Visual Exploration of Extreme Events in Large, Heterogeneous Climate Data

Thomas Nocke
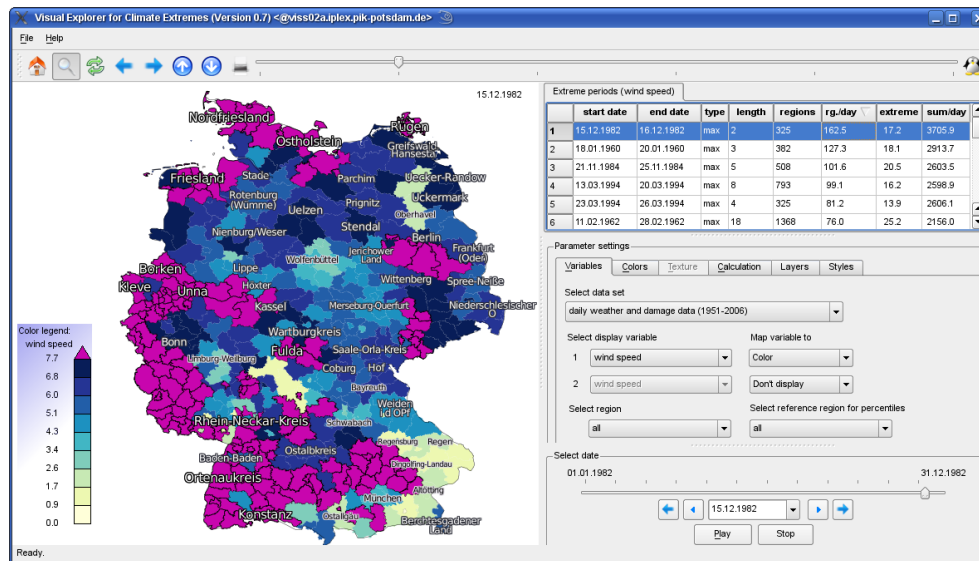

Fig. 1. Visual Explorer for Climate Extremes, linking a map highlighting extreme regions and a sortable table with extreme properties

Abstract— Interactive visual analytics of large, heterogeneous time-dependent, geo-referenced data is still a challenging problem. On the one hand, visual analytics methods are not yet established in science context such as climate research. On the other hand, the combined application of interactive methods provided by visualization and geographic information systems and of automated mining methods is still hampered for non-expert users. As a solution, this paper presents an interactive tool for the analysis of extreme events in heterogeneous climate data sets. To support extreme event detection and comparison, it provides interactive extreme definition and parallel calculation, an interactive map linked with a sortable table of extreme events with their properties and a variety of mapping opportunities. The tool integrates ensemble data of different grid types and time scales and is adaptable to different user groups and computing resources.

Index Terms—Visual analytics, visualization, geographic information systems, extreme events, climate research.

---

1 INTRODUCTION

On the one hand, the typical procedure for the visual analysis of large climate data sets by climate researchers is to reduce / aggregate the data to a small amount and then visualize it as a static map. On the other hand, for the presentation of climate data, online portals have been designed that represent a large number of (partially interactive) climate related maps with few analysis functionalities only (e.g. the Climate Wizard, http://www.climatewizard.org). However, an interactive visual exploration of heterogenous climate data with a multitude of interactive representations (see e.g. Kehrer et al. [2]) has not yet become practice in this field. This has different reasons:

- the advantages of sophisticated visual analytics methods and tools are rarely known in scientific practice,

- data heterogeneity such as

    - different grid types (regular and rotated grids, region based shape data, scatted data)

    - different time scales (hourly, daily, monthly, annual)

    - ensemble data (different climate scenarios)

    is not easily resolvable by scientist users with a single system,

- the size of data to be queried interactively requires fast data structures and caching mechanisms.

To fill this gap, we introduce a new tool for the visually based analysis of extreme events in large, heterogeneous climate data.

1.1 Background

In a joint project, German insurance companies[1] and climate researchers investigate extreme events in meteorological and meterology-driven insurance damage data, analyse dependencies and propagate this knowledge to future extreme events. To inter-compare spatial and temporal extension of extreme events, a variety of raw data from different sources (measurements, insurance damage costs, re-simulation of past events, simulation of future climate) needs to be investigated, such as

- monthly past damage and meteorologic measurement data (439 German districts and selected stations, 56 years, 14 variables),

- *Thomas Nocke is PostDoc at Potsdam Institute for Climate Impact Research, E-mail: nocke@pik-potsdam.de.*

---

[1] Gesamtverband der Deutschen Versicherungswirtschaft e.V. (GDV)

- daily past flood related data (CLM model, 37x49 grid resolution, 30 years, 9 variables).

- daily future meteorology data (Star model, 439 German districts, 54 years, 11 variables, 9 scenarios).

1.2 Challenges

Data size and heterogeneity is most challenging in this context. Due to the specific metadata characteristics, a certain visualization or geographic information system might be the best choice. In practice, however, a scientist typically is familiar with one visualization system only. Thus, the user's flexibility of using interactive visualization techniques is strongly restricted, and she/he tends to static solutions.

Furthermore, the visual mapping of different climate variables and of extremes is not trivial. This includes the mapping of extreme and non-extreme values in focus & context as well as the visual comparison of different variables / time steps.

Moreover, the tool designed must be scalable to different use cases with varying user groups, tasks and available resources, such as the data analysis by a climate researcher on a Linux visualization server machine or the result presentation or restricted data analysis by an insurance expert on a typical Windows desktop/notebook computer. Finally, data privacy is an important issue, because different user groups have different data set access rights (e.g. to see a reduced temporal data resolution only).

2 THE VISUAL EXPLORER FOR CLIMATE EXTREMES

Major tasks to be supported by the developed tool are (1) the interactive definition of extremes, (2) the fast and easy-to-use access to a multitude of extreme events in space, time, ensembles, and (3) the visual comparison of data extremes from different data sources. As a solution to the described requirements, we introduce a integrated, interactive visual analytics tool.

2.1 Approach

To enable the user to interactively navigate through a large number of extreme events very quickly, we provide a table representing extreme period properties and a linked geographical map representation (see Fig. 1). In addition, using time slider and animation, she/he can navigate through the temporal extend of the events, or analyze post- and preconditions of extreme events or extreme event transitions. Parts of the data without extremes can be inspected as well, for instance if an extreme event might be expected in a certain time period and was not represented by the data / the extreme value limits.

To support the comparison of regional distributions of different variables at an extreme event period, we support side-by-side image based comparison and bivariate mapping using coloring and texturing. In this context, the application of color/texture maps separating extreme values from non-extreme values improves the visual comparability of different kinds of variables.

To support users in selecting suited visual mappings, we provide a mechanism that uses a-priori knowledge about certain data variables and typical mappings in scientific practice. Therefore, the tool provides appropriate default settings both for extreme calculation and visual mappings for those variables. In particular, this includes variable specific percentiles for extreme value limit calculation, types of color maps, number of color segments and extreme value colors. All those settings can be interactively reset and stored by a user, providing parallel computing mechanism for recalculation of extreme events and state-of-the-art color maps [1].

Furthermore, certain extreme events are related to certain geographical regions, for instance floods to river catchments or insurance relevant information only at grid cells located over land. Thus, we provide mechanisms to flexibly define geographic regions of interest. This can be done using shape files providing region classification (river catchment, federal states, climatic regions, etc.) or interactively on the map. For the visual mapping of such regions, we provide a focus & context approach. Currently, we provide focus regions fully saturated (both extreme values and non-extreme values) with large region labels if

available, and for the context, only regions with extreme values are represented, using reduced color saturation and smaller region labels.

Finally, to provide access by different user groups, the graphical user interface is adaptable with respect to the number of parameter settings presented to the user, and hidden parameters are set to default values. Furthermore, we a provide multi-lingual user interface (currently English and German).

2.2 Implementation

For a fast and portable implementation, the tool bases on C++/QT. Different kinds of data sources are represented by a homogeneous similar internal representation using an object-oriented class hierarchy and polymorhism. Currently, the tools imports NetCDF, Shape and ASCII data. To speed up recurrent time-intensive data reading and pre-calculation processes, we provide a data serialization mechanism, storing homogenized and precalculated data in binary files.

To avoid user overload, the visual map representation was homogenized, mapping different grid types of input data to GIS-shapes. For the interactive geographical map representation, we use the GIS rendering API of Mapnik (http://www.mapnik.org), which provides basic, but essential GIS functionality such as geographic projections and region labeling.

2.3 Example

Figure 2 illustrates the visual comparison of spatial structures of extreme events from two different data sources with differing grids and data ranges with in the tool.
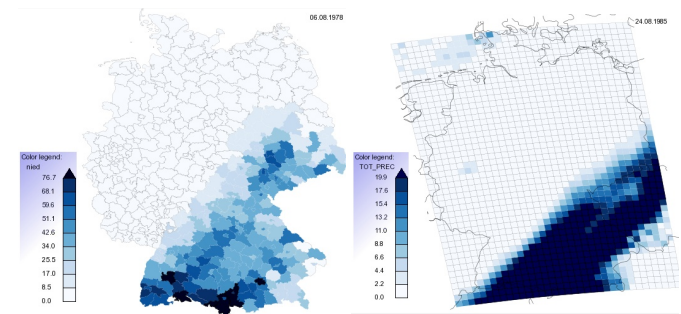

Fig. 2. Visual comparison of precipitation in measured data on district level (left) and of a simulated gridded data set (right)

3 CONCLUSION

This paper presented a new approach for visual analytics of extreme events in geo-referenced data, integrating large, heterogeneous data sources. An interactive, easy-to-use tool has been developed, which is adaptable to different use cases (user groups, resources). However, there are still challenges for future work. This includes the development of new visual mappings to both the map view (e.g. multiple maps for each time step of an extreme period) and the extreme event table (e.g. a graphical table). Moreover, for the representation of focus & context regions, we see potential for further developments. Finally, an explicit representation of the distribution of extreme events' temporal trends and frequencies seems to be promising.

REFERENCES

[1] C. Brewer, G. Hatchard, and M. Harrower. Colorbrewer in print: A catalog of color schemes for maps. *Cartography and Geographic Information Science*, 30(1):5–32, 2003.

[2] J. Kehrer, P. Filzmoser, and H. Hauser. Brushing moments in interactive visual analysis. *Computer Graphics Forum*, 29(3):813–822, 2010.