Preprocessing Climate Data for Access to Local Spatial Extremes

Yael M. Camacho Bonaparte
Polytechnic University of Puerto Rico

Houssain Kettani, Faculty Advisor
Polytechnic University of Puerto Rico

George Ostrouchov and Robert Sisneros, Mentors
Computer Science and Mathematics Division
Oak Ridge National Lab

Background

Changes in weather patterns, referred to as climate changes are related to changes in the global energy balance. The energy received from the sun is transported around the globe by different climate events such a hurricanes, high and low pressure systems, jet streams, and ocean currents. Today's climate simulations are based on complex mathematical models, are composed of collections of very large computer codes running on supercomputers, and produce very large quantities of detailed data. In order to better understand climate change, we must explore and analyze climate events extracted from such large simulation data sets. Because of the data size, tracking of continuous and sustained changes in climate events needs to be automated. The main result of this research is a stand-alone preprocessing code that can be run on climate data to access local spatial information for computing local extremes. This information is then added to the original input data for further exploration and analysis. We use VisIt, a parallel data analysis and visualization system for large data [1], for further processing of the data.

Climate Simulation Data

Our data are a large collection of weather snapshots at 6-h intervals from a climate simulation run. Each file contains 4 time steps, giving one day of simulation data from around the world. There are 15 weather-related variables in each time step. The data are considered ultra-large, because of the dimensions of the variables and the size of each file; each file is a simulation of one day and has a size of about 2 GB. We had available 1 year, 1 month and 14 days of simulated data, which is nearly 1TB in size. The simulation data is divided into many one-day simulation files using the Network Common Data Form (NetCDF) format [2]. Given that the data are ultra-large and multidimensional we used the lens cluster [3] to process the data in parallel.

We use VisIt software to analyze and visualize the simulation data while searching for different climate events, for example hurricanes. However, VisIt currently lacks the ability to do local spatial smoothing of the data. The availability of local spatial smoothing gives access to computing local extremes and the ability to extract local spatial information. Differencing and thresholding can then be used by VisIt to extract various features [4].

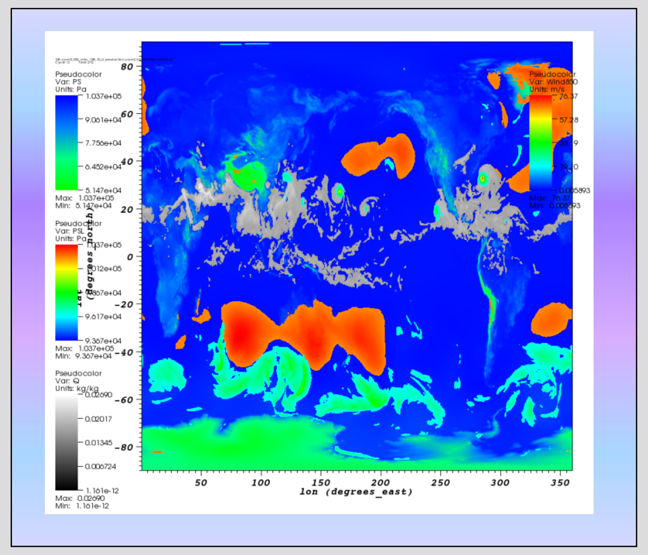

Figure 1: Features extracted by VisIt using thresholds [4] based on data preprocessed by this project.

Preprocessing

In this project, we developed a C code to do local smoothing of NetCDF format spatial time series data. The availability of local spatial smoothing gave us access to local extremes with standard VisIt tools such as thresholding. Taking the difference between the original and a spatially smoothed version results in a "local" definition of extremes. The scale of the extremes depends on the scale of the smoothing (bandwidth). For a given bandwidth, our code reads a variable from a NetCDF input data file and calculates the bandwidth average for each position. There are no boundary effects in this calculation, since the data wraps around the globe. Our code takes this into account. After calculating the local average it opens the file and writes the new calculated variable into the file. When the process finishes the input data contains useful information that can be further analyzed with VisIt.

Visualization

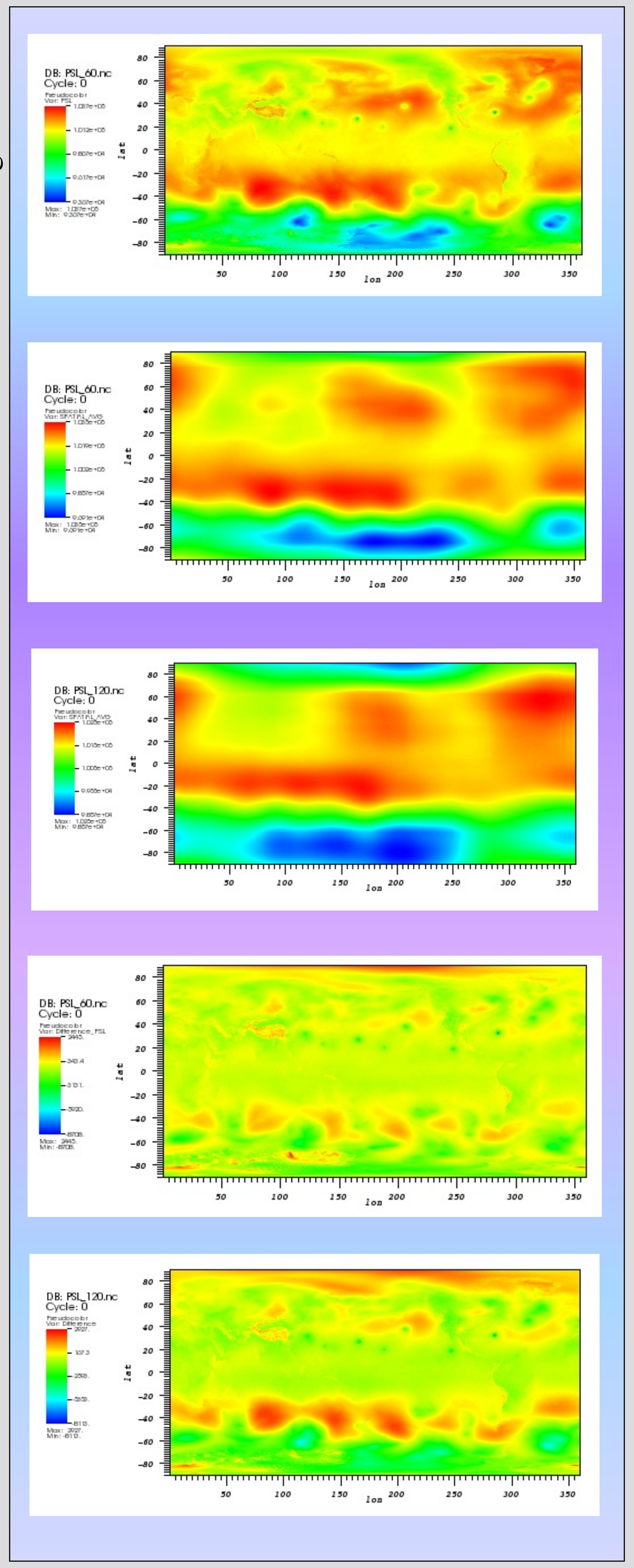

Figure 2: Sea Level Pressure (PSL) – Bandwidth of 60 makes high pressure systems more similar and extractable by thresholding.

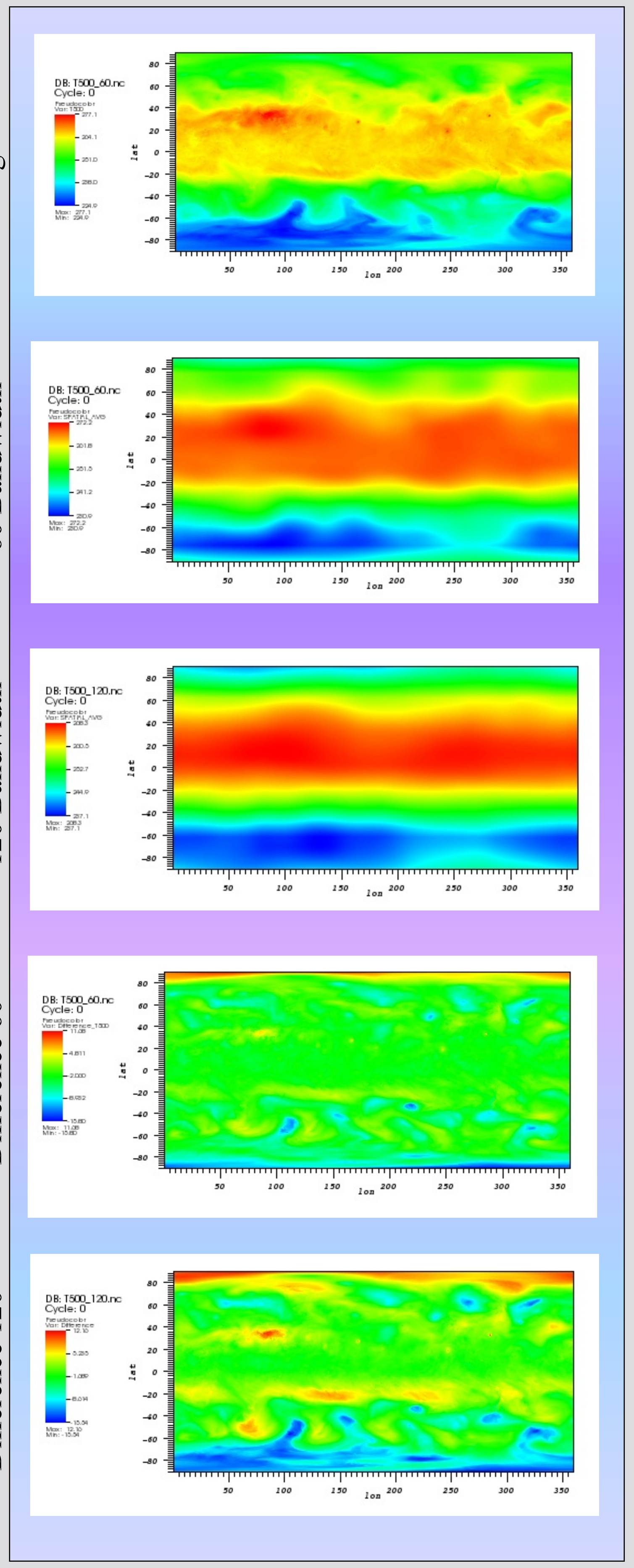

Figure 3: Temperature (T500) – Bandwidth of 120 highlights local temmperature anomalies.

Conclusion

We have developed a preprocessing tool that allows VisIt to extract local spatial features from climate simulation data. Preprocessing different variables with different bandwidth gives access to different climate features. This is the first step to bringing a local feature extraction capability to VisIt. The development of this tool required learning many new tools and concepts including the UNIX operating system, VisIt software, NetCDF file format, high-dimensional data, and spatial information. Overall, the experience was mainly educative.

Future Work

Local smoothing can be extended to the time domain into three-dimensional fields. Preprocessing of climate simulation data for local spatial features will eventually be incorporated into VisIt for real time processing in parallel. Better understanding of the simulated climate data will help identify different climate scenarios and automate the process. With appropriate data, automated identification capabilities can be built. Climate changes can be better analyzed and quantified with automated methods to identify climate change over longer time scales.

Acknowledgements

We would like to thank the Department of Energy (DOE) for providing the funds for The Faculty and Student Teams (FaST) program, Oak Ridge National Laboratory (ORNL) for providing the facilities and tools for this appointment, and Oak Ridge Institute for Science and Education (ORISE) along with Oak Ridge Associated Universities (ORAU), which made the necessary accommodations to make sure we had a pleasant experience. We thank our mentors, Dr. George Ostrouchov, Dr. Robert Sisneros from ORNL and Dr. Houssain Kettani from the Polytechnic University of Puerto Rico for providing guidance and support throughout the whole experience.

References

- [1] VisIt, data analysis and visualization software, https://wci.llnl.gov/codes/visit
- [2] NetCDF, Network Common Data Format, http://www.unidata.ucar.edu/software/netcdf
- [3] Lens, 32 node Linux cluster dedicated to data analysis and high-end visualization, http://www.olcf.ornl.gov/computing-resources/lens
- [4] Javier Colón Lazú, 2011, Extracting features from climate data using VisIt

OLCF
Oak Ridge Leadership Computing Facility

Polytechnic University Puerto Rico

ORISE

Office of Science
FaST

Oak Ridge National Laboratory