# Scalable Methods for Visual Analysis of Massive Movement Data

Natalia Andrienko and Gennady Andrienko
Fraunhofer Institute IAIS
Schloss Birlinghoven
53754 Sankt Augustin
Germany

One of common shortcomings of current geovisualisation techniques and tools is the lack of scalability with respect to data volume, while practical needs often require exploration and analysis of very large dataset. Our goal in one of ongoing research projects is to design and develop scalable techniques and tools to support visual analysis of very large collections of data about movement of discrete entities, i.e. change of the spatial positions of the entities over time. Purely visual and interactive techniques involving visual representation of individual data items are insufficient for this purpose. On the one hand, limitations are posed by the display size and resolution, the speed of rendering, and sometimes even the size of computer memory. On the other hand, even when the technology permits, a display of large number of data items may be incomprehensible because of human perceptual limitations. Hence, there is a need to combine visualisation with the use of the database technologies and computational methods of data processing and analysis.

In our study, we undertake a systematic approach to the design of methods and tools for visual analysis of massive movement data. The main idea is that software tools prepare and visualise the data so that the human analyst can detect various types of patterns by looking at the visual displays. In order to facilitate the detection of patterns, it is necessary to understand what types of patterns may exist in the data. We define the possible types of patterns in such a kind of movement data on the basis of an abstract model of the data as a mathematical function that maps entities and times onto spatial positions. Then, we look for data transformations, computations, and visualisation techniques that can facilitate the detection of these types of patterns and are suitable for very large datasets, possibly, not fitting in the computer memory. Under such a constraint, visualisation is applied to data that have been previously aggregated and generalised by means of database operations and/or computational techniques. Another reason for aggregation and generalisation may be preserving privacy, which is often an issue in data analysis.

According to the results of the systematic design, we are implementing several prototype tools, which will be demonstrated at the conference.

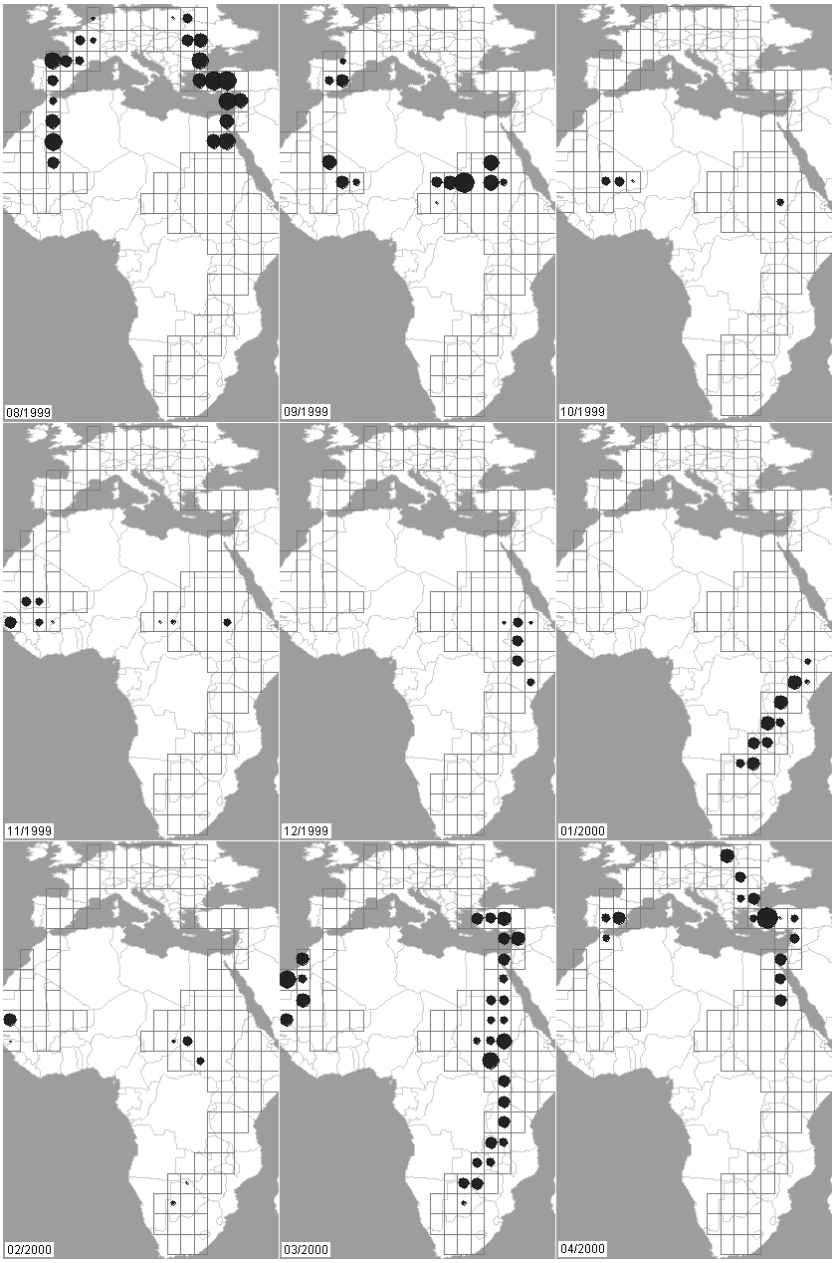


***Figure 1:*** Example: A visualisation of data (average speed) about seasonal migration of storks aggregated spatially by cells of a regular rectangular grid and temporally by months.