Visual integration of diverse environmental data: A case study in Central Germany

C. Helbig
Karsten Rink
A. Marx
Joerg A. Priess
M. Frank

See next page for additional authors

Follow this and additional works at: https://scholarsarchive.byu.edu/iemssconference

Helbig, C.; Rink, Karsten; Marx, A.; Priess, Joerg A.; Frank, M.; and Kolditz, Olaf, "Visual integration of diverse environmental data: A case study in Central Germany" (2012). International Congress on Environmental Modelling and Software. 87. https://scholarsarchive.byu.edu/iemssconference/2012/Stream-B/87

This Event is brought to you for free and open access by the Civil and Environmental Engineering at BYU ScholarsArchive. It has been accepted for inclusion in International Congress on Environmental Modelling and Software by an authorized administrator of BYU ScholarsArchive. For more information, please contact scholarsarchive@byu.edu, ellen_amatangelo@byu.edu.
Visual integration of diverse environmental data: A case study in Central Germany

C. Helbig\textsuperscript{a,b,c}, K. Rink\textsuperscript{a}, A. Marx\textsuperscript{a}, J. Priess\textsuperscript{a}, M. Frank\textsuperscript{c} and O. Kolditz\textsuperscript{a,b}

\textsuperscript{a}Helmholtz Centre for Environmental Research - UFZ, Permoserstraße 15, 04318 Leipzig, Germany (carolin.helbig@ufz.de, karsten.rink@ufz.de, andreas.marx@ufz.de, joerg.priess@ufz.de, olaf.kolditz@ufz.de)

\textsuperscript{b}Technical University Dresden, Faculty of Forest, Geo and Hydro Sciences, 01062 Dresden, Germany

\textsuperscript{c}University of Applied Sciences Leipzig, Faculty of Computer Science, Mathematics and Natural Sciences, Gustav-Freytag-Str. 42a, 04277 Leipzig, Germany (mfrank@imn.htwk-leipzig.de)

Abstract: Land use and climate changes are some of the most important phenomena which will influence our life in the coming decades. To analyze these phenomena and the projections of their impact, diverse and dynamic data are collected and complex projects for observation and measurement are being developed (e.g. GLUES and TERENO projects). The goal is to find an appropriate way to visualize measured data in combination with simulation results. We propose to develop a generic methodology that can be used beyond the envisaged case study region ‘Central Germany’ (Thuringia / Saxony / Saxony-Anhalt). The basis of the visualization consists of raster and vector data for climate, land use, hydrology, agrology, etc. Methods involved in the process of assimilation include up- and downampling as well as conversions regarding the reference system. The specific user groups of the application are scientists of different domains, stakeholders with expert knowledge and students of the corresponding research fields. They are strongly involved in the data analysis, so that the resulting interactive visualization could be developed according to their requirements. The presentation of the data is geo-referenced across different layers, which include measured data and simulated data for the focused investigation. It also provides additional information in terms of statistics with regard to the scenarios. Layers of simulated data, for example from different scenarios of land use, could be compared and changes could be examined in time response. The visualization will be integrated into an existing framework for data exploration and modeling.

Keywords: Visualization; Environmental Data; Central Germany; OpenGeoSys Project

1 INTRODUCTION

To understand the processes and consequences of land use and climate change it is necessary to analyze the results of the research conducted in various scientific fields. These results are available in the form of observed and simulated data from the fields of climate research, physical and anthropogenic geography, hydrology and socio-economic studies.

In order to adjust our land use management options to the changing environmental conditions it is necessary to organize a far-sighted planning at regional (Hauck and Priess
and global (Perez-Vega et al. [2012], Priess et al. [2010]) levels. Therefore it is required that projections and forecasts regarding land use (Hoymann [2010], Albrecht et al. [2009]), climate (Meinke et al. [2010]) and soil (Michael et al. [2010]) should be included. Various deviations and uncertainties of the corresponding models should be considered, too (Smiatek et al. [2009])). Furthermore, there are complex projects for observation and measurement of environmental data being developed, for instance TERENO (Zacharias et al. [2011]) and GLUES (Seppelt [2011]).

The visualization of diverse environmental data in combination with parameterization and analysis tools allows to combine the strengths of human judgment of visual representations with electronic data processing (Andrienko et al. [2011], Thomas and Cook [2005]). These approaches are subsumed under the term geospatial visual analytics. The development tends to the opportunity of seamless integration of 2D and 3D data in visual analytics tools (Breunig and Zlatanova [2011]). There are more and more applications and projects that deal with the combination and management of heterogeneous data (Mäs et al. [2011], Klimke et al. [2011]). The software framework which has been developed and used for geospatial visual analytics in this paper is OpenGeoSys Data Explorer (Kolditz et al. [2012], Rink et al. [2012], see 2.1).

2 METHODS

The basis of the development of the visualization is shaped by the requirements of the users. There are three different user groups: experts, stakeholders and students. The experts aim is to evaluate and analyze their research results with the help of analysis tools. The requirements of the stakeholders needs are to evaluate different scenarios which are the basis of their decisions. Finally, the user group of students is supposed to use the visualization in the course of study for better understanding of complex correlations. In this first step of developing the visualization methods, we focused on the user group of experts. For this matter we conducted interviews with experts to discover their requirements to the visualization. A further challenge is to design the visualization in such a way that it could be applicable to all of its three stages: the explorative analysis (data is available, different visualization methods are used for discovering hypotheses), confirmative analyses (will be used to prove hypotheses) and communication (as basis to present and discuss the results with other parties, Schuhmann and Müller [2000]).

The visualization will be developed in such a way that it can be used for a variety of
Figure 2: Combination of various input data for Central Germany

environmental data. Thus, various data can be visualized in a combined manner and unknown correlations between the data can be illustrated. NetCDF is a common file format for the data from the fields of climate, hydrology and geology. For that reason we implemented a NetCDF interface (see 2.2). The data consists of measured (discrete) and simulated (continuous) data. Figure 1 shows the combination of the continuous and discrete data in a geographic context. The layers of the various data (see 2.2) that is available for different scales (see 2.3) can be visualized for analysis.

2.1 OGS Data Explorer

OpenGeoSys (OGS) is an open source software for the numerical simulation of thermo-hydro-mechanical-chemical processes. The OGS Data Explorer is part of the current OGS version and a tool for 3D data exploration and mesh generation for FEM simulation. It provides a wide range of input and output formats. The user can import raster, vector, subsurface, graphic and observation data as well as data from databases. Import and export are handled using for interfaces (e.g. GMESH, http://geuz.org/gmsh/) and native files (e.g. xml). The export of the visualization is possible in such graphic formats as VTK, OpenSG and VRML.

The development of the OGS expansion is based on the interviews with experts of different domains, for instance climate, land use, hydrology and agrology. In the study region the data for land use is available in geo-tiff-format, the climate data in NetCDF-format. There is already an existing interface for geo-tiff, but it was necessary to implement a universally valid NetCDF interface.

2.2 Integration of heterogeneous data

The aim of this study is to combine various input data sets for the study area (see 2.3) from the fields of hydrology, land use, remote sensing, administrative divisions, climate and geology so that the output can be used for the numeric simulation of hydrological and terrestrial processes (see figure 2). One point to consider is the transformations between different coordinate systems. For example, NetCDF is usually available in the World Geodetic System 1984 (WGS 84), data used in hydrology is usually in Gauß-Krüger coordinate system. For the transformation from WGS 84 to Gauß-Krüger we use the Geospatial Data Abstraction Library (GDAL, http://www.gdal.org/). This tool is only partially suitable for NetCDF transformations because the transformation is limited to a single time step. Files with more time steps become unprofitable changed in their structure. Another point to consider is the up- and downsampling of various data sets to adjust their resolution. Depending on the specific task and the used method, this can be an advantage. For example, in the case with downsampling it can be profitable...
to use mean values. On the other hand downsampling can be a disadvantage, if used for measured data because detailed information can get lost. Up- and downsampling is made with Esri ArcMap (http://www.esri.de/). To compare different time slices of the same or even different data sets, a color table with predefined common colors for the specific matter is used.

2.3 Different scales

The study area is divided into different scales with an increasing data resolution: Germany, Central Germany (Saxony, Saxony-Anhalt, Thuringia) and the Bode catchment (see figure 3). For each scale there are different data sets available. For Germany, there is simulated climate data with temperature and precipitation rates from 1960 to 2100 based on different IPCC scenarios and observation data from meteorological stations from 1960 to 1990. In addition, information about its topology, water network and administrative divisions (e.g. federal states, administrative districts) is available. The data for Central Germany consists of land use data classified by CORINE from 1990 and 2006, information about protected areas with their different classifications (e.g. biosphere reserve, national park, landscape conservation area, nature protection area, natural park, FFH area, SPA area), information about the soil (e.g. main soil type, classifications by Working Group of the Federal States on Waste) and administrative information (e.g. population). For the Bode catchment we have data about boreholes, land cover, groundwater etc.

3 Preliminary results

To get an overview over the focused climate data in Germany and for its dynamics over time, we used Esri ArcMap to produce an animation over the different time steps. Figure 4 illustrates three slides of the animation depicts the difference of the moving average (starting in the year 1960) over ten years to the temperature expected by the IPCC scenario A1B. The moving average is calculated based on temperature in an interval of 10 years (e.g. 1962-1972 or 1963-1973). The difference of the running mean temperature shows how the temperature with respect to the year 1960 as a starting point decreases or increases. The first slide indicates the simulated mean difference of the temperature from 1962-1972, the second slide from 2025-2035 and the third slide from 2088-2098. This animation is used to identify significant areas where the temperature changes are particularly strong or weak, to give these areas special attention in the following data visualizations. These visualizations will help compare the measured and simulated data. The identified areas are the Alps, the low mountain range and the coast.

NetCDF is a common binary file-format used for climate data which consists of variables representing their physical quantities and have a number of dimensions which stand for the axis information, e.g. latitude, longitude and time. For the interface the NetCDF C++
Figure 4: Example slides from an animation that shows the difference of the running mean temperature from 1962-1972, 2025-2035 and 2088-2098 expecting the IPCC scenario A1B.

Figure 5: The bars represent the difference between simulated and measured temperature (mean values for the period from 1960 to 1990). The map shows the simulated temperature.

Interface library (http://www.unidata.ucar.edu/software/netcdf/docs/netcdf-cxx/) is used. The interface can read out the available variables of the selected file and select a suitable one. In the next step it is necessary to select the dimensions that are attributed to latitude and longitude. These dimensions are essential for displaying the file in OGS Data Explorer. If additional dimensions are available, they can be selected, too (e.g. the value of measured temperature). OGS creates a mesh out of the NetCDF file, the user can define the element type (quad or triangle); the intensity (z-value) can be used as material property of mesh elements or elevation.

To validate the simulated climate data, we combined the simulation results and the observation data from meteorological stations. For this matter we created a mesh of the simulated data by importing it from a NetCDF-file and using its z-value as elevation. To import the station data to OGS, we had to convert the data in the csv file to GMS (groundwater modeling system) format. For the comparison of simulated and measured data we calculate the difference between the value at the station and the value at the mesh node closest to the station. The absolute value of the temperature difference is represented by the length of the bars, the color signifies different class intervals (bins) for an easy-to-follow visualization of all stations.

In figure 5 we combined the differences between the simulated and measured temperature (bars) at a height of 2 meters with the visualization of the simulated temperature from 1960 to 1990 (texture) and added a shape file of Germany for a better visual orientation. The differences lie between the mean of the simulated temperature from 1960 to
Figure 6: The bars represent the difference between simulated and measured temperature (mean values for the period from 1960 to 1990). The mesh is based on the simulated temperature.

1990 and the mean of the measured temperature of the meteorological stations in this time. The visualization with texture and bars allows to analyze the data of the individual stations. Using a common difference map instead of this one will lead to the loss of information because of the interpolation. Furthermore, this kind of visualization is beneficial for analyzing the station data relating to other data (e.g. land use, precipitation, etc.). As can be shown, the station on the Brocken mountain in the middle of Germany shows the highest difference. The visualization demonstrates that there is a correlation between the level of temperature and the amount of the difference. In Central Germany, the higher bars occur in the regions with medium temperature (green). Figure 6 shows another facility of visualizing the differences and temperature by using the latter as elevation and adding the bars of the differences to the z-value of the points. This visualization provides the approval of the correlation.

Now we zoom into the area of interest and focus on the region of Central Germany where the land use and DEM data is available. The combination of the differences (bars) among the land use data classified by CORINE in 1990 (see figure 7) shows that the highest differences are always characteristic of forests and semi-natural areas (green color) or very close to them. The visualization of the differences (bars) together with the DEM (see figure 8) leads to the recognition that there are proportionally more high bars in mountain areas (black) than in other areas. Even in the western of Saxony-Anhalt there are more bars that represent a high difference.

4 Conclusion and Outlook

In this paper we described our present state of developing a visualization concept that integrates diverse environmental data from different scientific disciplines such as climate research, land use, hydrology, geology, etc. This concept is developed on the basis of specific data and applications which are presented in this paper. We used OpenGeoSys, the open source software for data integration and numerical analysis, as geospatial visual analytics tool. The study is conducted at three different scales (Germany, Central Germany and the Bode catchment) that differ in the availability of various data. The aim of the visualization is to combine, analyze and validate measured as well as simulated data and derive new insights out of it.

In the future development we plan to enlarge the visual analytic methods. For example we are going to provide tools to filter and highlight pre-defined patterns and uncertainties in the data to analyze it. In addition to these tools, we plan to integrate diagrams for
Figure 7: The bars represent the difference between simulated and measured temperature (mean values for the period from 1960 to 1990). The map shows the land use based on CORINE in 1990.

Figure 8: The bars represent the difference between simulated and measured temperature (mean values for the period from 1960 to 1990). The map shows the digital elevation model.

detailed information of the focused data in terms of statistics. To support user interaction with the data, we are going to integrate the functionality to modulate the scenarios by changing various parameters. The complete process of developing these concepts and tools will take place in cooperation with the prospective users and their requirements.

To extend the amount of the integrated data for our case study region, we plan to integrate simulated soil moisture data from the field of hydrology and simulated land use data from the field of landscape ecology. This part is already in progress. Since alternative energy resources become more and more important, we want to integrate for example data from wind forecast, geothermal energy and bio-energy into our case study to foster decision support for sustainable land management.

**ACKNOWLEDGMENTS**

The authors gratefully acknowledge climate data support of CERA (Climate and Environmental Retrieving and Archiving). The first author would also like to express her gratitude to the ESF (European Social Fund) "Knowledge and Knowhow Transfer" for the funding of the scholarship. We thank HIGRADE, the graduate school of UFZ, and the school for doctoral students at the HTWK. We are especially obliged to Prof. Lars Bernard from the Technical University of Dresden for this professional support.
REFERENCES


