Dynamic aggregation on grids for interactive analysis of multidimensional spatial information

Gennady Andrienko, Natalia Andrienko, and Ivan Denisovich

Fraunhofer AiS - Autonomous Intelligent Systems Institute
Schloss Birlinghoven, Sankt-Augustin, D-53754 Germany
http://www.ais.fraunhofer.de/and/
e-mail: gennady.andrienko@ais. fraunhofer.de
tel +49-2241-142486, fax +49-2241-142072

SUMMARY

In this paper we propose a dynamic aggregation procedure that bridges continuous and object-oriented representation models of cartographic information. We discuss problems of visual analysis of data from multiple grids. As a solution, we propose a procedure of dynamic aggregation of grids into a set of regularly-shaped polygons. The polygons are characterized by attributes that are derived from values in the cells of the original grids. For such polygons, numerous visualization and analysis methods can be applied. A possibility to change dynamically the resolution of the aggregation supports sensitivity analysis.

KEYWORDS: *exploratory data analysis, interactive visualization, dynamic aggregation*

Grid-based spatial 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. On the other hand, grid-based data are often used for the representation of derived information about spatial and spatio-temporal properties of discrete geographical objects. For example, density of roads or communication lines, parameters of seismic flow, pollution and pesticide accumulation data, and statistical parameters of accidents. There is even a tendency towards using grids for the representation of population-related data since this helps to cope with the problem of changing administrative boundaries.

In our recent paper [1] we introduced methods and tools for interactive analysis of grid data. However, the described methods are suitable mostly for analysis of individual grids. Although rendering of one of the grids in a semi-transparent mode makes it possible to represent two grids on the same map, the resulting image may be difficult to perceive and not sufficiently supportive for comparing the distributions of the phenomena and exploring the relationships between them. Besides, this approach is not extendable to three or more grids.

Many graphical data representation methods supporting the comparison of value distributions and analysis of relationships between variables have been developed in the area of statistics and, in particular, its offspring known as exploratory data analysis (EDA). Examples of such graphs are histogram, scatterplot, scatterplot matrix, box plot, parallel coordinate plot, etc. In EDA, these displays are implemented as highly interactive software tools [2,3]. It would be useful to apply the power of these tools for the exploration of multiple grids. However, direct application of these representation methods to values in grid nodes is typically impossible for the following reasons:

- the grids may differ in extent and resolution, so that there is no direct correspondence between the nodes of different grids;

- the grids are usually very detailed, and the representation of the values from all nodes would result in tremendous cluttering and overlapping of the elements on the graphs. Besides, this puts serious limits to display interactivity.

For the same reasons, one cannot also apply traditional cartographic representation methods suitable for two or more attributes to grid data.

To cope with these problems, we have implemented the technique of dynamic aggregation of grids in our system CommonGIS. The user can define a division of the territory into rectangular areas (cells), and, in response, the system summarizes data from multiple grids into a single table. The rows of the table correspond to the rectangles produced and the columns contain the values of the attributes for these rectangles, where each attribute is derived from one of the grids. The attribute values are defined as the averages from all values of the grid nodes that fit into the corresponding rectangles. If no grid node fits in some rectangle, the value is interpolated between the nearest nodes of the grid. It is important that the user may interactively change the desired resolution of the grid-to-table transformation. In the result, the table is dynamically modified: the number of rows increases or decreases, depending on the current resolution, and the attribute values are re-computed from the original grids.

The so transformed grid data can now be explored using any analytical tools suitable for spatially referenced attribute data, in particular, numerous tools available in CommonGIS: interactive diagram maps [4], multidimensional statistical graphics displays [5], tools for multiple-criteria decision making [6], and spatial data mining [7]. Especially interesting is the analysis of temporal sequences of grids using change maps [8], animated and dynamic maps, and time graphs [9].

All displays in CommonGIS are sensitive to changes in the tables providing data for the visualization. Therefore, they are automatically updated when the user changes the resolution of the grid-to-table transformation. Hence, the user can check whether the patterns and relationships observed on the displays depend on the resolution of the grid-derived table.

Let us demonstrate by examples how data from multiple grids can be analyzed using the grid-to-table transformation technique.

The map in Figure 1 represents the distribution of broadleaved and coniferous forests over Europe. The data come originally from two grids, each grid containing data for one forest type. The resolution of the original grids is 1607x1020. Using the grid-to-table transformation technique, the data have been summarized over 1700 (50x34) rectangular cells and put in a table as two attributes. Then, the values of these attributes for the cells have been represented on the map using bar charts. The map clearly shows that the north of Europe has higher total amounts of forests with great prevalence of coniferous forests. In Central and Eastern Europe broadleaved forests prevail although the amounts are much lower than on the north.

For a more detailed investigation of the data, let us increase the resolution of the grid-to-table transformation. The map in Figure 2 corresponds to the resolution 100x68. It allows us, for example, to consider the variation of the amounts and proportions of the two types of forests over the territory of Central and Eastern Europe. We see spatial clusters of prevailing coniferous forests in Alps, on the north around the border between Germany and Poland, and on the east. Clusters with substantial prevalence of broadleaved forests can be observed in southern France, in Italy, on the northwest of Balkan Peninsula and in the center of the Eastern Europe. Beyond these clusters, the amounts of forest of both types are rather low.

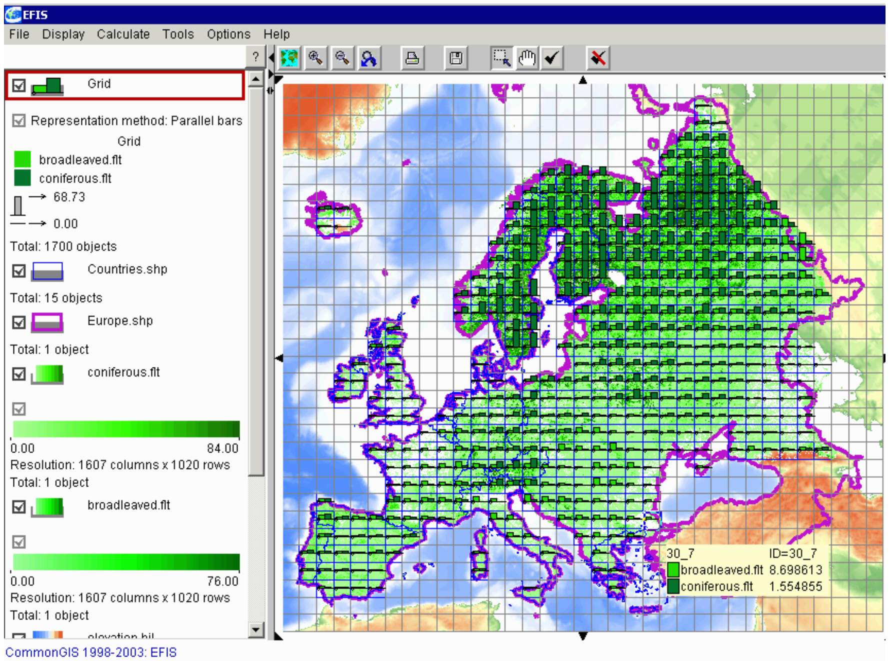

Figure 1. Data derived from two grids are represented by bar charts.

Figure 3 demonstrates that attribute data derived from grids using the grid-to-table transformation technique can be analyzed using various computational tools applicable to tables, for example, techniques of data mining such as cluster analysis.

We have used the grid-to-table transformation technique for combining data from 6 grids: amounts of forests of different types, non-forest land, and elevation. Then we have applied one of the cluster analysis methods available in the public domain data mining system Weka [10] to the resulting table with 6 attributes. The map in Figure 3 shows how the clustering algorithm has divided the territory of Europe into 3 clusters according to these attributes. It can be seen that, although the clustering algorithm did not use any spatial information, it roughly divided the territory of Europe into northern, middle, and southern parts, except some areas in the center that are closer by their characteristics to the northern territories.

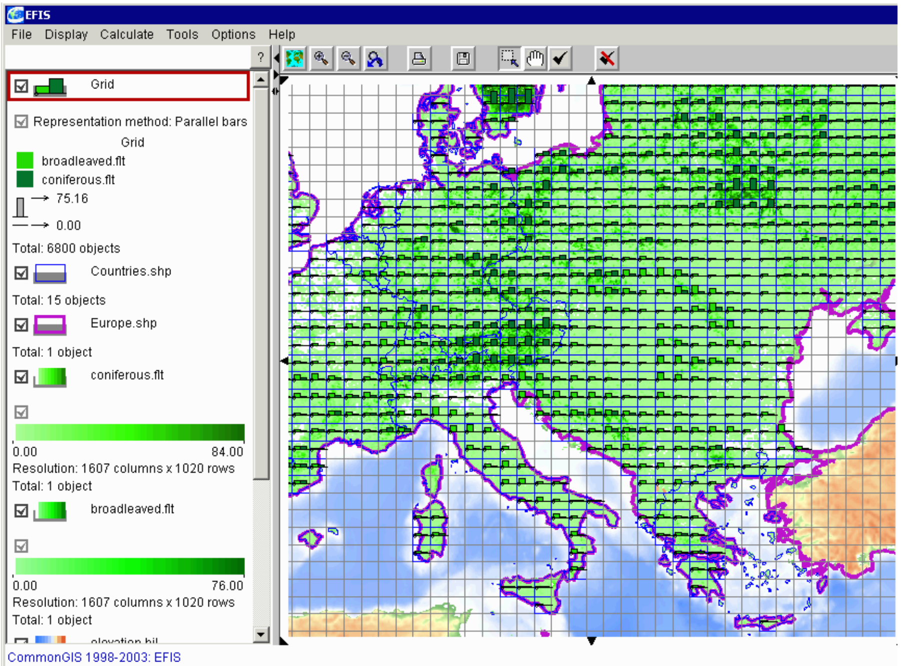

Figure 2. Grid-to-table transformation with higher resolution has been applied to the same data as in Figure 1.

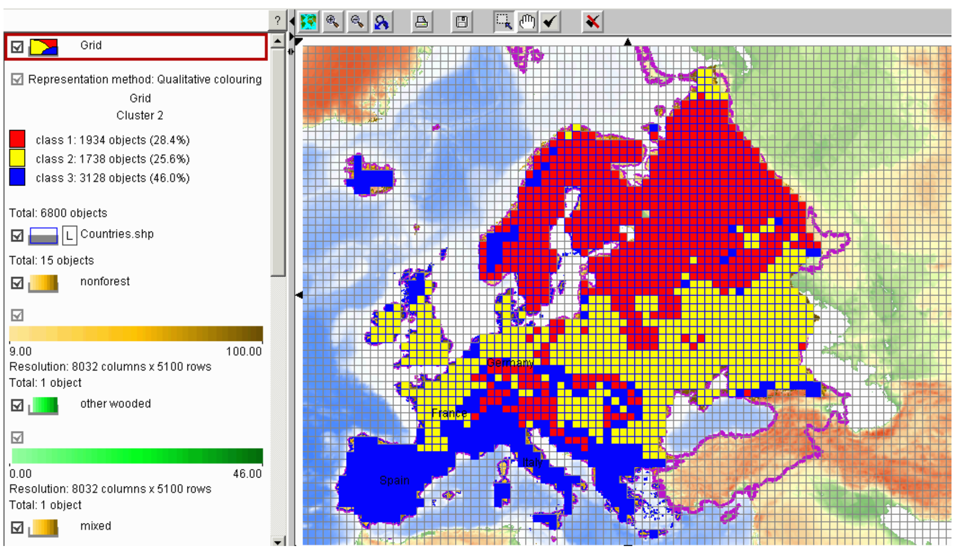

Figure 3. The map represents a result of cluster analysis of grid-derived data.

As can be seen from Figure 4, non-cartographical visualization techniques can be also useful for the analysis of grid-derived data. The histograms represent the statistical distributions of amounts of different forest types: broadleaved (upper left), coniferous (upper middle), mixed (upper right), as well as other wooded land (lower left) and non-forest land (lower right). The bars of the histograms painted using the colors of the clusters represented in Figure 3 so that the sizes of the bar segments are proportional to the numbers of objects from the respective clusters fitting in the corresponding value interval. In the result, the upper central and right histograms show us that the red cluster mostly consists of cells with medium and high amounts of coniferous and mixed forests while the cells in the yellow cluster have low amounts of these types of forests. From the lower right histogram, one can see that the yellow cluster corresponds to high amounts of non-forest land while the red cluster consists of cells with low amounts of non-forest land and, hence, with high amounts of forests. The blue cluster has similar amounts of coniferous and mixed forests as the yellow cluster and occupies an intermediate position between the red and the blue clusters with respect to the amount of non-forest lands.

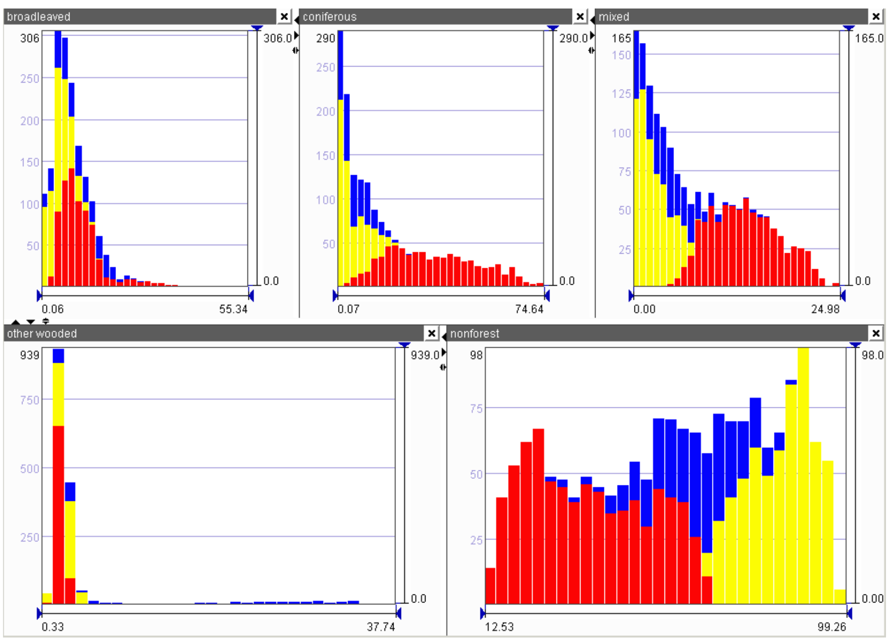

Figure 4. Grid-derived data are represented on linked interactive histograms. The histograms are painted according to the clusters shown in Figure 3.

These few examples indicate the great variety of opportunities for data analysis enabled by the grid-to-table transformation technique. There are also opportunities for further advancement of the technique itself. Thus, in constructing a table from grids, it may be useful to compute not only the average grid values in the cells but other statistics as well: minimum, maximum, variance, median or other percentiles, etc. It is also possible to do the transformation at a variable resolution: apply coarse aggregation to an entire territory but divide more important or more interesting parts of it into finer cells for seeing more details. These advancements are in the plans for the future development of CommonGIS.

We are also planning to use the grid-to-table transformation for enhancing the possibilities for analysis of multiple large grids in the Internet-based variant of the system. The idea is that the transformation is done on the server side, and the client receives the result, i.e. a table with aggregated data from several grids. The amount of data to be transferred over the Internet is thereby substantially reduced. The aggregated data provide a good first overview of the spatial variation of characteristics. After getting the first impression, the user has a possibility to request more detailed data (i.e. with lower degree of aggregation or even without any aggregation) for a part of the territory or for the entire territory.

The system CommonGIS for interactive visual analysis and decision making is freely available for research and educational purposes at the URL www.commongis.com.

The presented work is partly supported by the European Commission in project GIMMI (IST-2001-34245; http://services.txt.it/gimmi/).

BIBLIOGRAPHY

1. Andrienko, G., Andrienko, N., and Gitis, V. Interactive Maps for Visual Exploration of Grid and Vector Geodata. *ISPRS Journal of Photogrammetry and Remote Sensing.* 2003, v.57 (5-6), pp. 380-389

2. W.S. Cleveland, *Visualizing Data*, Hobart Press, Summit, New Jersey, 1983

3. A. Buja, J.A. McDonald, J. Michalak, W. Stuetzle, Interactive data visualization using focusing and linking. In *Proceedings IEEE Visualization'91*, Washington: IEEE Computer Society Press, 1991, 156-163

4. Andrienko, G. and Andrienko, N. Interactive Maps for Visual Data Exploration. *International Journal Geographical Information Science.* 1999, v.13 (4), pp.355-374

5. Andrienko, G. and Andrienko, N. Exploring Spatial Data with Dominant Attribute Map and Parallel Coordinates. *Computers, Environment and Urban Systems*. 2001, v.25 (1), pp.5-15

6. Jankowski, P., Andrienko, N., and Andrienko, G. Map-Centered Exploratory Approach to Multiple Criteria Spatial Decision Making. *International Journal Geographical Information Science.* 2001, v.15 (2), pp.101-127

7. Andrienko, N., Andrienko, G., Savinov, A., Voss, H., and Wettschereck, D. Exploratory Analysis of Spatial Data Using Interactive Maps and Data Mining. *Cartography and Geographic Information Science.* 2001, v.28 (3), pp. 151-165

8. Andrienko, N., Andrienko, G., and Gatalsky, P. Exploring Changes in Census Time Series with Interactive Dynamic Maps and Graphics. *Computational Statistics.* 2001, v.16 (3), pp.417-433

9. Andrienko, N., Andrienko, G., and Gatalsky, P. Tools for Visual Comparison of Spatial Development Scenarios. *Proceeding of Information Visualisation'2003* (London UK).

10. I.H.Witten, E.Frank, *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*, Morgan Kaufmann, 1999