THE TERRA DATA FUSION PROJECT

Allocation: Blue Waters Professor/180 Knh
PI: Larry Di Girolamo[1]
Collaborators: Guangyu Zhao[1], John Towns[2], Shaowen Wang[2], Yan Liu[2], Kent Yang[3]

[1]University of Illinois at Urbana-Champaign
[2]National Center for Supercomputing Applications
[3]The HDF Group

EXECUTIVE SUMMARY

The Terra Data Fusion Project that we initiated in 2015 continues to make significant progress through collaborative efforts among NASA, HDF Group, and NCSA. Large Terra data transfers (~1PB) involving the Blue Waters team and NASA are now complete. The software tool that fuses all the Terra radiance granules was developed and has produced one year of basic fusion data on Blue Waters. Besides the scientific usage demonstrated in the previous reports, the dataset has been further used to (1) characterize ice crystal roughness of cirrus clouds, resulting in a better understanding of ice cloud optical properties, and (2) gauge the effects of cloud heterogeneity on microphysical retrievals, concluding that cloud spatial heterogeneity is insufficient to explain the large biases in the MODIS *Re* standard product.

RESEARCH CHALLENGE

The Terra satellite was launched in 1999 and continues to collect Earth sciences data using five instruments: the Moderate-resolution Imaging Spectroradiometer (MODIS), the Multi-angle Imaging SpectroRadiometer (MISR), the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER), the Clouds and Earth's Radiant Energy System (CERES), and the Measurements of Pollution in the Troposphere (MOPITT). Terra data is among the most popular of NASA's datasets, serving not only the scientific community but also governmental, commercial, and educational communities.

The need for data fusion and for scientists to perform large-scale analytics with long records has never been greater [1]. The challenge is particularly acute for Terra, given its growing data volume (>1 petabyte), the storage of different instrument data at different NASA centers, the different data file formats and projection, and inadequate cyberinfrastructure [2]. We recently initiated the Terra Data Fusion Project to tackle two long-standing problems: (1) How do we efficiently generate and deliver Terra data fusion products; and (2) How do we facilitate the use of Terra data fusion products by the community in generating new products and knowledge through national computing facilities, and disseminate these new products and knowledge through national data sharing services?

Solutions to these questions will: 1) facilitate greater ease in creating new geophysical retrieval algorithms that provide greater accuracy than the current single-instrument algorithms, 2) provide an easy mechanism for users to access and process the entire Terra record, (3) greatly reduce error and redundancy in the science community among those researchers using multiple instrument datasets, (4) provide greater insight into geophysical processes through synergistic use of fusion products, and (5) provide a framework for fusion that could extend to other NASA missions and constellations. The end result will facilitate discovery and accelerate progress in Earth Science research. Use cases are presented below.

METHODS & CODES

Key steps in the Terra Data Fusion Project include: (1) transferring the entire Terra record (Level 1B radiance; >1 petabyte) to Blue Waters from NASA centers, (2) building software optimized for whole-mission processing on Blue Waters to create basic fusion products; (3) optimizing data granularity and HDF API settings that best support parallel I/O on Blue Waters; and (4) archiving and distributing Terra fusion products through existing NASA services.

Thus far, key steps (1) and (2) are accomplished. We successfully transferred mission-scale radiance data from all the five Terra instruments to Blue Waters. We built a software tool to merge all the Terra radiance granules into one basic fusion granule, which contains not only radiance measurements but also their uncertainties, geolocation, sun-view geometry, and observational time. The tool has produced one year of basic fusion data on the Blue Waters. We are currently working on key step (3) to prepare for mission-scale processing with help from the Blue Waters team. The overview of this project and progress report was given at the MISR science team meeting and data users' science symposium [3].

Initial science investigation using this Terra fusion dataset was carried out in two studies. Fusing MISR and MODIS products, we characterized the effective radius (*Re*) of the liquid cloud drop size distribution over oceans. Details were given in *Blue Waters 2015 Annual Report* and in [4]. Our results paint a radically different picture of the distributions of cloud drop sizes in our atmosphere compared to what was previously determined from the original MODIS data; e.g., 2 to 11 μm differences in the zonal means. We also processed the 15-year MISR data to examine the decadal trends in the Earth's color and texture [5]. Globally, we show that the Earth has been appearing relatively more blue (up to 1.6% per decade from both nadir and oblique views) and smoother (up to 1.5% per decade, only from oblique views) over the past 15 years.

RESULTS & IMPACT

We continue to explore the scientific usage of this dataset in two new studies. The first was that we used the dataset to generate one full year (2013) of MISR+MODIS cloud-element fusion data (~50 terabytes) and share the data with Prof. Ping Yang's group at the Texas A&M University via GridFTP. The fusion data were used to build ice crystal roughness parameters that are important to our meteorological understanding of ice clouds. Our intermediate results were presented at recent meetings and conferences [6–7]. Further scientific analyses are being conducted.

The second was that we examined the effect of cloud spatial heterogeneity on satellite-retrieved liquid cloud *Re* using 13 years of MISR and MODIS fusion for the month of January. We did find some dependency of *Re* on cloud spatial heterogeneity. Further stratification of the biases by both cloud spatial heterogeneity and cloud optical depth for various solar zenith angles and geolocations reveals that more heterogeneous and optically thick clouds tend to have larger *Re* bias, indicating cloud spatial heterogeneity is insufficient to explain the large biases in the MODIS *Re* standard product. Our results were presented at the MISR data users' science symposium [8].

WHY BLUE WATERS

Key advantages of using Blue Waters for access, usage, and distribution of Terra fusion products are that the Terra data and processing are local, with access and sharing that are global. It has been demonstrated that having the Terra data local, with processing tuned to a massively parallel system with excellent sharing services, in one of the largest storage and bandwidth computing facilities in the country, provides an optimum framework for large-scale processing, analytics, and mining of the entire Terra record. In addition, the project staff provides expertise critically needed to optimize workflows.

Terra is just one earth science data set. Fusion with other instrument records and meteorological reanalysis data (all of which are growing exponentially) for advancing earth science requires—and will continue to require—a Track-1 system that is accessible by the community.

PUBLICATIONS AND DATA SETS

Liang, L., L. Di Girolamo, and W. Sun, Bias in MODIS cloud drop effective radius for oceanic water clouds as deduced from optical thickness variability across scattering angles. *J. Geophys. Res. Atmos.*, 120:15 (2015), DOI: 10.1002/2015JD023256. [The main data set applied to this study is the fusion product generated by fusing the MISR Level1B radiance product (Version F03_0024) and the MODIS level2 cloud product (Collection 6).]

Zhao, G., et al., Regional changes in Earth's color and texture as observed from space over a 15-year period. *IEEE Trans. Geosci. Remote Sens*, 54:7 (2016), DOI: 10.1109/TGRS.2016.2538723. [The main data set applied to this study is the MISR level1B radiance product (Version F03_0024).]