.4.4. Test subjects

Participants are selected to represent the target population of people who are the likely users. The user group can include geographers, demographers, environmental scientists, epidemiologists, and others. For the present test, selected participants are geographers and environmental scientists who have experience in data analysis and the use of GIS. They are domain specialists who have knowledge about the data and have both the motivation and the qualifications to do a proper interpretation of the analysis.

Test subject domain knowledge
The dataset explored is a complex dataset on geography and economic development (Gallup et al. 1999), compiled to support analysis of the complex relationships between geography and macroeconomic growth. The dataset contains 48 variables on economy, physical geography, population and health for 150 countries. This dataset was explored in Chapter 2 and the results are used to validate the test participant's results.

6.4.5. Experimental procedure

The operational tasks described in table 6.3 are used in the experiment with sample cases from the dataset explored in the test. Target results are constructed for comparison with the participant's results. The participant's task will be completed if his or her result matches the test target results developed. The list of tasks, together with criteria for measuring whether they have been successfully completed, is available.

A test schedule is prepared to ease the planning and execution of the different test sessions. The schedule describes the location, time of each session, and the participant's name. The test environment consists of a computer installed with ArcGIS, Matlab software, and the prototype visualization tool. The test environment has been selected so that noise levels are minimum, in order to avoid disrupting the test. The test sessions are individual sessions in which the participant works in the presence of only the test administrator. Twenty participants, including geographers, cartographers and environmental scientists, with experience in data analysis and the use of GIS are invited to take part in the test. The dataset used is related to a general geographic problem, for which all the participants have the knowledge to conduct the analyses. Participants are allowed to ask questions during the test.

A script is made so that all participants are treated in the same way during the test session. The script describes the steps of the test in detail, and is read to each participant at the beginning of the session in order to ensure that all participants receive the same information. To allow the participants to refer back to the list of tasks as they attempt a task, a written description of the task is handed to each participant.

A logging sheet for each participant (at each session) is used to record timing, events, participant actions, concerns and comments. At the end of the session, a brief questionnaire is given to the participants to collect other comments they need to communicate.

An introduction is given at the beginning of each session. The introduction explains the purpose of the test, and introduces the test environment and the tools to be used. Participants are informed that the test consists of testing the design and tools, not their abilities. At the end of the introduction, participants' questions are answered. In the introduction, the participants are informed about the total number of tasks, but the tasks are given one at a time according to the random numbers assigned. To ensure that participants are at ease, are fully informed of any steps, and inquiries are answered, an introduction to each session is given. Participants are assured that they have the option to abandon any tasks that they are unable to complete. They are left to work quietly, without any interruption unless necessary. Participants are asked to report, as they work, any problems they find or things they don't understand.

The order of the task presentation is randomized. Individual test sessions are conducted using random numbers for randomizing the order of task presentation of the graphical representations for the 10 tasks, 3 to 6 graphical representations used for each task. For this final testing, that involves actual use of the different tools (graphical representation) and logging time for the tasks, we involve 20 participants as recommended by usability methods (Nielsen 1994b).

A pilot test is first run to test any anomalies, solve any timing problems, and tune the experimental set-up. The average time required to complete all the tasks is 90 minutes.

6.5. Conclusion

In this chapter we have presented an evaluation strategy for assessing the usability and usefulness of the visual-computational analysis environment designed in the previous chapters. The evaluation method emphasizes exploratory tasks and knowledge discovery support. Its is based on the examination of a taxonomy of conceptual visualization goal and tasks. These tasks were decomposed into operational visualization tasks and experimental tasks related a the particular dataset to be used in the evaluation. The test in the next Chapter will involve the experimental tasks derived here and relate them to the conceptual visualization goals. New representation forms used to visualize geospatial data such as the SOM use new alternative techniques to represent the attribute spaces. An important step in the design of such visualization tools will rely on understanding the way users make interpretations of the information spaces. The choice of a proper representation metaphor is crucial to the successful use of the tools. The methodology presented here will be used examine the effectiveness of the proposed representations for exploratory tasks, and for knowledge discovery support as compared to maps and parallel coordinate plots.

Chapter 7

Usability testing and results

7.1. Introduction

Usability testing or user testing is a fundamental usability method of assessing the effectiveness, usefulness and performance of a tool. In geovisualization particularly, user testing can provide insight into how a visual interface can support data exploration tasks. The use of new representation forms and interactive means to visualize geospatial data requires an understanding of the impact of the visual tools used for data exploration and knowledge construction. Thus user testing is becoming an important step in the improvement of the design of geovisualization environments. The lack of appropriate evaluation methodology in the geovisualization domain has, however, limited the number of user testing experiments so far.

Since the design of effective visualization tools will rely on understanding the way users make interpretations of the information spaces used to represent patterns and relationships in data, the choice of a representation method is crucial to the success of a visualization environment. One of the dominant approaches in geovisualization is the integration of several representation methods that provide different perspectives of the data in multiple linked views. Such an integration of views can be more effective if focused on the potential of the individual representations for specific conceptual visualization goals that can better support the exploration, evaluation and interpretation of patterns and ultimately support knowledge construction. Empirical testing of the visualization tools can provide such insights into the potential of particular visual displays.

Based on an evaluation strategy explored in Chapter 3 and the subsequent evaluation methodology developed in Chapter 6, which emphasizes the use of a taxonomy of visualization tasks, we have designed a user testing experiment in which the representation methods described in the previous chapters are compared with maps and parallel coordinate plots.

In this chapter we present the experimental procedures applied during the usability testing, and an analysis and discussion of the results. These results are

---

This chapter is based on:

Koua E. L., MacEachren, A. and Kraak M.J. (Submitted). Evaluating the usability of an exploratory geovisualization environment. Submitted to International Journal of Geographical Information Science.

organized according the visual tasks derived from the taxonomy of visualization conceptual goals and operations described in Chapter 6. The results are compared for the different representations: maps, parallel coordinate plots, and the SOM-based representations.

7.2. A brief summary of the methodology

The methodology used for the test was developed in the previous chapter. Here we provide a summary of its application during the test. The evaluation is based on a user-based and task-based usability method intended to address the exploratory nature of geovisualization. Based on a taxonomy of visualization tasks, experimental tasks were developed and related to the conceptual visualization and data analysis goals. The test participants are involved in evaluating the different graphical representations in terms of the derived specific experimental tasks, which are end-user tasks related to the general goal of finding patterns and relationships in the data.

Each representation method used to represent the tasks is assessed according to defined test measures. These measures include effectiveness/performance, usefulness, and user reactions (attitude). For effectiveness we examine the correctness of response and the time taken to complete each task. Usefulness is examined by user rating of the representations in terms of compatibility with the user's expectations of the tool for the given task, flexibility and ease of use, and the user's reported understanding of the tool. User reactions (attitude) are examined by user rating of the representations in terms of satisfaction and preference for a given task. The objective is to measure and compare performance scores (effectiveness), the user's level of understanding, and the different ratings related to usefulness and user reactions for the different representation forms and conceptual visualization tasks. Maps and parallel coordinate plots are used for comparison with the SOM-based representations. While completing the tasks, users are asked to report their preferences and viewpoints about the representation forms. Since the evaluation focuses on individual tasks, which take an average of three minutes each, we did not include the time taken for questions or the level of guidance needed by participant for each task as measures.

7.3. Experiment

7.3.1. Test environment

The test environment consists of a computer installed with ArcGIS, Matlab software, and the prototype visualization tool. The environment was selected so that noise levels were minimum, in order to avoid disrupting the test. The prototype graphical user interface of the visualization environment was used to load the dataset and the different representations. The interface integrates the

use of maps, a parallel coordinate plot and SOM-based clustering representations such as the unified distance matrix representation, 2D/3D projection, 2D/3D surface, as well as component plane displays for the exploration of relationships among attributes of the dataset.

The test was organized in individual sessions in which the participant works in the presence of the test administrator. The individual SOM-based graphical representations were programmed to be used separately in a window with interactive features provided in the Matlab interface (zooming, panning, rotation and 3D view). ArcGIS was used for tasks involving maps, and a free and fully functional Java-based interactive parallel coordinate plot was used, with the basic features needed for the test (brushing, automatic display of names of countries and values of variables as the mouse moves over the data records, adding and removing attributes from the display).

Two forms were used to record user task performance and the different ratings (see Appendices B1 and B2). Task performance was reported by the test administrator on Appendix B1. User ratings on usefulness (compatibility, ease of use, understanding) and user reactions (satisfaction and preferences) were reported by the participants on Appendix B2 for the different tasks and representations used.

7.3.2. Pilot testing

The first two candidate users were used as pilot test subjects to ascertain any deficiencies in the test procedures, such as tasks descriptions, timing of each test session, the rating system, and instructions for test tasks. A revision was made based on the problems detected during pilot testing, particularly of the task description and timing.

7.3.3. The data explored

The test involves a complex dataset on geography and economic development (Gallup et al. 1999), compiled to support analysis of the complex relationships between geography and macroeconomic growth. The dataset contains 48 variables on economy, physical geography, population and health for 150 countries (see figure 3.8). This dataset was explored in Chapter 3 and the conclusions of the exploration are used to validate the test participant's results.

7.3.4. Participants

Twenty participants from an initial list of 25 who met the profile set for the test agreed to make time for the test. The profile of test participants was a target

population that included geographers, demographers, environmental scientists, and epidemiologists – likely users of such a geovisualization environment. The selected participants were GIScience domain specialists, with knowledge of such a dataset that involves geography and economic development. They also had both the motivation and the qualifications to do a proper interpretation of the analysis. They included geographers, cartographers, geologists, and environmental scientists, and all had had experience in data analysis and the use of GIS. Most of them are pursuing a PhD research. The selection of the sample size (20 participants) was based on recommendations from usability engineering literature (Nielsen 1994b) regarding final testing that involves actual use and logging time.

7.3.5. Experimental procedure

An introduction to the test was presented to each participant at the beginning of each session. The introduction explained the purpose of the test, the geovisualization environment, the tools to use, the dataset, the forms for reporting, and the different rating levels. Participants were informed about the total number of tasks, and assured that the test did not assess their abilities, but rather the tools used. At the end of the introduction, participants' questions were answered. Each tool was explained to the participants, their questions were answered, and they were asked to confirm that they understood how the tool works or to ask more questions. Participants were assured that they had the option to abandon any task that they were unable to complete, and were asked to report, as they worked, any problems they found or things they didn't understand. The introduction and all the steps of the test were contained in a script so that all the participants were treated in the same way during the session and received the same information.

During the test, participants were involved in the evaluation of the different graphical representations in terms of specific end-user tasks related to the operational visualization tasks described in Chapter 6, with sample cases from the dataset explored in the test. They were encouraged to interact with the interface. While completing the tasks, they were asked to report their preferences and viewpoints about the representation forms. Target results were constructed for comparison with the participant's results. The participant's task was judged completed if his or her result matched the test target results developed. The list of tasks, together with criteria for measuring whether they had been successfully completed, was available. The tasks were written on separate sheets and were given one at a time according to the random numbers assigned (see Appendices A1 and A2). The rationale behind the use of random numbers for the order of task presentation and the graphical representations for each of the 10 tasks (three to four graphical representations were used for each task) was to reduce the learning effect for the sample size.

The average time required to complete all the tasks was 90 minutes. On a logging sheet, the time for user task performance for each representation was recorded, as well as participants' opinions and comments. Participants were allowed to ask questions during the test.

7.4. Test results

The analysis of the test results is organized according to the usability measures described in Chapter 6: effectiveness/performance, usefulness, and user reactions (see figure 6.2). The results are also presented by experimental tasks and corresponding conceptual visualization goals. The tasks are grouped into clustering (cluster and categorize) and exploration (locate, identify, distinguish, compare, rank, distribution, associate, correlate).

7.4.1. Analysis of effectiveness and performance

1. Correctness of response

Correctness of response was used as a measure of performance. A task completed with the correct response is given 1 and a task not completed or completed with the wrong response is assigned 0. The analysis of the correctness of response shows that the parallel coordinate performed poorly compared with maps and SOM component planes (see figure 7.1). The SOM component plane display performed well for all tasks. The map performed well generally, except for task 6 (distribution), task 2 (identify) and task 8 (compare).

The component plane display performed better than maps and the parallel coordinate plot for visualization tasks such as '*identify'*, '*distribution'*, '*correlate'*, '*compare'* and '*associate'*. The maps were as good as component planes for tasks such as '*locate'*, '*distinguish'* and '*rank'*. For these tasks the parallel coordinate plot performed poorly, especially for '*rank'*, '*associate'* and '*distinguish'*.

For the tasks '*cluster'* and '*categorize'* (see figure 7.2), the SOM-based representations (unified distance matrix, 2D/3D surface and 2D/3D projection) performed equally well and far better than the parallel coordinate plot. For revealing the categories, the unified distance matrix was found less effective than the 2D/3D projection and 2D/3D surface. The 2D/3D projection was found to be more effective for finding the categories.

The percentage of completed tasks with correct response and the corresponding time taken is presented in table 7.1. The tasks were described in table 6.5 and 6.6.

Table 7.1 Percentage of completed task with correct response.

Task number	***Conceptual visualization goal***	***Representation method***	***Percentage of completed task with correct response***	***Average time taken to complete task (in seconds)***
Task 1 Indicate the poorest countries (reference to the 1995 GDP lower than 750)	***Locate***	Maps	100	41.40
		Parallel coordinate plot	75	86.33
		Component planes	100	41.10
Task 2 Identify possible relationships between the following attributes: population density in the coastal region and in the interior, and GDP per capita 95	***Identify***	Maps	60	177.00
		Parallel coordinate plot	55	99.50
		Component planes	90	104.37
Task 3 How does income (GDP 1995) of the countries vary across space? Define differences and similarities between the countries	***Distinguish***	Maps	100	69.80
		Parallel coordinate plot	35	126.63
		Component planes	95	107.00
Task 4 Define all the regions on the display, and draw boundaries. Define categories of countries such as rich, and poor countries on the display, and indicate the to which category South Africa belong. Are there African countries in this category? List the countries	***Categorize***	Unified distance matrix	80	86.71
		2D/3D Projection	95	93.45
		2D/3D Surface	90	98.06
Task 5 Find gaps in the data and indicate the different clusters	***Cluster***	Unified distance matrix	100	47.45
		2D/3D Projection	100	48.80
		2D/3D Surface	100	35.35
		Parallel coordinate plot	55	37.17
Task 6 What are the common characteristics of low-income countries (GDP lower than 750)?	***Distribution***	Maps	35	38.20
		Parallel coordinate plot	40	179.56
		Component planes	100	106.05
Task 7 Indicate the 5 lowest GDP countries and the 5 highest GDP	***Rank***	Maps	100	120.85
		Parallel coordinate plot	90	138.06
		Component planes	100	90.40
Task 8 Compare population density on coastal regions (within 100 km of the coastline) and inland regions (beyond 100 km from the coastline)	***Compare***	Maps	65	245.69
		Parallel coordinate plot	50	116.40
		Component planes	95	98.10
Task 9 Form relationships between the countries in the geographic tropics and their economic development (GDP 1995) as compared with the other countries	***Associate***	Maps	80	129.13
		Parallel coordinate plot	55	91.83
		Component planes	100	60.90
Task 10 Examine economy development (GDP 95) across the countries: landlocked countries and countries that have access to the sea	***Correlate***	Maps	85	114.39
		Parallel coordinate plot	50	57.45
		Component planes	100	64.25

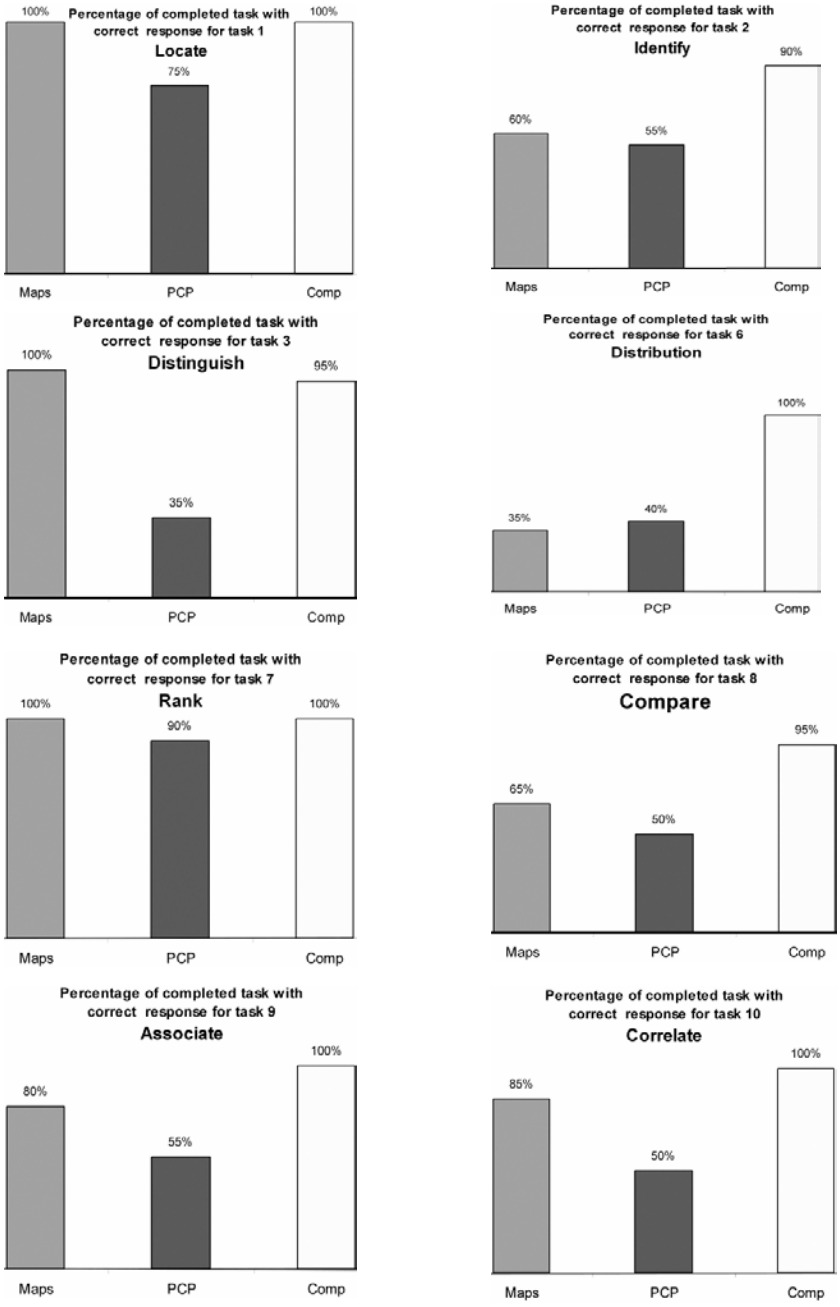


Figure 7.1. Percentage of completed task with correct response for the different visualization tasks. PCP= Parallel coordinate plot, Comp=SOM component plane display. The tasks have been organized in two groups: clustering tasks (tasks number 4 and 5), and detail exploration tasks (tasks number 1, 2, 3, 6, 7, 8, 9, 10). In this figure only the detail exploration tasks are presented.

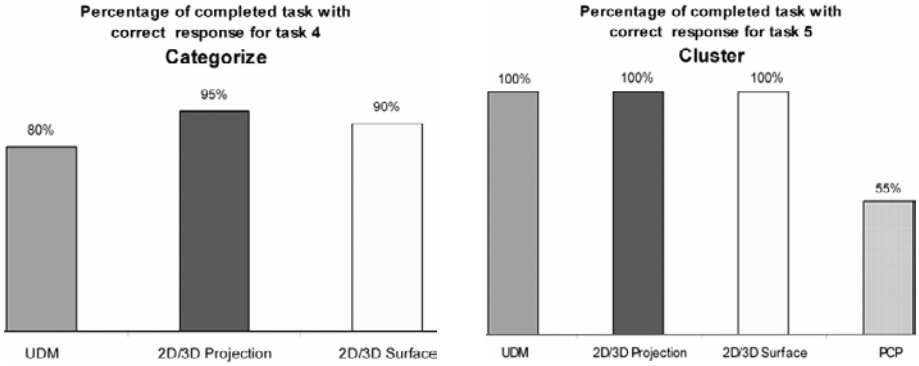


Figure 7.2. Percentage of completed task with correct response for the different visualization tasks. UDM=Unified distance matrix, PCP= Parallel coordinate plot. The tasks have been organized in two groups: visual grouping tasks (tasks number 4 and 5), and detail exploration tasks (tasks number 1, 2, 3, 6, 7, 8, 9, 10). In this figure only the visual grouping tasks are presented.

Further analysis of the correctness of response measure was conducted using a pair-wise comparison of the mean scores for the different representations for each conceptual visualization goal examined. Statistics of the paired sample tests are presented in table 7.2. The paired sample tests show significant differences ($p<0.05$) in the mean scores for the different tasks. For the task '*locate'*, the map and the component plane display performed equally well (with 100% successful task completion by users), compared with the parallel coordinate plot (75% successful task completion by users). For this task, a significant difference was found between the map and the parallel coordinate plot ($P=0.021$), and between the component plane display and the parallel coordinate plot ($P=0.021$).

For the task *'identify'*, the map and parallel coordinate plot performed relatively poorly (60% and 55% successful task completion by users respectively), compared with the component plane display (90%). The component plane display shows a significant difference to the map ($p=0.030$) and to the parallel coordinate plot ($p=0.005$).

The map and the component plane display performed well for the task *'distinguish'* (100% and 95% successful task completion by users respectively), with no significant difference. Both show a significant difference ($p=0.000$) to the parallel coordinate plot, which performed poorly for this task (35% successful task completion by users).

The component plane display performed far better than the map and the parallel coordinate plot for the task '*distribution'*. The difference in the mean scores of the component plane display is significant with the map ($P=0.000$) and the parallel coordinate plot ($p=0.000$). The map and parallel coordinate plot performed poorly (35% and 40% successful task completion by users respectively).

The three representations (map, parallel coordinate plot and component plane display) performed equally well for the task *'rank'*.

For the tasks *'compare', 'associate'* and *'correlate'*, the component plane display performed better (100% successful task completion by users for '*correlate'* and '*associate'*, and 95% for '*compare'*) than the map and the parallel coordinate plot. The map performed relatively well (85% and 80% successful task completion by users respectively) for '*correlate'* and '*associate'* but relatively poorly for the task '*compare'* (65% successful task completion by users). The results show significant differences between the component plane display and the map ($p=0.010$ for '*compare'*, $p=0.042$ for '*associate'*) and no significant difference for *'correlate'*. The difference is significant between the component plane display and the parallel coordinate plot for '*compare'* ($p=0.001$), '*associate'* ($p=0.001$), and for '*correlate'* ($p=0.000$). The map and the parallel coordinate plot show no difference for the tasks '*compare'* and '*associate'*, but a significant difference for '*correlate'* ($p=0.031$), with 85% successful task completion by users for the map and 50% for the parallel coordinate plot.

No significant difference was found for the task '*categorize', one of the* visual grouping tasks (cluster and categorize), for which the representation methods explored included the unified distance matrix, the 2D/3D projection, the 2D/3D surface, and the parallel coordinate plot (only examined for the task cluster). Since labelling data items is not appropriate in the parallel coordinate plot, it was not applied for the task '*categorize'*. All SOM-based representation methods for these two tasks performed well (100% successful task completion by users) compared with the parallel coordinate plot (55%). These representations show a significant difference ($p=0.001$) compared with the parallel coordinate plot for the task '*cluster'*.

Table 7.2. Paired samples test for correctness of response. (*) indicates that the difference is significant (P<0.05). Udm = unified distance matrix, Pcp = parallel coordinate plot, Comp = SOM component plane display, Proj = 2D/3D projection, Surf = 2D/3D surface plot.

Conceptual visualization goal	Representation method	Paired differences					t	df	P-value Sig. (2-tailed)
		Mean	Std. deviation	Std. error mean	95% confidence interval of the difference				
					Lower	Upper			
Locate	Map - Pcp	0.25	0.444	0.099	0.042	0.458	2.517	19	0.021*
	Pcp - Comp	-0.25	0.444	0.099	-0.458	-0.042	-2.517	19	0.021*
Identify	Map - Pcp	0.05	0.510	0.114	-0.189	0.289	0.438	19	0.666
	Map - Comp	-0.3	0.571	0.128	-0.567	-0.033	-2.349	19	0.030*
	Pcp - Comp	-0.35	0.489	0.109	-0.579	-0.121	-3.199	19	0.005*
Distinguish	Map - Pcp	0.65	0.489	0.109	0.421	0.879	5.940	19	0.000*
	Map - Comp	0.05	0.224	0.050	-0.055	0.155	1.000	19	0.330
	Pcp - Comp	-0.6	0.598	0.134	-0.880	-0.320	-4.485	19	0.000*
Categorize	Udm - Proj	-0.15	0.489	0.109	-0.379	0.079	-1.371	19	0.186
	Udm - Surf	-0.1	0.308	0.069	-0.244	0.044	-1.453	19	0.163
	Proj - Surf	0.05	0.394	0.088	-0.134	0.234	0.567	19	0.577
Cluster	Udm - Pcp	0.45	0.510	0.114	0.211	0.689	3.943	19	0.001*
	Proj - Pcp	0.45	0.510	0.114	0.211	0.689	3.943	19	0.001*
	Surf - Pcp	0.45	0.510	0.114	0.211	0.689	3.943	19	0.001*
Distribution	Map - Pcp	-0.05	0.759	0.170	-0.405	0.305	-0.295	19	0.772
	Map - Comp	-0.65	0.489	0.109	-0.879	-0.421	-5.940	19	0.000*
	Pcp - Comp	-0.6	0.503	0.112	-0.835	-0.365	-5.339	19	0.000*
Rank	Map - Pcp	0.1	0.308	0.069	-0.044	0.244	1.453	19	0.163
	Pcp - Comp	-0.1	0.308	0.069	-0.244	0.044	-1.453	19	0.163
Compare	Map - Pcp	0.15	0.671	0.150	-0.164	0.464	1.000	19	0.330
	Map - Comp	-0.3	0.470	0.105	-0.520	-0.080	-2.854	19	0.010*
	Pcp - Comp	-0.45	0.510	0.114	-0.689	-0.211	-3.943	19	0.001*
Associate	Map - Pcp	0.25	0.716	0.160	-0.085	0.585	1.561	19	0.135
	Map - Comp	-0.2	0.410	0.092	-0.392	-0.008	-2.179	19	0.042*
	Pcp - Comp	-0.45	0.510	0.114	-0.689	-0.211	-3.943	19	0.001*
Correlate	Map - Pcp	0.35	0.671	0.150	0.036	0.664	2.333	19	0.031*
	Map - Comp	-0.15	0.366	0.082	-0.321	0.021	-1.831	19	0.083
	Pcp - Comp	-0.5	0.513	0.115	-0.740	-0.260	-4.359	19	0.000*

2. Time to complete tasks

Time to complete the tasks was used as a second variable for the performance measure. The analysis of the time taken to complete the tasks reveals some important differences between the different representations used (see figure 7.3). Table 7.3 shows detailed statistics of the time spent on the different tasks and representations.

Table 7.3. Descriptive statistics for the time spent on completing the tasks.

Task number	Conceptual visualization goal	Representation method	N	Minimum	Maximum	Mean	Std. deviation
Task 1	**Locate**	Maps	20	11	117	41.40	30.275
		Parallel coordinate plot	15	20	260	86.33	65.976
		Component planes	20	4	191	41.10	45.588
Task 2	**Identify**	Maps	13	61	476	190.62	119.362
		Parallel coordinate plot	11	62	290	180.91	76.470
		Component planes	18	24	251	110.17	59.848
Task 3	**Distinguish**	Maps	20	16	208	69.80	50.671
		Parallel coordinate plot	7	48	286	144.71	93.660
		Component planes	20	4	519	107.00	124.980
Task 4	**Categorize**	Unified distance matrix	16	17	211	92.13	62.049
		2D/3D projection	19	16	233	98.37	63.949
		2D/3D surface	18	19	243	98.06	55.943
Task 5	**Cluster**	Unified distance matrix	20	4	114	47.45	34.705
		2D/3D projection	20	6	373	48.80	78.648
		2D/3D surface	20	5	110	35.35	30.901
		Parallel coordinate plot	11	4	102	40.55	28.500
Task 6	**Distribution**	Maps	7	6	260	109.14	99.607
		Parallel coordinate plot	9	62	621	179.56	177.256
		Component planes	20	5	319	106.05	78.451
Task 7	**Rank**	Maps	20	25	336	120.85	72.836
		Parallel coordinate plot	18	78	254	138.06	60.910
		Component planes	20	38	179	90.40	36.787
Task 8	**Compare**	Maps	13	94	711	245.69	161.916
		Parallel coordinate plot	9	48	272	129.33	78.187
		Component planes	19	4	196	103.26	54.690
Task 9	**Associate**	Maps	16	23	246	129.13	62.976
		Parallel coordinate plot	12	10	199	91.83	57.588
		Component planes	20	9	250	60.90	61.053
Task 10	**Correlate**	Maps	18	23	284	114.39	73.989
		Parallel coordinate plot	10	44	225	114.90	57.047
		Component planes	20	5	143	64.25	36.009

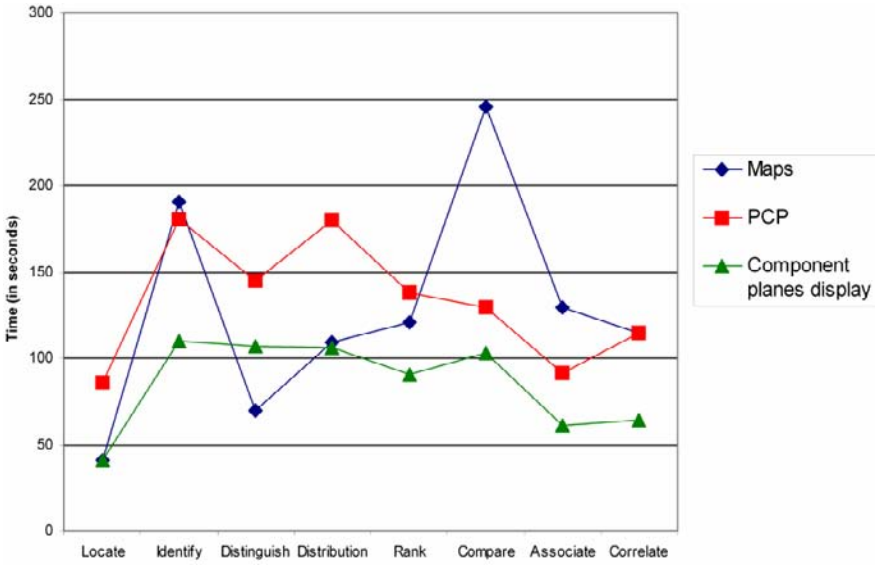


Figure 7.3. Time to complete tasks for three exploratory tools: map, parallel coordinate plot (PCP), and component plane display.

In general the component plane display required less time than the maps and the parallel coordinate plot for the different tasks. The map was faster for '*distinguish'*, but a far slower medium for comparison tasks (see figure 7.3). For the tasks of clustering, all the representation methods used (unified distance matrix representation, 2D/3D projection, 2D/3D surface plot and parallel coordinate plot) required less time to perform the tasks (between 35 seconds and 50 seconds). In figure 7.4 the time spent using the different representations is presented for each task.

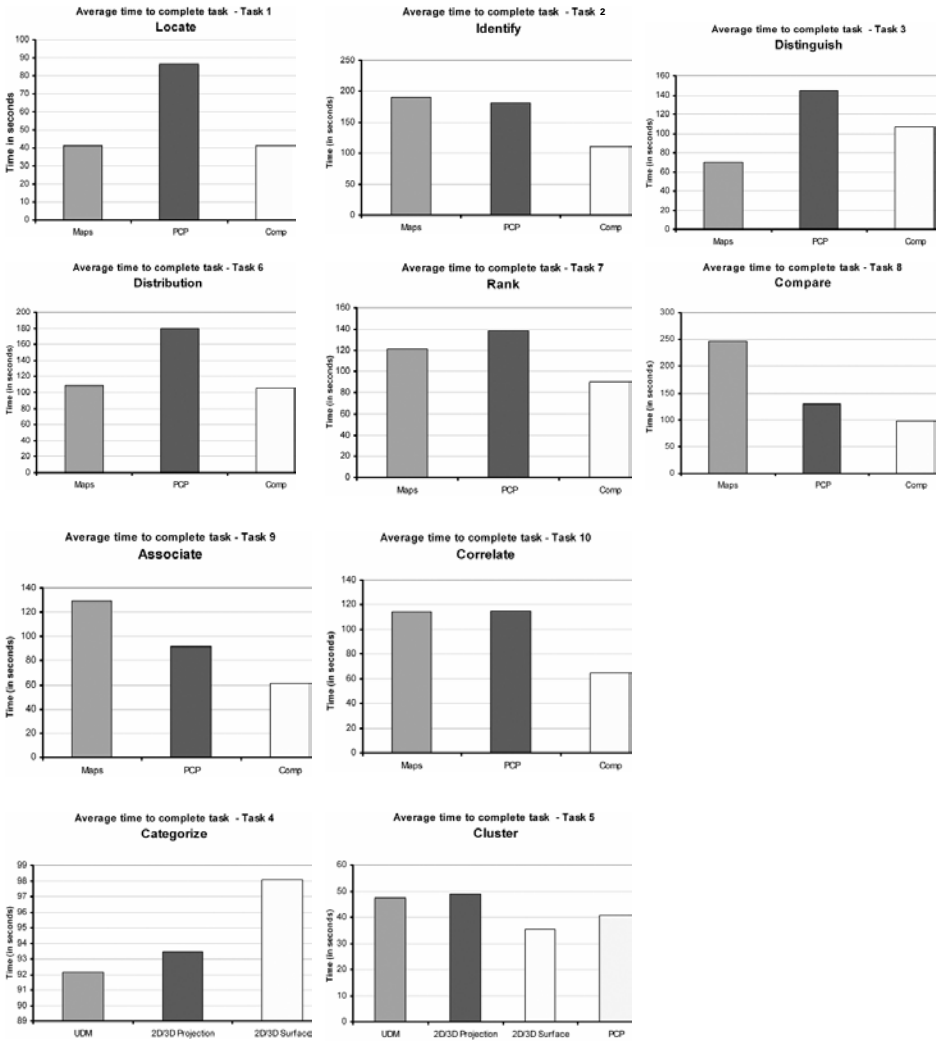


Figure 7.4. Time taken to successfully complete the tasks. The tasks have been organized in two groups: detail exploration tasks (tasks number 1, 2, 3, 6, 7, 8, 9, 10), and visual grouping tasks (tasks number 4 and 5). PCP= Parallel coordinate plot, Comp=SOM component plane display, UDM=Unified distance matrix.

Further analysis was conducted using a paired-wise t-test with the different representations to compare the mean scores for the time taken to complete the tasks. Detailed statistics of the paired-wise sample test are presented in table 7.4.

The paired sample test shows no significant differences between the representations for the tasks '*distribution'*, '*cluster'* and '*categorize'*.

For the task '*locate'*, no significant difference was found between the map and the component plane display. Both representations required just 41 seconds to complete the task. The parallel coordinate plot required double the time needed by the map and the component plane display for the same task. Thus a significant difference was found between the parallel coordinate plot and the map ($p=0.005$) and the component plane display ($p=0.002$).

The '*identify'* task took relatively longer with all the representations (an average of 190 seconds with the map, 180 seconds with the parallel coordinate plot, and 110 seconds with the component plane display). For this task, no significant difference was found between the map and the parallel coordinate plot, or between the map and the component plane display. The difference is significant between the component plane display and the parallel coordinate plot ($p=0.020$).

For the task '*distinguish'*, a significant difference in the mean score for the time spent on the tasks was found between the map and parallel coordinate plot ($p=0.028$). While the map requires a little more than one minute (69 seconds) to complete the task, the parallel coordinate plot requires more than two minutes (144 seconds). The difference in time spent using the component plane display (107 seconds) and the map (69 seconds) was not found significant.

For the task '*rank'*, the participants spent on average 90 seconds for the component plane display, 120 seconds for the map, and 138 seconds for the parallel coordinate plot. The mean scores for the time spent were only significantly different between the component plane display and the parallel coordinate plot ($p=0.003$). No significant difference was found between the map and component plane display.

The map required a lot more time (245 seconds) than the component plane display (103 seconds) and the parallel coordinate plot (129 seconds) for the task '*compare'*. A significant difference was found between the map and the component plane display ($p=0.012$).

The task '*associate'* required on average 60 seconds for the component plane display, 91 seconds for the parallel coordinate plot, and 129 seconds for the map. The results show a significant difference between the map and the component plane display ($p=0.020$).

For the task '*correlate'*, the map and parallel coordinate plot required on average the same amount of time (114 seconds). The component plane display required 64 seconds, roughly half the time needed with the map and parallel coordinate plot. This shows a significant difference between the component plane display and the map ($p=0.016$) and the parallel coordinate plot ($p=0.006$).

Table 7.4. Paired samples test for the time taken to complete the tasks. (*) indicates that the difference is significant (P<0.05). Udm = unified distance matrix, Pcp = parallel coordinate plot, Comp = SOM component plane display, Proj = 2D/3D projection, Sur f= 2D/3D surface plot.

Task		Paired differences					t	df	P-value Sig. (2-tailed)
		Mean	Std. deviation	Std. error mean	95% confidence interval of the difference				
					Lower	Upper			
Locate	Map - Pcp	-46.267	53.526	13.820	-75.909	-16.625	-3.348	14	0.005*
	Map - Comp	0.300	52.374	11.711	-24.212	24.812	0.026	19	0.980
	Pcp - Comp	50.200	52.757	13.622	20.984	79.416	3.685	14	0.002*
Identify	Map - Pcp	22.000	132.990	42.055	-73.136	117.136	0.523	9	0.614
	Map - Comp	83.167	140.597	40.587	-6.164	172.498	2.049	11	0.065
	Pcp - Comp	75.273	90.459	27.274	14.502	136.044	2.760	10	0.020*
Distinguish	Map - Pcp	-90.000	82.595	31.218	-166.388	-13.612	-2.883	6	0.028*
	Map - Comp	-37.200	118.804	26.565	-92.802	18.402	-1.400	19	0.178
	Pcp - Comp	3.143	151.201	57.149	-136.695	142.981	0.055	6	0.958
Categorize	Udm - Proj	-1.133	58.481	15.100	-33.519	31.252	-0.075	14	0.941
	Udm - Surf	-5.313	83.028	20.757	-49.555	38.930	-0.256	15	0.801
	Proj - Surf	2.882	82.722	20.063	-39.649	45.414	0.144	16	0.888
Cluster	Udm - Proj	-1.350	74.025	16.553	-35.995	33.295	-0.082	19	0.936
	Udm - Surf	12.100	43.086	9.634	-8.065	32.265	1.256	19	0.224
	Udm - Pcp	11.273	42.626	12.852	-17.364	39.910	0.877	10	0.401
	Proj - Surf	13.450	64.779	14.485	-16.868	43.768	0.929	19	0.365
	Proj - Pcp	25.273	106.432	32.091	-46.229	96.775	0.788	10	0.449
	Surf - Pcp	-6.636	35.870	10.815	-30.734	17.461	-0.614	10	0.553
Distribution	Map - Pcp	13.000	107.480	76.000	-952.672	978.672	0.171	1	0.892
	Map - Comp	-1.571	142.078	53.701	-132.972	129.829	-0.029	6	0.978
	Pcp - Comp	67.667	193.987	64.662	-81.445	216.778	1.046	8	0.326
Rank	Map - Pcp	-14.889	84.948	20.023	-57.133	27.355	-0.744	17	0.467
	Map - Comp	30.450	75.701	16.927	-4.979	65.879	1.799	19	0.088
	Pcp - Comp	55.944	67.753	15.969	22.252	89.637	3.503	17	0.003*
Compare	Map - Pcp	126.333	245.452	100.205	-131.253	383.919	1.261	5	0.263
	Map - Comp	132.077	160.423	44.493	35.134	229.020	2.968	12	0.012*
	Pcp - Comp	30.444	110.263	36.754	-54.312	115.200	0.828	8	0.432
Associate	Map - Pcp	40.444	83.539	27.846	-23.769	104.658	1.452	8	0.184
	Map - Comp	66.250	101.841	25.460	11.983	120.517	2.602	15	0.020*
	Pcp - Comp	32.833	83.512	24.108	-20.228	85.895	1.362	11	0.200
Correlate	Map - Pcp	25.222	98.776	32.925	-50.704	101.148	0.766	8	0.466
	Map - Comp	54.500	86.186	20.314	11.641	97.359	2.683	17	0.016*
	Pcp - Comp	54.000	48.360	15.293	19.406	88.594	3.531	9	0.006*

To better view the range, highest and lowest values, extreme values, outliers and the median, box plots are shown in figures 7.5 and 7.6.

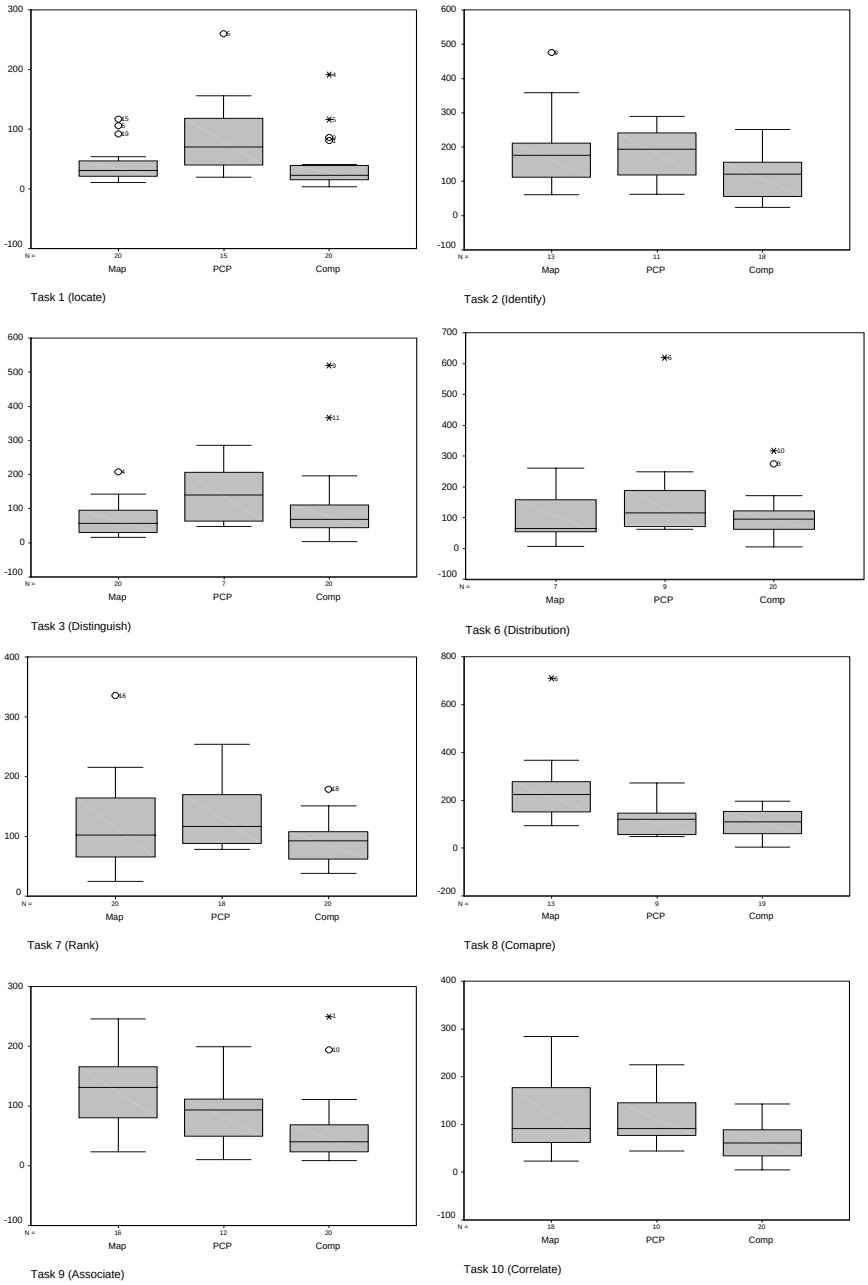


Figure 7.5. Box plots for the time spent on completing the tasks using the exploration tools. The box plots reveal the range of values, outliers (*), extreme values (0), as well as the median.

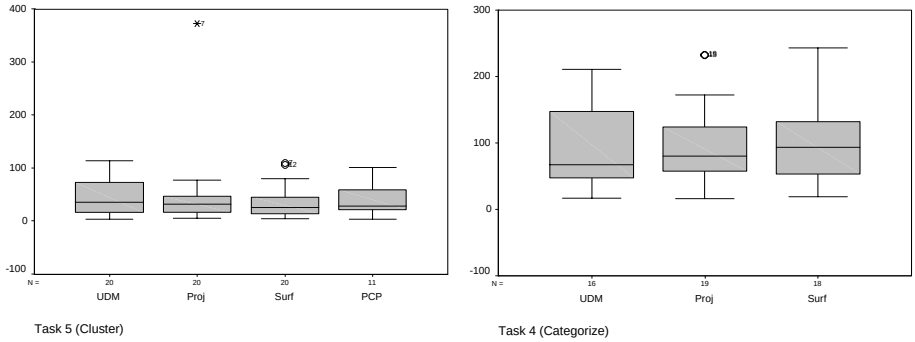


Figure 7.6. Box plots for clustering and categorization tools. UDM=Unified distance matrix, Proj= 2D/3D Projection, Surf= 2D/3D surface, PCP= Parallel coordinate plot.

7.4.2. Usefulness and user reactions

Usefulness and user reactions were reported using a five-point scale on the form presented in Appendix B2 (5 = very good, 4 = good, 3 = fairly good, 2 = poor, 1 = very poor). Usefulness includes compatibility, ease of use/flexibility, and perceived user understanding, and user reactions include user satisfaction and preferences.

A combined view of the different measures of usefulness and user reactions is presented in figure 7.7 for the tasks.

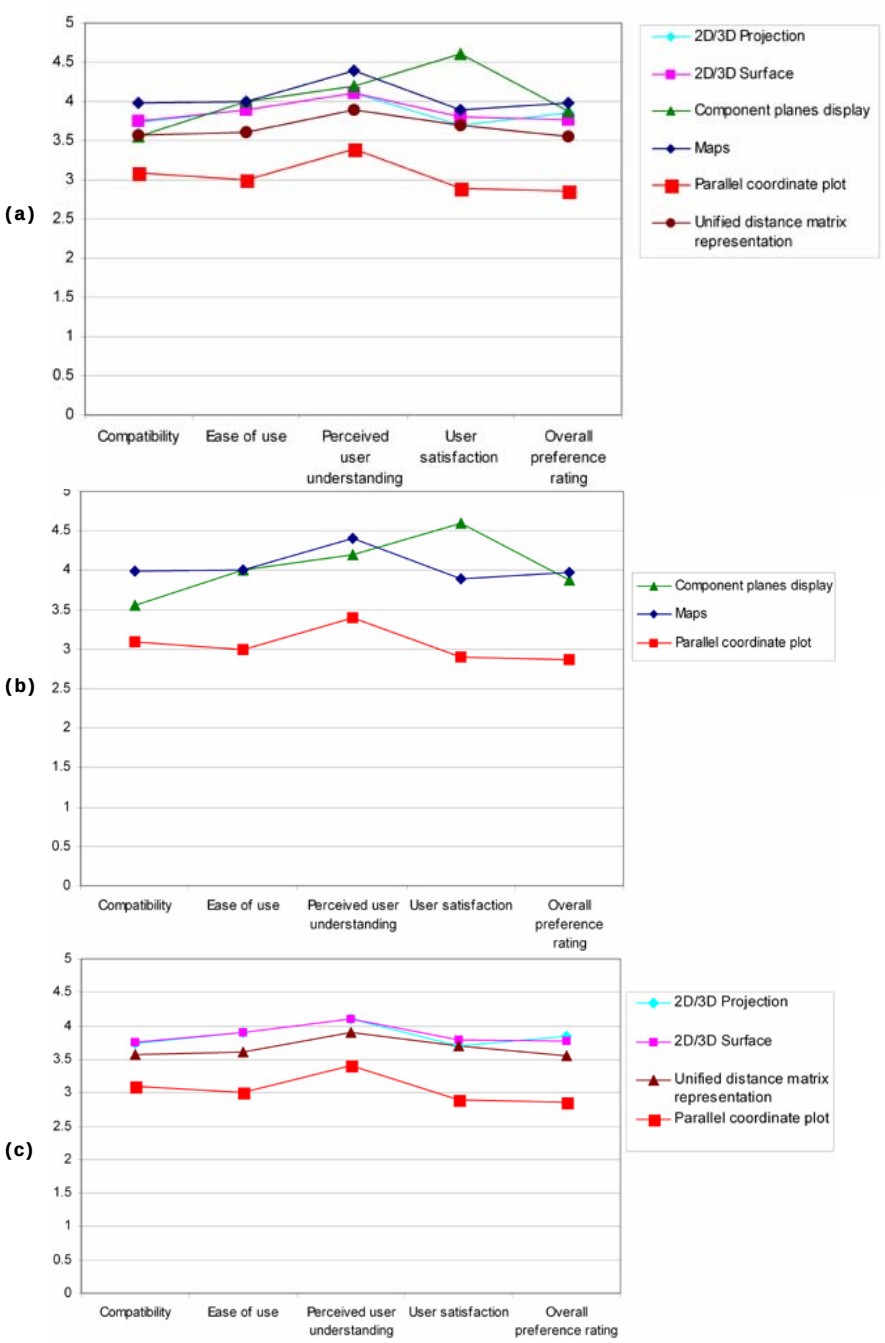


Figure 7.7. Overall ratings of the representations for all the different tasks combined: (a) shows all the representations for all the tasks; (b) shows tools used for detailed exploration tasks; and (c) shows tools used for visual grouping (clustering) tasks. The vertical axis represents the rating scale (5 = very good, 4 = good, 3 = fairly good, 2 = poor, 1 = very poor).

Figure 7.7b shows clear separation between the lines representing the parallel coordinate plot and those representing the map and the component planes. The parallel coordinate plot scored lower for all the usefulness and user reaction variables (compatibility, ease of use, perceived understanding, user satisfaction), and was rated much lower than the other representations. Figure 7.7b shows that the parallel coordinate plot scored poorly for user satisfaction (less than 3), ease of use (3), and was difficult to understand. The component plane display and the map were rated equally for ease of use and in the overall preference rating. As one could expect, the map was better rated for compatibility with users' expectations of the tasks. However, the users were more satisfied with the component plane display for their exploration results. Both the map and the component plane display were easily understood (4.2 and 4.4 on average on the five-point scale respectively).

For the clustering and visual grouping (categorization) tasks, the SOM-based representation tools (unified distance matrix representation, 2D/3D projection, 2D/3D surface) were better rated (between 3.5 and 4) for the different variables (compatibility, ease of use, perceived understanding, user satisfaction, and preference) than the parallel coordinate plot. The parallel coordinate plot scored much lower for these visual grouping tasks.

Detailed statistics on the compatibility, ease of use, user understanding, user satisfaction and user preference are presented in table 7.5.

To further analyze the differences in user rating for the representations, a paired sample test was conducted. Tables 7.6 and 7.7 show the significant differences between the representations for all the variables. T1 to T10 represent the task numbers corresponding the conceptual visualization goals (see table 7.1).

Table 7.5. Statistics for compatibility, ease of use, perceived user understanding, user satisfaction and the overall preference rating of the representations.

		Compatibility				Flexibility (ease of use)				Perceived user understanding				User satisfaction				User preference (overall rating)			
Task		Mean	Median	Std. deviation	Variance	Mean	Median	Std. deviation	Variance	Mean	Median	Std. deviation	Variance	Mean	Median	Std. deviation	Variance	Mean	Median	Std. deviation	Variance
T1	MAP	4.85	5	0.366	0.134	4.8	5	0.410	0.168	4.9	5	0.308	0.095	4.8	5	0.523	0.274	4.8	5	0.410	0.168
	PCP	3.75	4	1.118	1.250	3.75	4	1.164	1.355	3.85	4	1.226	1.503	3.65	4	1.040	1.082	3.4	3	1.142	1.305
	COMP	4.3	4	0.733	0.537	4.1	4	1.119	1.253	4.45	5	0.686	0.471	4.25	4	0.716	0.513	3.8	4	0.951	0.905
T2	MAP	3.25	3.5	1.118	1.250	3.3	3.5	1.129	1.274	3.65	4	1.137	1.292	3.3	3.5	1.261	1.589	3.3	3	1.302	1.695
	PCP	2.95	3	1.050	1.103	2.9	3	1.119	1.253	3.25	3.5	1.118	1.250	2.8	3	1.152	1.326	2.95	3	1.395	1.945
	COMP	3.75	4	1.164	1.355	3.8	4	1.105	1.221	3.85	4	1.089	1.187	3.6	3.5	1.273	1.621	3.65	4	1.182	1.397
T3	MAP	4.55	5	0.686	0.471	4.35	5	0.813	0.661	4.8	5	0.410	0.168	4.55	5	0.759	0.576	4.65	5	0.671	0.450
	PCP	2.55	2.5	1.099	1.208	2.4	2	1.046	1.095	3.15	3	1.226	1.503	2.3	2	1.218	1.484	2.2	2	0.951	0.905
	COMP	3.85	4	1.137	1.292	3.75	4	1.070	1.145	4	4	1.214	1.474	3.95	4	1.191	1.418	3.65	4	1.040	1.082
T4	UDM	3.45	4	1.234	1.524	3.6	4	1.314	1.726	3.85	4	1.040	1.082	3.7	4	1.081	1.168	3.7	4	1.261	1.589
	PROJ	3.65	4	0.875	0.766	3.75	4	0.910	0.829	3.9	4	0.968	0.937	3.65	4	0.988	0.976	3.75	4	0.851	0.724
	SURF	3.4	3	1.188	1.411	3.5	3.5	1.100	1.211	3.9	4	1.021	1.042	3.45	3	0.999	0.997	3.55	4	0.945	0.892
T5	UDM	3.7	4	0.923	0.853	3.65	4	1.040	1.082	3.95	4	0.887	0.787	3.7	4	0.865	0.747	3.4	3.5	1.231	1.516
	PROJ	3.8	4	1.105	1.221	3.95	4	1.146	1.313	4.2	4	0.951	0.905	3.8	4	0.951	0.905	3.95	4	0.887	0.787
	SURF	4.1	4	0.641	0.411	4.25	4	0.851	0.724	4.3	4	0.571	0.326	4.2	4	0.768	0.589	4	4	0.858	0.737
	PCP	2.95	3	1.191	1.418	2.75	3	1.333	1.776	3.25	3.5	1.446	2.092	2.6	2	1.314	1.726	2.526	2	1.124	1.263
T6	MAP	3.15	3.5	1.531	2.345	2.95	3	1.638	2.682	4.05	5	1.504	2.261	3	3	1.622	2.632	3.1	3	1.553	2.411
	PCP	2.9	3	1.294	1.674	2.9	3	1.210	1.463	3.2	3	1.281	1.642	2.6	3	1.231	1.516	2.85	3	1.309	1.713
	COMP	4.3	4	0.733	0.537	4.3	5	1.031	1.063	4.25	5	0.967	0.934	4.35	4	0.671	0.450	4.368	4	0.761	0.579
T7	MAP	4.4	5	0.940	0.884	4.4	5	0.883	0.779	4.65	5	0.587	0.345	4.35	5	1.089	1.187	4.45	5	0.887	0.787
	PCP	3.75	4	1.070	1.145	3.65	4	1.137	1.292	4	4	1.026	1.053	3.8	4	1.196	1.432	3.55	3.5	0.999	0.997
	COMP	3.85	4	0.988	0.976	3.95	4	0.999	0.997	4.05	4	1.050	1.103	3.95	4	1.050	1.103	3.579	4	1.170	1.368
T8	MAP	3.65	4	1.089	1.187	3.9	4	1.071	1.147	4.2	4	0.768	0.589	3.55	4	1.191	1.418	3.7	4	1.302	1.695
	PCP	2.9	3	1.294	1.674	2.85	3	0.988	0.976	3.45	3.5	1.191	1.418	2.75	3	1.164	1.355	3	3	1.214	1.474
	COMP	4.1	4	0.788	0.621	4	4	0.858	0.737	4.35	4.5	0.745	0.555	3.95	4	0.945	0.892	4.15	4	0.813	0.661
T9	MAP	3.95	4	1.146	1.313	3.95	4	0.999	0.997	4.3	4.5	0.801	0.642	3.9	4.5	1.373	1.884	4	4.5	1.298	1.684
	PCP	3	3	1.214	1.474	2.8	3	1.105	1.221	3.2	3	1.196	1.432	2.8	2	1.105	1.221	2.7	2.5	1.174	1.379
	COMP	4.35	4.5	0.745	0.555	4.15	4	0.813	0.661	4.6	5	0.681	0.463	4.1	4	0.912	0.832	4	4	1.026	1.053
T10	MAP	4.15	4	0.988	0.976	4.05	4	0.999	0.997	4.5	5	0.688	0.474	4	4	1.257	1.579	3.85	4	1.226	1.503
	PCP	3.05	3	1.099	1.208	3.21	3	1.134	1.287	3.6	4	1.231	1.516	2.85	3	1.182	1.397	2.7	2.5	1.174	1.379
	COMP	4.3	5	0.865	0.747	4.1	4	0.968	0.937	4.4	4.5	0.681	0.463	4.1	4	0.852	0.726	4.25	4	0.851	0.724

Table 7.6. Statistical significant differences between the representations for compatibility, ease of use measure, and perceived user understanding.

Compatibility									
		Paired differences					t	df	P-value
		Mean	Std.	Std. error mean	95% confidence interval of the difference				
					Lower	Upper			
T1	MAP - PCP	1.1	1.210	0.270	0.534	1.666	4.067	19	0.001
	MAP - COMP	0.55	0.605	0.135	0.267	0.833	4.067	19	0.001
T2	PCP - COMP	-0.8	1.361	0.304	-1.437	-0.163	-2.629	19	0.017
T3	MAP - PCP	2	0.858	0.192	1.598	2.402	10.420	19	0.000
	MAP - COMP	0.7	1.218	0.272	0.130	1.270	2.570	19	0.019
	PCP - COMP	-1.3	1.658	0.371	-2.076	-0.524	-3.508	19	0.002
T5	UDM - SURF	-0.4	0.754	0.169	-0.753	-0.047	-2.373	19	0.028
	UDM - PCP	0.75	1.209	0.270	0.184	1.316	2.775	19	0.012
	PROJ - PCP	0.85	1.531	0.342	0.133	1.567	2.482	19	0.023
	SURF - PCP	1.15	1.137	0.254	0.618	1.682	4.524	19	0.000
T6	MAP - COMP	-1.15	1.631	0.365	-1.913	-0.387	-3.153	19	0.005
	PCP - COMP	-1.4	1.392	0.311	-2.051	-0.749	-4.499	19	0.000
T7	MAP - PCP	0.65	1.309	0.293	0.037	1.263	2.221	19	0.039
	MAP - COMP	0.55	1.099	0.246	0.036	1.064	2.238	19	0.037
T8	MAP - COMP	-0.45	0.887	0.198	-0.865	-0.035	-2.269	19	0.035
	PCP - COMP	-1.2	1.436	0.321	-1.872	-0.528	-3.736	19	0.001
T9	MAP - PCP	0.95	1.572	0.352	0.214	1.686	2.703	19	0.014
	PCP - COMP	-1.35	1.348	0.302	-1.981	-0.719	-4.477	19	0.000
T10	MAP - PCP	1.1	1.586	0.355	0.358	1.842	3.101	19	0.006
	PCP -COMP	-1.25	1.482	0.331	-1.944	-0.556	-3.771	19	0.001
Ease of use									
		Paired differences					t	df	P-value
		Mean	Std.	Std. error mean	95% confidence interval of the difference				
					Lower	Upper			
T1	MAP - PCP	1.05	0.999	0.223	0.583	1.517	4.702	19	0.000
	MAP - COMP	0.7	0.801	0.179	0.325	1.075	3.907	19	0.001
T2	PCP - COMP	-0.9	1.553	0.347	-1.627	-0.173	-2.592	19	0.018
T3	MAP - PCP	1.95	1.050	0.235	1.459	2.441	8.305	19	0.000
	MAP - COMP	0.6	1.188	0.266	0.044	1.156	2.259	19	0.036
	PCP - COMP	-1.35	1.182	0.264	-1.903	-0.797	-5.107	19	0.000
T5	UDM - SURF	-0.6	0.754	0.169	-0.953	-0.247	-3.559	19	0.002
	UDM - PCP	0.9	1.410	0.315	0.240	1.560	2.854	19	0.010
	PROJ - PCP	1.2	1.673	0.374	0.417	1.983	3.207	19	0.005
	SURF - PCP	1.5	1.277	0.286	0.902	2.098	5.252	19	0.000
T6	MAP - COMP	-1.35	1.663	0.372	-2.128	-0.572	-3.630	19	0.002
	PCP - COMP	-1.4	1.818	0.407	-2.251	-0.549	-3.444	19	0.003
T7	MAP - PCP	0.75	1.446	0.323	0.073	1.427	2.319	19	0.032
T8	MAP - PCP	1.05	1.605	0.359	0.299	1.801	2.926	19	0.009
	PCP - COMP	-1.15	1.348	0.302	-1.781	-0.519	-3.814	19	0.001
T9	MAP - PCP	1.15	1.599	0.357	0.402	1.898	3.217	19	0.005
	PCP - COMP	-1.35	1.387	0.310	-1.999	-0.701	-4.353	19	0.000
T10	MAP - PCP	0.789	1.548	0.355	0.043	1.536	2.222	18	0.039
	PCP - COMP	-0.84	1.675	0.384	-1.650	-0.035	-2.191	18	0.042
Perceived user understanding									
		Paired differences					t	df	P-value
		Mean	Std.	Std. error mean	95% confidence interval of the difference				
					Lower	Upper			
T1	MAP - PCP	1.05	1.234	0.276	0.472	1.628	3.804	19	0.001
	MAP - COMP	0.45	0.510	0.114	0.211	0.689	3.943	19	0.001
T3	MAP - PCP	1.65	1.089	0.244	1.140	2.160	6.773	19	0.000
	MAP - COMP	0.8	1.281	0.287	0.200	1.400	2.792	19	0.012
	PCP - COMP	-0.85	1.387	0.310	-1.499	-0.201	-2.741	19	0.013
T5	UDM - SURF	-0.35	0.587	0.131	-0.625	-0.075	-2.666	19	0.015
	PROJ - PCP	0.95	1.761	0.394	0.126	1.774	2.412	19	0.026
	SURF - PCP	1.05	1.395	0.312	0.397	1.703	3.367	19	0.003
T6	PCP - COMP	-1.05	1.791	0.400	-1.888	-0.212	-2.622	19	0.017
T7	MAP - PCP	0.65	1.226	0.274	0.076	1.224	2.371	19	0.028
	MAP - COMP	0.6	1.231	0.275	0.024	1.176	2.179	19	0.042
T8	MAP - PCP	0.75	1.410	0.315	0.090	1.410	2.380	19	0.028
	PCP - COMP	-0.9	1.252	0.280	-1.486	-0.314	-3.214	19	0.005
T9	MAP - PCP	1.1	1.071	0.240	0.599	1.601	4.593	19	0.000
	PCP - COMP	-1.4	1.231	0.275	-1.976	-0.824	-5.085	19	0.000
T10	MAP - 0PCP	0.9	1.447	0.324	0.223	1.577	2.781	19	0.012
	PCP - COMP	-0.8	1.322	0.296	-1.419	-0.181	-2.707	19	0.014

Table 7.7. Statistical significant differences between the representations for user satisfaction and overall preference rating.

User satisfaction

		Paired differences					t	df	P-value
		Mean	Std.	Std. error mean	95% confidence interval of the difference				
					Lower	Upper			
T1	MAP - PCP	1.15	1.089	0.244	0.640	1.660	4.721	19	0.000
	MAP - COMP	0.55	0.605	0.135	0.267	0.833	4.067	19	0.001
	PCP - COMP	-0.6	1.231	0.275	-1.176	-0.024	-2.179	19	0.042
T3	MAP - PCP	2.25	1.118	0.250	1.727	2.773	9.000	19	0.000
	PCP - COMP	-1.65	1.694	0.379	-2.443	-0.857	-4.355	19	0.000
T5	UDM - SURF	-0.5	0.761	0.170	-0.856	-0.144	-2.939	19	0.008
	UDM - PCP	1.1	1.410	0.315	0.440	1.760	3.488	19	0.002
	PROJ - PCP	1.2	1.609	0.360	0.447	1.953	3.335	19	0.003
	SURF - PCP	1.6	1.314	0.294	0.985	2.215	5.446	19	0.000
T6	MAP - COMP	-1.35	1.843	0.412	-2.213	-0.487	-3.275	19	0.004
	PCP - COMP	-1.75	1.482	0.331	-2.444	-1.056	-5.280	19	0.000
T8	PCP - COMP	-1.2	1.473	0.329	-1.889	-0.511	-3.644	19	0.002
T9	MAP - PCP	1.1	1.744	0.390	0.284	1.916	2.820	19	0.011
	PCP - COMP	-1.3	1.380	0.309	-1.946	-0.654	-4.212	19	0.000
T10	MAP - PCP	1.15	1.599	0.357	0.402	1.898	3.217	19	0.005
	PCP - COMP	-1.25	1.517	0.339	-1.960	-0.540	-3.684	19	0.002

Overall user preference rating

		Paired differences					t	df	P-value
		Mean	Std.	Std. error mean	95% confidence interval of the difference				
					Lower	Upper			
T1	MAP - PCP	1.4	1.314	0.294	0.785	2.015	4.765	19	0.000
	MAP - COMP	1	0.725	0.162	0.660	1.340	6.164	19	0.000
T3	MAP - PCP	2.45	0.759	0.170	2.095	2.805	14.433	19	0.000
	MAP - COMP	1	1.414	0.316	0.338	1.662	3.162	19	0.005
	PCP - COMP	-1.45	1.538	0.344	-2.170	-0.730	-4.216	19	0.000
T5	UDM - PCP	0.895	1.560	0.358	0.143	1.647	2.500	18	0.022
	PROJ - PCP	1.421	1.539	0.353	0.679	2.163	4.025	18	0.001
	SURF - PCP	1.421	1.261	0.289	0.813	2.029	4.911	18	0.000
T6	MAP - COMP	-1.368	1.707	0.392	-2.191	-0.546	-3.495	18	0.003
	PCP - COMP	-1.421	1.677	0.385	-2.229	-0.613	-3.693	18	0.002
T7	MAP - PCP	0.9	1.483	0.332	0.206	1.594	2.714	19	0.014
	MAP - COMP	0.842	1.463	0.336	0.137	1.547	2.509	18	0.022
T8	PCP - COMP	-1.15	1.663	0.372	-1.928	-0.372	-3.092	19	0.006
T9	MAP - PCP	1.3	1.949	0.436	0.388	2.212	2.982	19	0.008
	PCP - COMP	-1.3	1.218	0.272	-1.870	-0.730	-4.772	19	0.000
T10	MAP - PCP	1.15	1.755	0.393	0.328	1.972	2.930	19	0.009
	PCP - COMP	-1.55	1.468	0.328	-2.237	-0.863	-4.722	19	0.000

The analysis of the statistics in tables 7.5, 7.6 and 7.7 shows some significant differences between the representations for the different tasks with regard to the measures of usefulness (compatibility with the user's expectations for the task, ease of use, the user's understanding of the tool for the task), and user reactions (user satisfaction, and the overall preference rating of the tools).

1. Compatibility with the user's expectations for the different tasks

For compatibility with the user's expectations of the tool for the tasks, the map was found more suitable (mean = 4.85 and median = 5 on the five-point scale) for the tasks *'locate', ''distinguish' and 'rank'*. The component plane display was found more appropriate for the tasks '*identify*', '*distribution*', '*compare*', '*associate*' and '*correlate*'. The parallel coordinate plot was rated generally poor (2 on the five-point scale) or fairly good (3 on the five-point scale) for all the tasks. The best ratings of the parallel coordinate plot were for the tasks '*rank*' and '*locate*', where the mean score = 3.75 and the median = 4 (same result for both tasks).

These results for compatibility confirm the performance analysis presented in section 7.4.1. for correctness of response and time taken. Some graphs of the compatibility rating for all the tasks are presented in figure 7.8.

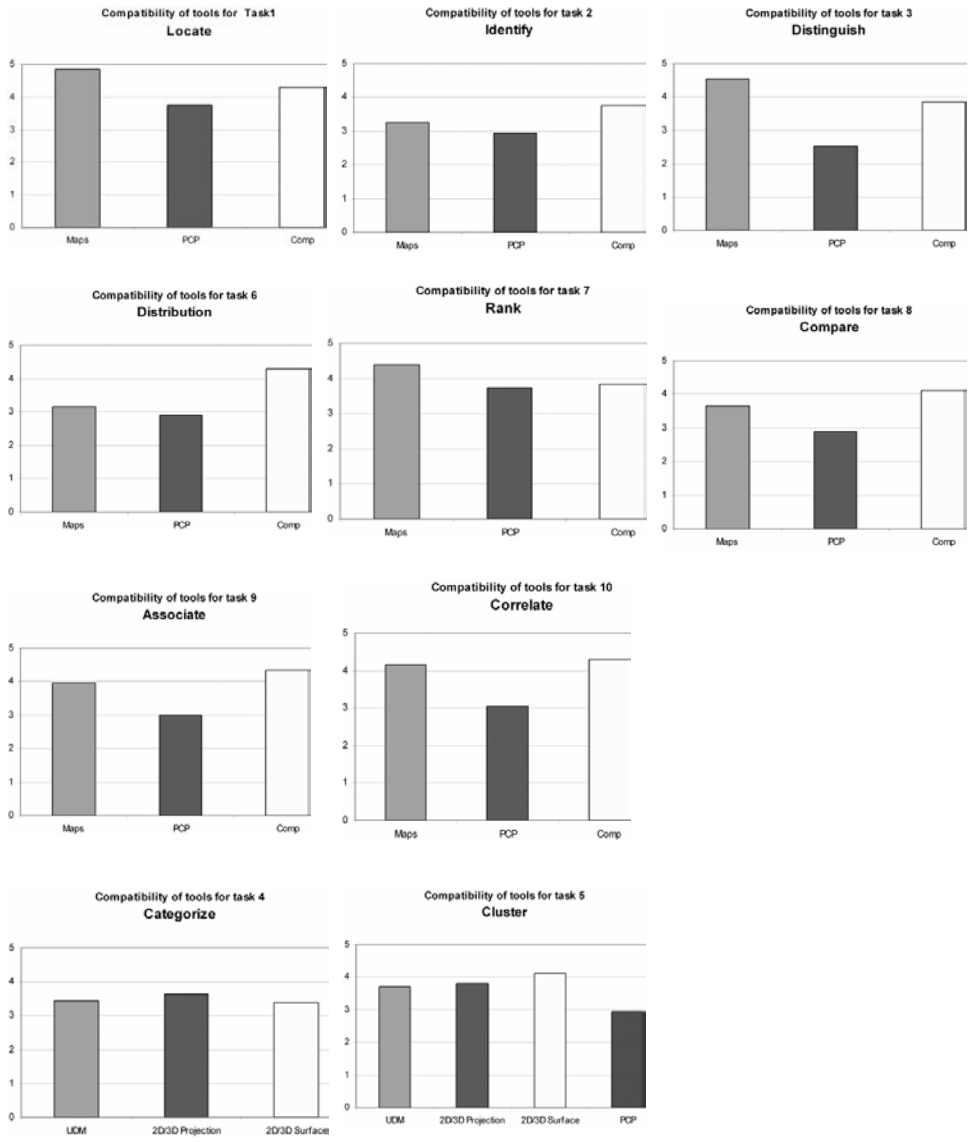


Figure 7.8. Compatibility rating of the representations for the each visualization task. Scale: 5 = very good, 4 = good, 3 = fairly good, 2 = poor, 1 = very poor. The tasks have been organized in two groups: detail exploration tasks (tasks number 1, 2, 3, 6, 7, 8, 9, 10), and visual grouping tasks (tasks number 4 and 5). PCP= Parallel coordinate plot, Comp=SOM component plane display, UDM=Unified distance matrix.

For the task *'locate'*, the results show a significant advantage for the map compared with the parallel coordinate plot (p=0.001) and the component plane display (p=0.001).

For the task *'identify'*, no significant difference was found between the map and the component plane display. Both were rated above fairly good on average (mean = 3.25 and median = 3.5 for the map, and mean = 3.75 and median = 4 for the component plane display) and better than the parallel coordinate plot. For this task, a significant difference was found between the parallel coordinate plot and the component planes (p=0.017). The map and the parallel coordinate plot did not show any significant difference.

For the task *'distinguish'*, the map was rated best suitable representation (mean = 4.55, median = 5) with a significant difference compared with the component plane display (p=0.019) and the parallel coordinate plot (p=0.000). The component plane display was also found more compatible for this task than the parallel coordinate plot (p=0.002).

For the task *'distribution'*, the component plane display was rated best suitable representation (mean = 4.3, median = 4) with a significant difference compared with the map (p=0.005) and parallel coordinate plot (p=0.000). The map and the parallel coordinate plot were rated fairly good (mean = 3.15, median = 3.5 for the map, and mean = 2.9, median = 3 for the parallel coordinate plot).

For the task *'rank'*, the map was found very suitable (mean = 4.4, median = 5) with a significant difference compared with the component plane display (p=0.037) and the parallel coordinate plot (p=0.039). Both the component plane display and the parallel coordinate plot were also rated good (mean = 3.85, median = 4 for the component plane display, and mean = 3.75, median = 4 for the parallel coordinate plot).

The component plane display was found more suitable for the task *'compare'* (mean = 4.1, median = 4), with a significant difference compared with the map (p=0.035) and the parallel coordinate plot (p=0.001). The parallel coordinate plot was found poor for this task (mean = 2.9, median = 3). The map was not found particularly good for comparing several attributes (mean = 3.65, median = 4).

The map and the component plane display were found equally good (no significant difference) for the tasks *'associate'* and *'correlate'*, with the component plane display slightly better (4.35 and 4.3 of mean score for the component plane display, and 3.95 and 4.15 mean score for *'associate'* and *'correlate'* respectively). The parallel coordinate plot was found fairly good for these two tasks. The map was significantly more suitable than the parallel coordinate plot for the task *'associate'* (p=0.014) and the task *'correlate'* (p=0.006). The component plane

display was significantly more suitable than the parallel coordinate plot for the task '*associate'* (p=0.000) and for the task '*correlate'* (p=0.001).

For the task '*cluster'* and '*categorize'*, the tools used (unified distance matrix, 2D/3D projection, 2D/3D surface and parallel coordinate plot) were generally suitable for the tasks. No difference was found between the SOM-based clustering tools for the task '*categorize'*. A significant difference was found between each of the SOM-based tools and the parallel coordinate plot for the task '*cluster'* (p=0.012 with the unified distance matrix; p=0.023 with the 2D/3D projection; p=0.000 with the 2D/3D surface). No significant difference was found between the unified distance matrix and the 2D/3D projection. However, a significant difference was found between the unified distance matrix and the 2D/3D surface plot (p=0.028), with the 2D/3D surface more compatible for the users (mean = 4.1, median = 4).

2. Flexibility / ease of use

As with compatibility, the map was found easier for the tasks '*locate'*, '*distinguish'* and '*rank'*. The component plane display was found easier to use for the tasks '*identify'* and '*distribution'*. The parallel coordinate plot was generally found difficult to use, especially for the tasks '*distinguish'*, '*associate'*, and '*compare'*, but less difficult to use for the tasks '*rank'* and '*locate'* (see figure 7.9).

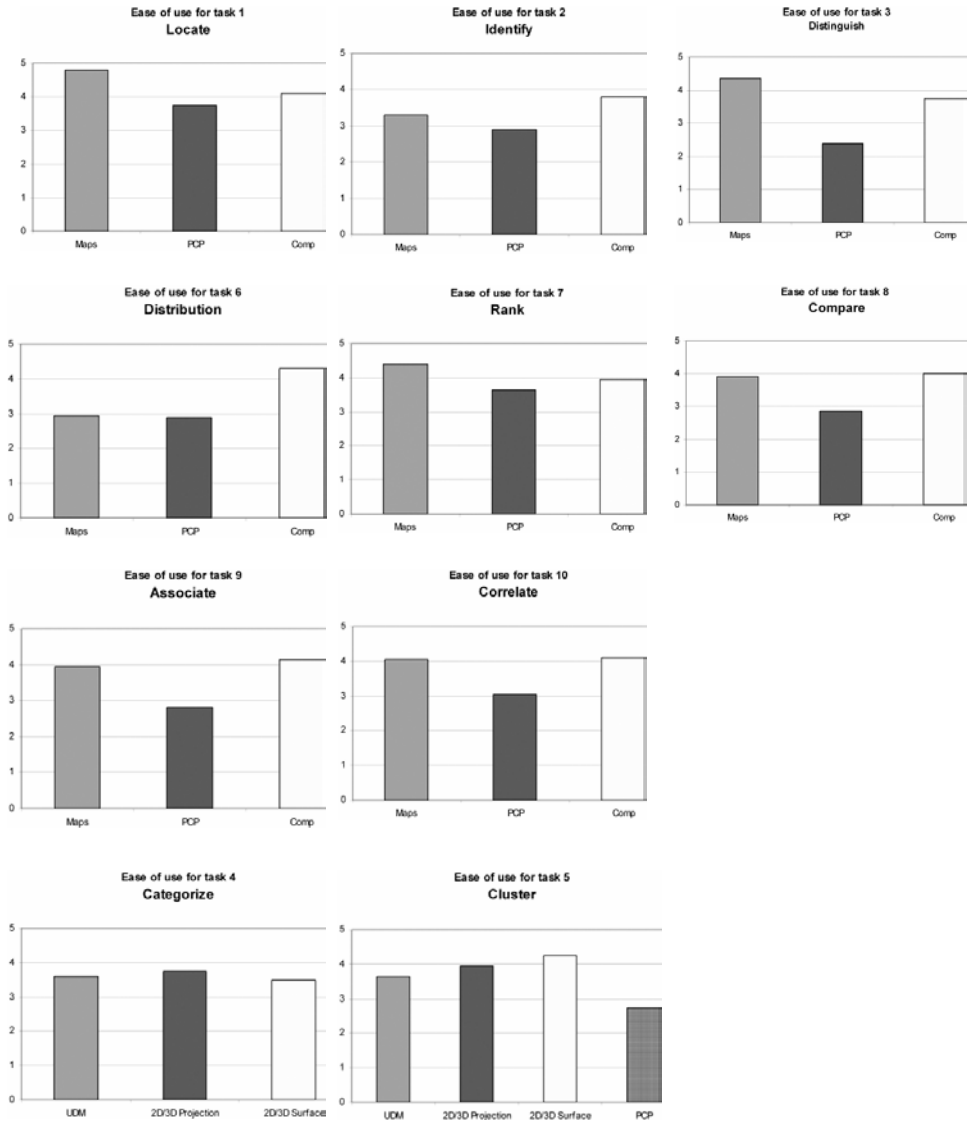


Figure 7.9. Rating of ease of use for the representations for the each visualization task. Scale: 5 = very good, 4 = good, 3 = fairly good, 2 = poor, 1 = very poor. The tasks have been organized in two groups: detail exploration tasks (tasks number 1, 2, 3, 6, 7, 8, 9, 10), and visual grouping tasks (tasks number 4 and 5). PCP= Parallel coordinate plot, Comp=SOM component plane display, UDM=Unified distance matrix.

For the task '*locate*', the map was found very easy to use (mean = 4.8, median = 5), with a significant difference compared with the component plane display ($p=0.001$) and the parallel coordinate plot ($p=0.000$). The component

plane display was also found easy to use (mean = 4.1, median = 4), and significantly easier than the parallel coordinate plot (p=0.018).

All the tools were found relatively easy to use for the task '*identify'*. The component plane display was found easier for the task, with a significant difference compared with the parallel coordinate plot (p=0.018). No significant difference was found between the map and the parallel coordinate plot, or between the map and the component plane display (see figure 7.9).

The map was easier for the task '*distinguish'* (mean = 4.35, median = 5) compared with the component plane display (mean = 3.75, median = 4), and the parallel coordinate plot (mean = 2.4, median = 2). The parallel coordinate plot was found difficult to use for this task. The analysis of the mean scores shows a significant difference between the map and the component plane display (p=0.036), between the map and the parallel coordinate plot (p=0.000), and between the component plane display and the parallel coordinate plot (p=0.000).

The component plane display was found easier for the task '*distribution'* (mean = 4.3, median = 5), compared with the map (mean = 2.93, median = 3) and the parallel coordinate plot (mean = 2.9, median = 3). This result shows a significant difference between the component plane display and the map (p=0.002) and the parallel coordinate plot (p=0.003). The parallel coordinate plot was found relatively difficult.

For the task '*rank'*, the map was found much easier (mean = 4.4, median = 5), with a significant difference compared with the parallel coordinate plot (p=0.032). No significant difference was found between the map and the component plane display, which was also found easy to use (mean = 3.95, median = 4).

Both the map and the component plane display were found easy to use for the task '*compare'* (mean = 3.9, median = 4, and mean = 4 and median = 4 for the map and component plane display respectively). The parallel coordinate plot was found difficult (mean = 2.85, median = 3) and shows a significant difference with the map (p=0.009) and with the component plane display (p=0.001).

The tasks '*associate'* and '*correlate'* were rated similarly for the representations (with no statistical difference). The component plane display and the map were found easy for both tasks (median = 4 for both the map and the component plane display). The parallel coordinate plot was found fairly difficult (median = 3 for both tasks). Significant difference was found between the map and the parallel coordinate plot (p=0.005 and p=0.039 for the tasks '*associate'* and '*correlate'* respectively). The component plane display also shows a significant difference to the parallel coordinate plot (p=0.000 and p=0.042 respectively) for the tasks '*associate'* and '*correlate'*.

For the tasks '*cluster'* and '*categorize'*, the tools used (unified distance matrix, 2D/3D projection, 2D/3D surface and parallel coordinate plot) were generally fairly easy to use. No difference was found between the SOM-based clustering tools for the task '*categorize'* with regard to ease of use. A significant difference was found between each of the SOM-based tools and the parallel coordinate plot for the task '*cluster'* ($p=0.010$ with the unified distance matrix, $p=0.005$ with the 2D/3D projection, $p=0.000$ with the 2D/3D surface). No significant difference was found between the unified distance matrix and the 2D/3D projection. However, a significant difference was found between the unified distance matrix and the 2D/3D surface plot ($p=0.002$), with the 2D/3D surface easier (mean = 4.24, median = 4).

3. Perceived user understanding of the representations used

The map and the component plane display were generally well understood for all the tasks (see figure 7.10). The parallel coordinate plot was not well understood for some of tasks such as '*compare'*, '*associate'*, '*distinguish'*, '*distribution'* and '*correlate',* but relatively well understood for the task '*rank'*.

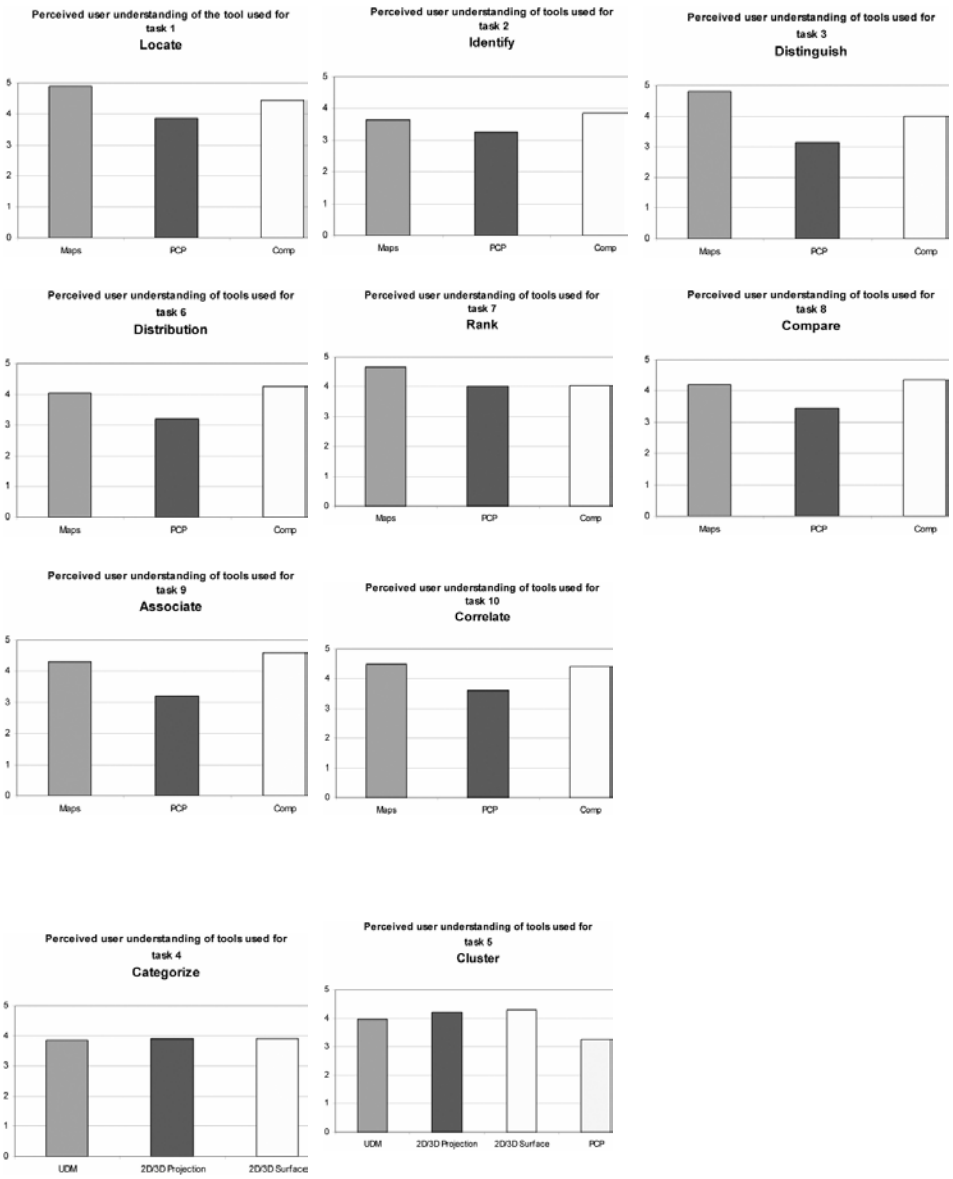


Figure 7.10. Rating of perceived user understanding of the representations for the different visualization tasks. Scale: 5 = very good, 4 = good, 3 = fairly good, 2 = poor, 1 = very poor. The tasks have been organized in two groups: detail exploration tasks (tasks number 1, 2, 3, 6, 7, 8, 9, 10), and visual grouping tasks (tasks number 4 and 5). PCP= Parallel coordinate plot, Comp=SOM component plane display, UDM=Unified distance matrix.

For the task '*locate'*, the map was better understood (mean = 4.9, median = 5) with a significant difference compared with the component plane display (p=0.001) and the parallel coordinate plot (p=0.001). The component plane display and the parallel coordinate plot were also well understood for this task (median = 4 and 5 respectively for parallel coordinate plot and the component plane display).

For the task '*identify'*, all the tools (map, parallel coordinate plot, component plane display) were well understood. No significant difference was found between them.

The map was found much easier for the task '*distinguish'* (mean = 4.8, median = 5), with a significant difference to the component plane display (p=0.012) and the parallel coordinate plot (p=0.000). The component plane display was also well understood (mean = 4, median = 4) compared with the parallel coordinate plot (median = 3), with a significant difference (p=0.013).

For the task '*distribution'*, the map and the component plane display were similarly well understood (with no statistical difference) compared with the parallel coordinate plot. A significant difference was found between the component plane display and the parallel coordinate plot (p=0.017).

All the tools were well understood for the task '*rank'*, with the map significantly better than the component plane display (p=0.017) and the parallel coordinate plot (p=0.028).

For the tasks '*compare'*, '*associate'* and '*correlate'*, the component planes and the map were equally well understood and better than the parallel coordinate plot. The result shows a significant difference between both the map and the component plane display compared with the parallel coordinate plot for each of the three tasks. The map shows a significant difference to the parallel coordinate plot (p=0.028 for the task '*compare'*, p=0.000 for the task '*associate'*, and p=0.012 for the task '*correlate'*). The component plane display shows a significant difference to the parallel coordinate plot (p=0.005 for the task '*compare'*, p=0.000 for the task '*associate'*, and p=0.014 for the task '*correlate'*).

For the tasks '*cluster'* and '*categorize'*, the tools used (unified distance matrix, 2D/3D projection, 2D/3D surface and parallel coordinate plot) were generally understood. No difference was found between the SOM-based clustering tools for the task '*categorize'* with regard to perceived user understanding. A significant difference was found comparing the 2D/3D projection with the parallel coordinate plot (p=0.026), and comapring the 2D/3D surface with the parallel coordinate plot (p=0.003). No significant difference was found between the unified distance matrix and the 2D/3D projection. A significant difference was found between the unified distance matrix and the 2D/3D surface plot (p=0.015), with the 2D/3D

surface better understood for the task '*cluster'* (mean = 4.3, median = 4) than the unified distance matrix.

4. User satisfaction

Figure 7.11 shows the rating of user satisfaction for the different tasks and representations used. In general users were satisfied with the component plane display and the map. The parallel coordinate plot was not satisfactory for the tasks '*distinguish'*, '*associate'*, '*correlate'* and '*distribution'*.

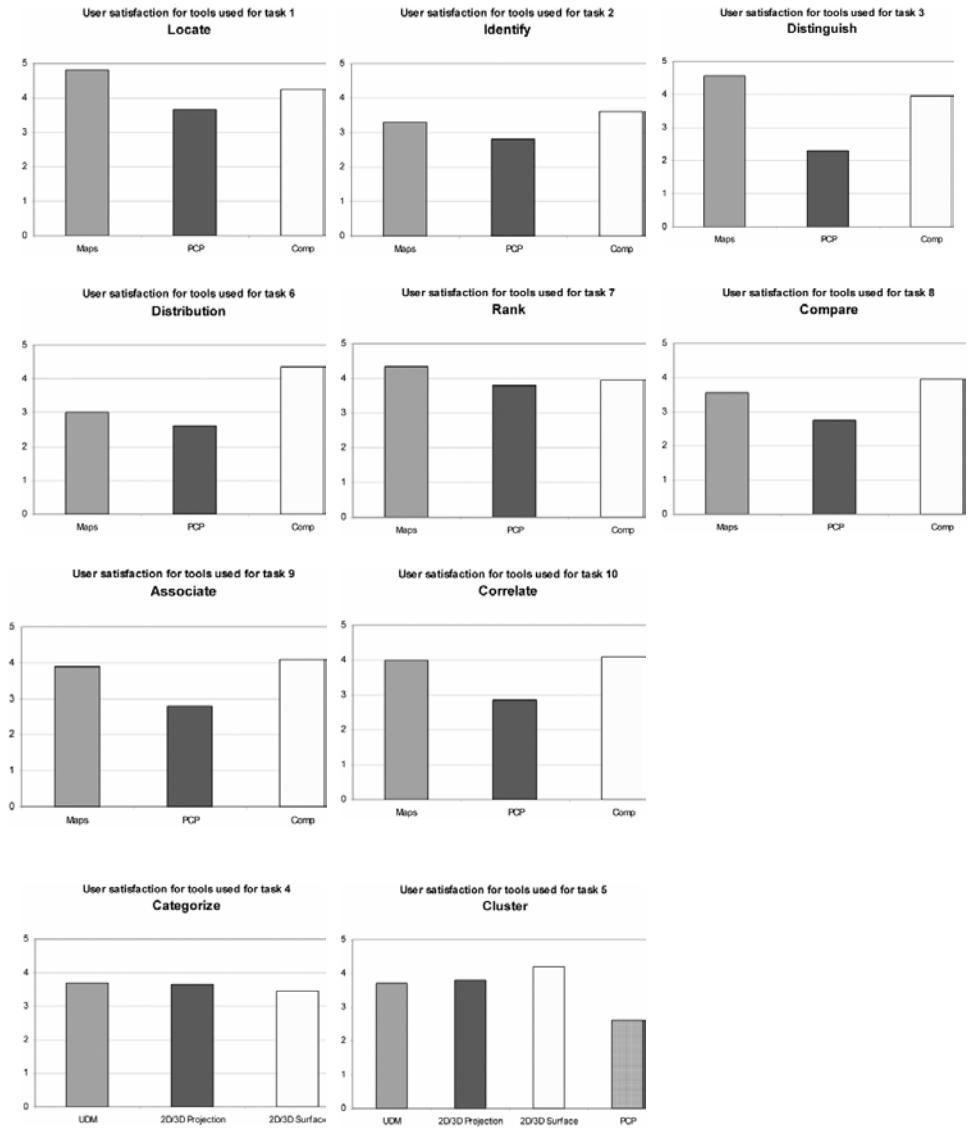


Figure 7.11. Rating of user satisfaction for the representations for the different visualization tasks. Scale: 5 = very good, 4 = good, 3 = fairly good, 2 = poor, 1 = very poor. The tasks have been organized in two groups: detail exploration tasks (tasks number 1, 2, 3, 6, 7, 8, 9, 10), and visual grouping tasks (tasks number 4 and 5). PCP= Parallel coordinate plot, Comp=SOM component plane display, UDM=Unified distance matrix.

For the task *'locate'*, users were generally satisfied with the map (mean = 4.8, median = 5), the component plane display (mean = 4.25, median = 4), and parallel coordinate plot (mean = 3.65, median = 4). These results show some

significant differences between the map and the parallel coordinate plot (p=0.000), the map and the component plane display (p=0.001), and between the component plane display and the parallel coordinate plot (p=0.042).

No significant difference between the tools was found for the task '*identify'*. All the representations were fairly satisfactory for this task.

The map and the component plane display were fairly satisfactory for the task '*distinguish'* (mean = 3.3 and 3.6 respectively for the map and the component plane display), with no significant difference in the mean score values. The parallel coordinate plot was found not satisfactory (mean = 2.8). This result shows a significant difference between the map and the parallel coordinate plot (p=0.000), and between the component plane display and the parallel coordinate plot (p=0.000).

For the task '*distribution'*, the component plane display was found very satisfactory (mean = 4.35), with a significant difference compared with the map (p=0.004) and the parallel coordinate plot (p=0.000). The map and the parallel coordinate plot were found not satisfactory (mean = 3 and 2.6 respectively).

No difference was found between the tools for the task '*rank'*. They were all found satisfactory.

For the task '*compare'*, the map and the component plane display were found satisfactory (mean = 3.95 for the component plane display and 3.55 for the map). The parallel coordinate plot was rated not satisfactory (mean = 2.75). A significant difference was found between the component plane display and the parallel coordinate plot (p=0.002).

The representations were rated similarly for the two tasks '*associate'* and '*correlate'* with regard to user satisfaction (see figure 7.11). The component planes and the map were equally rated for user satisfaction for these two tasks as satisfactory (mean = 3.9 and 4 for the map, and 4.1 and 4.1 for the component plane display respectively for '*associate'* and '*correlate'*). The mean scores for the parallel coordinate plot are low (2.8 and 2.85 for '*associate'* and '*correlate'* respectively). This shows a significant difference between the map and the parallel coordinate plot (p=0.011 and 0.005 respectively for '*associate'* and '*correlate'*), and between the component plane display and the parallel coordinate plot (p=0.000 and 0.002 respectively for '*associate'* and '*correlate'*).

For the tasks '*cluster'* and '*categorize'*, the tools used (unified distance matrix, 2D/3D projection, 2D/3D surface and parallel coordinate plot) were generally more satisfactory compared with the parallel coordinate plot for the task '*cluster'*. No difference was found between the SOM-based clustering tools for the task '*categorize'* with regard to user satisfaction. A significant difference was found between the SOM-based clustering tools and the parallel coordinate plot: with the

2D/3D projection (p=0.003), the 2D/3D surface (p=0.000), and the unified distance matrix (p=0.002). A significant difference was found between the unified distance matrix and the 2D/3D surface plot (p=0.008), with the 2D/3D surface being found more satisfactory for the task '*cluster'* (mean = 4.2, median = 4) than the unified distance matrix.

5. User preference rating

The overall preference rating of the tools for the different tasks revealed that the map was preferred for the tasks '*locate'*, '*distinguish'* and '*rank'*. The component plane display was preferred for the tasks '*identify'*, '*distribution'*, '*compare'* and '*correlate'*. The map and the component plane display were generally equally rated with regard to preference for the task '*associate'*. The parallel coordinate plot was generally not preferred for the different tasks (see figure 7.12).

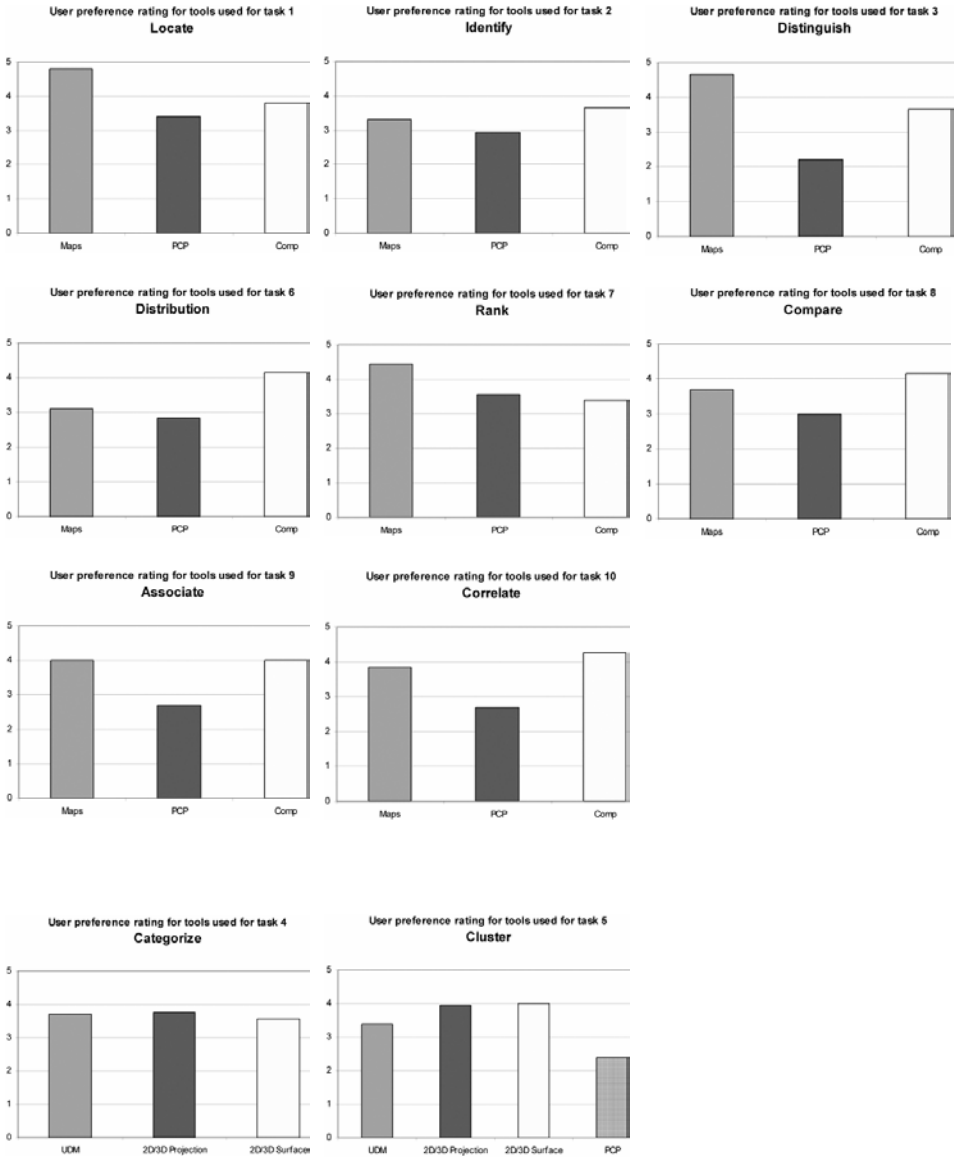


Figure 7.12. User preference rating of the representations for the different visualization tasks. Scale: 5 = very good, 4 = good, 3 = fairly good, 2 = poor, 1 = very poor. The tasks have been organized in two groups: detail exploration tasks (tasks number 1, 2, 3, 6, 7, 8, 9, 10), and visual grouping tasks (tasks number 4 and 5). PCP= Parallel coordinate plot, Comp=SOM component plane display, UDM=Unified distance matrix.

For the task '*locate'*, the map was preferred (mean = 4.8, median = 5) with a significant difference compared with the component plane display (mean = 3.8, median = 4) and the parallel coordinate plot (mean = 3.4, median = 3). The difference between the mean scores for the map compared with both the component plane display and the parallel coordinate plot is significant (p=0.000).

All the three tools (map, component plane display and the parallel coordinate plot) show no difference for the task '*identify'*.

The map was also preferred for the task '*distinguish'* (mean = 4.65, median = 5) compared with the component plane display (mean = 3.65, median = 4) and the parallel coordinate plot (mean = 2.2, median = 2). A significant difference was found between the map and the component plane display (p=0.005), and the parallel coordinate plot (p=0.000). The component plane display also shows a significant difference to the parallel coordinate plot in terms of preference (p=0.000).

The component plane display was preferred for the task '*distribution'* (mean = 4.37, median = 4), compared with the map (mean = 3.1, median = 3) and the parallel coordinate plot (mean = 2.85, median = 3). This result shows a significant difference between the component plane display, the map (p=0.003) and the parallel coordinate plot (p=0.002).

The map was preferred for the task '*rank'* (mean = 4.45, median = 5) compared with the component plane display (mean = 3.58, median = 4) and the parallel coordinate plot (mean = 3.55, median = 3.5). This shows a significant difference between the map and the component plane display (p=0.022), and the parallel coordinate plot (p=0.014).

The component plane display was generally preferred for the task '*compare'* (mean = 4.15, median = 4), compared with the map (mean = 3.7, median = 4), and the parallel coordinate plot (mean = 3, median = 3). A significant difference was found between the component plane display and the parallel coordinate plot (p=0.006). The difference between the map and the component plane display is not significant.

No difference in preference was found between the map and the component plane display for the task '*associate'* (mean = 4 for both). Both were preferred to the parallel coordinate plot, which shows a mean score of 2.7 and a median of 2.5. This result shows a significant difference between the map and the parallel coordinate plot (p=0.008) and between the component plane display and the parallel coordinate plot (p=0.009).

For the task '*correlate'*, the component plane display shows a mean score higher than the map (4.25 for the component plane display and 3.85 for the map). This result shows no significant difference. A significant difference was found between

the map and the parallel coordinate plot (p=0.009) and between the component plane display and the parallel coordinate plot (p=0.000).

7.5. Discussions

The analysis of the test results presented in the previous section reveal some important differences between the SOM-based representations, the map and the parallel coordinate plot according to the taxonomy of visualization tasks used for the evaluation. Each representation method by its inherent structure seems to emphasize particular attributes and support a particular set of visual tasks or inferences (Wehrend and Lewis 2000).

Maps were more effective for certain visual tasks such as locate and distinguish, but less effective for the tasks of comparison, correlation, and for relating many attributes (see figure 7.3). Although easy to use in general for all the test participants (it provides a good visual representation of the real world that the participants are used to), a major problem with the map was that it uses a single view for limited attributes, which is not appropriate for investigating many attributes for the dataset in a reasonable time. Many maps were needed to map more variables in order to complete some of the tasks. For visual comparison, the map was not as effective as the component plane display. It required more time for tasks that involve viewing relationships, since differences between classes geographically are not noticeable despite the colour scheme used for classification. One of the most important comments from the test participants was that tasks are difficult to complete with the maps if no prior hypothesis has been given.

Component plane displays were found to offer fast visual perception and were also found easier for finding relationships and understanding the patterns. This representation was especially effective and suitable for tasks involving visual composition (Zhou and Feiner 1998), such as associate, correlate, identify, and compare. Participants reported that the component plane display did not require much effort to view the patterns and to relate different attributes in a single view. Relationships between the attributes were found to be very apparent in component planes. This ability to permit immediate information extraction at a single glance with less attention is one of the measures of the quality of a visualization (Bertin 1983). The component plane display was less effective for the task of ranking among similar data items because of the clustering. More selection is needed for such tasks. Participants needed some guidance in using the component planes, but generally found the tool easier to use after a short introduction.

Parallel coordinate plots required the participants to keep track of a lot of information before they could summarize answers for the tasks. This is an important issue in visual encoding and perception (Cleveland and McGill 1984;

Cleveland and McGill 1986), key elements in knowledge construction using visual representations. This difficultly in keeping track of the information perceived makes the parallel coordinate plot difficult for the test participants to understand (see figure 7.10). Some participants reported they found the parallel coordinate plots confusing: too many lines were used and thus the picture provided was not clear. Much effort was needed, patterns were difficult to see, and it required more time to examine a particular variable. This was critical for its effectiveness, and this may explain the poor results in the user rating (compatibility, ease of use, understanding, satisfaction and preference rating). The visual processing of graphical displays by users (visual recognition and visual grouping) is an important factor in graphical perception (Cleveland 1993). The display of the parallel coordinate plot was found difficult to understand, although good for relating multiple variables, with its dynamic, interactive features. It was particularly inappropriate for tasks such as cluster, distinguish, and locate for patterns that are found at different locations, tasks that are related to visual attention (Zhou and Feiner 1998).

Among the clustering tools, the 2D/3D surface was found to be more comprehensible for visual grouping (proximity, similarity), and helpful for finding small differences within clusters, although it was reported that the use of fuzzy boundary made it a bit difficult to see cluster borders. The 2D/3D surface is generally better viewed than the unified distance matrix. The 2D/3D projection was much better viewed for representing proximity among data items. The unified distance matrix was found clear and helpful with the use of the hexagonal grid. These SOM-based tools for visual clustering were found better than the parallel coordinate plot.

7.6. Conclusion

The usability testing reported in this chapter has provided some insight into the performance, usefulness and usability of the SOM-based representations (unified distance matrix, 2D/3D projection, 2D/3D surface, and component plane display) compared with the map and the parallel coordinate plot for specific visual tasks.

To investigate the usability of the different representations, a test is needed to examine the subject's ability to perform visual tasks such as identifying clusters and relating the visual features to problems in the data exploration domain. This was realized by applying the visual taxonomy-based evaluation methodology developed in Chapter 6 in order to compare the use of SOM-based representations with that of maps and parallel coordinate plots.

For visual grouping and clustering, the SOM-based representations performed better than the parallel coordinate plot. For detailed exploration of attributes of the dataset, correlations and relationships, the SOM component plane display was found more effective than the map for visual analysis of the patterns in the data

and for revealing relationships. The map was generally a better representation for tasks that involve visual attention and sequencing (locate, distinguish, identify, rank).

The results of this test can serve as a guideline for designing geovisualization tools that integrate different representations such as maps, parallel coordinate plots and other information visualization techniques.

The integration of visual tools can for example use tools such as the SOM component plane display for visual processing of relationships and correlations in the data. Results of user' exploration can be presented in maps as the final output of the exploration process.

Chapter 8

Conclusions and recommentations

8.1. Conclusions of the research

The approach presented in this thesis focuses on the effective application of computational algorithms to extract patterns and relationships in large geospatial data, and the visual representation of the derived information, in order to facilitate knowledge construction. Based on this approach to combine visual and computational analysis, a prototype visualization environment was developed to implement the different methods proposed and contribute to the analysis of large volumes of geospatial data. The development of the prototype involved a number of design issues, including a usability framework that involved users in usability inspection and testing at different stages of the development process. The design emphasized the effective use of visual variables used in the graphical representations to represent basic visualization tasks that were derived from a taxonomy of visual exploration operations. The self-organizing map (SOM) demonstrates interesting capabilities in features extraction, clustering, and the projection of the dataset. The representation of the SOM (grid) provides opportunities for exploring the attribute space and, when integrated with appropriate visual exploration tools, for supporting exploratory visualization and the knowledge discovery process. An important issue was to integrate the different representational approaches into a user interface structured to enhance exploration and provide more flexibility and control for spatial analysis purposes. This was realized with the link between the attribute space visualization based on the SOM graphical representations, the geographical space with maps representing the SOM results, and other graphics such as parallel coordinate plots, in multiple views. This provides alternative perspectives for the better exploration, evaluation and interpretation of patterns and ultimately for supporting knowledge construction. Effective use of visual variables used in the design of the graphical representations was important for facilitating knowledge construction. Cartographic methods were used to improve the use of colour (colour scheme) and representation issues. Interactive manipulation (zooming, rotating, panning, filtering and brushing) of the graphical representations was an important factor in enhancing user goal-specific querying and selection from the general patterns extracted to more specific user selection of attributes and spatial locations for exploration, hypothesis generation, explanation and knowledge construction, and to support the cognitive activities involved.

An application with a large spatio-temporal dataset was explored. The SOM algorithm was used to uncover the structure, patterns, relationships and trends,

and to portray spatio-temporal patterns in a visual form that can allow better understanding of the derived structures and the geographical processes. Some techniques to specifically address temporal representations were explored.

A goal of the research was to characterize the overall effectiveness of the proposed representational approaches used in the exploratory geovisualization environment. An empirical usability test was conducted to assess the usability and usefulness of the different representations as compared with the use of maps and parallel coordinate plots. The study assessed how users understand the representations, the design concepts, the ability to support specific tasks, as well as the extent to which they support particular user goals, and finally the overall effectiveness of the design of such an exploratory visual-computational environment. The evaluation method emphasizes exploratory tasks and knowledge discovery support. The test involved a representation of intended users and a number of basic visualization tasks derived from a taxonomy developed in Chapter 6. The assessment provided some insight into, and understanding of, the effectiveness and usefulness of the representations for exploratory visual analysis, interpretation and understanding of the structure in the dataset. The usability study conducted in Chapter 7 found that the SOM computational analysis, with the appropriate visual exploration tools, could support exploratory visualization and the knowledge discovery process.

The SOM-based representation techniques for clustering (unified distance matrix, 2D/3D projection, 2D/3D surface) were found to be appropriate for visual grouping and for revealing summaries and the global patterns in the data. The SOM component plane display was found to offer fast visual perception and was easy for finding relationships and understanding the patterns in the data. This representation was especially effective and suitable for tasks involving visual composition, such as associate, correlate, identify and compare. It does not require much effort to view the patterns or to relate different attributes in a single view. In the component plane display, relationships between the attributes of a large dataset become very apparent to users. For design purposes, this study has shown that information visualization techniques can be effectively integrated with maps and other graphics such as parallel coordinate plots in order to enhance visual exploration. The general strategy derived from the evaluation of the proposed approach is that general clustering tools are needed in geovisualization. From the general patterns extracted, more detailed exploration can be carried out using maps and other information spaces such as the SOM-based component plane visualization. Since the map was found more effective for visual attention and sequencing (for tasks such as locate, distinguish and rank), it can better be used to represent the results of the information extraction process. For the exploration of relationships and correlations, other tools such as the SOM component plane display can be used to effectively support visual exploration.

The approach presented here provides opportunities to improve geographical analysis and support visual exploration and knowledge discovery in the context of

large geospatial datasets. One of the advantages of the SOM is that the algorithm is fast and effective for extracting patterns and relationships in very large datasets. Based on a similarity analysis, the algorithm was found to be effective in searching for correlations among operating variables. This can be achieved using the SOM component plane visualization, which enables the understanding of processes through visual representation, allowing several variables and their interactions to be inspected simultaneously. Patterns, relationships, irregularities and distributions can be effectively visualized.

New representation forms used to visualize geospatial data such as the SOM use new alternative techniques to represent the attribute spaces. An important step in the design of such visualization tools will rely on understanding the way users make interpretations of the information spaces. The choice of a proper representation metaphor is crucial to the successful use of the tools. The link between the attribute space using visualization tools and maps in multiple views can provide multiple perspectives for exploration, evaluation and interpretation of patterns, and ultimately support knowledge construction.

The results of the research have provided some answers to some of the questions put forward in geovisualization research, related to the increasing problems posed by the exploration of large volumes of geospatial data. Specifically, the research has offered insights on the specific objectives and questions presented in chapter 1 of this thesis:

1. The integration of computational and visual tools was proved to be useful and effective for visual exploration of patterns and relationships in large geospatial data. Tools for the attribute space visualization such as the SOM can well be integrated with the maps.

2. The SOM can support visual data mining by offering effective patterns extraction and visualization features.

3. The usability test results presented in Chapter 7 revealed that the SOM was generally found significantly better tool for analyzing and understanding patterns and relationships in data. Furthermore, each of the visual tools examined (map, SOM representations or parallel coordinate plot) was found to emphasize particular visual tasks. The SOM was found particularly useful for visual comparison and visual composition tasks such as associate, correlate, identify and compare. The map was found to be better for tasks involving visual attention such as locate, distinguish. These results offer opportunity for design of integrated visual tools in multiple views that can support exploratory tasks. The SOM can be used as a processing tool with which users can interact, visually detect patterns and relationships. The output of this process can better be presented using maps.

4. The research has also shown that non-geographic information spaces such as the SOM representations can be combined with geographic maps to improve visual interaction where the volumes of geospatial data are larges.

5. An application to time related representation problems in geospatial data was explored. The approach examined in the thesis offers several representational methods (component plane display, the visualization of trajectories, and projections) that facilitate the detection of changes in spatio-temporal data. These representations were combined with maps in a cartographic animation design to enhance change detection.

6. A task-based usability evaluation method was developed based on a taxonomy of visualization goals and tasks. The taxonomy was used to develop low-level tasks that were used to evaluate and compare the different representations. This evaluation method can serve in the assessment of any time of geovisualization environment.

7. Overall the approach of combining computational and visual approaches was found appropriate and effective to contribute to exploratory analysis of large geospatial data, and to support knowledge construction, as proposed in figure 3.1.

These results and answers to the research questions described above provide some guidelines for geovisualization design. The research shows that visual exploration can be enhanced by combining the attribute space and the geographic space visualizations. To be effective, this integration of visual tools needs to be done appropriately since these tools support different visual tasks.

Based on the usability test results, the integration of map and other representations techniques such as parallel coordinate plot and the SOM-based visualization of the attributes space should reflect the potential of each visual tool. This was found in the taxonomy-based test performed in Chapter 7. The attribute space visualization is effective as a visual data mining data allowing the user to select, filter, and output results. The results of this process can be viewed in maps.

A number of related issues were not covered within this research, and could be interesting for further research in the field. Some of these issues are outlined in the next section.

8.2. Recommendations

A number of further research issues related to the work presented in this thesis can be identified. More in depth research on the perceptual issues in the design of

geovisualization not examined in this research will be required since perceptual issues are important for the success of the geovisualization design.
Other applications of the method can be suggested for remote sensing image classification, combining unsupervised learning (SOM) and supervised learning (learning vector quantization) to understand processes and patterns during the classification tasks for complex image data.

We suggest a number of research goals that need to be further explored.

8.2.1. Issues related to visual perception and visual information processing

For new geovisualization designs, it is necessary to assess this capability of the graphical representations, since they use new representation forms, metaphors and representational spaces. Most geovisualization tools, however, are not grounded on perceptual theory in the design, and no framework for assessing different aspects of perception of the graphical representations exists for geovisualization design.

The most dominant research into understanding visualizations is Bertin's work (1967).A wide range of research work has been conducted on encoding, perception and the representation of graphs (Cleveland and McGill 1984; Cleveland and McGill 1986; MacEachren 1995). In cartographic visualization, a model of how insight is provided to geographers and earth scientists was presented (MacEachren and Ganter 1990).

To study the effectiveness of visualization environments, it is useful to understand the process of human perception. Understanding the perception of certain visual properties can lead to their effective use in visualization design. Some visualization tasks or functionalities may require more time or more human cognitive processing effort than others. A framework is needed that can guide the design of visualization, taking into consideration the perceptual theories. The use of visual variables for the perception of categories or clusters, including the effect of colour and size in a number of forms and combinations, is an important aspect to investigate in geovisualization research.

8.2.2. Issues on remote sensing image classification

In remote sensing, new data acquisition techniques offer tremendous opportunities that result in more and more geospatial data. New high-resolution sensing systems (e.g., IKONOS), synthetic aperture radar (SAR) and laser-based LIDAR systems can achieve spatial resolution down to 1 m and to the sub-metre level. In addition to spatial resolution, remote sensing systems are also improving with respect to spectral resolution. New hyperpectral sensor systems such as the

airborne visible infrared imaging spectrometer (AVIRIS) capture over 200 electromagnetic bands, generating a very detailed spectral signature for each pixel. These advances are creating new applications of remote sensing as well as challenges in information extraction from such large amounts of data and imagery files. New approaches are needed for uncovering and understanding the geographical patterns or processes.

Artificial neural networks have been applied in a number of studies on remote sensing image classification. These studies have mostly revealed that the neural network is superior to conventional classifiers, often recording overall accuracy improvements in the range of 10 to 20 percent (Liu et al. 2001). Liu et al. (2001) provide three main reasons why increasing application of neural networks in remote sensing classification can produce more accurate results than conventional approaches:

1. Neural network classifiers, which make no a priori assumption about the data distributions, are able to learn non-linear and discontinuous patterns in the distribution classes.
2. Neural networks can readily accommodate collateral data such as textural information, slope, aspect and elevation.
3. Neural networks are quite flexible and can be adapted to improve performance for particular problems.

In particular, experiments on the SOM in remote sensing image classification demonstrated rapid convergence, and showed good results compared with other methods (Gahegan and Takatsuka 1999; Luo and Tseng 2000; Evangelou et al. 2001). This is due to its ability to capture the probability distribution of the inputs. The SOM has been implemented for classification, including hyperspectral image analysis, in a number of image processing software packages.

Although the use of artificial neural networks and particularly SOM in remote sensing image classification has been successful, there are still a number of important issues in their applications. A better understanding of the training parameters (number of iterations, training time, neighbourhood selection, etc.) can support the understanding of the classification process and improve results.

The approach proposed in this thesis for combining computational and visual support for the exploration of large geospatial data, can be used to develop techniques to support the understanding and the manipulation of training parameters used to achieve results in remote sensing image classification. Visualization can provide insight into the workings of a network by transforming these parameters into more easily understood visual representations (Craven and Shavlik 1991). Some of the most recent work in this area is that of using a family of artificial neural networks (Gopal et al. 2001), known as fuzzy adaptive resonance theory (ART) networks.

The importance of spatial data mining is growing with the increasing incidence and importance of large geospatial datasets, including remote sensing images. The SOM (unsupervised learning) can be combined with learning vector quantization (LVQ), a supervised learning algorithm that can be seen as the supervised version of the SOM, in a framework for spatial data mining and visualization in order to improve geospatial analysis of complex remote sensing image data. The unsupervised SOM can first be used to cluster regions, and constructs a topology-preserving representation of the statistical distribution of all input data; then the LVQ (supervised) can be used to combine the outputs generated by the SOM training to tune this representation and better discriminate between pattern classes.

References

Agrawal, R., T. Imielinski and A. Swami (1993). Mining Association Rules between Sets of Items in Large Databases. ACM SIGMOD International Conference on Management of Data, Washinton, D.C.

Alhoniemi, E., J. Hollmen, O. Simula and J. Vesanto (1999). Process monitoring and modeling using the Self-Organizing Map. Integrated Computer-Aided Engineering 6(1): 3-14.

Allinson, N., H. Yin, L. Allinson and J. Slack (2001). Advances in Self-Organizing Maps. London, Springer-Verlag.

Andrienko, N., G. Andrienko and P. Gatalsky (2000). Mapping spatio-temporal data for exploratory analysis. Geo-informatics magazine 3.

Andrienko, N., G. Andrienko and P. Gatalsky (2003). Visual data exploration using space-time cube. International Cartographic Conference, Durban, South Africa.

Behme, H., W. D. Brandt and H. W. Strube (1993). Speech recognition by hierarchical segment classification. International Conference on Artificial Neural Networks (ICANN 93), Amsterdam, Springer verlag.

Bertin, J. (1967). Semiologie Graphique: les diagrammes, les reseaux, les cartes. Paris, Gauthier-Villars.

Bertin, J. (1981). Graphics and Graphic Information Processing. Berlin, Walter de Gruyter.

Bertin, J. (1983). Semiology of graphics: diagrams, networks, maps. Madison, WI, University of Wisconsin press.

Blok, C. (2001). Dynamic Visualisation Variables and Their Control in Animations of Geospatial Data. 20th International Cartographic Conference, Beijing, China.

Blok, C., B. Kobben, T. Cheng and A. A. Kuterema (1999). Visualization of relationships between spatial patterns in time by cartographic animation. Cartography and Geographic Information Science 26(2): 139-151.

Bloom, D. E. and J. D. Sachs (1998). Geography, demography and economic growth in Africa, Center for International Developemnt (CID), Harvard University.

Brewer, C. A. (1994). Color Use Guidelines for Mapping and Visualization. Visualization in Modern Cartography. A. M. MacEachren and D. R. F. Taylor. Tarrytown, NY, Elsevier Science: 123-147.

Cabena, P., P. Hadjnian, R. Stadler, J. Verhees and Z. Alessandro (1998). Discovering data mining: From concept to implementation. New Jersey, Prentice Hall.

Card, S. K., J. D. Mackinlay and B. Shneiderman (1999). Readings in Information Visualization. Using Vision to Think. San Francisco, Morgan Kaufmann Publishers.

Cartwright, W., J. Crampton, G. Gartner, S. Miller, K. Mitchell, E. Siekierska and J. Wood (2001). Geospatial Information Visualization Uer Interface Issues. Cartography and Geoinformation science 28(1).

Chandrasekaran, B., J. R. Josephon and V. Benjamins (1998). The ontology of tasks and methods. 11th Workshop on Knowledge Acquisition, Modeling and Management, Alberta, Canada.

Chen, C. (1999). Information visualization and Virtual Environments. London, Springer-Verlag.

Cleveland, W. S. (1993). A Model for Studying Display Methods of Statistical Graphics. Journal of Computational and Graphical Statistics 2: 323--364.

Cleveland, W. S. and R. McGill (1984). Graphical perception: Theory, experimentation and

application to the development of graphical methods. Journal of the American Statistical Association 79: 531-554.

Cleveland, W. S. and R. McGill (1986). An experiment in graphical perception. International Journal of Man-Machine Studies 25: 491-500.

Cook, D., J. J. Majure, J. Symanzik and N. Cressie (1996). Dynamic Graphics in a GIS: Exploring and Analyzing Multivariate Spatial Data Using Linked Software. Computational Statistics: Special Issue on Computeraided Analysis of Spatial Data 11(4): 467-480.

Cottrell, M., E. de Bodt and M. Verleysen (2001). A statistical tool to assess the reliability of Self-Organizing Maps. Advances in Self-Organizing Maps. N. Allinson, H. Yin, L. Allinson and J. Slack. London, Springer-verlag.

Craven, M. W. and J. W. Shavlik (1991). Visualizing learning and Computation in Artificial Neural Networks. International Journal on Artificial Intelligence Tools 1: 399-425.

Cromley, E. K. and S. L. McLafferty (2002). GIS and public health. New York, The Guilford Press.

Croner, C., L. Pickle, D. Wolf and A. White (1992). A GIS approach to hypothesis generation in epidemiology. ASPRS/ACSM technical papers. A. W. Voss. Washinton, DC, ASPRS/ACSM. 3: 275-283.

Deichmann, V. (1999). Geographic aspects of inequality and poverty.

Demartyines, P. and J. Herault (1997). Curvilinear Componet Analysis: A Self-organizing Neural Network for nonlinear mapping of data sets. IEEE Transactions on Neural Networks 8(1).

Dittenbach, M., D. Merkl and A. Rauber (2000). The Growing Hierarchical Self-Organizing Map. International Joint Conference on Neural Networks 2000 (IJCNN'2000), Como, Italy.

Dodge, M. and R. Kitchin (2001). Mapping cyberspace. London, Routledge.

Domik, G. O. (1993). Scientific Visualization. ED-MEDIA '93, World Conference on Educational Multimedia and Hypermedia, Orlando, Florida.

Domik, G. O. and B. Gutkauf (1994). User Modeling for Adaptive Visualization Systems. IEEE Visualization '94, IEEE Computer Society.

Dorling, D. and S. Openshaw (1992). Using computer animation to visualize space-time patterns. Environment and Planning B: Planning and Design 19: 639-650.

Dykes, J. A. (1997). Exploring Spatial Data Representation with Dynamic Graphics. Computers & Geosciences 23(4): 345-370.

Eberts, R. E. (1994). Four approaches to human-computer interaction. User interface design. R. E. Eberts. New Jersey, Prentice-Hall.

Edsall, R. M., M. J. Kraak, A. M. MacEachren and D. J. Peuquet (1997). Assessing the effectiveness of temporal legends in environmental visualization. GIS/LIS'97, Cincinnati.

Egbert, S. L. and T. A. Slocum (1992). EXPLOREMAP: An Exploration System for Choropleth Maps. Annals, Association of American Geographers . 82(2): 275-288.

Eick, S. G. (1997). Engineering perceptaually effective visualizations for abstract data. Scientific visualization: overview, methodologies and techniques. G. M. Nielson, H. Hagen and H. Muller. Los Alamitos, CA, IEEE computer Society Press: 191-210.

Evangelou, I. E., D. G. Hadjimitsis, A. A. Lazakidou and C. Clayton (2001). Data Mining and Knowledge Discovery in Complex Image Data using Artificial Neural Networks. Workshop on CRGD: Complex Reasoning on Geographical Data, Paphos-Cyprus, Deductive Constraint Databases for Intelligent Geographical Information Systems.

Fabrikant, S. I. (2001a). Evaluating the usability of the scale metaphor for querying semantic information spaces. Spatial Information Theory: Foundations of Geographic Information Science. D. R. Montello. Berlin, Germany, Springer Verlag: 156-171.

Fabrikant, S. I. (2001b). Visualizing region and scale in information spaces. Proceedings of the 20th International Cartographic Conference, Beijing, China.

Fabrikant, S. I. and B. Buttenfield (2001). Formalizing semantic spaces for information access. Annals of the Association of American Geographers 91(2): 263-280.

Fabrikant, S. I., M. Ruocco, R. Middleton, D. R. Montello and C. Jorgensen (2002). The first Law of Cognitive Geography: Distance and Similarity in Semantic Space. GIScience 2002, Boulder, CO.

Fabrikant, S. I. and A. Skupin (2004). Cognitively Plausible Information Visualization. Exploring GeoVisualization. J. Dykes, A. MacEachren and M. J. Kraak. Amsterdam, Elsevier.

Fayyad, U., G. Piatetsky-Shapiro and P. Smyth (1996). From data mining to knowledge discovery in databases. Artificial Intelligence Magazine. 17: 37-54.

Foody, G. M. (1999). Applications of the Self-Organizing feature map Neural Network in community data analysis. Ecological-Modelling 120(2-3): 97-107.

Frohlich, J. (1997). Neural Networks with Java. Web resources.

Fuhrmann, S., P. Ahonen-Rainio, R. Edsall, S. I. Fabricant, E. L. Koua, C. Tolon, C. Ware and S. Wilson (2004). Making useful and useable Geovisualization: design and evaluation issues. Exploring Geovisualization. J. Dykes, A. M. MacEachren and M. J. Kraak. Amsterdam, Elsevier.

Gagné, R. M. (1977). The conditions for learning. New York, Rinehart and Winston.

Gahegan, M. (1998). Scatterplots and scenes: Visualization techniques for exploratory spatial analysis. Computer Environment and Urban Systems 22(1): 43-56.

Gahegan, M. (2000a). On the application of inductive machine learning tools to geographical analysis. Geographical Analysis 32(2): 113-139.

Gahegan, M. (2000b). Visualization as a tool for GeoComputation. GeoComputation. S. Openshaw and R. J. Abrahart. New York, Taylor & Francis.

Gahegan, M. (2001). Data mining and knowledge discovery in the

geographical domain. Intersection of Geospatial Information and Information Technology, Content and Knowledge Distillation, Committee of the Computer Science and Telecommunications Board.

Gahegan, M. and B. Brodaric (2002). Computational and Visual Support for Geographical Knowledge Construction: Filling the gaps between Exploration and Explanation. Symposium on Geospatial Theory, Processing and applications, Ottawa, Canada.

Gahegan, M., M. Harrover, T. M. Rhyne and M. Wachowicz (2001). The integration of geographic visualization with Databases, Data mining, Knowledge Discovery Construction and Geocomputation. Cartography and Geographic Information Science 28(1): 29-44.

Gahegan, M. and M. Takatsuka (1999). Dataspaces as an organizational concept for the neural classification of geographic datasets. Fourth International Conference on GeoComputation, Fredericksburg, Virginia, USA.

Gallistel, C. R. (1990). Representations in animal cognition: An introduction. Cognition 37: 1-22.

Gallup, L. J., J. D. Sachs and A. Mellinger (1999). Geography and economic development, Center for International Development, Havard University.

Galton, A. (2001). Space, Time and the Representation of Geographical Reality. TOPOI 20: 173-187.

Girardin, L. (1995). Mapping the virtual geography of the World Wide Web. Fifth International World Wide Web conference, Paris, France.

Gopal, S., W. Liu and C. Woodcock (2001). Visualization Based on the Fuzzy ARTMAP neural Network for mining Remotely Sensed data. Geographic data mining and knowledge discovery. H. J. Miller and J. Han. London, Taylor and Francis.

Gould, P. R. (1995). The coming plague. New York, Blackwell.

Green, M. (1998). Toward a perceptual science of multidimensional data visualization: Bertin and beyond.

Groth, R. (1999). Data mining: a hands-on approach for business professionals. New Jersey, Prentice hall.

Guimaraes, G. (2000). Temporal knowledge discovery with Self-Organizing Neural Networks. International Journal of Computer Science and Systems.

Hägerstrand, T. (1970). What about people in Regional Science? Papers of the Regional Science Association 24:: 7-21.

Hägerstrand, T. (1982). Diorama, path and project. Tijdschrift voor Economische en SocialeGeographie 73: 323 - 339.

Haklay, M. and C. Tobon (2003). Usability evaluation and PPGIS: toward a user-centered design approach. International Journal of Geographical Information Science 17(6): 577-592.

Hannon, B. and M. Ruth (1994). Dynamic modeling. New York, Springer-Verlag.

Harrover, M. (2002). Visualizing Change: using cartographic animation to explore remote-sensed data. Cartographic Perspectives 39: 30-42.

Hartson, H. R. and D. Hix (1989). Towards empirically derived methodologies and tools for HCI development. International Journal of Man-Machine Studies 31: 477-494.

Haykin, S. (1994). Neural Networks: A comprehensive foundation. New Jersey, Prentice Hall.

Hinton, G. E., J. L. McClelland and D. E. Rumclhart (1986). Distributed representations. Parallel Distributed Processing: Explorations in the Microstructure of Cognition. D. E. Rumclhart and J. L. McClelland. Cambridge, MA, MIT Press: 77-109.

Hirtle, S. C. and J. Jonides (1985). Evidence of hierarchies in cognitive maps. Memory and Cognition 13(3): 208-217.

Hix, D. and H. R. Hartson (1993). Developing user interfaces: Ensuring usability through product and process. New York, John Wiley & Sons.

Hodge, D. C. and D. G. Janelle (2000). Information, place and cyberspace: issues on accessibility. New York, Springer.

Inselberg, A. (1985). The Plane with Parallel Coordinates. The Visual Computer. 1: 69-91.

Jacobson, R. (1999). Information design. London, The MIT Press.

Jollife, I. T. (1986). Principal Component Analysis, Springer-Verlag.

Kaski, S. (1997). Data exploration using self-organizing maps. Acta Polytechnica Scandinavica, Mathematics, Computing and Management in Engineering Series, Helsinki University of Technology, Finland. 82.

Kaski, S. and T. Kohonen (1996). Exploratory data analysis by the self-Organizing Map: Structure of welfare and poverty in the world. Third International Conference on Neural Networks in the Capital market, London, England.

Keim, D. and H. Kriegel (1996). Visualization technique for mining large database. Journal of Computational and Graphical Statistics 8(6): 923-938.

Keim, D. A. (2002). Information Visualization and Visual Data Mining. IEEE transactions on Visualization and Computer Graphics 7(1): 100-107.

Keim, D. A., M. C. Hao, J. Ladisch, M. Hsu and U. Dayal (2001). Pixel Bar Charts: A New Technique for Visualizing Large Multi-Attribute Data Sets without Aggregation. INFOVIS: 113-.

Keller, P. and M. Keller (1992). Visual clues: Practical Data Visualization. Los Alamitos, CA, IEEE Computer Scociety Press.

Kienegger-Domik, G. O. (1995). Intelligent Visualization Systems in Educational Environments. World Conference on Educational Multimedia and Hypermedia, (ED-MEDIA 95), Graz, Austria.

Knapp, L. (1995). A task analysis approach of geographic data. Cognitive aspects of Human-Computer Interaction for Geographic Information Systems. T. L. Nyerges, D. M. Mark, R. Laurini and M. J. Egenhofer. Kluwer, The Netherlands, Nato ASI 83: 355-371.

Kohonen, T. (1989). Self-Organization and Associative memory, Spring-Verlag.

Kohonen, T. (1991). Self-Organizing maps optimisation approaches. Artificial Neural Networks.

Kohonen, T. (1995). Self-Organizing maps, Spring-Verlag.

Kohonen, T. (1997). Exploration of very large databases by self-organizing maps. ICNN'97, International Conference on Neural Networks, Piscataway, NJ, IEEE Service Center.

Kohonen, T. (2001). Self-Organizing Maps. Berlin, Spring-Verlag.

Kohonen, T., S. Kaski, K. Lagus, J. Salojärvi, V. Paatero and A. Saarela. (2000). Self-Organization of a Massive Document Collection. IEEE Transactions on Neural Networks, Special Issue on Neural Networks for Data Mining and Knowledge Discovery 11(3): 574-585.

Kolb, D., A. (1984). Experiential learning: Experiences as the source of learning and development. New Jersey, Prentice-Hall.

Koua, E. L. (2002). Self-organizing Maps for Geospatial Information Visualization. 98th Annual meeting of the Ameraican Association of Geographers, Los Angeles, USA.

Koua, E. L. (2003a). 'Self-Organizing Maps' voor de representation en visualization van complexe ruimtelijke gegevens (Exploring Self-organizing Maps for representation and visualization of complex geospatila datasets). Dutch Cartographic Journal.

Koua, E. L. (2003b). Using Self-organizing Maps for Information Visualization and Knowledge Discovery in large geospatial datasets. 21th International Cartographic Conference, Durban, South Africa, ICA.

Koua, E. L. and M. J. Kraak (2004a). Evaluating Self-organizing Maps for Geovisualization. Exploring Geovisualization. J. Dykes, A. M. MacEachren and M. J. Kraak. Amsterdam, Elsevier.

Koua, E. L. and M. J. Kraak (2004b). Geovisualization to support the exploration of large health and demographic survey data. International Journal of Health Geographics 3:12.

Koua, E. L. and M. J. Kraak (2004c). Integrating computational and visual analysis in exploratory visualization and knwoledge discovery in the exploration of health statistics. Advances in Spatial Data handling II: Proceedings of the 11th International Symposium on Spatial Data Handling. P. F. Fisher. New York, Springer Verlag.

Koua, E. L. and M. J. Kraak (2004d). Self-Organizing Maps for exploratory visualization and Knowledge Discovery in large geopatial datasets. Self-Organising Maps: Applications in Geographic Information Sciences. P. Agarwal and A. Skupin. New York, Wiley & Sons.

Koua, E. L. and M. J. Kraak (2004e). A usability framework for the design and evaluation of an Exploratory Geovisualization Environment. 8th International Conference on Information Visualization, IEEE Computer Society.

Kraak, M. J. (1998). Exploratory cartography, maps as tools for discoveries. Enschede, ITC.

Kraak, M. J. (2000a). About maps, cartography, geovisualization and other graphics. Geoinformatics journal 3.

Kraak, M. J. (2000b). Visualisation of the time dimension. Publications on Geodesy Netherlands Geodetic Commission 47.

Kraak, M. J. (2003). The space-time cube revisited from a geovisualization perspective. 21st International Cartographic Conference (ICC 03), Durban, South Africa.

Kraak, M. J. and A. Brown (1996). Web cartography: Developments and prospects. London, Taylor & Francis.

Kraak, M. J., J. C. Muller and F. Ormeling (1995). GIS-cartography: visual decision support for spatio-temporal data handling. International Journal of Geographical Information Systems 9(6): 637-645.

Krathwohl, D. R. (1993). Methods of educational and social science research: An integrated approach. New York, Longman.

Landau, B. and R. Jackendoff (1993). What and where in spatial language and spatial cognition. Behavioral and brain sciences 16: 217-265.

Lang, K. J. and M. J. Witbrock (1988). Learning to tell two spirals apart. 1988 Connectionist Models Summer school, Morgan Kaufmann.

Liu, W., S. Gopal and C. Woodcock (2001). Spatial data mining for classification, visualization and interpretation with ARTMAP Neural Network. Geographic

data mining and knowledge discovery. H. J. Miller and J. Han. London, Taylor and Francis.

Lohse, J. (1991). A cognitive model for the perception and understanding of graphs. SIGCHI conference on Human factors in computing systems: Reaching through technology, New Orleans, Louisiana, United States.

Luo, J. H. and D. Tseng, C. (2000). Self-Organizing Feature Map for Multispectral Spot Land Cover classification. ACRS 200.

MacEachren, A. M. (1994). Visualization in modern cartography: setting the agenda. Visualization in modern cartography. D. R. F. Taylor. Oxford, UK, Pergamon: 1-12.

MacEachren, A. M. (1995). How maps work: representation, visualization, and design. New york, The Guilford Press.

MacEachren, A. M. (2000). An evolving cognitive-semiotic approach to geographic visualization and knowledge construction. Information Design Journal 10(1): 26–36.

MacEachren, A. M., I. Brewer, G. Cai and J. Chen (2003). Visually-enabled geocollaboration to support data exploration and decision-making. 21st International Cartographic Conference (ICC) 'Cartographic Renaissance', Durban, South Africa.

MacEachren, A. M., M. Gahegan, W. Pike, I. Brewer, G. Cai, E. Lengerich and F. Hardisty (2004). Geovisualization for knowledge construction and decision support. Computer Graphics and Applications, IEEE 24(1): 13-17.

MacEachren, A. M. and J. H. Ganter (1990). A pattern identification approach to cartographic visualization. Cartographica 27(2): 64-81.

MacEachren, A. M. and M. J. Kraak (2001). Research challenges in geovisualization. Cartography and Geoinformation science 28(1).

MacEachren, A. M., M. Wachowicz, R. Edsall, D. Haug and R. Masters (1999). Constructing knowledge from multivariate spatiotemporal data: integrating geographical visualization with knowledge discovery in databases methods. International Journal of Geographical Information Science 13(4): 311-334.

Malczewski, J. (1999). GIS and multicriteria decision analysis. New York, John Wiley & Sons.

Mallach, E. G. (1994). Understanding decision support and expert systems. Chicago, Irwin.

Mao, J. and A. K. Jain (1995). Artificial neural Networks for feature extraction and multivariate data projection. IEEE transactions on neural Networks 6(2): 296-317.

Marr, D. (1982). Vision. New york, W.H. Freeman co.

Martinez, W. L. and A. R. Martinez (2002). Computational Statistics handbook with MALTAB. Boca Raton, Chapman & Hall/CRC.

McCarthy, B. (1987). The 4MAT system. IL, Excel Inc.

McCormick, B., T. A. DeFanti and M. D. Brown (1987). Visualization in scientific computing. ACM SIGGRAPH Computer Graphics, special issue 21(6).

McKendry, J. E. (1991). The role of geography in extending biodiversity gap analysis. Applied Geography 11: 135-152.

McNamara, T. P. (1986). Mental representations of spatial relations. Cognitive Psychology 18: 87-121.

Michalski, R. S. and K. A. Kaufman (1997). Machine Learning and Data mining: Methods and Applications, John Wiley & Sons publications.

Miller, H. J. and J. Han (2001). Geographic data mining and knowledge discovery. London, Taylor and Francis.

Mizoguchi, R., J. Vanwelkenhuyen and M. Ikeda (1995). Task ontology for Reuse of Problem Solving. Towards Very Large Knowledge Bases. N. J. I. Mars. Amsterdam, IOS Press: 46-59.

Mockler, R. J. and D. G. Dologite (1992). An introduction to experts systems: knowledge-based systems. New york, Macmillan Publishing Company.

Monmonier, M. (1990). Strategies for the visualization of geographic time-series data. Cartographica 27(1): 30-45.

Monmonier, M. (1992). Authoring Graphics Scripts: Experiences and Principles. Cartography and Geographic Information Systems 19(4): 247-260.

Morse, E., M. Lewis and K. A. Olsen (2000). Evaluating visualizations: using a taxonomic guide. International Journal of Human-Computer Studies 53: 637-662.

Munro, P. (1991). Visualizations of 2D hidden unit space. Pittsburgh, PA, University of Pittsburgh.

Muntz, R. R., T. Barclay, J. Dozier, C. Faloutsos, A. M. MacEachren, J. L. Martin, C. M. Pancake and M. Satyanarayanan (2003). IT road map to a Geospatial Future, report of the Committee on Intersections Between Geospatial Information and Information Technology. Washington, D C, National Academics Press.

Nielsen, J. (1994a). Heuristic evaluation. Usability Inspection Methods. J. Nielsen and R. L. Mack. New York, NY, John Wiley & Sons.

Nielsen, J. (1994b). Usability Engineering. San Francisco, Morgan Kaufmann.

Nielsen, J. and R. L. Mack (1994). Usability Inspection Methods. New York, John Wiley & Sons.

Nielsen, J. and R. Molich (1990). Heuristic evaluation of user interfaces. ACM CHI'90, Seattle, WA.

Nielson, G. M., H. Hagen and H. Muller (1997). Scientific Visualization. Overviews, methodologies, techniques. Washington, IEEE Computer Society.

Norman, D., A. and S. Draper, W. (1986). User centered system design: New perspectives on Human-Computer Interaction. New Jersy, Lawrence Erlban Associates Publishers.

Ogao, P. J. and M. J. Kraak (2002). Defining visualization operations for temporal cartographic animation design. International Journal of Applied Earth Observation and Geoinformation 4: 11-22.

Oja, E. and S. Kaski (1999). Kohonen Maps. Amsterdam, Elsevier.

Olsson, L. (1989). Integrated Resource Monitoring by means of Remote Sensing, GIS and Spatial Modelling in Arid Environments. Soil Use and Management, 5(1): 30-38.

Openshaw, S. (1995a). Developing automated and smart spatial pattern exploration tools for geographical systems applications. The Statistician 44(1): 3-16.

Openshaw, S. (1995b). Human systems modelling as a new grand challenge area in science, : . Environment and Planning A 27: 262-279.

Openshaw, S. (2000). GeoComputation. GeoComputation. S. Openshaw and R. J. Abrahart. London, Taylor and Francis,: 1-33.

Openshaw, S. and S. Alvanides (2001). Designing zoning systems for representation of socio-economic data. Time and Motion of Socio-Economic Units. I. Frank, J. Raper and J. Cheylan. London, Taylor and Francis.

Openshaw, S., A. Cross and M. Charlton (1990). Building a prototype geographical correlates machine. International Journal of Geographical Information Systems 4(4): 297-312.

Openshaw, S. and C. Openshaw (1997). Artificial Intelligence in geography. Chichester, John Wiley & Sons.

Openshaw, S. and I. Turton (1996). A parallel Kohonen algorithm for the classification of large spatial datasets. Computers-and-Geosciences 22(9): 1019-1026.

Peterson, M. P. (1995). Interactive and animated cartography. Englewood Cliffs, Prentice Hall.

Peuquet, D., J. and M. Kraak, J. (2002). Geobrowsing: Creative thinking and knowledge discovery using geographic visualization. Information Visualization 1: 80-91.

Peuquet, D. J. (1984). A conceptual framework and comparison of spatial data models. Cartographica 21(4): 66-113.

Peuquet, D. J. (1994). It's About Time: A Conceptual Framework for the Representation of Temporal Dynamics in Geographic Information Systems. Annals of the Association of American Geographers 84(3): 441 - 461.

Polson, P. G., C. Lewis, J. Rieman and C. Wharton (1992). Cognitive walkthroughs: A method for theory- based evaluation of user interfaces. International Journal of Man-Machine Studies 36: 741-773.

Pratt, L. Y. and J. Mostow (1991). Direct transfer of learned information among neural netwoks. 9th National Conference on Artificial Intelligence, Anahcim, CA.

Qian, L., M. Wachowicz, D. J. Peuquet and A. M. MacEachren (1997). Delineating operations for visualization and analysis of space-time data in GIS. GIS / LIS, Cincinnati.

Ravden, S., J. and G. Johnson, I. (1989). The evaluation checklist. Evaluating usability of human-computer interfaces: A practical method. S. J. R. G. I. Johnson. Chichester, Ellis Howard.

Rhind, J. (1993). Managing environmental data. Mapping Awareness and GIS in Europe 7(2): 3-7.

Rigol, J. P., C. H. Jarvis and N. Stuart (2001). Artificial Neural Networks as a tool for spatial interpolation. International Journal of geographical Information Science 15(4): 323-343.

Roddick, J. F., K. Hornsby and M. Spiliopoulou (2001). YABTSSTDMR - Yet Another Bibliography of Temporal, Spatial and Spatio-Temporal Data Mining Research. SIGKDD Temporal Data Mining Workshop, San Francisco, CA, ACM.

Roddick, J. F. and B. G. Lees (2001). Paradigms for Spatial and Spatio-Temporal Data Mining. Geographic Data Mining and Knowledge Discovery. H. J. Miller and J. Han. London, Taylor and Francis: 33-49.

Rogers, A. (1974). Statistical analysis of spatial dispersion. London, Pion limited.

Rouse, W. B. and K. R. Boff (1987). System design: Behavioral perspectives on designers, tools, and organizations. Amsterdam, Elsevier Science Publishing.

Roussinov, D. and C. H. (1998). A Scalable Self-organizing Map Algorithm for Textual Classification: A Neural Network Approach to Thesaurus Generation. In Communication and Cognition. Artificial Intelligence 15(1-2): 81-112.

Rowland, G. (1993). Designing and instructional deisgn. Educational Technology Research and development 41(1).

Rubin, J. (1994). Handbook of usability testing: How to plan, design, and conduct effective tests. New York, John Wiley & Sons, Inc.

Russel, S. and P. Norvig (1995). Artificial intelligence: A modern approach. New jersey, Prentice hall International.

Salton, G. and M. J. McGrill (1983). Introduction to modern information retrieval. New York, MacGraw-Hill.

Sammon, J., W. Jr. (1969). A nonlinear mapping for data structure analysis. IEEE Transactions on Computers C-18(5): 401-409.

Schaale, M. and R. Furrer (1995). Land surface classification by Neural Networks. InternationalJourna of Remote Sensing 16(16): 3003-3031.

Sheperd, A. (1989). Analaysis and training in information technology tasks. Task analysis for human-computer interaction. D. Diaper. Chichester, Ellis Horwood.

Shneiderman, B. (1997). Designing the user interface : strategies for effective human - computer interaction, Addison-Wesley.

Sibley, D. (1988). Spatial applications of exploratory data analysis. Norwich, Geo Books.

Skidmore, A. (1995). Neural Network and GIS: GIS users need to approach Neural Networks with a good deal of caution; they certainly do not take the sweat out of analysing a complex dataset. The Australian Geographic Information Systems Applications Journal 11: 53-55.

Skidmore, A., B. J. Turner, W. Brinkhof and E. Knowles (1997). Performance of a Neural Network: mapping forests using GIS and Remote Sensed data. Photogrammetric Engineering and Remote Sensing 63(5): 501-514.

Skupin, A. (2003). A novel map projection using an Artificial Neural Network. 21st International Cartographic Conference (ICC), 'Cartographic Renaissance;, Durban, South Africa.

Skupin, A. and B. P. Buttenfield (1997). Spatial Metaphors for Visualizing Information Spaces. Proceedings AUTO-CARTO 13, ACSM/ASPRS.

Skupin, A. and S. Fabrikant (2003). Spatialization Methods: A Cartographic Research Agenda for Non-Geographic Information Visualization. Cartography and Geographic Information Science. 30(2): 99-119.

Slocum, T., A. (1999). Thematic Cartography and Visualization. New Jersey, Prentice Hall.

Slocum, T., A. and S. L. Egbert (1993). Knowledge acquisition from choropleth maps. Cartography and Geographic Information Science 20(2): 83-95.

Slocum, T. A., C. Blok, B. Jiang, A. Koussoulakou, D. R. Montello, S. Fuhrmann and N. R. Hedley (2001). Cognitive and usability issues in geovisualization: a research agenda. Cartography and Geographic Information Science 28(1): 61-76.

Spence, R. (2001). Information visualization. Harlow, Addison-Wesley.

Strehl, A. and J. Ghosh (2002). Relationship-based clustering and visualization for multidimensional data mining. INFOMS Journal on Computing 00(0): 1-23.

Sweeney, M., M. Maguire and B. Shackel (1993). Evaluating user-computer interaction: a framework. International Journal of Man-Machine Studies 38.

Timo, H. (1997). Self-Organizing Maps in Natural Language Processing. Helsinki, Helsinki University of Technology, Department of Computer Science and Engineering.

Tobler, W. (1970). A Computer Movie Simulating Urban Growth in the Detroit Region. Economic Geography 46(2): 234-240.

Torgerson, W. S. (1952). Multidimensional Scaling, I: theory and method. Psychometrika 17(401- 419).

Tso, B. and P. M. Mather (2001). Classification methods for remotely sensed data. London, Taylor and Francis.

Tukey, J. (1977). Exploratory Data Analysis, Addison-Wesley.

Turner, B. L., G. Hyden and R. Kates (1993). Population growth and agricultural change in Africa. Gainesville, University Press of Florida.

Turner, B. L. and H. Schwarz (1980). Trends and interrelatonships in food, population, and energy in eastern Africa: A preliminary analysis, Clark University. 1.

Ultsch, A. (1993). Self-organizing Neural Networks for Visualization and Classification. Information and Classification. O. Opitz, B. Lausen and R. Klar. Berlin, Springer-Verlag: 307-313.

Ultsch, A. and H. Siemon (1990). Kohonen's self-organizing feature maps for exploratory data analysis. Proceedings International Neural Network Conference INNC'90P, Dordrecht, The Netherlands.

Van der Voort, M., M. Dougherty and S. Watson (1996). Combining Kohonen maps with ARIMA time series models to forecast traffic flow. Transportation-Research-Part-C:-Emerging-Technologies 4C(5): 307-318.

Vesanto, J., J. Himberg, E. Alhoniemi and J. Parhankangas (1999). Self-Organizing Map in MatLab: the SOM toolbox. Matlab DSP conference, Espoo, Finland.

Wachowicz, M. (2000). The role of geographic visualization and knowledge discovery in spatio-temporal modeling. Publications on Geodesy 47: 27-35.

Walter, S. D. (1993). Visual and statistical assessment of spatial clustering in mapped data. Statistics in medecine 12: 1275-1291.

Ware, C. (1999). Information Visualization: Perception for design. San Francisco, Morgan Kaufman Publishers.

Wehrend, S. and C. Lewis (2000). A problem-oriented classification of visualization techniques. IEEE Visualization.

Weijan, W. and D. Fraser (1996). Spatial and temporal Classification with Multiple Self-Organizing Maps. Society of Photo-optical instrumentation 2955: 307-314.

Wejchert, J. and G. Tesauro (1990). Neural Network Visualization. Advances in Neural Information Processing Systems. San Matco, CA, Morgan Kaufmann. 2: 465-472.

Weldon, J., L. (1996). Data mining and visualization. Database programming and design 9(5).

Wilkinson, G., C. Kontoes and C. N. Murray (1993). Recognition and inventory of oceanic clouds from satellite data using an artificial Neural Network technique, Dimethylsulphide: oceans, atmosphere and climate. international symposium, (Kluwer, for CEC; EUR 14796), Belgirate, 1992.

Wilkinson, L. (1999). The grammar of graphics. New York, Springer.

Xu, L. (2001). An overview on unsupervised learning from data mining perspective. Advances in Self-Organising Maps. N. Allinson, H. Yin, L. Allinson and J. Slack. London, Springer.

Yin, H. (2001). Visualisation induced SOM (ViSOM). Advances in Self-Organising Maps. A. N., H. Yin, L. Allinson and J. Slack. London, Springer.

Zhou, M. X. and S. K. Feiner (1998). Visual task characterization for automated visual discourse synthesis. Computer Human Interaction, Los Angeles, CA.

Author's bibliography

1. Koua E. L. and Kraak M.J. (2004). A Usability framework for the design and evaluation of an exploratory geovisualization environment. In: Proceedings 8th International Conference on Information Visualization. 14-16 July 2004 London. IEEE Computer Society Press, 2004. pp 153-158.

2. Koua E. L. and Kraak M. J. (2004). Integrating computational and visual analysis for the exploration of health statistics. In: SDH 2004: Proceedings of the 11th international symposium on spatial data handling: advances in spatial data handling II. : 23-25 August 2004, University of Leichester. / ed. by P.F. Fisher. - Berlin etc.: Springer, 2004. pp. 653-664.

3. Koua, E. L. and Kraak, M. J. (2004). Evaluating Self-organizing Maps for Geovisualization. In: Exploring Geovisualization. J. Dykes, A. M. MacEachren and M. J. Kraak (eds.). Amsterdam: Elsevier.

4. Fuhrmann, S.; Ahonen-Rainio, P.; Edsall, R.; Fabrikant, I. S.; Koua, E. L.; Tolon, C.; Ware, C.; Wilson, S. (2004). Making Useful and Useable Geovisualization: design and Evaluation issues. In: Dykes, J., A.M. MacEachren & M. J. Kraak (eds). Exploring geovisualization. Amsterdam: Elseviers

5. Koua E. L. and Kraak, M. J (2004). Geovisualization to support the exploration of large health and demographic survey data. International Journal of Health Geographics 2004, 3:12.

6. Koua E. L. and Kraak, M.J. (2004). Self-organizing maps for exploratory visualization and knowledge discovery in large geopatial data. In: Agarwal, P. and Skupin, A. (eds.) Self-Organising Maps: applications in geographic information sciences. New York: Wiley & Sons.

7. Koua E. L. & Kraak, M.J. (2003). Exploring spatio-temporal patterns in large geospatial data using Self-Organizing Maps. In Review: Cartography and geographic Information Science Journal (ACSM).

8. Koua E. L. and Kraak M. J. (2004). Alternative visualization of large geospatial data. Cartographic Journal vol 41 (3).

9. Koua, E. L. (2003). Using Self-Organizing Maps for Information Visualization and knowledge discovery in large geospatial data sets. Proceedings ICA International Cartographic Conference ICC 2003, Durban, South Africa.

10. Koua, E. L. (2003). 'Self-organizing Maps' voor de representation en visualization van complexe ruimtelijke gegevens. (Self-organizing Maps for representation and visualization of complex geospatial datasets). Kartografisch Tijdschrift, Vol. 29, Nr. 2, pp. 5-9.

11. Koua, E. L. (2002). Self-Organizing Map for Geospatial Information Visualization - 98th Annual Meeting of the American Association of Geographers, Los Angeles, CA, Mar. 20-23, 2002.

12. Koua E. L., MacEachren, A. and Kraak M.J. (Submitted). Evaluating the usability of an exploratory geovisualization environment. Submitted to International Journal of Geographical Information Science.

Appendix A1. Random numbers of task presentation (20 participants and 10 tasks)

Tasks	Participants																			
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
	4	2	6	9	6	1	6	6	3	5	9	3	10	8	3	2	5	7	4	5
	3	5	8	5	8	8	5	8	7	1	3	8	7	5	10	7	1	3	5	7
	6	10	1	2	5	3	1	5	1	4	4	7	3	6	6	10	10	9	1	3
	1	8	10	10	3	7	3	3	2	7	7	10	8	9	1	6	6	5	10	4
	8	9	2	1	7	4	2	7	4	6	10	6	1	2	2	1	8	2	6	8
	5	4	7	6	4	2	8	4	8	2	6	1	9	3	4	4	7	10	2	1
	10	3	4	3	9	10	10	2	5	8	5	5	5	4	8	8	3	4	8	9
	7	7	3	7	10	5	9	10	9	9	1	4	2	7	5	5	9	1	3	2
	9	6	9	4	2	6	4	9	10	10	8	2	6	10	7	9	4	6	7	10
	2	1	5	8	1	9	7	1	6	3	2	9	4	1	9	3	2	8	9	6

Appendix A2. Random numbers for the task presentation and the graphical representations used for each task

Task	Participants																			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
	Representation number																			
1	2	2	1	2	3	3	2	3	3	2	3	3	3	3	1	2	3	2	2	2
	1	1	2	3	1	1	1	1	1	1	2	2	2	2	2	3	1	1	3	3
	3	3	3	1	2	2	3	2	2	3	1	1	1	1	3	1	2	3	1	1
2	3	3	3	1	2	2	3	2	2	3	1	1	1	1	3	1	2	3	1	1
	2	2	1	2	3	3	2	3	3	2	3	3	3	3	1	2	3	2	2	2
	1	1	2	3	1	1	1	1	1	1	2	2	2	2	2	3	1	1	3	3
3	2	2	1	2	3	3	2	3	3	2	3	3	3	3	1	2	3	2	2	2
	3	3	3	1	2	2	3	2	2	3	1	1	1	1	3	1	2	3	1	1
	1	1	2	3	1	1	1	1	1	1	2	2	2	2	2	3	1	1	3	3
4	5	5	4	5	6	6	5	6	6	5	6	6	6	6	4	5	6	5	5	5
	6	6	6	4	5	5	6	5	5	6	4	4	4	4	6	4	5	6	4	4
	4	4	5	6	4	4	4	4	4	4	5	5	5	5	5	6	4	4	6	6
5	2	5	2	4	4	5	5	2	4	2	4	5	6	4	4	4	4	6	4	4
	5	6	6	6	2	6	4	4	2	6	2	4	5	5	2	2	5	2	6	2
	6	4	5	5	6	4	6	6	5	5	6	6	4	6	6	6	2	5	2	5
	4	2	4	2	5	2	2	5	6	4	5	2	2	2	5	5	6	4	5	6
6	3	3	3	1	2	2	3	2	2	3	1	1	1	1	3	1	2	3	1	1
	2	2	1	2	3	3	2	3	3	2	3	3	3	3	1	2	3	2	2	2
	1	1	2	3	1	1	1	1	1	1	2	2	2	2	2	3	1	1	3	3
7	3	3	3	1	2	2	3	2	2	3	1	1	1	1	3	1	2	3	1	1
	1	1	2	3	1	1	1	1	1	1	2	2	2	2	2	3	1	1	3	3
	2	2	1	2	3	3	2	3	3	2	3	3	3	3	1	2	3	2	2	2
8	2	2	1	2	3	3	2	3	3	2	3	3	3	3	1	2	3	2	2	2
	3	3	3	1	2	2	3	2	2	3	1	1	1	1	3	1	2	3	1	1
	1	1	2	3	1	1	1	1	1	1	2	2	2	2	2	3	1	1	3	3
9	1	1	2	3	1	1	1	1	1	1	2	2	2	2	2	3	1	1	3	3
	2	2	1	2	3	3	2	3	3	2	3	3	3	3	1	2	3	2	2	2
	3	3	3	1	2	2	3	2	2	3	1	1	1	1	3	1	2	3	1	1
10	2	2	1	2	3	3	2	3	3	2	3	3	3	3	1	2	3	2	2	2
	1	1	2	3	1	1	1	1	1	1	2	2	2	2	2	3	1	1	3	3
	3	3	3	1	2	2	3	2	2	3	1	1	1	1	3	1	2	3	1	1

Appendix B1. Logging sheet and effectiveness / user performance form (used by the test administrator)

Date: _____ Logged by: _____ Participant number: _____
Start time: _____ End time: _____

Task	*Representation method used*	*Correctness of response*	*Time (in seconds)*	*Other comments*
1	Maps	☐ Yes ☐ No		
	Parallel coordinate plot	☐ Yes ☐ No		
	Component planes	☐ Yes ☐ No		
2	Maps	☐ Yes ☐ No		
	Parallel coordinate plot	☐ Yes ☐ No		
	Component planes	☐ Yes ☐ No		
3	Maps	☐ Yes ☐ No		
	Parallel coordinate plot	☐ Yes ☐ No		
	Component planes	☐ Yes ☐ No		
4	Unified distance matrix	☐ Yes ☐ No		
	2D/3D projection	☐ Yes ☐ No		
	2D/3D surface	☐ Yes ☐ No		
5	Unified distance matrix	☐ Yes ☐ No		
	2D/3D projection	☐ Yes ☐ No		
	2D/3D surface	☐ Yes ☐ No		
	Parallel coordinate plot	☐ Yes ☐ No		
6	Maps	☐ Yes ☐ No		
	Parallel coordinate plot	☐ Yes ☐ No		
	Component planes	☐ Yes ☐ No		
7	Maps	☐ Yes ☐ No		
	Parallel coordinate plot	☐ Yes ☐ No		
	Component planes	☐ Yes ☐ No		
8	Maps	☐ Yes ☐ No		
	Parallel coordinate plot	☐ Yes ☐ No		
	Component planes	☐ Yes ☐ No		
9	Maps	☐ Yes ☐ No		
	Parallel coordinate plot	☐ Yes ☐ No		
	Component planes	☐ Yes ☐ No		
10	Maps	☐ Yes ☐ No		
	Parallel coordinate plot	☐ Yes ☐ No		
	Component planes	☐ Yes ☐ No		

Appendix B2. Evaluation form used by participants

Task 1- 3

Subjective measures: usefulness, user reactions (5= very good, 4= Good, 3= fairly good, 2= poor, 1= very poor)

Task	Representations and tools used	Usefulness															User reaction (attitude)										
		Compatibility of tool with user's expectations					Flexibility / ease of use of tool for the task					Perceived user's understanding of the tool					User satisfaction					User preference (rating)					User opinion and comments
		5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	5	4	3	2	1	
Task 1	Maps																										
	Parallel Coordinate Plot																										
	Component planes																										
Task 2	Maps																										
	Parallel Coordinate Plot																										
	Component planes																										
Task 3	Maps																										
	Parallel Coordinate Plot																										
	Component planes																										

Appendix B2. Evaluation form used by participants

Task 4 - 6

Subjective measures: usefulness, user reactions (5= very good, 4= Good, 3= fairly good, 2= poor, 1= very poor)

Task	*Representations and tools used*	*Usefulness*															*User reaction (attitude)*										
		Compatibility of tool with user's expectations					*Flexibility / ease of use of tool for the task*					*Perceived user's understanding of the tool*					*User satisfaction*					*User preference (rating)*					*User opinion, com-ments*
		5	*4*	*3*	*2*	*1*	*5*	*4*	*3*	*2*	*1*	*5*	*4*	*3*	*2*	*1*	*5*	*4*	*3*	*2*	*1*	*5*	*4*	*3*	*2*	*1*	
Task 4	Unified distance matrix																										
	2D/3D projection																										
	2D/3D surface																										
Task 5	Unified distance matrix																										
	2D/3D projection																										
	2D/3D surface																										
	Parallel coordinate plot																										
Task 6	Map																										
	Parallel Coordinate Plot																										
	Component planes																										

Appendix B2. Evaluation form used by participants

Task 7 - 10

Subjective measures: usefulness, user reactions (5= very good, 4= Good, 3= fairly good, 2= poor, 1= very poor)

		Usefulness															*User reaction (attitude)*										
Task	***Representations and tools used***	***Compatibility of tool with user's expectations***					***Flexibility / ease of use of tool for the task***					***Perceived user's understanding of the tool***					***User satisfaction***					***User preference (rating)***					***User opinion, com-ments***
		5	*4*	*3*	*2*	*1*	*5*	*4*	*3*	*2*	*1*	*5*	*4*	*3*	*2*	*1*	*5*	*4*	*3*	*2*	*1*	*5*	*4*	*3*	*2*	*1*	
Task 7	Maps																										
	Parallel Coordinate Plot																										
	Component planes																										
Task 8	Maps																										
	Parallel Coordinate Plot																										
	Component planes																										
Task 9	Map																										
	Parallel Coordinate Plot																										
	Component planes																										
Task 10	Map																										
	Parallel Coordinate Plot																										
	Component planes																										

Abbreviations

ANN	Artificial Neural Network
ART	Adaptive Resonance Theory
AVIRIS	Airborne Visible/Infrared Imaging Spectrometer
CDC	Centers for Disease Control and Prevention
EDA	Exploratory Data Analysis
ENVI	Environment for Visualizing Images, from RSI (Research Systems Inc.)
FAO	Food and Agriculture Organization of the United Nations
FEWS	Famine Early Warning System
GDP	Gross Domestic Product
GIS	Geographic Information Systems
GIScience	Geographic Information Science
GPS	Global Positioning System
HCI	Human-Computer Interaction
ICA	International Cartographic Association
ITC	International Institute for Geo-information Science and Earth Observation
KDD	Knowledge Discovery in Databases
LIDAR	LIght Detection And Ranging
LVQ	Learning vector Quantization
MDMV	Multidimensional multivariate visualization
MDS	Multidimensional Scaling
NDVI	Normalized Difference Vegetation Index
PCA	Principal component analysis
PCP	Parallel Coordinate Plot

SAR	Synthetic Aperture Radar
SOM	Self-Organizing Map
UCD	User-Centered Design
U-matrix	Unified Distance Matrix
USAID	United State Agency for International Development

Summary

Due to the advances in data acquisition techniques, volumes of geospatial data are rapidly increasing and the structure of datasets is becoming more complex. These large volumes of data are difficult to process with common geospatial analysis techniques. The extraction of patterns and the discovery of new knowledge may be difficult with such large and complex datasets, as certain patterns may remain hidden. One of the major research areas in geovisualization is the exploration of such complex geospatial data for the purpose of uncovering and understanding patterns or processes. New approaches in spatial analysis and visualization are needed to represent such data in a visual form that better stimulates pattern recognition and hypothesis generation, allows better understanding of structures and processes, and supports knowledge construction.

Information visualization and abstract information spaces are increasingly used as a means to visualizing such complex data. One attempt has been the use of artificial neural networks as a technology that is especially useful in situations where the numbers are vast and the relationships are often unclear or even hidden. In particular, the self-organizing map (SOM) neural network is often used as a means of organizing complex information spaces, and for the extraction of patterns and the creation of abstractions where conventional methods may be limited.

In this research, we explore the integration of computational and visual approaches, to contribute to the analysis of complex geospatial data. Computational analysis based on the SOM is used in a framework for data mining, knowledge discovery and spatial analysis, for uncovering the structure, patterns, relationships and trends in the data. Graphical representations supported by information visualization techniques and cartographic methods are then used to portray derived structures and patterns in a visual form that allows better understanding of the structures and the geographic processes. Different techniques for representation, visualization and interaction are combined for the better exploration of patterns and relationships in the data. The framework is informed by current understanding of the effective application of visual variables for cartographic and information design, by developing theories on interface metaphors for geospatial information displays, and by previous empirical studies of map and information visualization effectiveness. It is used to facilitate the knowledge construction process by supporting user's exploratory tasks in a number of ways, including a scenario for better use of the representational spaces. The ultimate goal is to support visual data mining and exploration, and gain insights into underlying distributions, patterns and trends, and thus contribute to enhancing the understanding of geographic processes and support knowledge construction.

The framework guided the initial design decisions of a prototype exploratory geovisualization environment. This design is based on a user-centred approach to structure an interface that integrates several representation forms, and visualization and interaction techniques, along with cartographic methods for the effective use of visual variables with which the visualization is depicted. The visualization environment incorporates several graphical representations of SOM output. These include a distance matrix representation, 2D and 3D projections, 2D and 3D surfaces, and component plane visualization with which correlations and relationships can be easily explored. Multiple views are used to simultaneously present interactions between several variables over the space of the SOM, maps, and other graphics such as parallel coordinate plots.

Some applications of the method are explored with different datasets. First a simple case of the exploration of a known dataset related to socio-economic data in a region of the Netherlands (the Overijssel) is used to describe the basic graphical representations. An example case of the exploration of complex attributes relationships is examined with a dataset containing complex relationships between geography and economy development, and with which a number of hypotheses are tested. With the user interface development, a specific case of the exploration of a large database on health statistics on Africa is explored to investigate the options of the interface, and the integration of the different graphical representations. In addition, a particular case of the exploration and representation of spatio-temporal patterns is examined with a dataset related to the production of food (cereals) in Africa over the last 40 years. The objective with the later case is to represent and visualize underlying space-time dynamics, and interactions between several variables. Some spatio-temporal representation techniques are proposed to support the exploration of the time-related geographic trends and patterns, and allow visual change detection in such processes. They include the use of component plane visualization, the visualization of trajectories, and projections using the time dimension.

A usability evaluation methodology based on a taxonomy of exploratory tasks and visualization operations is developed to assess the effectiveness of the proposed exploratory geovisualization environment. A subsequent empirical usability testing is conducted and involves different options of map-based and interactive visualizations of a SOM output with the exploration of a socio-demographic dataset. The study emphasizes the visual exploration and knowledge discovery processes.

The usability test results and answers to the research questions provide some guidelines for geovisualization design that integrate different representations such as maps, parallel coordinate plots and other information visualization techniques. The research shows that visual exploration can be enhanced by combining the attribute space and the geographic space visualizations. To be effective, this integration of visual tools needs to be done appropriately since these tools are found to support different visual tasks. For visual grouping and clustering, visual

analysis and comparison of the patterns in the data, and for revealing relationships, the SOM was found more effective than the map.

The usability test results suggest that the integration of map and other representations techniques such as parallel coordinate plot and the SOM-based visualization of the attributes space should reflect the potential of each visual tool. The attribute space visualization is effective as a visual data mining tool allowing the user to select, filter, and output results. The results of this process can be viewed in maps, since the map was generally a better representation for tasks that involve visual attention and sequencing (locate, distinguish, rank).

Keywords: Geovisualization, Information visualization, Self-organizing map, Spatial analysis, Data mining, Knowledge discovery, Exploratory visualization, Visual exploration, Knowledge construction.

Samenvatting (Summary in Ducth)

Door vooruitgang op het gebied van gegevensinwinningstechnieken neemt de hoeveelheid en complexiteit van de beschikbare ruimtelijke gegevens enorm toe. Deze grote hoeveelheden zijn moeilijk te verwerken met behulp van de huidige ruimtelijke analyse technieken. Doordat sommige patronen verborgen zullen blijven is de extractie van nieuwe kennis uit dergelijke datasets lastig. Een van de belangrijkste onderzoeksdoelen van de geovisualisatie is de exploratie van dergelijke complexe ruimtelijke gegevenssets met als doel het ontdekken en begrijpen van aanwezige patronen of processen. Een nieuwe aanpak in de analyse en visualisatie is nodig om de gegevens zo visueel te representeren dat dit leidt tot patronenherkenningen, het formuleren van hypothesen stimuleert, het beter begrijpen van structuren en processen bevordert, en de kennis doet toenemen.

Informatie visualisatie en abstracte informatieruimtes worden steeds meer toegepast om dergelijke complexe gegevens te visualiseren. Een van de mogelijkheden is het gebruik van kunstmatige neurale netwerken als een technologie die met name nuttig is in situaties waarin de gegevenshoeveelheid enorm is en de relaties in de dataset onduidelijk of zelfs verborgen zijn. In het bijzonder wordt de zelforganiserende kaart (self-organizing map = SOM) gebruikt om complexe informatieruimtes te organiseren, om patronen te extraheren en voor de aanmaak van (visuele) abstracties, waar conventionele methodes te beperkt zijn.

In dit onderzoek staat de integratie van de computationele en visuele benadering centraal om zo bij te dragen aan de analyse van complexe grote hoeveelheden ruimtelijke gegevens. De computationele benadering is hier gebaseerd op de SOM. Deze wordt gebruikt voor data mining, ruimtelijk analyse, het ontrafelen van structuren, patronen, relaties en trends in de data, en de ontdekking van kennis en. Grafische representaties vervolgens gebruikt om afgeleide structuren en patronen weer te geven in een visuele vorm die het beter begrijpen van de structuren en geografische processen toelaat. Hierbij spelen informatie visualisatietechnieken en cartografische methodes een belangrijke rol. Verschillende technieken voor representatie, visualisatie en interactie zijn in een prototype gecombineerd voor een betere toegankelijkheid van (verborgen) patronen en relaties in de data. De bovenstaande benadering is gefundeerd op de huidige kennis op het gebied van de toepassing van de cartografische regelgeving, informatie ontwerp, interface metaforen en op empirische studies naar het gebruik van de effectiviteit van de diverse weergave methoden zodat exploratieve taken worden ondersteund. Het uiteindelijke doel is de ondersteuning van een beter begrip van geografische processen en de verrijking van kennis door het interactieve aanbieden van gereedschappen ter exploratie van de gegevens.

De eerder genoemde aspecten vormde de basis voor het initiële ontwerp van het prototype van een exploratieve geovisualisatie omgeving. Dit ontwerp is

gefundeerd op een zogenaamde gebruikergeoriënteerde benadering zodat er een geschikte interface gemaakt kon worden. Deze interface omvat verschillende visuele representatie vormen en interactie technieken die een effectief gebruik van cartografische methoden en technieken mogelijk maakt. Onder de visuele representaties bevinden zich verschillende grafische representaties van SOM uitvoer. Dit zijn een afstandsmatrix, 2D en 3D projecties en 2D en 3D surfaces, een componenten vlak. Daarnaast zijn ook standaard kaarten en diagrammen zoals de parallelle coördinaten plot beschikbaar zodat correlaties en relaties snel bekeken kunnen worden. Hiertoe zijn meerdere aan elkaar gekoppelde vensters gebruikt, waarbij interactie in een venster onmiddellijk leidt tot reactie in de grafische representaties in de andere vensters

Het functioneren van het prototype is getoetst aan de hand van verschillende in complexiteit toenemende datasets. Als eerste is gewerkt met een eenvoudige en bekende dataset met socio-economische gegevens van de provincie Overijssel om de grafische SOM representaties te beschrijven. Een meer complexe dataset met gegevens per land over de economische ontwikkeling en de geografie is gebruikt om een aantal werk hypotheses te toetsen. Ten behoeve van een verdere ontwikkeling van de interface van het prototype en de verdere integratie van verschillende grafische representaties is gebruik gemaakt van een grote database met gezondheidsstatistieken van Afrika. Tenslotte is er nog een dataset van de voedselproductie over de laatste veertig jaar in Afrika gebruikt om de toepassing van het systeem voor de exploratie van ruimte-tijd gegevens te beoordelen. Hier worden tevens nieuwe representatie methoden voorgesteld die de exploratie van dergelijke tijdsreeksen kunnen vereenvoudigen en zijn toegespitst op het herkennen van veranderingen. Naast de componenten vlakken zijn hiertoe ook tijdspaden gebruikt.

Een bruikbaarheidsonderzoek gebaseerd op een taxonomie van exploratieve taken en visualisatie operaties is ontwikkeld om de effectiviteit van de voorgestelde geovisualisatie omgeving te toetsten. Hiertoe is een empirische gebruikerstest uitgevoerd op basis van exploratieve taken waarbij verschillende kaart en SOM gebaseerde visualisaties zijn gebruikt.

De testresultaten en antwoorden op de onderzoeksvragen geven nieuwe richtlijnen voor het ontwerp van een geovisualisatie omgeving waarbij verschillende representaties zoals kaarten, parallel coördinaten plots en informatie visualisatie vormen geïntegreerd moeten worden. Het onderzoek toont aan dat de visuele exploratie verbeterd kan worden door de combinatie van de attribuut ruimte met de geografische ruimte. Om effectief te zijn moet deze integratie wel volgens bepaalde richtlijnen worden uitgevoerd omdat ieder van de representaties in de verschillende ruimtes geschikt blijken te zijn voor bepaalde taken. Zo bleek voor clustering en visuele analyse, het vergelijken van patronen in de gegevens en voor het tonen van relaties de SOM representatie beter te functioneren dan de kaart. De test resultaten suggereren dat de integratie van de verschillende representaties overeen moeten komen met de potentie van ieder van de

gereedschappen. De visualisatie van de attribuutruimte is effectief als een visuele data mining gereedschap die de gebruiker laat selecteren, filteren. Dit resultaat kan vervolgens bekeken worden in kaarten omdat deze uitblinkt in visueel overzicht en taken als lokaliseren, onderscheiden en ordenen.

Trefwoorden: geovisualisatie, informatie visualisatie, zelf-organiserende kaart, ruimtelijke analyse, data mining, exploratieve visualisatie, kennis constructie.

ITC dissertation list

1. **Akinyede, Joseph O.,** 1990, Highway cost modelling and route selection using a geotechnical information system
2. **Pan He Ping**, 1990, 90-9003757-8, Spatial structure theory in machine vision and applications to structural and textural analysis of remotely sensed images
3. **Bocco Verdinelli, G.**, 1990, Gully erosion analysis using remote sensing and geographic information systems: a case study in Central Mexico
4. **Sharif, M**, 1991, Composite sampling optimization for DTM in the context of GIS
5. **Drummond, J.**, 1991, Determining and processing quality parameters in geographic information systems
6. **Groten, S.**, 1991, Satellite monitoring of agro-ecosystems in the Sahel
7. **Sharifi, A.**, 1991, 90-6164-074-1, Development of an appropriate resource information system to support agricultural management at farm enterprise level
8. **Zee, D. van der**, 1991, 90-6164-075-X, Recreation studied from above: Air photo interpretation as input into land evaluation for recreation
9. **Mannaerts, C.**, 1991, 90-6164-085-7, Assessment of the transferability of laboratory rainfall-runoff and rainfall - soil loss relationships to field and catchment scales: a study in the Cape Verde Islands
10. **Ze Shen Wang**, 1991: 90-393-0333-9, An expert system for cartographic symbol design
11. **Zhou Yunxian**, 1991, 90-6164-081-4, Application of Radon transforms to the processing of airborne geophysical data
12. **Zuviria, M. de**, 1992, 90-6164-077-6, Mapping agro-topoclimates by integrating topographic, meteorological and land ecological data in a geographic information system: a case study of the Lom Sak area, North Central Thailand
13. **Westen, C. van**, 1993, 90-6164-078-4, Application of Geographic Information Systems to landslide hazard zonation
14. **Shi Wenzhong**, 1994, 90-6164-099-7, Modelling positional and thematic uncertainties in integration of remote sensing and geographic information systems
15. **Javelosa, R.**, 1994, 90-6164-086-5, Active Quaternary environments in the Philippine mobile belt
16. **Lo King-Chang**, 1994, 90-9006526-1, High Quality Automatic DEM, Digital Elevation Model Generation from Multiple Imagery
17. **Wokabi, S.**, 1994, 90-6164-102-0, Quantified land evaluation for maize yield gap analysis at three sites on the eastern slope of Mt. Kenya
18. **Rodriguez, O.**, 1995, Land Use conflicts and planning strategies in urban fringes: a case study of Western Caracas, Venezuela
19. **Meer, F. van der**, 1995, 90-5485-385-9, Imaging spectrometry & the Ronda peridotites
20. **Kufoniyi, O.**, 1995, 90-6164-105-5, Spatial coincidence: automated database updating and data consistency in vector GIS
21. **Zambezi, P.**, 1995, Geochemistry of the Nkombwa Hill carbonatite complex of Isoka District, north-east Zambia, with special emphasis on economic minerals

22. **Woldai, T.**, 1995, The application of remote sensing to the study of the geology and structure of the Carboniferous in the Calañas area, pyrite belt, SW Spain
23. **Verweij, P.**, 1995, 90-6164-109-8, Spatial and temporal modelling of vegetation patterns: burning and grazing in the Paramo of Los Nevados National Park, Colombia
24. **Pohl, C.**, 1996, 90-6164-121-7, Geometric Aspects of Multisensor Image Fusion for Topographic Map Updating in the Humid Tropics
25. **Jiang Bin**, 1996, 90-6266-128-9, Fuzzy overlay analysis and visualization in GIS
26. **Metternicht, G.**, 1996, 90-6164-118-7, Detecting and monitoring land degradation features and processes in the Cochabamba Valleys, Bolivia. A synergistic approach
27. **Hoanh Chu Thai**, 1996, 90-6164-120-9, Development of a Computerized Aid to Integrated Land Use Planning (CAILUP) at regional level in irrigated areas: a case study for the Quan Lo Phung Hiep region in the Mekong Delta, Vietnam
28. **Roshannejad, A.**, 1996, 90-9009284-6, The management of spatio-temporal data in a national geographic information system
29. **Terlien, M.**, 1996, 90-6164-115-2, Modelling Spatial and Temporal Variations in Rainfall-Triggered Landslides: the integration of hydrologic models, slope stability models and GIS for the hazard zonation of rainfall-triggered landslides with examples from Manizales, Colombia
30. **Mahavir, J.**, 1996, 90-6164-117-9, Modelling settlement patterns for metropolitan regions: inputs from remote sensing
31. **Al-Amir, S.**, 1996, 90-6164-116-0, Modern spatial planning practice as supported by the multi-applicable tools of remote sensing and GIS: the Syrian case
32. **Pilouk, M.**, 1996, 90-6164-122-5, Integrated modelling for 3D GIS
33. **Duan Zengshan**, 1996, 90-6164-123-3, Optimization modelling of a river-aquifer system with technical interventions: a case study for the Huangshui river and the coastal aquifer, Shandong, China
34. **Man, W.H. de**, 1996, 90-9009-775-9, Surveys: informatie als norm: een verkenning van de institutionalisering van dorp - surveys in Thailand en op de Filippijnen
35. **Vekerdy, Z.**, 1996, 90-6164-119-5, GIS-based hydrological modelling of alluvial regions: using the example of the Kisaföld, Hungary
36. **Pereira, Luisa**, 1996, 90-407-1385-5, A Robust and Adaptive Matching Procedure for Automatic Modelling of Terrain Relief
37. **Fandino Lozano, M.**, 1996, 90-6164-129-2, A Framework of Ecological Evaluation oriented at the Establishment and Management of Protected Areas: a case study of the Santuario de Iguaque, Colombia
38. **Toxopeus, B.**, 1996, 90-6164-126-8, ISM: an Interactive Spatial and temporal Modelling system as a tool in ecosystem management: with two case studies: Cibodas biosphere reserve, West Java Indonesia: Amboseli biosphere reserve, Kajiado district, Central Southern Kenya
39. **Wang Yiman**, 1997, 90-6164-131-4, Satellite SAR imagery for topographic mapping of tidal flat areas in the Dutch Wadden Sea
40. **Saldana-Lopez, Asunción**, 1997, 90-6164-133-0, Complexity of soils and Soilscape patterns on the southern slopes of the Ayllon Range, central Spain: a GIS assisted modelling approach

41. **Ceccarelli, T.**, 1997, 90-6164-135-7, Towards a planning support system for communal areas in the Zambezi valley, Zimbabwe; a multi-criteria evaluation linking farm household analysis, land evaluation and geographic information systems
42. **Peng Wanning**, 1997, 90-6164-134-9, Automated generalization in GIS
43. **Lawas, C.**, 1997, 90-6164-137-3, The Resource Users' Knowledge, the neglected input in Land resource management: the case of the Kankanaey farmers in Benguet, Philippines
44. **Bijker, W.**, 1997, 90-6164-139-X, Radar for rain forest: A monitoring system for land cover Change in the Colombian Amazon
45. **Farshad, A.**, 1997, 90-6164-142-X, Analysis of integrated land and water management practices within different agricultural systems under semi-arid conditions of Iran and evaluation of their sustainability
46. **Orlic, B.**, 1997, 90-6164-140-3, Predicting subsurface conditions for geotechnical modelling
47. **Bishr, Y.**, 1997, 90-6164-141-1, Semantic Aspects of Interoperable GIS
48. **Zhang Xiangmin**, 1998, 90-6164-144-6, Coal fires in Northwest China: detection, monitoring and prediction using remote sensing data
49. **Gens, R.**, 1998, 90-6164-155-1, Quality assessment of SAR interferometric data
50. **Turkstra, J.**, 1998, 90-6164-147-0, Urban development and geographical information: spatial and temporal patterns of urban development and land values using integrated geo-data, Villaviciencia, Colombia
51. **Cassells, C.**, 1998, Thermal modelling of underground coal fires in northern China
52. **Naseri, M.**, 1998, 90-6164-195-0, Characterization of Salt-affected Soils for Modelling Sustainable Land Management in Semi-arid Environment: a case study in the Gorgan Region, Northeast, Iran
53. **Gorte B.G.H.**, 1998, 90-6164-157-8, Probabilistic Segmentation of Remotely Sensed Images
54. **Tenalem Ayenew**, 1998, 90-6164-158-6, The hydrological system of the lake district basin, central main Ethiopian rift
55. **Wang Donggen**, 1998, 90-6864-551-7, Conjoint approaches to developing activity-based models
56. **Bastidas de Calderon, M.**, 1998, 90-6164-193-4, Environmental fragility and vulnerability of Amazonian landscapes and ecosystems in the middle Orinoco river basin, Venezuela
57. **Moameni, A.**, 1999, Soil quality changes under long-term wheat cultivation in the Marvdasht plain, South-Central Iran
58. **Groenigen, J.W. van**, 1999, 90-6164-156-X, Constrained optimisation of spatial sampling: a geostatistical approach
59. **Cheng Tao**, 1999, 90-6164-164-0, A process-oriented data model for fuzzy spatial objects
60. **Wolski, Piotr**, 1999, 90-6164-165-9, Application of reservoir modelling to hydrotopes identified by remote sensing
61. **Acharya, B.**, 1999, 90-6164-168-3, Forest biodiversity assessment: A spatial analysis of tree species diversity in Nepal
62. **Akbar Abkar, Ali**, 1999, 90-6164-169-1, Likelihood-based segmentation and classification of remotely sensed images

63. **Yanuariadi, T.**, 1999, 90-5808-082-X, Sustainable Land Allocation: GIS-based decision support for industrial forest plantation development in Indonesia
64. **Abu Bakr**, Mohamed, 1999, 90-6164-170-5, An Integrated Agro-Economic and Agro-Ecological Framework for Land Use Planning and Policy Analysis
65. **Eleveld, M.**, 1999, 90-6461-166-7, Exploring coastal morphodynamics of Ameland (The Netherlands) with remote sensing monitoring techniques and dynamic modelling in GIS
66. **Yang Hong**, 1999, 90-6164-172-1, Imaging Spectrometry for Hydrocarbon Microseepage
67. **Mainam, Félix**, 1999, 90-6164-179-9, Modelling soil erodibility in the semiarid zone of Cameroon
68. **Bakr, Mahmoud**, 2000, 90-6164-176-4, A Stochastic Inverse-Management Approach to Groundwater Quality
69. **Zlatanova, Z.**, 2000, 90-6164-178-0, 3D GIS for Urban Development
70. **Ottichilo, Wilber K.**, 2000, 90-5808-197-4, Wildlife Dynamics: An Analysis of Change in the Masai Mara Ecosystem
71. **Kaymakci, Nuri**, 2000, 90-6164-181-0, Tectono-stratigraphical Evolution of the Cankori Basin (Central Anatolia, Turkey)
72. **Gonzalez, Rhodora**, 2000, 90-5808-246-6, Platforms and Terraces: Bridging participation and GIS in joint-learning for watershed management with the Ifugaos of the Philippines
73. **Schetselaar, Ernst**, 2000, 90-6164-180-2, Integrated analyses of granite-gneiss terrain from field and multisource remotely sensed data. A case study from the Canadian Shield
74. **Mesgari, Saadi**, 2000, 90-3651-511-4, Topological Cell-Tuple Structure for Three-Dimensional Spatial Data
75. **Bie, Cees A.J.M. de**, 2000, 90-5808-253-9, Comparative Performance Analysis of Agro-Ecosystems
76. **Khaemba, Wilson M**., 2000, 90-5808-280-6, Spatial Statistics for Natural Resource Management
77. **Shrestha, Dhruba**, 2000, 90-6164-189-6, Aspects of erosion and sedimentation in the Nepalese Himalaya: highland-lowland relations
78. **Asadi Haroni, Hooshang**, 2000, 90-6164-185-3, The Zarshuran Gold Deposit Model Applied in a Mineral Exploration GIS in Iran
79. **Raza, Ale**, 2001, 90-3651-540-8, Object-oriented Temporal GIS for Urban Applications
80. **Farah, Hussein**, 2001, 90-5808-331-4, Estimation of regional evaporation under different weather conditions from satellite and meteorological data. A case study in the Naivasha Basin, Kenya
81. **Zheng**, **Ding,** 2001, 90-6164-190-X, A Neural - Fuzzy Approach to Linguistic Knowledge Acquisition and Assessment in Spatial Decision Making
82. **Sahu, B.K.**, 2001, Aeromagnetics of continental areas flanking the Indian Ocean; with implications for geological correlation and Gondwana reassembly
83. **Alfestawi, Y.,** 2001, 90-6164-198-5, The structural, paleogeographical and hydrocarbon systems analysis of the Ghadamis and Murzuq Basins, West Libya, with emphasis on their relation to the intervening Al Qarqaf Arch
84. **Liu, Xuehua,** 2001, 90-5808-496-5, Mapping and Modelling the Habitat of Giant Pandas in Foping Nature Reserve, China

85. **Oindo, Boniface Oluoch**, 2001, 90-5808-495-7, Spatial Patterns of Species Diversity in Kenya
86. **Carranza, Emmanuel John**, 2002, 90-6164-203-5, Geologically-constrained Mineral Potential Mapping
87. **Rugege, Denis**, 2002, 90-5808-584-8, Regional Analysis of Maize-Based Land Use Systems for Early Warning Applications
88. **Liu, Yaolin**, 2002, 90-5808-648-8, Categorical Database Generalization in GIS
89. **Ogao, Patrick**, 2002, 90-6164-206-X, Exploratory Visualization of Temporal Geospatial Data using Animation
90. **Abadi, Abdulbaset M.**, 2002, 90-6164-205-1, Tectonics of the Sirt Basin – Inferences from tectonic subsidence analysis, stress inversion and gravity modelling
91. **Geneletti, Davide**, 2002, 90-5383-831-7, Ecological Evaluation for Environmental Impact Assessment
92. **Sedogo, Laurent G.**, 2002, 90-5808-751-4, Integration of Participatory Local and Regional Planning for Resources Management using Remote Sensing and GIS
93. **Montoya, Lorena**, 2002, 90-6164-208-6, Urban Disaster Management: a case study of earthquake risk assessment in Carthago, Costa Rica
94. **Ahmad, Mobin-ud-Din**, 2002, 90-5808-761-1, Estimation of Net Groundwater Use in Irrigated River Basins using Geo-information Techniques: A case study in Rechna Doab, Pakistan
95. **Said, Mohammed Yahya**, 2003, 90-5808-794-8, Multiscale perspectives of species richness in East Africa
96. **Schmidt, Karin**, 2003, 90-5808-830-8, Hyperspectral Remote Sensing of Vegetation Species Distribution in a Saltmarsh
97. **Lopez Binnquist, Citlalli**, 2003, 90-3651-900-4, The Endurance of Mexican Amate Paper: Exploring Additional Dimensions to the Sustainable Development Concept
98. **Huang, Zhengdong**, 2003, 90-6164-211-6, Data Integration for Urban Transport Planning
99. **Cheng, Jianquan**, 2003, 90-6164-212-4, Modelling Spatial and Temporal Urban Growth
100. **Campos dos Santos, Jose Laurindo**, 2003, 90-6164-214-0, A Biodiversity Information System in an Open Data/Metadatabase Architecture
101. **Hengl, Tomislav**, 2003, 90-5808-896-0, PEDOMETRIC MAPPING, Bridging the gaps between conventional and pedometric approaches
102. **Barrera Bassols, Narciso,** 2003, 90-6164-217-5, Symbolism, Knowledge and management of Soil and Land Resources in Indigenous Communities: Ethnopedology at Global, Regional and Local Scales
103. **Zhan, Qingming**, 2003, 90-5808-917-7, A Hierarchical Object-Based Approach for Urban Land-Use Classification from Remote Sensing Data
104. **Daag, Arturo S.**, 2003, 90-6164-218-3, Modelling the Erosion of Pyroclastic Flow Deposits and the Occurrences of Lahars at Mt. Pinatubo, Philippines
105. **Bacic, Ivan**, 2003, 90-5808-902-9, Demand-driven Land Evaluation with case studies in Santa Catarina, Brazil
106. **Murwira, Amon**, 2003, 90-5808-951-7, Scale matters! A new approach to quantify spatial heterogeneity for predicting the distribution of wildlife
107. **Mazvimavi, Dominic**, 2003, 90-5808-950-9, Estimation of Flow Characteristics of Ungauged Catchments. A case study in Zimbabwe

108. **Tang Xinming,** 2004, 90-6164-2205, Spatial object modeling in fuzzy topological spaces: with applications to land cover change.
109. **Kariuki, P.C.**, 2004, 90-6164-221-3, Spectroscopy and swelling soils: an integrated approach.
110. **Morales Guarin, J.M.,** 2004, 90-6164-222-1, Model-driven design of geo-information services.
111. **Mutanga, O.,** 2004, 90-5808-981-9, Hyperspectral remote sensing of tropical grass quality and quantity.
112. **Sliuzas, R.V.,** 2004, 90-6164-223-X, Managing informal settlements: a study using geo-information in Dar es Salaam, Tanzania.
113. **Lucieer, Arko**, 2004, 90-6164-225-6, Uncertainties in Segmentation and their Visualisation.
114. **Corsi, Fabio**, 2004, 90-8504-090-6, Applications of existing biodiversity information: Capacity to support decision-making.
115. **Tuladhar, Arbind**, 2004, 90-6164-224-8, Parcel-based Geo-information System: Concepts and Guidelines.
116. **Elzakker, Corné van**, 2004, 90-6809-365-7, The use of maps in the exploration of geographic data.
117. **Nidumolu, Uday Bhaskar**, 2004, 90-8504-138-4, Integrating Geo-information models with participatory approaches: applications in land use analysis.

Curriculum vitae

Etien Luc Koua was born on 10 June 1968 in Bongouanou, Côte d'Ivoire. After gaining his Baccalaureat in mathematics and natural sciences at high school, he entered the National Polytechnic Institute in Yamoussoukro, Côte d'Ivoire, in 1989 and graduated in Computer Science in 1992. He worked from 1992 to 1995 as a system analyst and programmer, involved in the development of software for private and public organizations in Côte d'Ivoire. In 1995, he attended a specialization course on Computer Graphics and Interactive Multimedia in Groningen and, in 1997, received his Master of Science degree in Instructional Systems Design, with a specialization in Human-Computer Interaction and Courseware Engineering, at the University of Twente in the Netherlands. Since 1998, he has been involved in the FHA regional project for West and Central Africa in collaboration with the Payson Center for International Development and Technology Transfer, Tulane University, and has participated in several regional development projects. He has been actively involved in operational research, institutional development and capacity building, information systems development, instructional technology, and database and GIS development for public health mapping and monitoring in West and Central Africa. In May 2001, he started his PhD research at ITC.