PACE Data:
Products, Formats, Distribution

Fred Patt

August 4, 2022

Purpose

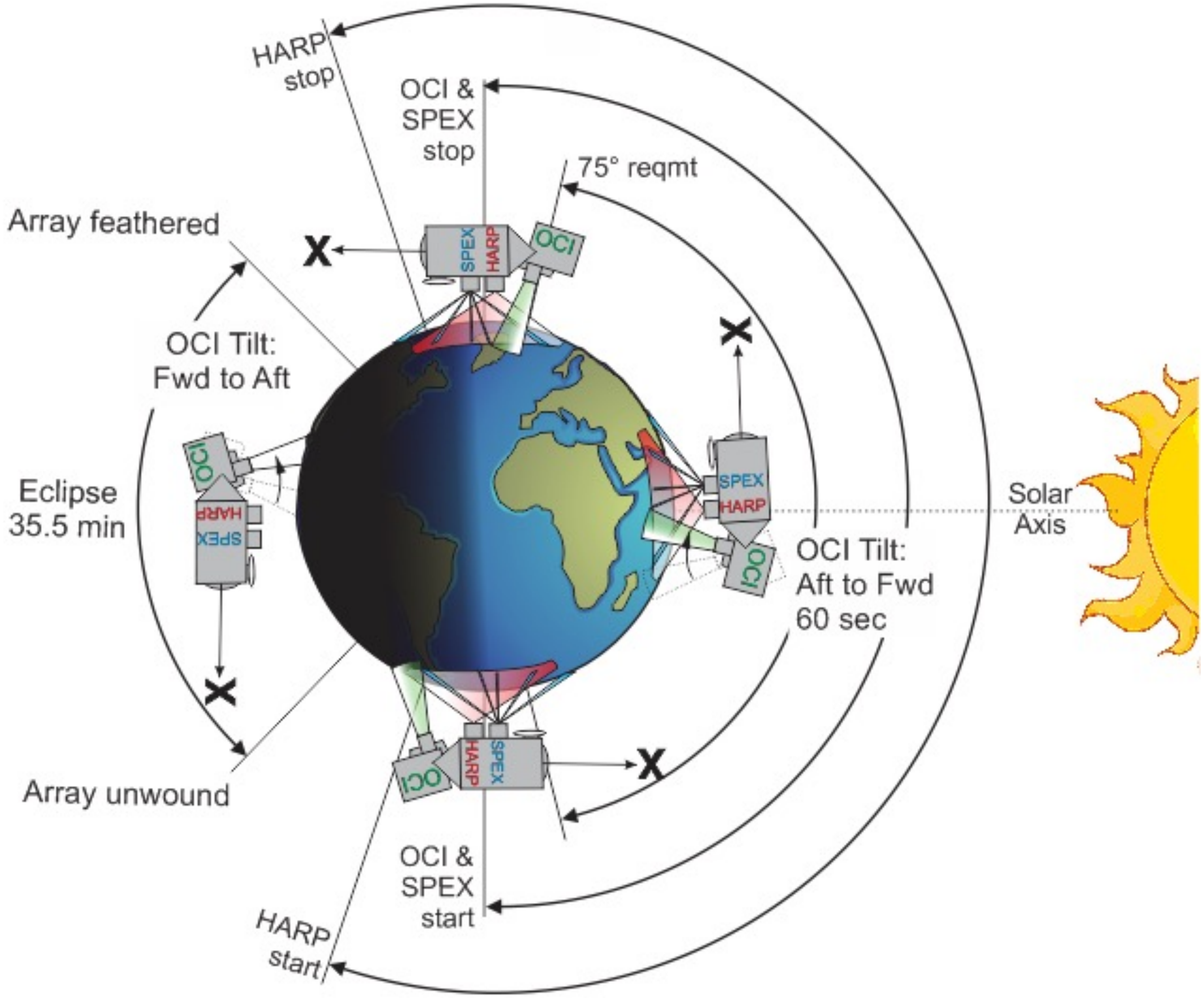
Describe how PACE science data makes its way from

instruments on the spacecraft

to

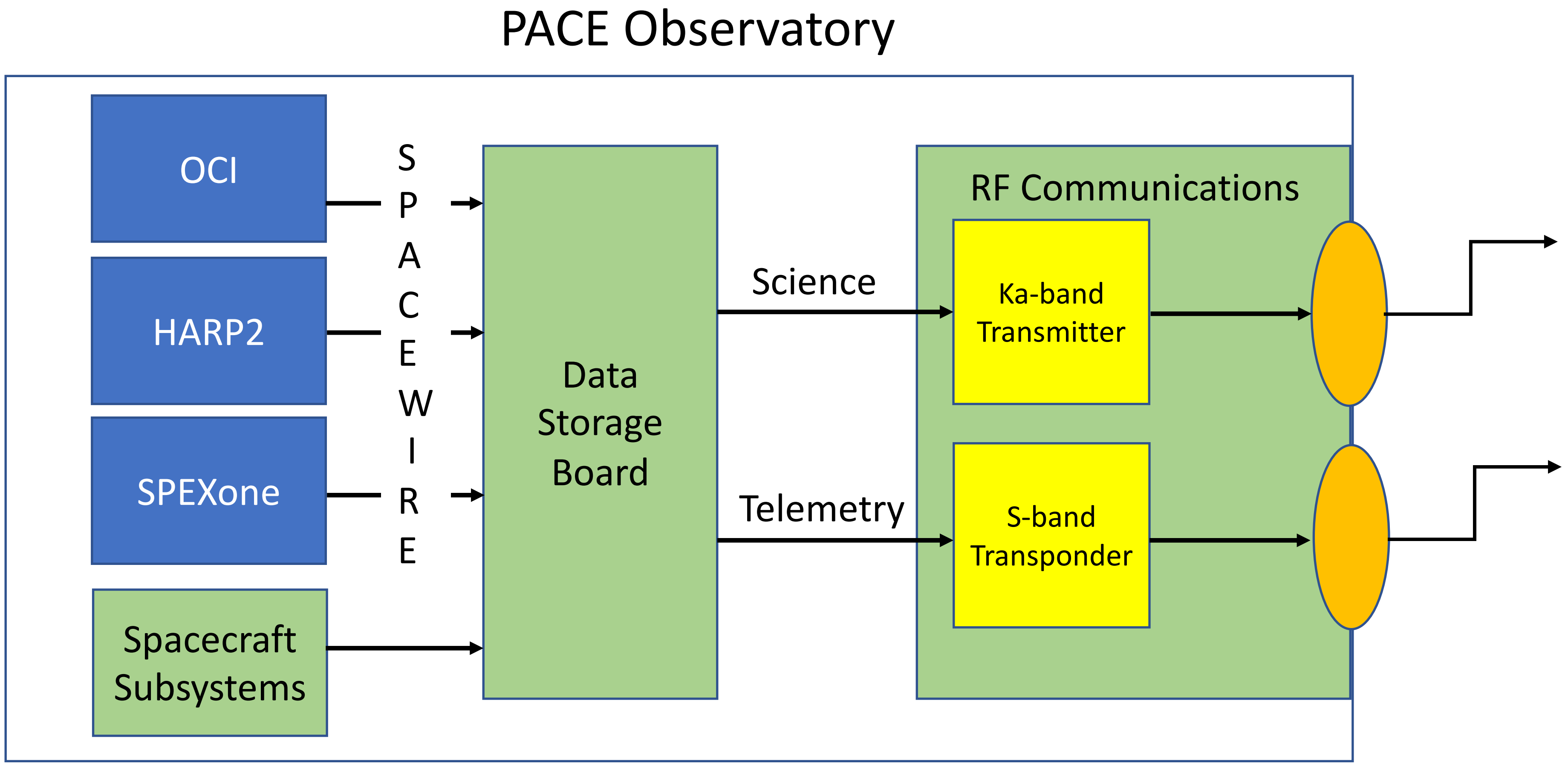
scientifically meaningful data products in the hands of the international research community

Science Data Collection

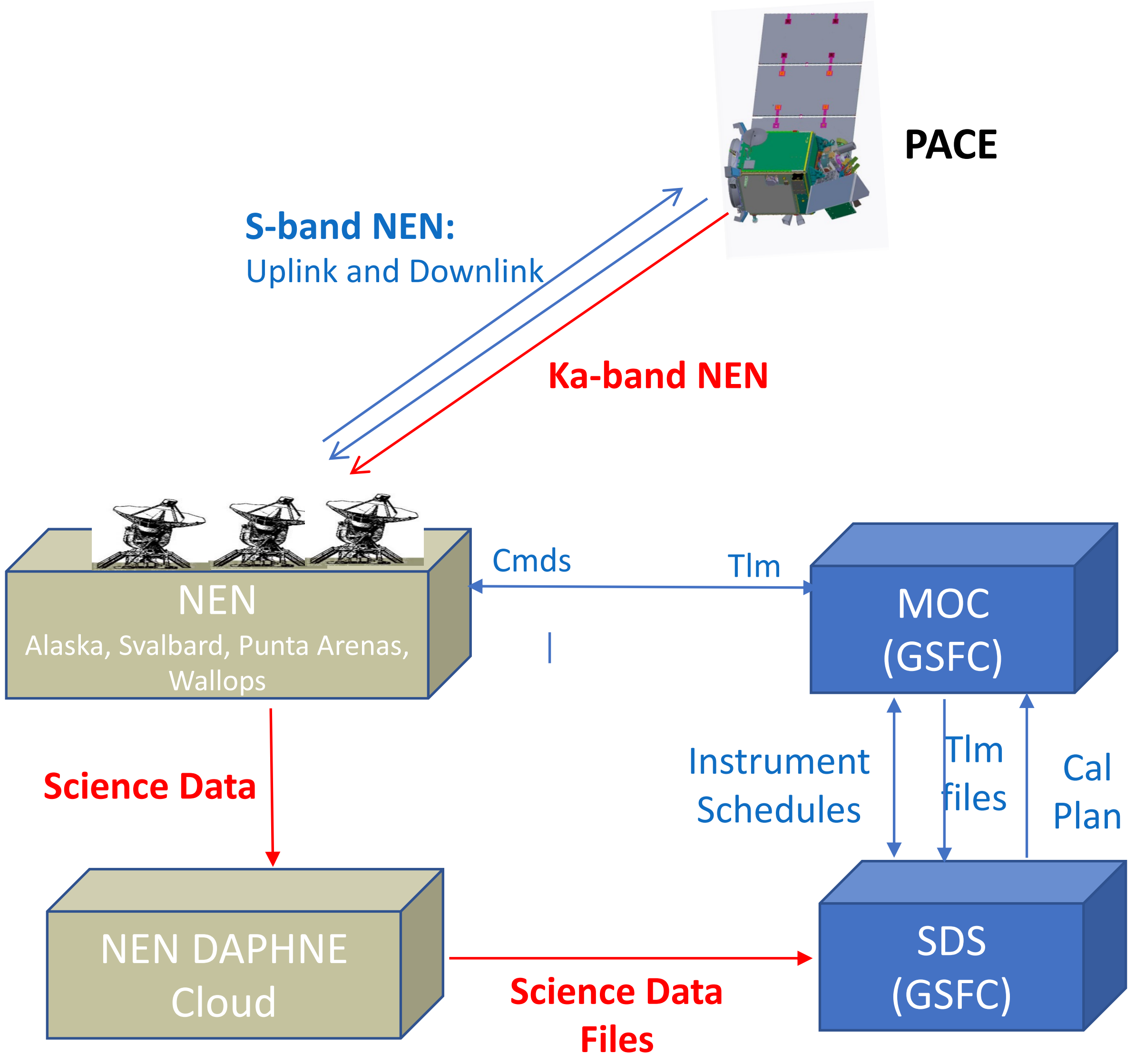


Credit: Jamie Eitnier

Onboard Data Storage and Transmission



Data Downlink and Delivery



Data Product Levels

Data product levels are standardized for all NASA Earth science missions:

- Level 0 – Raw instrument / spacecraft data (CCSDS[1] packets)
- Level 1A – Reconstructed, unprocessed instrument data
- Level 1B – Calibrated, geolocated instrument data in physical units
- Level 1C – Co-registered Level-1B data
- Level 2 – Science data products at instrument resolution
- Level 3 – Temporally and spatially composited (i.e., mapped) science data products

All product above Level 0 are in NetCDF, a platform-independent, self-describing file format.

[1]Consultative Committee for Space Data Systems

GODDARD EARTH SCIENCES

Why are there so many data product Levels?

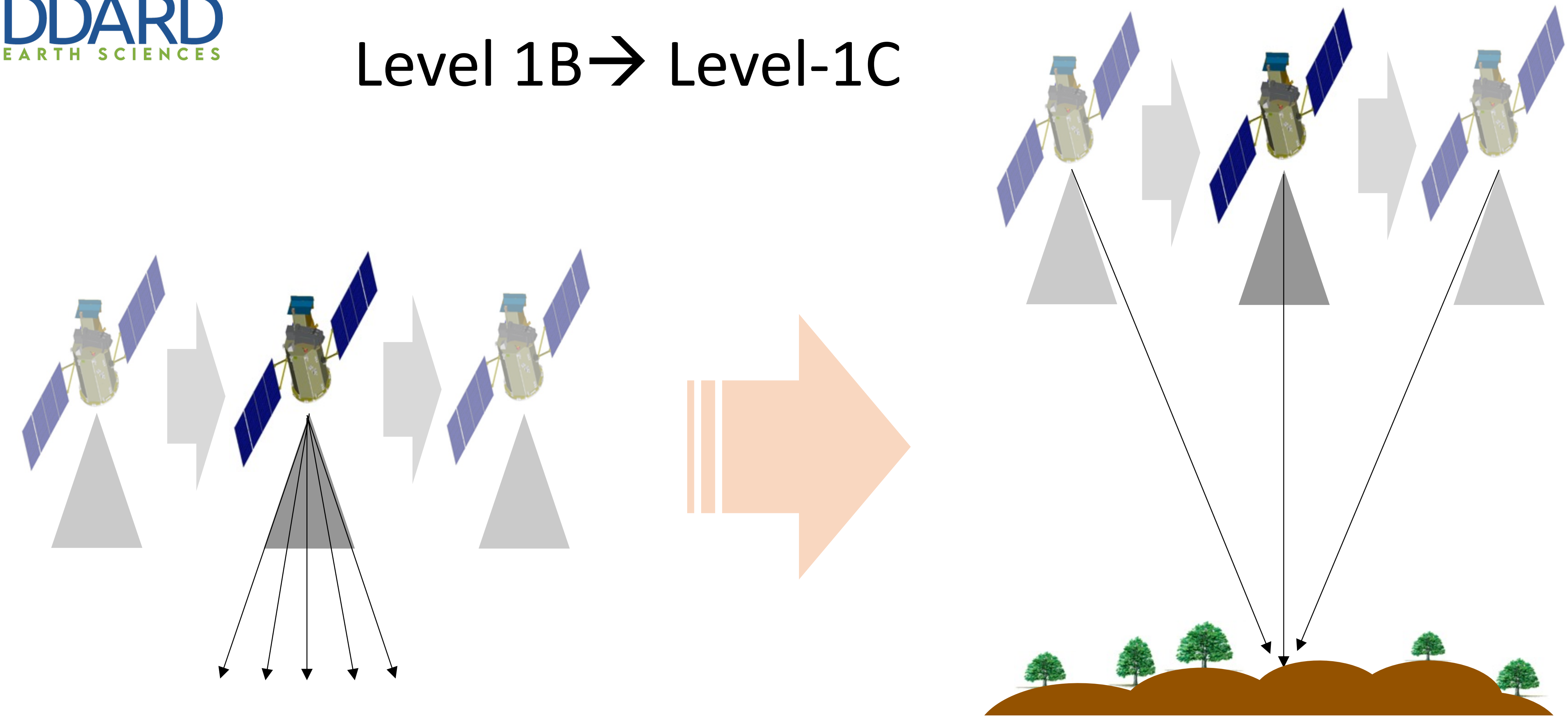
Each data product level represents a logical step in the data processing.

- Level 0 is all of the data generated by the spacecraft – must save!
  - If needed, all downstream product levels can be regenerated from Level-0.
  - Typically stored in multiple copies and locations to protect against loss.
- Level 1A re-creates the as-measured data structure
  - CCSDS packets typically break up the data.
  - For OCI, the measurement structure is one scan (telescope rotation).
  - For HARP2 and SPEXpone, the structure is one detector image.
  - Level 1A processing also includes measurement time referencing and collection of associated spacecraft telemetry (e.g., navigation).
  - Data "granules" (e.g., N minutes) are set at this step.

Many data product levels (cont.)

- Level 1B products incorporate calibration and geolocation
  - The raw measurements are converted to physical units (e.g., reflectance).
  - All known instrument corrections are applied (e.g., temporal variation, temperature, linearity).
  - Full spectral and spatial resolution is maintained.
  - Geolocation of each measurement is computed using the spacecraft navigation data and instrument geometry model.
- Level 1C products co-register the measurements taken at multiple viewing angles (HARP2 and SPEXone).
  - Required for science algorithms based on multiple view angles.
  - L1C grid is based on a nadir view angle.

Level 1B→ Level-1C



Level 1B:
Calibrated, geolocated data in spacecraft reference frame

Level 1C:
Calibrated, geolocated data *mapped to location of interest to input to L2 algorithms*

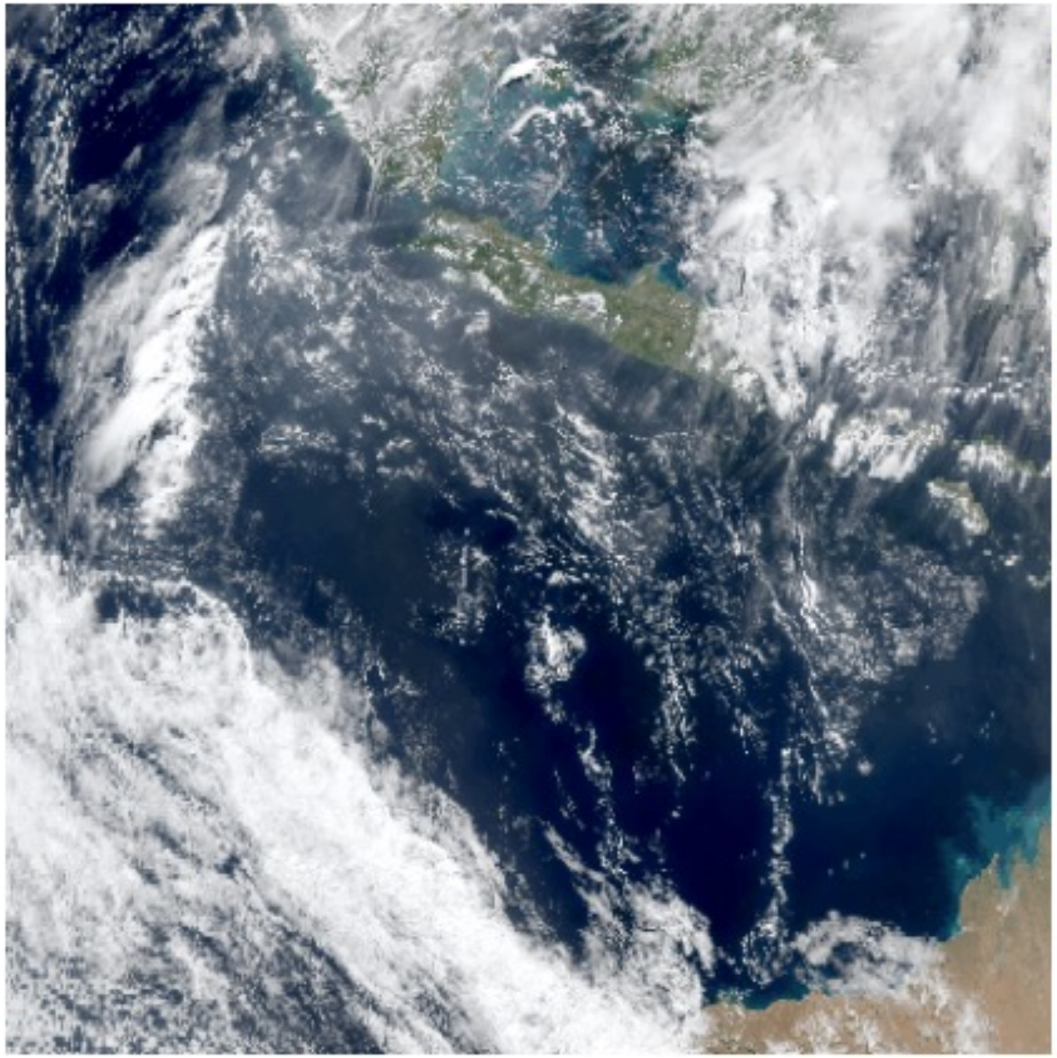
Credit: Kirk Knobelspiesse

Many data product levels (cont.)

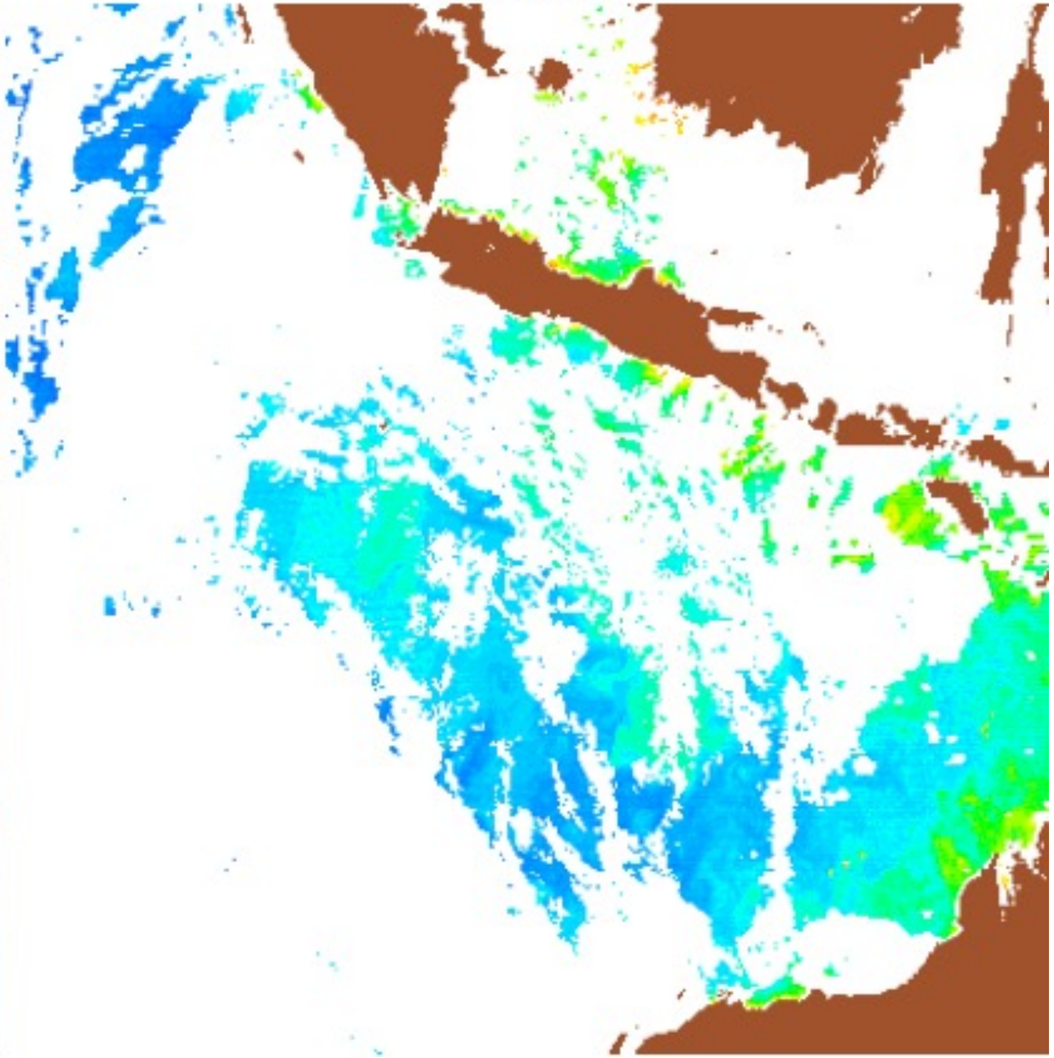
- Level 2 are the first science products
  - Calibrated data are processed to retrieve geophysical parameters.
  - OCI L2 products are generated for ocean color, aerosol and cloud properties.
  - L2 products are also generated for HARP2 and SPEXone.
  - Most algorithms require additional (ancillary) data from external sources
  - In most cases, parameters are retrieved at full instrument resolution.
  - Quality flags indicate algorithm issues and conditions.
- Level 3 products are spatially and temporally composited and resampled to a geographic coordinate system.
  - Daily products combine all Level-2 products for a day.
  - Products are temporally composited over progressively longer time scales (8 days, month, season, year, mission).
  - Coordinate systems are either equal-area (binned) or equal-angle (mapped)

Level 1B and Level 2 Images

Quasi True Color



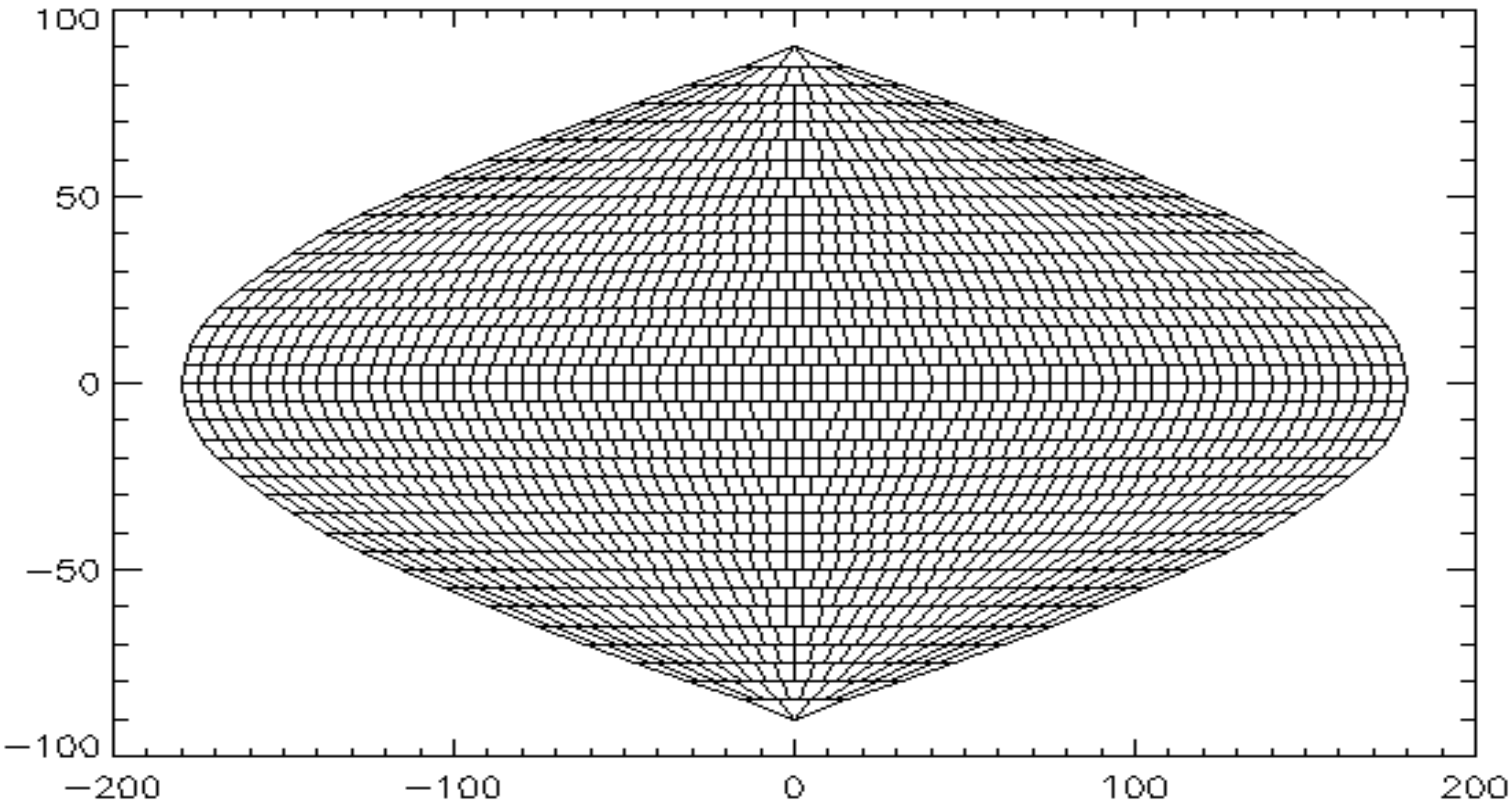
Chlorophyll



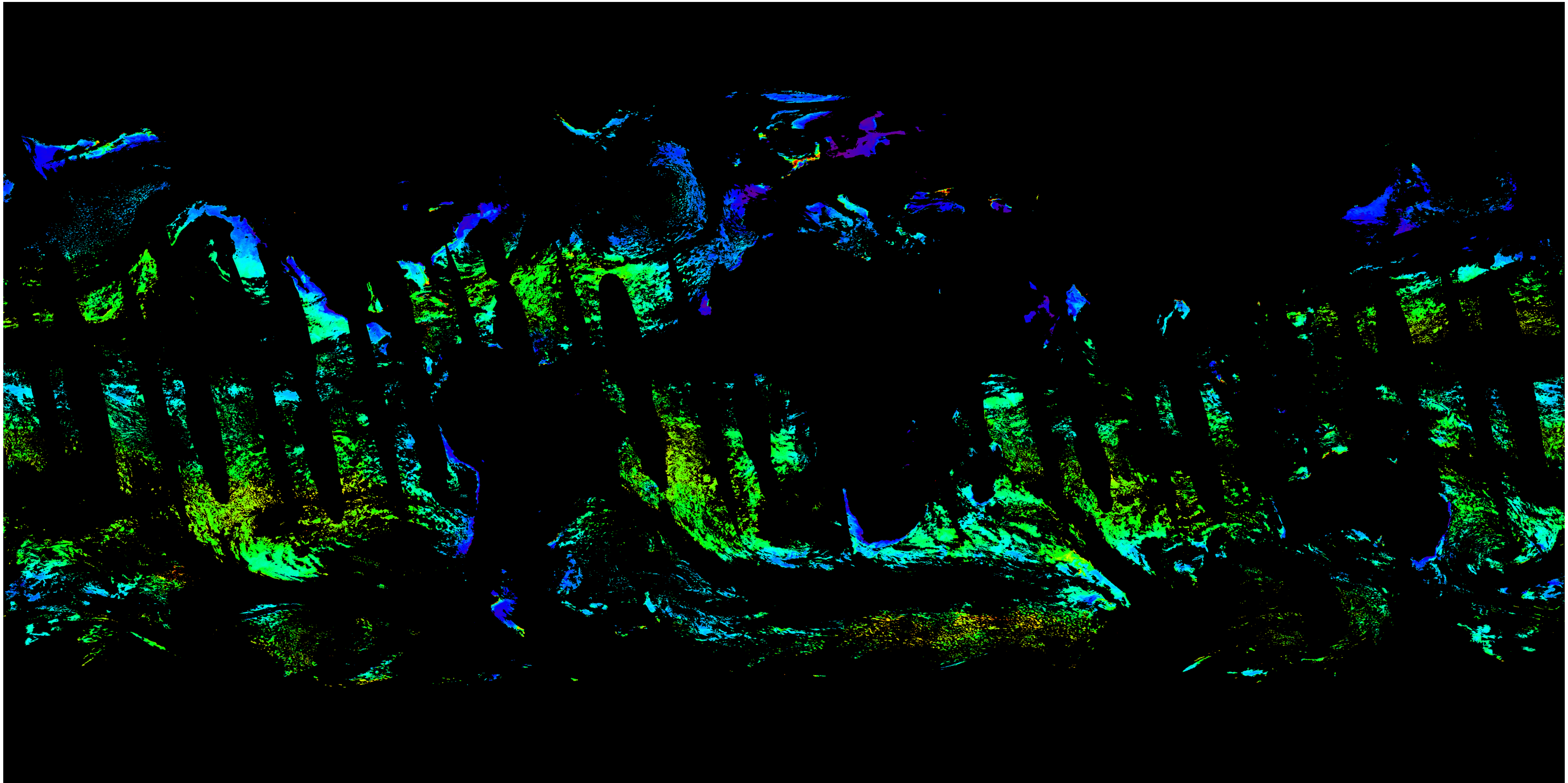
Level 3 Binning vs Mapping

- The initial Level 3 processing is ***binning***.
  - All of the products for the specified time range are collected and averaged in bins in a geographic reference frame.
  - Our binning projection is sinusoidal equal-area.
  - L2 observations are selected or excluded based on quality flags.
  - The most common bin resolution we use is 4.6 km.
- The binned products are then ***mapped*** to a standard projection.
  - Our standard map projection is equi-rectangular (Platte Carre).
  - These are the most commonly-requested products.
- Binning is ***collection*** and ***selection***, while mapping is ***projection***.

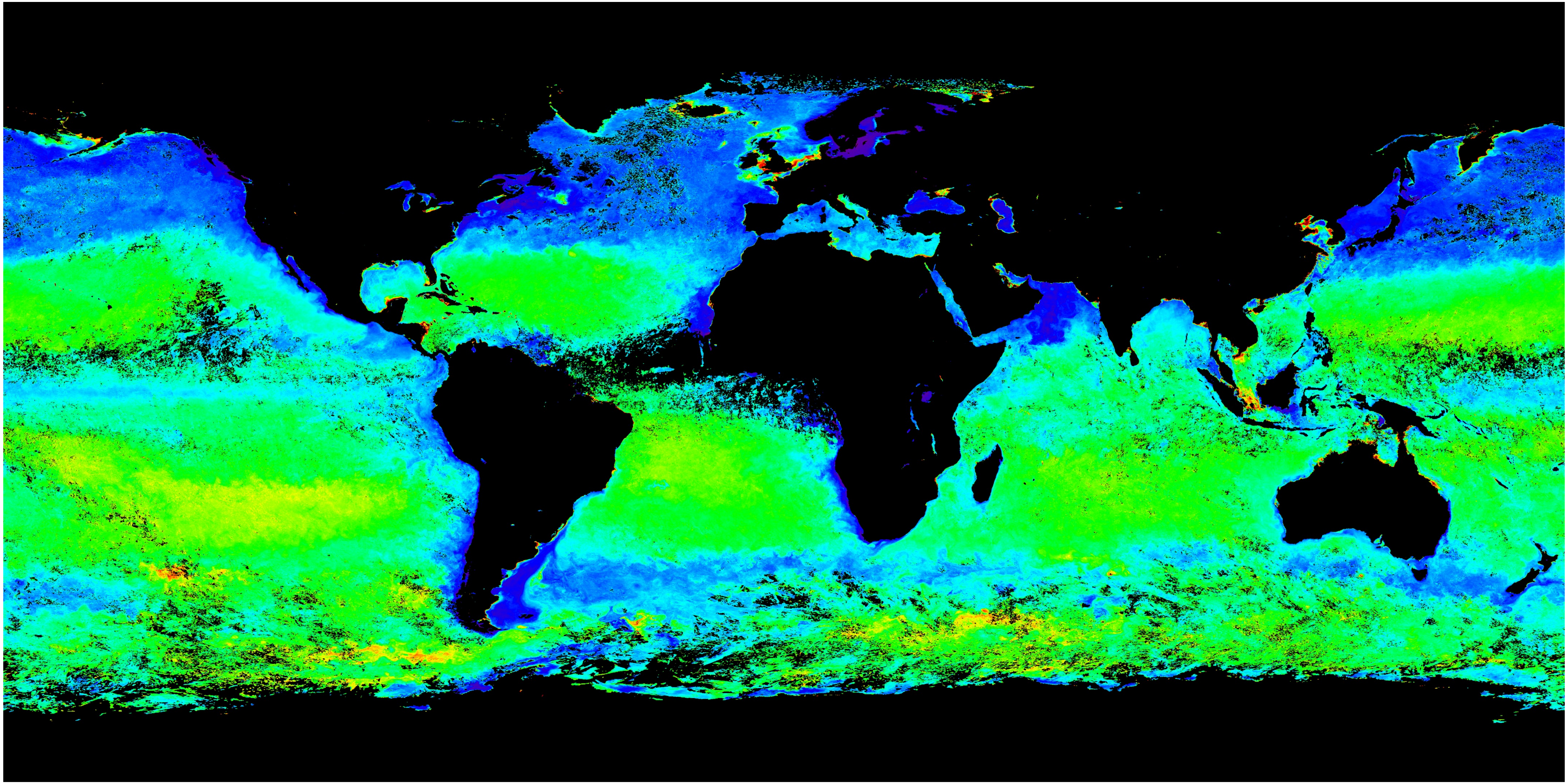
Sinusoidal Equal-area Projection



Level 3 Daily (486 nm Rrs)



Level 3 Monthly (486 nm Rrs)



Why NetCDF?

- NetCDF is a platform-independent, self-describing data format.
  - Platform independence allows products generated on any computer to be read on any other computer.
  - Self describing means the user does not need to understand the inner structure of the file to read the data.
- NetCDF supports attributes to describe the products at the file, group and parameter levels
- Internal compression at the parameter array level allows for efficient data storage.
- NetCDF file access does require use of tools and/or libraries.
  - It is well supported by standard languages and analysis packages.

Sample NetCDF Parameter Attributes

```
float chlor_a(number_of_lines, pixels_per_line) ;
        chlor_a:long_name = "Chlorophyll Concentration, OCI Algorithm" ;
        chlor_a:units = "mg m^-3" ;
        chlor_a:standard_name = "mass_concentration_of_chlorophyll_in_sea_water" ;
        chlor_a:_FillValue = -32767.f ;
        chlor_a:valid_min = 0.001f ;
        chlor_a:valid_max = 100.f ;
        chlor_a:reference = "Hu, C., Lee Z., and Franz, B.A. (2012). Chlorophyll-a
algorithms for oligotrophic oceans: A novel
```

Sample Group Attributes (Flag Percentages)

```
group: flag_percentages {
  // group attributes:
        :ATMFAIL = 6.698958f ;
        :LAND = 31.29364f ;
        :PRODWARN = 0.03301942f ;
        :HIGLINT = 16.81254f ;
        :HILT = 0.f ;
        :HISATZEN = 21.20025f ;
        :COASTZ = 0.7254022f ;
        :SPARE = 0.f ;
        :STRAYLIGHT = 25.56421f ;
        :CLDICE = 63.61963f ;
        :COCCOLITH = 0.00572401f ;
        :TURBIDW = 0.2648902f ;
        :HISOLZEN = 0.f ;
        :LOWLW = 0.6002185f ;
        :CHLFAIL = 0.01656289f ;
        :NAVWARN = 0.f ;
        :ABSAER = 0.f ;
        :MAXAERITER = 0.2386197f ;
        :MODGLINT = 43.01252f ;
        :CHLWARN = 0.01076153f ;
        :ATMWARN = 0.2804571f ;
        :SEAICE = 0.f ;
        :NAVFAIL = 0.f ;
        :FILTER = 0.f ;
        :BOWTIEDEL = 12.875f ;
        :HIPOL = 5.021407f ;
        :PRODFAIL = 92.06805f ;
  } // group flag_percentages
```

PACE Simulated and Proxy Data

- Efforts are well underway to simulate OCI, HARP2 and SPEXone data to test the data processing and science algorithms at SDS.
  - OCI PyTOAST simulation at Level 1B by SDS
  - OCI L1B simulation by the GSFC Global Modeling and Assimilation Office
  - HARP2 L1A and L1C simulations by UMBC
  - SPEXone L1A and L1C simulations by SRON
- A common test day of March 21, 2022 has been agreed to for the initial Day-in-the-Life (DITL) test at SDS
  - SDS provided simulated navigation data to all teams.
  - Prior simulation efforts also performed for different dates.
- Precursor instrument data (e.g., airborne HARP and SPEX) have also been used as proxy data sources.

Simulated and Proxy Data Access on OceanColor Web

https://oceancolor.gsfc.nasa.gov/data/pace/test-data/

OceanColor WEB

ABOUT MISSIONS DATA DOCS SOFTWARE & TOOLS SERVICES GALLERY FORUM

PACE Mission ▸ Similated and Proxy Data

Simulated and Proxy Data

For OCI, simulated data at Level-1B is being generated to support processing software development and testing. Simulated Level-1C data has been produced for SPEXone, and there is a SPEX airborne instrument that provides valuable proxy data for testing algorithms. Proxy data is also available for HARP2, including data from an airborne version, and the HARP cubesat that began operations in April 2020.

OCI Level-1B

Several simulators have been developed to produce OCI Level-1B data.

- pyTOAST Simulation

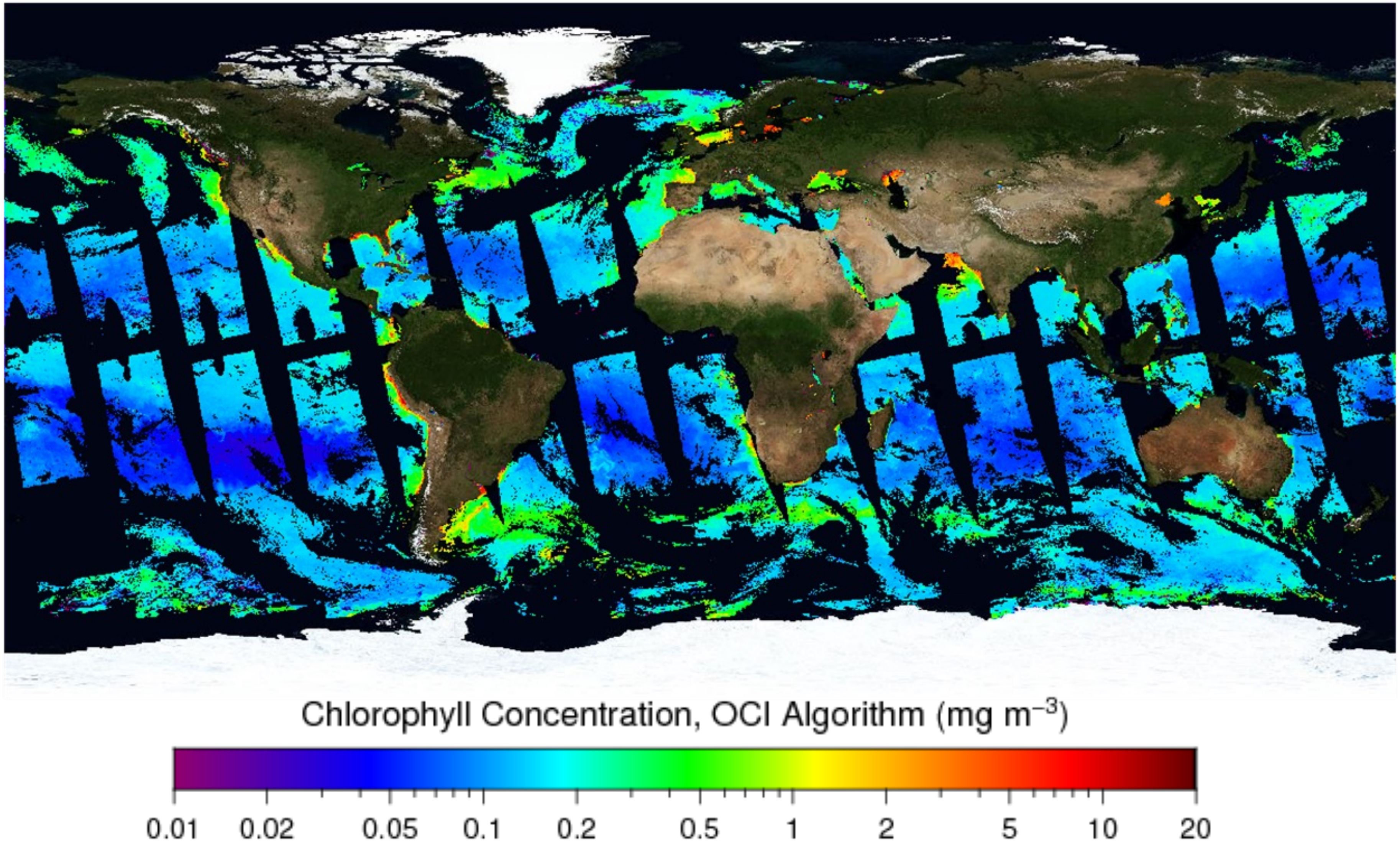
  The python top-of-atmosphere simulation tool (pyTOAST) utilizes retrieved surface and atmospheric properties and top-of-atmosphere (TOA) radiances from MODIS and VIIRS, pre-computed radiative-transfer look-up tables for the OCI spectral response, and spectral libraries of land and clouds to produce realistic radiometry in the standard Level-1B format of OCI. The pyTOAST simulator is computationally efficient, and thus allows for large scale production for testing retrieval software mechanics and data flow. OCI Level-1B pyTOAST data is available for a 5-day period from 21-25 March 2019. Read more...

- GMAO Simulation (coming soon)

SPEXone Level-1C

- SPEXone Simulated Data

Chlorophyll-A from PyTOAST for March 21, 2022



Questions?