(Toward) Analysis Ready and Cloud Optimized Data Formats for NASA Airborne Facility Instruments

Michele Thornton, Scott Pearson, Rupesh Shrestha, Yaxing Wei, Chris Lindsley, Bruce E Wilson

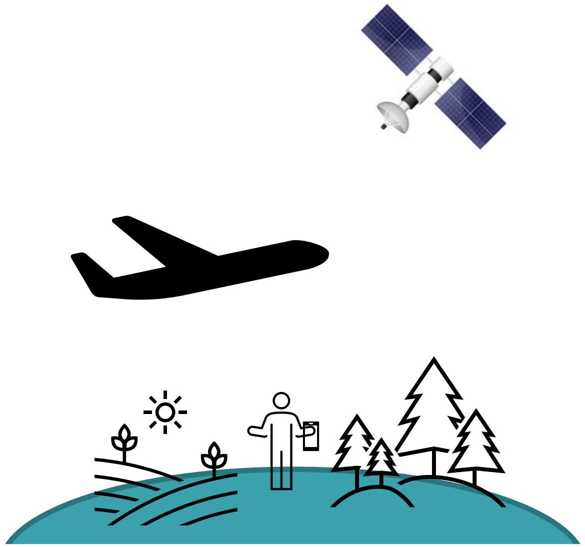
AGU Fall Meeting, 2022: IN35A

December 14, 2022

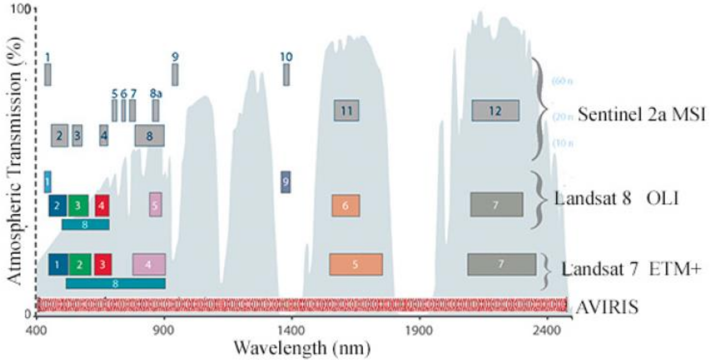
The Oak Ridge National Laboratory Distributed Active Archive Center for Biogeochemical Dynamics operates under an interagency agreement between NASA and the U.S. Department of Energy

NASA Airborne Science Program

- NASA's Airborne Science Program is responsible for providing aircraft systems that further science and advance the use of satellite data
- Primary objectives:
  - **Satellite Calibration and Validation**
  - **Support New Sensor Development**
  - **Contribute to Process Studies**
    - **High-resolution measurements of complex systems**
    - **Coupled to global satellite observations**
  - **Develop the Next-Generation of Scientists and Engineers**



ORNL DAAC: NASA Airborne Facility Instrument Data



Imaging Spectrometers

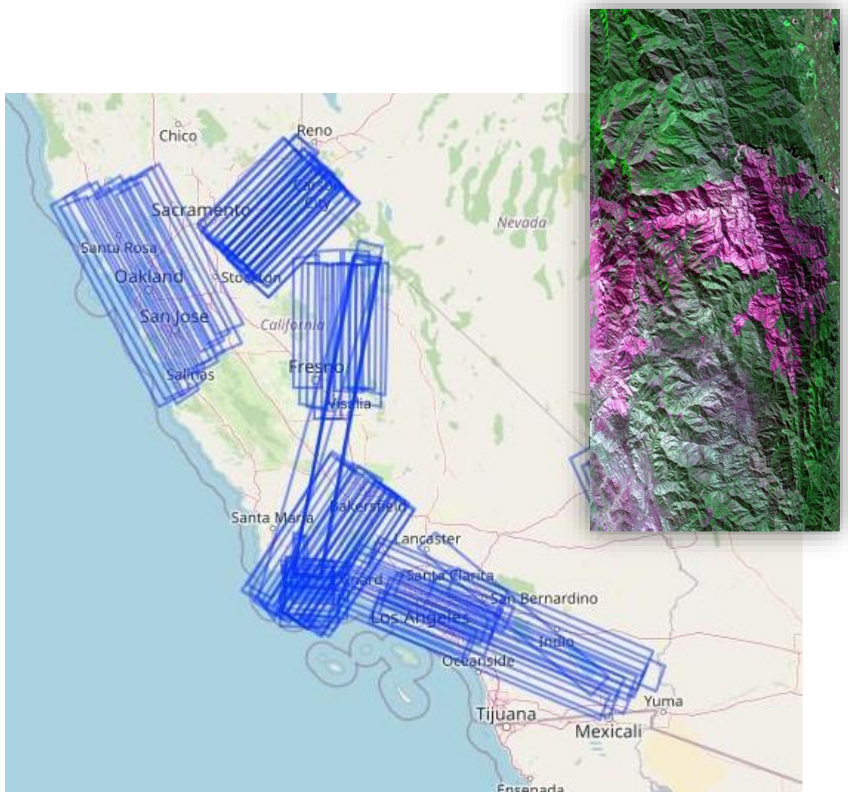
1. **MASTER** (MODIS/ASTER Airborne Simulator)
2. **AVIRIS-Classic** (Airborne Visible/Infrared Imaging Spectrometer)
3. **AVIRIS-Next Generation**

	MASTER	AVIRIS-C	AVIRIS-NG
Research Center	Airborne Sensor Facility (ASF) at NASA Ames Research Center – PI: Dr. Simon Hook	The Jet Propulsion Laboratory (JPL) – PI: Dr. Robert Green	The Jet Propulsion Laboratory (JPL) – PI: Dr. Robert Green
Spectral Range/Resolution	~ 500 – 13,000nm 50 “bands” VSWIR, mid-TIR	~ 400 – 2,500nm @ 10nm 224 “bands” VSWIR	~ 400 – 2,500nm @ 5nm 425 “bands” VSWIR
Products / Formats	Radiance (L1B) in HDF-4, Reflectance (L2) in ENVI Ancillary files	Radiance (L1B) ENVI, Reflectance (L2) ENVI Ancillary and supporting files	Radiance (L1B) ENVI, Reflectance (L2) in ENVI Ancillary and supporting files
NASA Campaigns (ORNL DAAC)	HyspIRI Prep, SARP, DFRC, WDTS, FIREX-AQ	HyspIRI Prep, SISTER	HyspIRI Prep, Delta-X, ABoVE, SHIFT, SISTER
Operational	1998 - present	1989 - present	2009 - present

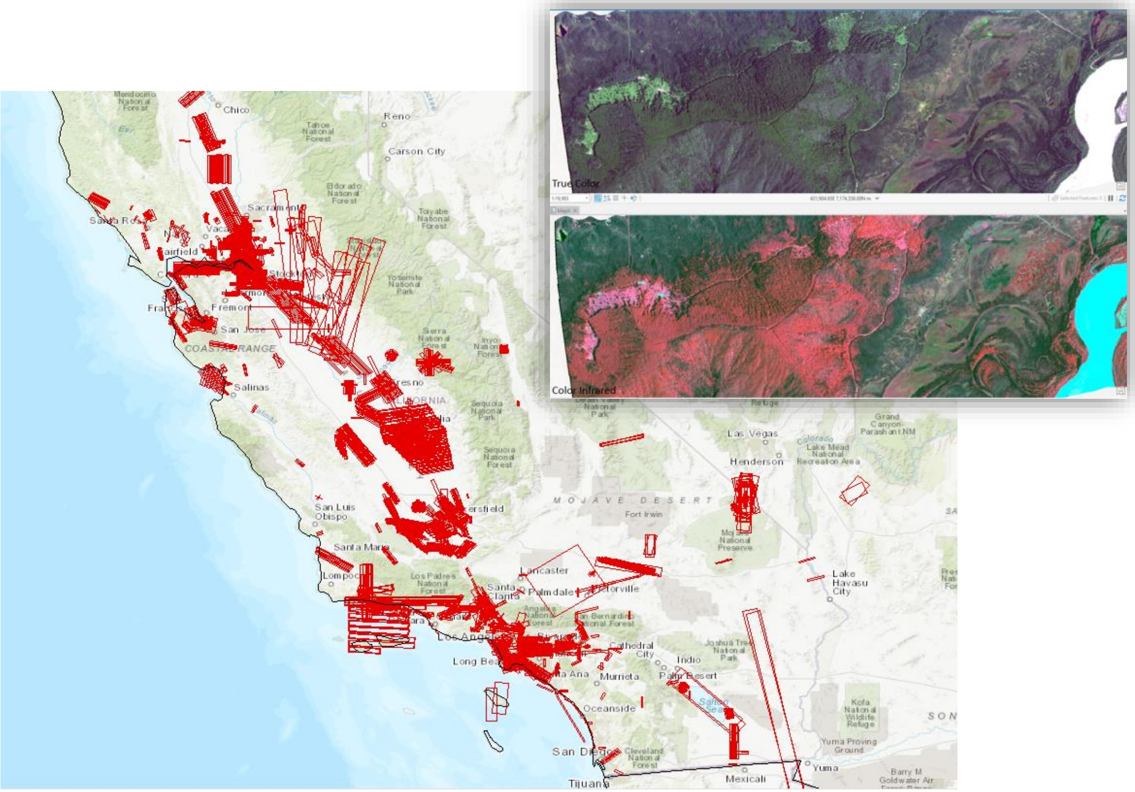
ORNL DAAC: NASA Airborne Facility Instrument Data

	MASTER	AVIRIS-C	AVIRIS-NG
Data Availabillity	1999 - present	~1989 - present	2014 - present
No. Flight Lines	~ 6.560	~ 9,347	~ 6,710

MASTER Flight Paths: sample



AVIRIS-NG Flight Paths: sample



ORNL DAAC Airborne Faculty Instrument Data

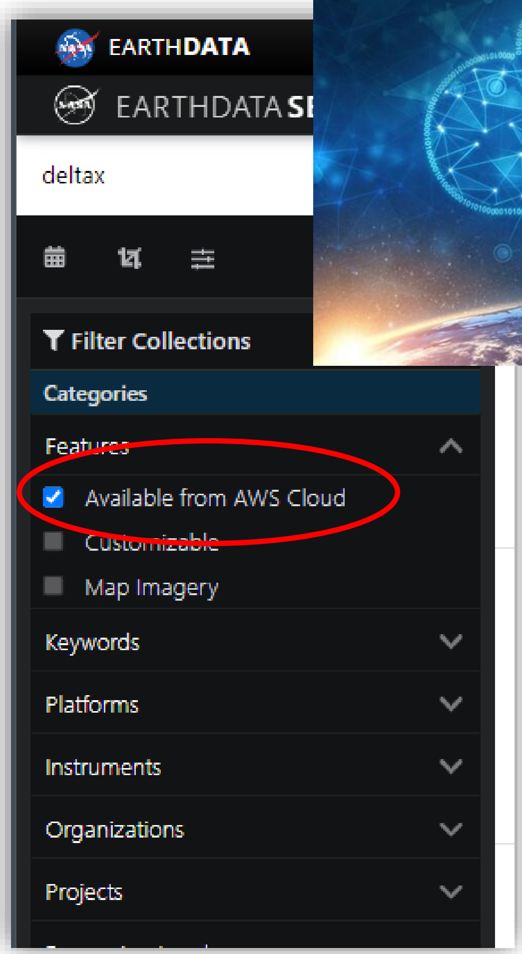
- Archive Challenges
  - File formats and structure
    - Legacy, “dated”, and proprietary file formats (hdf4, ENVI)
    - data structure: rotated grids (authoritative/historic AVIRIS data, North is not “up”)
      - Not GIS “friendly”
  - Data are (too) big/cumbersome for the “download and analyze” workflow
  - Changes/improvements occurred over time
    - Different versions of radiometric corrections

ORNL DAAC Airborne Faculty Instrument Data

- Use Cases
  - Discovery of coincident/concurrent data from airborne, orbital, field measurements
  - Need to “mosaic” and harmonize (normalize) adjacent, overlapping flight path data
    - BRDF, glint corrections
  - Data reduction / transformation needs for overlapping or coincident data
    - subsetting by spatial extent / temporal extent
    - including pixel-level extraction
    - subsetting by band
  - Analysis of “big” data

ORNL DAAC Airborne Faculty Instrument Data Cloud Migration

- Earthdata Cloud Benefits
- Direct S3 Single Access Point
  - when working in AWS US-West-2 region
- Coincident data are co-located
  - indexed metadata
  - discovery and analysis of Coincident Data (orbital, airborne, field)
- Allows for Cloud-optimized formats
  - COG, netCDF, Zarr, Kerchunk
  - facilitate cloud-based data analysis
- Integration with EOSDIS-wide tools
  - Harmony
  - OPeNDAP
  - ArcGIS Server



ORNL DAAC Airborne Faculty Instrument Data Cloud Migration

- Should we convert to COG or a Zarr is the wrong framing
  - Use case dependent
    - Cloud Optimized GeoTIFF (COG)
      - Allows HTTP GET range requests, allows clients to ask for portions of a file
      - Visualization
    - Zarr
      - format for the storage of chunked, compressed, N-dimensional arrays
      - Designed from the ground up to work well with object store
      - Xarray works really well with Zarr
      - Includes _ARRAY_DIMENSIONS
    - Kerchunk : “zero-copy mapping” credit Ryan Abernathy, Columbia University
- No single solution to cloud-optimized formats
- Depends on problem, problem context, and problem solver

ORNL DAAC Airborne Facility Instrument Data Cloud Migration

Historic Data

- NASA Earthdata Archive long-term preservation
- data provenance
- file and collection metadata
- S3 access and cloud compute capabilities

Cloud Optimized

- Create cloud-optimized mirror-copy of historic data using Community Standards
- Coordinate with and consider NASA ESDIS-wide tools and services
- Develop or point to Notebook Tutorials for common cloud-optimized needs

User Community

- Leverage NASA Facility Instrument and Science Teams to expose analysis and methods to the broader user community
- Develop and provide opportunities to user community through workshops, webinars, and learning resources

ORNL DAAC Airborne Facility Instrument Data Cloud Migration

Notebook Example:
AVIRIS-NG to Cloud-optimized formats

Introduction

NASA's AVIRIS-NG comes as ENVI image format, which might not be amenable to some software and the cloud. This tutorial demonstrates simple conversions to various formats such as GeoTIFFs, CoGs, netCDF, and zarr.

Sample file

We will use one Level 2 reflectance file `ang20170706t180635_corr_v2p9_img`, which can be downloaded from this portal. The `corr` in the file name means that this is a orthocorrected, scaled reflectance image file. The `v2p9` is a version marker. For more information about the file name, read the filenaming convention.

Converting to netCDF

```
import rioxarray
avirisng_f = 'ang20170706t180635_corr_v2p9_img'
avirisng = rioxarray.open_rasterio(avirisng_f, driver='ENVI', parse_coordinates=True)
avirisng.attrs['long_name'] = 'orthocorrected scaled reflectance'
```

Now, the `avirisng` variables contains all bands and attribute information as xarray.

We will add a longitude and latitude coordinates as well.

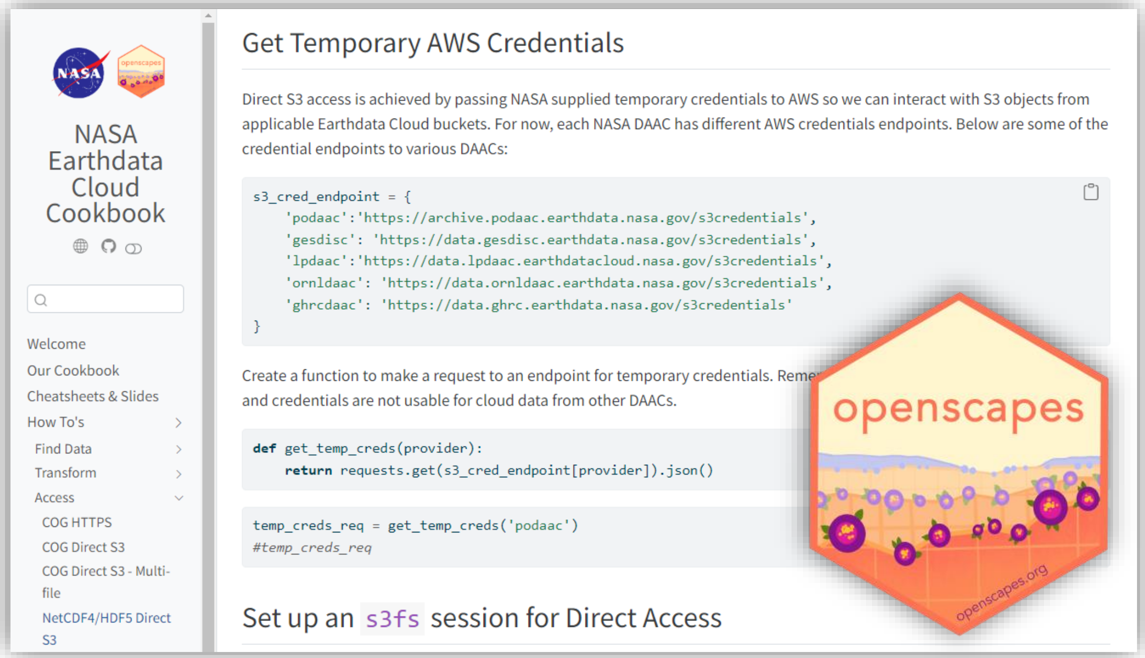
```
from pyproj import Transformer
import xarray as xr

transformer = Transformer.from_crs(avirisng.rio.crs, "epsg:4326", always_xy=True)
lon1, lat1 = transformer.transform(avirisng.xc, avirisng.yc)
avirisng = avirisng.assign_coords(lon=xr.DataArray(lon1, dims=avirisng.xc.dims))
avirisng = avirisng.assign_coords(lat=xr.DataArray(lat1, dims=avirisng.yc.dims))
```

We will first give a name ("corr") to the dataarray and then convert to a netCDF4 file.

https://github.com/rupesh2/aviris_conversion

Tutorial Example:
NetCDF/HDF5 File - S3 Direct Access



Get Temporary AWS Credentials

Direct S3 access is achieved by passing NASA supplied temporary credentials to AWS so we can interact with S3 objects from applicable Earthdata Cloud buckets. For now, each NASA DAAC has different AWS credentials endpoints. Below are some of the credential endpoints to various DAACs:

```
s3_cred_endpoint = {
    'podaac':'https://archive.podaac.earthdata.nasa.gov/s3credentials',
    'gesdisc': 'https://data.gesdisc.earthdata.nasa.gov/s3credentials',
    'lpdaac':'https://data.lpdaac.earthdatacloud.nasa.gov/s3credentials',
    'ornldaac': 'https://data.ornldaac.earthdata.nasa.gov/s3credentials',
    'ghrcdaac': 'https://data.ghrc.earthdata.nasa.gov/s3credentials'
}
```

Create a function to make a request to an endpoint for temporary credentials. Reme and credentials are not usable for cloud data from other DAACs.

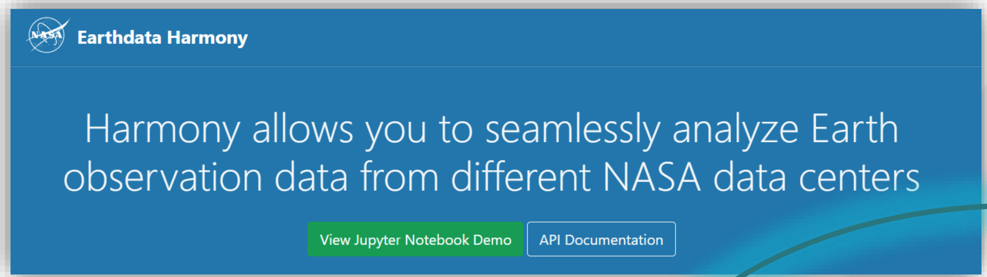
```
def get_temp_creds(provider):
    return requests.get(s3_cred_endpoint[provider]).json()
```

```
temp_creds_req = get_temp_creds('podaac')
#temp_creds_req
```

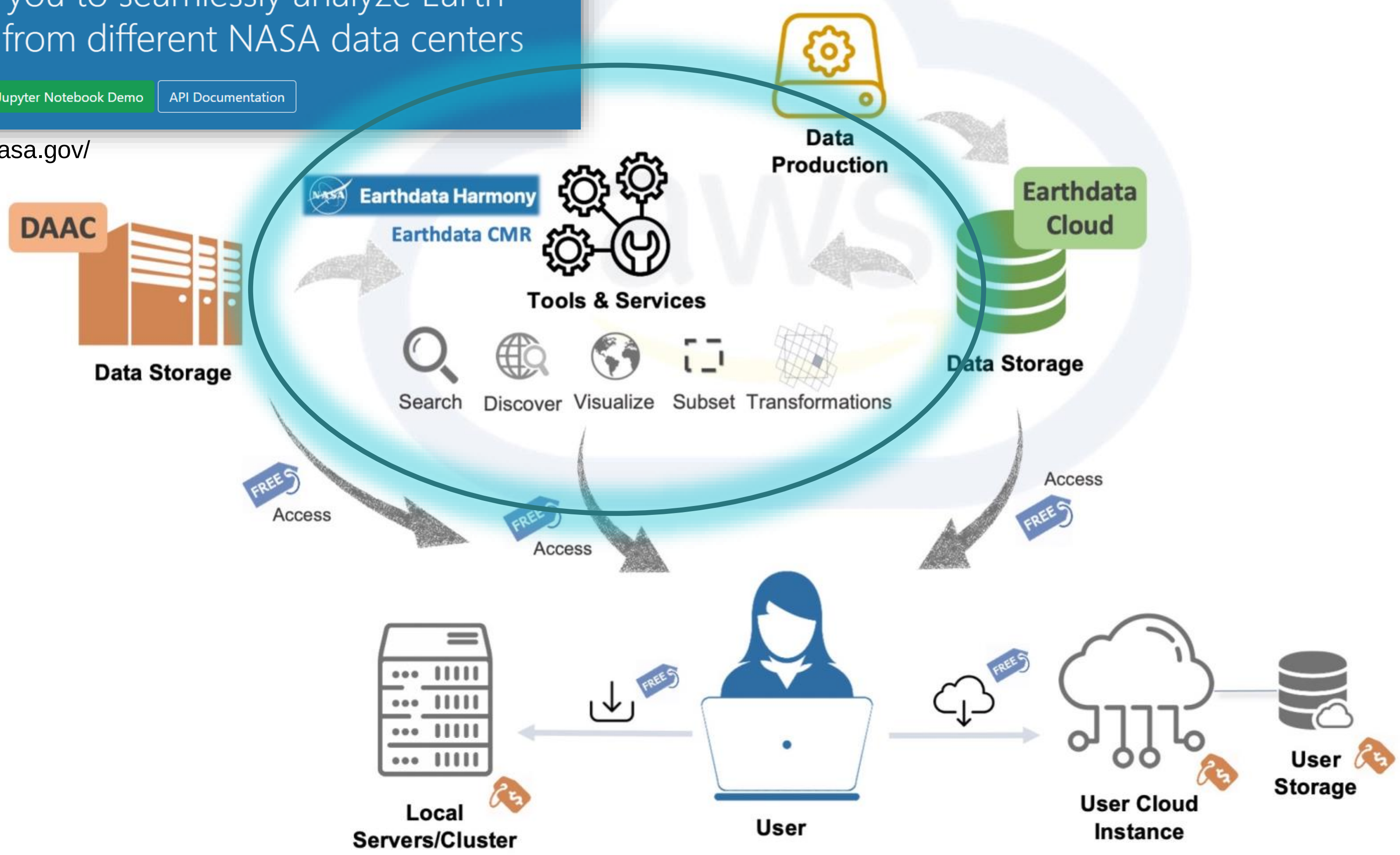
Set up an `s3fs` session for Direct Access

https://nasa-openscapes.github.io/earthdata-cloud-cookbook/how-tos/access/Earthdata_Cloud__Single_File__Direct_S3_Access_NetCDF4_Example.html

ORNL DAAC Airborne Facility Instrument Data Cloud Migration – Enterprise Services



https://harmony.earthdata.nasa.gov/



https://www.earthdata.nasa.gov/learn/articles/harmony-in-the-cloud

Leverage the Experience and Methods of Science Teams

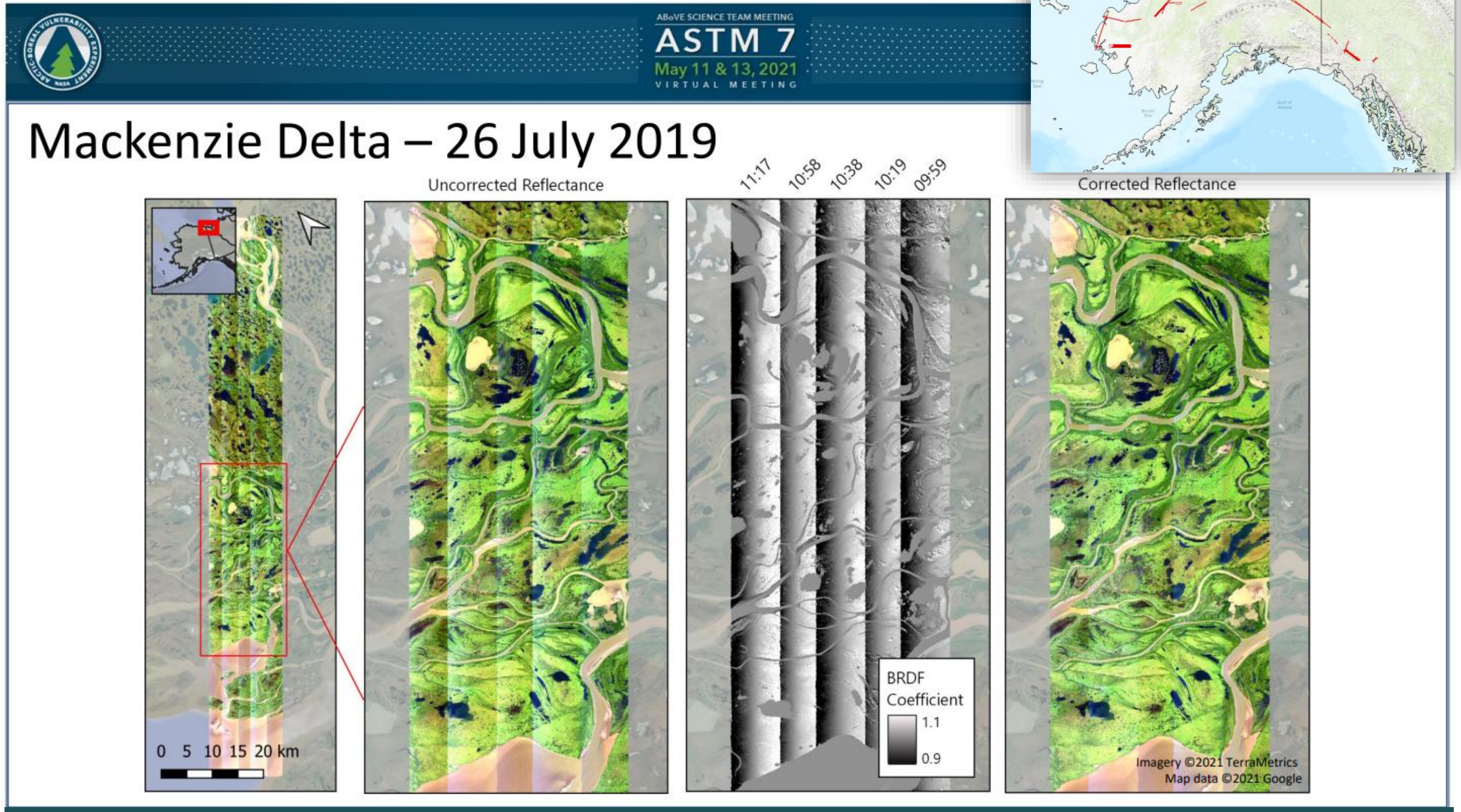
- ABoVE (Arctic Boreal Vulnerability Experiment)
  - Extensive use of NASA Facility Instruments

Bidirectional Reflectance Distribution Function (BRDF)

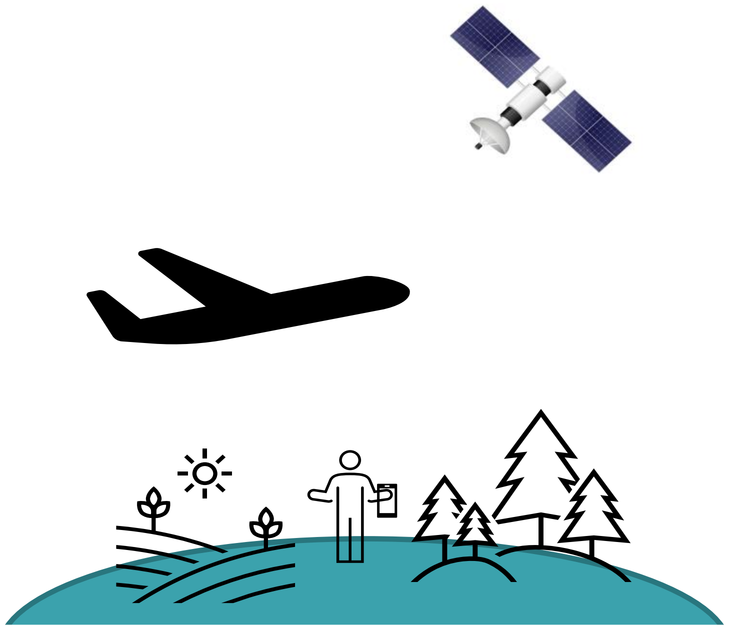
The same target has varying brightness based on solar and sensor viewing geometries

"BRDF correction of ABoVE AVIRIS-NG flightline using FlexBRDF"

Natalie Queally, Zhiwei Ye, Kyle Kovach, Ryan P. Pavlick, Fabian Schneider, Philip A. Townsend



https://cce-datasharing.gsfc.nasa.gov/files/conference_presentations/Talk_Queally_35_27.pdf



(Toward) Analysis Ready and Cloud Optimized Data Formats for NASA Airborne Facility Instruments

Questions?