Session Introduction

SIT Chair

CEOS SIT Technical Workshop 2020

Session and Agenda Item #2.2.1

9 September 2020

CEOS

Committee on Earth Observation Satellites

SST-VC ARD Perspectives

Edward M. Armstrong[1] , Anne O'Carroll[2], Christo

Whittle[3]

CEOS SST-VC SST-VC co-leads

CEOS SIT Technical Workshop 2020

Session and Agenda Item #2.2.2

9 September 2020

[1]NASA Jet Propulsion Laboratory, California Institute of Technology,
[2]EUMETSAT, [3]SANSA / CSIR

Introduction

- A working ARD definition review
- ARD survey for Oceans
  - Results of the MODIS Aqua/Terra Surface Temperature (SST) ARD assessment and recommendations
  - Review of assessment criteria and recommendations
  - ARD delivery and services for L1/L2 - a thought experiment
- Ways forward with other ocean datasets
- Summary

An ARD definition

- Time series data pre processed for further scientific (and interdisciplinary) analysis
  - o Ready for
    - Time series analysis
    - Subsetting
    - Regridding/Reprojection/Recombination/Reformatting
    - New compute platforms (e.g., cloud, noSQL database)
  - o Processed for
    - Environmental and physical corrections at pixel level
    - Big Data format
  - o Documented for
    - Data processing, quality, and error assessment
- ARD foundation definition separate from deployment or services

- Cloud based AWS MUR in Zarr format
  - Created by Chelle Gentemann (Farallon Institute) et al.
  - Part of the Amazon Web Service (AWS) Open Data Registry on their cloud platform
    - https://registry.opendata.aws/mur/
    - GitHub: https://github.com/pangeo-gallery/osm2020tutorial
  - Provides a complete ecosystem based on Pangeo, Xarry, Dask, Jupyter and Binder to manipulate the MUR Zarr dataset on a few or >100 CPUs
  - Compute credits provided by Amazon
  - Supported by the NASA IMPACT program

- Goal: assess the ARD readiness of **GHRSST MODIS Aqua/Terra SST Level 2P** datasets
  - DOIs: **10.5067/GHMDA-2PJ19**, **10.5067/GHMDT-2PJ19**
  - Together represent multi decadal SST observations starting in 1999 (Terra) and 2002 (Aqua)
    - Combined 100 TB size
  - Recently reprocessed to v2019.0 with improved cloud screening
  - Have been available in a GHRSST specification since 2006
    - Complete CF/ACDD/ISO granule metadata
    - Pixel level uncertainty, quality flagging and auxiliary information

The Assessment

- Using the CAR4L Surface Temperature template
  - Four sections: General, Per Pixel, Radiometric/Atmos Corrections, Geo Corrections
  - Total 36 assessment factors
- Overall nearly every category sufficiently addressed (More details next slide)

	Threshold	Target
General Metadata	15 Pass/ 1 Partial Pass / 0 Fail	11 Pass / 3 Partial Pass/ 2 Fail
Per pixel	7 Pass / 0 PP / 1 Fail / 5 NA	2 Pass/ 4 PP / 2 Fail / 5 NA
Rad/Atmos Corrections	5 Pass/ 0 PP/ 0 Fail / 2 NA	5 Pass/ 0 PP /0 Fail / 2 NA
Geo Corrections	1 PP	1 PP

Assessment Review

- Based on these assessment results we rank these MODIS SST datasets with a "grade" of: B+ / A-
- Positive factors
  - Complete granule level metadata in CF /ACDD in netCDF4 data model
  - Complete pixel level quality and uncertainty information
  - Well described coordinate systems
  - Bonus: all variables described with *coverage_content_type* attribute (ISO 19115 adapted to ACDD v1.3)
    - defines variable as physical measurement, quality information, auxiliary information etc.
  - Bonus: variables described with CF standard names where appropriate

- Negative factors
  - Algorithm and other provenance information not completely documented
  - Quality level "condition" tracing not readily available
  - Solar and satellite viewing angles missing
- Other observations
  - Some factors not relevant
    - Projection information.
      - These are unprojected L2 products
      - Pixels don't "line up" as in gridded L3/L4
    - Geometric Correction not as applicable
  - Some factors are more important than others

- Level 2 data are ungridded. But a preferable format for power users that have requirements for the native data space/time resolution.
- With an assumption that they are stored in a cloud environment so processing could occur next to the native data. Necessary services for L2 manipulation would be:
  - A granule discovery service (e.g, Earthdata search, STAC)
  - A subsetting service (e.g., Harmony API)
  - A reprojection service (e.g., Harmony API with GDAL)
  - A reformatting service (e.g., Harmony API)
- In principal if you had all of these and compute assumptions were met one could create a gridded "data cube" at the exact spatial and temporal regions of interest.

Applications to other SST datasets

- The CARD4L assessment can be applied to other GHRSST datasets as well
- The GHRSST catalog currently represents over 90 datasets
- A ongoing activity in the SST-VC will be to engage other GHRSST dataset producers of popular datasets to do an assessment
- Because of the common data/model the answer should mostly be the same
  - But there are nuances in the algorithm documentation/uncertainty derivation/auxiliary information that a survey can draw out
  - Possible outcome: Consideration of consistent definitions and terminology e.g. harmonisation of SST definitions such as SSTskin, SSTsub-skin so that different sensors are delivering data formulated (and with uncertainties assessed) in the same way.

Summary

- Provided a review of an ARD definition and a working example of a L4 SST ARD
- Presented the results of an assessment of L2P MODIS Aqua/Terra SST datasets
  - Overall very positive outcome
    - Coordinate systems, uncertainty information, quality information documented
    - Standardized metadata for machine readability
  - The foundation state of these datasets are ARD and lend themselves to further data manipulation services (thought experiment)
  - CAR4L assessment could be improved with a quantitative scoring and factor weighting
- More ARD assessments will be performed in the SST-VC community and collaboration with others is welcomed
- There are ongoing complementary NASA/Interagency ARD activities such as CEOS COVERAGE, CEOS Interoperability Lab, NASA/ESA MAAP etc.
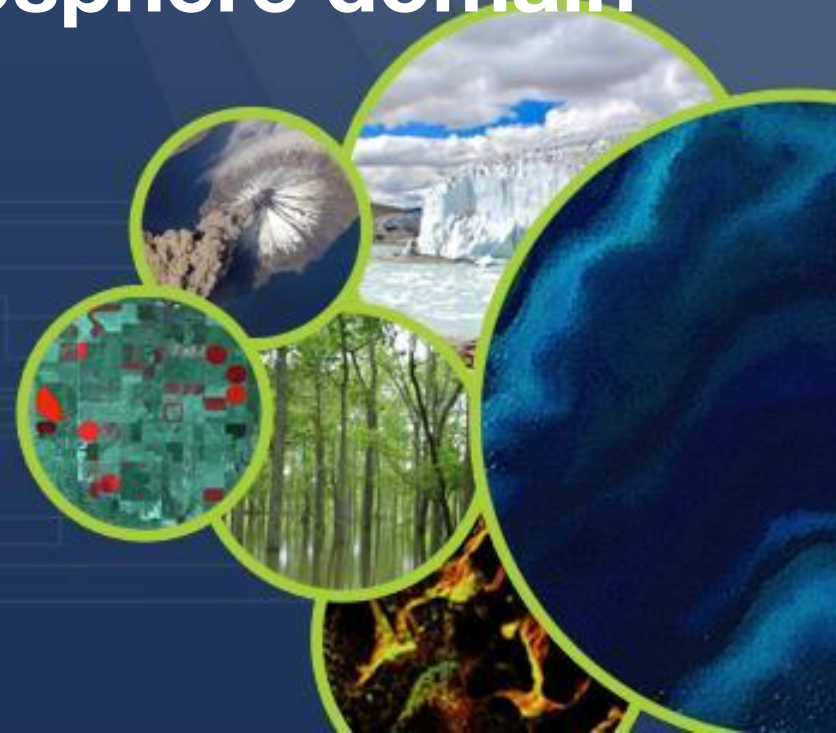
Atmospheric Products

- **The Atmospheric Composition Virtual Constellation (AC-VC) goal is to collect and deliver data to improve monitoring, assessment, and predictive capabilities for changes in the <u>ozone layer</u>, <u>air quality</u>, and <u>climate forcing</u>**

- **The atmospheric constituents products are typically column/profiles of trace and greenhouse gases**
  - ozone profiles
  - aerosol optical depth
  - column-averaged carbon dioxide
  - …

- **Most atmospheric composition instruments are sampling systems, not imagers**

- **A constellations of polar orbiting (LEO) and geostationary (GEO) air-quality sensors is being deployed**
  - The provision of standardized LEO/GEO air-quality ARD will be critical to support direct analysis using multi-sensor data
  - The operational products of current missions like TROPOMI/Sentinel-5 Precursor already cover several characteristics expected from analysis ready data

CARD4L Description Document and TROPOMI/S5P L2 products

- **General Metadata**
  - Compliant

- **Quality Metadata**
  - Compliant

- **Measurement-based / Radiometric Calibration**
  - Not applicable

- **Geometric Calibration**
  - Atmospheric products are usually provided in the satellite-grid and not in a regular-grid

- **Solar and View Angle Correction**
  - This "Correction" is not applicable. The radiative transfer modeling used for atmospheric retrievals takes into account solar and view angles

- **Atmospheric Correction**
  - Not applicable

CARD4L PFS Requirements and TROPOMI/S5P L2 products

- **General Metadata**
  - Traceability, Metadata Machine Readability, Data Collection Time, etc.
    - Compliant
  - Geometric Correction Methods, Spectral Bands, etc.
    - Not applicable

- **Per-Pixel Metadata**
  - Metadata Machine Readability, Solar and Viewing Geometry, etc.
    - Compliant
  - Incomplete Testing, Saturation, etc.
    - Not applicable

- **Radiometric and Atmospheric Corrections**
  The name of this requirement is misleading for atmospheric products, but the content applies
  - Measurement, Measurement Uncertainty
    - Compliant

- **Geometric Corrections**
  - Not applicable

Summary

- **The usability of Atmospheric Level-2 products as ARD is limited, Level-3 (gridded data) and Level-4 (fluxes, sources/sinks) are needed to create useful ARD products**
  - Atmospheric instruments often provide incomplete spatial coverage and/or temporal coverage on time scales over which the trace gas or aerosol concentrations can be changed by local production/loss or transport.
  - The observed concentration of an atmospheric trace gas or the abundance of an aerosol species is determined by both its local sources and sinks and by the ambient wind field.
  - In general, atmospheric Level-2 products can't be ingested in Geospatial Data Cubes

- **Considerable work is needed to adapt and extend the CARD4L Framework (Definition, PFS and Product Assessment) to the Atmosphere domain**
  - AC-VC propose to perform a study of the requirements for specific atmospheric ARD products including reactive gases, aerosols and greenhouse gases.
  - That study could evolve into a white paper and/or a roadmap for a future atmospheric ARD Framework.

- **https://drive.google.com/file/d/1PErXKmywnO0oY4bnVo6TXusAxJfldk_a/view?usp=sharing**

- CEOS Coastal Observations and Applications Study Team (CEOS-COAST): Initiated at 33rd CEOS Plenary to define a CEOS Coastal Strategy and facilitate engagement with key international stakeholders

  - **Bridging land and aquatic observations within CEOS, helping to integrate across multiple CEOS entities and domains, both thematic and technical.**

  - **Leveraging CEOS Systems, services and interoperability approaches, including the CEOS Analysis Ready Data (ARD) framework already demonstrated for terrestrial and oceanic applications.**

  - **Facilitating the broader utilization of Earth observations for greater societal benefits within coastal zones and enhancing CEOS engagement with external stakeholders such as GEO, IOC/GOOS, UN Environment, WMO and the UN Decade of Ocean Science for Sustainable Development (2021-2030).**

- CEOS Coastal Observations and Applications Study Team (CEOS-COAST): Initiated at 33rd CEOS Plenary to define a CEOS Coastal Strategy and facilitate engagement with key international stakeholders

  - **Bridging land and aquatic observations within CEOS, helping to integrate across multiple CEOS entities and domains, both thematic and technical.**

  - **Leveraging CEOS Systems, services and interoperability approaches, including the CEOS Analysis Ready Data (ARD) framework already demonstrated for terrestrial and oceanic applications.**

Goals of this session (Emphatic yes to All!):
- Review outcomes from the SST-VC ocean domain ARD survey and what CEOS ARD for SST might look like.
- Share the work on the USGS Aquatic Reflectance Demonstration Product and related Product Family Specification
- Explore what changes may be needed to the CARD4L Framework to allow it to go 'beyond land'; agree on those steps.
- Present and discuss the proposal to have LSI-VC work with other VCs to generalise the CARD4L Definition so that it is applicable to all domains.
- Identify options to expand the CEOS ARD for Land Framework to Oceans and Atmosphere. Identify leads from non-land areas (Coasts, Ocean, Atmosphere) to work through CEOS ARD for those areas.

- *What*: Successful realization and utilization of trans-boundary Analysis Ready Data (ARD) will be one of the unifying factors & primary engines that drives the COAST effort forward (NB: heritage/non-ARD data sets are of vital importance as well)

- *Why*: Because there have been significant challenges to date, and more so great opportunities looking ahead, in bridging and coordinating geophysical satellite data across the land-sea (~aquatic) <u>interface</u>, as well as *in situ* and diverse biological (and ultimately socio-economic) data sets – ARD provides a common framework!

- *How*: As COAST moves toward implementation, we will guide and leverage current and planned capabilities and approaches within CEOS for ARD, i.e., SST-VC, CARD4L/LSI-VC, USGS Aquatic Reflectance Demo, COVERAGE and WGISS/SEO
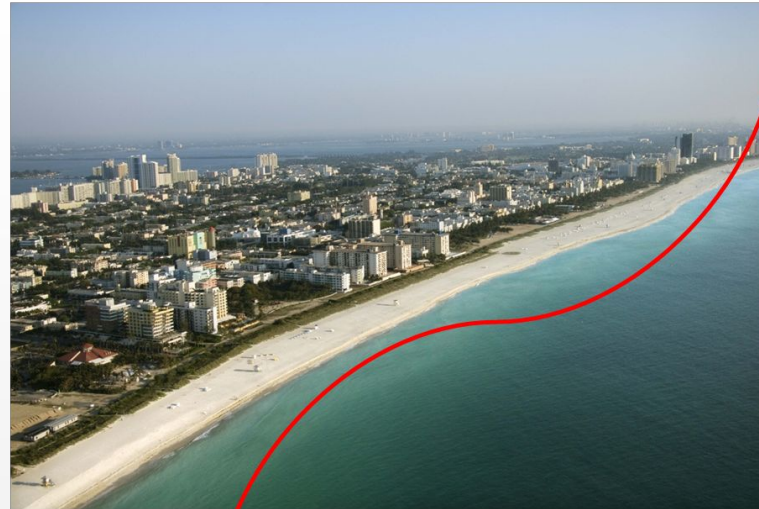
Cross-Cutting Needs:
- Analysis ready data
- Tools, products & services
- User-centric web portals

Products needed:
- Land cover/use (impervious surfaces)
- Shoreline mapping/elevation
- Precipitation and Discharge
- Sediment and Nutrient loadings
- Habitat/water quality maps
- et al.

Partners/Stakeholders
- Blue Planet
- AquaWatch
- UN Environment
- IOC/WMO

Products needed:
- Land cover/use
- Bathymetry/elevation
- Shoreline mapping
- Waves and Tides
- Flood Maps
- et al.

COAST Project Component
Land to Sea Impacts (~ biological/ecological)

COAST Project Component:
Sea to Land Impacts (~ physical forcing)

Ecosystems, Water Quality & Habitats

- Sediment loading (benthic habitat impacts)

- Coastal eutrophication (SDG 14.1.1. et al.)

Coastal Disasters/Hazards: Flooding & Inundation

- Large-scale coastlines: urbanized, rural/agricultural, mixed use

- Small-island states: Coral-reef lined

COAST - all about the interfaces: Processes, Data, Geography, Science & People

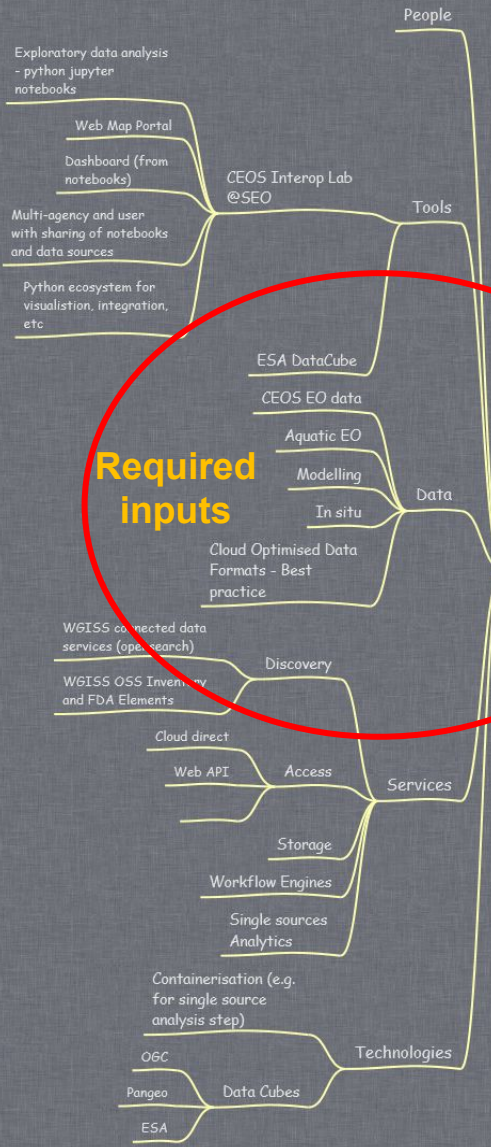CEOS COAST Tools and Services: Stage 1 — mind map

Outcomes — Not today!

Agree Principles
- Demonstrate capability not build a production system
- Validate solutions through implementation experiments
- Demonstrate *interoperability/ integratability* between CEOS agencies not point solutions
- Demonstrating integration will need to remain tractable given the many challenges being hit *but* solutions should be developed on the basis of going global
- Keep in mind advancing the globally applicable goal and the identification and removal of barriers rather than a wizz bang point solution

Objectives (All objectives relate to the specifics of the COAST Pilot projects)
- Demonstrate Discovery, access and Integration of key EO datasets
- Demonstrate ARD Discovery and Use for COAST applications
- Demonstrate Discovery and access to in situ
- Discovery best practices for collaborative CEOS project development in COASTal regions

Approach
- Use CEOS Open Search to find relevant data — L2 USGS and ESA for EO data (others?)
- Use a data *pipeline* to acquire (or use in place) ARD
- Use data cube technology to integrate multiple EO datasets for common analysis — Sat to Sat data

Experiments (Inventory minimum outcome)
- Use python data handling tools to obtain and prepare insitu data that "matches the EO data" — Will be ad hoc in many cases / Likely to lead to ongoing work on insitu data Discovery and Interoperability across cEOS Agencies (and beyond?)
- Examples of python data analysis tools for integration of the above — Sat to in situ data / data to modelling output
- Comparison of algorithm(s) applied to different test sites where similar data (in situ, EO, modelling) are available
- Derived products available for visualisation (either from storage or ondemand analysis)
- Where possible all experiments should utilise multiple agency sources and analysis services in an interoperable system
- We may only get as far as inventorying the different approaches used in each agency pilot and seeing what is common. We may not implement a common approach in Stage I but ultimately that is the goal.

Required inputs — Inventory of Resources

People

Tools
- CEOS Interop Lab @SEO
  - Exploratory data analysis - python jupyter notebooks
  - Web Map Portal
  - Dashboard (from notebooks)
  - Multi-agency and user with sharing of notebooks and data sources
  - Python ecosystem for visualisation, integration, etc

Data
- ESA DataCube
- CEOS EO data
- Aquatic EO
- Modelling
- In situ
- Cloud Optimised Data Formats - Best practice

Discovery
- WGISS connected data services (open search)
- WGISS OSS Inventory and FDA Elements

Services
- **Access**
  - Cloud direct
  - Web API
  - Storage
  - Workflow Engines
  - Single sources Analytics
- **Technologies**
  - Containerisation (e.g. for single source analysis step)
  - OGC
  - Pangeo
  - ESA
  - Data Cubes

What will make COAST happy? Trans-boundary CEOS ARD!

CEOS-COAST & CEOS ARD

Key findings per CEOS-COAST White Paper #3: Cross-Cutting Tools, Systems & Services:

- COAST should leverage the CEOS Analysis Ready Data (ARD) framework, WGISS Interoperability Lab, Virtual Constellations, GHRSST, COVERAGE et al. developed approaches, tools, and services as available/fit for purpose

- COAST should embrace…. the level of data that application developers desire, referred to as Analysis and Access of ARD in the CEOS Interoperability Terminology Report

- Many COAST relevant ARD discussions were initiated or continued within CEOS sub-communities (CEOS/WGISS, GEO AquaWatch, LSI-VC, SST-VC) with COAST a likely benefactor of streamlined, open source, formats trending toward achievement of FAIR data principles (Findable, Accessible, Interoperable, and Reusable/Reproducible)

- GEO AquaWatch, in association with COAST, is providing recommendations to CEOS for core ARD requirements to be adopted for the aquatic side - including addressing when available land requirements / approaches incompletely characterize or fulfill aquatic data product needs, or are not applicable, at least not w/o corollary modifications to suit the aquatic realm. This includes actively supporting the Aquatic Reflectance Product Family Specification (PFS) Science Expert Review Panel led by N. Pahlevan, S. Labahn, C. Barnes and A. Siqueira.
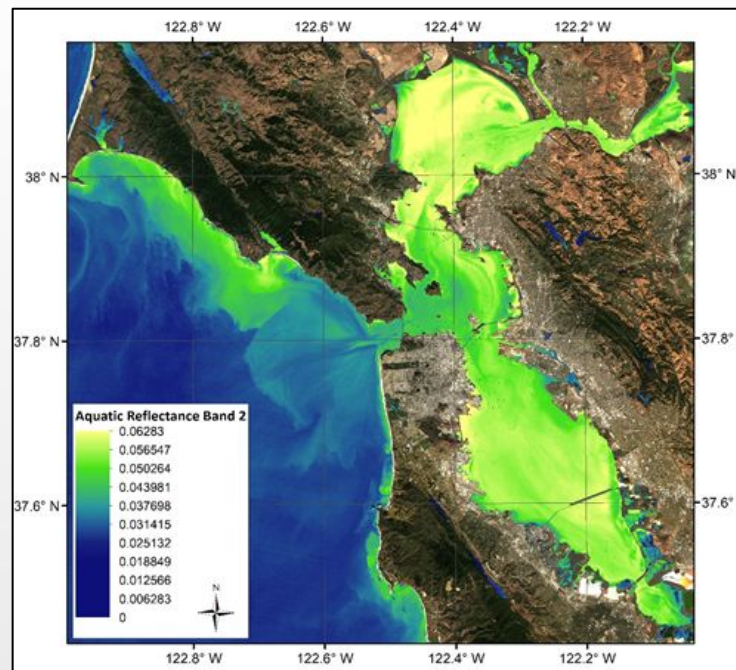
Introduction

- **Landsat Aquatic Reflectance is a provisional on-demand USGS product that provides nondimensional normalized remote-sensing reflectance for coastal and inland waters, assuming a perfectly Lambertian surface**

- ***Region*: Global**

- ***Sensor*: Landsat 8 Operational Land Imager**

- ***Date availability*: April 2013 - present**

- ***Application*: coastal regions compared to highly eutrophic/turbid inland waters**

Landsat 8 Provisional Aquatic Reflectance Science Product
OLI Band 2 (Blue band) San Francisco Bay, CA

- **Potential to make a valuable contribution to:**
  - Environmental monitoring capabilities for coastal and inland aquatic ecosystems
  - Aquatic science

History of Algorithm Development

- **Principal Investigator and USGS-NASA 2018-2023 Landsat Science Team member: Dr. Nima Pahlevan (SSAI at NASA/GSFC)**

- **The atmospheric correction algorithm is directly derived from the Sea-viewing Wide Field-of-View Sensor (SeaWiFS) Data Analysis System (SeaDAS) package distributed by NASA's Ocean Biology Processing Group**

- **The remote-sensing reflectance output of SeaDAS is multiplied by Pi (i.e., ~3.142) to obtain a dimensionless reflectance referred to as Aquatic Reflectance**
  - PI recommendation to make the product unitless
  - "Aquatic Reflectance" is an evolving product in the U.S. and international aquatic communities

Product Characteristics

- **The Aquatic Reflectance product is available for Landsat 8 visible bands only (i.e., Bands 1-4)**

- **The near infrared (NIR) and short-wave infrared (SWIR) bands are consumed by atmospheric compensation and hence are not delivered in the final product (PI recommendation)**

- **The processing flags (containing detailed information about the atmospheric correction), Landsat Level-2 Pixel Quality Assessment (QA), and metadata are also delivered within the product file package**

Description	Range	Valid Range	Scale Factor	Unit	Fill Value	Data Type
Aquatic Reflectance (bands 1 to 4)	-4,720 to 31,420	0 to 31,420	0.00001	Unitless	-9999	INT16
Processing Flags	0 to $2^{31}-1$	0 to $2^{31}-1$	NA	NA	-9999	INT32
Level 2 Pixel Quality Assessment	0 to 65535	1 to 2047	NA	Bit Index	1 (bit 0)	UINT16
Level 2 XML Metadata file	NA	NA	NA	NA	NA	NA

Pixel QA Processing Flags

Bit	Attribute	Description
0	ATMFAIL	Atmospheric correction failed
1	LAND	Designated by the input land file (off for Landsat 8 processing)
2	PRODWARN	Intermediate product value falls outside the min/max range
3	HIGLINT	If glint coefficient > glint threshold (0.005). This is not checked for land or night pixels.
4	HILT	Flags high total radiance (Lt) measured by sensor (not used)
5	HISATZEN	View angle > threshold (60 deg)
6	COASTZ	Controlled via the bathymetry file, based on the latitude/longitude of the current pixel
7	SEADAS_CLOUD	Cloud based on SeaDAS algorithm
8	CLOUD_SHADOW	Cloud shadow based on the cloud mask algorithm
9	CLOUD	Cloud based on the cloud mask algorithm
10	COCCOLITH	Flagged based on the coccolithophore algorithm coefficients, detects coccolithophore blooms. (Coccolithophores are a group of phytoplanktons that inhabit a wide variety of marine environments and are distinctive by their production of small calcium plates or coccoliths which are organized around each living cell as an outer covering.)

Bit	Attribute	Description
11	TURBIDW	Flags turbid water. If LwB4 / FonomB4 > 0.0012 Where LwB4: water-leaving band 4 value FonomB4: nominal F0 value for band 4 F0: the mean solar flux
12	HISOLZEN	Solar angle > threshold (70 deg)
13	Unused	NA
14	LOWLW	If water-leaving radiance band 3 (LwB3) value < 0.15 (nlwmin)
15	CHLFAIL	Flags a failure in the chlorophyll calculation
16	NAVWARN	Mainly used for fill pixels
17	Unused	NA
18	RRSWARN	R_{rs} < min_valid or R_{rs} > max_valid
19	MAXAERITER	Number of iterations for the aerosol corrections exceeds the maximum threshold (10)
20	MODGLINT	glint coefficient > minimum glint threshold (0.0001)

Bit	Attribute	Description
21	CHLWARN	chlorophyll < min_valid or chlorophyll > max_valid
22	ATMWARN	If the value is invalid, then set ATMWARN; also set if MAXAERITER is set. Based on ancillary QC flags (ozone, water vapor, RH, wind speed, surface pressure)
23	Unused	NA
24	SEAICE	Controlled via the ice file, based on the latitude/longitude of the current pixel
25	NAVFAIL	If there is an invalid latitude/longitude value
26	FILTER	
27	Unused	NA
28	Unused	NA
29	HIPOL	If degree of polarization for any band is > hipol (0.5)
30	PRODFAIL	Processing failure (data values are outside the legal allowed, too many bands, illegal values, etc.)
31	Unused	NA

Data Access

- **Now available on-demand through the USGS EROS Science Processing Architecture (ESPA) for Landsat 8 Collection 1 and eventually with Landsat 8 Collection 2 in mid-2021**
  - https://espa.cr.usgs.gov/

Other Resources

- **USGS Landsat Missions Website: usgs.gov/landsat**
  - Package Content, Auxiliary Data, Caveats and Constraints & Documentation
  - Contact custerv@usgs.gov for additional product support

CARD4L
Product Family Specification

- **Aquatic Reflectance CARD4L PFS currently in development**
  - Lead by: **Nima Pahlevan** (NASA) & **Chris Barnes** (USGS/ KBR)
  - Coordinated by: **Andreia Siqueira** (GA)
  - Expert Review Panel from the aquatic user community

- RIT	**- EPA**	**- GEO Aquawatch**
- NASA	**- Kent State**	**- USGS**
- U of MN	**- NOAA**	
- ESA	**- EO Map Australia**	
- CSIRO	**- IVOS**	

- **Met with review panel in June and September, 2020**
  - Further discussion on atmospheric adjacency effects and BRDF
  - Working on scheduling another review panel
- **Next steps**
  - Tech edit and distribute final draft PFS to LSI-VC for review
  - Review and endorse final AR PFS at LSI-VC-11 (target)

- Beyond Land: Options to expand the CEOS ARD for Land Framework to Oceans and Atmosphere

  o Improving/Testing the Definition

  o Additions to the framework (e.g., advisory notes).

  o Options to trial the framework for Coastal, Ocean and Atmosphere data products

  o Improving ARD services and deployment

Additional materials

CEOS Analysis Ready Data for Land (CARD4L) are satellite data that have been processed to a minimum set of requirements and organized into a form that allows immediate analysis with a minimum of additional user effort and interoperability both through time and with other datasets.

The definition of CARD4L is not exclusive or prescriptive. A range of data products will be produced by CEOS Agencies to meet the needs of a diverse user community and many of those products are analysis ready data to their users. However, this definition of CARD4L reflects the attributes of fundamental measurement products for the majority of global remote sensing users with land imaging applications, and are the minimum level required to support time series analysis and data interoperability. <u>CARD4L will therefore be geophysical measurements that are comparable in space and time, with sufficient per-pixel (observation) metadata to enable users to select 'observations of interest' for their analyses.</u>