Multidisciplinary Research on Geographical Information in Europe and Beyond
Proceedings of the AGILE'2012 International Conference on Geographic Information Science, Avignon, April, 24-27, 2012
ISBN: 978-90-816960-0-5
Editors: Jérôme Gensel, Didier Josselin and Danny Vandenbroucke

A Visual Analytics Approach to Validate Geoscientific Simulation Ensembles

Andrea Unger
Helmholtz Center Potsdam GFZ
Telegrafenberg
Potsdam, Germany
unger@gfz-potsdam.de

Volker Klemann
National Oceanographic Centre
6 Brownlow Street
Liverpool UK
volman@noc.ac.uk

Sven Schulte
Helmholtz Center Potsdam GFZ
Telegrafenberg
Potsdam, Germany
schulte@gfz-potsdam.de

Doris Dransch
Helmholtz Center Potsdam GFZ
Telegrafenberg
Potsdam, Germany
dransch@gfz-potsdam.de

Abstract

A common analysis step in the assessment of geoscientific simulation models is to identify suitable parameter values of the model, which replicate the behavior of the real process with sufficient accuracy. This is a challenging step because observed and modeled values, which are heterogeneous in temporal and spatial resolution and often varying in preciseness, need to be compared as well as multiple models in time and space. To handle these challenges for the specific application of modeling sea-level heights in the last 100,000 years, we introduce a tailored multiple view visualization concept that handles, on the one hand, the multitude of models that needs to be dealt with, and, on the other hand, allows to assess spatio-temporal variations and varying preciseness in observed and modeled data. The concepts were developed in close cooperation with researchers from visualization and the earth sciences. The software tool that implements the concepts is already in use for model and data analysis in the described application.
Keywords: Visualization, visual analysis, geoscientific modeling, simulation model assessment, simulation model comparison, spatio-temporal data, heterogeneous data.

Flanking data analysis in Earth-System Sciences by advanced visualization methods is a striking feature due to rising complexity, amount and variety of available data.

In this poster, we present a visual analysis concept that is tailored to support earth scientists in a common, but challenging analysis step in model assessment: finding suitable parameter values for a geoscientific simulation model. This step is performed to ensure that the simulation model represents the real process with sufficient accuracy. The sought parametrization can often not be directly determined. Thus, the accuracy of a model is assessed by comparing the model output – the resulting simulation data – to observation data of the real process. To identify suitable parameter values, scientists thus need to compare multiple model outputs.

The assessment of geoscientific models is challenging because modeled and observation data, which are to be compared, are heterogeneous in temporal and spatial resolution and often varying in preciseness. Moreover, also multiple models have to compared in time and space.

In our example application, sea-level variations in the last 100,000 years are studied. In this application, available observation data are derived from fossil samples found along the coasts worldwide. They give indications about the sea level variations at a certain geographical location. While the location is relatively precise, the age of the sample is determined with a substantial error due to the dating method. The fossil's indication of former sea level can be best described by a possibility range, which is bounded (peat of a salt marsh) or half-bounded (the living condition of a shell is limited by the mean sea level). This impreciseness in the data hinders a gross analysis by means of statistical methods [2].

Our visualization addresses these challenges by bringing together all necessary data, including model parameters, model outputs, observation data as well as the results of automated analysis, and making them available in an interactive environment of multiple linked views. The concepts were derived in close cooperation between researchers from the fields of visualization and the Earth sciences and realize a previously introduced conceptual framework [3]. Assessing the validity within a model ensemble directly extends prior work that focused on the assessment of a single model parameterization [1].

Our visualization tool employs an overview and detail visualization strategy (see Figure 1): An overview gives a general impression of the dependency between parameterizations and accuracy of the output for the model set. Additional detail views allow to assess spatio-temporal variations for individual model outputs.

The overview is provided by scatterplots that directly relate parameter values and model accuracy for a set of model parameterizations. The accuracy value is determined by an analytical component, which bundles statistical methods to automatically perform an analytical comparison of modeled and observed val-

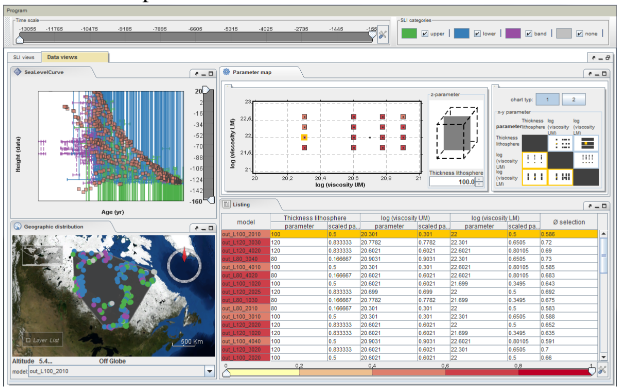

Figure 1: A screenshot of our visual analysis tool. The views on the right hand side give an overview on the dependency between model parameters and model output accuracy. On top, a scatterplot shows the variations in the model accuracy (encoded by red color) in relation to two parameter values (mapped to the axes of the scatterplot). Cube and scatterplot matrix serve for the selection of two parameters from the parameter set. The table below is a redundant view of the parameter and accuracy values for each model, encoding the accuracy of model parametrization again by red color. On the left, a time-value-plot and a virtual globe show a detailed comparison of modeled and observed values.

ues. First, for each sample of the model outputs, a "goodness of fit" is derived. This value reflects whether the sample from the model output matches the corresponding observation. This goodness of fit of individual samples is the basis to compute the average fit for each model parametrization. The overview visualization does not directly include the spatio-temporal context, as the main goal is to show the dependency between parametrization and fit.

To explore spatio-temporal variations, we use two strategies: As one strategy, the user can interactively adjust the selection of data samples that is considered in the average fit. This is done in several views that allow to explore the spatio-temporal distribution and varying quality of the observation data.

As the second strategy, we support the user in comparing the observation data with a single model output, which is selected from the overview. This detail visualization comprises a comparison of individual observed and modeled sea level heights both in time (using a time-value plot) and space (using a virtual globe). Overview and detail views are closely linked to provide an instantaneous back and forth between them.

Feedback from geoscientists implies that the tool is a valuable component in the validation task, as it provides direct access to large, heterogeneous data and the interrelations among model parametrization, goodness of fit of model outputs, time, space, and varying quality of observation data.

Our implementation, which employs the efficiency of databases to retrieve and store data, is currently able to interactively explore over 50 models, each characterized by three numerical parameters and over 14,000 output samples, and observation data containing 14,000 samples with around 30 attributes each. It now serves as an analysis tool for the evaluation of model simulations as well as for the evaluation of imprecise fossil samples in studies of past sea-level variations.

Although our visualization tool was developed with a focus on one application, the visualization concepts address common challenges in the analysis of heterogeneous data with spatio-temporal context and varying quality. As a next step, our aim is to generalize these concepts to further geoscientific models.

References

[1] Doris Dransch, Patrick Köthur, Sven Schulte, Volker Klemann, and Henryk Dobslaw. Assessing the quality of geoscientific simulation models with visual analytics methods - a design study. *International Journal of Geographical Information Science*, 24(10):1459–1479, 2010.

[2] Volker Klemann and Detlef Wolf. Using fuzzy logic for the analysis of sea-level indicators with respect to glacial-isostatic adjustment: An application to the richmond-gulf region, hudson bay. *Pure and Applied Geophysics*, 164:683–696, 2007. 10.1007/s00024-007-0191-x.

[3] Andrea Unger, Sven Schulte, Volker Klemann, and Doris Dransch. Poster: Visual and Analytical Methods to Assess Geoscientific Model Configurations. Poster at IEEE Conference on Visualization (Vis), 2011.

Multidisciplinary Research on Geographical Information in Europe and Beyond
Proceedings of the AGILE'2012 International Conference on Geographic Information Science, Avignon, April, 24-27, 2012
ISBN: 978-90-816960-0-5
Editors: Jérôme Gensel, Didier Josselin and Danny Vandenbroucke