Preprocessing Climate Data for Access to Local Spatial Extremes

Yael M. Camacho Bonaparte, and Houssain Kettani,
Electrical and Computer Engineering and Computer Science Department
Polytechnic University of Puerto Rico
San Juan, PR 00919

George Ostrouchov, and Robert Sisneros
Computer Science and Mathematics Division
Oak Ridge National Laboratory
Oak Ridge, TN 37831

Abstract

Since the beginning of time, humans have been interested in the weather, to protect life, property, or plan everyday activities. Over the past decade, climate simulation has become substantially more accurate due to extensive use of sophisticated mathematical models, the availability of space-based remote sensing tools, and the exponential increases in the computing power afforded by state-of-the-art supercomputers. To better understand weather, we must analyze climate simulation output for continuous and sustained changes in weather patterns, referred to as climate changes. The VisIt [1] data analysis and visualization software package provides complex capabilities for such data analysis; however it currently lacks the ability to do local smoothing of input data. The availability of local spatial smoothing gives access to computing local extremes (as opposed to global extremes) and the ability to extract local spatial features. The focus of our research is on creating a stand-alone preprocessing code, written in a programming language called C, to calculate local spatial information; in our case we focused on getting local averages of a variable. The multidimensional simulated climate data contains 15 weather-related variables from around the world, which were simulated at 6-h intervals. 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 ~2GB. Our code reads data from each file, calculates the spatial average, and then adds the spatial average as a new variable to the data file, thus making it available for further analysis. Given that the data files are ultra-large, we used a supercomputer called Lens, for faster processing. Lens is a 32 node Linux cluster dedicated to data analysis and high-end visualizations, quite suitable for running our code with the simulated climate data. The main result of this research is a preprocessing code that adds useful spatial information to the simulated climate data, for further analysis.

Key Words: Climate, Simulation, Preprocessing

.

I. Introduction

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 as: 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, for further processing of the data.

II. Simulated Data

Our data are 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 2GB. 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) [2] format. Given that the data are ultra-large and multidimensional we used the lens cluster [3] to process the data in parallel.

We used 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 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.

III. 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 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.

IV. Visualization

After preprocessing the data, we used VisIt to get the difference between the original data and the local average. In this case, we are plotting the original Sea Level Pressure (PSL) variable along with 60 and 120 bandwidth and their respective differences. Bandwidth of 60 makes high pressure systems more similar and extractable by thresholding using VisIt.

.

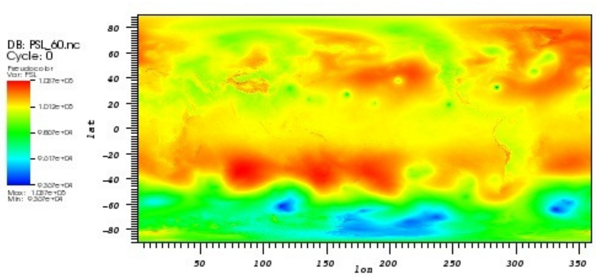

Figure 1 - Original

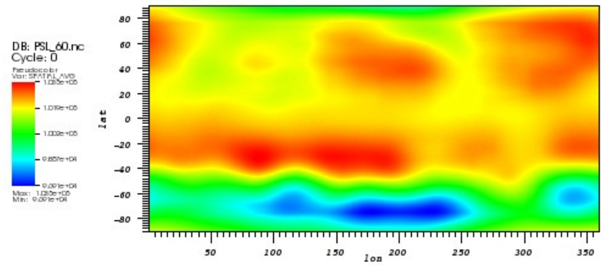

Figure 2 - 60 Bandwidth

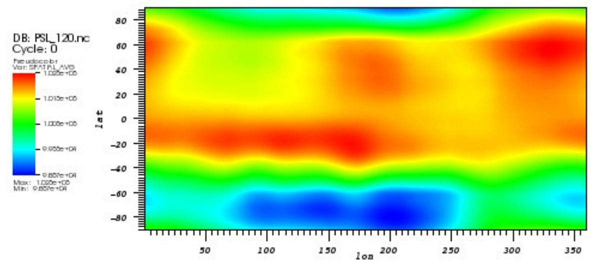

Figure 3 - 120 Bandwidth

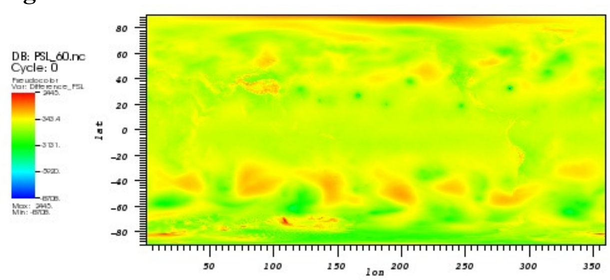

Figure 4 - Difference 60

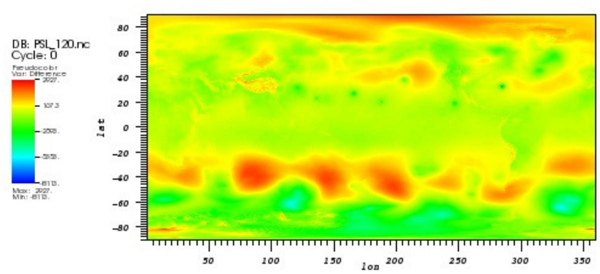

Figure 5 - Difference 120

V. 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.

VI. 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.

VII. Acknowledgements

This work was funded in part by the Department of Energy (DOE) under the Faculty and Student Teams (FaST) program at Oak Ridge National Laboratory (ORNL). It was also funded in part by the Department of Defense and the Extreme Scale Software Center at the Oak Ridge National Laboratory.

VIII. 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/