Geo-visualisation support for environmental modelling and decision making

Gennady Andrienko and Natalia Andrienko

Fraunhofer Institute AIS, Schloss Birlinghoven, 53754 Sankt Augustin, Germany
{Gennady,Natalia}.andrienko@ais.fraunhofer.de
http://www.ais.fraunhofer.de/and

Outputs of environmental simulation models usually have a form of multidimensional tables that contain values of several attributes referring to different moments of time, different locations in space, and different scenarios of possible management. We consider analytical questions arising in the comparison of scenarios and propose interactive methods for visual analysis, which incorporate data transformation and data aggregation. This approach allows us to manage various aspects of the complexity of simulation results: high dimensionality, abrupt temporal changes, and great variability of values.

1. Introduction.

Simulation modelling is widely used nowadays for gaining an understanding of environmental processes and for managerial decision making. An outcome of a simulation model often consists of values of multiple attributes referring to different moments of time and different locations in space. Typically, an analyst or decision maker needs to compare results of several simulations corresponding to different scenarios of possible development or management.

In our recent paper (Andrienko et al, 2003a) we considered the types of analytical tasks arising in such a comparison. Thus, we distinguished the analysis of characteristics at selected time moments in selected places, the investigation of spatial distributions of attribute values at selected time moments, the study of the dynamics of characteristics at selected locations over time periods, and the exploration of the overall spatio-temporal patterns.

A good data visualisation tool is expected to support these analytical tasks on different levels of consideration: the whole data set, dynamically defined subsets of interest, and individual instances with particular characteristics. The tool must be able to provide a rapid overview of the data, zooming on demand to dynamically defined subsets of interest, various transformations of the data and their graphical representations, search for objects with specific characteristics, etc. In response to these requirements, we have suggested a group of complementary information visualisation techniques that collectively provide a suitable coverage of the space of possible analytical tasks. The group includes interactive dynamic maps, time graphs, temporal aggregation, and dynamically linked statistical graphics techniques.

In the current paper, we generalize our experiences gained from applying interactive geo-visualisation techniques to the analysis and comparison of simulation results in two domains, pesticide accumulation and forest development. The results are multidimensional datasets with attribute values referring, besides space and time, to one of several scenarios and to certain problem-specific dimensions such as, for example, tree species and age groups in forest simulations. Such data are quite difficult to analyze, first of all due to large volumes and high dimensionality, which stem from the interplay of the following aspects:

- the number of simulation scenarios considered simultaneously;
- the number of locations or spatial compartments for which the simulation has been performed;
- the number of time moments the simulation results refer to (the length of time series);
- the number of characteristic attributes involved in the evaluation and comparison of the simulated processes.

Hence, tool scalability, i.e. the capability to cope with large multidimensional datasets, becomes a critical issue. There are also other difficulties caused by specific properties of the data. Thus, they often contain abrupt temporal changes (for example, in the result of forest cutting or other management activities); data values may vary within wide intervals while the variation is often extremely skewed.

As a response to these challenges, we suggest several new methods for visual data analysis, which are based on data transformation and data aggregation.

2. Interactive data transformation and aggregation.

We shall start with an example analysis of pesticide accumulation in 5 agricultural management scenarios with different active ingredients and crops.

On Figure 1 five time graphs show the results of simulating the dynamics of pollutant concentration in soil resulting from different agricultural management strategies. The simulations have been performed for selected crops and active ingredients of pesticides, 5 pairs in total, for about 1000 territory compartments in Lombardy, Italy (the compartments, called "soil polygons" by domain analysts, are defined so that soil properties are homogeneous within each of them). One can notice very high variability of outputs and significant overplotting of lines on the display due to a skewness of value distribution: the bulk of values is close to zero while some values are extremely high. These distribution peculiarities suggest that a logarithmic data transformation could be helpful. This is consistent with the general practice applied by soil contamination analysts: they usually consider several value intervals with breaks at 0.001, 0.01, 0.1, 1, 10, etc.

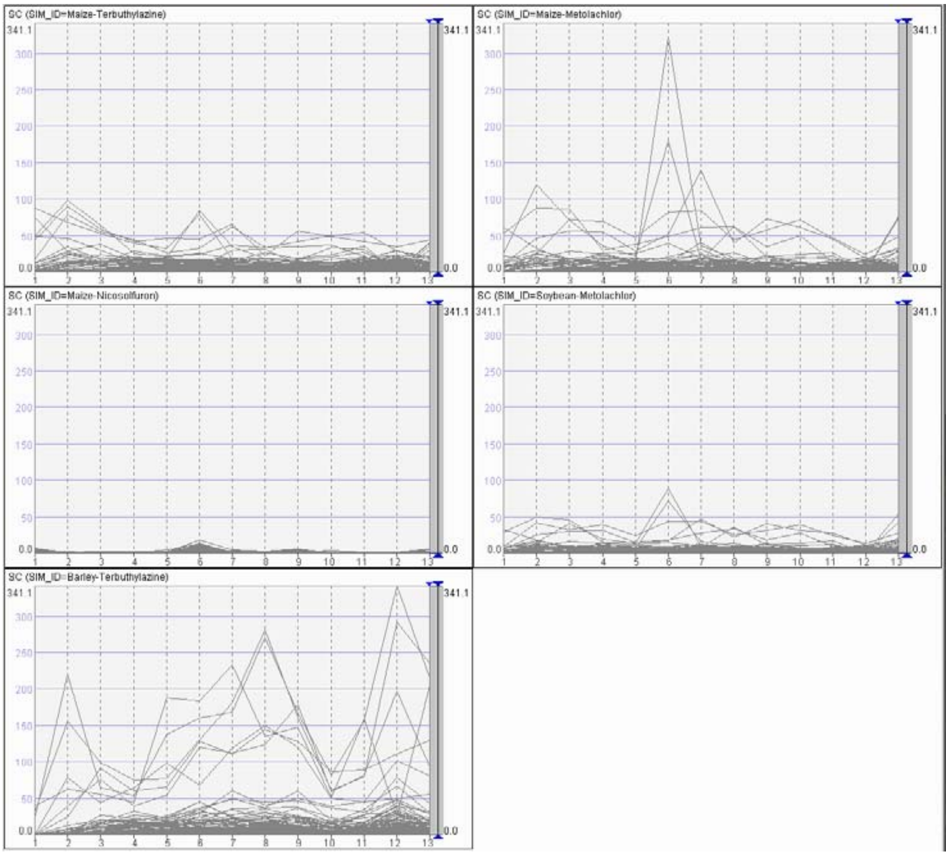

Figure 1. Five time graphs show the results of simulating the dynamics of pollutant concentration in soil resulting from different agricultural management strategies.

On Figure 2 the pollutant concentration values have been transformed into decimal logarithms (to avoid missing values in the transformation results, all original 0.0 values have been replaced by 0.00001). The graphical presentation of the results of the logarithmic transformation has exposed important features of the five scenarios that were completely hidden in the original visualisation due to the outliers. Additionally, the running median line is drawn on each time graph.

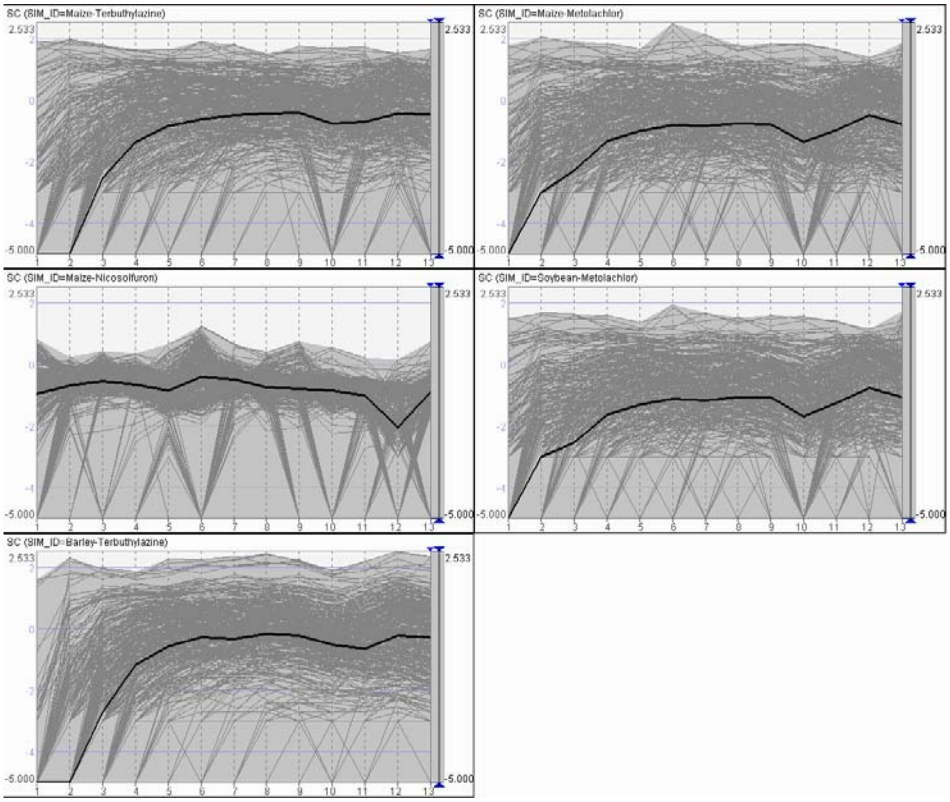

Figure 2. The pollutant concentration values have been transformed into decimal logarithms.

For studying general properties of the temporal dynamics of soil contamination on the level of the whole dataset, it is suitable to compute positional statistical measures (quantiles) for each time moment. Examples of positional measures are median, quartiles, deciles (which are presented in this example), percentiles, etc. After connecting corresponding positions for consecutive time moments, we are getting an aggregated representation of the pesticide accumulation (Figure 3).

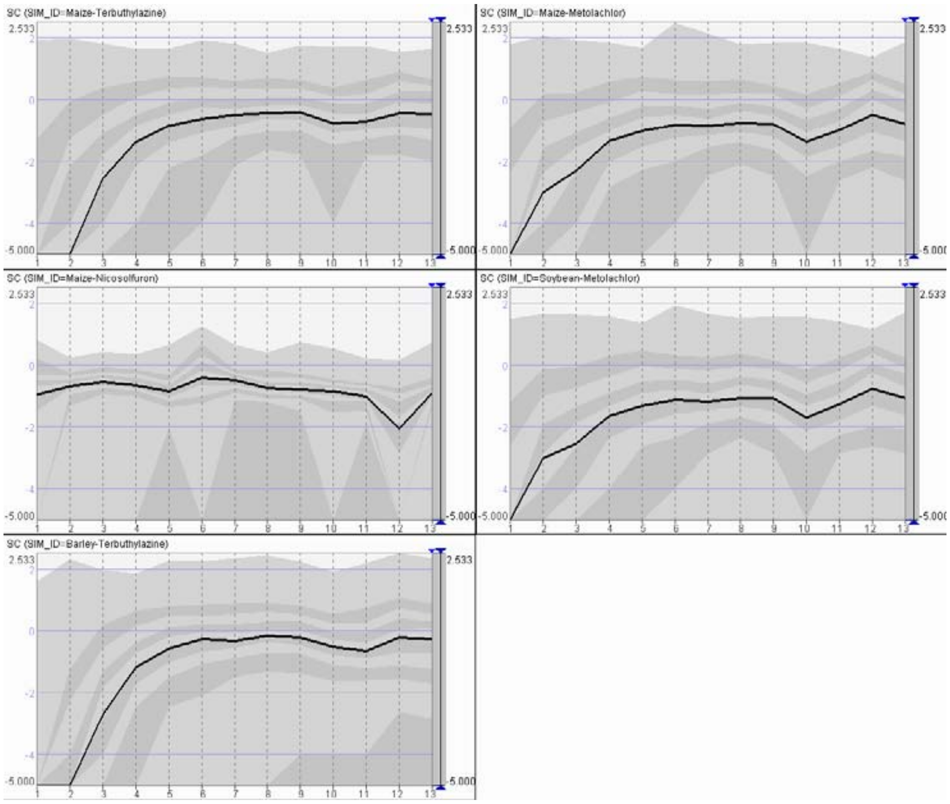

Figure 3. An aggregated representation of pesticide accumulation in the five scenarios.

Another aggregated representation is based on dividing the attribute value range into subintervals (Figure 4). In this example, the division is applied to the logarithmic transformation results, and the break values are -3 (corresponds to the original value 0.001) -2 (0.01), -1 (0.1), -0.1 and 0.1 (correspond to the original interval around 1.0), and so on. Then, the tool has counted for each year how many soil polygons have their transformed values fitting in each of the so defined intervals. The stacked bars below each graph show the resulting counts. Green colors correspond to low values, yellow – to medium values, and red – to high and extremely high concentrations. It is important that the break values have been defined interactively, using a device that provides an immediate feedback during the modification of values.

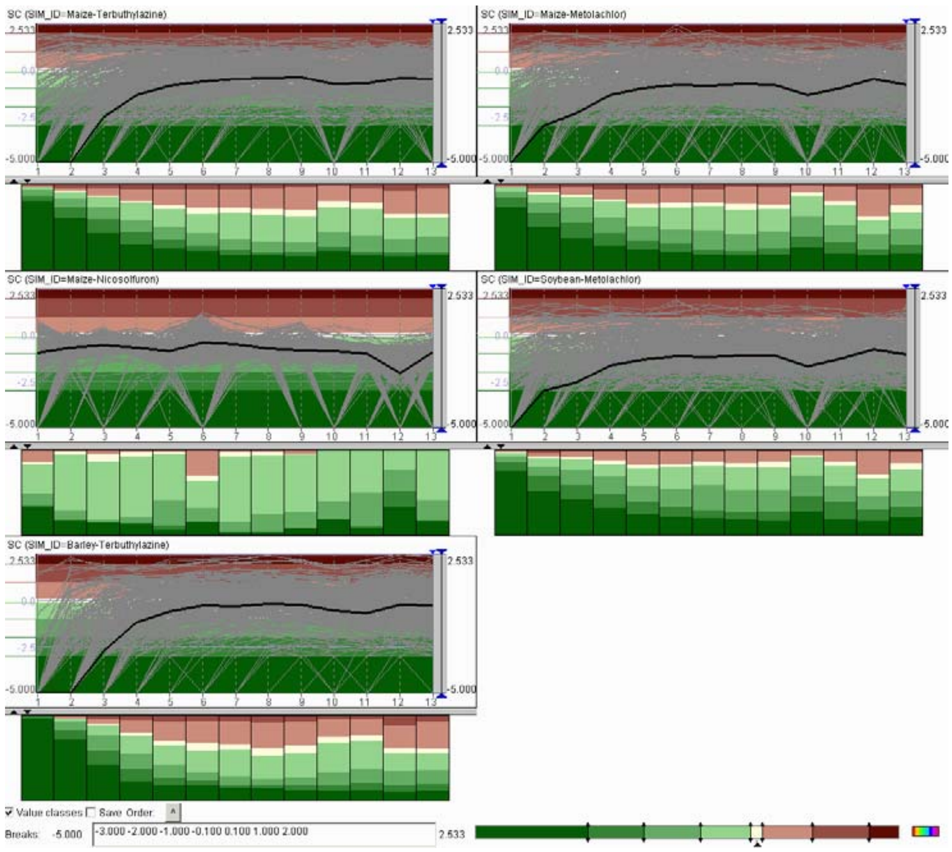

Figure 4. Aggregated representation based on dividing the attribute value range into subintervals.

After defining meaningful break values, it may be useful to remove the original time graphs and focus the attention on the aggregated representations (Figure 5). This visualisation provides a statistical summary of the dynamics of pesticide concentrations in the five scenarios. It effectively supports scenario comparison, which is important for informed decision making. For example, one can notice that one of the scenarios (the second from top display on the left) is characterized by medium concentrations during the whole simulated period except the year 6, in which the number of districts with high concentrations is rather large. Other simulations have quite a different behavior.

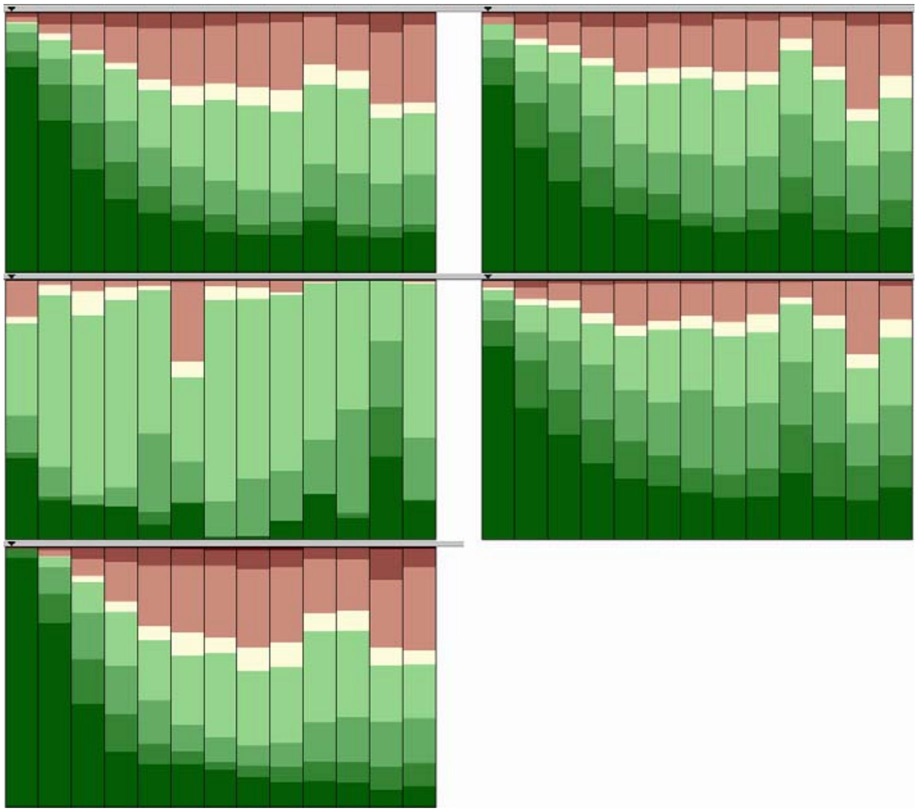

Figure 5. A statistical summary of the dynamics of pesticide concentrations in the five scenarios.

Figure 6 shows how an aggregated representation of forest structure could be used for studying the dynamics of forest biodiversity. In the course of simulating a natural development of a forest during 200 years without any management activities, the amount of forest (basal area, m^2/Ha) has been computed for 6 species and 25 age groups in each forest compartment and each year. For the purpose of analyzing the dynamics of the forest biodiversity, a so-called dominant attribute classification has been used. The method shows, for each year, what species and age group occupied the largest area in each compartment. Compartments with low total amount of forest (i.e. below a certain value called presence threshold) are classified as empty. Additionally, a dominance threshold has been introduced for separating mixed forest compartments from those more homogeneously covered. The values of the two thresholds have been defined interactively, after making several trials and inspecting the results.

The three maps demonstrate the spatial distribution of dominant species for 3 time moments: the initial state, the state after 100 years, and the final state. Color hues indicate the dominant species and the degrees of darkness show the dominant age groups, with lighter shades corresponding to younger trees. The black color is used for the compartments classified as empty and white represents mixed forests. In addition to the maps, the diagrams on the left provide aggregated information about dominant species in all forest compartments at a given time moment. Each diagram contains colored cells grouped into 6 rows (one per species) and

25 columns (one per age group). Colors are the same as on the maps. Within a cell, a white horizontal line is drawn if there is at least one compartment where the corresponding species and age group combination is present, with regard to the presence threshold. A white rectangle is drawn if there is at least one forest compartment where the corresponding species and age group combination prevails. Hence, the diagrams simultaneously serve as a color legend for the maps and provide aggregated information about the presence of species and age groups in the whole forest.

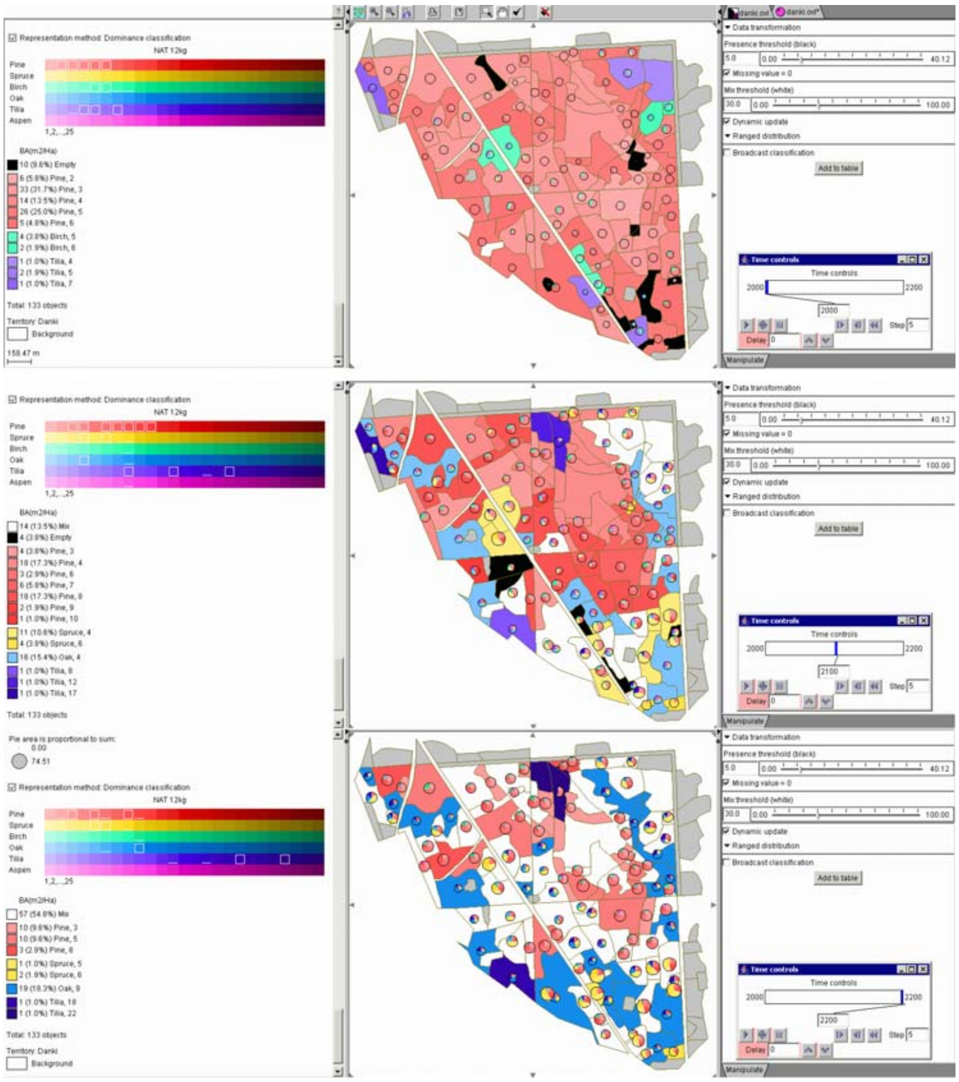

Figure 6. An aggregated representation of forest structure.

3. Discussion and conclusions.

A common property of the proposed methods is the utilization of non-trivial aggregations for reducing the dimensionality and complexity of data and providing a synoptic representation of a scenario as a whole. There is a certain similarity to OLAP tools (Thomsen 2002), which also apply data aggregation. However, the aggregation methods in OLAP are limited, for the efficiency reasons, to simple counts, sums, minimal and maximal values, and averages. This may be sufficient for transaction analysis and reporting in business applications if attribute values are distributed normally, with relatively low variance, and without outliers (see Wilkinson (1999) for the discussion on OLAP applicability). If these assumptions about data properties are not fulfilled, the aggregation results may be useless or even misleading.

Therefore, the analysis of complex spatio-temporal processes requires other approaches, such as using positional statistical measures (i.e. various quantiles), value transformation, or/and separate consideration of different value subintervals. Extremely important is the analysis of the aggregation sensitivity by means of interactive and dynamic displays. A challenging task is development of database models, architectures, and software that could efficiently support such operations with very large data sets. At the same time, there is a pressing need in scalable and powerful methods for interactive visual analysis of complex multidimensional data sets.

Another challenging task is helping users to orient in the large variety of existing data analysis methods and to utilize these methods effectively. As a first step in this direction, we developed an operational classification of spatio-temporal analytical questions and matched the existing methods to the tasks (Andrienko et al, 2003b). We are going to extend our framework by involving the scalability issues. On this basis, we are planning to build a knowledge-based tool for guiding users in choosing appropriate data analysis methods, applying these methods to user's data, an interpreting their results. There is also a need in developing a theoretical and methodological framework that links data analysis to decision making.

References

N. Andrienko, G. Andrienko, and P. Gatalsky, "Tools for Visual Comparison of Spatial Development Scenarios", In Banissi, E. et al (Eds.) *IV 2003. Seventh International Conference on Information Visualization*, Proceedings, 16-18 July, 2003, London, UK. IEEE Computer Society, Los Alamitos, California, 2003, pp. 237-244

N.Andrienko, G.Andrienko, and P.Gatalsky, "Exploratory Spatio-Temporal Visualization: an Analytical Review", *Journal of Visual Languages and Computing*, 2003, v.14 (6), pp. 503-541

E.Thomsen, "*OLAP Solutions. Building Multidimensional Information Systems*", NY: Wiley, 2002

L.Wilkinson, "*The Grammar of Graphics*", Springer-Verlag, New York, 1999

Acknowledgements

The paper is based on the results achieved in projects GIMMI (IST-2001-34245, Geographical Information and Mathematic Model Interoperability, 2002-2004, coordinator – TXT e-solutions, Italy), NEFIS (QLK5-CT-2002-30638, Network for a European Forest Information Service, 2003-2005, coordinator – EFI, Finland), and SILVICS (INTAS, Silvicultural Systems for Sustainable Forest Resources Management, 2002-2005, coordinator – Wageningen University, the Netherlands). We are going to extend this work within the integrated project OASIS (IST-2003-004677, Open Advanced System for Improved crisis management, 2004-2008, coordinator – EADS France), in which we are coordinating a subproject focused on intelligent decision support. We acknowledge the funding of the European Commission and friendly cooperation of our partners and colleagues.