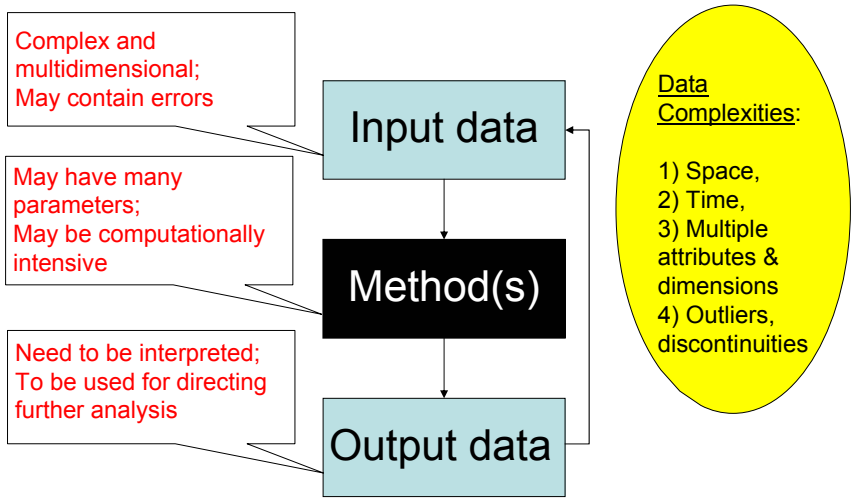
Visual Analytics and Data Mining in S-T-applications

Gennady Andrienko & Natalia Andrienko
Fraunhofer Institute AIS
Sankt Augustin
Germany
http://www.ais.fraunhofer.de/and

Fraunhofer Institut Autonome Intelligente Systeme

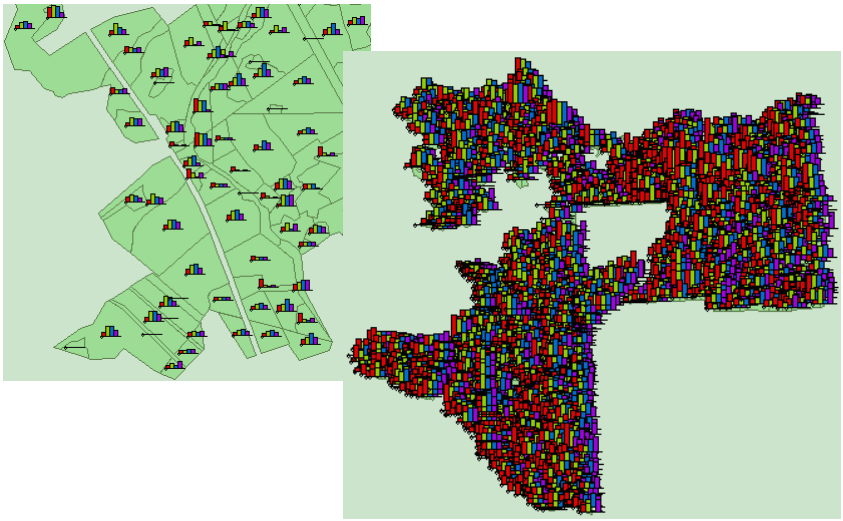
A View on S-T Data Mining



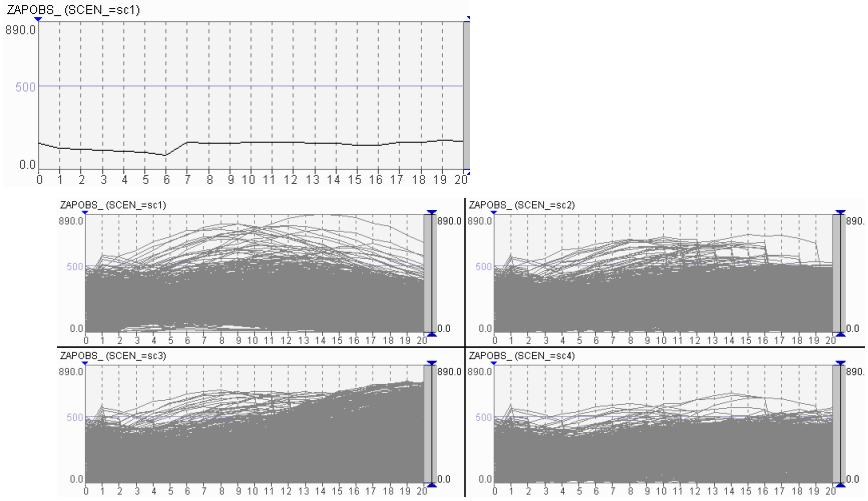
Complexities

- Number of attributes
- Length of time series
- Number of spatial objects
- High dimensionality
- Abrupt temporal changes
- Great variability of values

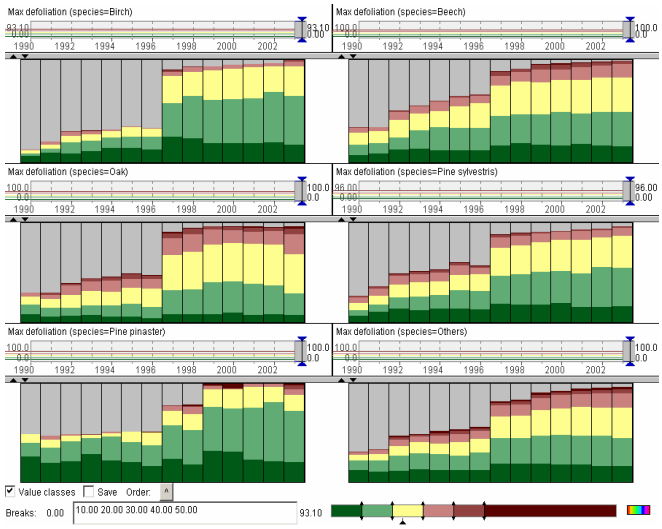
Complexities: example 1



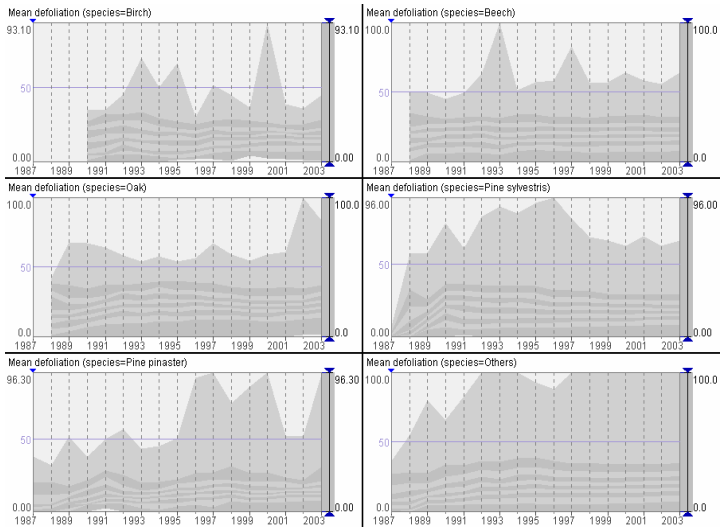
Complexities: example 2



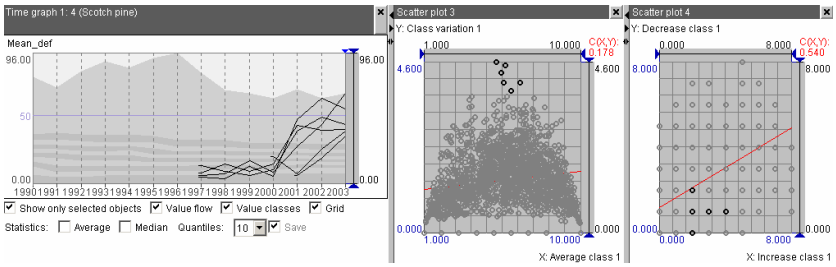
Aggregation method 1: by intervals



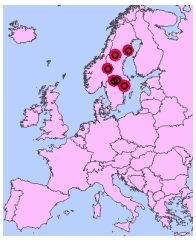
Aggregation method 2: by quantiles



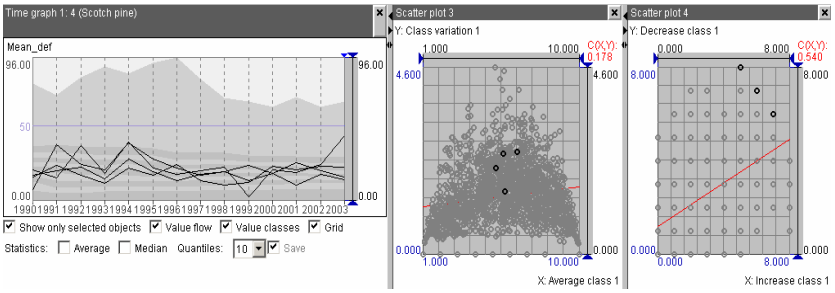
Attend to particulars: high variation



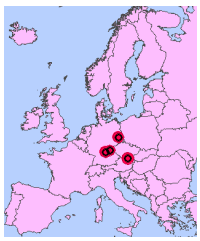
1. Aggregate time graph by quantiles
2. Save counts
3. Visualise e.g. on a scatter plot
4. Select items with high variation



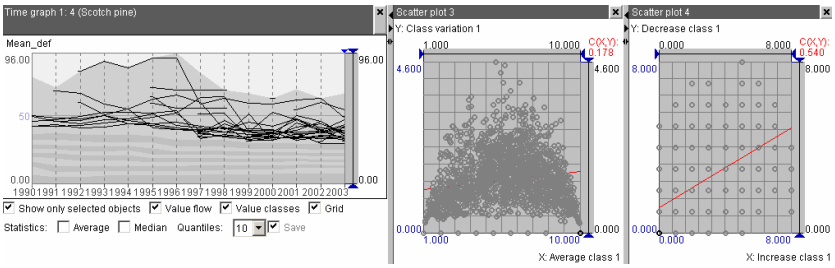
Attend to particulars: high fluctuation



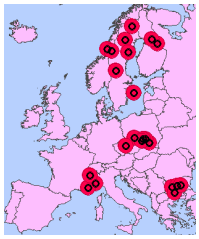
- Select items with maximal number of jumps between quantiles



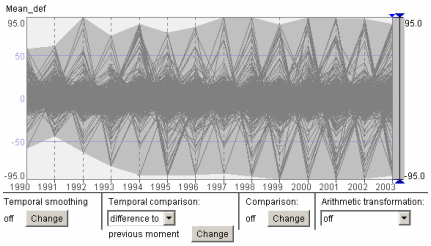
Attend to particulars: stable extremes

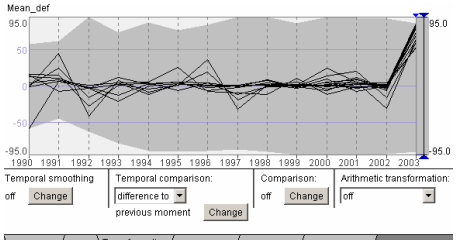


- Select items being always in the topmost 10%

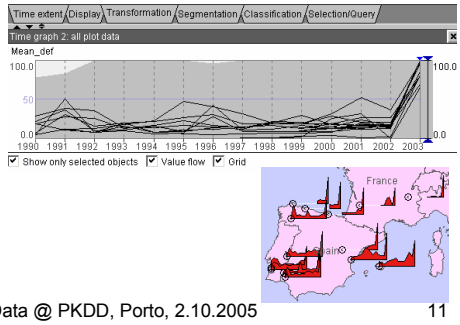


Attend to particulars: extreme changes

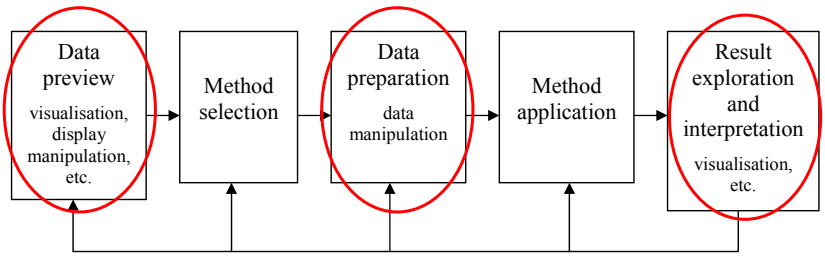




1. Transform the time graph to show changes
2. Select extreme changes in a specific year (here 2003)



Visual Analytics in Data Mining



Requirements to Visual Analytics

- Space- and Time-awareness
- Work with complex multidimensional data
- Support for uncertain and missing data
- Scalability
- Support and encouraging of several complementary views on the same data
- Dynamic linking and coordination of several data displays
- From the overall view to particulars of interest
- From idea generation to hypothesis testing using statistical methods, followed by reporting

Potentially useful tools for MSTD

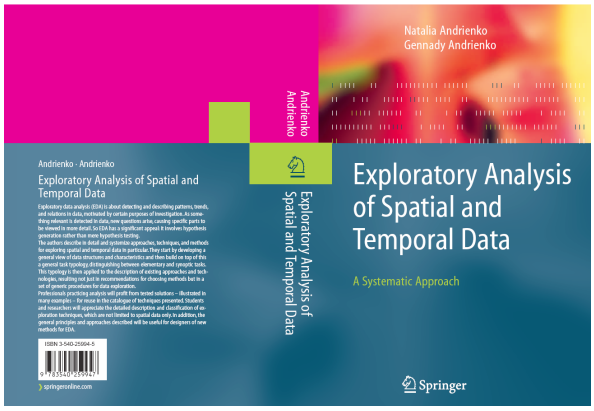
- Information visualisation tools, for example, HCE & TimeSearcher from HCIL, Univ. Maryland
- Geovisualisation tools, for example GeoVistaStudio (Penn State Univ.) and Descartes/CommonGIS (Fraunhofer Institute AIS)
- Graphical statistics tools, for example, Manet & Mondrian (Augsburg Univ.)
  - ☞ Usually such systems are research prototypes that implement innovative ideas, but provide restricted functionality and limited user support

Still open issues (for all tools!)

- Work with qualitative (non-numeric) data
- Work with fuzzy, uncertain, and missing data
- Continue scalability efforts
- Support in processing and management of findings: recording, structuring, browsing, searching, checking, combining, interpreting…
- Help in visual communication of derived data, constructed knowledge, and recommended decisions
- Adaptability to user, data, tasks, and hardware
- Embedding intelligence into software for helping users and avoiding cognitive overload

EDA: from Practice to Practical Theory

- ✓Data
- ✓Tasks
- ✓Tools
- ✓Principles



to appear ≈ end 2005

Visual Analytics & Data Mining

1. Do they need each other?
2. How to benefit from combining two scientific disciplines and related technologies?
3. How to develop each of two scientific disciplines for achieving a synergy?