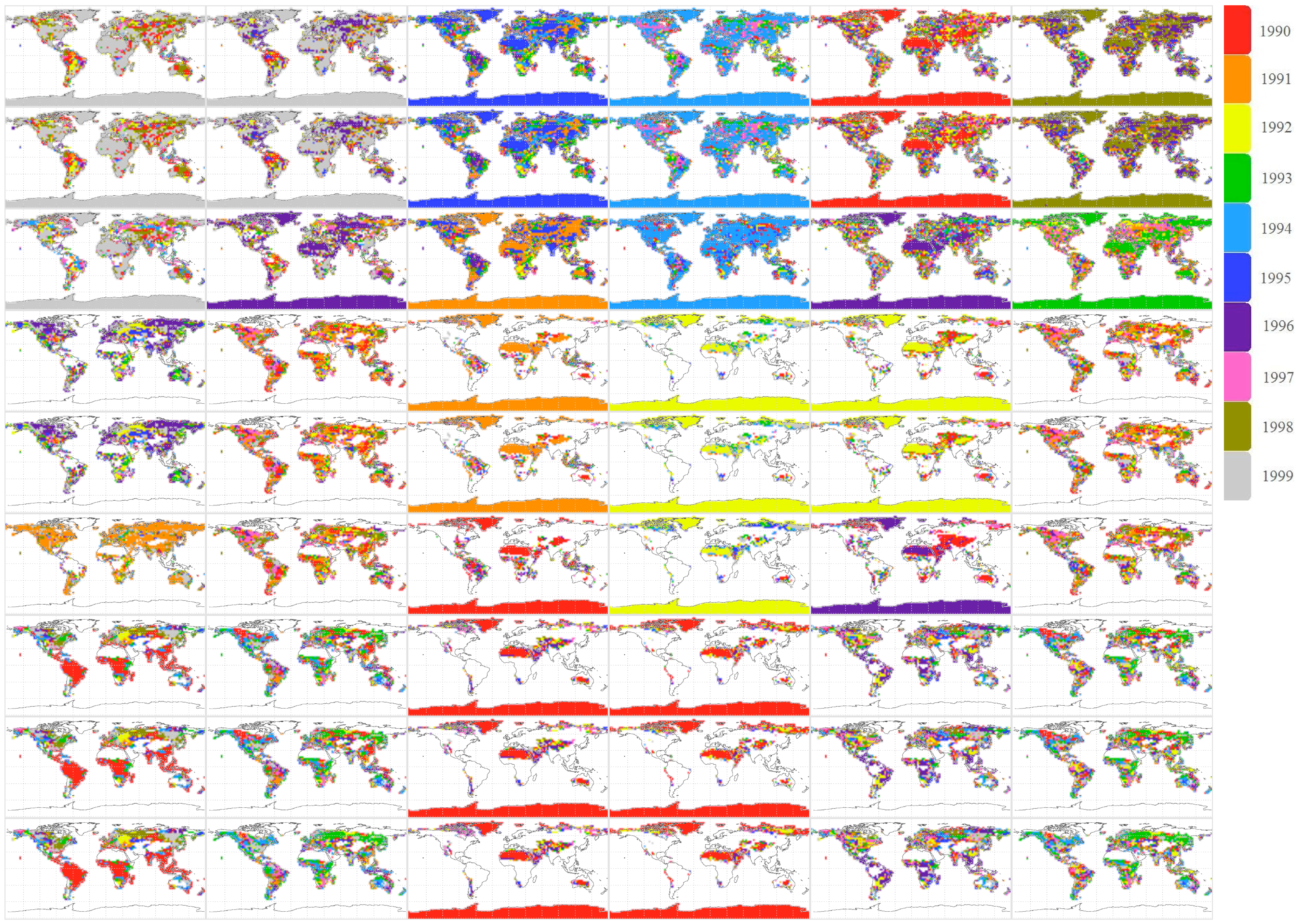
Time-Varying Multivariate Visualization for Understanding Terrestrial Biogeochemistry

Robert Sisneros+, Markus Glatter+, Brandon Langley+, Jian Huang+, Forrest Hoffman* and David Erickson III*

+UNIVERSITY OF TENNESSEE, KNOXVILLE *OAK RIDGE NATIONAL LABORATORY

We draw our research motivation from the current SciDAC thrust to develop accurate simulations of the global coupled climate-carbon cycle that model the interactions and feedbacks between the terrestrial biosphere and the climate system. In particular, the C-LAMP (the Carbon-Land Model Intercomparison Project) project is our current focus. C-LAMP bridges the climate modeling community and the measurement community to allows thorough tests and comparisons of terrestrial biogeochemistry models through a set of carefully crafted experiments. Well defined metrics have been established for comparison of model results against best available observational datasets, and models are graded on their scientific performance with respect to these metrics. Parallel visualization tools and diagnostics of large scale are particularly needed for uncovering model differences and discovering ways for improving individual models.

The visualization needs of C-LAMP model comparison are very demanding. There are a large number of variables involved in each simulation run. Exacerbated by the need to study multiple runs in a cohesive manner, the combinatorial space that needs to be explored is overwhelming, even just to study two variables from two simulation runs due to the ever increasing spatial and temporal resolutions in each variable. Our research on ultra-scale visualization led to a successful prototype system that combines two recent trends in large-data visualization: query-driven visualization and summarizing visualization. While our inherently parallel query-driven visualization system (presented on IEEE Visualization Conference'2006) provides a scalable visualization infrastructure, our new method of attribute-subspace based concurrent visualization (presented on EuroVis Conference'2008) produces the final summarizing visualization.



In our research, we visualize three different types of runs: (i) control run (C-LAMP 1.2), (ii) varying climate transient run (C-LAMP 1.3) and (iii) varying climate, CO_2 and N deposition transient run (C-LAMP 1.3). Under each type of run, we study two models from NCAR Community Climate System Model Version 3: CCSM3-CASA and CCSM3-CN. Hence we have in total 6 different simulation scenarios to visualize. The time span is the decade from 1990 till 1999. Three variables are considered: net ecosystem exchange of carbon (NEE), net primary production (NPP) and total leaf area index (TLAI).

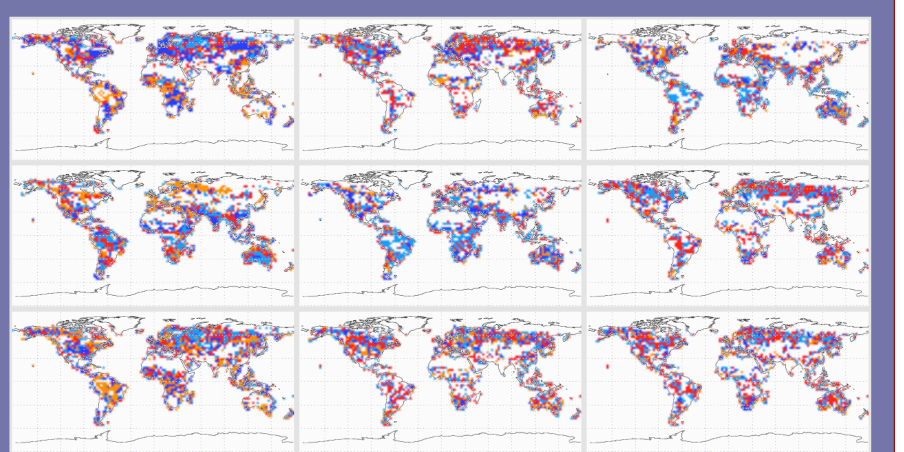
We use a location-specific summarizing method to visualize extreme and normal relative patterns among multiple concurrent attributes. Our motivation is to guide a user's attention to a much-reduced subset of a large and otherwise incomprehensible multivariate dataset. In order to do such guidance in an exploratory fashion, we base our methods on relative distribution patterns of multiple variables. In the overall problem space formed by all concurrent attributes, we term each subset of interest in our framework an attribute-specific subspace. We normalize all attributes to a canonical range. Also, the user supplies attribute target values that allow us to determine if the value at any location is "good", i.e. close to its target value. Common examples are: heaviest seasonal rainfall around the globe or longest period of drought around the globe. At each location, classification is based on which attribute at this location is the closest to its target value. Each location is classified and assigned to at most one attribute-specific subspace. When all values on a location are sufficiently far away from the target values, this location is thresholded and assigned to no subspace.

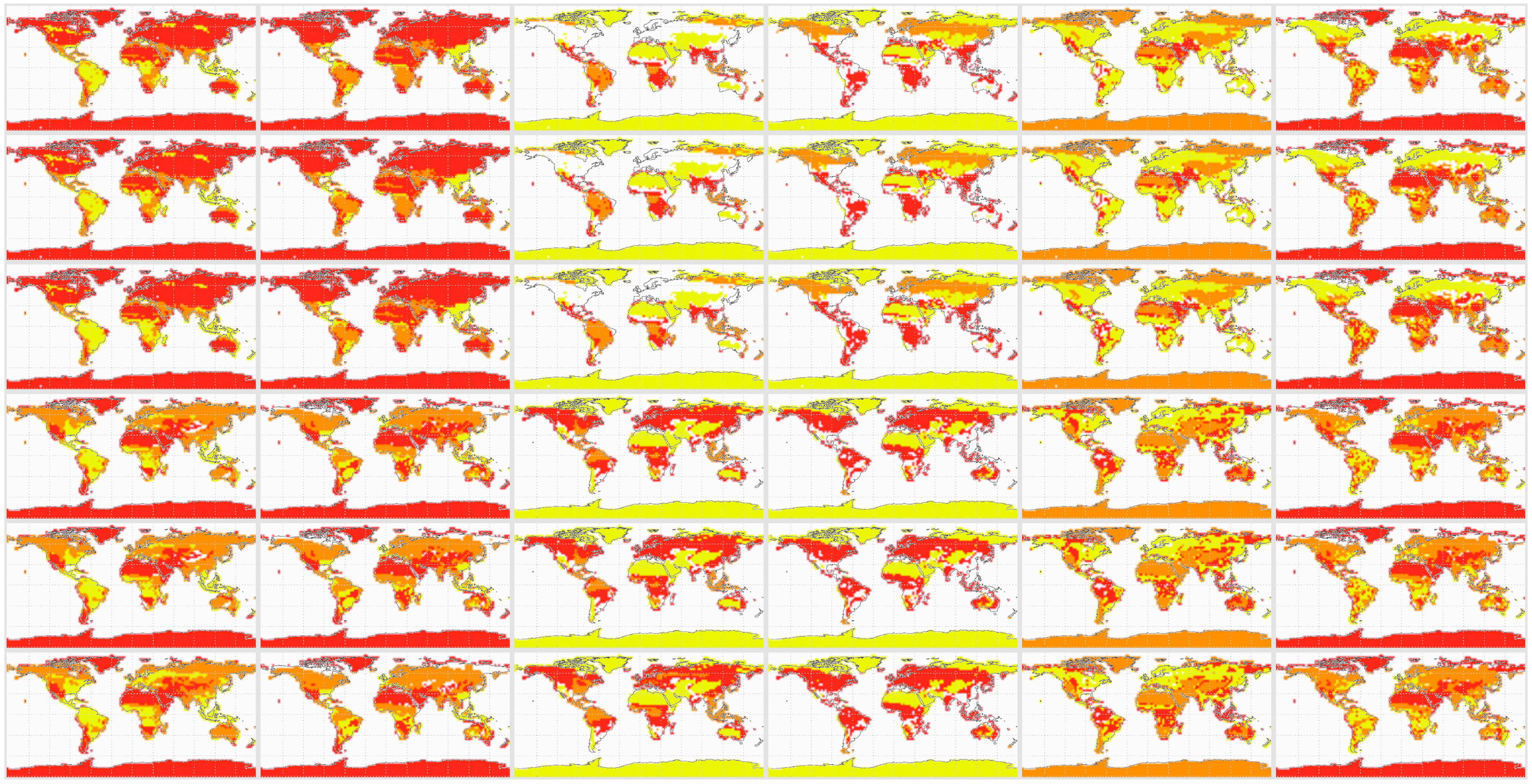
Runtime visualization data accesses are factored into compound boolean range queries based on the target values chosen for the individual variables. The parallel query-driven infrastructure can efficiently handle terascale data using 20 to 30 compute nodes for most use cases. The system has been shown to scale to much larger computing systems.

In the top and far-right figures, the left two, middle 2 and right 2 columns show cases where target values for all variables considered are global maximum, global minimum and mid-way between global extremes, respectively. **The Top Figure**. The top 3 rows show relative distribution of NEE's yearly averages in 1990-1999 from C-LAMP 1.2, 1.3 and 1.4 runs, respectively. The middle and bottom 3 rows are for NPP and TLAI, respectively. Columns #1, #3, #5 are from CASA model; #2, #4, and #6 are from CN model.

The Far-Right Figure. The organization of this figure is the same as the top figure. The variables shown are relative strength among NEE (red), NPP (orange) and TLAI (yellow). The top three rows are from January 1990, and the bottom three rows are July 1990.

The Right Figure. Pearson's correlation coefficient between pairs of variables for each simulation scenario over 1990-1999: red (1.2 CN), orange (1.2 CASA), yellow (1.3 CN), green (1.3 CASA), light blue (1.4 CN) and blue (1.4 CASA). The three rows (top to bottom) show NEE-NPP, NEE-TLAI and NPP-TLAI, respectively. The columns (left to right) are: positive, negative and no correlation, respectively.





The primary source of funding for this work is provided through Institute of Ultra-Scale Visualization (www.ultravis.org) under the auspices of the SciDAC program within the U.S. Department of Energy (DOE). This work is also supported in part by NSF grant CNS-0437508 and a DOE Early Career PI grant awarded to Jian Huang (No. DE-FG02-04ER25610). The Carbon-Land Model Intercomparison Project (C-LAMP) are partially sponsored by the Climate Change Research Division (CCRD) of the Office of Biological and Environmental Research (OBER) and the Computational Science Research and Partnerships (SciDAC) Division of the Office of Advanced Scientific Computing Research (OASCR) within DOE Office of Science. This research used resources of the National Center for Computational Science (NCCS) at Oak Ridge National Laboratory (ORNL), which is managed by UT-Battelle, LLC, for DOE under Contract No. DE-AC05-00OR22725. We are also in debt for collaborative support provided by Sean Ahern, Visualization Task Leader of ORNL and a co-PI of DOE SciDAC VACET Center (www.vacet.org).