Visual Mining of Spatial Time Series Data

Gennady Andrienko, Natalia Andrienko, and Peter Gatalsky

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

Abstract. CommonGIS is a system comprising a number of tools for visual data analysis. In this paper we demonstrate our recent developments for analysis of spatial time series data.

1 Introduction

Analysis of time series data is one of the most important topics in data mining research. A number of computational methods have been recently developed [1]. Researchers in statistical graphics have also paid a significant attention to this problem, proposing interactive visualisation techniques for time-series [2,3]. These papers introduced basic graphical and interaction primitives for enhancing analytical capabilities of time series plots, or time graphs:

1. Interactive access to values via graphics by pointing on a segment of a line.
2. Tools for overlaying lines for their comparison, with a possibility to distort the lines for better fit.
3. The possibility to select lines with particular characteristics, such as specific values at a given time moment or interval, specific profiles, etc.
4. Dynamic linking between a plot and other displays (scatter-plots, histograms, maps etc.) via highlighting and selection.

However, most of these tools are applicable to analysis of a small number of lines but are not suitable for studying large collections of time series data. Usually such tools are implemented as stand-alone prototypes with a limited number of available complementary displays.

We designed and developed novel methods of time series analysis in our system CommonGIS [4,5], which is unique among both commercial and research software systems as a composition of well-integrated tools that can complement and enhance each other thus allowing sophisticated analyses. The system includes various methods for cartographic visualisation, non-spatial graphs, tools for querying, search, and classification, and computation-enhanced visual techniques. A common feature of all the tools is their high user interactivity, which is essential for exploratory data analysis.

The main features of CommonGIS related to analysis of spatial time series are:

1. A variety of interactive mapping techniques combined with statistical graphics displays and computations.

2. Animated maps and other time-aware map visualisation techniques [6,7].
3. Novel information visualisation tools (dynamic query, table lens, parallel coordinate plots etc.) dynamically linked to maps and graphics via highlighting, selection, and brushing.
4. Interface to data mining tools [8].

2   Functionality

Let us describe a short example scenario of data analysis that demonstrates some of the new methods. The time graph in Figure 1, upper left, displays results of simulating forest development in 2,600 forest compartments over 100 years with 5-years interval, specifically, the tree biomass in tons per hectare. To eliminate line overplotting, individual lines have been replaced by the median line and 10 percentile lines (10%, 20%, etc.), which form virtual flows, or envelopes (Figure 1, upper right).
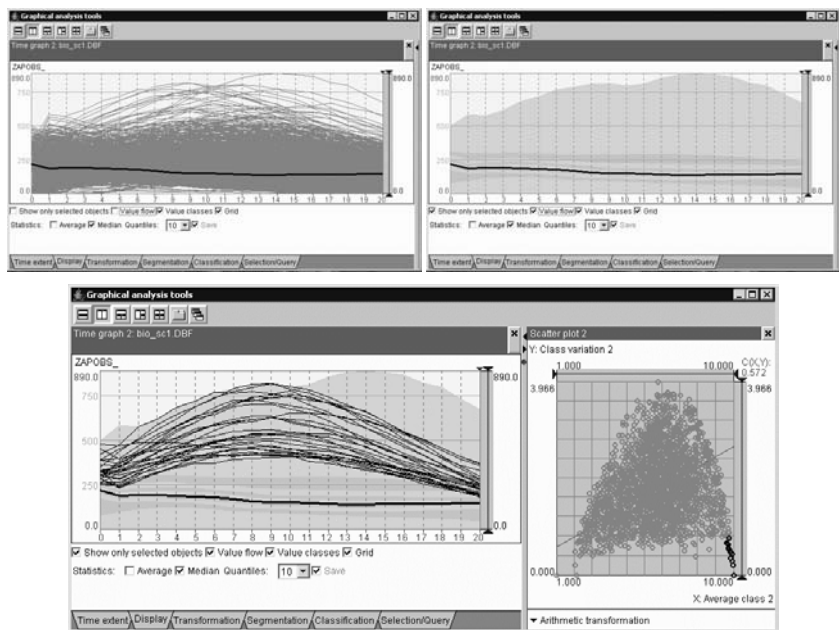

Fig. 1. Individual lines (left), median and 10 envelopes (right), and a scatter plot showing the average envelope and variance, that was used to select particular lines to be shown on the time graph (bottom).

Then, for each individual line, the system has computed the average envelope and the variance. These values have been represented on a scatter plot (Figure 1, bottom right). Using the dynamic link between the scatterplot and the time graph, one may select and explore lines with particular characteristics. Thus, in Figure 1, the lines mostly belonging to the highest envelope have been selected through the scatterplot and, as a result, highlighted on the time graph.

Further analysis can be done using thematic maps. Thus, figure 2 shows a map of relative changes of the tree biomass over the simulated period. The darkest colour shows compartments with biomass increase by more than 10%, the lighter colours correspond to small changes (±10%), decrease (from 10% to 50% of losses) and significant drops (more than 50% of losses). The system allows the user to transform data (e.g. calculate changes, smooth time series, etc.) and to manipulate visualisation properties. Thus, class breaks may be selected interactively on the basis of such criteria as statistical properties of value distribution, the number of objects in classes, or spatial distributional patterns.
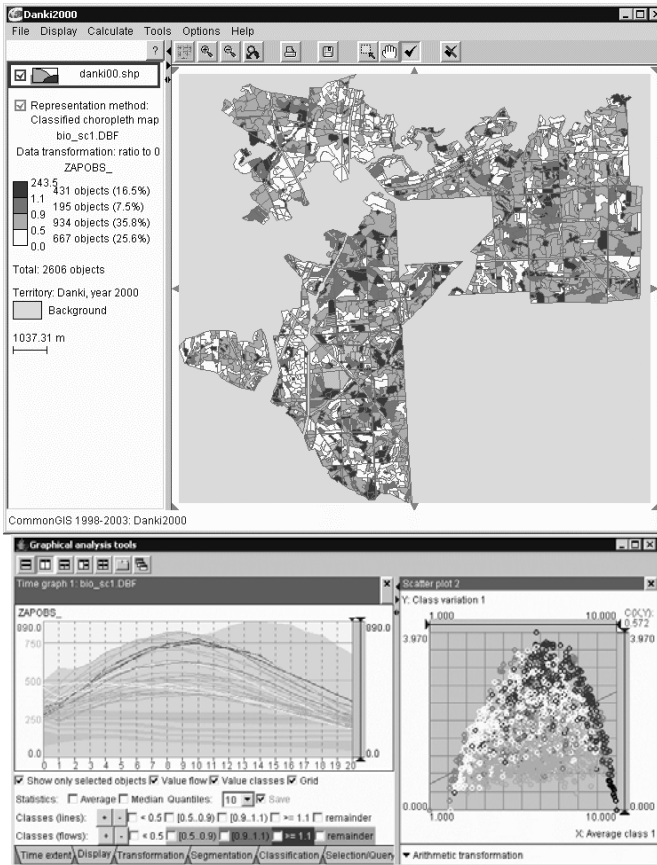

Fig. 2. The map divides forest compartments into classes according to the degree of biomass change. The class colours are projected onto the scatter plot and on the selected lines on the time graph.

3   Discussion and Future Work

We have proposed new interactive methods for analysis of spatial time series data. The methods are scalable with regard to the number of time-series considered simul-

taneously. However, we see the need in further research for supporting analysis of long time-series and for analysis of multi-attribute time series.

The methods described in this paper will be further developed in a recently approved integrated project OASIS (Open Advanced System for Improved crisiS management, 2004–2008, IST-2003-004677). The focus of further work is the adaptation of the time series analysis for needs and purposes of decision support (analysis of situation, formalization and evaluation of decision strategies, multi-criteria selection and decision sensitivity analysis).

Acknowledgements

References

1. E.Keogh and S.Kasetty, On the need for time series data mining benchmarks: a survey and empirical demonstration, Data Mining and Knowledge Discovery, 2003, 7 (4), 349-371.
2. Unwin, A. R., and Wills, G., Eyeballing Time Series. Proceedings of the 1988 ASA Statistical Computing Section, 1988, pp.263-268
3. H.Hochheiser and B.Shneiderman, Dynamic query tools for time series data sets: Timebox widgets for interactive exploration, Information Visualization, 2004, 3 (1), pp.1-18.
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, N., Andrienko, G., Informed Spatial Decisions through Coordinated Views, Information Visualization, 2003, v.2 (4), pp. 270-285
6. Natalia Andrienko, Gennady Andrienko, and Peter Gatalsky, Exploratory Spatio-Temporal Visualization: an Analytical Review, Journal of Visual Languages and Computing, 2003, v.14 (6), pp. 503-541
7. 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.
8. 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