VIIRS Cloud Mask Science Processing Algorithm (CLOUDMASK_SPA) User's Guide

Version 1.5.08.04

December 2018

GODDARD SPACE FLIGHT CENTER
GREENBELT, MARYLAND

VIIRS Cloud Mask Science Processing Algorithm

CLOUDMASK_SPA

General

The NASA Goddard Space Flight Center's (GSFC) Direct Readout Laboratory (DRL), Code 619.1, developed this software for the International Planetary Observation Processing Package (IPOPP). IPOPP maximizes the utility of Earth science data for making real-time decisions by giving fast access to instrument data and derivative products from the NOAA-20 [Joint Polar Satellite System (JPSS)], Suomi National Polar-orbiting Partnership (SNPP), Aqua, and Terra missions.

Users must agree to all terms and conditions in the Software Usage Agreement on the DRL Web Portal before downloading this software.

Software and documentation published on the DRL Web Portal may occasionally be updated or modified. The most current versions of DRL software are available at the DRL Web Portal:

https://directreadout.sci.gsfc.nasa.gov/?id=software

Questions relating to the contents or status of this software and its documentation should be addressed to the DRL via the Contact DRL mechanism at the DRL Web Portal:

https://directreadout.sci.gsfc.nasa.gov/?id=dspContent&cid=66

Algorithm Wrapper Concept

The DRL has developed an algorithm wrapper to provide a common command and execution interface to encapsulate multi-discipline, multi-mission science processing algorithms. The wrapper also provides a structured, standardized technique for packaging new or updated algorithms with minimal effort.

A Science Processing Algorithm (SPA) is defined as a wrapper and its contained algorithm. SPAs will function in a standalone, cross-platform environment to serve the needs of the broad Direct Readout community. Detailed information about SPAs and other DRL technologies is available at the DRL Web Portal.

Software Description

This software package contains the Visible Infrared Imaging Radiometer Suite (VIIRS) Cloud Mask Science Processing Algorithm (CLOUDMASK_SPA). The SPA supports both the SNPP and NOAA-20 (also called JPSS-1) missions; it can process VIIRS inputs from either spacecraft. The VIIRS Cloud Mask algorithm takes as input VIIRS I1, I2, I4, I5, M1, M4, M5, M7, M8, M9, M10, M11, M12, M13, M14, M15, and M16 band Sensor Data Record (SDR) products; the VIIRS M-Band Terrain-corrected Geolocation product; the VIIRS Active Fires Application Related Product (ARP); and meteorological ancillaries. The SPA produces the mission-compliant VIIRS Cloud Mask IP HDF5 product. The SPA functions in two modes: Standalone, or as an IPOPP plug-in.

Software Version

Version 1.6 of the DRL algorithm wrapper was used to package the SPA described in this document. The VIIRS Cloud Mask algorithm has been ported from the IDPS OPS (Interface Data Processing Segment Operational Code) Version 1.5.08.04.

Enhancements to this SPA include support for NOAA-20 (JPSS-1) VIIRS data processing. The NOAA-20 algorithm is an Engineering Version.

This software will execute on a 64-bit computer. This software has been tested on a computer with 32GB of RAM and a CentOS Linux 7 X86_64 operating system.

Credits

The VIIRS Cloud Mask algorithm was provided to the DRL by the JPSS Mission. This algorithm was ported to run outside of the IDPS by the DRL in collaboration with the Land Product Evaluation and Algorithm Test Element (LPEATE).

Prerequisites

To run this package, you must have the Java Development Kit (JDK) or Java Runtime Engine (JRE) (Java 1.6.0_25 or higher) installed on your computer, and have the Java installation bin/ subdirectory in your PATH environment variable. This package contains 64-bit binaries statically pre-compiled on an x86-compatible 64-bit computer running under Fedora 14, using gcc 4.5.1.

Program Inputs and Outputs

The SPA uses the following inputs:

a) VIIRS I1, I2, I4, I5, M1, M4, M5, M7, M8, M9, M10, M11, M12, M13, M14, M15, and M16 band Sensor Data Record (SDR) products;
b) VIIRS M-Band Terrain-corrected Geolocation product;
c) VIIRS Active Fires ARP Product; and
d) meteorological ancillaries.

The SPA produces the mission-compliant VIIRS Cloud Mask Intermediate Product (IP) HDF5 product as output.

Installation and Configuration

NOTE: Due to limited resources, as well as the many variables that impact scientific integrity and algorithm stability, the DRL will soon no longer support the Standalone Mode for SPA processing. We strongly encourage you now to run SPAs in IPOPP Mode exclusively, that is, from within the IPOPP processing framework. IPOPP will autonomously:

- discover and register raw sensor data;

- retrieve ancillaries from the DRL's real-time and archived ancillary repositories;
- register ancillaries in its Ancillary File Cache;
- schedule SPA executions;
- fulfill science data/ancillary requests from SPAs;
- generate science data products; and
- manage the IPOPP file system.

Installing into an IPOPP Framework: This SPA can also be installed dynamically into an IPOPP framework to automate production of VIIRS Cloud Mask IP data products. The SPA installation process will install SPA service(s) into IPOPP. An SPA service is an IPOPP agent that provides the mechanism necessary for running an SPA automatically within the IPOPP framework. Once this SPA is installed, users must enable the service(s) corresponding to this SPA along with any other prerequisite service(s). Instructions for installing an SPA and enabling its services are contained in the IPOPP User's Guide (available on the DRL Web Portal). The SPA services associated with this SPA are listed in Appendix A.

Installing as a Standalone Application:

NOTE: If you have a previous version of this SPA installed, delete the SPA/CloudMask directory before decompressing and un-archiving the new SPA tar file.

Download the CLOUDMASK_1.5.08.04_SPA_1.6.tar.gz and CLOUDMASK_1.5.08.04_SPA_1.6_testdata.tar.gz (optional) files into the same directory.

Decompress and un-archive the CLOUDMASK_1.5.08.04_SPA_1.6.tar.gz and CLOUDMASK_1.5.08.04_SPA_1.6_testdata.tar.gz (optional) files:

```
$ tar -xzf CLOUDMASK_1.5.08.04_SPA_1.6.tar.gz
$ tar -xzf CLOUDMASK_1.5.08.04_SPA_1.6_testdata.tar.gz
```

This will create the following subdirectories:

```
SPA
    CloudMask
        algorithm
        ancillary
        mode
        station
        testdata
        testscripts
        wrapper
```

Software Package Testing and Validation

The testscripts subdirectory contains test scripts that can be used to verify that your current

installation of the SPA is working properly, as described below. Note that the optional CLOUDMASK_1.5.08.04_SPA_1.6_testdata.tar.gz file is required to execute these testing procedures.

Step 1: cd into the testscripts directory.
Step 2: There is a script named run-vcm inside the testscripts directory.
To run the VIIRS Cloud Mask algorithm, use

```
$ ./run-vcm
```

A successful execution usually requires three minutes or more, depending on the speed of your computer. If everything is working properly, the scripts will terminate with a message such as:

```
Output viirs.cmip is /home/ipopp/drl/SPA/CloudMask/testdata/output/IICMO.h5
```

You can cd to the output directory to verify that the science products exist. Test output product(s) are available for comparison in the testdata/output directory. These test output product(s) were generated on a 64-bit PC architecture computer running CentOS 7. The output products serve as an indicator of expected program output. Use a comparison utility (such as diff, h5diff, etc.) to compare your output product(s) to those provided in the testdata/output directory. Locally generated files may differ slightly from the provided output files because of differences in machine architecture or operating systems.

If there is a problem and the code terminates abnormally, the problem can be identified using the log files. Log files are automatically generated within the directory used for execution. They start with stdfile* and errfile*. Other log and intermediate files may be generated automatically within the directory used for execution. They are useful for traceability and debugging purposes. However it is strongly recommended that users clean up log files and intermediate files left behind in the run directory before initiating a fresh execution of the SPA. Intermediate files from a previous run may affect a successive run and produce ambiguous results. Please report any errors that cannot be fixed to the DRL.

Program Operation

In order to run the package using your own input data, you can either use the run scripts within the wrapper subdirectories, or modify the test scripts within the testscripts subdirectory.

To Use the Run Scripts

Identify the 'run' scripts: The wrapper directory within this package contains one subdirectory named CloudMask. The subdirectory contains an executable called 'run'. Execute 'run' within the correct wrapper subdirectory to generate the corresponding product. For instance, the 'run' within wrapper/CloudMask is used for creating VIIRS Cloud Mask outputs. Note that to execute 'run', you need to have java on your path.

Specify input parameters using <label value> pairs: To execute the 'run' scripts, you

must supply the required input and output parameters. Input and output parameters are usually file paths or other values (e.g., an automatic search flag). Each parameter is specified on the command line by a <label value> pair. Labels are simply predefined names for parameters. Each label must be followed by its actual value. Each process has its own set of <label value> pairs that must be specified in order for it to execute. Some of these pairs are optional, meaning the process would still be able to execute even if that parameter is not supplied. The three types of <label value> pairs that the CLOUDMASK_SPA uses are:

a) Input file label/values. These are input file paths. Values are absolute or relative paths to the corresponding input file.

b) Output file label/values. These are output files that are produced by the SPA. Values are absolute or relative paths of the files you want to generate.

c) Parameter label/values. These are parameters that need to be passed into the SPA (e.g., scan time).

The following tables contain labels, and their descriptions, required by the CLOUDMASK_SPA.

Input File Labels	Description	Source
viirs.gmtco	VIIRS Moderate Resolution Band Terrain-Corrected Geolocation input HDF5 file path	1. The C-SDR_SPA can be used to create these products. 2. Real time VIIRS SDR and Geolocation products over the eastern US region are available from the DRL ftp site at: SNPP: ftp://is.sci.gsfc.nasa.gov/gsfcdata/npp/viirs/level1/<GMTCO|SVIxx|SVMxx>_npp_dyyyyMMdd_thhmmssS_ehhmmssS*.h5 NOAA-20: ftp://is.sci.gsfc.nasa.gov/gsfcdata/jpss1/viirs/level1/<GMTCO|SVIxx|SVMxx>_j01_dyyyyMMdd_thhmmssS_ehhmmssS*.h5 Where yyyy, MM, dd represents the year, month, and day of month for the start of the swath; the first hh, mm, ss, S represents the hour, minutes, seconds, and 10th of a second for the start of the swath and the second hh, mm, ss, S represents the end time of the swath. 3. VIIRS SDR and Geolocation products for other locations and times are available for download at www.class.noaa.gov
viirs.svixx {xx = 01, 02, 04, 05}	VIIRS Imagery Resolution Band Ix {x = 1, 2, 4, 5} SDR input HDF5 file path	
viirs.svmxx {xx = 01, 04, 05, 07 to 16}	VIIRS Moderate Resolution Band Mx {x = 1, 4, 5, 7 to 16} SDR input HDF5 file path	
ncep_met	NCEP Numerical Weather Prediction Gridded Binary (GRIB)	For recent files go to: ftp://is.sci.gsfc.nasa.gov/ancillary/temporal/global/gd

Input File Labels	Description	Source
	File. This can be either a Global Data Assimilation System (GDAS1, 6 hourly, 1 degree global) analysis field file or a Global Model Forecast Fields (GFS) file (The GDAS and GFS files must be in grib1 format). Refer to the "NOTES" section below for more details regarding these meteorological ancillary file inputs.	as/(for GDAS) ftp://is.sci.gsfc.nasa.gov/ancillary/temporal/global/gfs (for GFS) For archived files go to: ftp://is.sci.gsfc.nasa.gov/ArchivedAncillary/temporal/global/gdas/
ssmi_nise	National Snow and Ice Data Center (NSIDC) Near-real time Ice and Snow Extent (NISE) (1 degree, global, daily)	For recent files go to: ftp://is.sci.gsfc.nasa.gov/ancillary/temporal/global/nise/NISE_SSMIF13_yyyyMMdd.HDFEOS For archived files go to: ftp://is.sci.gsfc.nasa.gov/ArchivedAncillary/temporal/global/nise/NISE_SSMIF13_yyyyMMdd.HDFEOS Where yyyy, MM, dd represents the year, month, and day for the NSIDC NISE ancillary file
Parameters	**Description**	
scantime	Parameter representing the inputs' scan start time in yyyyMMdd format where yyyy, MM, dd represents the year, month, and day of month for the start of the swath (e.g. 20140901)	

Output File Labels	Description	Output Format Description
viirs.cmip	VIIRS Cloud Mask IP output HDF5 file path	Refer to JPSS Common Data Format Control Book (CDFCB) Vol IV-Part 1 – IPs, ARPs, and Geolocation Data https://jointmission.gsfc.nasa.gov/documents.html

Execute the 'run': The following is an example of the command line to run the VIIRS Cloud Mask algorithm from the testscripts directory:

```
$  ../wrapper/CloudMask/run \
ncep_met ../testdata/input/gdas1.PGrbF00.140901.18z \
ssmi_nise ../testdata/input/NISE_SSMIF13_20140901.HDFEOS \
scantime 20140901 \
viirs.gmtco ../testdata/input/GMTCO_npp_d20140901_t1738560_e1740201_b14746_c20140908200702642321_noaa_ops.h5 \
   viirs.svm01 ../testdata/input/SVM01_npp_d20140901_t1738560_e1740201_b14746_c20140908200638045758_noaa_ops.h5 \
   viirs.svm04 ../testdata/input/SVM04_npp_d20140901_t1738560_e1740201_b14746_c20140908200714715029_noaa_ops.h5 \
   viirs.svm05 ../testdata/input/SVM05_npp_d20140901_t1738560_e1740201_b14746_c20140908200735223402_noaa_ops.h5 \
   viirs.svm07 ../testdata/input/SVM07_npp_d20140901_t1738560_e1740201_b14746_c20140908200729256422_noaa_ops.h5 \
   viirs.svm08 ../testdata/input/SVM08_npp_d20140901_t1738560_e1740201_b14746_c20140908200803979409_noaa_ops.h5 \
```

```
viirs.svm09 ../testdata/input/SVM09_npp_d20140901_t1738560_e1740201_b14746_c20140908200743076006_noaa_ops.h5 \
viirs.svm10 ../testdata/input/SVM10_npp_d20140901_t1738560_e1740201_b14746_c20140908200925075349_noaa_ops.h5 \
viirs.svm11 ../testdata/input/SVM11_npp_d20140901_t1738560_e1740201_b14746_c20140908200858348915_noaa_ops.h5 \
viirs.svm12 ../testdata/input/SVM12_npp_d20140901_t1738560_e1740201_b14746_c20140908200844775167_noaa_ops.h5 \
viirs.svm13 ../testdata/input/SVM13_npp_d20140901_t1738560_e1740201_b14746_c20140908200931364181_noaa_ops.h5 \
viirs.svm14 ../testdata/input/SVM14_npp_d20140901_t1738560_e1740201_b14746_c20140908200652259426_noaa_ops.h5 \
viirs.svm15 ../testdata/input/SVM15_npp_d20140901_t1738560_e1740201_b14746_c20140908200911898965_noaa_ops.h5 \
viirs.svm16 ../testdata/input/SVM16_npp_d20140901_t1738560_e1740201_b14746_c20140908200729274370_noaa_ops.h5 \
viirs.svi01 ../testdata/input/SVI01_npp_d20140901_t1738560_e1740201_b14746_c20140908200814185302_noaa_ops.h5 \
viirs.svi02 ../testdata/input/SVI02_npp_d20140901_t1738560_e1740201_b14746_c20140908200817203819_noaa_ops.h5 \
viirs.svi04 ../testdata/input/SVI04_npp_d20140901_t1738560_e1740201_b14746_c20140908200756074531_noaa_ops.h5 \
viirs.svi05 ../testdata/input/SVI05_npp_d20140901_t1738560_e1740201_b14746_c20140908200946032012_noaa_ops.h5 \
viirs.cmip ../testdata/output/IICMO.h5
```

A successful execution usually requires three minutes or more, depending on the speed of your computer and the size of the input. If execution fails, you will see an error message indicating the cause of failure (e.g., a file cannot be found, or a label cannot be recognized). Correct it and run again. If the problem has some other cause, it can be identified using the log files. Log files are automatically generated within the directory used for execution. They start with stdfile* and errfile* and can be deleted after execution. Other log and intermediate files may be generated automatically within the directory used for execution. They are useful for traceability and debugging purposes. However it is strongly recommended that users clean up log files and intermediate files left behind in the run directory before initiating a fresh execution of the SPA. Intermediate files from a previous run may affect a successive run and produce ambiguous results. The 'run' can be executed from any directory the user chooses. This can be done by prefixing it with the file path for the 'run' script.

NOTES:

1. Either GDAS or Global Model Forecast Fields (GFS) files may be used for the "ncep_met" label. Try to use a GDAS file that is within ±3 hours of the SDR observation time. If that file is not available (as is often the case for real-time processing), use a GFS file instead. The naming convention for grib1 gfs files is gfs.thh.yymmdd.pgrbfxx (here yymmddd and hh represent analysis time, and xx represents forecast time step). Thus a file named gfs.t12.100201.pgrbf03 corresponds to 1500 hours (12+3) UTC on February 1, 2010. If you have to choose GFS data as input, you should attempt to use a file that is within ±1.5 hours of the SDR file. If there is more than one such GFS file, use the one with the smaller forecast time step. For example, if your data time is 15 UTC, you should try to use the 3 hour forecast field from the 1200 UTC model run, instead of the 9 hour forecast field from the 0600 UTC run. If no GDAS or GFS file is available using the above logic, use a GDAS file that is closest in time but within ±7 days of the granule time.
2. The dates for the NSIDC Near-real time Ice and Snow Extent (NISE) datasets (specified by the "ssmi_nise" label) should be as close as possible to the dates of the L1B granules. It is recommended to use an ancillary file that is within ±14 days of the granule time. The dates for the NISE ancillary files are encoded in the filenames as NISE_SSMIF13_yyyymmdd.HDFEOS. The ssmi_nise files are required for all products.
3. The data products generated by this SPA may be visualized with the DRL's H2G_SPA (Hierarchical Data Format [HDF] to Georeferenced Tagged Image File

Format [GeoTIFF] Converter Science Processing Algorithm). H2G is designed specifically for Direct Readout applications to create geolocated GeoTIFF images, jpeg browse images, and png browse images for parameter datasets in SNPP, NOAA-20 and EOS products. H2G_SPA and its User Guide are available for download from the DRL Web Portal. Please refer to Appendix A for information on enabling image production for this SPA in IPOPP.

To Use the Scripts in the testscripts Directory

One simple way to run the algorithms from the directory of your choice using your own data is to copy the run-vcm script from the testscripts directory to the selected directory. Change the values of the variables like WRAPPERCREFL, INPUTHOME and OUTPUTHOME to reflect the file paths of the wrapper directories and the input/output file paths. Then modify the input/output file name variables. Run the script to process your data.

Appendix A
SPA Services

Installation of this SPA in IPOPP mode will make the SPA services listed in Table A-1 available to IPOPP. These services along with any other prerequisite services (listed in Table A-2) will need to be enabled to allow IPOPP to automate production of the CLOUDMASK_SPA data products. Furthermore, users who wish to generate image products from the data products generated by this SPA will need to enable the image-generating services listed in Table A-3. The SPAs containing the Prerequisite and the image-generating services listed in Tables A-2 and A-3 can be downloaded from the DRL Web Portal, in case they are not already available in your IPOPP installation. Details about these other SPAs are available in the respective SPA User's Guide. Please refer to the IPOPP User's Guide for instructions on how to install an SPA in IPOPP and enable the corresponding services.

Table A-1. SPA Services

Services for this SPA	**Data Products produced**	
CloudMask[1]	**Product Name**	**Destination (when installed in IPOPP)**
	VIIRS Cloud Mask IP	$HOME/drl/data/pub/gsfcdata/<missiondir>/viirs/level2/IICMO_ppp_d*yyyyMMdd*_t*hhmmssS*_e*hhmmssS**.h5[2]

[1]The SPA service(s) for SNPP processing is available on the "SNPP-VIIRS" tab of the IPOPP dashboard, while the corresponding SPA service for NOAA-20 processing is available on the "JPSS-1-VIIRS" tab.

[2]Where *yyyy*, *MM*, *dd* represents the year, month and day of month for start of the swath; the first *hh*, *mm*, *ss*, *S* represents the hour, minutes, seconds and 10th of a second for the start of the swath and the second *hh*, *mm*, *ss*, *S* represents the end time of the swath. 'ppp' is the spacecraft field and is either 'npp' for SNPP or 'j01' for NOAA-20. <missiondir> is the mission specific directory. It is 'npp' for SNPP and 'jpss1' for NOAA-20.

Table A-2. Prerequisite SPA Services

Prerequisite SPA services	SPA in which they are available
VIIRS_C-SDR	C-SDR_SPA

Table A-3. Image-generating SPA Services

Image-generating SPA services	SPA in which they are available
vcmmaskh5-geotiff	H2G_SPA
vcmphaseh5-geotiff	

NOTE: Please refer to the H2G_SPA User's Guide for more details about the image products, including their locations and filename patterns when they are generated in IPOPP. The SPA service(s) for SNPP processing is available on the "SNPP-VIIRS" tab of the IPOPP dashboard, while the corresponding SPA service for NOAA-20 processing is available on the "JPSS-1-VIIRS" tab.