

CONFERENCE OF EUROPEAN STATISTICIANS

**Joint UNECE/EUROSTAT Work Session on Methodological Issues Involving the Integration of
Statistics and Geography**
(Tallinn, Estonia, 25-28 September 2001)

Topic (ii): New technological solutions, including those based on online data access

ORGANIZATION OF MAPPING SERVICES ON THE INTERNET: POSSIBLE SOLUTIONS

Submitted by German National Research Center for Information Technology¹

Contributed paper

I. INTRODUCTION

1. Cartographers have long adhered to the view of the map as primarily a medium of communication (MacEachren 1995). Recently, influenced by ideas of scientific visualization and exploratory data analysis, they started to pay more attention to the role of the map as a tool to support visual thinking and decision making (MacEachren 1994, MacEachren and Kraak, 1997). To play this role effectively, the map needs two principal additions: interaction and dynamics.

2. Within the AiS Knowledge Discovery research group we are developing Java based tools for exploratory thematic mapping of spatially referenced data. Different Java applets were developed for support of analysis of various types of spatially related information. In the second section of this paper we shall discuss three main approaches to geographic visualization on the Internet:

- i) Server-based architectures;
- ii) Client-based architectures;
- iii) Different variants of client-server architectures.

3. In the first approach, static raster images are generated on the server side and then transmitted to the client. This approach is traditionally used in "Internet Map Server" solutions of traditional GIS companies. It has very strong limitations on the interactivity and dynamics of maps, and to our way of thinking is practically not usable for exploratory data analysis.

4. Since 1996, we are developing various Java-based GIS applets with client-server architectures. In the third section of the paper we shall describe two such variants, namely Descartes and CommonGIS, where the latter can be seen as a successor of the former. Both systems support the automatic design of thematic maps, and they are able to support visual data exploration using interactive manipulation of screen maps. Yet they exploit two different variants of client-server architectures for web mapping.

II. ARCHITECTURES FOR WWW MAPPING

5. Most GIS vendors propose solutions for Internet mapping on the basis of their existing packages. The typical architecture involves a GIS running as a CGI script (Grossley & Boston, 1996, Plewe, 1997) on an Internet server computer (see Fig. 1). For example, versions 2 of the CIESIN Demographic Viewer

¹ Prepared by Gennady Andrienko, Natalia Andrienko and Hans Voss.

and others (for example, formerly available Canadian census WWW gateway, and Pennsylvania Statistics) have similar architecture. The interaction between the user and the system is done through CGI forms displayed by a web browser on an Internet-connected client computer. CGI forms provide a rudimentary graphical user interface. A user's input by means of a form is transformed into commands to the GIS. In response, the GIS generates a map and sends it to the client computer as a raster image, typically in GIF format. The limitations of this approach are obvious:

- restricted user interface;
- low interactivity (even a standard operation like zooming may not be very comfortable to use);
- insufficient quality of graphics (limited not only by the speed of the Internet connection but also by some restrictions of the GIF format);
- substantial delays when large raster images are sent from the server to the client in each transaction.

6. Simple interactivity for server-side mapping can be implemented in the JavaScript language. A script may enable a user to select one out of a set of previously prepared GIF images to be shown on the page. For example, if somebody wants to highlight one of 10 objects presented on a map, he should prepare 11 maps: one without highlighting and 10 highlighting one object each. Once all the images have been downloaded to the client computer the script allows reasonably fast interaction with them on the screen. In the same way Peterson (1999) implemented simple dynamic mapping and animation. The advantage of this approach is that JavaScript programs can be interpreted by most browsers. The language is very simple and can be used by non-programmers. However, visualizations are limited by raster GIF files, and interaction capabilities are very restricted.

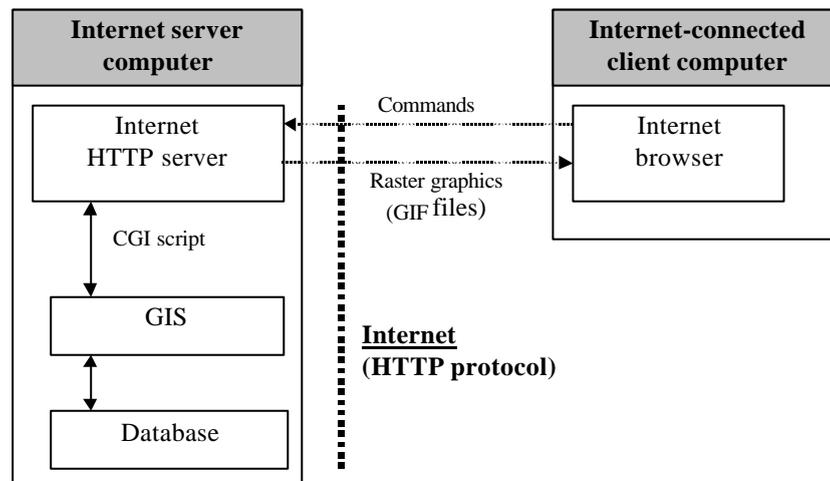


Fig. 1 Internet mapping: server-side GIS running as CGI script

7. Another solution consists in the use of Java applets (Fig. 2). Java is a programming language that can be interpreted by all common Internet browsers on various hardware platforms. Java applets embedded in web pages are loaded by a browser from an Internet server and then executed on a client without requiring any installation. Applets can access data stored on the server through the HTTP protocol. Advantages of this solution are the opportunities for providing an advanced user interface, unlimited user interaction, and generation of high-quality graphics.

8. This approach also has certain shortcomings. First, a critical issue is the size of the Java applet: downloading a big applet may take a long time. This may seriously constrain the functionality to be provided to the client. Secondly, applets may not run fast enough on some hardware and software platforms. Thirdly, data security is not guaranteed: any other software can access database files accessed

by the applet. Although these files can be encrypted, this is not a perfect solution: it is possible to open the encryption mechanism through decompiling the code of the applet. For these reasons most existing mapping applets only provide very basic map-viewing capabilities, and they are not used for real problem solving (Sorokine & Merzliakova, 1998). On the pro side we can observe that the general increase of speed of the Internet allowed the placing of more functionality into Java applets. For example, look at a collection of spatio-temporal visualizations at the URL <http://borneo.gmd.de/and/timemaps.html> and <http://borneo.gmd.de/and/time/> (Andrienko, Andrienko and Gatalsky, 2000).

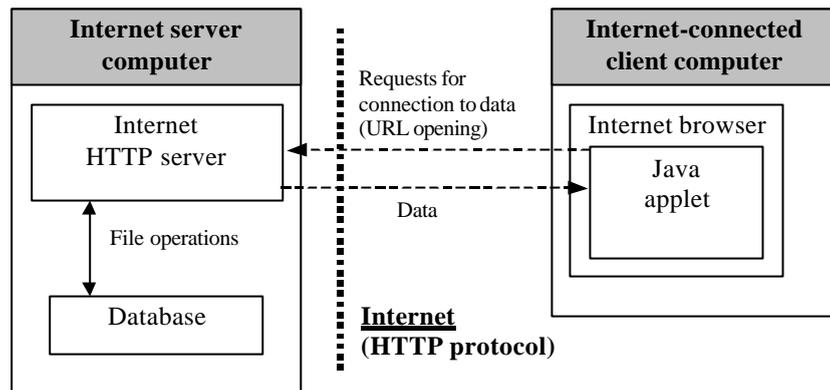


Fig. 2 Internet mapping: client-side GIS Java applet

9. Sometimes, instead of Java, other interpretable programming languages such as Tcl/Tk, are used. An example is the system for visual analysis of spatially referenced data named CDV (Dykes, 1997 & 1998). In such a case, to make the system work on the Internet, special software called browser plug-ins must be installed on the client computer. This requirement may actually be a serious limitation because end users may have insufficient competence to carry out the installation or be unwilling to do it.

10. In general, the available mapping software is rather weak with regard to the cartographic presentation of spatially referenced (statistical) data. Most GIS developers seemed to consider this function as marginal. This situation seems to be changing now that the value of GIS for geo-marketing applications has clearly come into focus. In contrast, our thematic mapping software Descartes, the first developments of which reach back over more than six years now, was primarily designed to support visualexploration of spatially referenced data through automated cartographic visualization. Correctness of map presentations, i.e. compliance with established principles of graphic and cartographic design, is very important for proper data interpretation and, as a consequence, for problem solving. As it cannot be presumed that all potential end users are familiar with these principles, Descartes incorporates such knowledge about thematic data mapping, and on this basis acts itself as a professional cartographer providing users with automatically built maps that appropriately visualize the user-selected data.

11. The maps built by Descartes are not just static pictures. The system offers a wide range of interactive tools to manipulate the presentations in order to make them more expressive and to create more beneficial conditions for revealing significant relationships, patterns, and regularities.

12. The server of Descartes provides database access, data manipulation (fulfilment of queries, calculations, transformation of structure of tables etc.), and map design (selection of presentation methods and setting of their parameters using the internal knowledge base on thematic mapping). The client provides the user interface, draws the vector-based maps, and supports interactive manipulation of these maps.

13. The client is a Java applet implemented in Java 1.1.8. Most Internet browsers such as Netscape versions from 4.0 and Internet Explorer versions from 4.0 support this version of Java. When a user loads the Internet page with the applet, it activates the server part of the system and starts communication with it through a socket connection. This approach permits the combining of high user interactivity and dynamic data displays on the client side with powerful functionality on the server side while the size of the applet can be kept relatively small. A general scheme of Descartes' functionality is presented in Fig. 3.

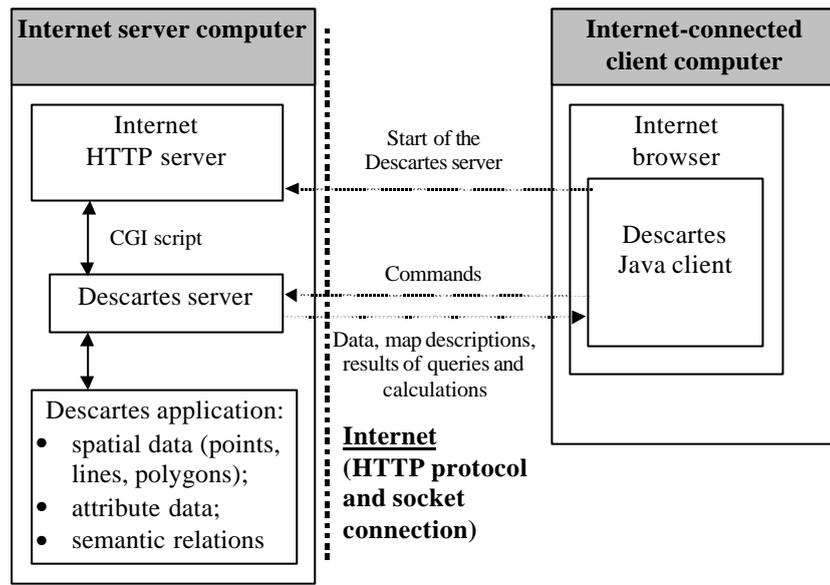


Fig. 3 Internet mapping in Descartes: client-server architecture with the client in Java

14. However, this approach also exhibits a drawback: some local networks are protected by firewalls that prohibit socket operations and thus Descartes applications could not be run across the firewall. Another problem is access to large data sets. The existing implementation with full loading of all data to the client seriously limits applicability and speed of the system.

15. To solve these problems, a new variant, actually a successor of the system, was developed. The name of this successor is "Descartes XXI", or also "CommonGIS", according to the EC-funded project with the same name which supported a large part of the development of this new version (CommonGIS: Esprit project 28983, November 1998 - June 2001). The functionality is again distributed between client and server (both components are implemented in Java). The server only runs on demand, i.e. when the client needs some new data (for example, as result of shifting or zooming a map). In this case, the client starts a new copy of the server software and passes full context information to it (loaded map and table, map extent, derived attributes, etc.). Respectively, the server sends back data only for specified regions and topics. The system can be configured: depending on parameters and needs of an application, some components may be either included into the server, or into the client, or in none or both (see Figure 4).

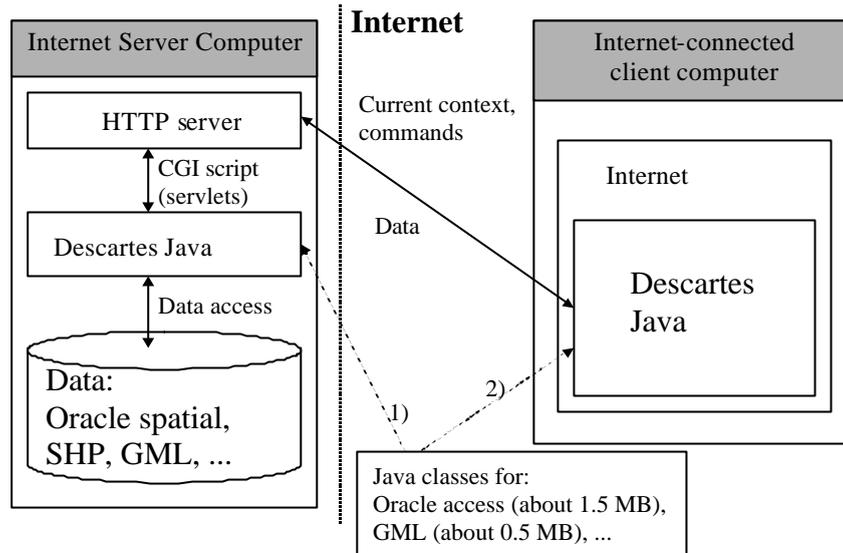


Fig. 4 Internet mapping in Descartes XXI / CommonGIS: client-server architecture with the server starting on demand

III. CLIENT-SERVER SYSTEMS FOR WEB MAPPING

16. We will address below the potential of interactive web mapping by having a closer look at the systems Descartes and CommonGIS respectively. It should be borne in mind that most current GIS specifically support the analysis and manipulation of spatial objects, but also exhibit some shortcomings with regard to thematic mapping support:

- Their users bear full responsibility for a correct selection of presentation techniques. Consequently, they must proficiently know principles of cartographic presentation, have to spend considerable time and effort in map production, and cannot entirely concentrate on their task of analysis and problem solving;
- GIS are typically complex and difficult to operate; significant training is required;
- GIS do not sufficiently support the further exploitation and exploration of produced maps, i.e. their use in data analysis.

17. Both, Descartes and CommonGIS overcome these shortcomings. They provide an intelligent mapping support and a full spectrum of functions for interactive visual data analysis. Some traditional GIS functions (manipulation of layers, data access through a map, etc.) are also available.

18. From an architectural viewpoint, they have communicating client and server components. Descartes uses a server which runs permanently while the client is active. A CommonGIS client starts its server each time for every new communication.

III.1 Descartes: Knowledge-based System for Visualization Design

19. Descartes was created in 1996 on the basis of our earlier system Iris which was also a knowledge-based system for map design implemented in C++ for Windows (Andrienko and Andrienko 1997). The detailed description of Descartes is available in the paper (Andrienko and Andrienko 1999a). Demonstrators of Descartes with dozens of applications (example data sets) are available from several servers:

- <http://borneo.gmd.de/and/java/iris/> at GMD AiS.KD, Sankt-Augustin, DE

- <http://www.mimas.ac.uk/descartes/> at KINDS (Knowledge Interfaces to National Data Sets) Project, Manchester, UK
- <http://geolibrary.uidaho.edu/> at University of Idaho, Moscow, Idaho, USA

20. The server of Descartes (a C++ program available for various dialects of Unix and Windows) is used for accessing and transforming data and for designing visualizations on the basis of data characteristics and relationships. The user interface (client) of Descartes is implemented as a Java applet. It provides table display, map displays (most of traditional thematic mapping styles are implemented), supplementary graphics displays (dot plots, scatter plot, box plot, parallel coordinates, etc.), and supports interactive manipulation with maps and graphics.

21. Descartes includes many map manipulation techniques such as highlighting, brushing, visual comparison and interactive classification (Andrienko and Andrienko 1999a), cross-classification (Andrienko and Andrienko 1999), classification on the basis of dominant attribute (Andrienko and Andrienko 2001), etc. Database querying and mathematical calculations are also supported. Their results can be immediately visualized on dynamic maps.

22. Descartes uses so-called data characterization for visualization design, i.e. it performs best when certain information about the semantic characteristics of the data (meta data) are provided. The characterization of the data is done during the process of building an application. In general, to create an application, one should:

- Import geographical data into Descartes' internal format which has a structure isomorphic to the format of ArcView's shape files, but is about twice as compact due to the use of less precise float numbers. Descartes includes a variety of converters from most popular GIS data formats: SHP and E00 by ESRI, MIF/MID by MapInfo, DXF by AutoCad, etc.;
- Import table data into DBF format which is used by Descartes;
- Link records of the DBF file to contours or points of some layer in geographical data set. The linkage is done through common identifiers. This means that the table contains a column with identifiers of spatial objects the data in each record refer to. The same identifiers are indicated in the file with coordinates or contours of the spatial objects;
- Describe the semantics of the table data (see (Andrienko and Andrienko 1999b) for a more detailed description and full list of data characteristics). This is basically done by grouping of columns referring to the same attribute and then by specification of the attribute. The columns referring to the same attribute are differentiated by values of a parameter. For example, columns with population division refer to different values of parameters "gender" and "age group".

23. The process of making an application is supported by a special Java applet called Application Builder, which interactively guides the user through the whole process, and permits the viewing of currently constructed maps and tables at almost any time during the process.

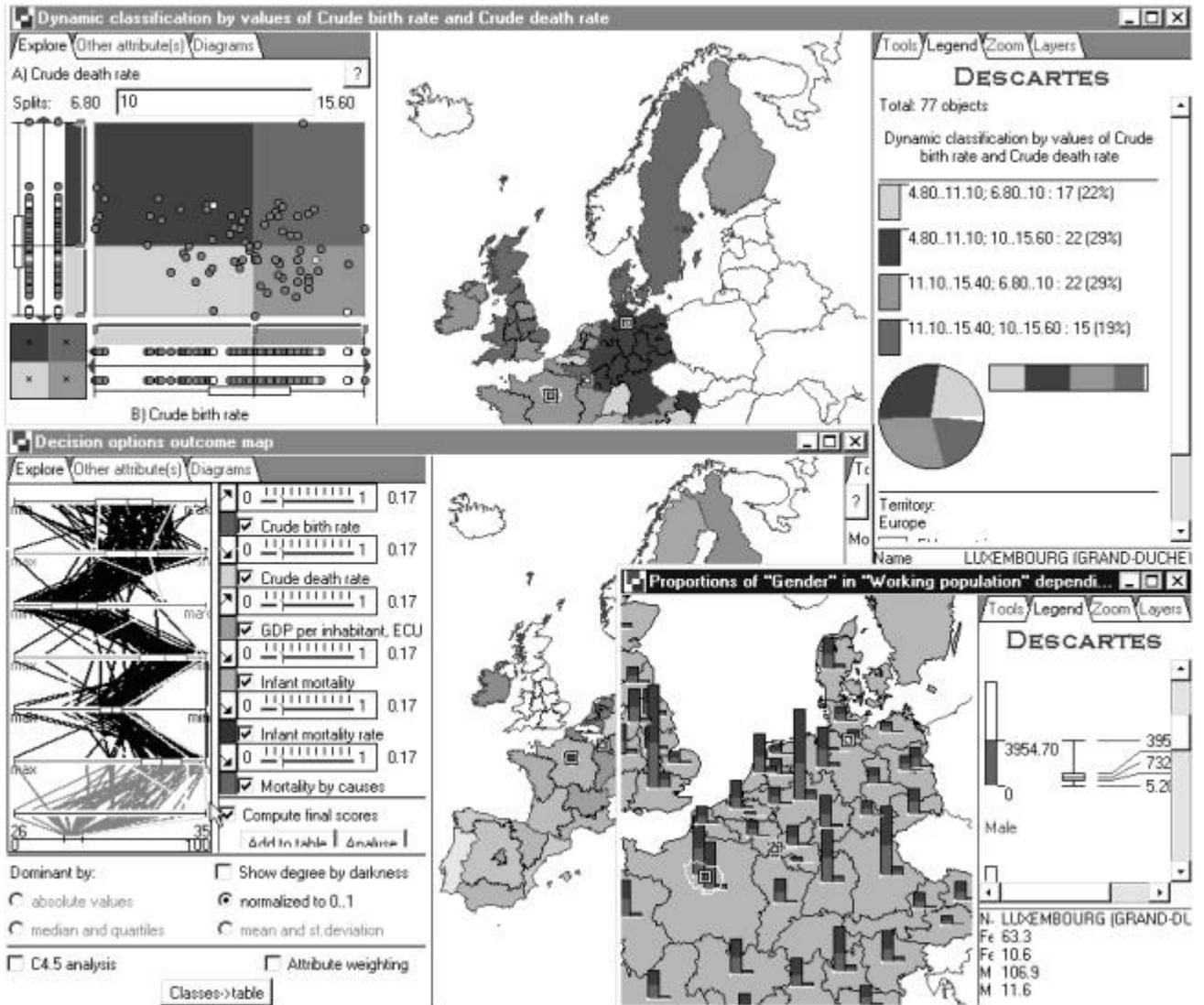


Fig. 5 Multiple window user interface of Descartes: maps and statistical graphics linked by dynamic highlighting and brushing

24. The main shortcoming of Descartes is that it uses the socket connection technology for communication between the server and the client and, as discussed above, its functioning is sometimes prohibited by firewalls. This problem was solved with CommonGIS.

III.2 CommonGIS - Common Access to Geographically Referenced Data

25. Aside from the firewall problem the CommonGIS system also solved another problem, namely providing access to map and table data stored in relational databases. It thus provides:

- ways to work with distributed data stored on multiple web servers;
- possibility of connecting to various relational databases;
- efficient work with large data sets due to optimization of access to data;
- relatively easy extension through adding user defined types of spatial objects.

26. The CommonGIS system can be directly applied to most standard map and table data formats. In particular, it can work with any ODBC/JDBC-compliant database, such as Oracle 8i, SHP files (ArcInfo), GML files (Geography Mark-up language by OpenGIS consortium), WKB implementation of the simple-

features specifications (OpenGIS consortium) and others. Vector maps can be combined with raster (JPG and/or GIF) and with visualization of grid data. It is possible to load thematic data directly from MS Word/Excel via clipboard. The system contains a dialog module that acquires data semantics from users on demand, i.e. at the moment when the semantics are important for the visualization design.

27. A demonstrator of the CommonGIS prototype is available at the GMD server <http://borneo.gmd.de/descartes/CommonGISApplet/>

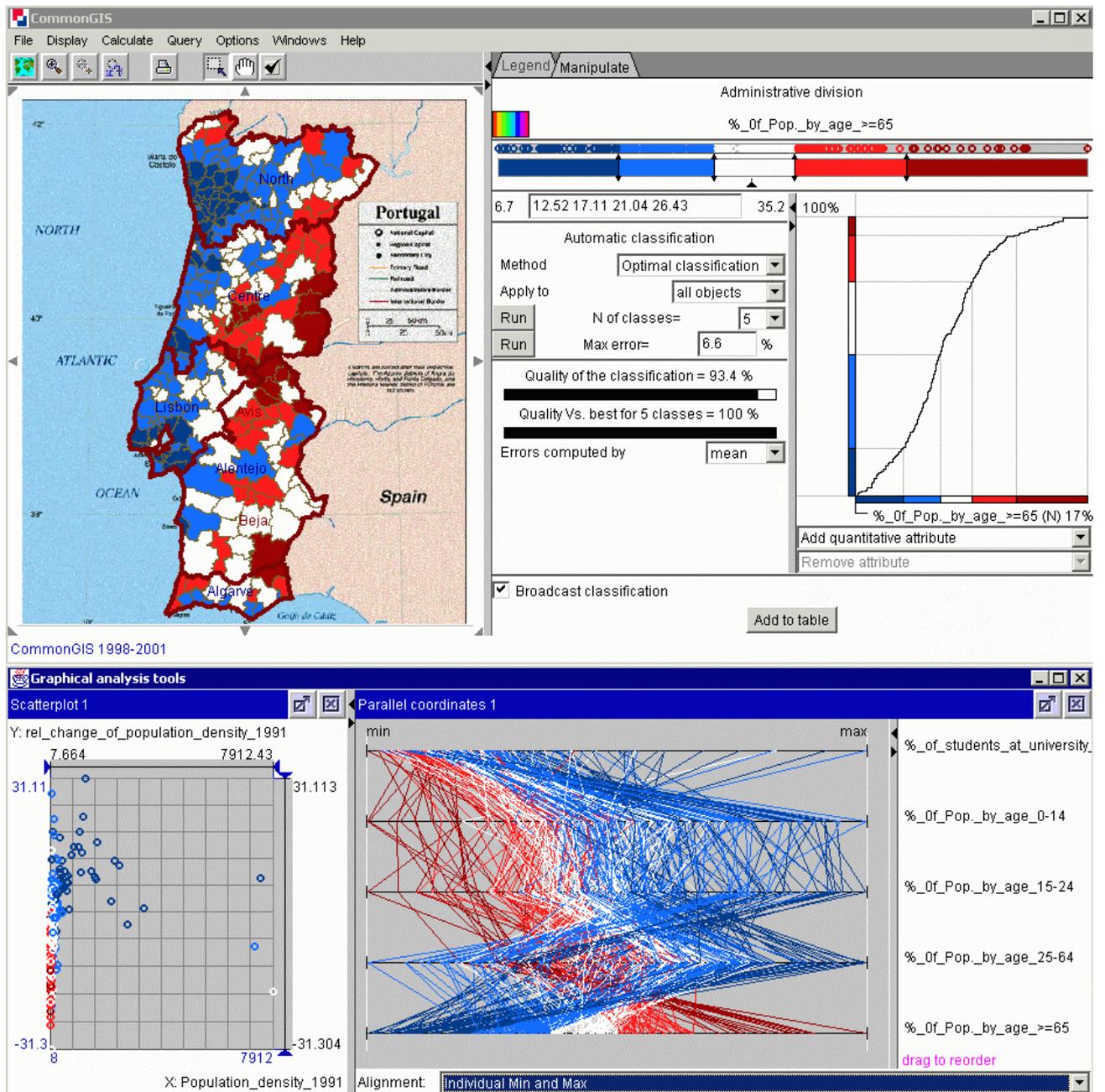


Fig. 6 User interface of the CommonGIS system

IV. CONCLUSIONS

28. Development cycles in the web mapping world are fast. Yet we have still seen only a few truly interactive thematic mapping systems. This may be so because plug-ins are generally not considered a solution for many customers, and Java applets are also not yet sufficiently trusted in terms of security, efficiency, or other reasons. Partially, we would claim, this is so because the true value of real interactive mapping has not yet been conceived.

29. The effective usability of Descartes, and its successor CommonGIS, for exploratory data analysis and decision-making has been confirmed by many who have used it. For example, Descartes/CommonGIS is now provided as the web-based exploratory tool for the UK MIMAS project, which makes UK census data available for educational purposes at universities (<http://www.kinds.ac.uk/kinds/>). Or, an extension of Descartes was built and applied in order to visualize part of the thematic data from the New Cronos database of Eurostat with time-series animation.

REFERENCE

1. Andrienko, G. and Andrienko, N. (1997). Intelligent Cartographic Visualization for Supporting Data Exploration in the IRIS System, *Programming and Computer Software*, 23 (5), pp.268-282
2. Andrienko, G. and Andrienko, N. (1999a). Interactive Maps for Visual Data Exploration. *International Journal Geographical Information Science*, 13 (4), pp.355-374
3. Andrienko, G. and Andrienko, N. (1999b) "Knowledge Engineering for Automated Map Design in DESCARTES", In C.B.Medeiros (ed.), *Advances in Geographic Information Systems. Proceedings of the 7th International Symposium ACM GIS'99, Kansas-City, November 5-6, 1999, NY: ACM Press*, pp.66-72
4. Andrienko, N., Andrienko, G., and Gatalsky, P. (2000). Visualization of Spatio-Temporal Information in the Internet. In A.M.Tjoa, R.R.Wagner, A.Al-Zobaidie (eds.) *Second International Workshop on Web-Based Information Visualization: WebVis 2000, Greenwich, London, UK, 48 September, 2000. 11th International Workshop on Database and Expert Systems Applications, Proceedings. IEEE Computer Society, Los Alamitos, California, 2000*, pp.577-585
5. Andrienko, G. and Andrienko, N. (2001). Exploring spatial data with dominant attribute map and parallel coordinates, *Computers, Environment and Urban Systems*, 25 (1), pp.5-15
6. Dykes, J.A. (1997). Exploring Spatial Data Representation with Dynamic Graphics. *Computers & Geosciences*, 23, 345-370.
7. Dykes, J.A. (1998). Cartographic Visualization: Exploratory Spatial Data Analysis with Local Indicators of Spatial Association using Tcl/Tk and cdv. *The Statistician*, 47, 485-497
8. MacEachren, A.M. "How Maps Work: Representation, Visualization, and Design" (NY: The Guilford Press), 1995
9. MacEachren, A.M. "Visualization in modern cartography: setting the agenda". In *Visualisation in Modern Cartography* (NY: Elsevier Science Inc.), pp.1-12, 1994
10. MacEachren, A.M. and Kraak, M.-J. "Exploratory cartographic visualization: advancing the agenda". *Computers and Geosciences*, 23 (4), pp.335-344, 1997
11. Grossley, D. and Boston, T. (1995). A Generic Map Interface to Query Geographical Information using the World Wide Web. *World Wide Web Journal*, 1996, 723-735
12. Peterson, M.P. (1999). Active Legends for Interactive Cartographic Animation. *International Journal Geographic Information Science*, 13, 375-383
13. Plewe, B. (1997). *GIS Online. Information Retrieval, Mapping, and the Internet*. Santa Fe: OnWorld Press.
14. Sorokine, A. and Merzliakova, I., 1998, Interactive Map Applet for Illustrative Purposes. In *Proceedings of 6th International Symposium on Advances in Geographic Information Systems* (pp.46-51), New York: ACM Press.