Visualization in Earth System Science

Michael Böttinger, Deutsches Klimarechenzentrum (DKRZ), Martin Schultz, Max-Planck-Institute for Meteorology and Joachim Biercamp, Deutsches Klimarechenzentrum (DKRZ)

From Climate to Earth System Models

Traditionally, climate is defined as the statistical collective of the weather conditions of a specified area during a specified interval of time, usually several decades. This definition is currently undergoing a change to place more emphasis on the exchange of energy, momentum, and mass between the different compartments of the Earth System. Although weather is experienced as a pure atmospheric phenomenon with high temporal variability, the long-term changes of mean weather conditions are driven by the dynamics of slowly changing components of the climate system: e.g. the ocean, sea and land ice, and the biosphere.

Several feedbacks between the climate system compartments have been identified. Two prominent examples show links between atmospheric chemistry and the physical atmosphere: the photochemical formation of ozone in the troposphere, which then acts as a greenhouse gas (Figure 1), and the influence of aerosols and their chemical composition on the formation and properties of clouds. In contrast to short-term weather forecasting, where slowly varying components can be prescribed as boundary conditions, coupled global 3-dimensional models of the full physical ocean-atmosphere-sea-ice system are needed for simulations on longer time scales. In the future, we will see even further integration of chemical, biological, and socio-economic models into traditional climate models, ultimately leading to a comprehensive global modelling system termed the Earth System Model (ESM) (Figure 2).

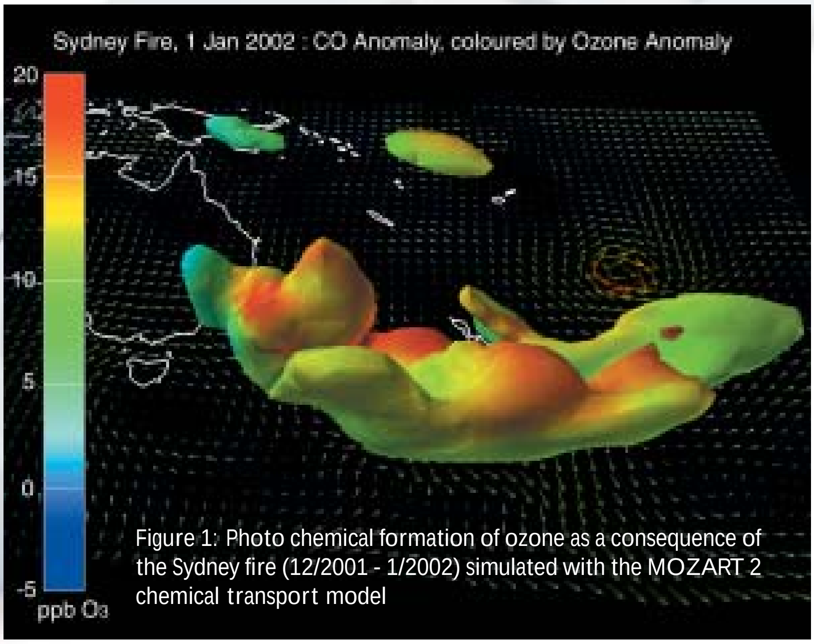

Figure 1: Photo chemical formation of ozone as a consequence of the Sydney fire (12/2001 - 1/2002) simulated with the MOZART 2 chemical transport model

The challenge

The output from such models can be described as a multitude of time dependent 2D and 3D data sets, each consisting of several scalar and vector variables.

The individual data sets may have different time intervals, and they do not necessarily share a common computational grid. For example, the ocean component may run on an Arakawa-C grid with shifted poles (Figure 3) [1], while the atmosphere is simulated on an almost regular Gaussian grid. New grid structures such as the triangular grid from the GME model of the German Weather Service (DWD) [2] are being developed to remedy some of the problems encountered with regular grids, e.g. the singularity at the poles. Many models use vertically non-linear coordinate systems. Data from atmospheric models is often stored on pressure levels or so-called hybrid levels (topography following levels at the bottom of the model, pressure level at top and a mixture of both in between). Some models like isopycnal ocean models even use time dependent vertical coordinates.

The enhanced supercomputer technology allows us to refine the spatial resolution and to add more processes and variables (e.g. tropospheric and stratospheric chemistry) to the models. The higher resolution has many advantages: small scale processes, which otherwise have to be parameterized, can be simulated directly, and local topographic effects on the atmospheric or the ocean circulation are better resolved.

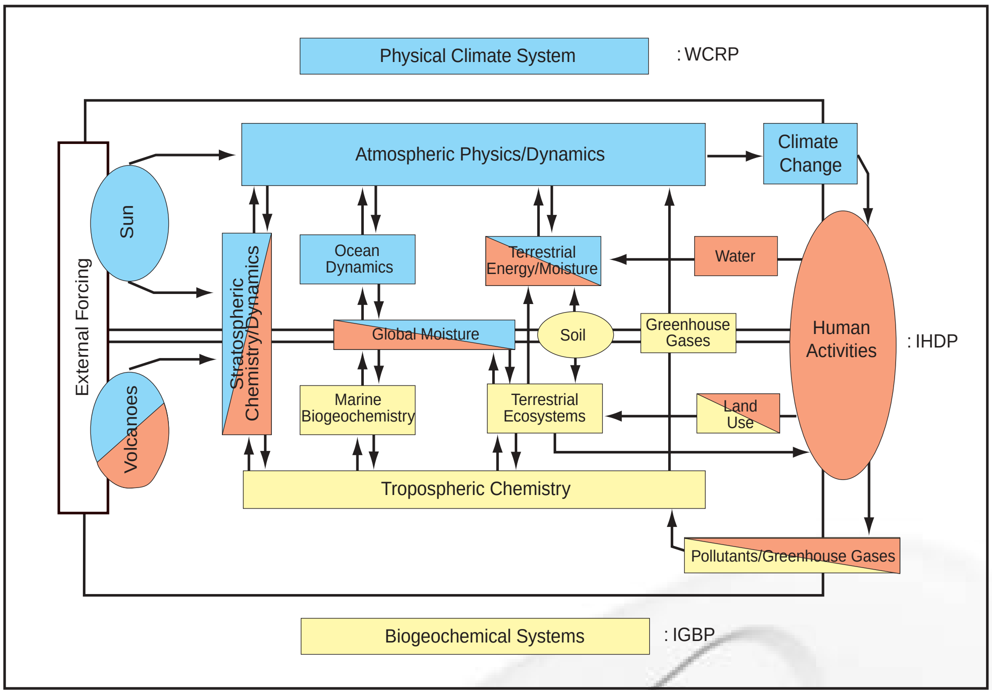

Figure 2: The Earth System: A modified "Bretherton diagram" highlighting some of the linkages between social systems, biogeochemical systems, and the physical climate system. Courtesy Guy Brasseur, Max Planck Institute for Meteorology

So-called Monte-Carlo-simulations (repeating the same simulation with slightly altered initial or boundary conditions) and multi-model simulations are used to quantify the variability of the model simulations and the probability of specific results. Such ensembles further increase the amount and the complexity of model output.

Analysing model results has become a real challenge for geoscientists: the amount of data that is produced by global and regional models has risen exponentially over the past decade (Figure 4). The data volumes generated by an ESM mandate that visualization applications read data in different formats and on different meshes without prior conversion. Data on different geometries and with different dimensions must be compared to each other and to observational data, which is often inhomogeneous in time and space (Figure 5). Geoscientists need data processing and visualization tools, which help them understand the Earth System. These tools must be fast and flexible in order to efficiently support the search for new phenomena and feedbacks, and they must also be able to produce high-quality graphics, which can directly be used in publications (Figure 6).

The reality (1) - data structures and formats

Ten to fifteen years ago, when many of today's visualization packages were being developed, most atmosphere and ocean models had horizontally regular or rectilinear grids. Most of the visualization packages available today are not prepared to deal with irregular grids, which are quickly becoming the standard in Earth System modelling. Hence, such data needs to be interpolated prior to its visualization, which can lead to unacceptable alterations of results.

Over the past decades, two quasi-standard formats have been established, which allow us to write self-describing and machine independent data sets for Earth System science: GRIB and NetCDF. A third format, HDF, is widely used to store satellite data.

Many data sets from atmospheric models have been generated in the GRIB format ([3],[4]). A GRIB record consists of a short descriptive header plus one horizontal layer of one variable. The GRIB header contains a limited metadata set, which for example allows us to read the files without knowing the grid size and structure in advance.

The more general NetCDF format [5] facilitates random access of individual data records and allows for arbitrary metadata within the file itself. NetCDF is not specifically designed for Earth System requirements, therefore additional metadata conventions are needed to guarantee common names, units etc. In 1995 the COARDS convention [6] has been designed for global atmospheric and oceanographic research data sets. Because of limitations like the restriction to rectilinear grids this standard is not sufficient for many recent models. The CF convention [7] currently being developed is an extension of COARDS which for example adds support for non-rectilinear grids. NetCDF/CF will likely become the quasi standard in Earth System modelling [8][9].

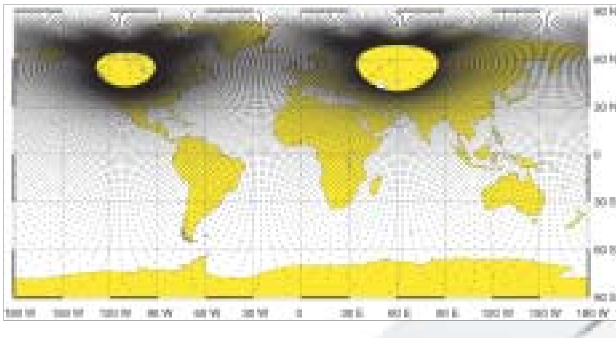

Figure 3: Example for a curvilinear 302x132 grid with the HOPE-C ocean model (GI4). The flexible position of the poles allows for a locally increased grid resolution in the area of interest: in this example the northern North Atlantic (here 30-40 km resolution). Courtesy Uwe Mikolajewicz, Max Planck-Institute for Meteorology

The reality (2) - visualization software

In terms of algorithm development, the problem "visualization of data that describe the four-dimensional space-time world" is mostly solved. Techniques for the display of time dependent scalar or vector fields have been developed and published years ago. Many of them have been included in commercial and non-commercial data visualization software [10]. But why are these techniques so rarely used by the scientists who generate the data?

Publications of climate researchers seldom include 3D representations of their data. How many 3D images of 3D data are contained in the 881 pages of the last IPCC report [11]? We found none. While we commit that 3D visualization may be more widely used for the interactive exploration of data than for the production of reproducible, quantitatively and scientifically exact graphics, it is nevertheless striking to observe the nearly complete lack of 3D images in climate-related publications. One of the factors, which may explain this phenomenon is the fact that the spatial structure of 3D still images cannot be easily printed. Well-established methods like shadow casting - which would improve the depth perception of 3D images - are not yet available within any scientific visualization package.

Most interactive 3D visualization solutions have very limited 1D and 2D capabilities. Such plots are essential for the quantitative analysis of model output. It is often desirable to interactively explore the data in 3D (e.g. find the ideal location of a slice), before producing a publication quality 2D plot. Currently, this approach usually requires the use of more than one software package with an intermediate "off-line" data extraction step.

Commercial Software:
In the late 1980s / beginning of the 1990s several commercial companies started to develop all-purpose 3D data visualization software. Interactive script and command languages like PV-Wave or IDL were extended to provide more 3D functionalities. Modular visual programming environments like AVS, Iris Explorer and IBM Data Explorer or end-user applications like Wavefront's Data Visualizer were expected to revolutionize the way scientific data is displayed [12], and it was believed that they would spread widely enough to provide profits for the companies.

The situation today: Wavefront is gone. IBM stopped further developments of Data Explorer and released it as open source. AVS (now AVS/Express) was rewritten with a more object-oriented approach - which makes it even harder to use. SGI's Iris Explorer was taken over by NAG.

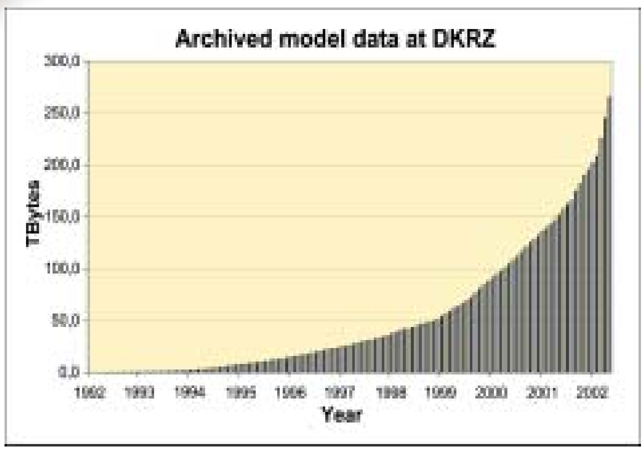

Figure 4: Archived climate model data stored at DKRZ 1992-2002 (without duplicates).

Major developments in these software packages were made years ago. In the last years the advances from release to release were relatively small. Newly developed visualization methods have rarely been implemented; they can only be added by the user with significant effort. The usability has not increased very much. Modules for specific visualization methods (e.g. volume rendering) are very often only applicable for specific data types or grids and users have to be increasingly familiar with the software in order to know in advance which combination of modules might work with their data. Documentation and online help are still an issue: important details are very often not covered. PV-Wave and IDL have grown in functionality. For example, IDL now features an object oriented graphics engine. Both programs are still commercially supported and they are widely used. However, these languages suffer from legacy codes and concepts. In order to use object graphics in IDL, the user has to learn many details about 3D rendering, which distract from the actual task of visualizing scientific data.

None of the commercial products offers a "ready-to-use" visualization application for Earth System modelling; users have to write their own software and learn about the programming language, available libraries, or visual programming environments.

Free software:
Based on the insight that commercial software only solved some of the typical problems in the visualization of geophysical data, some scientific institutes decided to develop their own visualization software tailored to their needs (e.g. GrADS, FERRET, NCAR Graphics). These tools were developed in close cooperation with the geoscientists and have therefore enjoyed large acceptance in the community. But considering the challenges of coping with today's model output, it becomes quickly clear that practically all of the freely available plotting packages are too limited in their data model or in the types of plots that are offered. For example: it is impossible to read two data sets with different vertical or horizontal grids into the GrADS software, and to compare the results in one plot. FERRET, GrADS and other software packages can read COARDS NetCDF files, but they often fail if the data was written with other conventions like CF. Except for OpenDX, which appears rather slow with large data sets and requires a large learning effort, hardly any free software package is able to deal with irregular grids. As opposed to most other programs, Vis5D [13] is a very usable and efficient 3D data visualization application for atmospheric data, but it supports only horizontally regular grids, it has only very limited 1D or 2D capabilities, and it has no built-in GRIB or NetCDF data importer. Also, control over the appearance of a plot is fairly limited.

No visualization package fully supports the data formats and conventions applied in ESMs. To use them, a considerable amount of data conversion and module writing is required. Most software is difficult to use for geoscience applications and requires a large learning effort. Expertise in both visualization and Earth System research is needed in order to make full use of the available software, or to develop custom-tailored applications, which can then be used by a small fraction of the community. Bridging the gap between Earth System science and visualization is rarely rewarded, because - from a science perspective - too much time is spent on technical details, and - from the visualization perspective - the focus is too narrow.

Conclusion - what is needed

At present, a patchwork of different tools is needed in order to produce the desired visualization results from ESM output. Data files must be replicated in order to allow for their visualization, and a lot of image manipulation is needed in order to yield publishable results.

Facing the quickly growing data volumes produced with ESMs and their increasing complexity, visualization and data processing must converge into a single system for 1D to 3D visualization. The following list contains key requirements for such a system:

- Cover the whole range from static 1D plots to interactive state-of-the-art 3D visualization methods
- High interactive 3D performance (hardware acceleration)
- Data importer/browser for all major file formats used (GRIB, NetCDF/CF, HDF, IEEE, ASCII)
- Interface to DBMSs
- Capable of dealing with huge data sets
- Automatic use of metadata, including some "understanding" of the meaning of physical quantities
- Access to processing functionality, mathematical functions, statistics
- Support for different geometrical grids, extendable to support future grid definitions
- Multiple data sets on different grids, interpolation between grids
- Easy to use, easy to learn

- Interactive use via GUI and script/batch mode processing
- Extendable to allow for special functions
- Include World Maps and 3D topography, extendable to include e.g. vegetation maps
- Automatic mapping of geo-registered data, arbitrary map projections
- Publishing quality image output with WYSIWYG preview
- Platform independent
- Freely available in order to allow use in education and in developing countries

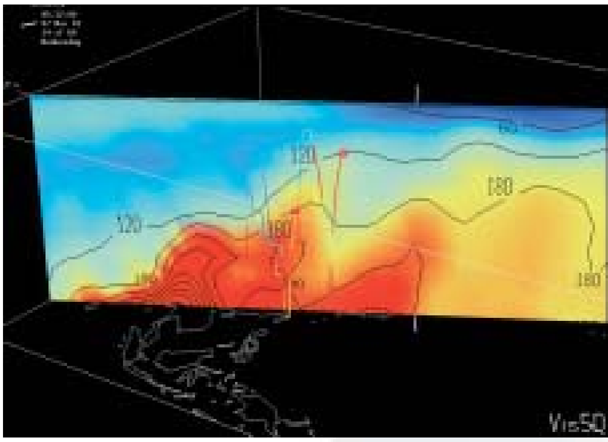

Figure 5: Comparison of airborne CO measurements and model results (TRACE-P).

Currently, climate research institutions and research projects rarely provide any significant funding for the development of suitable visualization applications. Infrastructure programmes like PRISM [8] and ESMF [9] do have data processing and visualization on their agenda, but the resources attributed to these issues are far too low. Processing and visualization of Earth System data must be recognized as a challenging engineering problem similar to the construction of a sophisticated scientific instrument. Because many visualization concepts are not known to geoscientists today, they should receive better and earlier training in data processing and visualization techniques (e.g. at university courses). A highly efficient, easy-to-use, and flexible visualization tool would likely boost the productivity of scientists working in climate research. The building blocks are out there: now the architect and the sponsor are needed to bring it all together!

References

[1] http://www.mpimet.mpg.de/Depts/Klima/natcli/C-HOPE/C-HOPE.html

[2] Majewski, D. 1998: The new global icosahedral-hexagonal gridpoint model GME of the Deutscher Wetterdienst. ECMWF Seminar on Recent developments in numerical methods for atmospheric modelling. Reading, 7 - 11 September 1998. Seminar Proceedings, 172 - 201.

[3] http://www.wmo.ch/web/www/WDM/Guides/Guide-binary-2.html

[4] http://www.wmo.ch/web/www/DPS/FM92-GRIB-Edition2.pdf

[5] http://www.unidata.ucar.edu/packages/netcdf

[6] http://ferret.wrc.noaa.gov/noaa_coop/coop_cdf_profile.html

[7] http://www.cgd.ucar.edu/cms/eaton/cf-metadata/index.html

[8] Program for Integrated Earth System Modeling: http://prism.enes.org

[9] Earth System Modeling Framework: http://www.esmf.ucar.edu/

[10] Visualization packages supporting NetCDF: http://www.unidata.ucar.edu/packages/netcdf/software.html

[11] Climate Change 2001: The Scientific Basis - Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change (IPCC), IPCC 2001, ISBN: 0521014956 http://www.grida.no/climate/ipcc_tar/wg1/index.htm

[12] See also: Hibbard, W.: Confessions of a Visualization Skeptic, Computer Graphics 34(3), 11-13. 2000 (http://www.ssec.wisc.edu/~billh/vfmay2000.html)

[13] http://www.ssec.wisc.edu/~billh/vis5d.html, http://vis5d.sourceforge.net/

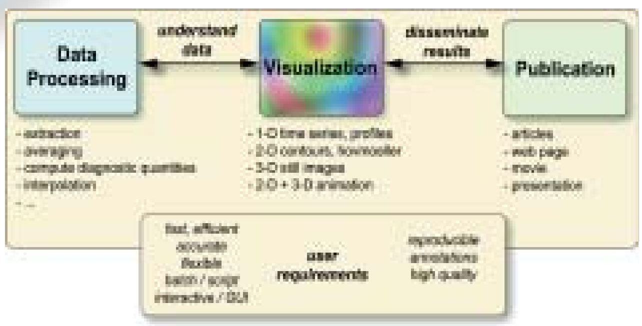

Figure 6: The role of visualization in the geosciences in between data analysis and publication.