THE TERRA DATA FUSION PROJECT

Allocation: Blue Waters Professor/210 Knh
PI: Larry Di Girolamo[1]
Co-PIs: John Towns[2], Shaowen Wang[1], Guangyu Zhao[1], Kent Yang[3], Yan Liu[1]

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

EXECUTIVE SUMMARY

We have led the Terra Data Fusion Project to success through collaborative efforts among NASA, the HDF Group, and NCSA. We generated and validated missionwide processed and calibrated (Level 1B [1]) fusion products (2.4 PB) on Blue Waters, which provides the necessary stepping stone for developing higher-level products and provides the framework for other flavors of fusion. For these data, we built an open-source tool that resamples Terra satellite data into a common grid adopted by any instrument or defined by any scalable map projection. The fusion data set has been further used to: (1) characterize ice crystal roughness of cirrus clouds, resulting in a better understanding of global ice cloud optical properties; (2) quantify regional biases in the retrieved cloud drop sizes of liquid water clouds induced from cloud heterogeneity; and (3) examine decadal changes in the Earth's radiance fields, revealing temporal and spatial variability at multiple scales.

RESEARCH CHALLENGE

The Terra satellite was launched in 1999 and 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 (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 data sets, 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 [2]. 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 projections, and inadequate cyberinfrastructure [3]. We recently initiated the Terra Data Fusion Project to tackle two long-standing problems: (1) How to efficiently generate and deliver Terra data fusion products; and (2) How to 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 problems 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) reduce error and redundancy in the science community among those researchers using multiple instrument data sets; (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, we have accomplished key steps (1), (2), and (3), and we are working with NASA on step (4).

We also developed an open source tool to resample or reproject the radiance fields stored in the Terra basic fusion product into a common grid adopted either by a Terra instrument or defined by any map projections. We built two additional tools specifically to facilitate large data processing on Blue Waters: a Globus Python interface data transfer tool, which interacts with the Globus CLI, and a Python PBS workflow manager, which is a lightweight module with a friendly interface that provides a programmatic way to define complex job chains.

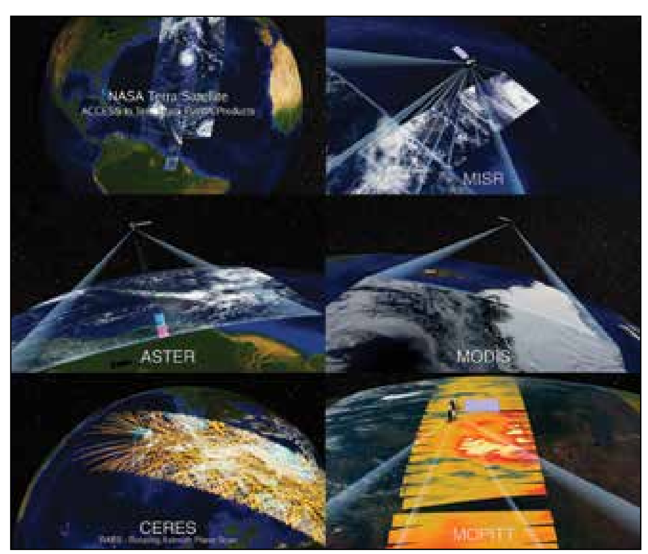

Figure 1: Image frames captured from an animation clip that dynamically displays and projects the radiance imageries generated from one single Terra basic fusion granule (Orbit 3671) onto Earth as being orbited by Terra for all of the five Terra instruments. See it on YouTube at *https://youtu.be/C2uyjRGwwOs*.

RESULTS & IMPACT

We successfully transferred mission-scale radiance data from all five Terra instruments to Blue Waters. We built a software tool to merge all Terra radiance granules into one Basic Fusion (BF) granule, which contains not only radiance measurements but also their uncertainties, geolocation, sun-view geometry, and observational time. The tool has produced missionwide basic fusion data (2.4 PB) on Blue Waters. We are currently working with NASA to transfer the entire data set to the Earthdata Cloud. We also produced missionwide metadata for the basic fusion product, which will be ingested into the NASA Common Metadata Repository such that the basic fusion product can be discovered and searched publicly. The overview of this project and progress report was given at the American Geophysical Union 2017 Fall Meeting [4].

In collaboration with the NCSA Advanced Visualization Lab, we developed a visualization tool that dynamically displays and projects the radiance imageries generated from one single BF granule onto a 3D Earth as being orbited by Terra for all of the five Terra instruments. We created and posted an animation clip on YouTube, with image frames from the clip given in Fig. 1. To the best of our knowledge, this is the first animation to simultaneously show five RGB imageries of all five Terra instruments on the same orbit. The tool not only helps us to validate and explore the BF data but will also have a profound educational influence among the broader scientific community.

We carried out science investigations using this Terra fusion data set primarily in two studies in the past year. We fused MISR and MODIS data to characterize biases in retrieved cloud drop effective radius inherent to MODIS-alone retrievals and further examined the underlying causes of the biases. 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. Our corrected cloud drop sizes are now in line with spot measurements from field campaigns. In the second study, we have been working closely with Prof. Ping Yang at Texas A&M University on a specialized MISR and MODIS fusion data set designed for retrieving ice cloud microphysical properties, including ice crystal roughness. Early results from his group were presented at several meeting and conference venues, showing an altitude (hence temperature) and regional dependence on ice crystal structure.

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 workflow.