

Installation Instructions for the Community Satellite Processing Package (CSPP) Hyper-Spectral Enterprise Algorithm Package (HEAP) Software Version 1.0.2

University of Wisconsin-Madison, Space Science and Engineering Center (SSEC)
Supported by the NASA/NOAA Joint Polar Satellite System (JPSS) Project

12 November 2020

Section 1: Introduction

1.1 Overview

This document contains instructions for installation and operation of the Community Satellite Processing Package (CSPP) release of the NOAA Hyper-Spectral Enterprise Algorithm Package (HEAP) software for retrieving atmospheric profiles of temperature, moisture, trace gases and cloud-cleared radiances from direct broadcast (DB) CrIS+ATMS Sensor Data Records (SDRs) for S-NPP and NOAA-20 and IASI+AMSUA/MHS Level1-b files for Metop-A, B, and C.

HEAP runs the NUCAPS (NOAA Unique Combined Atmospheric Processing System) algorithm. Its development is led by Walter Wolf and Lihang Zhou at the NOAA NESDIS Center for Satellite Applications and Research (STAR). This CSPP release was created in close collaboration with Chris Baret (the NOAA government science lead for NUCAPS from 2003 to 2013) and his staff of developers of the NUCAPS science code, Murty Divakarla and Nadia Smith, who are the scientists most familiar with the algorithmic and scientific underpinnings. For any inquiries or status updates, please contact letitia.soulliard@noaa.gov, Daniel.Lefevre@noaa.gov, and Murty.Divakarla@noaa.gov.

This CSPP release provides HEAP version 2.1, adapted and tested for operation in a real-time direct broadcast environment. The software contains binary executable files and supporting static data files, as well as input and output files for verification of a successful installation. The CSPP_HEAP software is available from the CSPP website:

<https://cimss.ssec.wisc.edu/cspp/>

Software, test data, and documentation may be downloaded from this web site. Please use the 'Contact Us' form on the website to submit any questions or comments about CSPP. Source code for HEAP is not included in this release and is presently not available for distribution.

CSPP_HEAP requires access to dynamic ancillary data, specifically Global Forecast System (GFS) forecast files in Gridded Information in Binary format (GRIB). These data can be fetched at runtime from our server, currently at https://jpssdb.ssec.wisc.edu/cspp_v_2_0/ancillary/, but this may be modified by changing environment variable JPSS_REMOTE Ancillary DIR. This automatic fetching is the default behavior, but it can be turned off for sites not connected to the internet (--quarantine mode); the user is then responsible for fetching and staging these data (**Section 2.3**).

1.2 System requirements

System requirements for the CSPP_HEAP v1.0.2 software are as follows:

- Intel or AMD CPU with 64-bit instruction support, with 2GB RAM,
- CentOS-6, -7 or -8 64-bit Linux (or other compatible 64-bit Linux distribution),
- Perl v5.010 ; for CentOS-7 make sure perl-core is also installed,
- Command-line utilities timeout, which, and uidgen,
- lftp utility for fetching ancillary data (unless running in quarantine mode),
- 6 GB of disk space (plus space for your own DB data and CSPP_HEAP products).

Linux terminal commands included in these instructions are for the bash shell.

1.3 Implementation Notes

The directory/file structure of CSPP_HEAP_1_0 (pruned for the purposes of this document) is shown below:

```
CSPP_HEAP_1_0/
|-- bin
|   |-- h5diff
|   |-- h5dump
|   |-- h5repack
|   |-- kai
|   |-- nagg
|   |-- ncdump
|   |-- ncgen
|   |-- wgrib
|   `-- wgrib2
|-- DAP
|   |-- HEAP_v2-1
|   |   |-- DOCS
|   |   |   |-- HEAP_Delivery_Memo.docx
|   |   |   |-- HEAP_EUM_v1.0.docx
|   |   |   |-- HEAP_PCF_PSF.docx
|   |   |   |-- HEAP_Production_Rules.docx
|   |   |   |-- HEAP_README.docx
|   |   |   |-- HEAP_SMM_v1.0.docx
|   |   |   `-- NUCAPS_ATBD_20130821.pdf
|   |   `-- OPS
|   |       |-- ancillary
|   |       |-- CDLFILES
|   |       |-- Common_Bin
|   |       `-- scripts
|   |-- lib
|   |-- scripts
|   |   |-- aggregate_swath.pl
|   |   |-- cspp_heap_env.sh
|   |   |-- heap_granule.pl
|   |   |-- run_heap.bash
|   |   `-- run_heap.pl
|   `-- version.txt
```

Everything beneath DAP/ comprises the delivery from NOAA/NESDIS/STAR; the remainder is the scripting and organization applied by the CSPP team to implement HEAP in the direct broadcast environment. Detailed documentation files from NOAA are provided in directory CSPP_HEAP_1_0/DAP/HEAP_v2-1/DOCS.

1.4 Input Data Requirements

CSPP_HEAP produces single Field-of-Regard (FOR) retrievals from CrIS + ATMS SDR calibrated and geo-located data and from IASI + AMSUA/MHS calibrated and geo-located Level 1-b data. Table 1 provides information on the satellites and instruments presently supported by the CSPP_HEAP software.

Table 1: CSPP_HEAP input satellite data requirements.

Satellite	Instrument	Direct Broadcast Processing Software	Required Satellite Data Input Filemasks
Suomi-NPP & NOAA-20 (JPSS-1)	ATMS	CSPP SDR v2+ (or IDPS SDR)	Antenna temperature, geolocation: TATMS*.h5, GATMO*.h5
	CrIS		Radiance, geolocation: SCRIS*.h5, SCRIF*.h5, GCRSO*.h5
Metop-A, Metop-B & Metop-C	IASI	AAPP v7.15 or higher	IASI_XXX_1C_{M01,M02}_*
	AMSUA		amsual1b_M0{1,2,3}_*.11b or ????.AMAX.M{1,2,3}.*
	MHS		mhsl1b_M0{1,2,3}_*.11b or ????.MHSX.M{1,2,3}.*

1.4.1 JPSS Data Inputs

CSPP is the Community Satellite Processing Package and, in this table, refers to the SDR generating components of CSPP available from:

<https://cimss.ssec.wisc.edu/cspp/>

IDPS is the Interface Data Processing Segment developed by Raytheon Intelligence and Information Systems. S-NPP and NOAA-20 products created by IDPS are available from the CLASS archive:

<https://www.class.noaa.gov/>

1.4.2 Metop Data Inputs

The required Metop satellite data inputs are generated by the AAPP software available from:

<https://www.nwpsaf.eu/site/software/aapp/>

The underlying HEAP algorithm requires that IASI L1C files be of size 22 or 23 Measurement Data Records (MDRs). Direct Broadcast IASI L1C files will often be longer, comprising the measurement data downlinked during an entire satellite overpass. Thus CSPP_HEAP needs to break an overpass-length IASI L1C file into files (granules) of alternating 22 and 23 MDRs. This may mean that some MDRs are not processed because they do not fit into an integral number of granules. Control over this is provided through command line --mode. The default is to prefer the middle of the overpass.

1.5 What's New?

This is CSPP_HEAP v1.0.2, a minor update to CSPP_HEAP v1.0. It adds support for SO2 retrievals for Suomi-NPP and NOAA-20, and it speeds up execution through a recompilation with -O2 optimizations, halving the execution time of the retrieval step. There is a new test data package, CSPP_HEAP_TESTDATA_V1.0.2.tar.gz. (**NOTE:** Metop-A, B, and C currently return climatological values for SO2).

1.6 Output Products

The NUCAPS retrieval algorithm creates files as described in the '*Hyperspectral Enterprise Algorithm Package (HEAP) System Maintenance Manual (SMM)*' (CSPP_HEAP_1_0/DAP/HEAP_v2-1/DOCS/HEAP_SMM_v1.0.docx); the most used files are:

```
NUCAPS-EDR_v2r0_<sat>_s????????????????_e????????????????_c?????????????????.nc
NUCAPS-CCR-AR_v2r0_<sat>_s????????????????_e????????????????_c?????????????????.nc
NUCAPS-OLR_v2r0_<sat>_s????????????????_e????????????????_c?????????????????.nc
```

where <sat> is either npp, j01, m01, m02, or m03.

Additionally, and by default, aggregated overpass-length output files with reshaped data dimensions are generated by CSPP_HEAP. These are named like NUCAPS-AGG-EDR*.nc and NUCAPS-AGG-CCR*.nc and are non-standard in both name and format. However we have found these aggregated overpass length files to be helpful to users, especially to those who wish to visualize retrievals with imaging software. The standard format files will always be delivered alongside them

The following tables show the primary data products in the NUCAPS-EDR file, a description of stability parameters, and values assigned to the retrieval quality flag.

Table 2: CSPP_HEAP primary data products in NUCAPS-EDR files

Description
Atmospheric temperature [K] at 100 pressure levels
Atmospheric moisture [g/g] at 100 pressure levels
Atmospheric ozone [ppb] at 100 pressure levels
Atmospheric liquid water [g/g] at 100 pressure levels
Carbon dioxide dry mixing ratio [ppm] at 100 pressure levels
Trace gas mixing ratios: CO, CH4, HNO3, N2O, SO2* [ppb] at 100 pressure levels. [*Metop-A, B, and C return climatological values for SO2]
Surface skin temperature [K]
Microwave surface emissivity
Column averaged CO2 per ATMS or AMSU FOV [ppm]
Cloud top pressure for up to two cloud layers [hPa]
Cloud top fractional coverage for up to two cloud layers
Stability parameters

Table 3: CSPP_HEAP Stability Parameter descriptions

Index	Description
1	Convective Available Potential Energy (CAPE) [J/kg]
2	Convective Inhibition (CIN) [J/kg]
3	Pressure of Lifting Condensation Level (LCL) in [mb]
4	Pressure of Equilibrium Level (EL) in [mb]
5	Pressure of Level of Free Convection (LFC) in [mb]
6	Temperature of Lifting Condensation Level (LCL) [K]
7	Temperature of Level of Free Convection (LFC) [K]
8	Potential Temperature of Lifted Parcel in [K]
9	Equivalent Potential Temperature of Lifted Parcel [K]
10	Lifted Index (LI)
11-16	Blank placeholders to allow more to be added at a later date.
Notes	
If there is no LFC, then 1, 2, 5, and 7 are not calculated. If there is no LCL, then 3 and 6 are not calculated. If there is no EL, then 4 is not calculated.	

Table 4: CSPP_HEAP Quality_Flag values for NUCAPS-EDR and NUCAPS-CCR files

Value	Meaning
0	Good
1	rejected by physical
2	rejected by MIT file
4	rejected by NOAA (regression) file
8	rejected by internal MIT
9	rejected by physical and internal MIT
16	rejected by internal NOAA
17	rejected by physical and internal NOAA
24	rejected by internal MIT and internal NOAA
25	rejected by physical, internal MIT, and internal NOAA
-9999	Missing

The Algorithm Theoretical Basis Document (ATBD) for NUCAPS can be obtained from:
https://www.star.nesdis.noaa.gov/jpss/documents/ATBD/ATBD_NUCAPS_v2.0.pdf

1.7 S-NPP Direct Broadcast Products Versus NOAA Archive Products

To accommodate the inclusion of CrIS Full Spectral Resolution (FSR) in the direct broadcast transmission of S-NPP data, it does not include CrIS Field-of-View (FOV) number 4 (beginning 1 December 2014) and number 6 (beginning 18 August 2015) in the High Rate Data (HRD). The un-truncated data sets are available from the NOAA CLASS archive. Therefore, the NUCAPS products created from direct broadcast inputs will differ from the products created using NOAA archive S-NPP inputs. The amount of differences in the retrievals will vary depending on the atmospheric environment of the observations.

1.8 Disclaimer

Original scripts and automation included as part of this package are distributed under the GNU GENERAL PUBLIC LICENSE agreement version 3. Binary executable files included as part of this software package are copyrighted and licensed by their respective organizations, and distributed consistent with their licensing terms.

The University of Wisconsin-Madison Space Science and Engineering Center (SSEC) makes no warranty of any kind with regard to the CSPP software or any accompanying documentation, including but not limited to the implied warranties of merchantability and fitness for a particular purpose. SSEC does not indemnify any infringement of copyright, patent, or trademark through the use or modification of this software.

There is no expressed or implied warranty made to anyone as to the suitability of this software for any purpose. All risk of use is assumed by the user. Users agree not to hold SSEC, the University of Wisconsin-Madison, or any of its employees or assigns liable for any consequences resulting from the use of the CSPP software.

Section 2: Installation and Configuration

2.1 Overview

This software package contains the CSPP_HEAP retrieval system based on the NOAA HEAP Delivered Algorithm Package (DAP) v2.1 which has been packaged by SSEC for direct broadcast users. In CSPP_HEAP_1_0/DAP/HEAP_v2-1/DOCS are NOAA documents that contain detailed information about the processing system, such as input satellite and ancillary data files and output data filename conventions and content specifications.

2.2 Installation of CSPP_HEAP Software

Download the following file from the CSPP website (<https://cimss.ssec.wisc.edu/cspp/>):

```
CSPP_HEAP_V1.0.2.tar.gz
```

Optionally, you can also fetch the test data package:

```
CSPP_HEAP_TESTDATA_V1.0.2.tar.gz
```

Install the software as shown below (a new directory named CSPP_HEAP_1_0 will be created):

```
tar -xzf CSPP_HEAP_V1.0.2.tar.gz
```

Set the CSPP_HEAP_HOME environment variable to the name of the directory where CSPP_HEAP is installed (the current working directory in this example):

```
export CSPP_HEAP_HOME=$PWD/CSPP_HEAP_1_0
```

Alternatively, edit file CSPP_HEAP_1_0/scripts/cspp_heap_env.sh to appropriately set the environment variable CSPP_HEAP_HOME for your system.

Source the environment setting script:

```
source $CSPP_HEAP_HOME/scripts/cspp_heap_env.sh
```

CSPP_HEAP is now ready to use. Test data files are included in the distribution tar file; no additional files are needed. Execute the main driver script now with no arguments:

```
run_heap.bash (or -h or -help)
```

If the installation has been successful to this point, you will be presented with CSPP_HEAP command line switches and options. See next page.

This does not mean that your installation is *processing* data correctly (see **Section 3.2**), but it does mean that the *environment* has been successfully set to start using CSPP_HEAP.

Options:

--version Show program's version number and exit
 -h, --help Show this help message and exit
 -n, --nothing Make the PCF files, but do not execute.

Mandatory Arguments:

At a minimum these arguments must be specified [i.e. no default].

-s MISSION, --sat=MISSION

The mission satellite data for HEAP algorithm.
 Possible values are: npp, j01, m01, m02, m03.

-i INPUT_FILES, --input_files=INPUT_FILES

Path to input files. May be directory name, or directory/filemask (use * and ?, not regexp). Wildcarded filemasks MUST be "quoted" (or OS expands). e.g. -i "/dataDir/{GATM0,GCRS0,SCRIS,TATMS}*t18*.h5", or -i "/dataDir/{IASI_xxx,amsua1b,mhsl1b}*20200504*" (BUT -i /dataDir really is the simplest way).

Extra Options:

These options may be used to customize behavior of this program.

-m MODE, --mode=MODE

Mode of processing of form X_Y with different options for JPSS and Metop MISSIONs. Y is IRMW for infrared + microwave (JPSS: CrIS+ATMS; Metop: IASI+AMSUA/MHS) or IRO for infrared-only. For JPSS, X is LR or HR, where LR means NSR (SCRIS) and HR means FSR (SCRIF). X for Metop selects how to disaggregate IASI L1C files into 22/23 MDRs per granule; S to include overpass start line, E to include end line, and M to prefer the middle of overpass. S,M,E matter only for DB overpasses; they are ignored, or can be omitted, for granular IASI L1C. [default: -mode=HR_IRMW (JPSS), -mode=M_IRMW (Metop)].

-o OUTPUT, --output=OUTPUT

Output option string of form OSR where O is for OLR (outgoing longwave radiation), S is for Subsetter, & R is for Retrieval. Any combination of these letter options is valid. [default: --output==R].

-w WORK_DIR, --work=WORK_DIR

The directory in which all activity will occur, this is also where output files will be written. [default: current directory].

-d DYNAMIC_ANC, --dynanc=DYNAMIC_ANC

Path to root of dynamic ancillary data tree; beneath here data are organized as per SSEC remote ancillary server for CSPP, i.e. %Y %m %d %j where %Y is 4-digit year, %m is 2-digit month, %d is 2-digit day of month and %j is 3-digit day of year. [default: environment var. \$CSPP_DYNAMIC_ANCIL_DIR].

-f N_FOVS, --fovs=N_FOVS

The number of fields of view per CrIS field of regard. Added to accommodate loss of FOV.4 from DB data stream starting 1-Dec-2014. For IDPS data, and DB data before 01-Dec-2014, you can use --fovs=9 (or -f 9) and for DB data before 18-Aug-2015 use --fovs=8 (or -f 8). This has no meaning for Metop missions. [default: --fovs=7 for npp, --fovs=9 for j01].

-p PROCESSORS, --processors=PROCESSORS

Number of cpus to use, parallel only within overpass. [default: --processors=1].

-v, --verbose

Each occurrence increases verbosity 1 level from [default: ERROR only] -v=WARNING -vv=INFO -vvv=DEBUG

-c, --clean

Deletes temporary working directory under WORK_DIR at process completion.

-q, --quarantine

[default: --clean; --noclean is the contrary option]. Processing machine is not connected to the web for the purpose of fetching required dynamic ancillary data. [default: --noquarantine].

-a, --aggregate

Aggregates NUCAPS-EDR and NUCAPS-CCR-AR output files into swaths of up to 30 minutes as a POST-PROCESSING step. Variables are reshaped to a non-standard form and files named NUCAPS-AGG-EDR & NUCAPS-AGG-CCR-AR to distinguish from standard format & length granules. [default: --agg, --noagg is the contrary].

-b, --break

Breaks (disaggregates) 'npp' and 'j01' SDRs into 32s granules as a PRE-PROCESSING step. If input SDRs are provided ALREADY as 32-second granules, choosing the contrary option, --nobreak, saves both time and temporary disk space. No effect on Metop processing. [default: --break, --nobreak or -nob is the contrary].

-x, --xfiles

Use non-DB filenames for AMSUA/MHS L1B data sources. DB filenames are like {amsua,mhs}l1b_M0{1,2,3}*.l1b. Non-DB filenames are like *.{AMAX,MHSX}.M{1,2,3}.* and under this option IASI L1C files are assumed granular, i.e. comprise 22/23 MDRs per file. [default: -nox, i.e. Direct Broadcast filenames].

2.3 Dynamic ancillary data for CSPP_HEAP

Dynamic ancillary data requirements are National Center for Environmental Prediction (NCEP) Global Forecast System (GFS) 0.5 degree GRIB2 format files named like `gfs.t12z.pgrb2.0p50.f006.20180814`. The software chooses the two GFS 3 hour forecast files that bound the input data time range. For the installation test (**Section 3.2**), these files are provided in the distribution, so you have the option to run disconnected from the internet (quarantine mode). In quarantine mode, the user is responsible for obtaining and staging these data. For an example of a full day of GFS data in GRIB2 format, see https://jpssdb.ssec.wisc.edu/cspp_v_2_0/ancillary/2020_06_02_154/ (which also includes other curated ancillary data to support all of our CSPP software packages).

If the preferred GFS files are not yet available at runtime, CSPP_HEAP will use files from the model run 6 hours previous. There will be some impact on retrieval quality, and a warning similar to the following will be sent to STDERR:

```
run_heap.pl::gfs_fetch_with_fallback: WARNING: Cannot link to
/path/to/gfs.t12z.pgrb2.0p50.f006.20200602 at /install_dir/scripts/run_heap.pl line 1052.
```

If two GFS forecast files that bound the input data time range prior to the 15-hour forecast time **cannot be found**, a warning is issued and the affected granules **are not processed**.

Section 3: Running the Software

3.1 CSPP_HEAP Driver Script

Bash script `$CSPP_HEAP_HOME/scripts/run_heap.bash` checks for environment variable `CSPP_HEAP_HOME` and then invokes the Perl script `$CSPP_HEAP_HOME/scripts/run_heap.pl`, which contains all the remaining logic to organize HEAP processing (i.e. create a temporary directory, fetch dynamic ancillary data requirements, de-aggregate input SDRs or Level1-b files into granules, execute the HEAP preprocessor and retrieval (and other optional steps), optionally aggregate the output into overpass length files, copy output NetCDF4 files to target directory, and purge the temporary working directory).

Script `run_heap.bash` requires, at a minimum, the name of the satellite, and the name of the directory containing the SDRs or Level1-b files (infra-red and micro-wave satellite data files in the same directory). By default, the dynamic ancillary data location is set by environment variable `CSPP_DYNAMIC Ancil_DIR`, but it can be over-ridden on the command line, e.g.,

```
run_heap.bash -s npp -i /your_npp_data -d /your_dynamic_anc_data
```

Many command line options are available for `run_heap.bash`. Overleaf is a table of annotated examples that the user may find helpful.

Table 5: CSPP_HEAP example command line executions

NOAA-20 with all non-mandatory arguments set to default
<code>run_heap.bash -s j01 -i /SDR/cris_atms</code>
Process the NOAA-20 (j01) CrIS and ATMS SDRs found in directory /SDR/cris_atms/. The processing mode is not specified and so defaults to HR_IRMW meaning that both CrIS and ATMS SDRs are used. The HR means “high-resolution” which, in this context, means use the full-spectral resolution CrIS SDRs, i.e. the SCRIF files. The working directory, which is also the output directory, is unspecified; it defaults to the current working directory, \$PWD. Use default 1 CPU, verbosity is unspecified and so is quiet.
Metop-C with all non-mandatory arguments set to default except #CPUs
<code>run_heap.bash -s m03 -i /L1B/iasi_amsua_mhs -p 6</code>
Process the Metop-C (m03) IASI, AMSUA and MHS L1B files found in directory /L1B/iasi_amsua_mhs/. The working directory, which is also the output directory, is unspecified; it defaults to the current working directory, \$PWD. Use 6 CPUs, verbosity is unspecified and so is quiet. Processing mode is not specified and so defaults to M_IRMW meaning that both IASI and a pair of AMSUA/MHS files are used. The M means “middle” so that the middle of the IASI overpass is preferred. [NOTE: Direct Broadcast (DB) Metop IASI files need to be disaggregated into granular form before being presented to HEAP, but the length of the DB IASI files may not be an integral number of IASI granules. Thus a decision must be made whether to discard scanlines at the start of the overpass, at the end of the overpass, or to discard some at both ends i.e. to prefer the middle of the overpass. The default is the latter.]
Suomi-NPP, infra-red only, output directory specified, long form commands, multi-CPU
<code>run_heap.bash --sat=npp --input=/SDR/cris --mode=HR_IRO --work=/EDR/out -p 8</code>
Process the Suomi-NPP files CrIS files in directory /SDR/cris/ using mode HR_IRO (high resolution CrIS, i.e. SCRIF files, and infra-red only) and with outputs written to directory /EDR/out/. In this example we used mostly the long-form of command line arguments. Use 8 CPUs (short form command), and verbosity is quiet.
Metop-B, infra-red only, with retrieval done in parallel using 4 CPUs
<code>run_heap.bash -s m01 -i . -d /data/dynanc -w work/m01 -m M_IRO -p 4 -vv</code>
Process the Metop-B (m01) IASI L1B files in current working directory, writing outputs to directory work/m01/ and using 4 CPUs with a moderate amount of processing information sent to stdout. The dynamic ancillary data are in /data/dynanc/ and, if not found, are downloaded to there from the remote site stored in environment variable \$JPSS_REMOTE_ANC_DIR
Metop-A with outgoing longwave radiation (OLR) and channel subset files also made
<code>run_heap.bash -s m02 -i . -o OLR -w /output -p 4 --quarantine</code>
Process the Metop-A (m02) IASI L1B files in current working directory, using default processing mode M_IRMW, writing outputs to directory /output/ and using 4 CPUs. Also generate the outgoing longwave radiation and channel subset NetCDF files. Run in quarantine mode, meaning that no dynamic ancillary data will be downloaded – these data will be sought in the directory specified by environment variable \$CSPP_DYNAMIC_ANCIL_DIR; if they are missing CSPP_HEAP will not process the affected granules.

NOAA-20 without disaggregating already disaggregated inputs, multi-CPU and verbose
<pre>run_heap.bash -s j01 -i /SDR/cris_atms -p 12 --nobreak -vvv</pre>
Process the NOAA-20 (j01) CrIS and ATMS SDRs found in directory /SDR/cris_atms/. Assume that the input SDRs are not aggregated and so save time by not disaggregating them unnecessarily. The processing mode is not specified and so defaults to HR_IRMW. The working directory, which is also the output directory, is unspecified; it defaults to the current working directory, \$PWD. Uses 12 CPUs, verbosity is set to maximum output.
Suomi-NPP with low spectral resolution inputs, keep the temporary working directory
<pre>run_heap.bash -s npp -i /SDR/cris_atms -m LR_IRMW -w /EDR/out -p 8 --noclean -v</pre>
Process the Suomi-NPP files CrIS and ATMS SDRs in directory /SDR/cris_atms/ using mode LR_IRMW (low resolution CrIS, i.e. SCRIS files, plus ATMS) and with outputs written to directory /EDR/out/. The temporary working directory is kept which can be useful for debugging and for accessing files that are not copied out to the working directory. Uses 8 CPUs (short form command), and verbosity is low but not silent.
Metop-A using granular, non-DB data
<pre>run_heap.bash -s m02 -i /imported -o OR -p 4 --xfiles</pre>
Process the Metop-A (m02) IASI L1B files in directory /imported/, using 4 CPUs in default processing mode M_IRMW but with granular IASI data and AMSUA and MHS files named like NSS.AMAX.M1.* and NSS.MHSX.M1.* (instead of like amsual1b_M01_*.l1b and mhs11b_M01_*.l1b). Outputs are written to current working directory and include the outgoing longwave radiation (OLR files).
NOAA-20 just running the pre-processor (an undocumented command option)
<pre>run_heap.bash -s j01 -i /SDR/cris_atms -o P -w /output -p 10</pre>
Pre-process the NOAA-20 (j01) CrIS and ATMS SDRs found in directory /SDR/cris_atms/ using 10 CPUs. Write to directory /output/ files named like NUCAPS-ALL-HR_v2r0_j01_*.nc. This might be useful for other processing algorithms that require just these files as quickly as possible.
Suomi-NPP using non-DB data and without post-aggregating the outputs
<pre>run_heap.bash -s npp -i /SDR/cris_atms -w /EDR/out -o RS --fovs=9 -noa</pre>
Process the Suomi-NPP files CrIS files in directory /SDR/cris_atms/ using default mode HR_IRMW (high resolution CrIS, i.e. SCRIF files, plus ATMS) and with outputs written to directory /EDR/out/. Make the channel subset files as well the retrievals. The SCRIF files are from NOAA CLASS and so have data in all 9 CrIS fields-of-view (--fovs=9). The final switch (-noa) indicates skipping the final step of aggregating the output NUCAPS-EDR and NUCAPS-CCR-AR files.

3.2 CSPP_HEAP Installation Verification

If you have downloaded the optional test data package, you can run a test overpass of each supported mission. In the following we assume that you have installed and configured CSPP_HEAP and it is ready to run (refer to **2.2 Installation of CSPP_HEAP Software**). Unpack the test data package.

```
tar -xzf CSPP_HEAP_TESTDATA_V1.0.2.tar.gz
```

Change directory and execute the verification script.

```
cd CSPP_HEAP_TESTDATA_1_0 ; ./verify_heap.bash
```

The verification script uses up to 6 CPUs at times; it takes about 30 minutes to an hour to complete (depending on the speed of your system). Success looks like this:

```
Running npp ...
work/npp : npp : EDR +++ PASSED
Running j01 ...
work/j01 : j01 : EDR +++ PASSED
Running m01 ...
work/m01 : m01 : EDR +++ PASSED
Running m02 ...
work/m02 : m02 : EDR +++ PASSED
Running m03 ...
work/m03 : m03 : EDR +++ PASSED
```

If this script reports PASSED for all cases, you are ready to start using CSPP_HEAP.

For an example of a processing transcript with more verbose output, see **Appendix A**.

NOTE: There can be small differences in data products if the user's system is not CentOS-6. The comparison script has been adjusted (relaxed) to allow for these small differences and report PASSED under CentOS-7 and CentOS-8. See **Appendix B**.

3.3 Visualizing HEAP Retrievals

The CSPP software suite includes a Sounder Quicklooks Package for plotting navigated sounder datasets derived from one of several software packages that create Level 2 atmospheric retrievals, including from HEAP. Retrievals of temperature, dew point, water vapor mixing ratio and relative humidity at a given atmospheric pressure level can be plotted from the NUCAPS Environmental Data Record (EDR) NetCDF files, as well as Skew-T Thermodynamic Diagrams at a given latitude/longitude.

The CSPP Sounder Quicklook Package is available for download from the CSPP website:

<https://cimss.ssec.wisc.edu/cspp/>

CSPP_HEAP Sounder Quicklook Version 1.0 supports unaggregated S-NPP, NOAA-20, Metop-A and Metop-B CSPP NUCAPS retrievals. A new version will be released shortly that supports all instruments in aggregated output format.

Figure 1 is the Sounder Quicklook plot of 500 hPa Dewpoint Temperatures created from the CSPP_HEAP test output NUCAPS-AGGR-EDR product file for Metop-C.

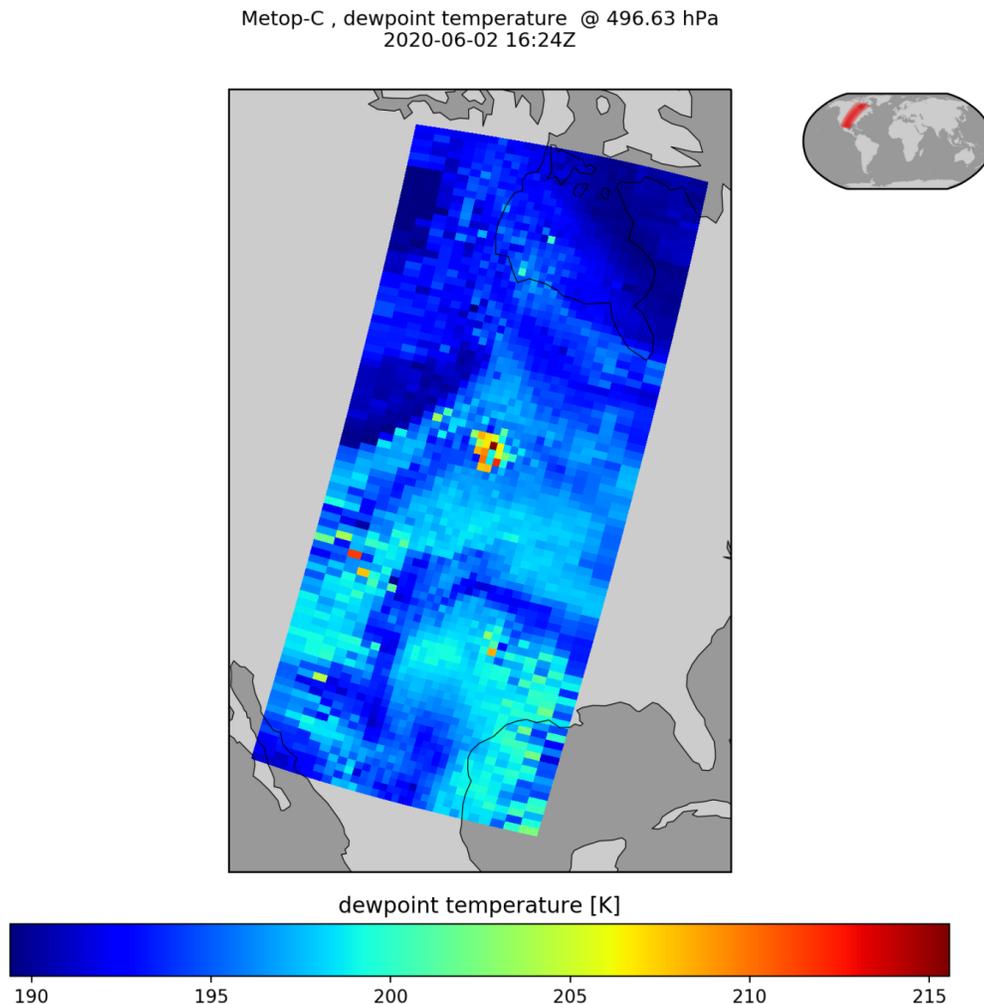


Figure 1: 496 hPa dew point temperature plot created from CSPP_HEAP retrievals from Metop-C (m03) on 2 June 2020.

Figure 2 is the CSPP Sounder Quicklook Skew-T Diagram created from the CSPP_HEAP test output NUCAPS-AGG-EDR product file for Metop-C at location 60.49 N, 88.63 W.

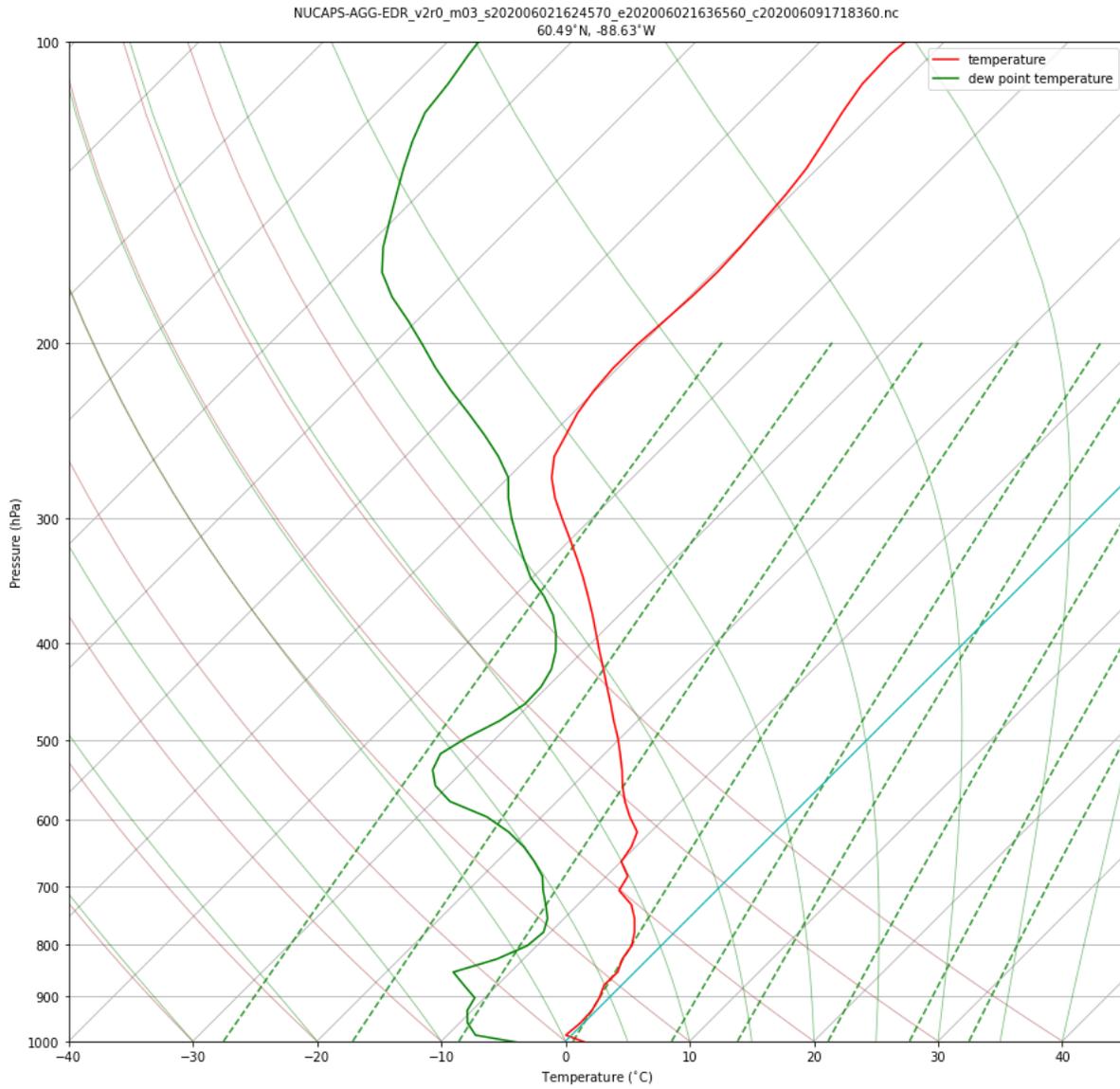


Figure 2: Skew-T diagram created from the CSPP_HEAP test output NUCAPS-AGG-EDR retrieval for 2 June 2020.

Appendix A: CSPP_HEAP NOAA-20 (j01) Example Output

In this example we are using mostly defaults except we have already downloaded matching GFS ancillary data into local directory `dynanc/` so we identify that directory and run in quarantine mode (`-d dynanc -q`). The input data is in local directory `db/sdr` and the results are written to directory `cspp/`. We use 8 CPUs (`-p 8`) and set verbosity level to 2 (`-vv`) which generates a fair amount of diagnostic output. Runtime should be less than 10 minutes on a modern desktop computer.

```
$ run_heap.bash -s j01 -d dynanc -p 8 -i db/sdr -w cspp -q -vv
```

```
-----
DEBUG %options
=====
'verbose' => 2,
'mode' => 'HR_IRMW',
'input_files' => 'db/sdr',
'processors' => 8,
'quarantine' => 1,
'aggregate' => 1,
'xfiles' => 0,
'gfs' => 3,
'break' => 1,
'dynanc' => '/data/jimd/Projects/CSPP_HEAP/developer/j01_test/dynanc',
'work' => '/data/jimd/Projects/CSPP_HEAP/developer/j01_test/cspp',
'clean' => 1,
'fovs' => 9,
'output' => 'PR',
'sat' => 'j01',
'nothing' => 0
DEBUG %options _END_
Making template HEAP_Preprocessor.PCF.template
Making template HEAP_Retrieval.PCF.template
-----
DEBUG %options
=====
'verbose' => 2,
'input_dir' => '/data/jimd/Projects/CSPP_HEAP/developer/j01_test/db/sdr',
'mode' => 'HR_IRMW',
'input_files' => [
  'GATM0_j01_d20200602_t1829066_e1840183_b13158_c20200602184251140810_cspp_dev.h5',
  'GCRSO_j01_d20200602_t1830479_e1838457_b13158_c20200602184819664222_cspp_dev.h5',
  'SCRIF_j01_d20200602_t1830479_e1838457_b13158_c20200602184819791155_cspp_dev.h5',
  'TATMS_j01_d20200602_t1829066_e1840183_b13158_c20200602184251141753_cspp_dev.h5'
],
'processors' => 8,
'quarantine' => 1,
'aggregate' => 1,
'process_dir' => [
  'swath_001'
],
'xfiles' => 0,
'break' => 1,
'gfs' => 3,
'dynanc' => '/data/jimd/Projects/CSPP_HEAP/developer/j01_test/dynanc',
'work' => '/data/jimd/Projects/CSPP_HEAP/developer/j01_test/cspp',
'output' => 'PR',
'fovs' => 9,
'clean' => 1,
'sat' => 'j01',
'nothing' => 0
DEBUG %options _END_
Running ... Preprocessor for swath_001 granule_001
Running ... Preprocessor for swath_001 granule_002
Running ... Retrieval for swath_001 granule_001
Running ... Preprocessor for swath_001 granule_003
Running ... Retrieval for swath_001 granule_002
Running ... Preprocessor for swath_001 granule_004
Running ... Preprocessor for swath_001 granule_005
Running ... Retrieval for swath_001 granule_003
```

```
Running ... Preprocessor for swath_001 granule_006
Running ... Retrieval for swath_001 granule_004
Running ... Preprocessor for swath_001 granule_007
Running ... Retrieval for swath_001 granule_005
Running ... Preprocessor for swath_001 granule_008
Running ... Retrieval for swath_001 granule_006
Running ... Retrieval for swath_001 granule_007
Running ... Retrieval for swath_001 granule_008
Running ... Preprocessor for swath_001 granule_009
Running ... Preprocessor for swath_001 granule_010
Running ... Retrieval for swath_001 granule_009
Running ... Preprocessor for swath_001 granule_011
Running ... Retrieval for swath_001 granule_010
Running ... Preprocessor for swath_001 granule_012
Running ... Retrieval for swath_001 granule_011
Running ... Preprocessor for swath_001 granule_013
Running ... Preprocessor for swath_001 granule_014
Running ... Retrieval for swath_001 granule_012
Running ... Preprocessor for swath_001 granule_015
Running ... Retrieval for swath_001 granule_013
Running ... Retrieval for swath_001 granule_014
Running ... Retrieval for swath_001 granule_015
Running ... /data/jimd/Projects/CSPP_HEAP/scripts/aggregate_swath.pl -s j01 -p EDR -i . -o ..
Running ... /data/jimd/Projects/CSPP_HEAP/scripts/aggregate_swath.pl -s j01 -p CCR-AR -i . -o ..
```

Appendix B: Verification of CentOS-7 Results Against CentOS-6 Reference

CSPP_HEAP was built under CentOS-6 and is distributed for use under CentOS-6, -7 and -8. Our testing has shown that results from CentOS-7 and CentOS-8 are identical, but there can be small differences with CentOS-6. We compared several variables drawn from runs of CSPP_HEAP using the test data package (CSPP_HEAP_TESTDATA_V1.0.tar.gz). Across the 5 satellite missions in the test data set, just 5 out of 8670 profiles returned a different Quality_Flag when processed with CentOS-7 compared to CentOS-6. Those 5 profiles were dropped from the analysis. The geolocation was always exact. In the column headings of the tables below, EXACT is the % of values machine-precision identical (CentOS-6 vs CentOS-7), CLOSE is the % within the stated threshold, and OUTLIER the % of outliers (i.e. outside of the stated threshold). RMS is the root mean square error of just the outliers; in K for temperatures, in ppm for average column CO₂, and as percentage errors for mixing ratio profiles (because of their dynamic range). Shown also are the MIN, MEAN and MAX of each variable, drawn from the CentOS-6 runs. We believe that these small differences impact neither the quality nor the utility of the HEAP retrievals in any significant way.

Table B1: Skin_Temperature (K). Threshold = 0.01 K.

SAT	MIN / MEAN / MAX (K)	EXACT(%)	CLOSE(%)	OUTLIER(%)	RMS (K)
npp	255.97 / 285.62 / 314.05	91.32	8.06	0.62	0.818
j01	235.93 / 295.22 / 323.68	89.26	9.57	1.17	1.451
m01	201.34 / 292.72 / 400.00	94.55	3.63	1.82	2.005
m02	249.74 / 314.45 / 400.00	95.80	3.62	0.58	0.038
m03	210.96 / 293.79 / 400.00	95.26	4.11	0.63	0.067

Table B2: Temperature (K). Threshold = 0.1 K.

SAT	MIN / MEAN / MAX (K)	EXACT(%)	CLOSE(%)	OUTLIER(%)	RMS (K)
npp	178.2 / 242.7 / 313.2	91.95	7.56	0.49	0.409
j01	177.3 / 243.7 / 334.1	90.40	8.82	0.78	1.469
m01	175.7 / 244.1 / 360.6	94.90	5.05	0.04	0.319
m02	192.3 / 244.6 / 341.0	95.95	4.05	0.00	NA
m03	174.6 / 243.7 / 353.9	95.52	4.47	0.01	0.128

Table B3: Mean_CO2 (ppm). Threshold = 0.01 ppm.

SAT	MIN / MEAN / MAX (ppm)	EXACT(%)	CLOSE(%)	OUTLIER(%)	RMS (ppm)
npp	398.89 / 406.37 / 422.97	90.21	9.24	0.56	0.402
j01	397.06 / 406.91 / 418.45	88.93	10.13	0.95	0.449
m01	397.26 / 410.01 / 452.58	94.26	3.93	1.82	0.155
m02	393.66 / 411.21 / 454.59	95.22	4.06	0.72	0.083
m03	366.45 / 409.44 / 546.21	94.93	4.74	0.33	0.053

Table B4: H2O_MR (g/kg). Threshold = 0.001 g/kg.

SAT	MIN / MEAN / MAX (g/kg)	EXACT(%)	CLOSE(%)	OUTLIER(%)	RMS (%)
npp	0.00007818 / 1.00198 / 27.2422	89.82	9.78	0.40	2.171
j01	0.00012933 / 1.46448 / 24.7815	88.46	10.97	0.57	13.798
m01	0.00004128 / 2.04804 / 106.715	94.03	5.05	0.92	3.000
m02	0.00046604 / 3.58805 / 107.913	95.23	4.02	0.76	0.102
m03	0.00005008 / 2.37781 / 90.2779	94.79	4.79	0.41	0.115

Table B5: O3_MR (ppb). Threshold = 1 ppb.

SAT	MIN / MEAN / MAX (ppb)	EXACT(%)	CLOSE(%)	OUTLIER(%)	RMS (%)
npp	13.2 / 1818.8 / 16759.6	89.93	9.72	0.34	0.761
j01	13.1 / 1771.2 / 9790.1	88.57	10.91	0.52	0.830
m01	9.9 / 1541.8 / 33323.2	94.02	5.14	0.84	0.362
m02	10.6 / 1380.7 / 10125.0	95.23	4.29	0.49	0.142
m03	2.4 / 1507.5 / 18848.5	94.79	4.92	0.29	0.147