Accessing and Using NPOESS Data

Suzanna (Barth) Diener, Ph.D., Raytheon Company, 16800 E,. CentreTech Parkway, Aurora CO 80011

I. ABSTRACT

The NPOESS data products are organized using some advanced features in the HDF5 format. Users should be aware of existing documentation, implementation of metadata and data structures, packaging of associated quality flags, and mapping the appropriate geolocation files to the observation data. A survey of current COTS software packages able to open the HDF5 products is presented, and coordination plans with other COTS packages such as IDL and Matlab is discussed.

II. HDF5

A. NPOESS PRODUCT DESCRIPTION

There are approximately 100 deliverable NPOESS data products through the Ground Integrated Data Processing System. These products include Raw Data Records, Sensor Data Records, and Environmental Data Records.

Of the 100 deliverable products, there are approximately 15 types of geolocation products associated with a particular SDR or EDR product. These geolocation products are listed in Table 1, and are delivered as secondary products with their primary SDR or EDR product. Because the geolocation data is stored primarily as IEEE float 64, a single geolocation product may be delivered with EDRs sharing a common geolocation method. This allows the user to avoid delivering multiple copies of duplicate files.

Geolocation Product	Product ID
ATMS-SDR-GEO	GATMO
CrIMSS-AUX-EDR	GCRIO
CrIS-SDR-GEO	GCRSO
VIIRS-MOD-EDR-GEO	GMGTO
VIIRS-MOD-GEO	GMODO
VIIRS-MOD-GEO-TC	GMTCO
VIIRS-MOD-MAP-IP	IVMIM
VIIRS-MOD-UNAGG-GEO	VMUGE
VIIRS-NCC-EDR-GEO	GNCCO
VIIRS-Aeros-EDR-GEO	
VIIRS-CLD-AGG-GEO	
VIIRS-DNB-GEO	GDNBO
VIIRS-IMG-EDR-GEO	GIGTO
VIIRS-IMG-GEO	GIMGO
VIIRS-IMG-GEO-TC	GITCO

Table 1
List of NPOESS Geolocation Products

A VIIRS geolocation file may range in size from 80 MB to 320 MB per granule. A single file for an SDR or EDR product may contain multiple granules of data, for any single data product type.

The NPOESS file naming convention consists of an ~80 character title, comprised of the following information: Data Product ID, Spacecraft ID, Data Start Date, Data Start Time, Data Stop Time, Orbit Number, Creation Date, Origin, Domain Description, and Extension. It is important to note that because this filename is dependent on creation time, each and every filename is unique the moment it is created. Thus, the file management of data must accommodate the continuously changing filenames. Note also, that the filename does not include the Granule ID number. For a description of filename fields, see Figure 1.

Field Position	Description	Applicable Values
1	Data Product ID This field provides a five character identifier for multiple data products contained in the file, each Data Product ID will be given, delineated by a dash, and are in the order that the products were placed in the HDFE file during creation.	Users See Appendix A, NPP/NPOESS Data Product Collection Short Names
1	Field Separator	_
1	Spacecraft ID	Npp NPP J1SS Orbit n1 NPOESS Latch1 n2 NPOESS Latch2
1	Field Separator	_
1	Data Start Date This field is identified with a 'd' followed by the year, month and day of the earliest granule described in the dataset of the file.	dYYYYMMDD YYYY=2000-2100 MM=01-12 DD=01-31
1	Field Separator	_
1	Data Start Time (UTC) This field is identified with a 't' followed by the hour, min, and sec of the earliest granule described in the dataset of the file.	tHHMMSS hh=00-23 MM=00-59 SS=00-60
1	Field Separator	_
1	Data Stop Time (UTC) This field is identified with an 'e' followed by the hour, minutes, and seconds of the latest granule described in the dataset of the file.	eHHMMSS hh=00-23 MM=00-59 SS=00-60
1	Field Separator	_
1	Orbit Number This field is identified with a 'b' followed by the orbit number that the dataset originated from. The orbit begins at the ascending node and the number is indicative of the earliest granule produced in the data product.	bxxxxx
1	Field Separator	_
1	Creation Date (UTC) This field is identified with a 'c' followed by the year, month, day, hour, min, and sec.	cYYYYMMDDHHMMSS YYYY=2000-2100 MM=01-12 DD=01-31 HH=00-23 MM=00-59 SS=00-60
1	Field Separator	_
1	Origin This field is a four character alpha-numeric identifier for the originator of the file.	Refer to list of the possible identifiers for this field, see Table 2-4.1, Origin Identifiers.
1	Field Separator	_
1	Domain Description This field is a three character identifier which indicates the domain from which the data originated. There are five recognized domain values for NPOESS. There is no specific domain for Field terminals.	Refer to list of the possible identifiers for this field, see Table 2-4.2, Domain Identifiers.
1	Field Separator	_
1	Extension The extension indicates the format of the file. All NPOESS/CDFS Data Products are delivered in the HDF5 file format.	h5

Table 2
Naming Convention for NPOESS Data File.

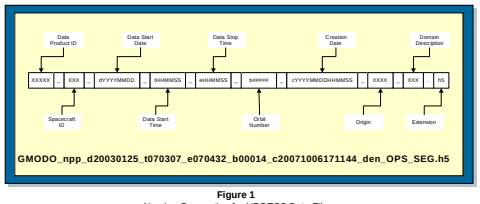

Figure 1
Naming Convention for NPOESS Data File

B. NPOESS HDF5 DATA STRUCTURES

The file's NPOESS HDF 5 data structure includes external metadata, internal metadata, and two parent-level directories. The first parent-level directory contains the actual data fields and associated quality flags, and the second contains information describing navigable points in the data from the first parent-level directory, such as granule start and end points. See Figure 2.

III. TOOLS

The following description of tools describes (a) Benefits of each tool, (b) Current abilities, and (c) Expected capabilities within the next development cycle. This is only a small selection of tools available to the user community, but it is intended to demonstrate accessibility of the NPOESS data.

A. HDFView 2.3, The HDF Group

The purpose of the HDF Group (THG) is to ensure the sustainable development of HDF technologies and the ongoing accessibility of HDF-stored data. It is a non-profit organization. THG advances education and science in public-purpose scientific research communities by developing, maintaining and servicing non-proprietary, freely available tools and technologies.

HDFView is a Java-based tool for browsing and editing the files formatted using the National Center for Supercomputing HDF4 and HDF5 standards. HDFView was implemented using JavaTM 2 Platform and designed to be machine-independent. The interface and features of HDFView are the same for all platforms.

HDFView allows users to browse through files starting with a hierarchical view using a tree diagram of all major objects. HDFView allows a user to descend through the hierarchy and navigate among the file's data objects and select individual file elements or objects. Users can easily expand or collapse folders to navigate the hierarchical structure of the HDF file. See **Figure 2**, for an illustration of the HDFView interface, useful for navigating through products.

The content of a data object is loaded only when the object is selected; this feature conserves RAM usage and allows the user to open only portions of the very large NPOESS files.

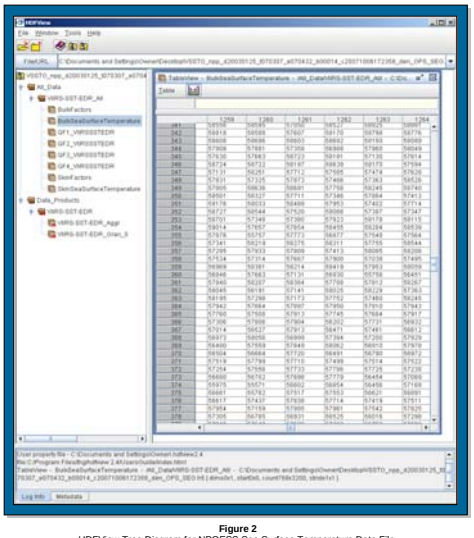

Figure 2
HDFView Tree Diagram for NPOESS Sea Surface Temperature Data File.

B. MATLAB (R2007b), The MathWorks, Inc.

MATLAB allows the user to access a GUI or command line scripts to access HDF5 files. It is supported by various flavors of the following platforms: Linux, Linux x86-64, Mac OS X (Intel), Mac OS X (PPC), Solaris 64, Windows, Windows x64.

The GUI enables the user to visualize an HDF4 structure, while automatically constructing the commands to open and import portions of the HDF4 data. HDF5 commands allow fast access to large files, working easily to perform batch processes for analyzing large amounts of data. Although the HDF4 file structure is viewable, in order to avoid loading the whole file, it is recommended that command-line scripts be used to open HDF5 files, for this version. 4 GB of RAM allows the user to comfortably handle NPOESS data files.

MATLAB excels in speed and ease of use for mathematical analysis and comparison of data. The visualization tool offers many color palettes and formats, capable of being applied to different portions of a data spectrum, for display in a single coherent image.

The MathWorks plans to continue supporting the HDF community as its standards evolve, and it is expected that the GUI will support HDF5 data formats in future versions of MATLAB.

C. IDL 6.4, ITT Visual Information Solutions, Inc.

IDL is a high-level language and tool used for manipulating and visualizing data. Built-in algorithm libraries allow image processing, math, statistics, and analysis routines. The array-oriented interpreted language enable the user to issue individual commands on the fly, with the speed of optimized C-language for array operations. IDL 7.0 runs on Windows XP/Vista, Mac OS X, Linux and SPARC Solaris Operating Systems.

IDL includes an HDF browser with a graphical user interface (GUI) allows users to browse through HDF5 (HDF4, HDF-EOS, and NetCDF) files starting with a hierarchical view using a tree diagram of all major objects; additionally, the GUI includes black and white thumbnail image displays as the user browses through multiple data objects within the NPOESS HDF5 file. The HDF Browser also provides a template for the extraction of HDF data and metadata into IDL, and IDL had no difficulty in accessing both NPOESS HDF5 external and internal metadata items. The content of a data object is loaded only when the object is selected; this feature conserves RAM usage and allows the user to open only portions of the very large NPOESS files.

See **Figure 3**, for an illustration of the IDL GUI interface, useful for navigating through products. IDL also allows command line scripting to enable batch processing of large amounts of data.

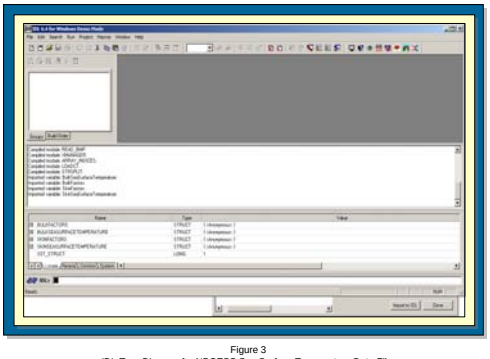

Figure 3
IDL Tree Diagram for NPOESS Sea Surface Temperature Data File.

IDL 7.0 supports a wide variety of file formats including HDF5 5-1.6.3, HDF-EOS 2.8, and HDF 4.1r5, plus other CDF and netCDF scientific data formats. HDF5 and the scientific data formats are core features of IDL. IDL, Research Systems, works closely with The HDF Group to ensure needed functionality and updates for new HDF5 releases as incorporated in IDL in future releases.

IV. SUMMARY

HDFView, MATLAB, and IDL currently support HDF5 data files and structures. All three tools provide GUIs which allow the user to click on single files, and IDL and MATLAB also include command line and scripting languages to open many files at once or in sequence. Due to the size of a single granule of an NPOESS data product, and its associated geolocation file, all tools do well with about 4 GB RAM, in order to comfortably navigate through the NPOESS products. Tools readily access external metadata, and generally allow different methods of access to internal metadata. Bit Level Quality Flags are packed into bytes and are available for interpretation through the NPOESS External Data Format Control Book. HDFView, MATLAB, and IDL all are plan to closely continue their support of HDF5 data.

Bibliography

NPOESS Common Data Format Control Book – External Volume I – Overview. NO. D34862-01, Rev. B, 8/27/07.

HDFView 2.3, User's Guide. http://hdf.ncsa.uiuc.edu/hdf-java , HDF Group, National Center for SuperComputing Applications, 2007. html/hdfview/UsersGuide/index.html

IDL 6.4, http://www.ittvis.com/idl/ , ITT Visual Information Solutions, 2007.

MATLAB R2007b, http://www.mathworks.com/products/matlab/, The MathWorks, 2007.