

VISUAL EXPLORATION AND EVALUATION OF CLIMATE-RELATED SIMULATION DATA

Thomas Nocke, Michael Flechsig, Uwe Böhm

Potsdam Institute for Climate Impact Research (PIK)
Telegraphenberg A31
14473 Potsdam, Germany

ABSTRACT

Large, heterogeneous volumes of simulation data are calculated and stored in many disciplines, e. g. in climate and climate impact research. To gain insight, current climate analysis applies statistical methods and model sensitivity analyzes in combination with standard visualization techniques. However, there are some obstacles for researchers in applying the full functionality of sophisticated visualization, exploiting the available interaction and visualization functionality in order to go beyond data presentation tasks. In particular, there is a gap between available and actually applied multi-variate visualization techniques. Furthermore, visual data comparison of simulation (and measured) data is still a challenging task. Consequently, this paper introduces a library of visualization techniques, tailored to support exploration and evaluation of climate simulation data. These techniques are integrated into the easy-to-use visualization framework SimEnvVis - designed as a front-end user interface to a simulation environment - which provides a high level of user support generating visual representations.

1 INTRODUCTION

Increasing computing power enables modelers to generate larger simulation data sets. When trying to identify their underlying properties – such as patterns and statistical relations between the variables – exploration increasingly becomes the bottleneck. Furthermore, there is an increasing demand for methods which support the evaluation of the reliability of conclusions based on simulation data.

For instance, in climate research a combination of statistical (e. g. clustering) and visualization methods (e. g. function plots, 2D-maps) are applied to explore the simulation data and to evaluate the reliability of models (e. g. analyzing multi-run experiment output). However, the application of visualization focuses on standard techniques. These provide easily understandable plots such as time series graphs, bar charts and (animated) colored 2D-maps,

ensuring their comprehension by a large audience. However, these plots are mainly restricted to uni-variate data and analysis of special parts of a large data set at once, often leading to a loss of overview. Thus, they are only of limited use detecting possibly interesting, partly unknown features in the multi-variate data, such as model inconsistencies or specific model configurations of interest. Furthermore, standard model comparison and evaluation techniques typically display difference plots or juxtaposed images, and may be ineffective for in-depth data diagnosis tasks. Here, drawbacks occur because the underlying data values are no longer visible (in difference images) and comparing individual values or simulation data for small regions may be difficult or even impossible in juxtaposed images.

Typical systems in the climate context for producing graphical representations range from Excel and Matlab, general purpose visualization systems and geographic information systems (e. g. ArcGIS (<http://www.esri.com/software/arcgis>)) to special purpose visualization solutions. However, there is a gap between the available methodology of sophisticated visualization systems and the applied visualization functionality in this field. In particular, due to the size of the simulation data, researchers tend to display their data sets statically in one image. Thus, the immense potential of interaction functionality (brushing & linking, overview & detail, focus & context) is exploited only marginally.

Therefore, the approach of this paper is to provide a variety of visualization and interaction methods to handle, analyze and evaluate simulation data effectively. For this purpose, a library of highly interactive visualization techniques has been compiled, providing both standard techniques and techniques adapted to the special requirements of the analysis of climate and climate impact model data. Besides a variety of tailored standard visualization (spatial and temporal scalar and vector data visualization) this paper presents special solutions for the comparative visualization of model output, and for multi-variate visualization for multi-run simulation experiments.

Furthermore, problems occur in making the variety of in the climate research context partly unknown techniques available for the heterogeneous data sets (coupled multi-dimensional and multi-variate scalar and vector data on different grids). To support users in generating suitable representations, mechanisms to guide them and to semi-automatically "design visualizations" need to be provided. Therefore, the easy-to-use, modular framework SimEnvVis has been developed that enables modelers from different research backgrounds (meteorology, physics, biology,...) to get access to the various, state-of-the-art visualization techniques.

2 Related work

Visualization techniques representing simulation data have been developed for many years, and a wide variety of techniques - especially for spatial and temporal data - has been elaborated, see e. g. (Wong and Bergeron 1997) and (Mueller and Schumann 2003) for an overview. However, there are still challenges concerning certain tasks when analyzing simulation data. This is especially true comparing spatial patterns in model runs with each other and with measured data visually (Pagendarm and Post 1995). This "comparative visualization" allows the analysis of data differences between data sets on the image (Zhou et al. 2002) and on the data level (Gelin 2002). However, typical systems in the context of model simulations still focus on image level comparison.

Furthermore, the analysis of multi-run simulation experiments in high dimensional model parameter spaces is still a challenging problem. Visualizations in this context range from overview about variable statistics (bar charts, e.g. rank correlations in tornado graphs), showing parts of the parameter space (2D visualization, scatterplots) to visualization of whole high-dimensional parameter spaces (scatterplot matrices, parallel coordinates, star glyphs), see (Cooke and van Noortwijk 2000) for an overview. Although first interaction techniques are used (e. g. (Cooke and van Noortwijk 2000) apply interactive filtering to parallel coordinates to focus on model states of interest), there is still a high potential for using sophisticated interaction techniques (e. g. reordering and highlighting) to improve navigation in high dimensional spaces.

Moreover, there is an increasing demand for information about model uncertainties. To apply them in the (visual) data analysis process and to communicate simulation results and their uncertainty to decision-makers, tailored visualization techniques need to be developed which intuitively code the data and its uncertainty in the same image. Therefore, techniques displaying uncertainties have been proposed (Pang et al. 1997), (Djurcilov et al. 2001), (Griethe and Schumann 2005), but are not yet widespread in simulation background such as climate modelling.

Besides the general purpose development of visualization methods for simulation data, it is necessary to adapt to the application context and its specific conventions and tasks. For meteorological and climate data in particular, a variety of visualization techniques have been developed (see (Schröder 1997) and (Treinish 1999) for an overview) and adapted to the background specifics (see e. g. (American Meteorological Society 1993)). Techniques for the visualization of meteorological and climate data have for instance been built for scattered data (Treinish 1994), for scalar data on 2D-regular grids (Stier et al. 2005), for scalar data on 3D-regular grids (Riley et al. 2003), for vector data (Griebel et al. 2004) and for data in their temporal context (Saito et al. 2005). Additionally, some methods have been elaborated to visualize the data in a multi-variate way (Macédo et al. 2000), (Doleisch et al. 2004), (Stier et al. 2005). However, the application of sophisticated interaction techniques in meteorological and climate studies is restricted to a few examples (Treinish 1999), (Macédo et al. 2000), (Doleisch et al. 2004).

To give researchers access to such techniques, there is a variety of monolithic tools and toolkits which specialize in visualizing meteorological and climate data (Generic Mapping Tools (GMT, <http://gmt.soest.hawaii.edu>), GrADS (<http://grads.iges.org/grads>), MAGICS (<http://www.ecmwf.int/publications/manuals/magics>), McIDAS (<http://www.unidata.ucar.edu/software/mcidas>), Metview (<http://www.ecmwf.int/publications/manuals/metview/index.html>), Ocean Data View (<http://odv.awi-bremerhaven.de>), Vis5D (<http://vis5d.sourceforge.net>) and FERRET (<http://ferret.wrc.noaa.gov/Ferret>). Due to their easy-to-use interfaces to plug in simulation data sets, these tools are widespread among certain sub-disciplines (e. g. GMT and Ferret for oceanographic research). However, in comparison with current visualization systems and toolkits (e. g. AVS (<http://www.avs.com>), OpenDX (<http://www.opendx.org>), InfoVis-Toolkit (<http://ivtk.sourceforge.net>), Spotfire (<http://www.spotfire.com>), there are some limitations which hamper new insights into hidden structures in the multi-variate model output, especially because they focus on the spatial visualization of individual variables.

3 A visualization technique library

3.1 General approach

The approach of this work is to compile a library of visualization techniques, and thus, making a variety of methods available for climate modelers which are suitable for exploring and evaluating large sets of multi-variate data. Such a library reduces the requirements for a user to have knowledge of internal data structures and flows as well as of current visualization developments. To increase the user acceptance of such a library, background specific standard

techniques are provided, enriched by a set of alternative (multi-variate) visualizations. Challenges compiling such a library are the heterogeneity of input data sets (especially to achieve a high error tolerance, e.g. for simulation data with inconsistencies) and the amount of data (handling large data sets in the visualization process interactively).

Therefore, to reduce development costs, the functionality of existing visualization systems has been used, supplying easy-to-use interfaces. However, for large data, standard systems are subject to efficiency restrictions. Thus, optimized library techniques can be implemented directly, enabling direct access to data structures and hardware resources. This flexible library design allows the coupling of different systems and programming languages and the combination of generally applicable visual methods with the domain specific requirements (e.g. adapting terms or background dependent color maps), reusing existing visualization software if necessary.

Consequently, a library was compiled which supports a variety of input data characteristics (1D, 2D, 3D, ND, different grid types, ...) and analysis tasks, focusing on exploration and evaluation tasks, providing overview techniques for spatial and temporal data, enabling users to filter their data and dynamically reload details on demand.

3.2 Standard visualization techniques for climate model simulation output

To supply a range of standard visualizations with a manifold of functionalities to (climate) modelers, the following visualization modules for base data classes have been provided (using OpenDX):

- 1D visualization: a set of time series plots (analysis of single temporal variables and comparison of variables)
- 2D/2.5D visualization: combined color, isolines and height mapping (analysis of spatial distributions and relations)
- 2D flow visualization: streamlines, glyphs and color mapping (analysis of different features of a 2D vector field)
- 3D visualization: intersecting planes, parallel planes, isosurfaces, decomposition methods, volume rendering

On one hand, these techniques include mechanisms for adapting the generated representations to the background and to the data specifics, for instance providing suitable color maps, adding additional geographic information (for a better orientation), mapping the data to the globe (to analyze conditions around the poles) and changing between interpolated and discrete data mapping.

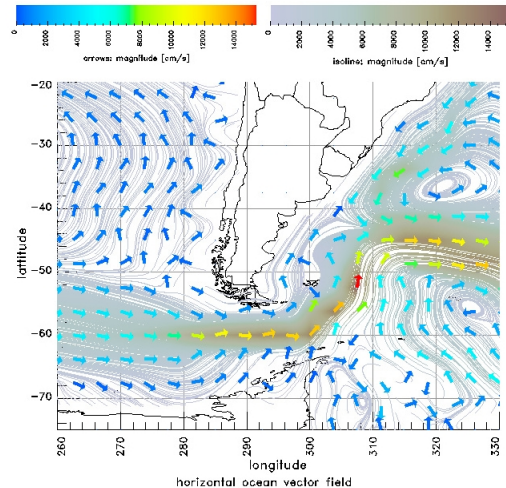


Figure 1: Streamline and arrow visualization of the horizontal ocean velocity field generated with the CLIMBER-3 model.

Figure 1 shows an example for a 2D flow visualization, tailored to display ocean vector fields. Here, the vertically aggregated horizontal ocean vector field generated with the CLIMBER-3 climate model (Montoya et al. 2005) has been mapped to streamlines and arrow glyphs of constant length, both color-coded by the vector field magnitude. This procedure avoids problems of glyph occlusion and hardly visible glyphs. The image illustrates a certain region of interest, interactively selected by the user, and represents extreme ocean flow conditions around the Cape Horn region. On the other hand, a variety of parameters have been supplied to enable a broad range of exploration tasks. For instance, a set of transparency maps has been provided, allowing the highlighting of extremes on demand.

A problem using standard techniques displaying gridded spatial data occurs at high resolutions. Then, the visual representations of single values or of value ranges of interest become very small (down to a single pixel representation on the display and smaller). As a solution to this problem, a mechanism to interactively highlight value ranges of interest has been integrated, using additional glyphs representing these value ranges of interest. Figure 2 illustrates this for the example of the LPJ vegetation model output (Sitch et al. 2003). Here, a regular grid with a resolu-

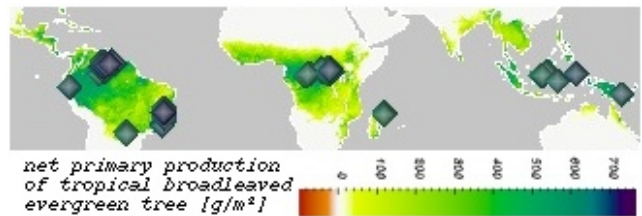


Figure 2: Colored Representation of net primary production of tropical broad-leaved evergreen trees, simulated with the vegetation model LPJ, highlighting extreme values between 570-770 g/m^2 using glyphs.

tion of 720x360 grid cells makes it nearly impossible to identify single data values. Thus, extreme values of *net primary production* are accentuated using small colored glyphs. Overlapping may not be avoidable, and therefore, as a first simple strategy, glyphs are ordered by their values back to front, displaying the most extreme values without occlusion.

4 Comparative visualization of climate related model output

Comparison of model output plays an important role in evaluating models. In particular, this includes inter-comparison of model simulations and to compare them with measured data. However, problems arise when comparing multi-variate data, e.g. when analyzing how well a simulation reproduces real world system states. Here, three problem classes of rising complexity can be identified:

1. comparison of simulation data from a single simulation run (different time slices, space regions, or variables),
2. inter-comparison of two simulation runs and comparison of simulated with measured data, and
3. comparison of a large set of simulation runs in an ensemble/multi-run experiment scenario.

A major challenge in all these classes is to compare the spatial patterns of the (partly 3D, time-dependent) data multi-variately.

Typically, either many juxtaposed images (image level) are used, or the data complexity is strongly reduced and data are compressed. This is done by the computation of aggregated variables, e.g. by the reduction of the spatial dimensionality and by the comparison of the temporal trends in the resulting time series only, by using clustering mapping multi-variate data into a single variable, or by reducing the dimensionality of a whole multi-run simulation ensemble using sensitivity or uncertainty measures. On one hand, global patterns can be easily perceived using image-level visualization, and at the same time, occlusions are avoidable, but small-size or even single grid point-related regional or local patterns are difficult to compare. However, a large number of variables reduces the efficiency of such an image level comparison.

On the other hand, abstraction strategies enable dynamic generation of simplified views onto the simulation data. However, to apply such abstraction methods and to get knowledge about the underlying data, strategies are required to show data slices or sub-sets on demand, preserving the spatial and temporal reference of the data. For instance, when using clustering, information about original data variables may be lost. Thus, to improve the user's confidence in such methods, cluster properties (e.g. cluster centroids)

and the original data need to be reloaded and displayed on demand (see (Nocke et al. 2004), (Nocke et al. 2005)).

Although data-level comparison methods that visualize several data sets in composed images (typically in their spatial reference), were introduced already in 1995 (Pagendarm and Post 1995), they are still not widespread (an exception is the explicit mapping of data differences only). However, there is potential to use such methods for evaluating simulation data, e.g. coding spatial/temporal uncertainties intuitively.

To classify visual comparison methods, we assume the following criteria:

- level of comparison: image level, data level,
- kind of data space to be compared: temporal phenomena, spatial phenomena, hybrid, multi-dimensional parameter spaces,
- number of dependent variables to be compared: single variable, bi-variate, multi-variate,
- representation space: time-series plots, 2D maps, 3D volume visualization, multi-variate visualization, hybrid,
- difference coding: explicitly (calc. of differences) vs. implicitly (displ. the original data values),
- data abstraction to be compared: original values vs. derived values (e.g. cluster properties),
- kinds of involved grids: similar vs. varying grids.

In the following sections, visualization examples for the problem classes mentioned above will be shown, illustrating different comparison strategies.

4.1 Comparison of simulation data from a single simulation run

Typically, animation techniques are used to visualize spatio-temporal data. However, this hampers the comparison of spatial structures and the tracking of interesting features. Instead, 3D visualization becomes a suitable approach, mapping time to one of the axes (see Fig. 3).

In doing so, features of interest can be tracked and easily compared in their temporal evolution. Fig. 3 (top-left) uses an interactively selected set of parallel slices to explore the dynamics of spatial pressure patterns, where extreme values are highlighted (brown for regions of high pressure and green for regions of low pressure). In Fig. 3 (bottom), the initial state of the same model variable and the extremes of changes related to this initial state are displayed, focusing on both positive changes (yellow to dark red), and negative ones (light to dark blue). To get an impression of the patterns of original values in the initial state, the related slice is displayed using a separate (gray-scale) color map. To track features in general, semi-transparent isosurfaces can be used, showing when data exceed (appearance of

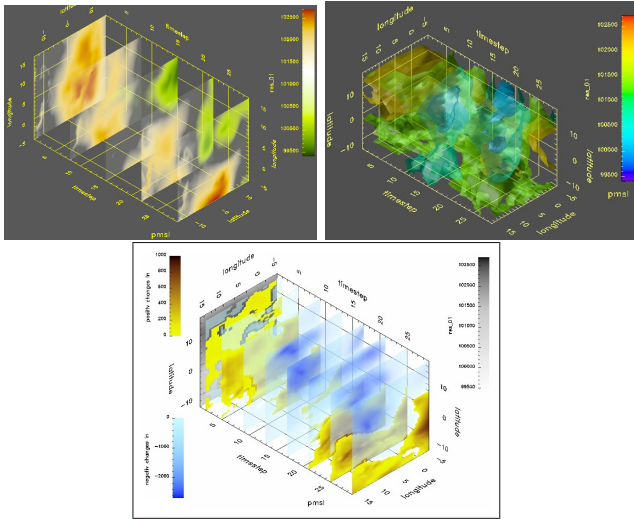


Figure 3: 3D comparative representation of spatio-temporal sea level pressure conditions simulated with the climate model CLM (Böhm et al. 2006); top-left: parallel slices displaying extremes of the original values; top-right: semi-transparent isosurfaces; bottom: differences of the extremes related to the initial time step.

colored image areas), or remain below (disappearance of colored image areas) a certain threshold, and the extent of exceedance for such pressure thresholds of interest (Fig. 3 top-right).

As a further family of methods to compare simulated data of an individual run, multi-variate metaphor-based iconic visualizations can be applied (Fig. 4, (Nocke et al. 2005)). Here, simulation data from a regular grid, reflecting the extremely low precipitation conditions, have been mapped to the locations of observation stations, to compare the multi-variate maize harvest conditions in the different regions. For this purpose, six parameters indicating the potential yield loss of maize have been derived from the simulated precipitation (Böhm et al. 2003). Furthermore, to identify regional agricultural differences, these parameters have been derived and clustered. They have been subjected to a multi-variate cluster analysis, to identify endangered regions based on multiple risk indicators. To get insight into the clustering results, the cluster centroids representing the mean properties of a certain cluster have been visualized (see Fig. 4). Therefore, maize icons were separated into six parts (one for each yield loss indicators), representing either bad conditions (brown, narrow cobs), middle (light yellow, middle-thick cobs) or good conditions (yellow, broad cobs) for the associated indicators.

In particular, such a multi-variate visualization provides an overview about dependencies between the individual indicators and enables users to compare their spatial patterns together in a single image. Furthermore, regions of identical cluster membership are merged by the developed visualization technique, resulting in larger icons. Thus, homogeneous

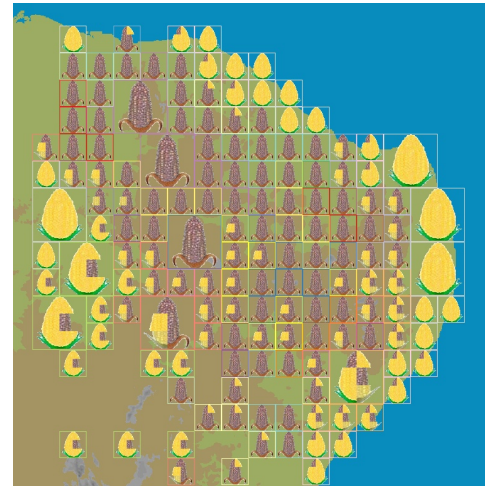


Figure 4: Spatial representation of cluster analysis results for a set of six parameters describing the potential drought risk for maize in the North-East of Brazil, derived from a REMO model simulation.

regions can be easily distinguished from heterogeneous ones, based on a multi-variate analysis of the underlying simulation data properties. However, this kind of representation for multi-variate clustered data is still restricted to one simulation or measured data set at a time (image level). Further research is needed to examine, whether icons can be used to compare two data sets in one image, too.

4.2 Inter-comparison of two simulation runs and comparison of simulated with measured data

Comparison of spatially represented data sets in one image is still a challenging task (e. g. considering different grids). A typical procedure is to map the data to the same grid and to visualize differences on this grid. However, this may lead to inaccuracies (smoothing of extreme values) and a loss of the original data values in the visualization. A further challenge is to visualize uncertainties of a model simulation in an easy-to-understand manner.

To illustrate these problems and propose first solutions, Fig. 5 compares clustered risk indicators as derived from a model simulation with those computed from measured station data. For this kind of visualization, the aggregated risk indicator rank was derived, identifying the overall risk of maize yield loss based on the six indicators as used in section 4.1 (see also Fig. 4). Then, in a first version, both the aggregated measurement- and simulation-related data value of a station are color-mapped to the upper and the lower part of a square and laid out to the station positions (see Fig. 5 left). Furthermore, the values of the cluster centroids a data point belongs to can be displayed for both data sets instead of the original values and augmented by two legends to evaluate if the (spatial) cluster structures of simulated data resembles that of the measured data. This representation allows the

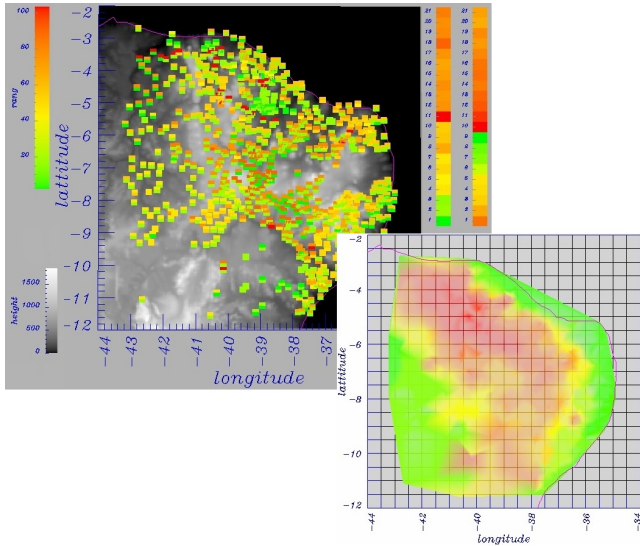


Figure 5: Comparison of (clustered) potential yield loss risk for maize, derives from measured data and simulated model results (REMO); left: using square glyphs; right: using color-coding (simulation data) and transparency (difference between simulation and observation) on a triangulated grid.

comparison of the simulation results at each station location with the value measured there, without losing information about the original values.

Drawbacks are occlusion problems in regions with a high station density, and difficulties in separating larger regions of good model predictions from bad ones. Furthermore, high values (bad conditions) catch the eye due to the color map used, focusing on extreme values. Thus, to get an overview of the simulation quality, a second technique using transparencies to map differences to triangulated station values has been compiled into the library (see Fig. 5 right). To preserve the absolute values, either simulation or measured data are mapped to color. The differences between both data sets are mapped using transparencies (low difference is represented by high opacity), more or less occluding the underlying grid. This gives an intuitive impression on the quality of the model results to represent regional risk indicator patterns. The representation clearly shows that certain regions with good (green) and medium (yellow) growing conditions for maize in coastal regions and in the interior are well reproduced by the model. It also depicts that the simulation over-estimated the potential yield loss for maize in the regions between coast and inland linked to an underestimation of the simulated rainfall there.

The problem gets even more complicated if the data sets to be compared are 3D. For isolated problems, special solutions can be elaborated. As a special case, we assume that differences between model runs are quite small and the modeler is interested in analyzing these small differences. Then, if there are only small spatial variations in the data, special visualization techniques can be built. Fig. 6 shows

two examples of an isosurface-based comparison of small differences between two model runs. Because such isosurfaces tend to penetrate each other, a first solution is to let the user translate the grids (and thus the isosurfaces) relative to each other (Fig. 6 left). This allows the comparison of overall shapes, however, there are still problems comparing single values or small regions. Thus, for flat isosurfaces (suitable e. g. for atmospheric variables), a method has been developed that analyzes, whether a certain point of one of the isosurfaces lies above or below the other isosurface or if the isosurfaces penetrate each other at this point. This information is then shown using red (above), blue (below) and white (cut) value (Fig. 6 left).

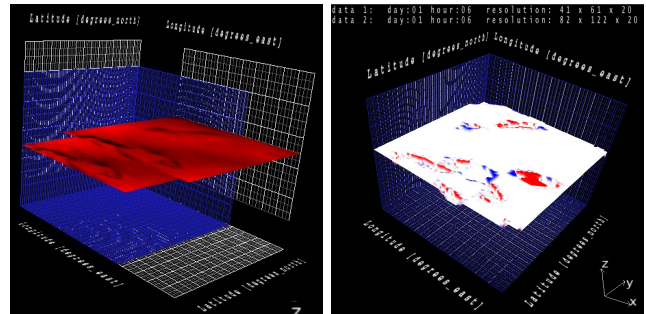


Figure 6: Comparison of two isosurfaces (temperatures, CLM model); left: displacement of the isosurfaces to avoid occlusions; right: color coding of one isosurface with the point-to-point difference between the two isosurfaces.

4.3 Comparative visualization of multi-run experiment outputs

To improve the understanding of model output from simulation ensembles, visual analysis in combination with statistical methods is commonly used (Cooke and van Noortwijk 2000). However, a large number of tested model parameters and a large number of simulation runs to be compared require alternative visualization approaches. Showing a certain plot may often lead to a loss of overview. Thus, our approach is to consequently follow Shneiderman's Mantra "Overview first, zoom and filter, then details on demand" (Shneiderman 1996), offering users a variety of views on their multi-run experiment output data. In particular, this includes a high degree of interactivity such as ordering and highlighting data.

To get an overview, the library provides a variety of techniques to visualize statistical aggregations of experiment run ensembles, for instance information on the sensitivity of the model results based on variations of the different parameters. To identify and compare the sensitivities, standard visualization techniques such as scatter plots can be used. However, using several such plots may lead to a loss of orientation between them. Thus, a graphical table has been implemented, mapping variable values to colored bars (one bar for each table cell), mapping low values to large

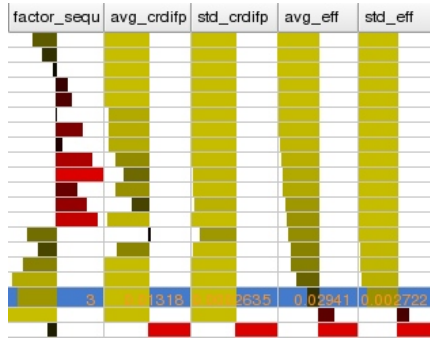


Figure 7: Visualization of sensitivity measures of a global sensitivity analysis (SWIM model (Hattermann et al. 2005)); sorted graphical table including simulation parameter numbers (factor_sequ), average and standard deviation for eff and crdifp.

bars from center to left (in yellow color), medium values to small bars (black) and high values to large bars from center to right. As an example, Figure 7 right displays the average and standard deviation of two model outputs (eff, crdifp) of a global sensitivity analysis (Morris 1991) of the hydrological model SWIM (20 model parameters considered, 8400 model runs). This graphical table allows a compact display of many variables at a glance. Interactive reordering helps to find hidden patterns within the data.

Based on such an overview, the user can select sensitive parameters and compare the related model runs in detail. Especially suitable for analyzing high-dimensional data (for both independent and dependent variables) with up to 15 variables are parallel coordinates. To efficiently use the screen space, this technique lays out the coordinate axes parallel and draws, according to the two variable values, a line between neighboring axes. Arranging several axes in this way and drawing lines between them leads to a bunch of polylines. This is illustrated in Fig. 8. Here, the most sensitive parameters for the SWIM model output variable crdifp have been selected and ordered by their significance (left to right), and are displayed with the two outputs crdifp (res_25) and eff (res_26). Then, to reveal patterns, the user can interactively change line transparencies and select certain variable value ranges. In this case, model runs of extreme low values of crdifp have been selected and the related lines have been color-coded, resulting in a certain polyline pattern (see Fig. 8).

Finally, if the user wants to get further details on patterns as well as on the single values, the data can be displayed in a graphical table. To illustrate this, a deterministic screening experiment in the space of the three most sensitive parameters was performed, and these parameters are displayed together with the run number and the two model output variables crdifp and eff (see Fig. 9). To display a maximal number of data records, the cells have been shrunk to the size of one pixel. Thus, the modeler gets an overview about the whole

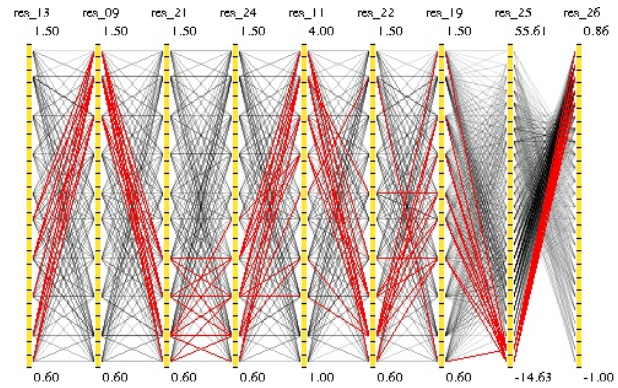


Figure 8: Parallel coordinate representation of the SWIM experiment scenario; the seven most sensitive parameters for the output crdifp are ordered by their significance (from left to right); two right columns: model outputs crdifp (res_25) and eff (res_26); selection of extreme low values of crdifp.

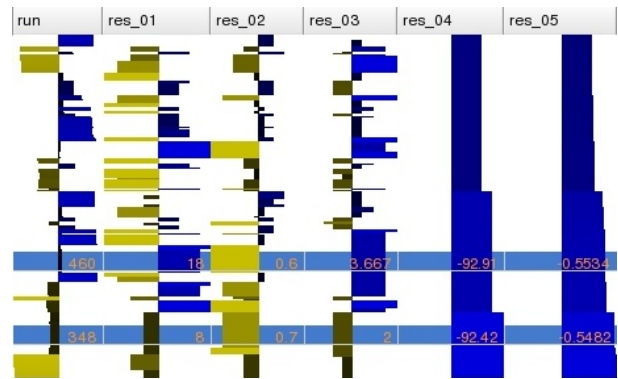


Figure 9: Graphical table representation of a deterministic screening experiment with SWIM (840 runs); sorted by model output crdifp (res_04); focused on runs with high values of crdifp.

simulation ensemble, can interactively re-sort columns and thus reveal dependency structures (e. g. identify multi-variate input conditions that lead to certain outputs data value ranges) and acquire values of interest by resizing the related rows.

5 The SimEnvVis Framework

Researchers are familiar with certain visualization environments and techniques, and typically focus on a limited set of tasks (e. g. find trends, present/communicate the results). To reduce the obstacles to use a new visualization framework, there must be a strong advantage doing so – in contrast to existing visualization tools such as Vis5D – especially providing sophisticated visualization techniques and easy-to-use interfaces to them. Therefore, the SimEnvVis framework¹ has been developed, providing a variety of the presented techniques to climate modelers. SimEnvVis is designed as a visualization server, handling server and

¹as a frontend of the Multi-Run Simulation Environment SimEnv: <http://www.pik-potsdam.de/software/simenv/>; input via NetCDF interface

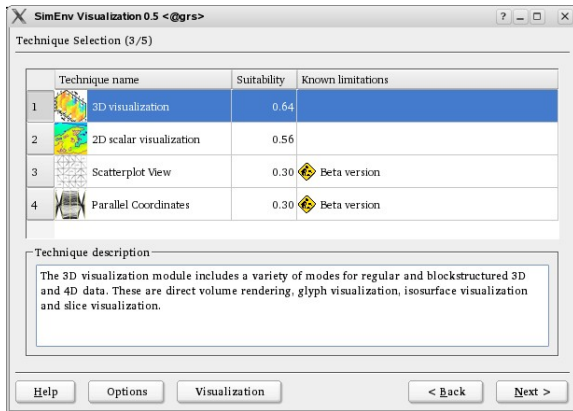


Figure 10: Selection of visualization techniques ordered by their suitability in the SimEnvVis framework.

client resource constraints on different hard and software platforms, ensuring a high portability.

Besides these technical issues, the manifold of integrated visualization techniques – partly not known to climate modelers – requires a substantial support selecting and parameterising these techniques. Our approach to provide a visualization design wizard, which hides internal system details and enables modelers to concentrate on the analysis process itself, providing a guideline generating easy-to-produce initial images. Figure 10 depicts part of the wizard, showing available visualization techniques ordered by an automatically calculated suitability.

Therefore, a two-step rule-based mechanism has been developed, which generates images based on metadata, user goals and profiles (see (Nocke and Schumann 2002), (Nocke and Schumann 2004), (Lange et al. 2006)). In a first step, the general suitability of a visualization technique is estimated based on important metadata only (e. g. no. of variables and data records). In a second step, these techniques are parameterized based on rules. This includes the selection of a visual mapping strategy (e. g. isolines or/and coloring), the selection of a mapping of certain variables to certain visual attributes (e. g. temperature to colors) and the fine-tuning of the parameters (e. g. selection of a color map). The following example illustrates two rules, adapting the suitability of a certain mapping strategy:

```
getMetaData(Attributes, "Number") > 3 ?[+=0.7];
hasGoal("overview") & hasGoal("trends")?[+=0.1];
```

To further increase the system acceptance to new users, parameter dialogs can be simplified on demand, focusing on main parameters only, using the others in default mode. Thus, learning phase can be simplified, still allowing full visualization control in later stages.

6 Discussion and conclusion

In this paper, a library of visualization techniques for modeling and simulation in climate background has been proposed.

Therefore, standard visualization as well as in this field non-typical representations have been provided and adapted to the specific application requirements. This library enables climate modelers to get new insides into the simulation data. The paper summarizes the state-of-art in climate visualization and contributes a systematic analysis of chances of comparative visualization and the application of alternative techniques for different types of comparison tasks. Finally, to support modelers generating images suitable for their current data and goal, the paper introduces the easy-to-use visualization framework SimEnvVis. It has been successful in the analysis of a variety of simulation data sets, and we have had positive feedback from climate researchers about the potential of the proposed techniques.

However, there are still challenges for future work. First of all, this includes improving the framework and the proposed techniques and further test them with the modeler's requirements. A further challenge is to not only use visualization as a front-end tool to simulation, but to apply it to all phases of the modeling process (e. g. illustrating the structure of the model itself, see e. g. (Schulz et al. 2006)).

ACKNOWLEDGMENTS

The authors want to thank Heidrun Schumann from the University of Rostock for a high level of support. Furthermore, we thank Ruppert Klein for supporting this work and Stefan Petri, Claus Rachimow and Ciaron Linstead for helpful discussions and proofreading. Finally, we thank the students involved in the implementation of the proposed techniques, including Mario Baalke, Ronny Kaeding, Karsten Wagenknecht, Matthias Schröder and Conrad Thiede.

REFERENCES

- American Meteorological Society 1993. Guidelines for using Color to Depict Meteorological Information: IIPS Subcommittee for Color Guidelines. *Bull. Amer. Meteor. Soc.* 74:1709–1713.
- Böhm, U. et al. 2006. CLM - the Climate Version of LM: Brief Description and long-term Applications. *COSMO Newsletter* (6): 225–235.
- Böhm, U., F.-W. Gerstengarbe, D. Hauffe, M. Kücken, H. Österle, and P. Werner. 2003. Dynamic regional climate modeling and sensitivity experiments for the northeast of Brazil. In *Global change and regional impacts*, ed. Gaiser, Krol, Frischkorn, and Araujo, 153–170. Springer, Berlin.
- Cooke, R., and J. van Noortwijk. 2000. Graphical Methods. In *Sensitivity Analysis*, ed. Saltelli, Chan, and Scott. Wiley.
- Djurcilov, S., K. Kim, P. Lermusiaux, and A. Pang. 2001. Volume Rendering Data with Uncertainty Information. In *Data Visualisation '01: Proceedings of the EG+IEEE*

VisSym in Ascona.

- Doleisch, H., P. Muigg, and H. Hauser. 2004. Interactive Visual Analysis of Hurricane Isabel with SimVis. IEEE Visualization (Vis'04) Contest, <http://vis.computer.org/vis2004contest/vrvis/>.
- Gelin, L. 2002. Evaluation of Added Value in Multidimensional Visualization for SMHI. In *European Working Group on Operational Workstations (EGOWS), Rome*.
- Griebel, M., T. Preusser, M. Rumpf, M. Schweitzer, and A. Telea. 2004. Flow Field Clustering via Algebraic Multigrid. In *Proceedings IEEE Visualization'04*.
- Griethe, H., and H. Schumann. 2005. Visualizing Uncertainty for Improved Decision Making. In *4th International Conference on Perspectives in Business Informatics Research (BIR'05)*. Sköde, Sweden.
- Hattermann, F., V. Krysanova, F. Wechsung, and M. Wattenbach. 2005. Runoff simulations on the macroscale with the ecohydrological model SWIM in the Elbe catchment - validation and uncertainty analysis. *Hydrological Processes* 19:693–714.
- Lange, S., T. Nocke, and H. Schumann. 2006. Visualisierungsdesign - ein systematischer Überblick. In *Simulation and Visualization'06 (in German)*, 113–128.
- Macédo, M., D. Cook, and T. Brown. 2000. Visual Data Mining In Atmospheric Science Data. *Data Mining and Knowledge Discovery* 4 (1): 69–80.
- Montoya, M., A. Griesel, A. Levermann, J. Mignot, M. Hofmann, A. Ganopolski, and S. Rahmstorf. 2005. The Earth System Model of Intermediate Complexity CLIMBER-3 α , Part I: description and performance for present day conditions. *Climate Dynamics* 25:237–263.
- Morris, M. 1991. Factorial plans for preliminary computational experiments. *Technometrics* 33 (2): 161–174.
- Mueller, W., and H. Schumann. 2003. Visualization Methods for Time-dependent Data - an Overview. In *Proc. of the 2003 Winter Simulation Conference, New Orleans*.
- Nocke, T., S. Schlechtweg, and H. Schumann. 2005. Icon-based Visualization using Mosaic Metaphors. In *Information Visualization (IV'05), London*, 103–109.
- Nocke, T., and H. Schumann. 2002, August. Meta data for visual data mining. In *Proc. Computer Graphics and Imaging (CGIM'02), Kaua'i, Hawaii, USA*.
- Nocke, T., and H. Schumann. 2004. Goals of Analysis for Visualiz. and Visual Data Mining Tasks. In *CODATA Worksh. Information, Presentation and Design, Praha*.
- Nocke, T., H. Schumann, and U. Böhm. 2004. Methods for the Visualization of Clustered Climate Data. *Computational Statistics* 19 (1): 75–94.
- Pagendarm, H.-G., and F. Post. 1995. Comparative Visualization - Approaches and Examples. In *Visualization in Scientific Computing*, ed. M. Göbel, H. Müller, and B. Urban. Springer, Wien.
- Pang, A., C. Wittenbrink, and S. Lodha. 1997. Approaches to Uncertainty Visualization. *The Visual Computer* 13 (8): 370–390.
- Riley, K., D. Ebert, C. Hansen, and J. Levit. 2003. Visually Accurate Multi-Field Weather Visualization. In *Proceedings IEEE Visualization 2003*, 279–286.
- Saito, T., H. Miyamura, M. Yamamoto, H. Saito, and Y. Hoshiya. 2005. Two-Tone Pseudo Coloring - Compact Visualization for One-Dimensional Data. In *Proceedings IEEE Symposium on Information Visualization (InfoVis'05)*, 173–180. Mineapolis, USA.
- Schröder, F. 1997. *Visualisierung meteorologischer Daten*. Springer (in German).
- Schulz, H.-J., H. Schumann, and T. Nocke. 2006. Visualizing and analyzing large systems of differential equations. In *Poster Presentation at the Winter Simulation Conference (WSC'06) in Monterey, USA*.
- Shneiderman, B. 1996. The Eyes Have It: A Task by Data Type Taxonomy for Information Visualization. In *Proceedings IEEE Symposium on Visual Languages '96, IEEE, Los Alamos*.
- Sitch, S. et al. 2003. Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the LPJ Dynamic Vegetation Model. *Global Change Biology* 9:161–185.
- Stier, P. et al. 2005. The aerosol-climate model ECHAM5-HAM. *Atmospheric Chemistry and Physics, European Geosciences Union* 5:1125–1156.
- Treinisch, L. 1999. Task-Specific Visualization Design. *IEEE Computer Graphics and Applications* 19 (5): 72–77.
- Treinisch, L. A. 1994. Severe Rainfall Events in Northwestern Peru: Visualization of Scattered Meteorological Data. In *Proceedings of Visualization'94*, ed. L. O'Connor and P. Storms.
- Wong, P. C., and R. D. Bergeron. 1997. 30 Years of Multidimensional Multivariate Visualisation. In *Scientific Visualisation*, ed. G. Nielson, H. Hagen, and H. Müller, 3–33. IEEE Computer Society, Los Alamitos.
- Zhou, H., M. Chen, and M. Webster. 2002. Comparative evaluation of visualization and experimental results using image comparison metrics. In *Proceedings IEEE Visualization'02*, 315–322.

AUTHOR BIOGRAPHY

THOMAS NOCKE is visualization researcher at the PIK. He was previously assistant at the University of Rostock. His research interests include Visualization and Visual Analytics of Climate Impact Data.

MICHAEL FLECHSIG is senior researcher at the PIK. His research interests include multi-run model simulation and experimentation environments, sensitivity and behavioral analyses and Monte-Carlo simulations.

UWE BÖHM is senior researcher at the PIK. His research focuses on dynamic regional climate modeling and combining uni- and multi-variate statistics for model evaluation.