PETASCALE PROCESSING OF SATELLITE EARTH OBSERVATIONS

Allocation: Blue Waters Professor/187 Knh
PI: Larry Di Girolamo[1]
Collaborators: Matias Carrasco Kind[2], Gregory Daues[2], Yulan Hong[1], Ralph Kahn[3], Lusheng Liang[4], Donald Petravick[2], John Towns[2], Kent Yang[5], Ping Yang[6], Yizhe Zhan[1], Guangyu Zhao[1]

[1]University of Illinois at Urbana–Champaign
[2]National Center for Supercomputing Applications
[3]NASA Goddard Space Flight Center
[4]NASA Langley Research Center/Science Systems and Applications, Inc.
[5]The HDF Group
[6]Texas A&M University

EXECUTIVE SUMMARY

Through collaborative efforts among NASA, HDF Group, and NCSA, the research team fully validated and enhanced the open source tool that resamples Terra satellite data into a common grid with a python interface to make it applicable beyond Terra to any Earth observation data. The Terra satellite, launched in 1999, continues to collect earth science data using five instruments: the Moderate-resolution Imaging Spectroradiometer (MODIS), the Multi-angle Imaging SpectroRadiometer (MISR), the Advanced Spaceborne Thermal Emission and Reflection Radiometer, the Clouds and the Earth's Radiant Energy System (CERES), and the Measurements of Pollution in the Troposphere. Terra data not only serve the scientific community but also the governmental, commercial, and educational communities.

The researchers further used the Terra data set to: (1) characterize the ice crystal roughness of cirrus clouds for a better understanding of ice cloud optical properties; (2) correct biases in the MODIS cloud effective radius retrievals through MISR and MODIS fusion; (3) evaluate MISR and CERES Arctic cloud albedo retrievals, indicating excellent consistency for certain solar zenith angles; and, (4) examine decadal changes in the Earth's radiance fields, revealing little cloud change in the global mean beyond calibration (Fig. 1) but large changes in certain regions. The research group extended the study by processing observations from spaceborne active sensors, showing a strong correlation between cloud occurrence frequency and climate regimes (Fig. 2).

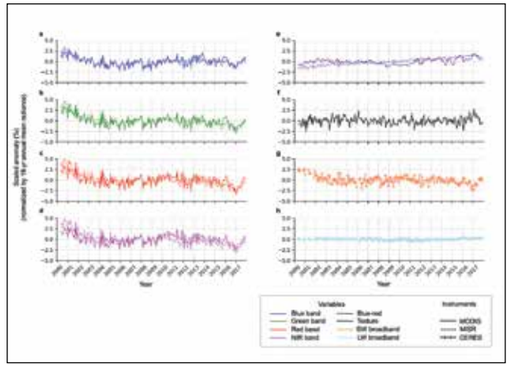

Figure 1: Time series of the deseasonalized and scaled anomalies of the monthly global mean spectral and broadband radiance and texture for MISR, MODIS, and CERES between 2000 and 2017.

RESEARCH CHALLENGE

The need for Terra data fusion and for scientists to perform large-scale analytics with long records has never been greater [1], given the growing data volume (greater than 1 petabyte), the storage of different instrument data at different NASA centers, the different data file formats and projections, and inadequate cyberinfrastructure [2]. The team initiated the Terra Data Fusion Project to tackle two long-standing problems: (1) the need for efficient generation and delivery of Terra data fusion products; and (2) the need to facilitate the use of Terra data fusion products by the community by generating new products and knowledge through national computing facilities, and to disseminate them through national data-sharing services. Solutions to these problems will significantly facilitate discovery and accelerate progress in earth science research.

METHODS & CODES

Key steps in the Terra Data Fusion Project include: (1) transferring the entire Terra record to Blue Waters from NASA centers (Level 1B radiance; greater than 1 petabyte); (2) building software optimized for whole-mission processing on Blue Waters to create basic fusion products; (3) optimizing data granularity and HDF user-defined settings that best support parallel I/O on Blue Waters; (4) archiving and distributing Terra fusion products through existing NASA services and commercial cloud services; and (5) driving development through processing and analysis of scientifically important use cases. Past Blue Waters reports have highlighted results on many of these key points. The research group has since extended their studies beyond Terra to include other satellite records and has developed an open source tool to resample or reproject the radiance fields into a common grid.

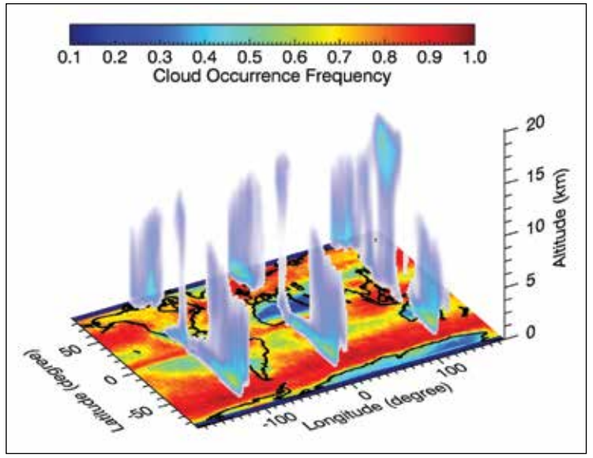

Figure 2: A 3D view of cloud occurrence frequency derived from the CloudSat/CALIPSO 2B–CLDCLASS–LIDAR product from 2007 to 2010.

RESULTS & IMPACT

The research team explored archiving and distributing Terra fusion products through existing NASA and commercial cloud services. They tested accessing and processing the basic fusion data on the Amazon cloud and found its processing time was slightly more than Blue Waters for the same settings but that costs to investigators' grants by using commercial clouds for a data-intensive project can be prohibitive. The team also fully validated and enhanced the open source tool that resamples or reprojects the radiance fields into a common grid with a Python interface to make the tool applicable beyond Terra to any Earth observation data.

Scientific investigation using the fusion data set was carried out primarily in four study areas. (1) The team characterized biases in the MODIS standard product of cloud drop effective radius (Re) and further examined the underlying causes. The results paint a radically different picture of the distributions of Re from what the original MODIS product provided. The bias-corrected cloud drop sizes are now in line with spot measurements from field campaigns and outputs from climate models. (2) The team retrieved ice cloud microphysical properties through MISR–MODIS fusion, showing regional dependence on ice crystal structure. Results have been published in an open access journal, *Remote Sensing*. (3) The group evaluated the MISR and CERES Arctic cloud albedo retrievals, showing excellent consistency between the instruments for solar zenith angles of less than 70°. These results were also published in *Remote Sensing*. (4) The team extended a global and regional radiance and texture trend study previously performed on Blue Waters [3] by using additional instruments: MODIS and CERES. The time series of the deseasonalized scaled anomalies of the monthly global mean spectral radiance and texture are shown in Fig. 1. The global blueing shown in the researchers' previous study [3] mainly results from a calibration drift in the MISR red band. There is no significant global trend in the Earth's reflected radiation field, but large regional changes were found.

The research team also analyzed cloud data from two spaceborne active sensors, CloudSat and CALIPSO, and found that large cloud occurrence frequencies are closely related to climate regimes at both horizontal and vertical scales (Fig. 2). A large frequency of high-level clouds at altitude ranges of 9 to 17 km in the tropics indicates the dominance of cirrus clouds, whereas shallow clouds with heights below 3 km are mostly observed in the subtropics. In the midlatitudes, clouds mostly occur below 5 km, although some clouds can reach to the tropopause (approximately 8 to 10 km).

WHY BLUE WATERS

The key advantages of using Blue Waters for access, usage, and distribution of Terra fusion products are that the Terra data and processing are local, whereas access and sharing are global. The research team demonstrated that having the Terra data local, with processing tuned to a massively parallel system with excellent sharing services on Blue Waters, provides an optimum framework for large-scale processing, analytics, and mining of the entire Terra record. In addition, the project staff provide expertise critically needed to optimize workflows.

PUBLICATIONS & DATA SETS

Y. Wang, S. Hioki, P. Yang, M. King, L. Di Girolamo, D. Fu, and B. A. Baum, "Inference of an optimal ice particle model through latitudinal analysis of MISR and MODIS data," *Remote Sens.*, vol. 10, no. 12, p. 1981, 2018, doi: 10.3390/rs10121981.

Y. Zhan, L. Di Girolamo, R. Davies, and C. Moroney, "Instantaneous top-of-atmosphere albedo comparison between CERES and MISR over the Arctic," *Remote Sens.*, vol. 10, no. 12, p. 1882, 2018, doi: 10.3390/rs10121882.