Recent Achievements in Visual Data Exploration and Analysis for Climate Science

Overview

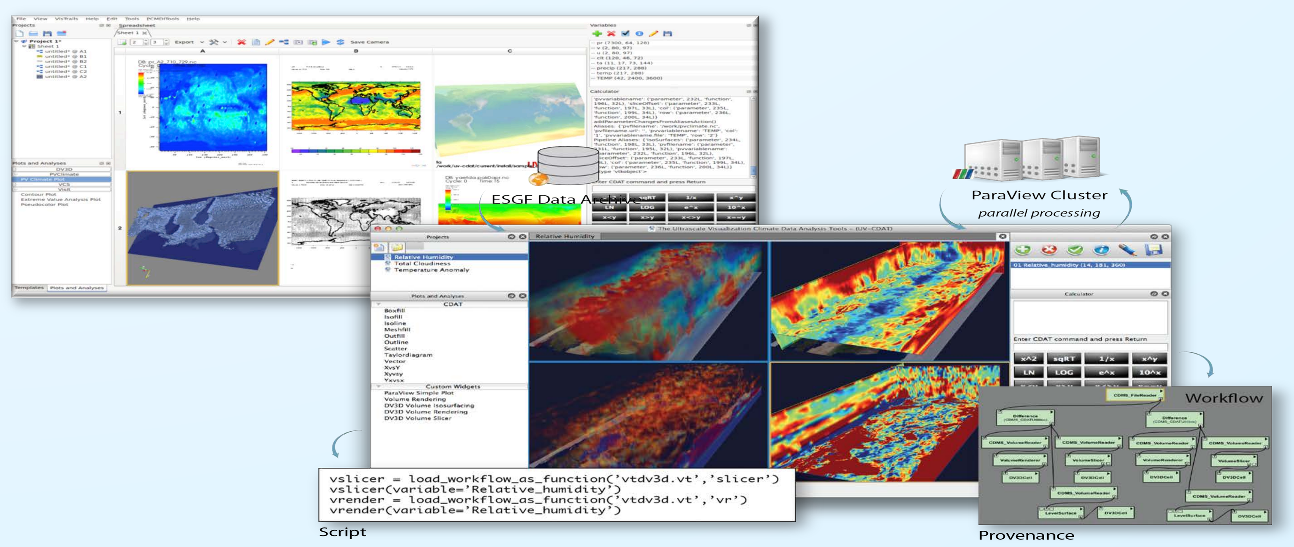
Climate science projects, like many areas of computational science, face substantial challenges that result from increasing data size and complexity. Legacy tools, which have served well for decades, are being replaced with new technologies better suited to run on advanced computational platforms and to enable knowledge discovery on increasingly more complex and challenging lines of scientific inquiry. We present recent results of several different projects, conducted in collaboration with the climate science community, that show the impact of these technological advances.

The climate science community has the opportunity to benefit from SDAV expertise and technologies in the areas of data management, analysis, visualization, and several different software products.

Our approach is to concurrently pursue applied R&D of new techniques needed by the climate science community, and to roll these into production-quality tools. We are leveraging a significant prior investment in visualization/analysis software infrastructure.

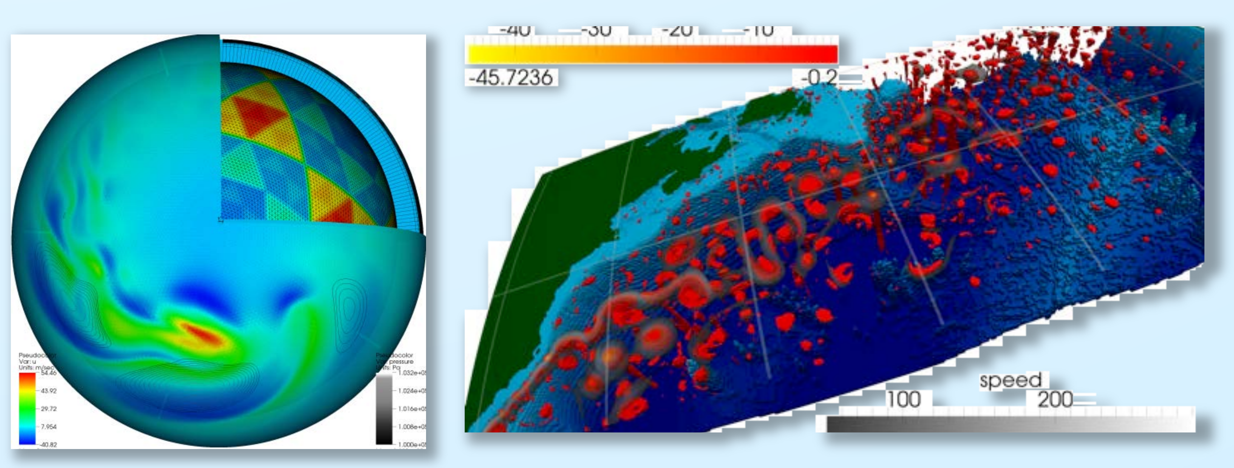
UV-CDAT: Community Software Infrastructure for Ultra-scale Climate Data Analysis

- An integrated software product for climate data management, analysis, and visualization. See http://www.uv-cdat.org/.
- Integrate DOE's climate modeling and measurements archives
- Develop infrastructure for national and international model/data comparisons
- Deploy a wide-range of climate data visualization, diagnostic, and analysis tools with familiar interfaces for very large, high resolution climate data sets (CDAT, VTK, R, VisIt, ParaView, DV3D, …)
- Workflow – data flows are directed graphs describing computational tasks
- Takes advantage of ESGF data management
- Many SDAV technologies deployed in UV-CDAT.



VisIt and ParaView

VisIt and ParaView are open source, parallel capable visualization tools that have been brought to bear on climate science visualization and analysis problems. Both tools are part of the SDAV portfolio, and both are included as part of the UV-CDAT software infrastructure.

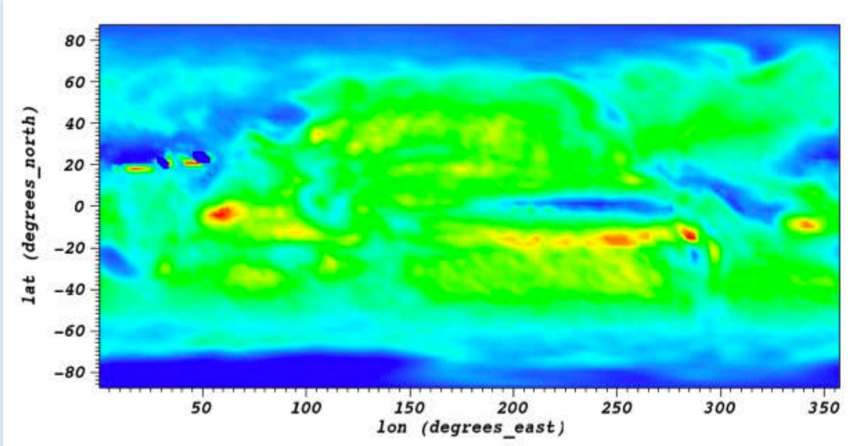


(left) VisIt visualization and rendering of geodesic grid output from the GCRM code. (right) ParaView rendering of eddies detected in POP model output.

Parallel I/O and Data Models

An important part of the SDAV portfolio is infrastructure (software libraries) for parallel I/O, along with the know-how to design and implement data models for the climate community. Our work includes application of Parallel NetCDF to the Global Cloud Resolving Model (GCRM) code, and can include ExaHDF5, a highly scalable I/O library that is ported to all modern DOE SC platforms.

Scalable Statistical Analysis for Climate Data Analysis

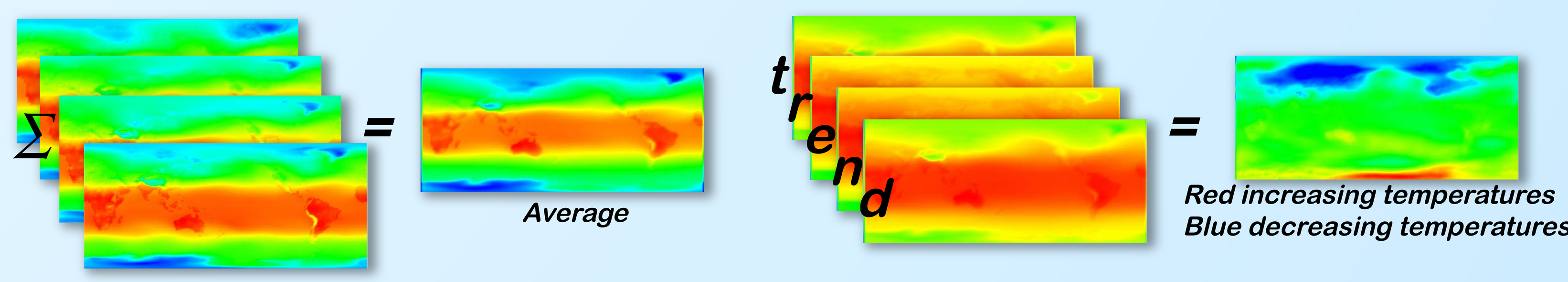


Scaling study

num CPUs	Annual Analysis	Monthly Analylsis
1	600.53	7,904.97
2	320.95	5,954.16
4	181.40	4,948.32
8	112.20	2,589.61
16	67.49	2,121.38
32	34.36	1,317.06
64	19.51	673.61
128	10.33	558.18
256	5.99	576.66
512	5.10	317.34
1024	4.59	166.41

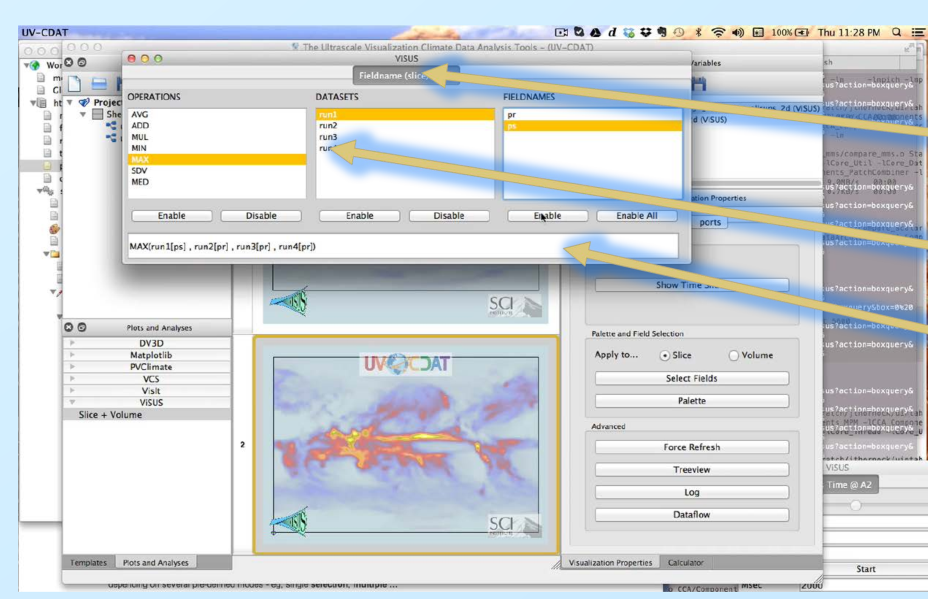
We have joined forces with both computer scientists, climate scientists, and statisticians to create frameworks for doing parallel, complex statistical analysis on the largest HPC resources. The resulting work has enabled climate scientists to perform analyses that were not possible beforehand. The statistical package R has been integrated into VisIt, and we have created two operators for performing analysis of extreme events. In particular, Generalized Extreme Value Analysis, and Peaks over Threshold. Both spatial and temporal parallelism techniques were implemented inside VisIt, and the statistical analysis was performed in R. Results of both GEV, and POT are shown below, as well as a scalability study. In order to provide climate scientists additional flexibility in performing analyses, we have also generalized the framework to allow for user-defined R and Python scripting kernels to be executed in parallel. The focus was on providing a set of template-like functions as an API that describes data access patterns. In this way, VisIt can be effectively used to do the parallel processing and management of very large data, and the scripting kernel can be executed in way that is straightforward for a domain scientist, and hides the complexities of parallel data processing. As an example, both of the GEV and POT operations were implemented using a ForEachLocationOverTime( ) template, and an R or Python kernel.

Statistical Fingerprinting of Climate Data



- Multiple global models of earth's climate
  - Models do not simulate all of the climatic phenomena
  - Contain different climatic forcing functions
- Science Goal: understand the affect of different phenomena and forcing on the earth's climate including human influence.
- SDAV Goal: new interactive analysis tools deployed in broadly used software environments – VisIt and UV-CDAT.
- Acknowledgements – B.D. Santer (LLNL) and the World Climate Research Programme's Working Group and Couple Modeling and the member climate modeling groups for producing the data and making it available.

Streaming Computation of Ensembles



The Data Processing Panel allows the user to interactively select which dataset should be combined in a particular ensemble and what aggregation function should be used. The user is provided with two options:

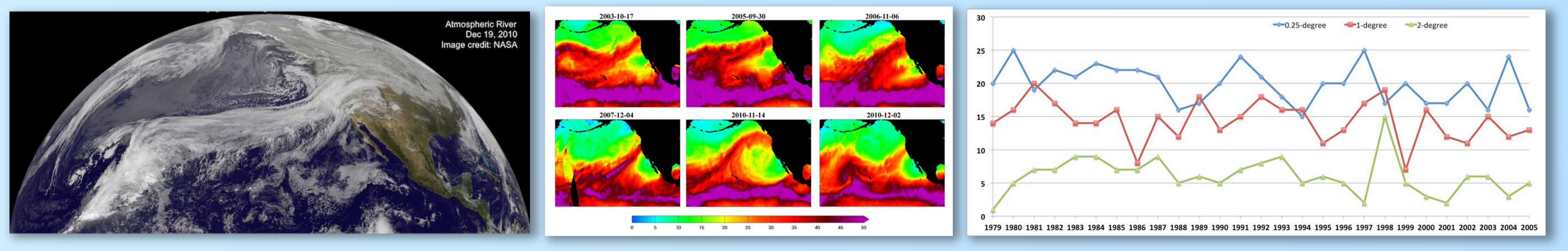
– a simple selection from a set of predetermined functions and parameters for automatic creation of an expression and;

– a free form expression that the user can type or edit from the one generated automatically

The results of the selected processing expression are presented in the ViSUS rendering windows, which is updated in real time as the user expression is updated.

The ViSUS client component has been embedded in the UV-CDAT framework to enable streaming computation of ensemble processing operators on both client- and server-side (depending on data location). The figure above shows a UV-CDAT session with the ViSUS rendering window visualizing the result of the data processing selection linked to the Data Processing Panel where the user can easily specify the processing formulate and the data sources, which can include multiple runs from multiple simulations combined as needed.

Detecting and Analyzing Atmospheric Rivers, an Extreme Weather Event

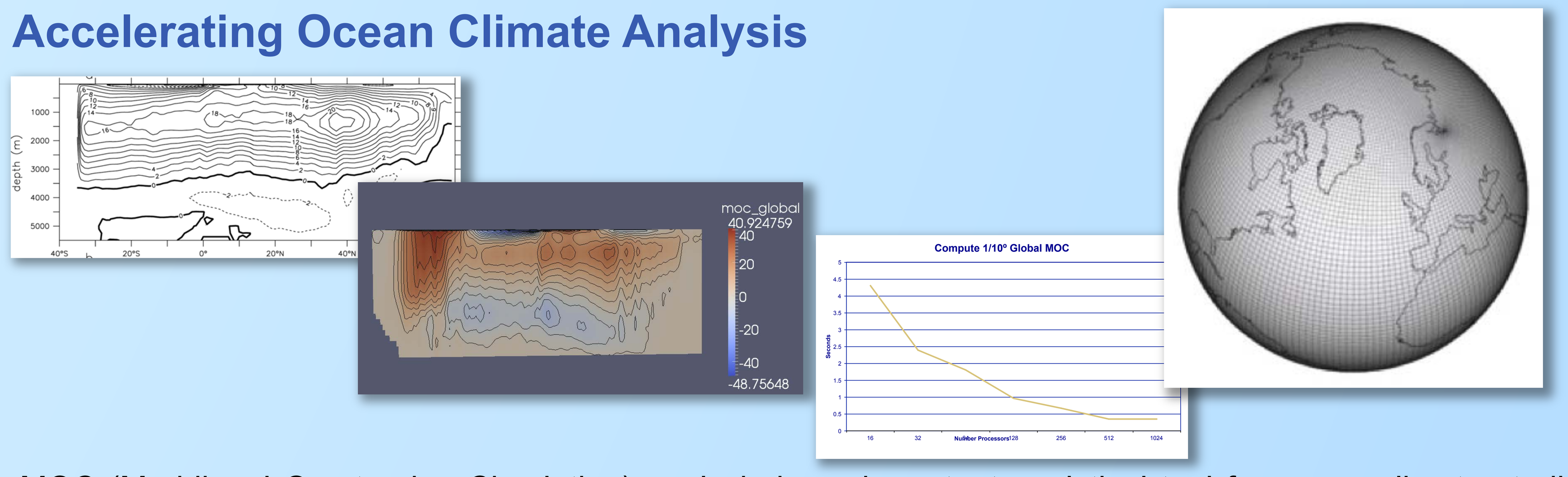


Extreme precipitation events on the western coast of North America are often traced to an unusual weather phenomenon known as atmospheric rivers (AR). These events refer to filamentary structures in the atmosphere that transport significant amounts of water over a long distance in narrow bands (left image).

We have developed an efficient atmospheric river detection algorithm based on image reconstruction techniques. Our detection algorithm is based on a thresholding condition on the total column integrated precipitable water vapor (prw) followed by a connected component labeling procedure to group the mesh points into connected regions from the tropics to the extratropics. We have applied this highly parallel technique both to satellite observations as well as large climate model output datasets (middle image). For example, our parallel AR detection code processes 716 GB of CAM5 integrated water vapor data corresponding to the western coast in ~30 minutes, where as a serial version takes ~22 days to process the same data.

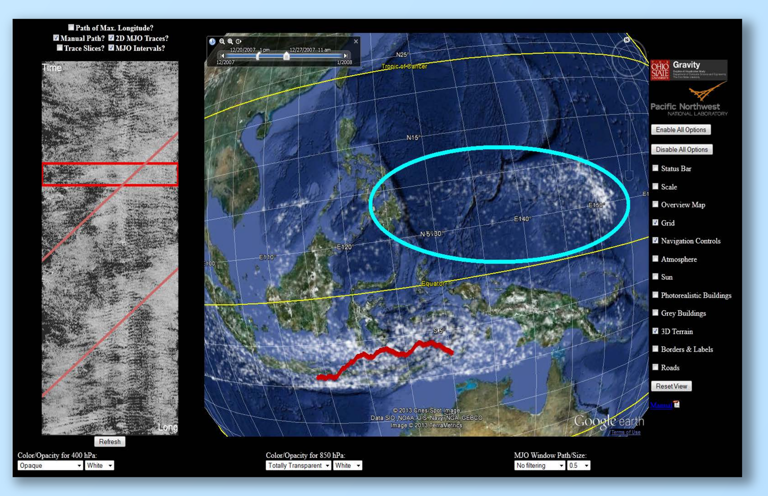
Using our detection code, domain scientists can compare the statistics and intensity of ARs in different climate models and in the same models at various resolutions (right image). They can also verify climate models with satellite observations.

Accelerating Ocean Climate Analysis



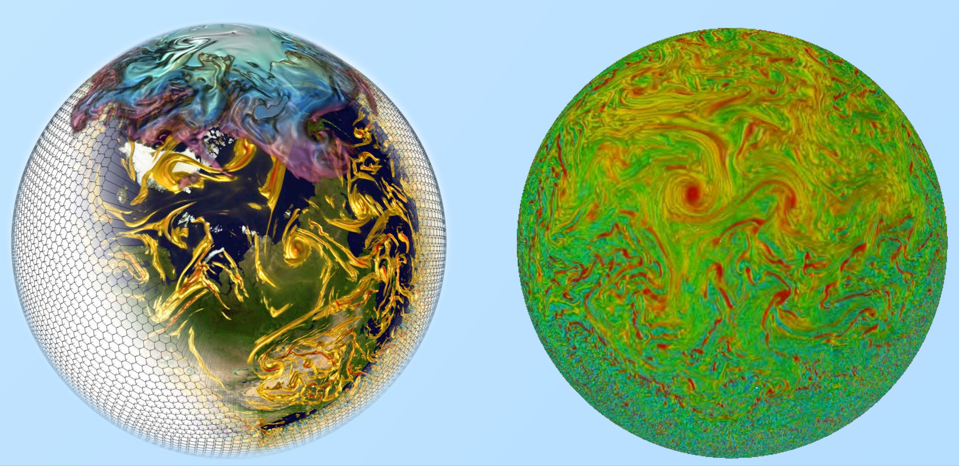
MOC (Meridional Overturning Circulation) analysis is an important analytical tool for ocean climate studies. To support POP, COSIM and Earth System Grid science, we have migrated the ParaView MOC implementation to a stand alone, parallel implementation of MOC analysis. This tool is intended to be used by the CESM/CCSM workflow and larger climate community. Additionally, we are in the final stages of accelerating POP and POP analysis with additional features: 1) integration of the stand alone MOC as *in situ* analysis, 2) acceleration of POP IO routines with MPI-IO and 3) performance analysis comparing MOC in different analytical workflows.

Feature Tracking and Visualization of the Madden-Julian Oscillation in Climate Model Output



The Madden-Julian oscillation (MJO) plays a significant role in intraseasonal weather variations over the Indian and Pacific Oceans. In this work, we developed an integrated analysis and visualization tool for simulated MJO episodes.Using a Web-based interface, the tool lets scientists more easily identify cloud and environmental processes associated with the MJO .By combining domain-knowledge-assisted feature tracking and global data overviews in both space and time, our tool enables climatologists to analyze their data more effectively.

Rendering Geodesic Grid Data



- A direct rendering solution for 3D geodesic grid data
- No need to decompose the cells into tetrahedra
- High quality and GPU accelerated
- Visualization of climate data output by the GCRM (Global Cloud Resolving Model) code.
- J. Xie, H. Yu, and K.-L. Ma. *Interactive Ray Casting of Geodesic Grids*. In EuroVis 2013 Proceedings, pp. 481-490

SDAV
Scalable Data Management, Analysis, and Visualization

Contributors: E. W. Bethel; (PnetCDF) R. Ross, W.-K. Liao; (GCRM Data Model) Prabhat, K. Schuchardt; (Scalable Statistics) D. Pugmire, H. Krishnan, Prabhat, M. Wehner; (Fingerprinting) A. Sanderson; (Streaming Computation) V. Pascucci, P.-T. Bremer; (Atmospheric Rivers) S. Byna, Prabhat, J. Wu, M. Wehner; (Ocean Climate Analysis) J. Woodring; (MJO Oscillation) H.-W. Shen; (Rendering) K.-L. Ma, J. Xie, H. Yu; (UV-CDAT) D. Williams.