Porting a parallel program from the NASA Center for Climate Simulation (NCCS) Discoverer supercomputer to desktops for validation of the Multi-sensor Aerosol Products Sampling System (MAPSS)

Southwestern Oklahoma State University

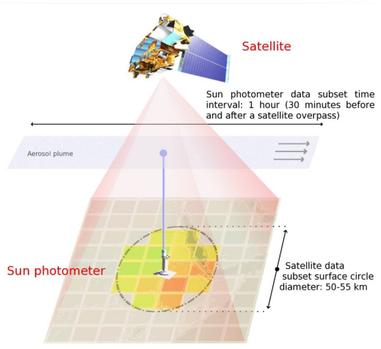
Charles Sleeper | Jack Guillory | Dr. Jeremy Evert | Department of Computer Science

Abstract

The National Aeronautics and Space Administration (NASA) produces nearly two gigabytes of data per second. NASA researchers leverage millions of dollars of computing hardware to analyze this data. NASA shares this data with the rest of the world. Advances in computer technology have provided modern desktop computers more powerful than the fastest supercomputers in the world from two and three decades ago. This provides many possibilities for greater use of NASA data. A lack of education materials for undergraduate research in high performance computing limits these possibilities. This research addresses this need by presenting the methodologies used to translate the NASA MAPSS software system from supercomputers and software engineers to desktops and undergrads. Undergraduate student researchers studied the MAPSS software system, created for the NASA Goddard Space Flight Center supercomputers, to conduct a validation study of NASA Earth Atmosphere aerosol data. Undergrads rewrote parts of the software allowing it to run on an Intel i7 processor running a Linux system. The students completed translation of four of the seven satellite sensors, and developed automation software allowing MAPSS to be portable between individual computers. The students provided documentation of this process allowing future students to complete the translation of the remaining sensor systems and the validation study. This should provide greater use of the data that streams from NASA every day.

Atmospheric Aerosol Data Collection

- Space-based sensor took measurements looking down to Earth, while ground-based sensors took measurements looking up. Statistics for each of the aerosol products involved extracting measurements that fall within 27.5 km of the chosen locations, and within 30 minutes of each satellite fly over.



Research Questions

Do NASA satellites sensor platforms produce similar readings to ground-based sensors?

Can NASA research be verified by undergraduate researchers?

What level of Cyber infrastructure resources are necessary to examine a data set of this size?

.

Timeline

- May 2016: SWOSU students received an Oklahoma NASA EPSCoR Research Initiation Grant. Students formed a summer research team. The students selected LaTeX for documentation of all publications, Github as their knowledge management tool. All students focused on their areas of expertise. Some began a literature review of the NASA MAPSS project. Other students developed a Python translation of the functionality of the MAPSS software. SWOSU student Jack Guillory took the lead on translating the MAPSS source code to run on a Linux distribution and standard desktop computer. Charles Sleeper began developing a Raspberry Pi Cluster to support future research partnerships with area High Schools.
- May 2016: The SWOSU HPC team participated in the first Oklahoma HPC competition. Students deployed a Raspberry Pi cluster. Students also developed code to find the determinant of a five thousand by five thousand matrix.
- June 2016: SWOSU students attended the Extreme Science and Engineering Discovery Environment (XSEDE) High Performance Computing (HPC) Workshop: Summer Boot Camp. This was in support of developing and optimizing code for HPC applications. Students were introduced to parallel computing concepts, openMP, the Message Passing Interface (MPI), CUDA, and openACC. Students toured the OU Supercomputing Center for Education & Research (OSCER), including their $2M Supercomputer. During the boot camp, students ran code on the Pittsburg Supercomputing Center Bridges $10M supercomputer, ran by Carnegie Mellon University.
- June 2016: SWOSU students began working through the NASA MAPSS software. The software continues to be in use after two decades of development. The team of software engineers at the NASA Goddard Space Flight Center had tuned the software for optimal performance on several supercomputers, including the Discoverer 12. Jack Guillory was able to deliver functionality with four of the nine modules within the software.
- August 2016: SWOSU students participated in a NSF CC* grant proposal. The SWOSU team coordinated a Cyber Infrastructure plan for three universities. During the fall 2017 Oklahoma Supercomputing Symposium, this plan was praised by members of the One OCII team as being one of the better written plans the state had seen in recent years.
- September 2016: SWOSU meets with Southeastern Oklahoma State University research faculty to begin collaboration on future NASA OK EPSoR Grants.
- December 2016: Oklahoma State University provided SWOSU with eight servers from a retired cluster. This cluster was deployed using centOS with the SLURM job scheduler. Students installed the operating system, set up the administrator accounts, and worked with university IT staff to enable access to the server from various classrooms.
- February 2017: The SWOSU HPC team reached out to Dr. Toshihisu Matsui for a copy of the NU-WRF model. The SWOSU COMSC 3133 Software Engineering students will begin analysis of the software. This software was used as one of the three components of the Student Cluster Competition at The 2016 International Conference for High Performance Computing, Networking, Storage and Analysis (SC16)

Research Challenges

- The MAPSS software contained hundreds of source files written in several programming languages over two decades by multiple authors.
- Some files for the software are automatically generated by the NASA supercomputer are not automatically generated within a standard installation of Linux Mint. These files were manually recreated from error messages

Results

- **Tool currently functions with:**
  - Cloud-Aerosol Lindar and Infrared Pathfinder Satellite Observation (CALIPSO)
  - Aerosol Robotic Network (Aeronet)
  - Multi-angle Imaging SpectroRadiometer (MISR)
  - Ozone Monitoring Instrument (OMI)
- **In progress**
  - Polarization and Directionality of the Earth's Reflectance's (Polder)
  - Visible Infrared Imaging Radiometer Suite (VIIRS)
  - Moderate Resolution Imaging Spectro radiometer (MODIS)
- **Next steps**
  - Run analysis on SWOSU Intel Cluster Computers and publish results
  - Run analysis on SWOSU Raspberry Pi cluster computers and expand analysis: Larger sensor radius, different sensor fusion models.

References

- Ichoku, C., D. A. Chu, S. Mattoo, Y. J. Kaufman, L. A. Remer, D. Tanre, I. Slutsker, and B. N. Holben, 2002: A spatio-temporal approach for global validation and analysis of MODIS aerosol products. *Geophys. Res. Lett.*, **29**, doi: 10.1029/2001GL013206.
- Chu, D. A., Y. J. Kaufman, C. Ichoku, L. A. Remer, D. Tanre, and B. N. Holben, 2002: Validation of MODIS aerosol optical depth retrieval over land. *Geophys. Res. Lett.*, **29**, doi: 10.1029/2001GLO13205.
- Remer, L. A., D. Tanre, Y. J. Kaufman, C. Ichoku, S. Mattoo, R. Levy, D. A. Chu, B. N. Holben, O. Dubovik, A. Smirnov, J. V. Martins, R.-R. Li, and Z. Ahmad, 2002: Validation of MODIS aerosol retrieval over ocean. *Geophys Res. Lett.*, **29**, doi: 10.1029/2001GL013204.