Data Quality at NOAA/NESDIS

NOAA
National Satellite and Information Service

WGISS-52 Virtual
October 21, 2021

Nancy Ritchey[1] and Lihang Zhou[2]

[1]Chief, Archive Branch, Data Stewardship Division, NCEI
[2]JPSS Data Products & Management Services, Deputy and Product Portfolio Manager

Executive Summary

- Challenge is to provide sufficient and consistent QA/QC information in a machine readable and actionable format, and curation of that information
- Recent progress on community driven efforts to develop QA/QC best practices

- Consistent QA/QC and Cal/Val processes across NESDIS satellite programs, promote use of best practices, incorporate lessons learned and seek feedback from peers, users and stakeholders
- Extensive documentation of QA/QC and Cal/Val on Program websites, including some operational product monitoring (ICVS)
- Overall product QA/QC and maturity information is reported in the metadata
- Data stewardship maturity assessments are utilized for some products

The Case for Data Quality Information

- The quality of data and information is important to support informed decisions on the (re)use of data

- Data quality information must be represented and communicated consistently to inform the use of the data

- Data quality information must be curated along with the data

- Currently no international standards or community guidelines for promoting the representation and sharing of quality information

Recent Progress

- [Call to Action for Global Access to and Harmonization of Quality Information of Individual Earth Science Datasets](#) (04May2021)

- [International Community Guidelines for Sharing and Reusing Quality Information of Individual Earth Science Datasets](#) (01Oct2021)

- SciDataCon 2021 session on *The State of Documenting and Reporting Data and Information Quality for Supporting Open Science* (19Oct2021)

- Research Data Alliance Annual meeting session on *Representing and Communicating Data Quality Information* (04Aug2021)

NOAA Satellite Data Quality - Process

- NESDIS Science Teams complete extensive QA/QC and Cal/Val

- Periodic Peer/Stakeholder Product Validation Reviews

- Annual User/Stakeholder feedback meetings

- Cal/Val activities planned and baselined annually, schedule tracked and reported routinely, risk and mitigation are identified and documented

NOAA Satellite Data Quality - Tools

- Integrated Calibration/Validation System (ICVS) monitoring capabilities
- Product maturity assessment and overall QA/QC flag is captured in the metadata
- More detailed information is reported for some products
- Some satellite products have undergone Data Stewardship Maturity Assessments, results are captured in the metadata, and available via OneStop

Satellite Data Quality - Examples

- Examples of available QA/QC and Cal/Val information
  - [JPSS/SNPP algorithm maturity](#)
  - [JPSS/SNPP product operational matrix](#)
  - [GOES-16 Peer/Stakeholder Product Validation Reviews](#)
  - [MiRS Data Quality Monitoring](#)
  - JPSS Product Cal/Val [plans](#)
  - DSMM assessment results for [JPSS ATMS SDR](#)
- Additional tools, information and assessment matrices follow

Integrated Calibration/Validation System (ICVS)

http://www.star.nesdis.noaa.gov/icvs/
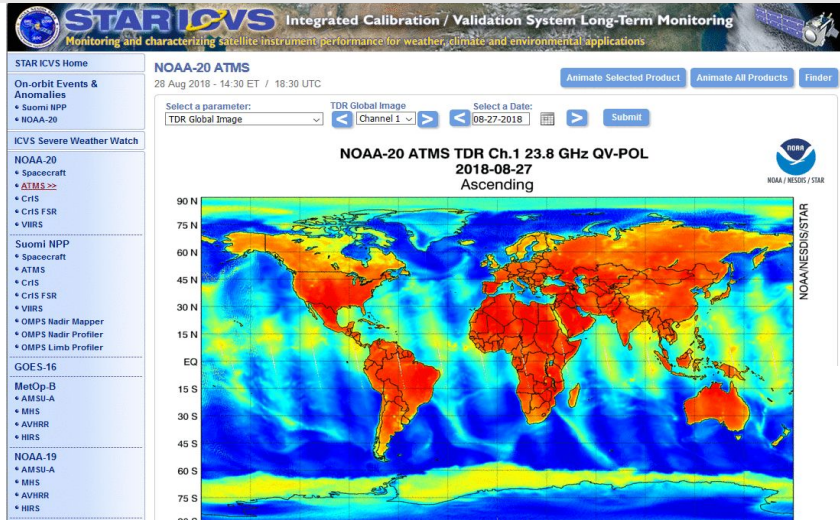
- Web-based system for continuous monitoring of S-NPP/JPSS and other satellite sensors' performance and the associated data product quality.

- New ICVS monitoring capabilities include
- Simultaneous Nadir Overpass Inter-sensor comparison
- Double Difference (DD) Monitoring: (O-B)sensor1 - (O-B)sensor2
- ICVS Clear/Sky Mask Machine-Learning Algorithm
- ICVS Severe Event Watch (iSEW) System
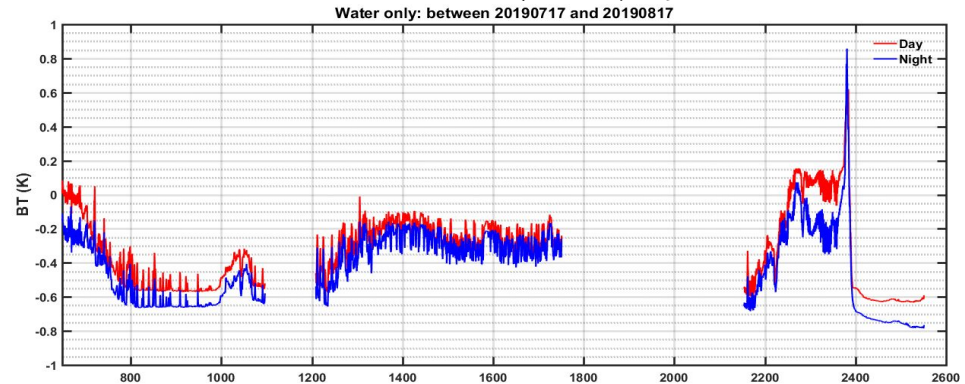
SNPP/NOAA-20 CrIS Data 32-Day Averaged Difference

NOAA Satellite and Information Service

JPSS Algorithms Maturity Matrix

- **Validation Data Sources:**
  - Ground-based Measurements; Campaigns of Opportunity (Field Campaigns), Model Analysis Fields (ECMWF, GFS); Correlative observations from other satellite sensors, SNOs; RT simulations, other algorithms
  - Lessons learned, setting up of validation data sources, validation and visualization tools developed for S-NPP have helped to expedite NOAA-20 product maturity

Beta—the product is minimally validated and may still contain significant errors, based on product quick looks using initial calibration parameters.

Provisional—product performance has been demonstrated through a large, but still (seasonally or otherwise) limited, number of independent measurements. The analysis is sufficient for limited qualitative determinations of product fitness-for-purpose, and the product is potentially ready for testing operational use.

Full—product performance has been demonstrated over a large and wide range of representative conditions, with comprehensive documentation of product performance, including known anomalies and their remediation strategies. Products are ready for operational use.

JPSS/SNPP Data Product Operational Matrix

Sensor	Data Product	Priority	ATBD	S-NPP Operational	S-NPP Maturity	NOAA-20 Operational	NOAA-20 Maturity
ATMS	ATMS TDR/SDR	1	●	✔	Validated	✔	Validated
CrIS	CrIS SDR	1	●	✔	Validated	✔	Validated
VIIRS	VIIRS SDR	1	●	✔	Validated	✔	Validated
OMPS	OMPS Total Column SDR	3	●	✔	Validated	✔	Validated
OMPS	OMPS Nadir Profiler SDR	3	●	✔	Validated	✔	Validated
OMPS	OMPS Total Column Ozone EDR	3	●	✔	Validated	✔	Validated
OMPS	OMPS Nadir Profiler Ozone EDR	3	●	✔	Validated	✔	Provisional
VIIRS	VIIRS Imagery	1	●	✔	Validated	✔	Validated
VIIRS	Ocean Color	2	●	✔	Validated	✔	Provisional
VIIRS	Sea Surface Temperature	2	●	✔	Validated	✔	Validated
VIIRS	VIIRS Polar Winds	2	●	✔	Validated	✔	Validated

Consistent Cal Val Process for S-NPP, NOAA-20 (and extended to GOES-16/17):

SDR/EDR Products - -> Beta - -> Provisional - -> Validated

Validation Maturity and Documentations

Validated Maturity End State	Assessment
Product performance has been demonstrated over a large and wide range of representative conditions (i.e., global, seasonal).	Validation against 4 months of global AERONET data
Comprehensive documentation of product performance exists that includes all known product anomalies and their recommended remediation strategies for a full range of retrieval conditions and severity level.	Yes
Product analyses are sufficient for full qualitative and quantitative determination of product fitness-for-purpose.	Yes
Product is ready for operational use based on documented validation findings and user feedback.	Product is in operations and user feedback did not reveal any issues
Product validation, quality assurance, and algorithm stewardship continue through the lifetime of the instrument	Yes

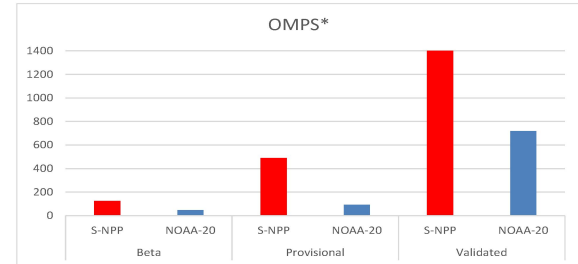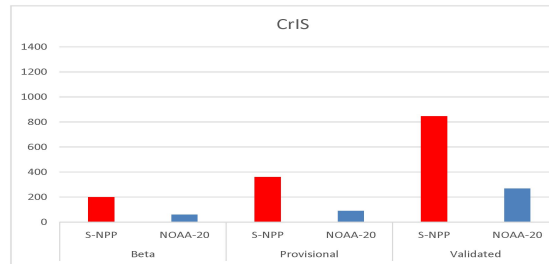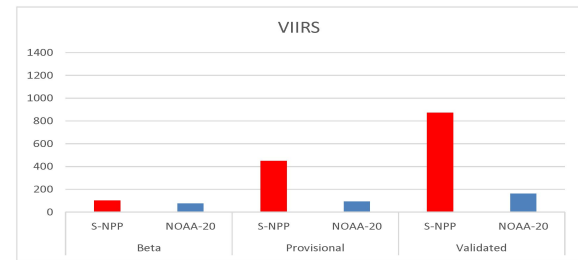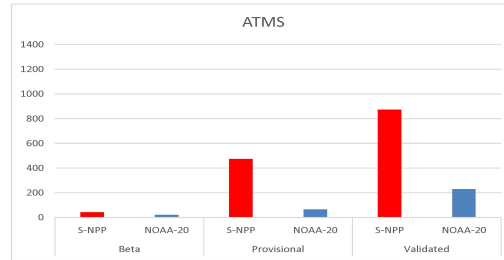| Science Maturity Check List	Yes ?
ReadMe for Data Product Users	Yes
Algorithm Theoretical Basis Document (ATBD)	Yes
Algorithm Calibration/Validation Plan	Yes, in preparation for J2
(External/Internal) Users Manual	Yes
System Maintenance Manual (for ESPC products)	Yes
Peer Reviewed Publications (Demonstrates algorithm is independently reviewed)	Yes
Regular Validation Reports (at least annually) (Demonstrates long-term performance of the algorithm)	As required

Review Team Chaired by NESDIS and Program Chief Scientist; composed of User Representatives, Product Operation Management, and Program Stakeholders… Users' feedback is required for the review

NOAA-20 SDR Maturity Achieved in Record Time

Key to the Successful Transition:

- Sensors are well-understood by the science teams

- Tools tested with S-NPP, Proxy data available, Algorithms baselined

- Mission-agnostic EDR algorithms: Once the SDRs were calibrated validated, the EDRs follow closely behind

- Well coordinated science teams

- Lessons learned from SNPP applied

Suomi NPP vs NOAA-20 Sensor Data Records (SDRs) Maturity Review Timeline Comparison

https://www.star.nesdis.noaa.gov/jpss/AlgorithmOperational.php

NOAA Satellite and Information Service

Product Maturity Matrix

CDR Name Here

maturity level as of mm/dd/yyyy

Climate Data Record (CDR) Maturity Matrix

Maturity	Software Readiness	Metadata	Documentation	Product Validation	Public Access	Utility
1	Conceptual development	Little or none	Draft Climate Algorithm Theoretical Basis Document (C-ATBD); paper on algorithm submitted	Little or None	Restricted to a select few	Little or none
2	Significant code changes expected	Research grade	C-ATBD Version 1+ ; paper on algorithm reviewed	Minimal	Limited data availability to develop familiarity	Limited or ongoing
3	Moderate code changes expected	Research grade; Meets int'l standards: ISO or FGDC for collection; netCDF for file	Public C-ATBD; Peer-reviewed publication on algorithm	Uncertainty estimated for select locations/times	Data and source code archived and available; caveats required for use.	Assessments have demonstrated positive value.
4	Some code changes expected	Exists at file and collection level. Stable. Allows provenance tracking and reproducibility of dataset. Meets international standards for dataset	Public C-ATBD; Draft Operational Algorithm Description (OAD); Peer-reviewed publication on algorithm; paper on product submitted	Uncertainty estimated over widely distributed times/location by multiple investigators; Differences understood.	Data and source code archived and publicly available; uncertainty estimates provided; Known issues public	May be used in applications; assessments demonstrating positive value.
5	Minimal code changes expected; Stable, portable and reproducible	Complete at file and collection level. Stable. Allows provenance tracking and reproducibility of dataset. Meets international standards for dataset	Public C-ATBD, Review version of OAD, Peer-reviewed publications on algorithm and product	Consistent uncertainties estimated over most environmental conditions by multiple investigators	Record is archived and publicly available with associated uncertainty estimate; Known issues public. Periodically updated	May be used in applications by other investigators; assessments demonstrating positive value
6	No code changes expected; Stable and reproducible; portable and operationally efficient	Updated and complete at file and collection level. Stable. Allows provenance tracking and reproducibility of dataset. Meets current international standards for dataset	Public C-ATBD and OAD; Multiple peer-reviewed publications on algortihm and product	Observation strategy designed to reveal systematic errors through independent cross-checks, open inspection, and continuous interrogation; quantified errors	Record is publicly available from Long-Term archive; Regularly updated	Used in published applications; may be used by industry; assessments demonstrating positive value

1 & 2	Research
3 & 4	IOC
5 & 6	FOC

CDRP-MTX-0008 V4.0 (12/20/2011)

Data Stewardship Maturity Matrix Defines Measurable, Five-Level Progressive Practices in Nine Quasi-Independent Key Components

Maturity Scale / Key Component	Level 1 - Ad Hoc Not Managed	Level 2 - Minimal Managed Limited	Level 3 - Intermediate Managed Defined, Partially Implemented	Level 4 - Advanced Managed Well-Defined, Fully Implemented	Level 5 - Optimal Level 4 + Measured, Controlled, Audit
Preservability	The state of being preservable				
Accessibility	The state of being publicly searchable and accessible				
Usability	The state of data product being easy to understand and use				
Production Sustainability	The state of data production being sustainable and extendable				
Data Quality Assurance	The state of data product quality being assured/screened				
Data Quality Control /Monitoring	The state of data product quality being controlled and monitored				
Data Quality Assessment	The state of data product quality being assessed				
Transparency /Traceability	The state of being transparent, trackable, and traceable				
Data Integrity	The state of data integrity being verifiable				

NOAA

Questions?