P578 EFFICIENT ACCESS TO RAW MEASUREMENTS AND PROCESSING COEFFICIENTS FOR NPOESS PREPARATORY PROJECT (NPP) AND JOINT POLAR SATELLITE SYSTEMS (JPSS) SENSOR DATA

Drew Saunders*[1], Jim Biard[1], Art Burden[1], Jeff Privette[2], Dan Baldwin[3], and Linda Copley[1]
1 - STG, Inc. Asheville, NC
2 - National Climatic Data Center (NCDC), Asheville, NC
3 - University of Colorado, Boulder, CO

1. ABSTRACT

The National Polar-orbiting Operational Environmental Satellite System (NPOESS) Preparatory Project (NPP) and Joint Polar Satellite System (JPSS) operational data products will provide global coverage of environmental conditions associated with the Earth's weather, atmosphere, oceans, land, and near-space environment. The detailed and comprehensive observations that will be acquired during these missions over the coming decades will provide valuable input for monitoring changes in climate. The National Oceanic and Atmospheric Administration (NOAA) Climate Data Record (CDR) Program at the National Climatic Data Center (NCDC) is mandated to produce data products from these and other satellites that are of sufficient quality and integrity to be used for climate research. CDR development requires the careful integration of current data with archived data from multiple sources, necessitating repeated reprocessing as discrepancies between measurements are resolved and changes in sensor performance are detected.

For the production of CDRs, researchers need to process from raw measurements. The raw NPP measurements or Raw Data Records (RDR) are inherently difficult to use for reprocessing. To provide raw data that are easier to process NCDC will create Climate RDRs. These Climate RDRs will be formatted as standard netCDF-4 files and will contain raw measurements and the coefficients required for converting the measurements to basic scientific and engineering units. The Climate RDRs will make the raw measurements more easily and efficiently accessible to a wider range of users, and will make reprocessing of NPP CDRs faster and simpler.

Climate RDRs will be produced for the Advanced Technology Microwave Sounder (ATMS), the Cross-track Infrared Sounder (CrIS), the Ozone Mapping and Profiler Suite (OMPS) Nadir Profile, and the Visible/Infrared Imager/Radiometer Suite (VIIRS). The Climate RDRs and CDRs will be archived and available from the NOAA Data Centers.

* *Corresponding author address:* Drew Saunders, NOAA National Climatic Data Center, 151 Patton Ave, Asheville, NC 28801;
e-mail: drew.saunders@noaa.gov

2. INTRODUCTION

2.1 Background

The study of climate change is a challenging undertaking. Trends of less than 1% per decade in a fundamental quantity such as sea-surface temperature, solar output, or cloud cover (for example) could correspond to massive changes in the Earth's environment over long time scales; yet such a signal must be detected in the presence of higher frequency natural variability that is orders of magnitude larger. In order to detect changes of this sort, it is critical to develop high quality, long time-base, wide coverage measurements of environmental variables that are useful for climate change monitoring and research. A measurement dataset that meets such criteria has been given the name Climate Data Record (CDR) (NRC 2001).

CDRs are a valuable international resource, and it is vital that they be managed and preserved. Proper stewardship and archival of these datasets can advance the pace and quality of climate research. Without it, various problems – such as changes in government funding, failure of a computer system, etc; could result in major setbacks.

2.2 The NOAA Climate Data Record Program

The National Research Council (NRC) in 2004 tasked NOAA to "embrace its new mandate to understand climate variability and change by

asserting national leadership for satellite-based Climate Data Record generation, applying new approaches to generate and manage satellite Climate Data Records; developing new community relationships, ensuring long-term consistency and continuity for a satellite Climate Data Record generation program." (NRC 2004) The NCDC Climate Data Record Program (CDRP) is part of NOAA's response.

The CDRP is an outgrowth of the Scientific Data Stewardship (SDS) Program, which had the goal "to ensure that satellite climate data are processed, archived, and distributed to users in a manner that is scientifically defensible for monitoring, diagnosing, understanding, predicting, modeling, and assessing climate variation and change." (Bates 2004) The CDRP is tasked with identification of environmental variable datasets that qualify as CDRs, and with stewardship and archival of those CDRs.

In order to satisfy the requirements of its mandate, NOAA must also place the input data needed for the production of the CDRs under stewardship and archival (NRC 2004).

2.3 The NPP Satellite

The NPP satellite is the first in a series of new polar-orbiting weather and climate observing platforms managed by the NPOESS and JPSS programs. NPP provides risk reduction with an opportunity to demonstrate and validate new instruments and processing algorithms, as well as to demonstrate and validate aspects of command, control, communications and ground processing capabilities prior to the launch of the first "full-scale" satellite in the series. The NPP satellite also ensures the continuity of critical environmental measurements, covering the gap between existing and future satellite missions (Leibee 2003).

The instruments on the NPP satellite are: the Advanced Technology Microwave Sounder (ATMS), the Cloud and Earth Radiant Energy System (CERES), the Cross-track Infrared Sounder (CrIS), the Ozone Mapping and Profiler Suite (OMPS), and the Visible/Infrared Imager/Radiometer Suite (VIIRS).

All of the operational data from the NPP satellite are scheduled to be archived. The raw sensor data files that are being archived are known as Raw Data Record (RDR) files, and contain the unmodified Consultative Committee for Space Data Systems (CCSDS) structured application packets acquired through the high-rate telemetry downlink from the NPP satellite. Each NPP RDR file contains one full measurement cycle, or granule, of raw sensor data. The sizes and production rates for the NPP RDR granules are found in Table 1.

Sensor	RDR Granule Size (kilobytes)	RDR Granule Period (sec)
ATMS	1108.41	32
CERES	1370.84	660
CrIS	5741.63	32
OMPS NP	1310.10	37.44
OMPS LP	5,240.13	37.44
VIIRS	236368.63	85.74

Table 1: NPP RDR Granule Sizes and Periods

3. CLIMATE RAW DATA RECORDS

3.1 Purpose

As the CDRP team at the NCDC began its analysis of CDR stewardship and production requirements, they determined that providing improved access to raw sensor data measurements and processing coefficients would provide a number of benefits to the program and the user community.

Throughout the life of an environmental observation satellite mission and beyond, sensor and algorithm knowledge increases as the system is studied. This enables more accurate calibration of the raw measurements as they are processed into higher level products (Privette 2008). To take full advantage of improvements in calibration, portions of CDR datasets must be periodically reprocessed from the raw, uncalibrated inputs. Additionally, researchers who are attempting to verify the accuracy of CDRs, analyze anomalies in the data, or develop improved calibrations for sensors will also need access to the raw measurements and processing coefficients.

Historically, the raw input data have been difficult to obtain and difficult to use. Extensive knowledge of the details of raw sensor data formats and satellite operational modes is required to successfully extract a raw measurement time series from a raw sensor data file. In some cases, specialized compression schemes have been used to maximize the information content in the data stream, demanding even more detailed knowledge of the format and processing in order to access usable raw measurements. Dealing with the bit ordering, or "endian-ness", of the raw data versus that of the computer system on which the data are being processed further complicates matters. The processing coefficients may also be stored in a non-standard binary format and be similarly difficult

to use.

The CDRP has decided to produce Climate Raw Data Records (C-RDRs) as a way to address these barriers to using raw measurements and processing coefficients. A C-RDR will be a dataset composed of raw measurements and processing coefficients that have been organized by measurement and packaged in a widely-used, platform-independent, binary data format. Processing (such as decompression) that is required to make original raw measurement values accessible will be performed during creation of C-RDRs, but no conversions of measurements to units will be done. C-RDRs will contain metadata that will clearly describe each raw measurement and processing coefficient set and identify its origin.

C-RDRs will not only make it easier for a wider range of users to access raw measurements and processing coefficients, they will improve the long-term usability of the data. As decades pass, it will become increasingly hard for someone without extensive domain knowledge to make use of the RDR datasets. By adhering to widely-adopted community standards regarding format and metadata, C-RDRs will be more likely to be accessible through backwards-compatibility features in software packages, and easier to port to new formats.

3.2 Format and Structure

C-RDR files will be written in network Common Data Form 4 (netCDF-4) format, which is binary, hierarchical, and self-describing. It is a specialization of the Hierarchical Data Format, version 5 (HDF5). NetCDF-4 features platform independence, efficient access, and user-transparent, lossless data compression. C-RDR files will make use of the group construct in netCDF-4, which allows the structure of the files to reflect the organization of the measurements.

The weather and climate research community makes wide use of netCDF, and have developed a number of usage metadata conventions for data stored using this format. The C-RDR files will conform to the Climate and Forecast (CF) convention, which specifies naming rules and metadata elements (known as attributes) to associate with each variable (the netCDF term for a unique set of measurement values).

In addition to the variable attributes specified by the CF convention, C-RDR raw measurement variables will each have an attribute named “source”. The source attribute is the key to establishing a strong and detailed link between the contents of the RDR files and the contents of the C-RDR files. It will be a character string that will clearly identify the location within the raw data stream of each measurement contained within the variable.

3.3 Metadata

Each C-RDR collection will have a fully compliant set of ISO 19115 metadata to aid in stewardship and archival. At the collection level, the metadata will be stored in eXtensible Markup Language (XML) files. Certain ISO 19115 metadata elements describe the contents of individual data files, and it makes sense to store those in the data files themselves.

At the data file level, there is also a netCDF convention for stewardship and archival - the Attribute Convention for Dataset Discovery (ACDD). There are applications developed for cataloging and distribution of netCDF files, and they use the ACDD elements. By including ACDD elements in the C-RDR files, their usability will be enhanced. The ACDD elements map directly into ISO 19115 elements, but the element names are different, and they are a subset of the elements that C-RDR files are likely to contain.

To avoid duplication of metadata elements within the C-RDR files, the current plan is to take a hybrid approach to the file-level metadata. C-RDR files will contain the full set of relevant ACDD elements, using the names specified by the convention. The files will also include ISO 19115 elements that are selected for inclusion at this level and that are not covered by the ACDD elements. The ISO 19115 elements will have names that are traceable to the standard. (Since ISO 19115 metadata are hierarchical and complex, they sometimes cannot be represented natively in netCDF-4 without resorting to complex types, which the CDRP team has decided to avoid.) A simple XML stylesheet-based translator can be used to generate XML output containing all of the file-level metadata in ISO 19115 format with ISO 19115 names.

4. NPP C-RDRS

4.1 Datasets

Sensors from four of the five instruments that make up the NPP payload were chosen for C-RDR production. They are ATMS, CrIS, OMPS Nadir Profile (OMPS NP), and VIIRS. The CERES instrument and OMPS Limb Profile sensor may be added in a later phase. The C-RDR data files for

each instrument will include a spacecraft diary, which is the record of satellite position, velocity, and attitude during the time span covered by the measurements in the file. There will also be one or more C-RDR files containing processing coefficients for each instrument. These files will contain the lookup tables and calibration equation coefficients needed to convert the instrument measurements from digital numbers to physical units. The C-RDR data files will be produced as the RDR files are received. The C-RDR coefficient files will be produced as new coefficients are made available by the NPP project.

The sizes of the NPP RDR files (one granule per file) were driven by the real-time operational requirements of the NPP Interface Data Processing Segment (IDPS). The C-RDR file sizes are not subject to the same constraints. The NPP C-RDR file sizes can be selected to maximize utility and minimize the number of files being archived. The currently planned sizes for the C-RDR data files are shown in Table 2. These sizes are calculated with no compression. Lossless compression within the netCDF-4 software will likely reduce the file sizes by 50% or better, but until files containing realistic test data are produced, the actual range of compressed file sizes are not known.

Sensor	C-RDR File Size (Megabytes)	C-RDR File Period (min)
ATMS	211.98	102
CrIS	1098.09	102
OMPS NP	214.15	102
VIIRS	945.47	6

Table 2: NPP C-RDR File Sizes and Periods

4.2 Production and Dissemination

The NPP C-RDRs will be produced using computing facilities within the NCDC. The production system will acquire science RDR files for the four NPP instruments from the Comprehensive Large-data Array Stewardship System (CLASS) soon after CLASS receives them from the IDPS. The system will manage gaps in the raw data stream, and will calculate quality metrics for the C-RDR files as they are created. These quality metrics will be stored in the C-RDR files as metadata attributes. As the C-RDR files are produced, they will be stored on systems within the NCDC and archived. They will be available internally for use in producing Fundamental CDR (F-CDR) products. The C-RDR files will be available to the larger community through subscription. The CDRP team is also investigating the use of Local Data Manager (LDM) or Integrated Rule-Oriented Data System (iRODS) to provide external users a means to obtain a live feed of C-RDR files as they are produced.

5. SUMMARY

By producing, archiving, and disseminating C-RDRs for the NPP satellite ATMS, CrIS, OMPS NP, and VIIRS instruments, the NCDC will provide the climate research community with efficient and measurement-oriented access to the raw sensor data and processing coefficients. The self-descriptive nature of the C-RDRs will enhance their usability both now and in the future.

6. ACKNOWLEDGEMENTS

The authors would like to thank the NASA Science Data Segment Land Product and Evaluation and Test Element, who shared expertise and a Linux port of the VIIRS operational software, and the Space Dynamics Laboratory at Utah State University who shared expertise and a Linux port of the CrIS/ATMS operational software.

7. REFERENCES

Bates, J. J., 2004: NOAA's Scientific Data Stewardship Program. *Proc. 13th Conference on Satellite Meteorology and Oceanography*, Norfolk, VA, Amer. Meteor. Soc., 5.1. [Available online at http://ams.confex.com/ams/13SATMET/techprogram/paper_79249.htm]

Leibee, J. E., and J. Kronenwetter, 2003: NPP: NASA's continuity mission for Earth observation. *Proc. SPIE Conf. on Sensors, Systems, and Next-Generation Satellites VI*, Vol. 4881, Agia Pelagia, Crete, Greece, The International Society for Optical Engineering, 86-94. [Available online at http://dx.doi.org/10.1117/12.462508]

NRC, 2001: *Climate Change Science: An Analysis of Some Key Questions*. The National Academies Press, 42 pp. [Available online at http://www.nap.edu/catalog.php?record_id=10139]

NRC, 2004: *Climate Data Records from Environmental Satellites: Interim Report*. The National Academies Press, 150 pp. [Available online at http://www.nap.edu/catalog.php?record_id=10944]

Privette, J. L., B. Barkstrom, J. J. Bates, T. Karl, B. Cramer, J. Kaye, D. Cecil, D. Young, C. Koblinsky, M. Tanner, G. Davis, M. F. Bonadonna, K. Boyd, 2008: Restoration of NPOESS climate capabilities - climate data records. *Fourth Symposium on Future National Operational Environmental Satellites*, New Orleans, LA, Amer. Meteor. Soc., P1.3. [Available online at http://ams.confex.com/ams/pdfpapers/131058.pdf]