Visualization at the Oak Ridge Leadership Computing Facility

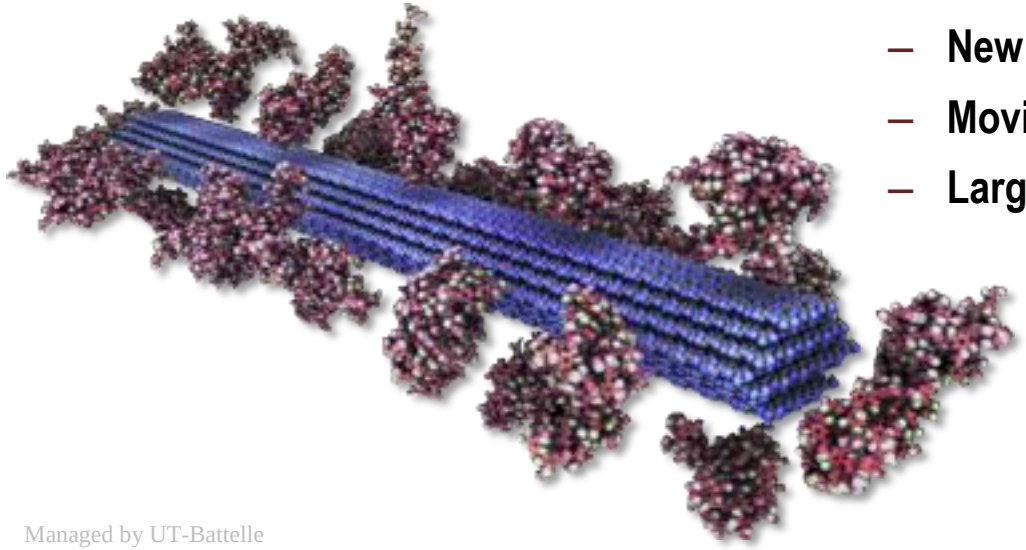
Presented by

Dave Pugmire

Oak Ridge National Laboratory

We provide data analysis and visualization services for ORNL's HPC users

- Many application domains
  - Magnetic confinement fusion
  - Climate
  - Material science
  - Nuclear energy
  - Astrophysics
  - Geographic information systems
  - …



- Focus on large data
  - Large, distributed analysis systems
  - Parallel tools: VisIt, Paraview, EnSight
  - Core competency in remote visualization
- Production visualization development team
  - Custom tools
  - New data exploration techniques
  - Movie/image generation
  - Large display support

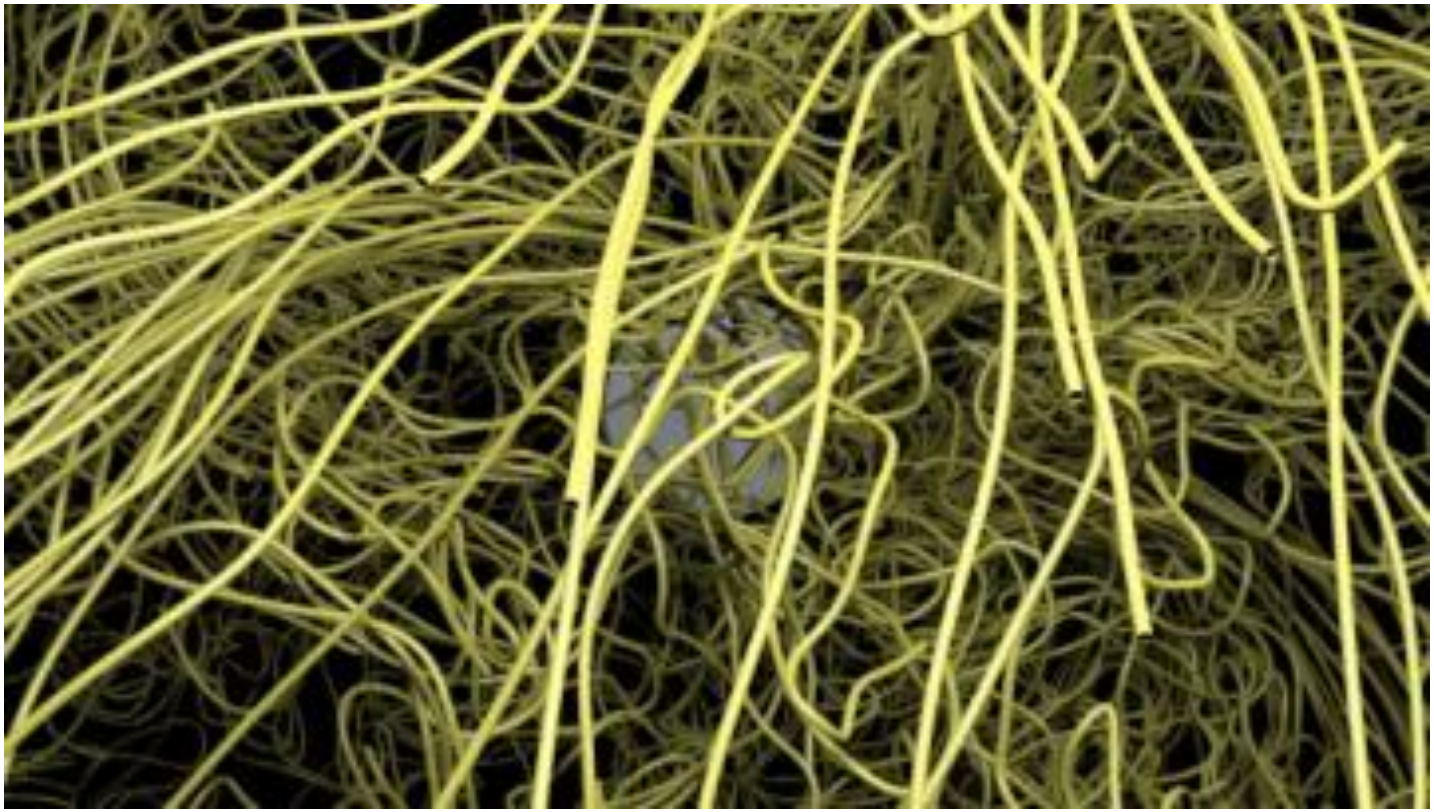
Preparing for Visualization and Analysis at the Exascale

- ORNL Vis team member Sean Ahern organized the "Scientific Discovery at the Exascale" workshop
- The two day workshop explored the implications of exascale architectures on the data management, analysis and visualization communites
- Recommendations include a focus on:
  - in situ processing frameworks
  - I/O middleware systems
  - fine grained algorithms to exploit future architectures
  - co-scheduling of analysis and simulation
- Full report available at *http://science.energy.gov/ascr/news-and-resources/program-documents*

Magnetic Fields in Core-Collapse Supernovae

Simulations performed on Jaguar by Eirik Endeve, Christian Cardall, Reuben Budiardja, and Anthony Mezzacappa

Winner of SciDAC 2011 Visualization Night People's Choice Award

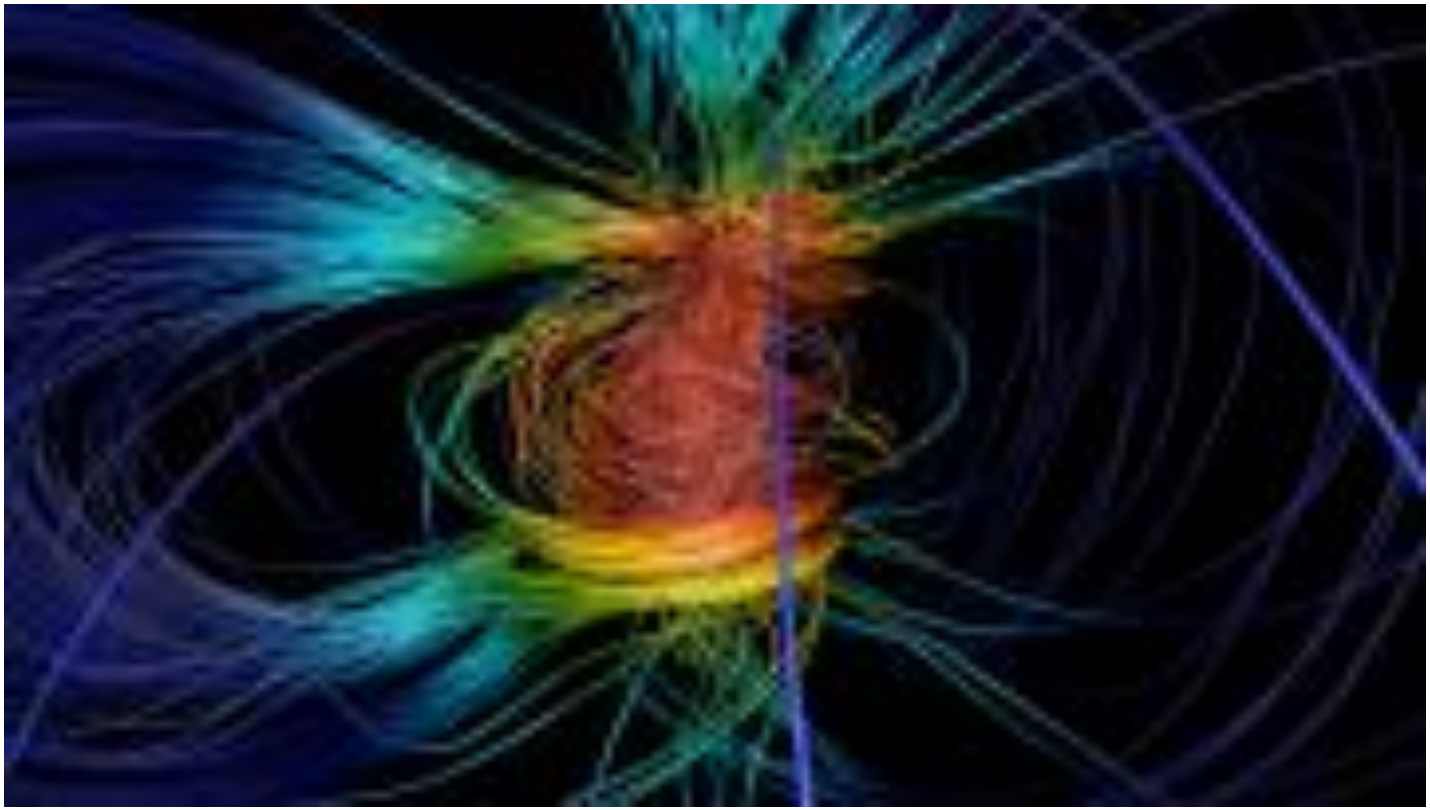


Visualization by Dave Pugmire

Magnetic Outflows from Active Galactic Nuclei

Simulations performed using FLASH by Paul Sutter, Paul Ricker, H.Y. Yang, and G. Foreman

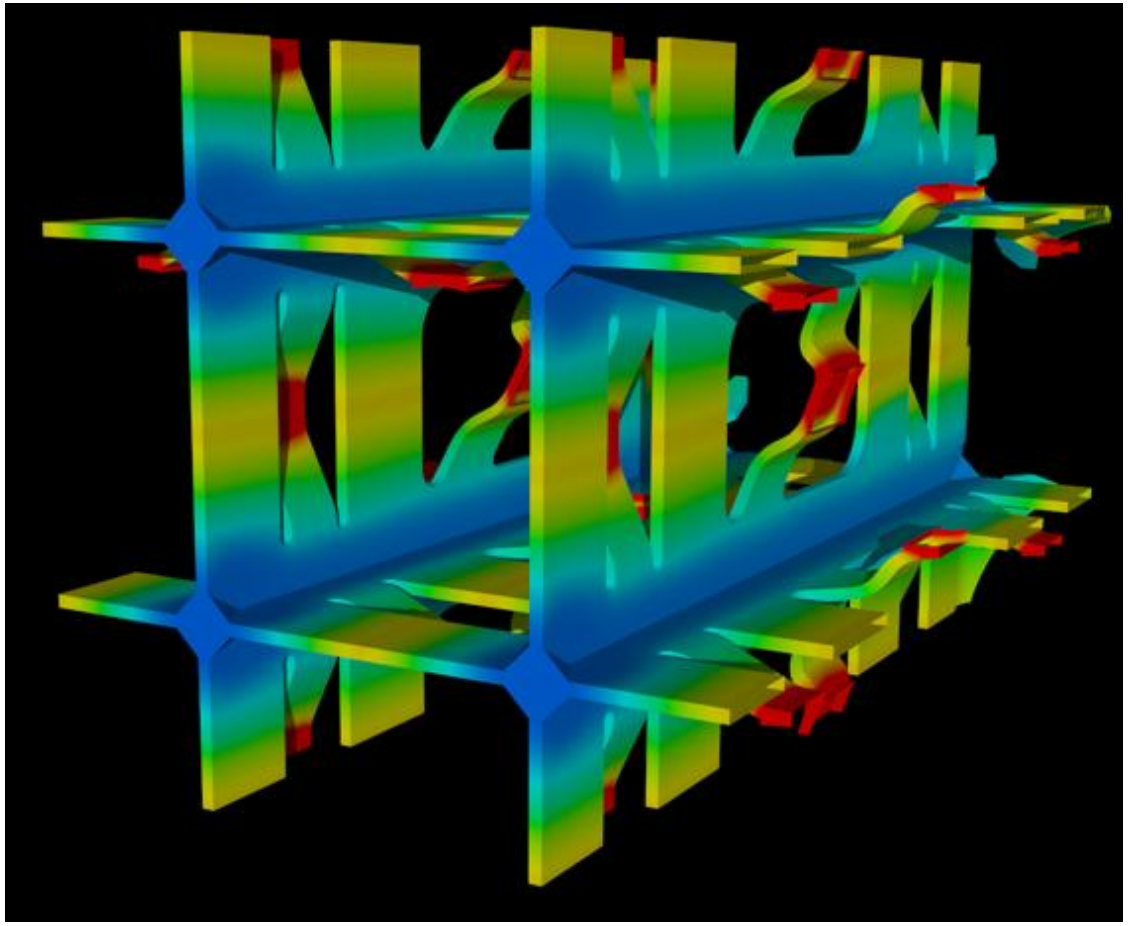
Winner of SciDAC 2011 Visualization Night People's Choice Award and awarded "Best Visual Presentation"



Visualization by Dave Pugmire

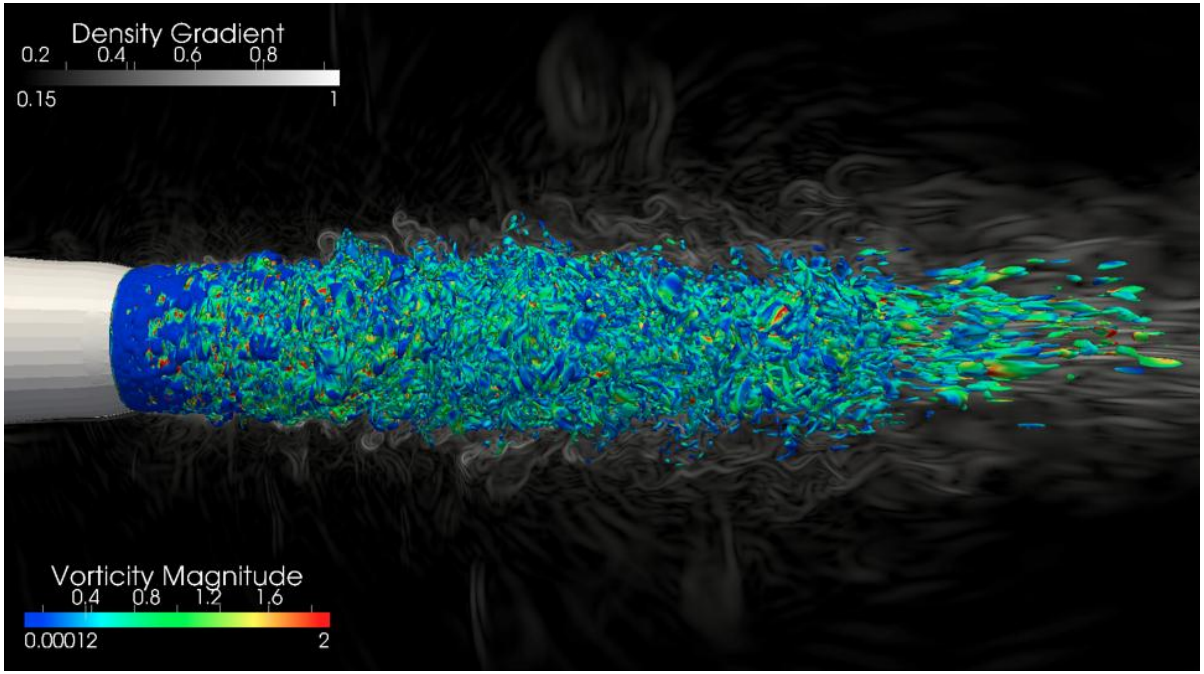
Visualization and Visualization Infrastructure Support for CASL

- Design of F5 CAVE immersive environment
- Development of software infrastructure for parallel visualization of large mesh problems
- Collaborations with project partners for workflow development and utilization of HPC resources



Visualization for Discretionary and Industrial Sector Projects

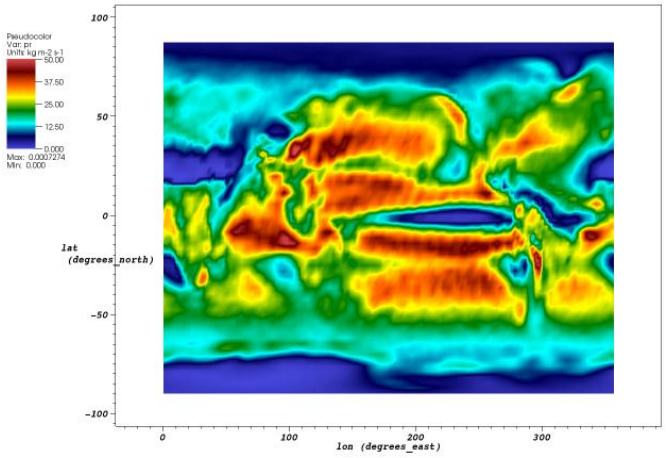
- Problem characterization and workflow planning
- Development of software infrastructure for parallel visualization
- Development of software infrastructure for remote visualization

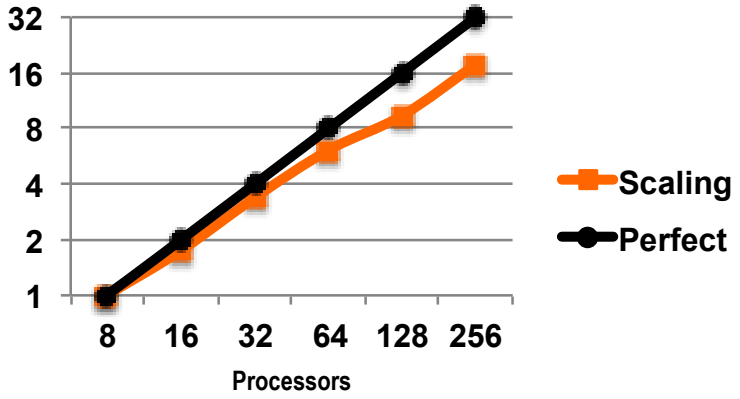


Integrating Statistical Tools into VisIt

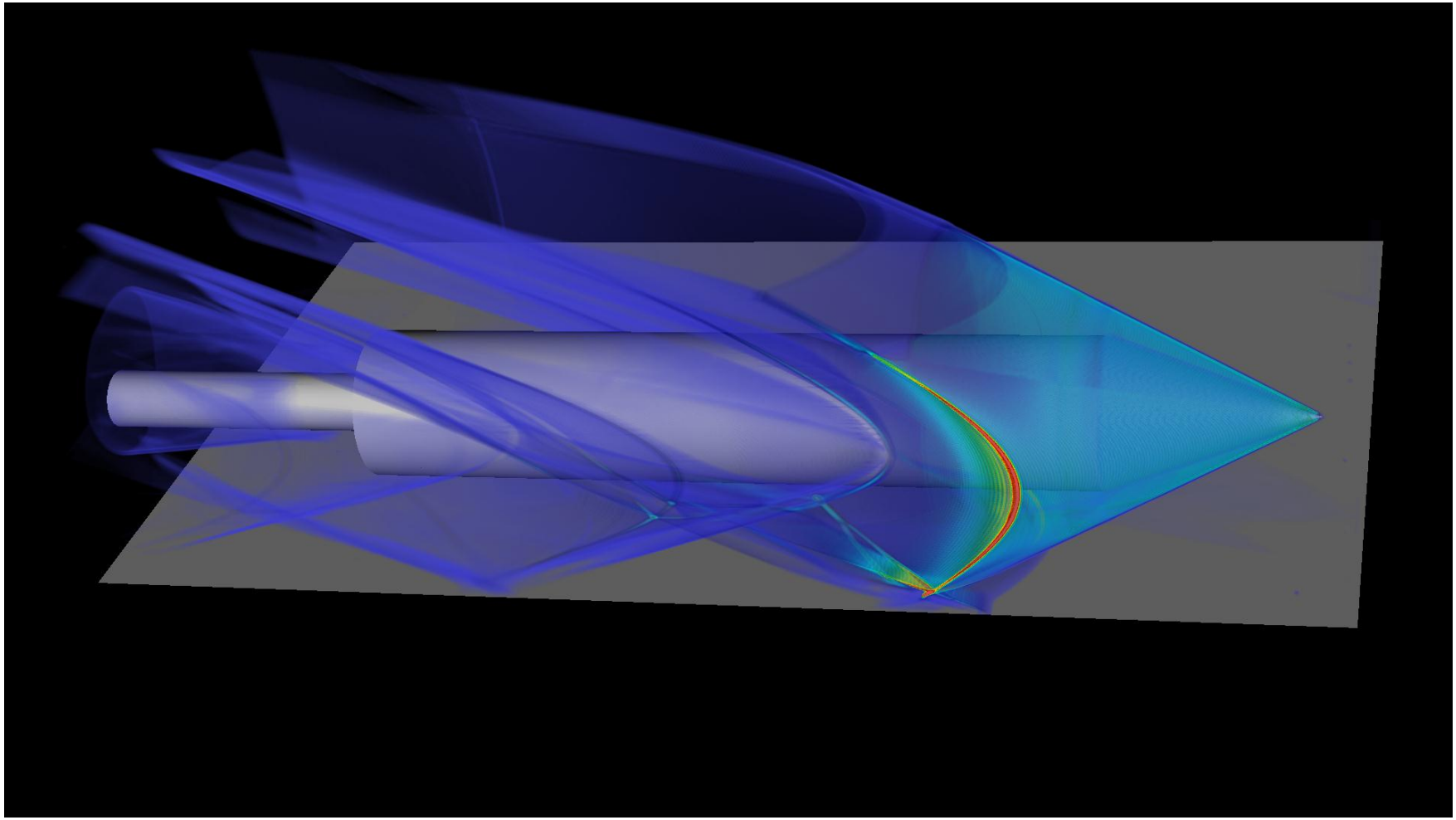
- R is the de facto standard for statistical analysis
- Recently, VTK has provided support for executing R code on VTK data structures
- Through funding from BER, statistical analysis will be made possible within VisIt, in parallel, using R

Scalability of Global Extreme Value analysis of 100 year precipitation data

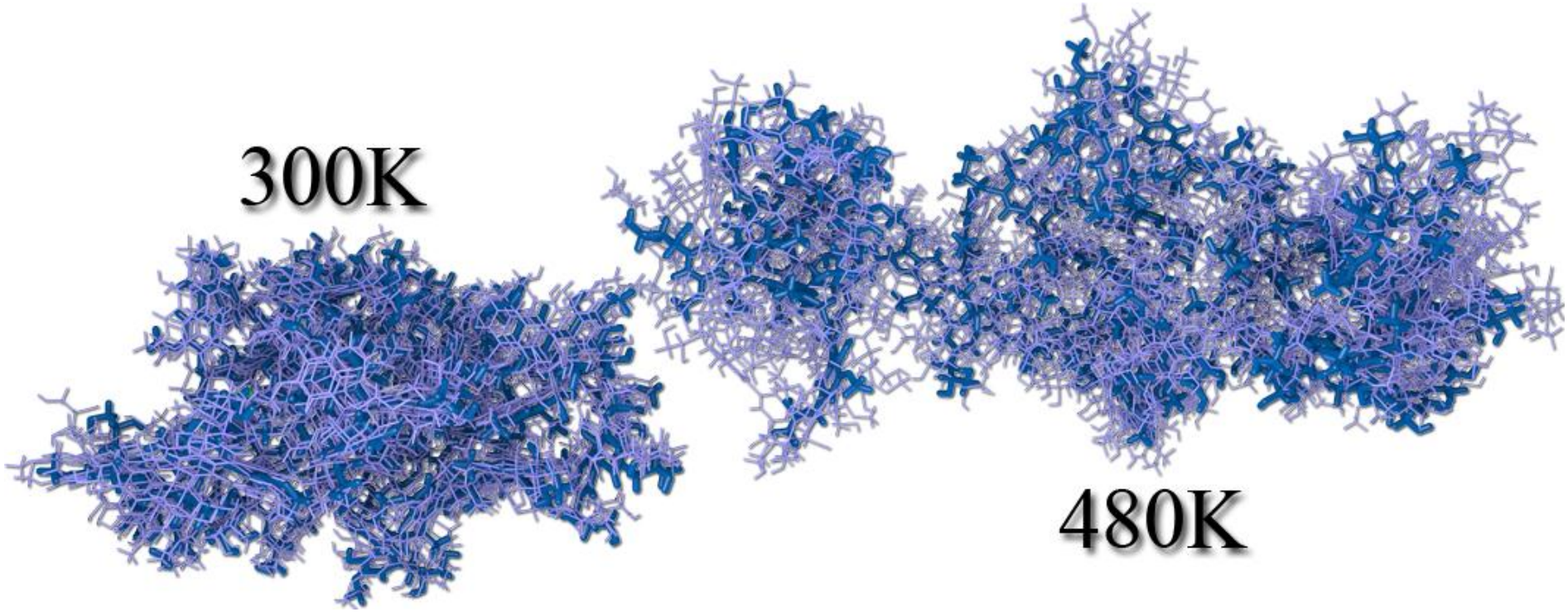




Methods to determine complex shock patterns in simulations using VisIt

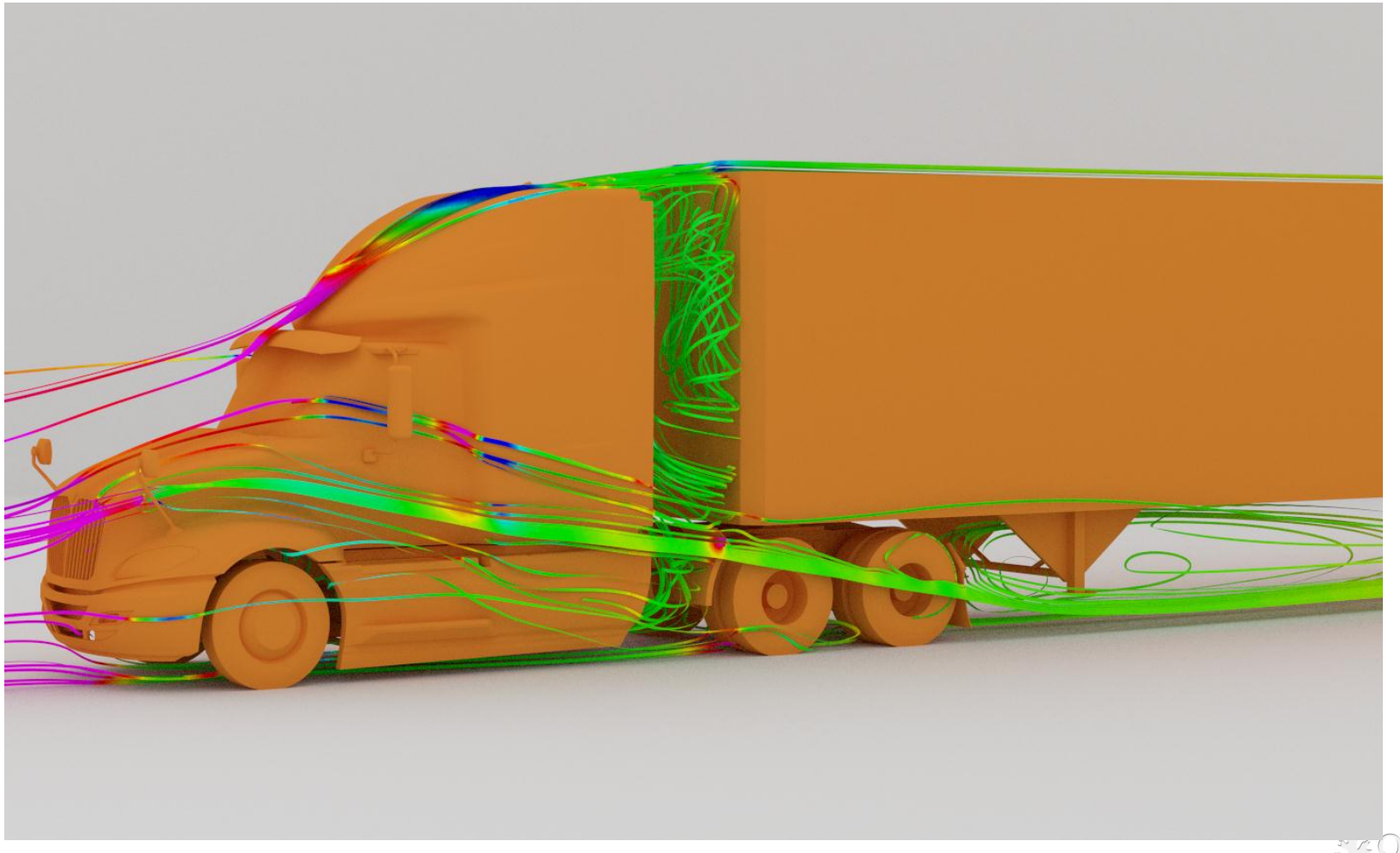


Time Dependent Molecular Visualization of Cellulosic Ethanol



Molecular structure at two different temperatures

Particle Tracing and Advanced Rendering in Support of Industrial Partners



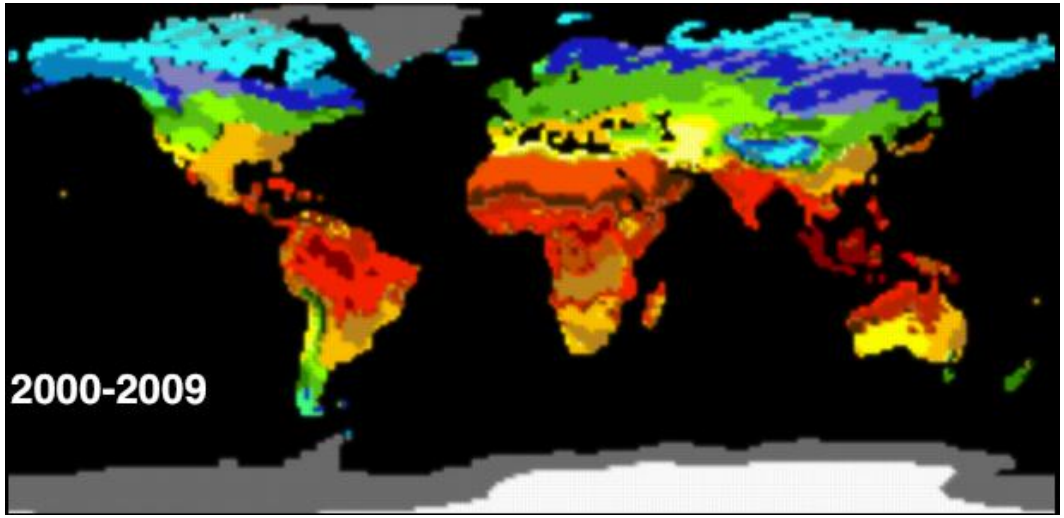
Overview Visualizations of Multivariate Data via Model-based Classification

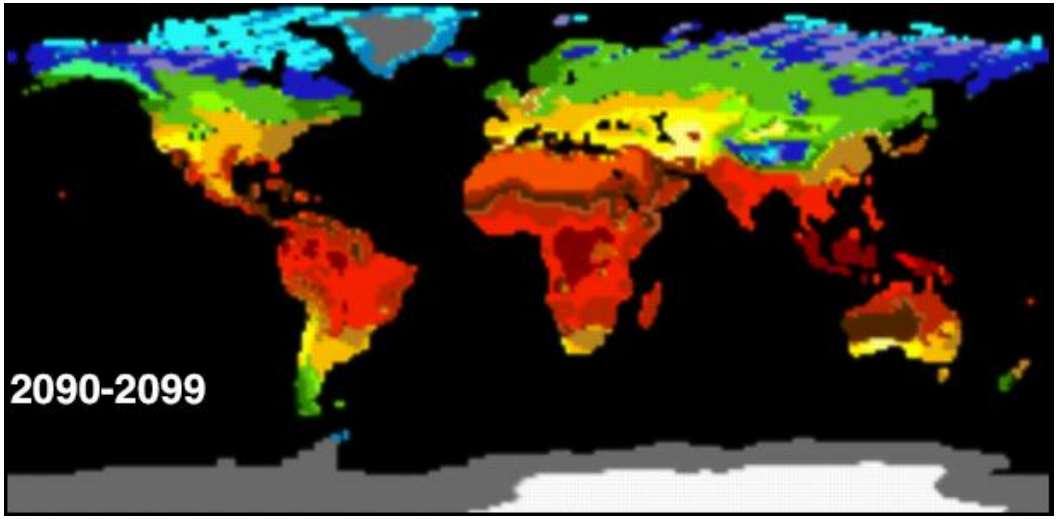
Summarize multiple user-defined high-dimensional inter-variable models

Spatial locations classified by distance to these models

Example use:

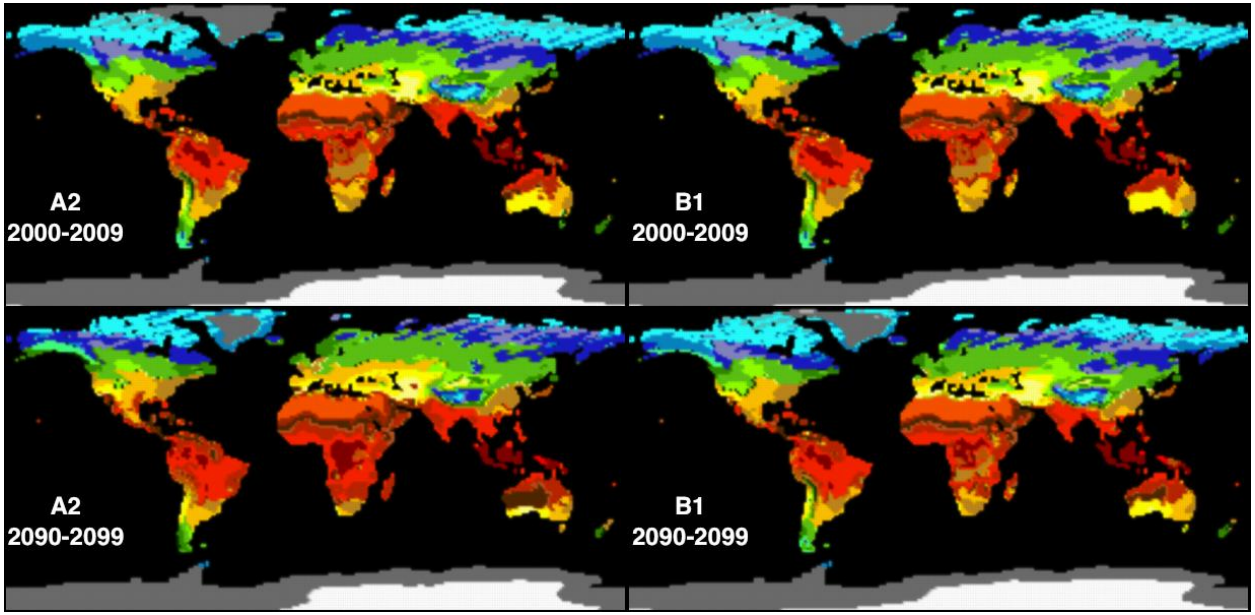
1. Ecoregionalization classification of climate data via 3 differentiating factors
2. 33 3-dimensional models specified concurrently
3. Allows for the comparison/analysis of initial and end stages of the simulation
4. Shows regional tendencies toward warmer regions in this particular high emissions scenario



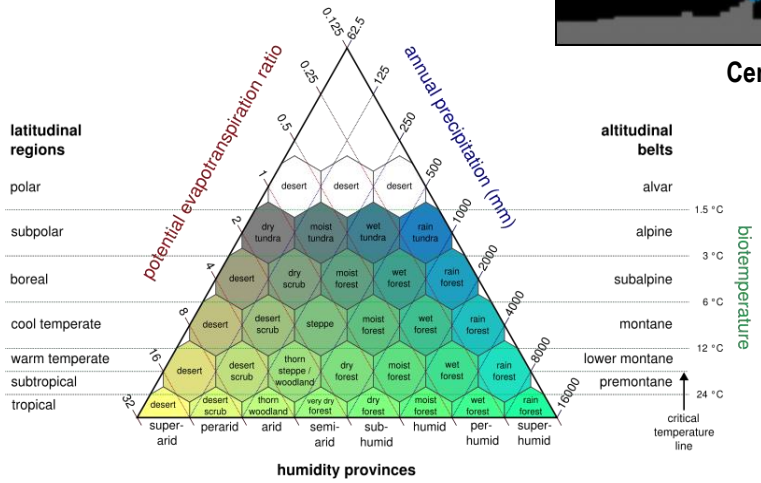


Multivariate Comparison of Climate Simulations

- A multivariate classification capability as a VisIt plugin
- Demonstrated with Holdridge life zones to compare climate simulations



Century comparison of Holdridge life Zones under climate scenarios A2 and B1

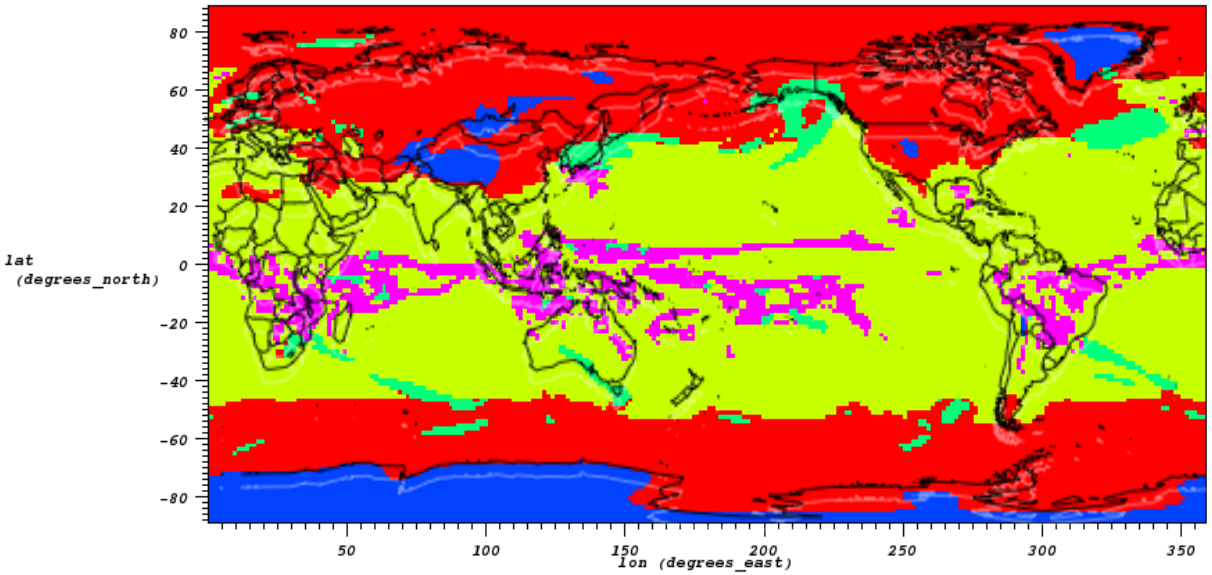


Holdridge Life Zones

- Track changes in multivariate climate that drive changes in ecology
- Compare ecology impact of climate scenarios

R. Sisneros, J. Huang, G. Ostrouchov, and F. Hoffman (2011). *Procedia Computer Science,* Vol. 4, p1582-1591.

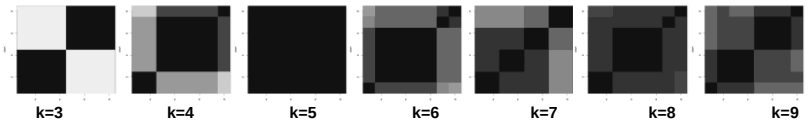
Clustering in data-parallel R for automated extraction of climate events



10 years of daily data:
Lat x Lon x Day for 5 variables:
119,603,200 x 5 matrix (3 GB)

- **Cluster in R without** lat, lon, **and** time **information (semi-supervised)**
- **Play resulting clusters as** lat **by** lon **in** time **with VisIt**
- **Sampling reduces clustering time by order of magnitude (cluster model parameter uncertainty)**
- **Random start agreement selects number of clusters (classification uncertainty)**

Pairwise agreement "max kappa correlation" of 8 starts

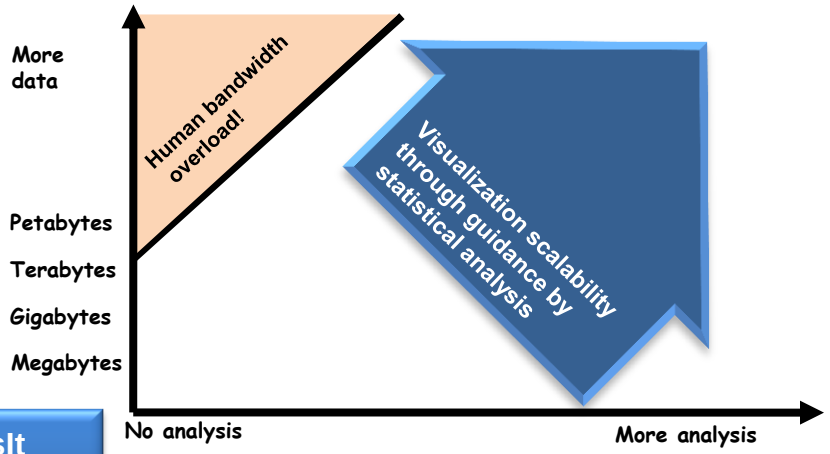


Building R-VisIt statistical framework for visualization of massive data sets

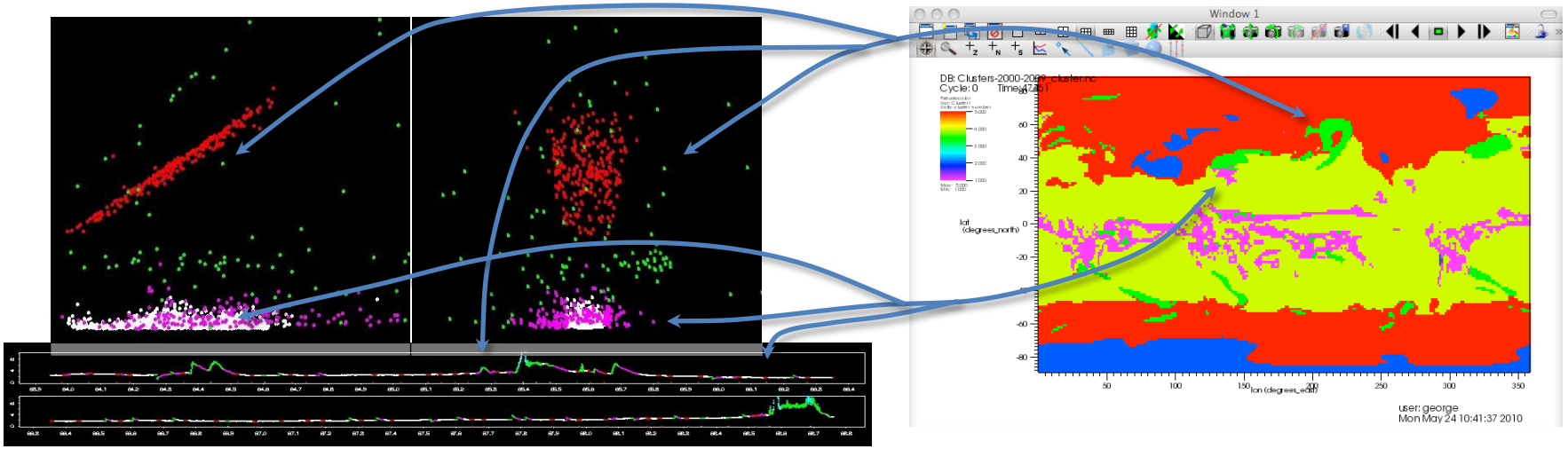
Browse a petabyte? Not humanly possible!

To view a petabyte at 100 MB/s takes 350 8-hour workdays! All year – only 8 weekends off!

Statistical analysis must select views or reduce to quantities of interest in addition to fast rendering

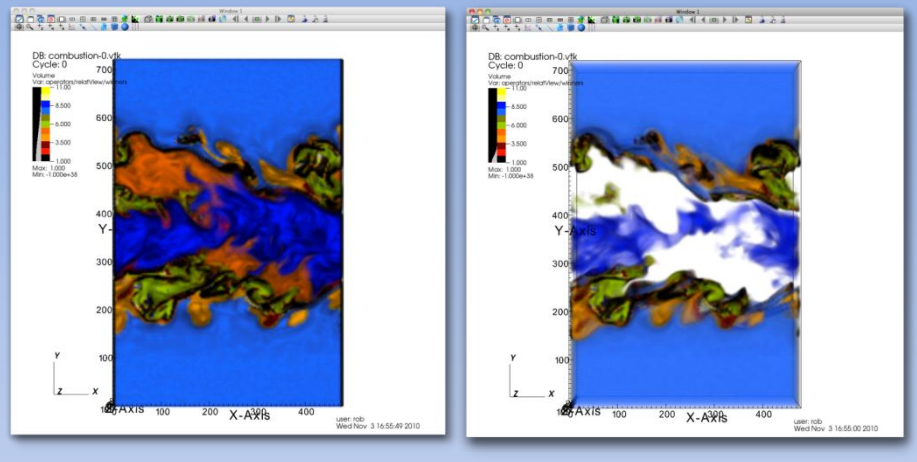


- Fit local models to segmented data with R inside VisIt
- R feature analysis in high-dimensional feature space
- Play statistically selected VisIt views of interest

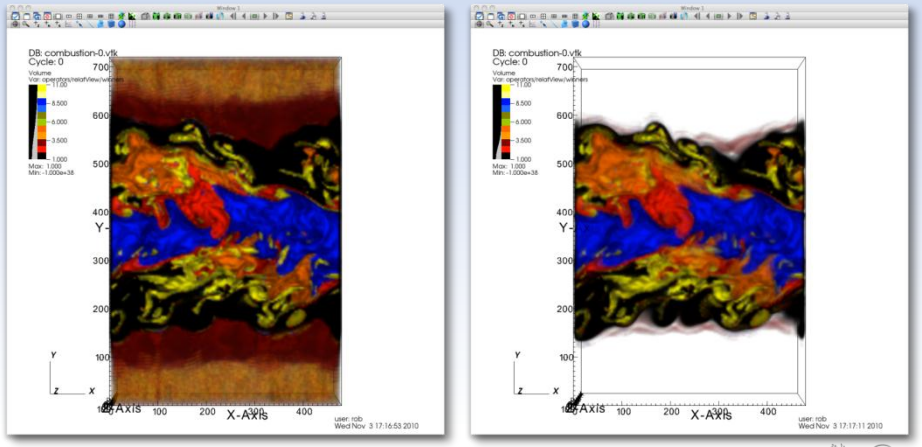


High-dimensional multivariate analysis

- A method to effectively study interactions between multiple variables by building and verifying hypothetical predictive models
- Multiple candidate models may be specified together, and our system classifies each data point by its distance to each candidate model. We then concurrently view all specified models
- Several distance metrics are considered, and calculations may take place in different analytical spaces

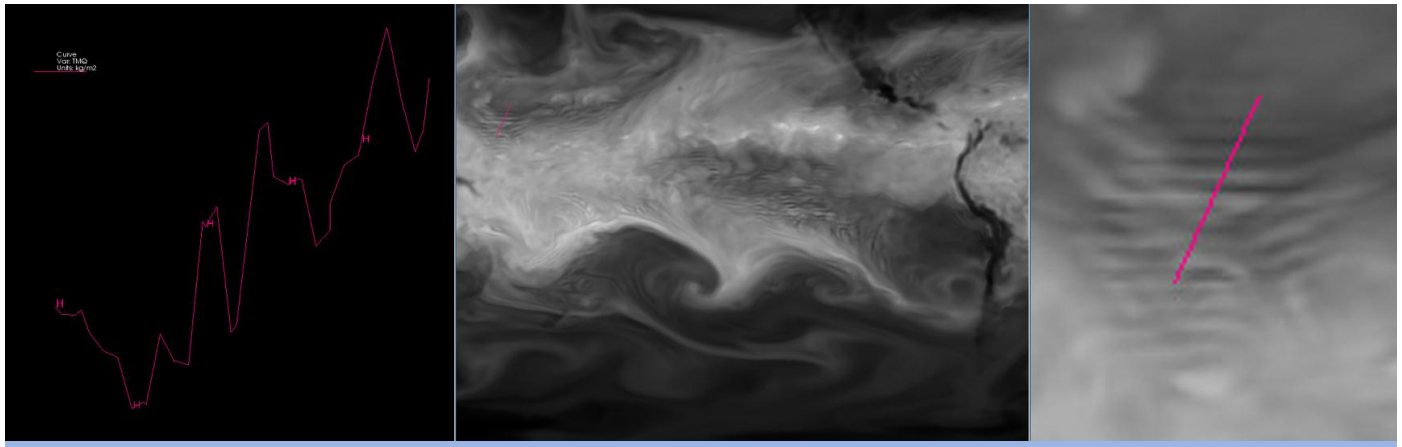


Classifications of a 5-dimensional combustion dataset based on 11 different models. In each pair of images, the rightmost are focused on models with minimal distances. Euclidean distances are calculated in *(top)* normal variable space and *(bottom)* probability density space



Climate science: Atmospheric water vapor in the Community Atmosphere Model

The T341 version of the Community Atmosphere Model, the atmospheric component of CCSM, runs with a resolution three times finer than typically used for climate simulation in order to resolve local features at the subglobal level (e.g., hurricanes). Scientists used data from a T341 model run to produce high fidelity visualizations; they observed a ringing effect emanating from regions with high values of integrated water vapor. Consulting with Jim Hack, the leader of Oak Ridge's Climate Science Institute and long-time climate model developer, visualization expert Jamison Daniel and climatologist Kate Evans created additional visualizations to determine the source of the problem

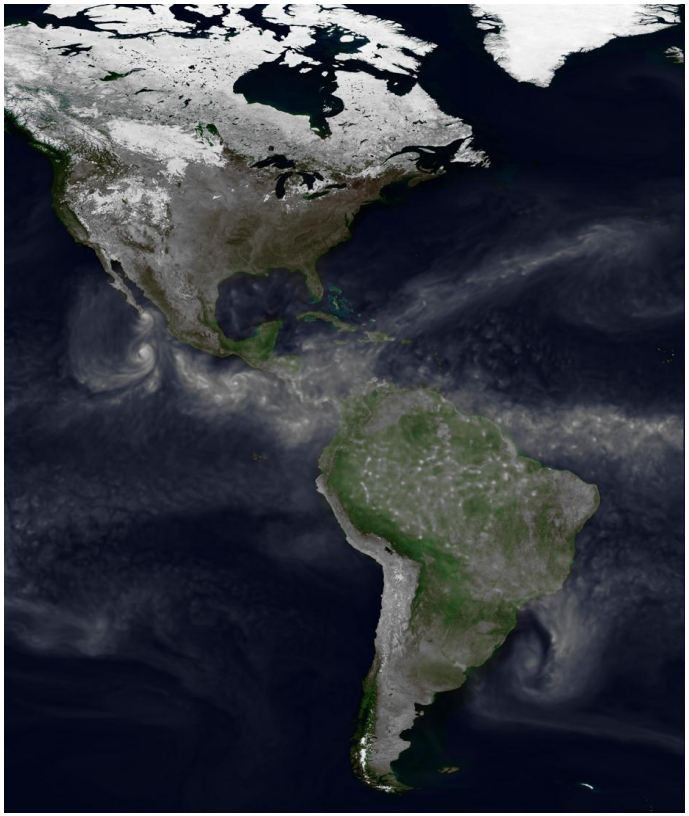


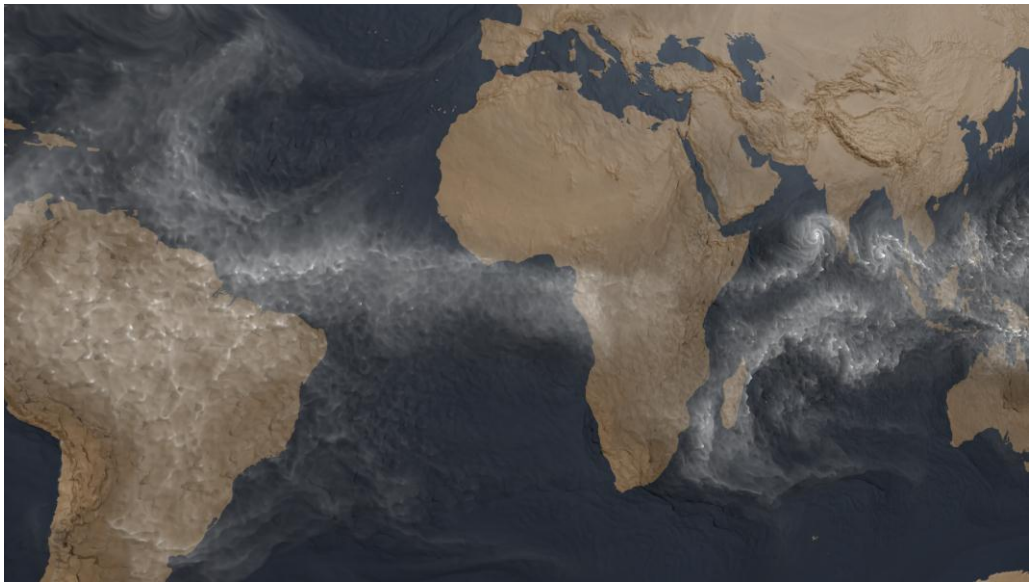
(Middle) A snapshot of total precipitable water from a T341 simulation. *(Left)* The level of precipitable water across one wavelike feature evident in the precipitable water field. The region across which the precipitable water is displayed on the left is designated by a purple line in the middle and on the right. This figure illustrates the scale of the wavelike feature and aided in the diagnosis of the “ringing” effects in the model.

Visualization of CCSM Community Atmospheric Models

The proper simulation of the distribution of water vapor in the climate system is essential to the accurate treatment of the hydrological cycle and the planetary radiation budget. This image show the simulated monthly-averaged distribution of the total column water vapor from a high-resolution configuration of the CCSM Community Atmospheric Model.

Visualization of CCSM Community Atmospheric Models

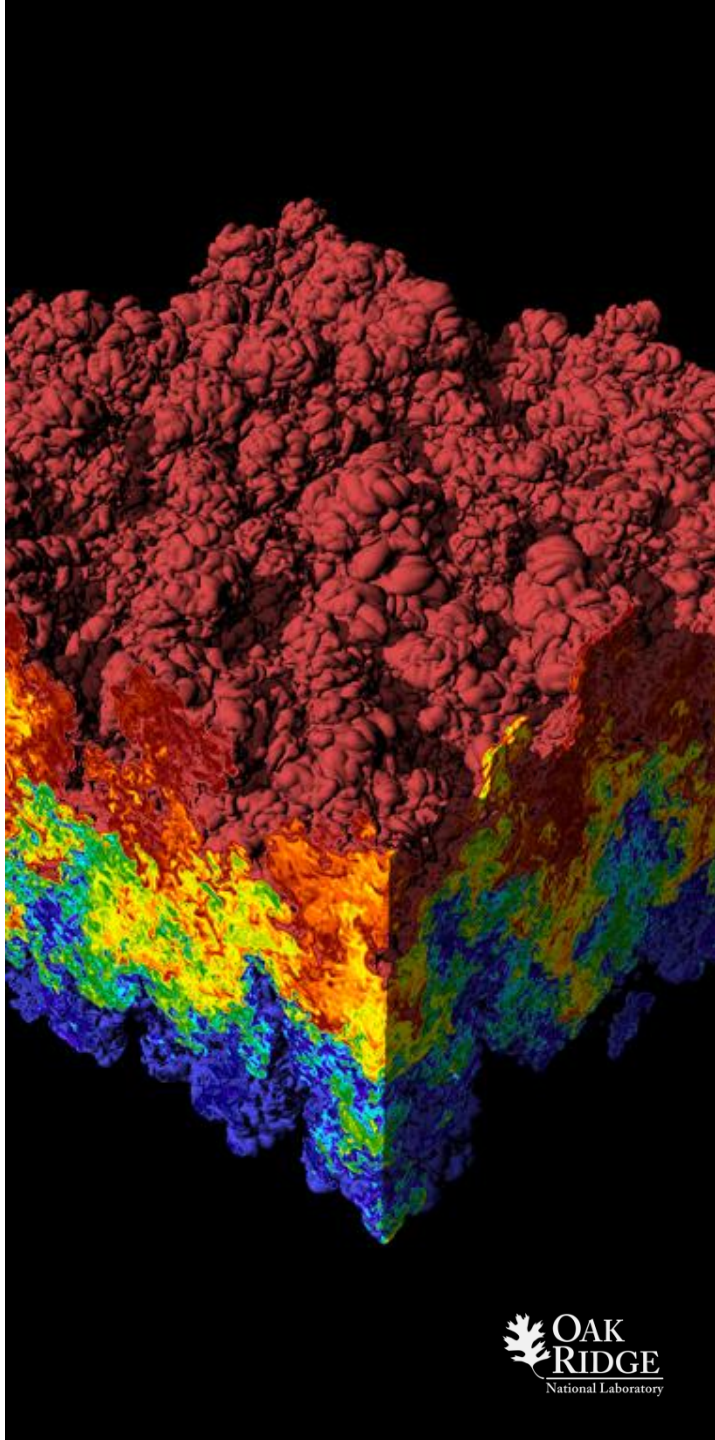




The proper simulation of the distribution of water vapor in the climate system is essential to the accurate treatment of the hydrological cycle and the planetary radiation budget. This image show the simulated monthly-averaged distribution of the total column water vapor from a high-resolution configuration of the CCSM Community Atmospheric Model.

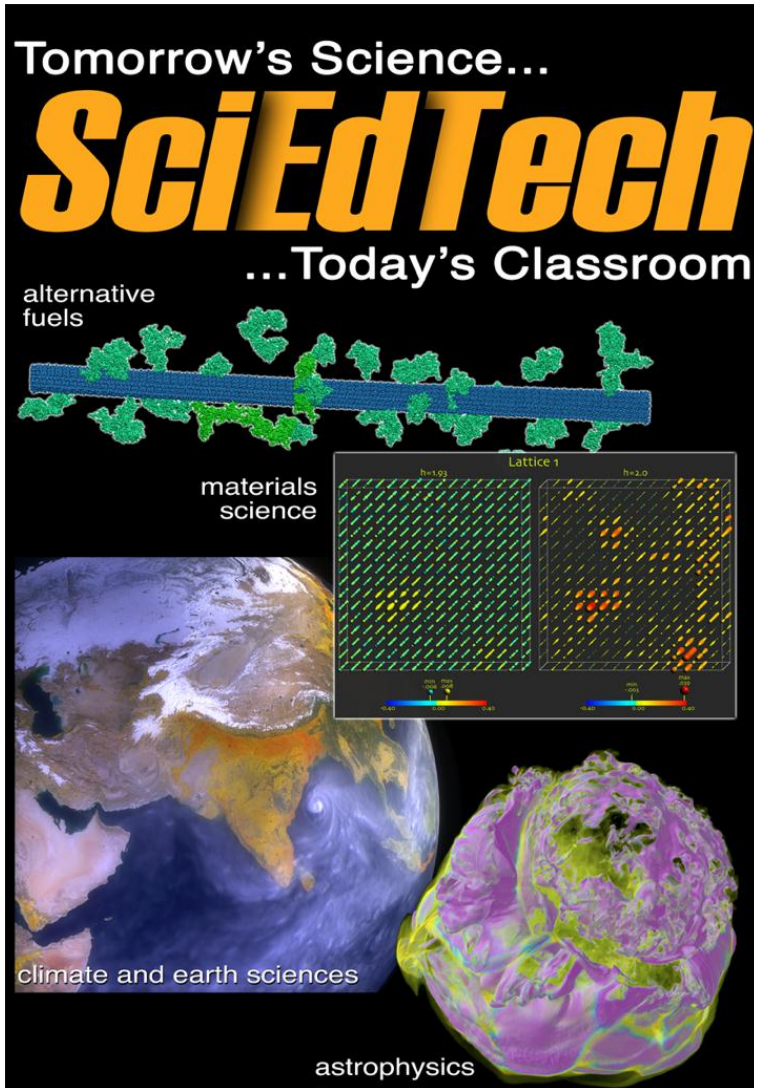
Co-developers of VisIt: An end user visualization and analysis tool for large data

- 5 major use cases: data exploration, data analysis, visual debugging, comparison, and presentation
- Demonstrated scaling to >100k cores
- Strong emphasis on building a usable, robust tool for scientists
- Open source project, with direct support from DOE (SciDAC, NEAMS, ASC, BER) and NSF (TeraGrid XD)
- visitusers.org wiki with 400+ pages of documentation and over 80,000 page views
- Over 100,000 downloads
- R&D 100 Award in 2005



Visualization as a Tool for Enhancing Middle School Science Education

- Utilization and subsetting of existing open science datasets
- Extension of open source visualization software for younger audiences
- Integration of visualizations with successful modular science curricula
- Collaboration between scientists, science educators, and students
- Quantification of success for augmented curricula



EVEREST facility

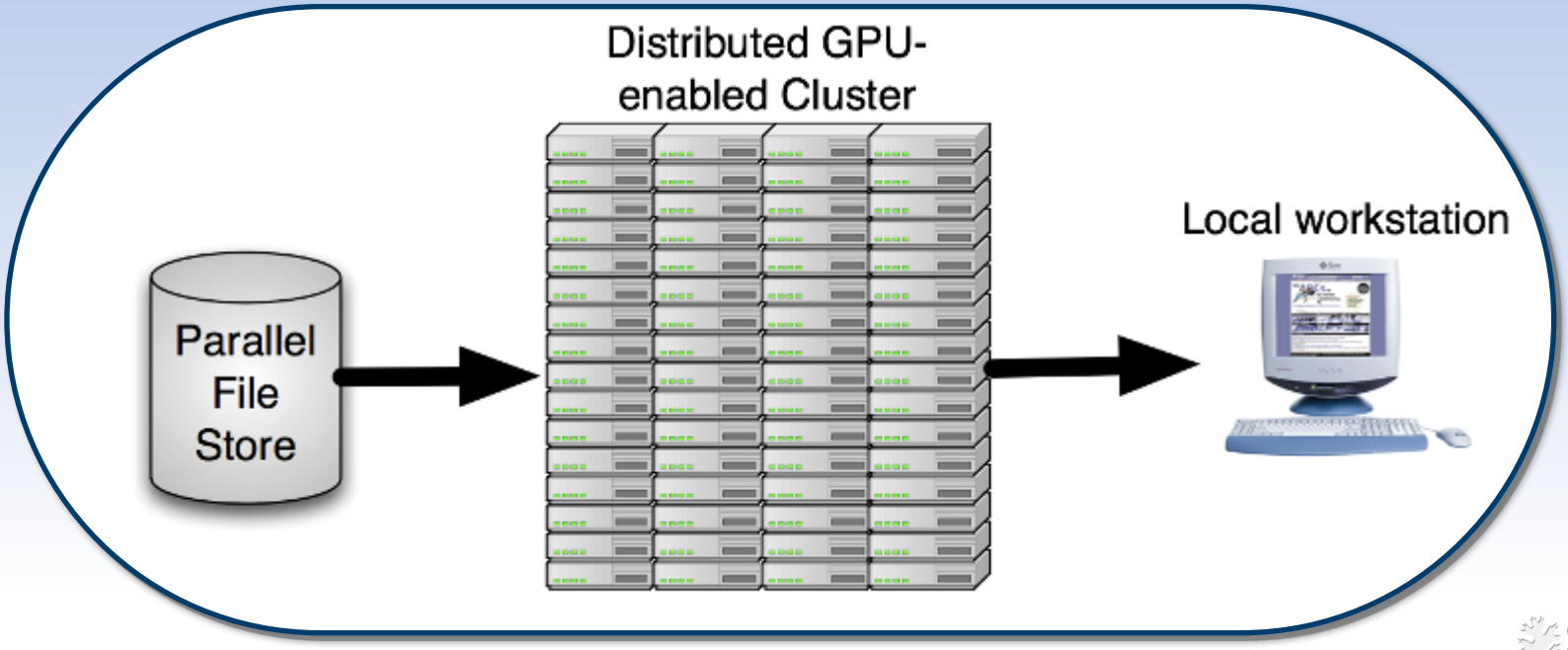
- 35 million pixel, 27-tile PowerWall
- 27 NVIDIA 8800 GTX GPUs, dedicated Linux cluster
- Interactive, large-scale, collaborative data analysis
- 30 feet by 8 feet

ORNL is one of the institutions in the BER LAB 10-05 project

- **Funded to develop and deliver software**
  - Supports high-throughput visual data analysis of ultra-high resolution data
  - Spatial and temporally high resolution simulation and observational data
- **Focus on addressing climate science questions**
  - Better understanding of extreme events and trends in climate change (near term)
  - Regional-scale drivers, attributions and consequences in a changing climate (long term)

Remote visualization for large data

- Largest datasets require use of institutional resources
- Reduces data movement issues
- Allows exploitation of multiple GPUs
- Provides visualization to remote users
- Exploited by VisIt, ParaView, EnSight



Contact

Sean Ahern
Scientific Computing Group
National Center for Computational Sciences
Oak Ridge National Laboratory
ahern@ornl.gov

Jamison Daniel
danieljr@ornl.gov

Dave Pugmire
pugmire@ornl.gov

Mike Matheson
mathesonma@ornl.gov

Ross Toedte
toedterj@ornl.gov