ELSEVIER

Interactive maps for visual exploration of grid and vector geodata

Gennady Andrienko[a,*], Natalia Andrienko[a], Valeri Gitis[b]

[a] *Fraunhofer AiS—Autonomous Intelligent Systems Institute, Schloss Birlinghoven, Sankt-Augustin D-53754, Germany*
[b] *IITP—Institute for Information Transmission Problems RAS, Bolshoi Karetnyi Lane, 19, 101447 Moscow, GSP-4, Russia*

Abstract

This article considers two types of tools used for the analysis of spatial data: computational tools for grid data and visualization-based interactive exploratory techniques. The main purpose is to demonstrate that computational and visual techniques are complementary, and their combined use in data analysis may produce a synergistic effect. To this end, we describe two example scenarios of data analysis in the evaluation of potential damage from earthquakes and in analysis of forest resources. The analyses were performed with the use of CommonGIS, featuring the availability of powerful geocomputational tools and interactive visualization-related facilities that can be efficiently used in combination in a desktop environment or via the World Wide Web.
© 2003 Elsevier Science B.V. All rights reserved.

1. Grid-based data analysis: goals and tools

Traditional approaches to the representation and processing of geodata in geographic information systems (GIS) assume that data are organized into semantically uniform layers, or themes. For example, data about a territory may consist of such layers as elevation, rivers, roads, cities, and economic districts. From the viewpoint of data representation, vector and grid-based layers are distinguished. Vector layers consist of discrete geographic objects in the form of points, lines, or areas that are specified by their coordinates. The objects may be characterized by various attributes; that is, there is a correspondence between the objects and particular attribute values. Grid-based data consist of values of an attribute specified in nodes of a regular rectangular grid. This format is most often used for conveying information about continuous spatial phenomena, such as surface elevation models, earth gravity, or magnetic fields. In addition, grid-based data are often used for the representation of derived information about spatial and spatio-temporal properties of discrete geographical objects, e.g., density of roads or communication lines, parameters of seismic flow (density of earthquakes, seismic activity, fractal dimension), and statistical parameters of accidents.

Further spatial analysis of grid data can yield predictions and expected trends. In seismologic applications, for example, grid analysis has resulted in predictions of earthquake time, coordinates, and

* Corresponding author. Tel.: +49-2241-142486; fax: +49-2241-142072.
E-mail address: gennady.andrienko@ais.fraunhofer.de (G. Andrienko).

energy (Gitis et al., 1994), prognosis of maximal possible magnitudes of expected earthquakes on the basis of geological and geophysical characteristics of the earth crust (Gitis, 1995), estimation of natural process parameters by geomonitoring time series (Gitis et al., 1997), and assessment of total economic loses resulting from natural and man-made disasters.

For analyzing grid data, particularly in combination with vector data, the following classes of analysis tools are often applied:

- Tools for the transformation of grid and vector data;
- Tools for plausible inference of spatial and spatio-temporal relationships;
- Visual (map-based) data exploration tools;
- Tools for explanation and validation of results.

In the following sections of this paper, we consider tools for data transformation (Section 2) and visual exploratory analysis (Section 3). In Section 4, we demonstrate synergistic use of the two groups of tools by example scenarios of grid and vector data analysis in different domains: evaluation of possible damage caused by an earthquake and analysis of forest resources in Europe. The analyses have been performed using our software system, CommonGIS, which can be accessed via the World Wide Web. In conclusion, we outline the directions for the planned future work on extending the analytical capabilities of CommonGIS.

2. Analytical transformations of grid and vector data

Possible operations for use with grid and vector data may be classified according to the following scheme.

1. Grid(s)⇒Grid. Change of resolution of a source grid, filtration of grids, building illumination models or gradients, flood modeling, user-defined pointwise transformations from several source grids to a new one.
2. Grid⇒Vector objects (geometry). Building isolines, triangulation or detection of polygons, as in deriving a coastline from elevation data.

3. Grid⇒Attributes of (existing) vector objects. Finding minimum/mean/maximum elevation for areas (units of territory division) or linear objects such as railways, oil and gas pipes, or power and communication lines; calculation of areas covered by parts of a raster that satisfy conditions of a query, for instance, percentages of territories of administrative districts where elevation exceeds 1000 m.
4. Vector data (geometry and, possibly, attributes)⇒ Grid. Calculation of {weighted} densities of events on the basis of point (individual events) or line (isolines) data, calculation of distances to sets of objects, buffer zones.
5. Vector objects⇒Vector objects (geometry). Building new vector objects from existing ones by means of such operations as intersection, union, subtraction, etc.
6. Vector data (geometry or/and attributes)⇒ Attributes of vector objects. Attributes may be derived from only geographic information (e.g., computation of lengths and areas of vector objects, counting objects fitting in areas), only values of attributes (transformation of absolute population numbers in different age groups into proportions) or from a combination of geographic information and attribute values (transformation from absolute population numbers into population densities by dividing population by area).

One can see that these operations have different computational complexity. Some of them produce easy-to-understand results, while outputs of other methods require support in interpretation. Generally, results of most analytical transformations should be visualized, i.e., presented in an easy-to-perceive form. At the same time, visualization is a powerful tool in its own right. It can effectively support, in particular, exploratory analysis of spatial data and phenomena. It can be used to empirically search for patterns and characteristic features of their spatial variation and relationships with other phenomena. Visualization of spatial information necessarily requires map displays, which can convey to the human eye important spatial features and relationships. Modern computers allow graphical data displays to be interactive and transformable in order to better support visual data analysis and exploration. A possibility to access both data and

visual data exploration system via the Web creates new exciting opportunities for analysis and problem solving.

The advantage of CommonGIS, in comparison with other systems supporting processing of vector and grid data, is the combination of geocomputational facilities (mainly ported from the system GeoProcessor, see Gitis et al., 1999), with advanced visualization-based interactive tools for exploratory analysis of spatial information. These two types of instruments can aptly complement each other in the course of data analysis.

3. Visual analysis of grid and vector data

In CommonGIS, we propose a set of interactive techniques for manipulating cartographic displays of spatial information that were specially designed for facilitating visual exploration of numeric grid data. Some of these techniques were earlier implemented for visual exploration of attribute data referring to vector objects (Andrienko and Andrienko, 1999).

Grid data are usually displayed in maps as raster images where colors of pixels encode attribute values. Attribute values and colors for pixels falling between grid nodes are determined through interpolation. A similar procedure of color encoding of numeric values is applied in producing choropleth maps used to represent values referring to areas of territory division. For choropleth maps, we have earlier designed and implemented a number of interactive techniques that allow the user to dynamically alter the color encoding and thereby produce various benefits that make it easier to locate particular values, simplify comparison of values, or expose initially unclear spatial patterns. The same techniques are applicable to grid data.

One of the interactive exploratory techniques is the "visual comparison" operation applied to the encoding of numeric attribute values with the use of the so-called diverging or double-ended color scale. Such a scale uses two different color hues for encoding values below and above a certain number, the so-called midpoint (Brewer, 1994). The distance of a value to the midpoint is proportionally encoded by the degree of darkness. A special third color is used for the midpoint value.

The "visual comparison" tool allows the user to move the midpoint of a diverging color scale and immediately observe the effect of this operation on the map. Visual comparison on choropleth maps is described in Andrienko and Andrienko (1999). The recent version of the system includes a similar tool for grid data as illustrated in Fig. 1. On the top, one can see a representation of a digital elevation model (the area of western Turkey) by a diverging color scale in which shades of blue are used for elevations below zero and shades of yellow for values above zero. Zero (i.e., the sea level) is the midpoint in this case. In the center, the midpoint value has been changed to 50 m; at the bottom, it is equal to 600 m. This operation may be used, in particular, to see what territory will be flooded if the sea level rises. In a more general case, especially when there are no natural thresholds in the data under analysis, the "visual comparison" operation may help an analyst to locate very high or very low values and detect patterns and trends in spatial variation of the data.

Another interactive technique applicable in both choropleth and grid mapping is manipulation of value classification. There are two principal opportunities for encoding numeric attribute values by colors. One is direct transformation of each individual value into a color so that, ideally, different values become represented by distinct colors (in practice, however, close values are often visually indistinguishable). For example, values may be encoded by proportional degrees of darkness. A variant of such encoding is the representation using shades of two colors, as in the diverging color scale shown above. The other opportunity is to divide the value range of the attribute into intervals and assign a distinct color to each interval. Hence, attribute values fitting in the same interval are identically represented. The latter procedure is known as classification and the intervals are called classes.

Classification, when implemented as an interactive, dynamic procedure, is a valuable instrument of exploratory analysis of spatial data (Andrienko et al., 2001b). Earlier, we have developed an interactive tool for classed choropleth maps that allows the user to add, move, and remove class breaks, with any slight change being immediately reflected on the map. Recently, we have implemented a similar tool for grid data. Fig. 2 shows how the tool operates. At the top, elevation values are divided into three classes
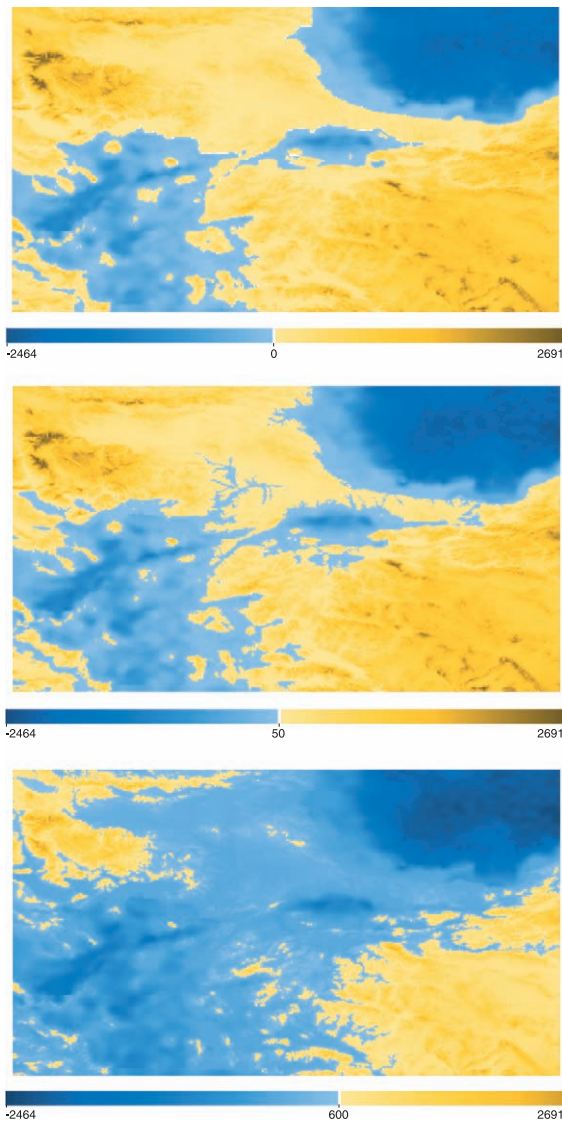
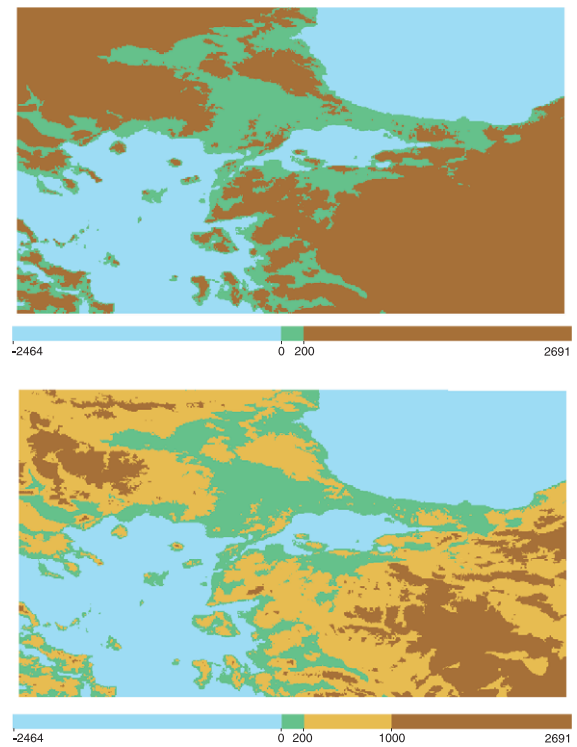Fig. 1. Moving the midpoint of a diverging color scale.

Fig. 2. Dynamic classification: adding a new class break changes the spatial pattern that may be perceived from the image.

One more interactive technique initially designed for attributes of vector objects but applicable to grid data as well is known as focusing (Andrienko and Andrienko, 1999). It allows the user to restrict the value range of an attribute that is represented on the map. The values that do not fit within the user-
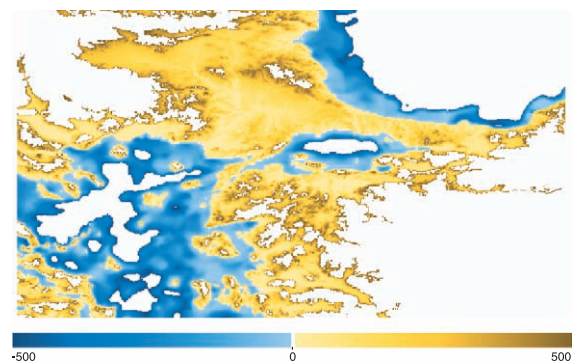

Fig. 3. Dynamic focusing: only elevations in range from − 500 to + 500 m are shown.

with class boundaries 0 and 200 m. At the bottom, the upper class has been further divided into two classes by introducing a new break, 1000 m. Classification may also be used for summarizing values of two attributes by cross-classification (Andrienko and Andrienko, 1999). For more than two attributes, the system applies classification by dominance among attribute values (Andrienko and Andrienko, 2001). We plan to apply similar techniques for combining data from two or more grids.

imposed limits are removed from view while the remaining values are shown with the maximum possible expressiveness. Thus, in a case of color encoding, the restricted number interval is matched with the whole color scale. This makes differences between close values more noticeable. In particular, the focusing tool may be used to remove outliers, i.e., very high or very low values, in comparison to the rest. Fig. 3 shows the effect of the application of focusing to grid data. Parts of the grid with elevations below -500 m and above $+500$ m have been hidden. As a result, it becomes possible to represent elevations between -500 m and $+500$ m with maximum distinctness.

Being applied to grid data, focusing has an additional effect. Hiding fragments of a raster image overlaying another data set uncovers the parts of the map beneath. This may help in investigations of relationships between the phenomenon represented by the raster and other phenomena that may be represented either by vector or grid data. For the same purpose, semitransparent display of a raster also may be used. In CommonGIS, it is possible to interactively vary the degree of transparency of a raster image representing grid data. The user may choose any transparency value from 0% (fully opaque) to 100% (fully transparent or invisible). In Fig. 4, the elevation model is overlaid with a raster image representing peak ground accelerations occurring due to an earthquake. The values are encoded using the scale of shades of gray where darker shades correspond to higher values. On the left, the image is opaque; therefore the elevations are not visible. On the right, the image is 40% transparent and the distribution of values in the two grids can be visually compared.

As seen in the examples presented herein, different color schemes may be used to visualize grid data. It should be noted that grids usually represent phenomena that are smoothly spread over a territory. This is opposite to choropleth mapping where values may be arbitrarily distributed and therefore produce the so-called patchwork maps. Smooth variations of values in grids make it possible to use more sophisticated color schemes than are usually applied in choropleth maps. For example, the traditional color scale for representation of elevations is built from several color bands with continuous increase or decrease of darkness within each individual band. For visualization of grid data, CommonGIS allows the user to choose from

a set of predefined color schemes or create her/his own color scale with an arbitrary number of color bands and customized variation of colors within the bands. For this purpose, different color spaces (RGB and HSB) and different functions for generating transition colors (linear, logarithmic, exponential) with variable granularity can be used. At any moment, the user may easily replace one color scale by another.

4. Examples of problem solving with grid and vector data

In this section, we describe the application of analysis and visualization tools to problems from two different domains: (1) evaluation of potential damage from earthquakes; and (2) analysis of forest resources in Europe. These applications clearly demonstrate the synergetic effect of combining interactive visualization tools with sophisticated methods of geographic computation.

4.1. Evaluation of possible damage from earthquakes

The first example demonstrates the use of CommonGIS for assessment of direct damage from a strong earthquake in western Turkey. The example is illustrative because the results were obtained on the basis of incomplete but available data.

Let us imagine that a decision maker needs to detect the cities in Turkey that may be seriously affected by a possible strong earthquake in order to allocate funding for seismic risk mitigation. For this purpose, data about population, buildings, and seismic activity need to be analyzed. The potential seismic impact is specified in the form of the grid of peak ground acceleration (α) built within the project GSHAP (Giardini et al., 1999). This model has been used for computing proportions of buildings with the lowest resistance to earthquakes that are likely to be heavily damaged-designated Type A according to the classification adopted in Turkey. To perform this analysis, two transformations $<\text{Grid} \Rightarrow \text{Grid}>$ were applied. First, a grid of maximum earthquake intensity $<I>$ was built according to the formula $\log(\alpha) = 0.3 * I + 0.014$ (Trifunac and Brady, 1975). The proportion of damaged buildings of the Type A, denoted here as $D(I)$, were then determined according

Fig. 4. Variation of image transparency.

to the empirically established nonlinear dependency (Coburn, 1995): $D(I \leq 5) = 0$, $D(5 < I \leq 6) = 0.05$, $D(6 < I \leq 7) = 0.5$, $D(7 < I \leq 8) = 0.75$, $D(8 < I \leq 9) = 0.9$.

The result of the computation is a new grid that may be visualized on a map. For exploring the spatial pattern of the potential damage, the user applies the interactive facilities for manipulating visual representation of grid data. The color scale is selected and, by applying the focusing tool, the map zero values (more precisely, values below 0.01, or 1%) are removed. As a result, the data appear on the map as a distinct area in which noticeable damage may be expected (Fig. 5). A prominent spatial pattern may be observed, with a thin west-to-east oriented strip and a wider crescent-shaped belt south of it. The user can make the layer semitransparent in order to see the relief in the endangered area.
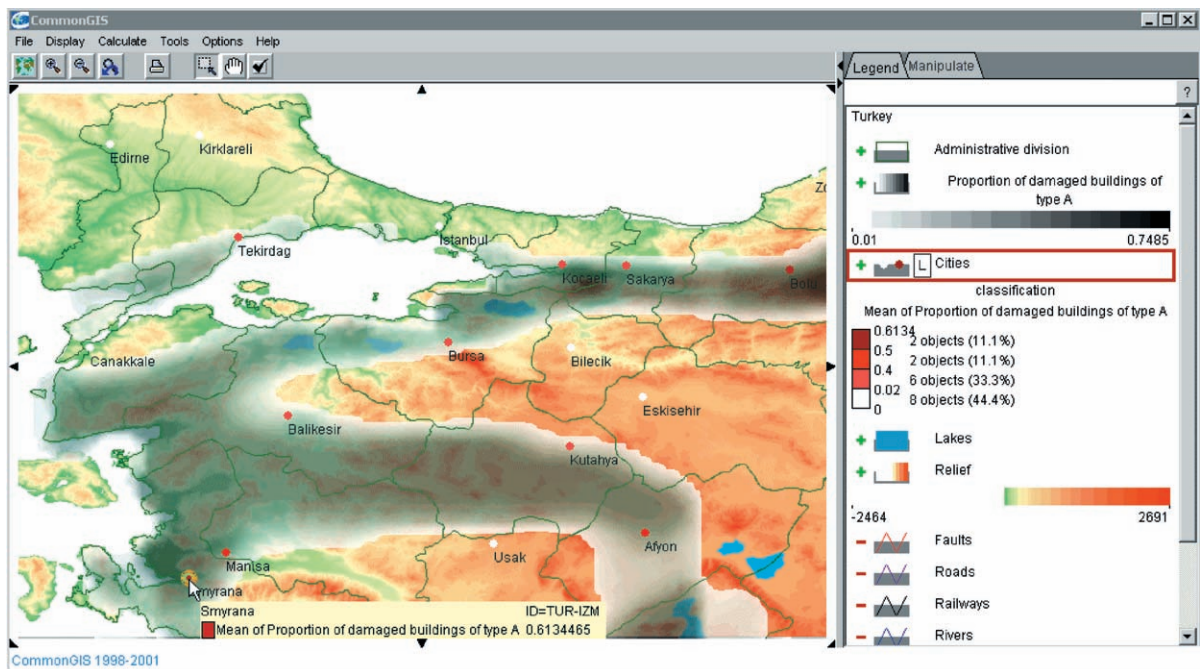

Fig. 5. The map of western Turkey showing relief, lakes, administrative division, and cities overlaid with a derived grid layer representing proportions of buildings of Type A that are likely to be heavily damaged as the result of a strong earthquake. The cities are classified according to the mean values of the damage.

By further applying the geocomputational tools of CommonGIS, the user finds for each big city (specified as a point object) the mean value of the proportion of damaged buildings within the radius of 10 km around the city center. This is a transformation of the type <Grid⇒Attributes of vector objects>. The user then applies the interactive classification tool in order to classify the cities according to the values of the so-derived new attribute. The mean proportion of damaged buildings varies from 0 to 0.61. The user defines four classes: cities with no or minor damage (from 0 to 0.02), cities with small to medium damage (from 0.02 to 0.4), cities with heavy damage (0.4 to 0.5), and cities with very heavy damage (over 0.5, or more than 50% of Type A buildings will be damaged). The distribution of the cities over the classes is $8-6-2-2$ or, in percentages, $44.4-33.3-11.1-11.1$. The classi-

fication results are represented on the map (see Fig. 5) with the color white used for minor damage, light red for medium damage, bright red for strong damage, and dark red for very strong damage. With the use of the map, the user can easily find the cities with the strongest expected damage. These are Smyrana and Bolu distinguished by the dark red color.

Additionally, the user requests the system to sum up the population number in each class and compute its proportion to the total urban population. The results of the computation are 46%, 32%, 13%, and 9%. Hence, 9% of the urban population of the area lives in two cities, Smyrana and Bolu, where more than 50% of Type A buildings are likely to be damaged due to a strong earthquake. These two cities are the most appropriate candidates for allocating the funding for seismic risk mitigation.
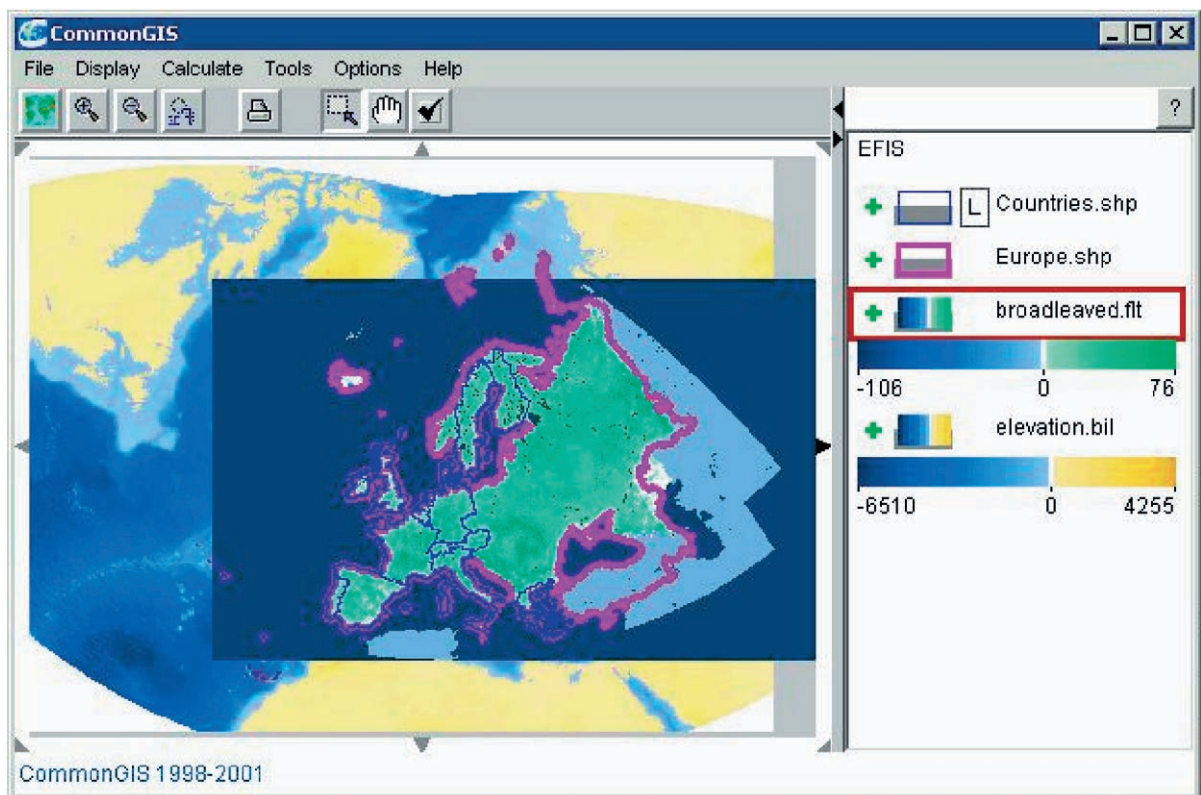

Fig. 6. Visualization of forest data using a diverging color scale exposes "special" values that encode coverage by water and absence of data, shown in dark and light blue, respectively.

4.2. Exploration of forest resource data

In this application, an analyst investigates the distribution of broadleaved forests in the countries of the European Union. Of particular interest are the forests situated in mountain areas. The data needed for the analysis are the elevation model of Europe and a grid of broadleaved forest densities derived from a satellite image. More precisely, each cell of the grid contains the percentage of the corresponding territory fragment covered by broadleaved forests. For Europe, the percentages vary between 0 and 76.

Visualization of the forest data using a diverging color scale makes the analyst notice that the data contain strange values: -1 (denoted on the map in Fig. 6 by light blue color) and -106 (dark blue). The values -106 occur where the land is covered by

water (seas and lakes) while the values -1 appear outside the territory of Europe and, apparently, denote absence of data. The analyst foresees that these special values may distort the results of the planned computations. For example, in calculating the mean densities of broadleaved forests by the countries of Europe, the values -106 occurring in lakes will be involved in the computations. Therefore, the analyst applies the tools for grid transformation available in CommonGIS in order to "clean" the data by replacing the negative values in the grid by zeros.

After cleaning the data, the analyst requests the system to perform the following query: find areas where broadleaved forests cover more than 10% of the territory and the elevation is 1000 m or more above the sea level. In the result, a new grid is built with values 1 in the cells satisfying the query and 0 where
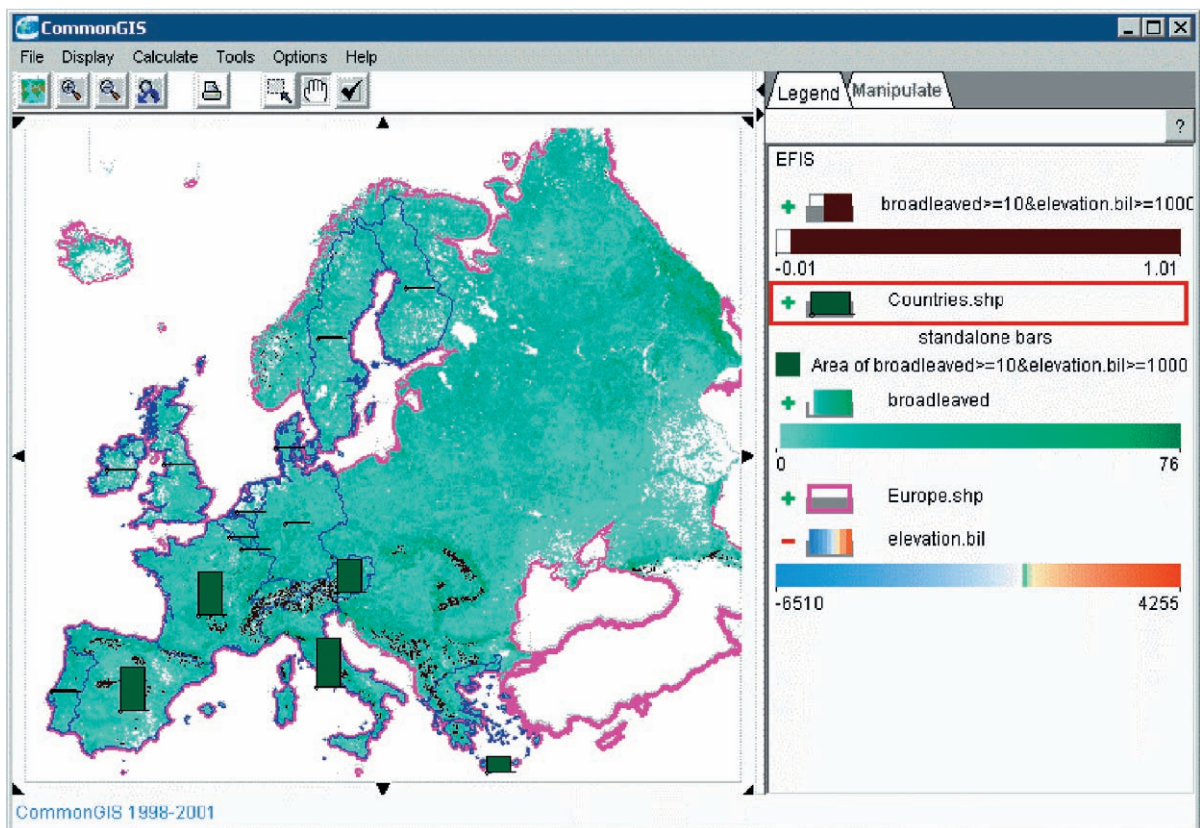

Fig. 7. Grid data about densities of broadleaved forests are represented by shades of green. The dark spots on the map mark the territories with elevation over 1000 m and more than 10% of coverage by broadleaved forests (a result of querying the data). The vertical bars for each country represent the total area of the territory satisfying the query.

the query is not satisfied. To show the new layer on the map, the system automatically chooses the so-called binary color scale, which represents nonzero values by some color while zero values are hidden. Hence, only territories satisfying the query conditions are marked on the map. In Fig. 7, the territories with more than 10% coverage by broadleaved forests occurring at elevations of more than 1000 m are shown in black.

Finally, for each of the EU countries (specified by their vector boundaries), the analyst lets the system calculate the total size (area) of its territory satisfying the query. The system produces a new attribute that can be visualized on the map, e.g., by bars, as shown in Fig. 7. The heights of the bars are proportional to the sizes of the territories satisfying the query. It may be seen that only four countries have notable amounts of broadleaved forests situated in mountain areas: Italy, Spain, France, and Austria.

These two example scenarios of problem solving demonstrate that visual analysis of spatial (grid and vector) data supported by appropriately designed interactive tools aptly complement the use of computational tools for data processing. The advantageous feature of CommonGIS is the combination of computational and visual techniques for analysis of spatial data.

5. Software implementation

The system CommonGIS is implemented in the Java programming language, and thus can work both locally on a computer where it is installed and over the Internet as an applet. The system requires significant amounts of memory to be available because grid data files are typically very large. Therefore, for running the system via the Internet, it is recommended to use the Java plug-in instead of the native Java implementations available in the major Web browsers. The plug-in supports the latest release of the Java language and makes it possible to increase the amount of (virtual) memory available to an applet.

A system demonstrator with earthquake data is available at the URL http://www.ais.fraunhofer.de/and/EFISApplet2002/marmara.html. The forest data analysis application can be accessed at the URL http://www.ais.fraunhofer.de/and/EFISApplet2002/. Those who would like to test the tools of the system may need to be patient while the system loads large amounts of grid data over the Internet. CommonGIS can be downloaded for free for research and education purposes from the project web site, URL http://www.commongis.com.

6. Conclusion and directions for further work

In the future, we would like to extend the analytical facilities of CommonGIS with respect to grid data by adding noncartographic visualization techniques. For example, a very useful tool for revealing correlation between attributes is the scatter plot widely used in statistics. We expect that this display can be used for investigating relationships between two phenomena represented as grids. Of course, it is impossible to show on a scatter plot values referring to all locations since a grid corresponds to a continuous fragment of territory with an infinite number of locations. It is possible, however, to represent values for a finite number of sample locations. The sample locations can be automatically selected with regular (user-specified) spatial intervals between them. By varying the intervals, the user changes the level of detail of the derived discrete data represented on the plot. The user may also interactively select specific locations to be added to an automatically built scatter plot, such as locations where extreme values are reached.

The above-described procedure of data discretization can also be used to enable the application of data mining methods to grid data. In Andrienko et al. (2001c), we describe how exploratory analysis of spatial data may benefit from a combination of cartographic visualization with methods of data mining. We foresee that a similar synergistic effect of the two approaches can be achieved for grid-based data as well.

Spatio-temporal processes, i.e., phenomena changing in time, are often represented by temporal sequences of grids. In such cases, special methods for analysis of spatial and temporal data are needed. These methods should be able to support detection and evaluation of absolute and relative changes (see Andrienko et al., 2001a), detection of periodicity (Edsall et al., 2000) and temporal aggregation, and include convenient mechanisms for executing temporal queries and animating maps and graphics.

In this paper, we have introduced novel interactive visualization techniques for the exploration of grid data. We have demonstrated that computational and visual techniques are complementary, and their combined use in data analysis may produce a synergistic effect. For this purpose, we have described two example scenarios of data analysis in different domains. The analyses were performed with the use of CommonGIS, a unique system that provides both powerful geocomputational tools and interactive visualization-related facilities that can be efficiently used in combination.

Acknowledgements

References

Andrienko, G., Andrienko, N., 1999. Interactive maps for visual data exploration. International Journal Geographical Information Science 13 (4), 355–374.

Andrienko, G., Andrienko, N., 2001. Exploring spatial data with dominant attribute map and parallel coordinates. Computers, Environment and Urban Systems 25 (1), 5–15.

Andrienko, N., Andrienko, G., Gatalsky, P., 2001a. Exploring changes in census time series with interactive dynamic maps and graphics. Computational Statistics 16 (3), 417–433.

Andrienko, G., Andrienko, N., Savinov, A., 2001b. Choropleth maps—classification revisited. Proc. ICA 2001, Beijing, China. Chinese Society of Geodesy, Beijing, China, vol. 2, pp. 1109–1119.

Andrienko, N., Andrienko, G., Savinov, A., Voss, H., Wettschereck, D., 2001c. Exploratory analysis of spatial data using interactive maps and data mining. Cartography and Geographic Information Science 28 (3), 151–165.

Brewer, C.A., 1994. Color use guidelines for mapping and visualization. In: MacEachren, A.M., Taylor, D.R.F. (Eds.), Visualization in Modern Cartography. Elsevier, Tarrytown, NY, pp. 123–147.

Coburn, A., 1995. Disaster prevention and mitigation in metropolitan areas: reducing urban vulnerability in Turkey. In: Parker, R., Kreimer, A., Munasinghe, M. (Eds.), Informal Settlements, Environmental Degradation, and Disaster Vulnerability. The Turkey Case Study. World Bank and International Decade for Natural Disaster Reduction, Washington, DC, pp. 66–94.

Edsall, R.M., Harrower, M., Mennis, J., 2000. Visualizing properties of temporal and spatial periodicity in geographic data. Computers and Geosciences 26 (1), 109–118.

Giardini, D., Grünthal, G., Shedlock, K.M., Zhang, P., 1999. The GSHAP global seismic hazard map. Annali di Geofisica 42 (6), 1225–1230.

Gitis, V., 1995. GIS technology for the design of computer-based models in seismic hazard assessment. In: Carrara, A., Guzzetti, F. (Eds.), Geographical Information Systems in Assessing Natural Hazards. Kluwer Academic Publishing, Dordrecht, Netherlands, pp. 219–233.

Gitis, V.G., Osher, B.V., Pirogov, S.A., Ponomarev, A.V., Sobolev, G.A., Jurkov, E.F., 1994. A system for analysis of geological catastrophe precursors. Journal of Earthquake Prediction Research 3 (4), 540–555.

Gitis, V., Jurkov, E., Osher, B., Pirogov, S., Vainchtok, A., 1997. Information technology for forecasting geological processes and phenomena. Journal of Artificial Intelligence in Engineering 11 (1), 41–48.

Gitis, V., Osher, B., Dovgyallo, A., Gergely, T., 1999. GeoNet: an information technology for space-time WWW on-line intelligent geodata analysis. In: Peckham, R.J. (Ed.), Proc. 4th EC-GIS Workshop. JRC, Ispra, pp. 124–135.

Trifunac, M.D., Brady, A.G., 1975. On the correlation of seismic intensity scales with the peaks of recorded strong ground motion. Bulletin of the Seismological Society of America 65 (1), 139–162.