



ENABLING EARTH SCIENCE: A fused dataset from the Terra Satellite on the Open Storage Network

Terra, the flagship of NASA's Earth Observing System (EOS), was launched into orbit on December 18, 1999 and carries 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).

The Terra Basic Fusion dataset hosted on the Open Storage Network (OSN) is an easy-to-access record of the Level 1 radiance (in physical units) and geolocation, sun-view geometry, and observation time for each IFOV at its native resolution for all the instruments on the v mission for selected areas (described below) covering the years 2000-2015. The granularity of the data is one orbit and is stored in HDF5, conforming to CF conventions and accessible by netCDF-4 models. The dataset is designed to accelerate investigations addressing questions such as:

- How is the global Earth system changing?
- What causes these changes in the Earth system?
- How will the Earth system change in the future?
- How does Earth system science provide benefit society?

The Algorithm Theoretical Basis Document is published at <https://go.illinois.edu/Terra-fusion-doc>.

Distribution via the Open Storage Network

Each Terra fusion file contains one orbit of data from all five Terra instruments. The WRS-2 paths currently stored on the OSN for the first 15 years of the Terra mission are as follow:

- Popular: 233, 125, 143, 108 (Greenland, China, India, Japan,)
- Useful for Calibration: 188, 150 (Nigeria, Antarctica)
- Bermuda: 11, 12, 13, 14
- Additional China; 123, 124, 126
- Additional India: 144, 145, 146, 147

Paths can be visualised using this tool: https://10dup05.larc.nasa.gov/MISR_BROWSE/path, which has output as below:

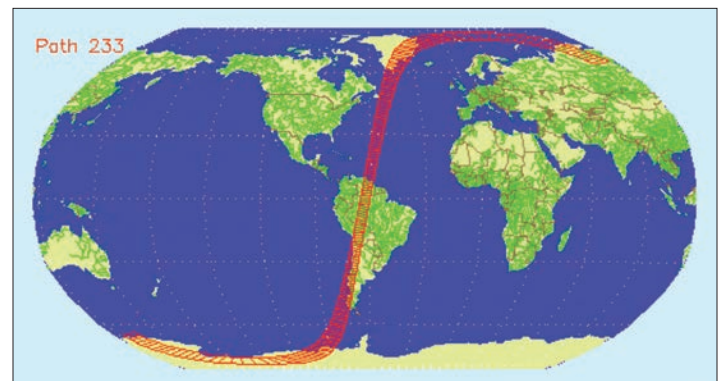
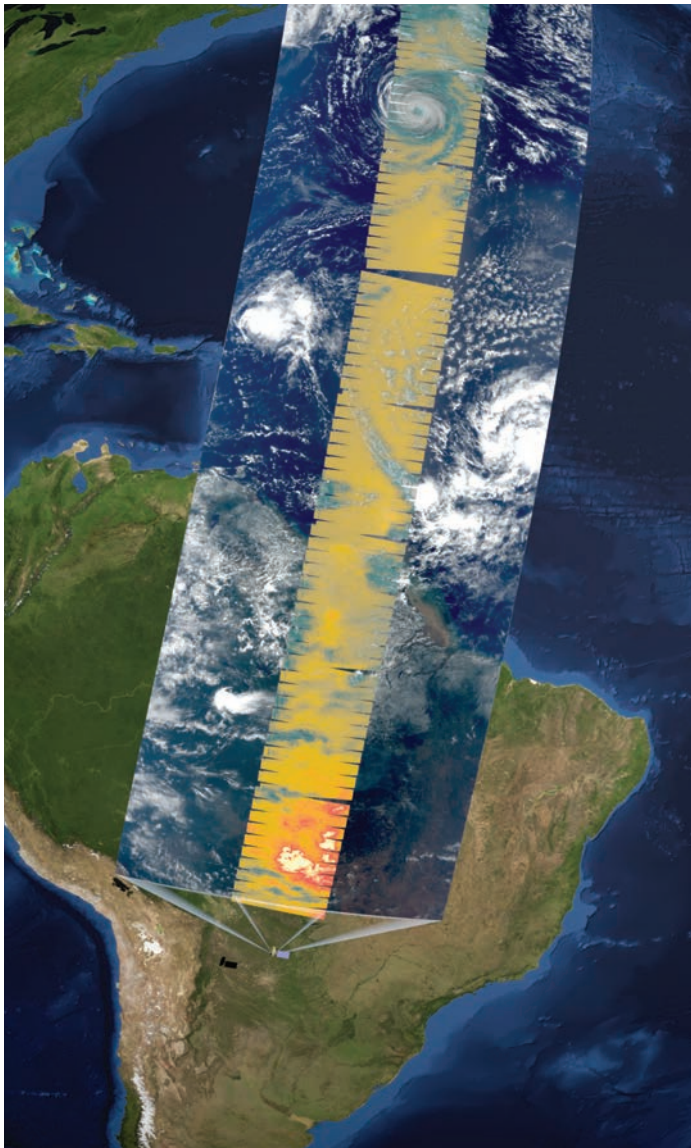


Figure - Path 233 includes Greenland

The record of Terra observations for one path is a collection of objects, each 15-40 GB in size. The total data volume for a path is about 10.5 TB over the entire mission. Each object is the information from one orbit of the Terra satellite, and contains all the instrumental information from that orbit. This data set is a Level 1B dataset, meaning data have been processed into sensor units. Object names contain the path number referred to above. Each object is in the format of an HDF5 file. Once downloaded, an HDF5 library is required to access the data.



Downloading the Data

The paths are hosted on the Open Storage Network (OSN). The OSN aims to provide a cyberinfrastructure service addressing specific data storage, transfer, sharing, and access challenges in scientific communities, thereby enabling and enhancing data-driven research collaborations across universities. The OSN seeks to make datasets widely available as a means to maximizing existing science investments. The OSN presents the Terra FUSION data for downloads from <https://ncsa.osn.xsede.org>. Unlike commercially hosted data, the OSN never charges for egress. Access is by standard web tools, including wget and curl. Even though the OSN is not part of Amazon, the Amazon s3 command line tools (awscli) can be used for inspection and download of data.

Example Usage

Identify Terra fusion contents present on the OSN:

```
aws s3 ls s3://TFusion/ --endpoint-url https://ncsa.osn.xsede.org --no-sign-request
```

Copy one specific orbit on path 124 (China) to your local machine:

```
aws s3 cp s3://TFusion/P124/TERRA_BF_L1B_09506_20011001024622_F000_V001.h5 . --endpoint-url https://ncsa.osn.xsede.org --no-sign-request
```

An advanced Fusion package, available on Github <https://github.com/TerraFusion>, forms an additional body of software which provides functionality such as registering the data from the various instruments to a common grid.

Agency and Foundation Support

The Terra Fusion work was supported from the NASA ACCESS program under contract NNX16AMO7A. The project was lead by Larry Di Girolamo, Professor of Atmospheric Sciences at the University of Illinois.

Distribution of the Data Sets described here is provided by The Open Storage Network (OSN), funded by NSF (awards 1747552, 1747493, 1747507, 1747490, 1747483) and the Schmidt Futures Foundation in 2018, is a storage substrate pilot to synergistically benefit from existing investments, and to support a range scientific and scholarly production. The initial OSN pilot is a distributed effort between JHU, SDSC at UCSD, NCSA at UIUC, RENCI at UNC-Chapel Hill and MGHPC. Project partners include PSC and TACC, the regional Big Data Innovation Hubs, and the project teams leading our several Science Use Cases.

Production of the basic fusion data sets, and scientific exploitation of the data sets is supported by resources provided by:

- The Blue Waters sustained-petascale computing project, which is supported by the National Science Foundation (awards OCI-0725070 and ACI-1238993) and the state of Illinois. Blue Waters is a joint effort of the University of Illinois at Urbana-Champaign and its National Center for Supercomputing Applications.
- The NSF-supported ROGER (Resourcing Open Geospatial Education and Research) cyberGIS supercomputer (Award#: 1429699).

