Water Vapour Climate Change Initiative (WV_cci) - CCI+ Phase 1

System Verification Report (SVR)

Ref: D3.3

Date: 19 August 2021

Issue: 3.0

For: ESA / ECSAT

Ref: CCIWV.REP.014

This Page is Intentionally Blank

Project	**:**	**Water Vapour Climate Change Initiative (WV_cci) - CCI+ Phase 1**
Document Title:		**System Verification Report (SVR)**
Reference	**:**	**D3.3**
Issued	**:**	**19 August 2021**
Issue	**:**	**3.0**
Client	**:**	**ESA / ECSAT**
Authors	**:**	Olaf Danne (Brockmann Consult), Michaela Hegglin (University of Reading), Hao Ye (University of Reading)
Copyright	**:**	Water_Vapour_cci Consortium and ESA

Document Change Log

Issue/ Revision	Date	Comment
0.1	April 2019	Initial issue
0.2	3 June 2019	First version for internal review
1.0	7 June 2019	First submitted version
1.1	31 August 2019	Revision based on v1.0 RIDs
2.0	31 July 2020	Scheduled document update
3.0	19 August 2021	Scheduled document update

TABLE OF CONTENTS

INDEX OF FIGURES

This Page is Intentionally Blank

1. INTRODUCTION

1.1 Purpose

Verification is the process to demonstrate that the system meets the specified requirements [1]. With regard to the ESA Climate Change Initiative (CCI) WV_cci project this means that the processing systems are able to generate the water vapour output products as specified in the Product Specification Document [2] and to provide the final data products to users. This comprises the identification of what shall be verified, the definition of how it shall be verified, the execution, and the reporting on pass or failure. Verification methods used are tests, inspection, and monitoring.

The systems to be verified consist of the dedicated subsystems for the generation of (a) TCWV products and (b) VRWV products with their corresponding subcomponents as described in the System Specification Document [3].

1.2 Scope

This System Verification Report (SVR) v3.0 documents the system verification activities of the ESA Climate Change Initiative (CCI) WV_cci project for its elements used over the project period.

2. SYSTEMS UNDER TEST

The systems under test are the dedicated subsystems for the generation of (a) TCWV products and (b) VRWV products with their corresponding subcomponents as shown in Figure 2-1. Both systems are described in more detail in [3].

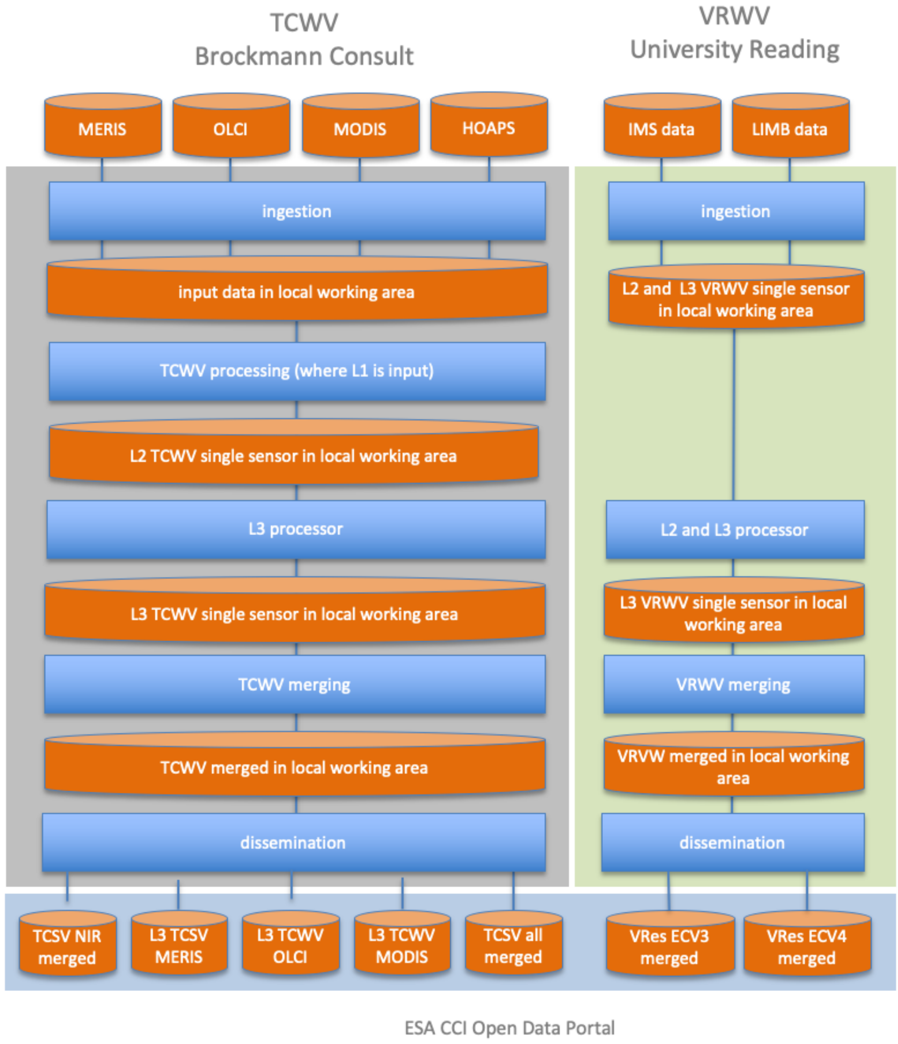


Figure 2-1: Symmetric system definition of the TCWV and VRWV processing systems. Taken from [4].

3. VERIFICATION APPROACH

3.1 TCWV Processing System

3.1.1 Preprocessing

The preprocessing steps applied in the TCWV retrieval are (see [3]):

- Pixel classification (i.e. cloud identification, land/water flagging) and reflectance band[1] subsetting for TCWV retrieval from MERIS and MODIS L1b data;
- Radiance-to-reflectance conversion in case of MERIS;
- Ingestion of ancilliary variables from ERA Interim datasets [5];
- Temporal interpolation and spatial collocation of ERA Interim onto L1b product grid.

3.1.1.1 Code sanity checks

A number of code sanity checks have been introduced into the pre-processing code. The most important ones are:

- the pixel classification fails if the input product is regarded as invalid (i.e. does not follow the corresponding L1b product specifications);
- the pixel classification fails if ancillary data (e.g. neural nets, MERIS L2 auxdata) cannot be loaded;
- the pixel classification fails if additional input products (e.g. MODIS cloud product, Land/Water mask product) do not follow specifications;
- the MERIS and OLCI radiance-to-reflectance conversion fails if geometry tie point grind (i.e. SZA) are missing in the source product;
- any of the preprocessing steps fails if the source product(s) have no geo-information;
- in case of MODIS, any of the preprocessing steps fails if the L1b source product is not flagged as a 'Day' product.

---

[1] *The term 'band' is used throughout this document. It follows the SNAP naming conventions and is equivalent to a NetCDF variable containing 2D raster data.*

3.1.1.2 Unit-level testing

The preprocessing modules have been subsequently unit-level tested. The reports are given below in a summary format. As for all other processing modules, all these tests must pass successfully in order to build and deploy the software on the dedicated processing systems (JASMIN for MODIS L2 processing, and Calvalus for all other steps).

```
[INFO] -------------------------------------------------------
[INFO]  T E S T S
[INFO] -------------------------------------------------------
[INFO] Running org.esa.s3tbx.idepix.core.util.IdepixIOTest
[INFO] Tests run: 7, Failures: 0, Errors: 0, Skipped: 0, Time
elapsed: 0 s
[INFO] Results:
[INFO] Tests run: 7, Failures: 0, Errors: 0, Skipped: 0

[INFO] -------------------------------------------------------
[INFO]  T E S T S
[INFO] -------------------------------------------------------
[INFO]Running
org.esa.snap.idepix.meris.dataio.nc4.IdepixMerisNc4WriterLoaded
AsServiceTest
writerPlugIn.Class = class
org.esa.snap.idepix.meris.dataio.nc4.IdepixMerisNc4WriterPlugIn
writerPlugIn.Descr = SNAP IDEPIX-MERIS NetCDF4 products
[INFO] Tests run: 1, Failures: 0, Errors: 0, Skipped: 0, Time
elapsed: 0.141 s
[INFO] Running org.esa.snap.idepix.meris.IdepixMerisOpTest
[INFO] Tests run: 1, Failures: 0, Errors: 0, Skipped: 0, Time
elapsed: 0.19 s
[INFO] Results:
[INFO] Tests run: 2, Failures: 0, Errors: 0, Skipped: 0

-------------------------------------------------------
 T E S T S
-------------------------------------------------------
[INFO] Running
org.esa.snap.idepix.modis.dataio.nc4.IdepixModisNc4WriterLoaded
AsServiceTest
writerPlugIn.Class = class
org.esa.snap.idepix.modis.dataio.nc4.IdepixModisNc4WriterPlugIn
writerPlugIn.Descr = SNAP IDEPIX-MODIS NetCDF4 products
```

```
[INFO] Tests run: 1, Failures: 0, Errors: 0, Skipped: 0, Time
elapsed: 0.169 sec

[INFO] Running org.esa.snap.idepix.modis.IdepixModisOpTest

[INFO] Tests run: 1, Failures: 0, Errors: 0, Skipped: 0, Time
elapsed: 0.169 sec

[INFO] Results :

[INFO] Tests run: 2, Failures: 0, Errors: 0, Skipped: 0


-------------------------------------------------------

 T E S T S

-------------------------------------------------------

[INFO] Running org.esa.snap.idepix.olci.IdepixOlciOpTest

[INFO] Tests run: 1, Failures: 0, Errors: 0, Skipped: 0, Time
elapsed: 0.238 s - in org.esa.snap.idepix.olci.IdepixOlciOpTest

[INFO] Running org.esa.snap.idepix.olci.IdepixOlciUtilsTest

[INFO] Tests run: 6, Failures: 0, Errors: 0, Skipped: 0, Time
elapsed: 0.047 s - in
org.esa.snap.idepix.olci.IdepixOlciUtilsTest

[INFO] Running
org.esa.snap.idepix.olci.TensorflowNNCalculatorTest


[INFO] Tests run: 3, Failures: 0, Errors: 0, Skipped: 0, Time
elapsed: 0.701 s - in
org.esa.snap.idepix.olci.TensorflowNNCalculatorTest

[INFO]

[INFO] Results:

[INFO]

[INFO] Tests run: 10, Failures: 0, Errors: 0, Skipped: 0


-------------------------------------------------------

T E S T S

-------------------------------------------------------

[INFO] Running org.esa.s3tbx.processor.rad2refl.Rad2ReflTest

[INFO] Tests run: 1, Failures: 0, Errors: 0, Skipped: 0, Time
elapsed: 0.17 s

[INFO] Results:

[INFO] Tests run: 1, Failures: 0, Errors: 0, Skipped: 0


-------------------------------------------------------

T E S T S

-------------------------------------------------------
```

```
[INFO] Tests run: 4, Failures: 0, Errors: 0, Skipped: 0, Time
elapsed: 0.169 s

[INFO] Running org.esa.s3tbx.meris.l2auxdata.AuxFileTest

[INFO] Tests run: 1, Failures: 0, Errors: 0, Skipped: 0, Time
elapsed: 0 s

[INFO] Running org.esa.s3tbx.meris.l2auxdata.UtilsTest

[INFO] Tests run: 2, Failures: 0, Errors: 0, Skipped: 0, Time
elapsed: 0.17 s

[INFO] Results:

[INFO] Tests run: 7, Failures: 0, Errors: 0, Skipped: 0
```

3.1.1.3 Monitoring

The preprocessing is executed on both JASMIN and Calvalus processing systems using the PMonitor Python framework (see [3]) initializing the parallel processing using a dedicated number of nodes. This framework provides various options for job monitoring as well as for logging and reporting on different levels of detail.

As an example, this is illustrated for one month (31 days) of MERIS preprocessing on Calvalus. This task consists of 62 jobs (31 jobs for pixel classification including radiance-to-reflectance conversion, and 31 jobs for ingestion of ERA Interim products with interpolation/colocation). Each of these jobs subsequently processes the single MERIS L1b of one day (usually ~13 products).

When the processing is initiated, a status file is generated which is monitoring the status of running, finished, and queued jobs. The file is updated every 10 seconds. Here it looks like this:

```
Every 10.0s: less tcwv_meris_preprocessing.status
Mon Apr 29 15:06:34 2019

62 created, 31 running, 31 backlog, 0 processed, 0 failed


Every 10.0s: less tcwv_meris_preprocessing.status
Mon Apr 29 15:10:14 2019

62 created, 11 running, 8 backlog, 43 processed, 0 failed


Every 10.0s: less tcwv_meris_preprocessing.status
Mon Apr 29 15:12:30 2019

62 created, 0 running, 0 backlog, 62 processed, 0 failed
```

Initially, 31 (of max. 32) jobs are running. These 31 jobs are doing the pixel classification. For each day, the pixel classification job must be finished before the ERA Interim ingestion job can be started, as the pixel classification output is needed as mandatory input. The waiting jobs are kept in a backlog. The PMonitor also takes care for this kind of processing conditions. After a while, a bunch of jobs (43) has been processed successfully, with a few jobs still running and a few in the backlog. The final state should always look like in the last line, with all created jobs finally processed successfully.

For each job, a dedicated log file is written, and in case of successful processing an entry is generated in a report file. Examples for those are given in Appendix 3: Processing logs and reports.

As a final check, it must be verified that the number of generated preprocessed products is as expected. For MERIS on Calvalus it is the same as the number of L1b input products:

```
ls /calvalus/eodata/MER_RR__1P/r03/2011/08/*/*.N1 |wc -l
444
```

```
ls /calvalus/projects/wvcci/idepix-era-
interim/meris/2011/08/*/*.nc |wc -l
444
```

For MODIS Terra on JASMIN we have to keep in mind that only 'Day' products are preprocessed. For the other products, the pixel classification processes will immediately terminate with an 'expected failure' with no output generated.

```
ls
/gws/nopw/j04/esacci_wv/odanne/WvcciRoot/L1b/MODIS_TERRA/2011/0
8/*/*.hdf |wc -l
8928
```

```
ls
/gws/nopw/j04/esacci_wv/odanne/WvcciRoot/IdepixEraInterim/MODIS
_TERRA/2012/08/01/*.nc |wc
3438
```

We have a total number of 8928 (31*288) L1b inputs, as there are 288 products per day (1 every 5 minutes), but the number of preprocessed products is just 3438. Now we need to verify that the number of missing products is consistent with the exclusion of non-'Day' products. This check is being logged in each single process, so we count the occurrence of this specific log output from all log files (pattern *.out):

```
grep "SKIP nightly product" *.out |wc
5490
```

This makes sense, as 5490 is the number of excluded products (8928 – 3438).

3.1.1.4 Visual inspection

There have been lots of visual inspections of single preprocessing result products, i.e. the pixel classification products should provide a correct flagging of land/sea pixels, and a 'reasonable' cloud detection[2]. After the ingestion of ERA Interim products, the ancillary variables for temperature, pressure, wind speed, and prior TCWV should be present in the result products, which in return serve as input products for the L2 TCWV retrieval. Figure 3-1 shows an example for MERIS: after the ERA Interim ingestion, the pixel classification product contains the additional variables t2m (temperature in 2m height), msl (mean sea level pressure), u10/v10 (horizontal wind) and tcwv (prior total column of water vapour). Together with the TOA reflectances in MERIS bands 13–15, they are mandatory input for the L2 TCWV retrieval.

For visual inspection of the generated NetCDF products we usually use the SNAP desktop application. For inspecting product metadata (global and variable attributes), ncdump is a useful tool as well.

[2] *An improvement/optimisation of the MERIS, OLCI and MODIS cloud detections is not part of this verification step. This is done in separate projects/activities (e.g. continuous work on the IdePix pixel classification processor at BC). For WV_cci, we use the cloud masking regarded as currently 'best available' (see text for more details).*

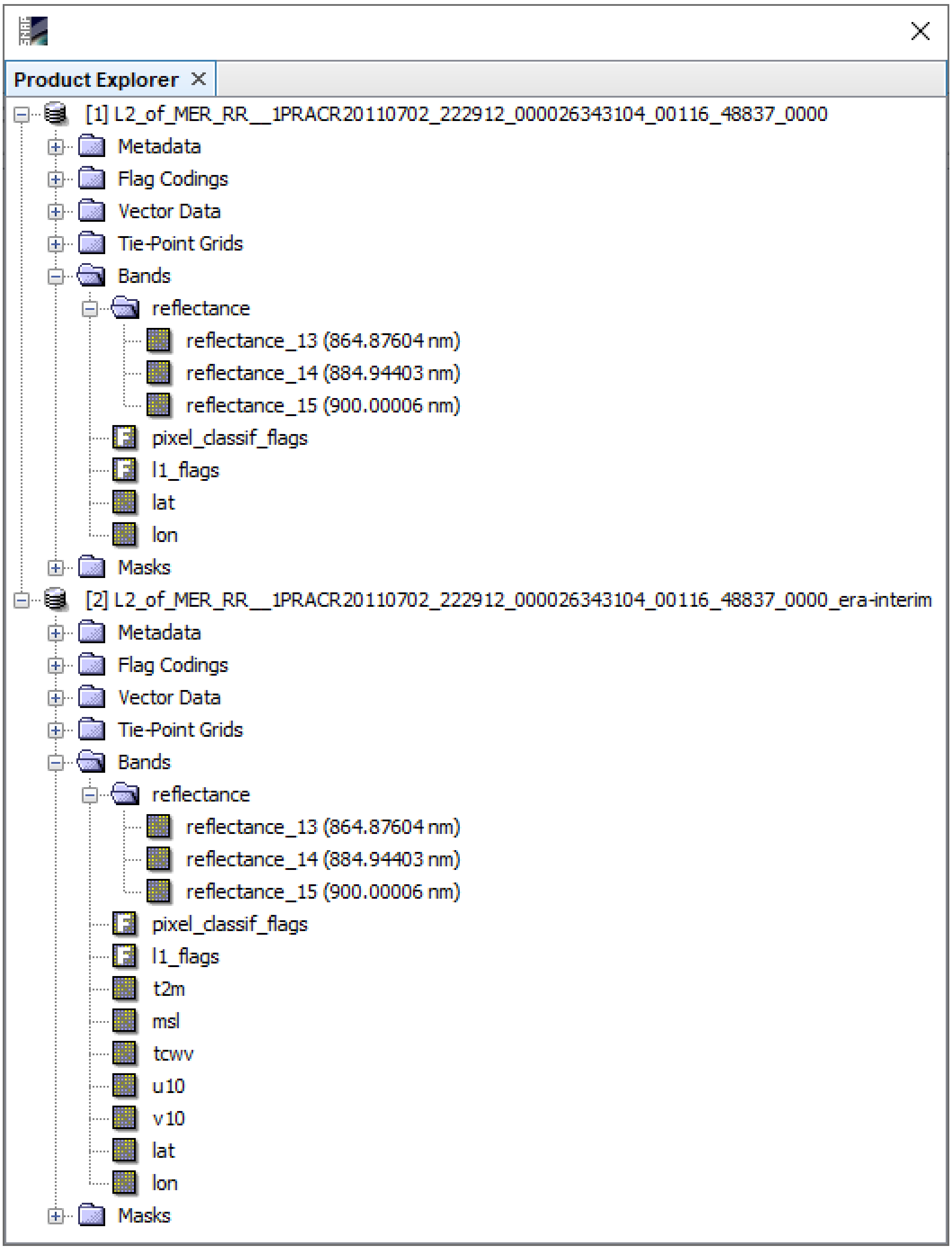


Figure 3-1: SNAP Product Explorer: MERIS IdePix pixel classification product [1], and same product after ingestion of ERA Interim ancillary variables [2].

Figure 3-2 shows the IdePix cloud classification for the MERIS RR product 20081023_085102. The cloud identification works obviously well over both land and water. The challenge in this example was the separation of clouds from sun glint which is present over the water surface in the left half of the image. The image on the right gives an idea of the distribution of the corresponding prior TCWV from ERA Interim.

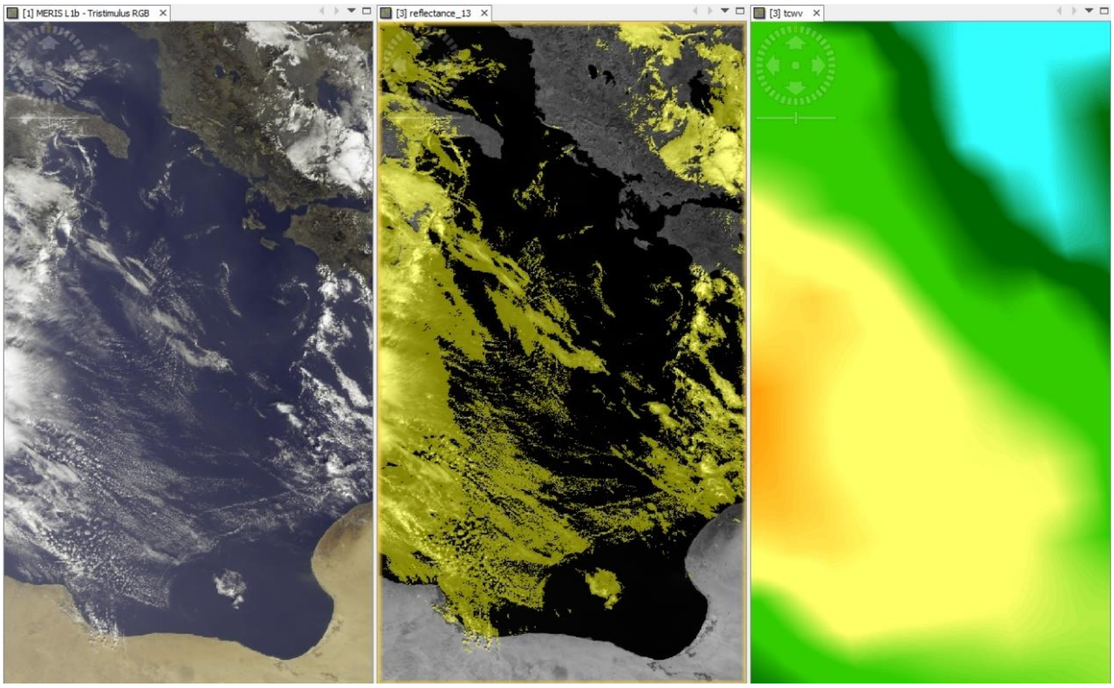


Figure 3-2: Example of a MERIS IdePix pixel classification product (20081023_085102) in the Mediterranean area with ingested ERA Interim ancillary variables: RGB (left), pixels flagged as cloud (yellow, centre), ERA Interim prior TCWV (right, on a scale 0–40 kg m^{-2}).

Figure 3-3 shows the cloud classification for the MODIS product MOD021KM.A2011196.1050. The upper right image shows the pixels flagged as cloud in MOD35 L2 cloud product, the lower images show pixels flagged as cloud by IdePix in a 'cloud conservative' mode (left) and a 'clear sky conservative' mode (lower right). In WV_cci, the MOD35 L2 cloud product is used as default cloud classification, as it is known to have good quality with a very comprehensive algorithm behind [6]. Moreover, for the MODIS Terra target years in WV_cci (2011–2017), the MOD35 L2 dataset is fully available from the NEODC database for the processing on the JASMIN cluster at Harwell/UK. In case single MOD35 L2 products are not available, the IdePix classification for MODIS is used as fallback. The algorithm behind it was developed at BC and is based on a neural net with some additional threshold tests for certain conditions. As we can see from Figure 3-3, the classification in 'cloud conservative' mode looks quite similar to the MOD35 L2 classification.

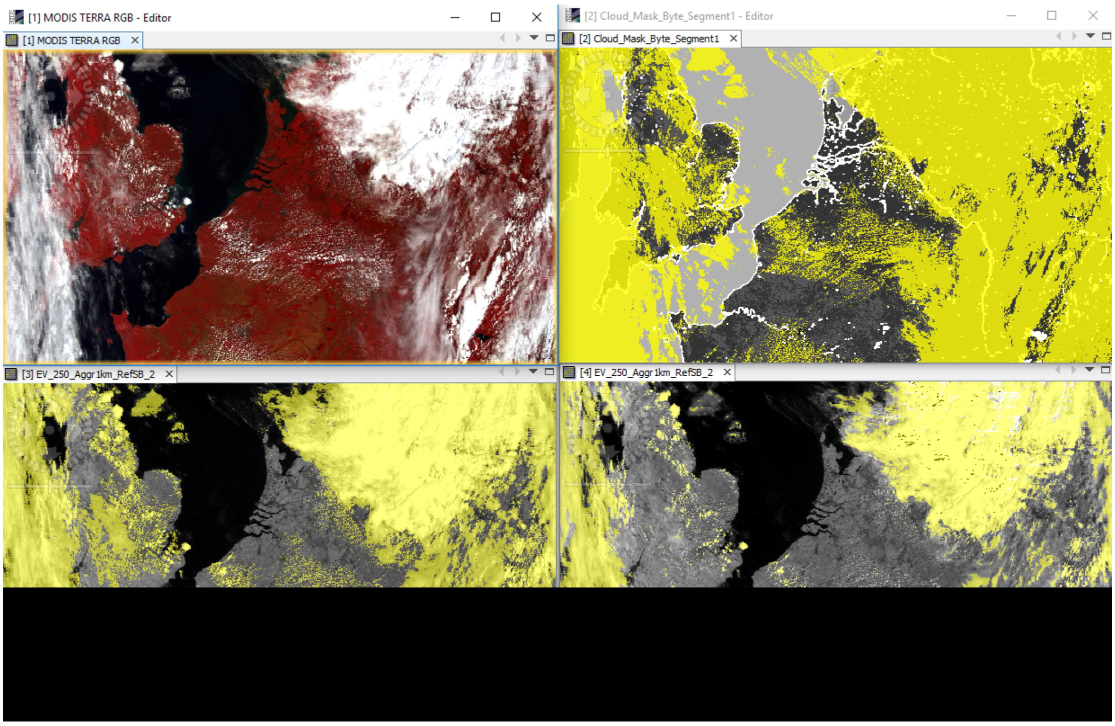


Figure 3-3: Example of cloud identification for MODIS (MOD021KM.A2011196.1050) over Western Europe: RGB (upper left), pixels flagged as cloud in MOD35 L2 cloud product (upper right), pixels flagged as cloud by IdePix in 'cloud conservative' mode (lower left) and 'clear sky conservative' mode (lower right).

Figure 3-4 shows the IdePix cloud classification for the OLCI RR product S3A_OL_1_ERR____20160715T100417. As for the other sensors, the cloud identification works obviously well over both land and water. It is rather conservative which is an advantage as the TCWV retrieval can be quite sensitive to undetected cloudy pixels which are not removed in advance.

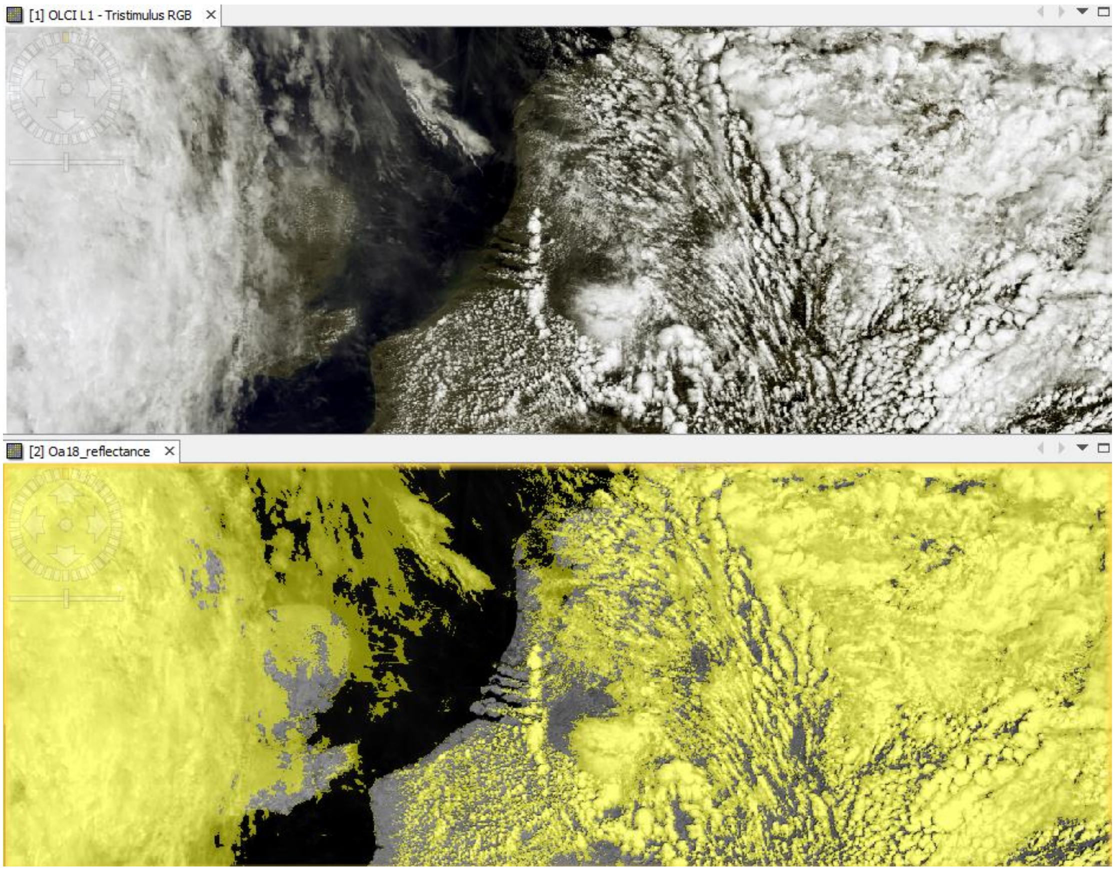


Figure 3-4: Example of an OLCI IdePix pixel classification product (S3A_OL_1_ERR____20160715T100417) over Western Europe: RGB (top), pixels flagged as cloud by IdePix (bottom).

3.1.2 TCWV L2 processing

3.1.2.1 Code sanity checks

A number of code sanity checks have been introduced into the TCWV L2 processing code. The most important ones are:

- the TCWV L2 processing completely fails if the input product is regarded as invalid (i.e. does not contain the input bands required for the retrieval);
- the TCWV L2 processing completely fails if any of the source products have no geo-information;
- the TCWV L2 processing completely fails if ancillary data (e.g. neural nets, MERIS L2 auxdata) cannot be loaded;

- the TCWV L2 processing fails for single pixels if one or more input reflectances are outside their valid range specified by the algorithm;
- the TCWV L2 processing fails for single pixels in case of retrieval failures (e.g. numerical issues);
- if a MOD35 L2 cloud product is available and used for processing, it must fulfil its product specifications. Otherwise the fallback (IdePix for MODIS) is used.

3.1.2.2 Unit-level testing

The TCWV L2 processing module has been comprehensively unit-level tested. Again, all these tests must pass successfully in order to build and deploy the software on the dedicated processing systems (JASMIN for MODIS L2 processing, and Calvalus for all other steps).

3.1.2.2.1 Testing of subcomponents

The TCWV L2 processing module consists of various subcomponents (data input/output, numerics for interpolation and optimal estimation, lookup table access etc.). The reports of the unit-level tests of these components are given below in a summary format.

```
-------------------------------------------------------
 T E S T S
-------------------------------------------------------
Running
org.esa.snap.wvcci.tcwv.dataio.erainterim.EraInterimProductRead
erLoadedAsServiceTest

Tests run: 1, Failures: 0, Errors: 0, Skipped: 0, Time elapsed:
0.191 sec


Running
org.esa.snap.wvcci.tcwv.dataio.erainterim.ScripGeocodingWriterL
oadedAsServiceTest

writerPlugIn.Class = class
org.esa.snap.wvcci.tcwv.dataio.erainterim.ScripGeocodingWriterP
lugIn

writerPlugIn.Descr = NetCDF following SCRIP convention

Tests run: 1, Failures: 0, Errors: 0, Skipped: 0, Time elapsed:
0.008 sec
```

```
Running
org.esa.snap.wvcci.tcwv.dataio.mod35.ModisMod35L2ReaderLoadedAs
ServiceTest

readerPlugIn.Class = class
org.esa.snap.wvcci.tcwv.dataio.mod35.ModisMod35L2ProductReaderP
lugIn

readerPlugIn.Descr = MODIS35 Format

Tests run: 1, Failures: 0, Errors: 0, Skipped: 0, Time elapsed:
0.002 sec


Running
org.esa.snap.wvcci.tcwv.dataio.mod35.ModisMod35L2UtilsTest

p.getStartTime() = 15-JUL-2011 10:55:00.000000

p.getEndTime() = 15-JUL-2011 10:55:00.000000

Tests run: 2, Failures: 0, Errors: 0, Skipped: 0, Time elapsed:
0.482 sec


Running
org.esa.snap.wvcci.tcwv.dataio.nc4.SnapWvcciNc4WriterLoadedAsSe
rviceTest

writerPlugIn.Class = class
org.esa.snap.wvcci.tcwv.dataio.nc4.SnapWvcciNc4WriterPlugIn

writerPlugIn.Descr = SNAP WV_cci NetCDF4 products

Tests run: 1, Failures: 0, Errors: 0, Skipped: 0, Time elapsed:
0 sec


Running
org.esa.snap.wvcci.tcwv.dataio.nc4.WvcciNc4ComplianceWriterTest

Tests run: 1, Failures: 0, Errors: 0, Skipped: 0, Time elapsed:
0 sec


Running
org.esa.snap.wvcci.tcwv.interpolation.TcwvInterpolationUtilsTes
t

Tests run: 6, Failures: 0, Errors: 0, Skipped: 0, Time elapsed:
0.151 sec


Running org.esa.snap.wvcci.tcwv.oe.OptimalEstimationTest

Tests run: 7, Failures: 0, Errors: 0, Skipped: 0, Time elapsed:
0.01 sec


Running org.esa.snap.wvcci.tcwv.oe.OptimalEstimationUtilsTest

Tests run: 6, Failures: 0, Errors: 0, Skipped: 0, Time elapsed:
0 sec


Running org.esa.snap.wvcci.tcwv.TcwvIOTest
```

```
Tests run: 1, Failures: 0, Errors: 0, Skipped: 0, Time elapsed: 0 sec


Running org.esa.snap.wvcci.tcwv.TcwvLutTest
time = 12.0 ms
Tests run: 3, Failures: 0, Errors: 0, Skipped: 0, Time elapsed: 0 sec


Running org.esa.snap.wvcci.tcwv.TcwvMerisLandLutTest
Tests run: 2, Failures: 0, Errors: 0, Skipped: 0, Time elapsed: 1.151 sec


Running org.esa.snap.wvcci.tcwv.TcwvMerisOceanLutTest
Tests run: 2, Failures: 0, Errors: 0, Skipped: 0, Time elapsed: 0.339 sec


Running org.esa.snap.wvcci.tcwv.TcwvModisLandLutTest
Tests run: 1, Failures: 0, Errors: 0, Skipped: 0, Time elapsed: 0 sec


Running org.esa.snap.wvcci.tcwv.TcwvModisOceanLutTest
Tests run: 2, Failures: 0, Errors: 0, Skipped: 0, Time elapsed: 0.432 sec


Running org.esa.snap.wvcci.tcwv.TcwvOpTest
Tests run: 1, Failures: 0, Errors: 0, Skipped: 0, Time elapsed: 0 sec


Running
org.esa.snap.wvcci.tcwv.util.MergeIdepixEraInterimOpTest
Tests run: 2, Failures: 0, Errors: 0, Skipped: 0, Time elapsed: 0.016 sec


Results :


Tests run: 39, Failures: 0, Errors: 0, Skipped: 0
```

3.1.2.2.2 Numerical verification against Python breadboard code

The most essential part of the verification of the TCWV retrieval is the per-pixel comparison of the computed TCWV values with the corresponding breadboard code which has been developed by Spectral Earth in line with the algorithm development and closely following the ATBD. This breadboard code was developed in Python and provides the full TCWV retrieval algorithm for single cloud-free pixels over ocean or over land, using the corresponding lookup tables. The TCWV L2 retrieval code used for the processing in WV_cci is basically a translation of the breadboard code into Java which allows a much more performant processing on the JASMIN and Calvalus platforms.

For the verification of the Java code, specific unit tests were developed which do a full TCWV retrieval for given input for single pixels over both land and water for all sensors (MERIS, OLCI and MODIS Terra. These tests must provide (within numerical tolerances) the same TCWV values as the breadboard code. In addition, these tests make use of all the subcomponents tested in the previous subsection, so this is another confirmation that the subcomponents work correctly.

The key lines of code in all these unit tests look like:

```
final TcwvResult result = TcwvAlgorithm.compute(…);

// breadboard TCWV result for ocean pixel is 28.007:
final double pythonBreadboardResult = 28.007;

assertEquals(pythonBreadboardResult, result.getTcwv(), 1.E-3);

System.out.println("MERIS OCEAN TCWV = " + result.getTcwv());
```

The reports of the 'full algorithm for one pixel' unit-level tests are given below in a summary format.

```
-------------------------------------------------------
 T E S T S
-------------------------------------------------------


Running org.esa.snap.wvcci.tcwv.TcwvFullAlgoForOneLandPixelTest
MERIS LAND TCWV = 7.169919573465825
OLCI LAND TCWV = 9.484049274929402
MODIS TERRA LAND TCWV = 6.513026760857524
Tests run: 3, Failures: 0, Errors: 0, Skipped: 2, Time elapsed:
23.147 sec


Running
org.esa.snap.wvcci.tcwv.TcwvFullAlgoForOneOceanPixelTest
MERIS OCEAN TCWV = 28.007106576980792
OLCI OCEAN TCWV = 33.396730957209482
MODIS TERRA OCEAN TCWV = 46.25291580393138
```

```
Tests run: 3, Failures: 0, Errors: 0, Skipped: 1, Time elapsed:
9.175 sec
```

3.1.2.3 Monitoring

The TCWV L2 processing is also executed on both JASMIN (MODIS Terra) and Calvalus (MERIS) processing systems using the PMonitor Python framework.

As an example, this is again illustrated for one month (31 days) of MERIS TCWV L2 processing on Calvalus. This task consists of 31 'per day' jobs, each of these jobs subsequently processes the single preprocessed products (see previous section) of one day.

Again, when the processing is initiated, a status file is generated which is monitoring the status of running, finished, and queued jobs. Here it looks like this:

```
Every 10.0s: less tcwv_meris.status                     Mon Apr
30 16:16:21 2019

31 created, 31 running, 0 backlog, 0 processed, 0 failed


Every 10.0s: less tcwv_meris.status                     Mon Apr
30 16:24:14 2019

31 created, 11 running, 0 backlog, 20 processed, 0 failed


Every 10.0s: less tcwv_meris.status                     Mon Apr
30 16:42:10 2019

31 created, 0 running, 0 backlog, 31 processed, 0 failed
```

Initially, 31 (of max. 32) jobs are running. Here we have no conditions for sequential processing (no jobs in backlog). After a while, a bunch of jobs (11), and finally all created jobs were processed successfully again.

As for the preprocessing, for each job a dedicated log file is written, and in case of successful processing an entry is generated in a report file. Examples for those are also given in Appendix 3: Processing logs and reports.

As a final check, we need to verifiy the number of generated TCWV products. For MERIS on Calvalus it is the same as the number of L1b input products as well as the number of preprocessed products:

```
>ls /calvalus/eodata/MER_RR__1P/r03/2011/08/*/*.N1 |wc -l
>444
```

```
>ls /calvalus/projects/wvcci/idepix-era-
interim/meris/2011/08/*/*.nc |wc -l
>444
```

```
>ls /calvalus/projects/wvcci/tcwv/meris/l2/2011/08/*/*.nc |wc -
l
>444
```

For MODIS Terra on JASMIN it is the same as the number of preprocessed 'Day' products:

```
>ls
/gws/nopw/j04/esacci_wv/odanne/WvcciRoot/IdepixEraInterim/MODIS
_TERRA/2011/08/01/*.nc |wc -l
>3438
```

```
>ls
/gws/nopw/j04/esacci_wv/odanne/WvcciRoot/Tcwv/MODIS_TERRA/2011/
08/01/*.nc |wc -l
>3438
```

In case the numbers are not as expected, the missing final products (e.g. due to corrupt inputs) are identified by appropriate comparisons of directory listings, and a reprocessing is done if possible.

3.1.2.4 Visual inspection

A number of TCWV L2 products from MERIS, OLCI and MODIS has been visually inspected. At verification stage the checks are:

- the products must contain all bands as specified in the PSD;
- the TCWV values should overall be in an expected range. For example, they should be roughly in line with the TCWV prior pattern (however, there can be significant differences on smaller scales, of course);
- the TCWV images should not contain significant artefacts (unless the specific reason is known);
- the TCWV images should be in line with the preprocessed products (i.e. cloudy pixels should have been removed from the retrieval).

Figure 3-5 shows the TCWV L2 retrieval for MERIS for the same product as in Figure 3-2. The right image shows TCWV retrieved for non-cloudy pixels. TCWV values over water are higher than over land in this case, as would be expected from the prior field shown in the centre image.

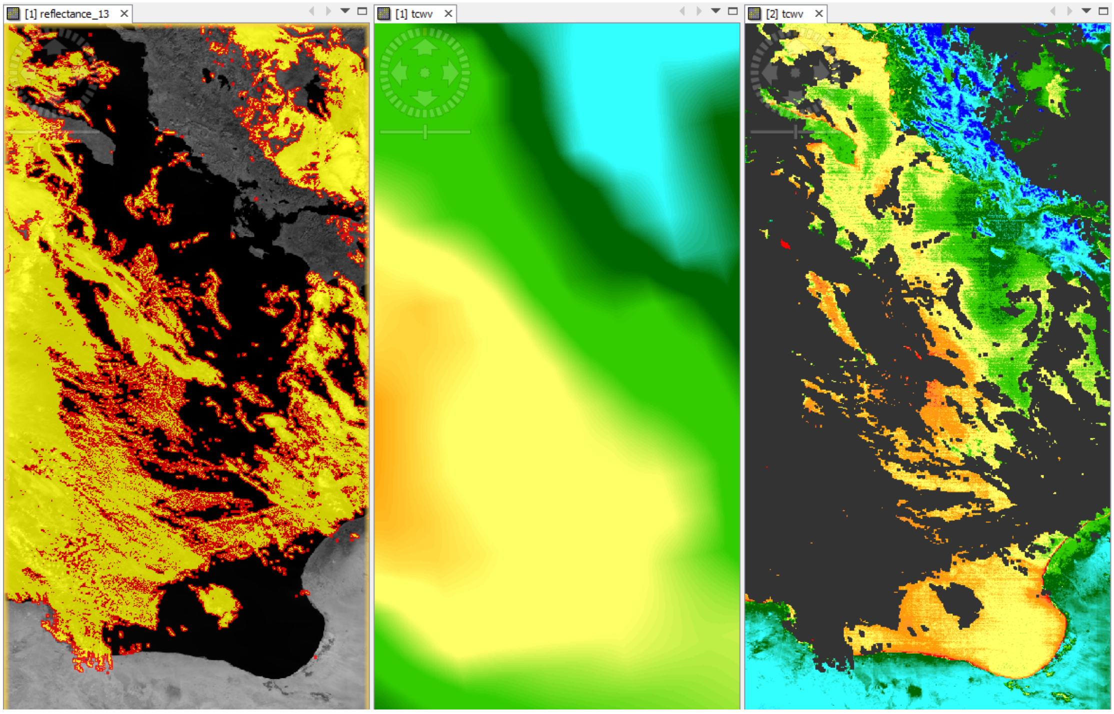


Figure 3-5: MERIS TCWV L2 retrieval for the same product as in Figure 3-2 (20081023_085102): left image shows pixels flagged as cloud (yellow) and cloud buffer (red), centre image shows again ERA Interim prior TCWV, right image TCWW retrieved from WV_cci algorithm (same scale, 0–40 kg m^{-2}).

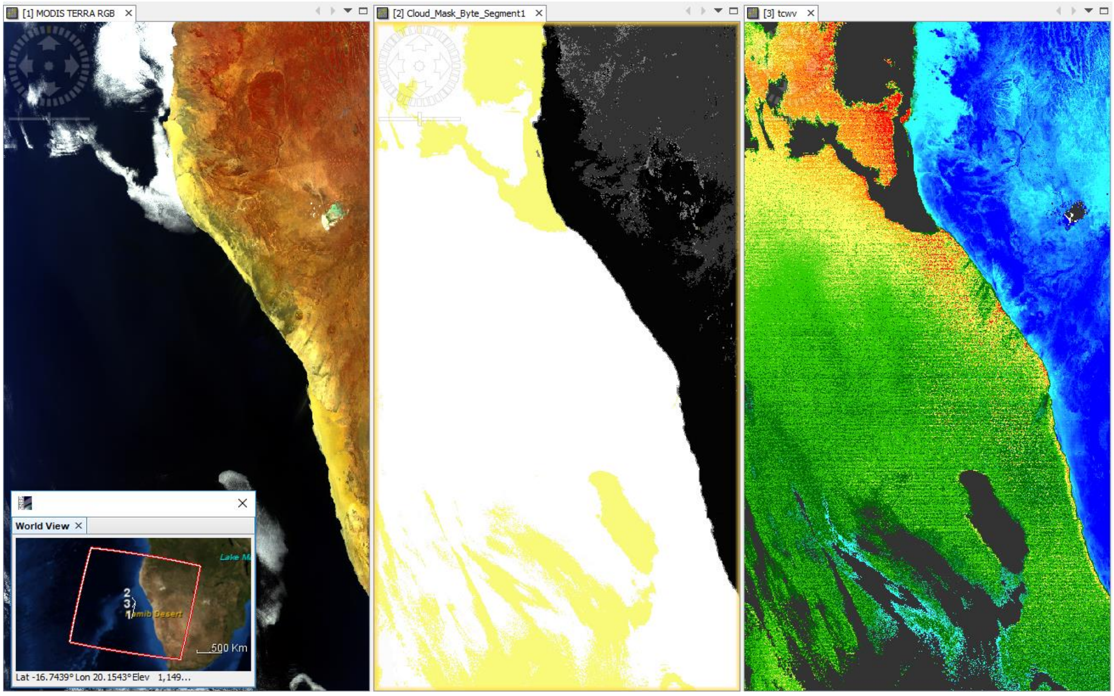


Figure 3-6: MODIS TCWV L2 retrieval for product MOD021KM.A2011196.0930: RBG (left) pixels flagged as cloud (yellow, centre), and TCWW retrieved from WV_cci algorithm (right, on a scale from 0–40 kg m^{-2}).

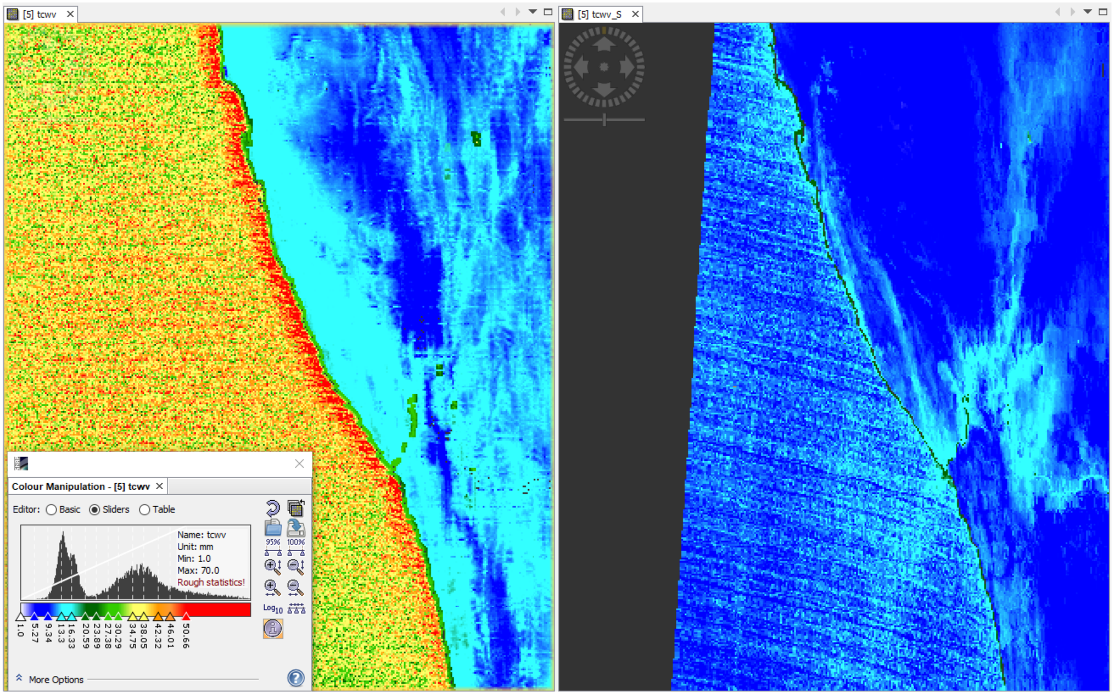


Figure 3-7: Comparison of MERIS and MODIS TCWV L2 retrieval: collocation of ‘Dataset 1’ products MER_RR__1PRACR20110715_075104 and MOD021KM.A2011196.0930) along the Namibian coast: MODIS left, MERIS right. TCWW values are on the same scale.

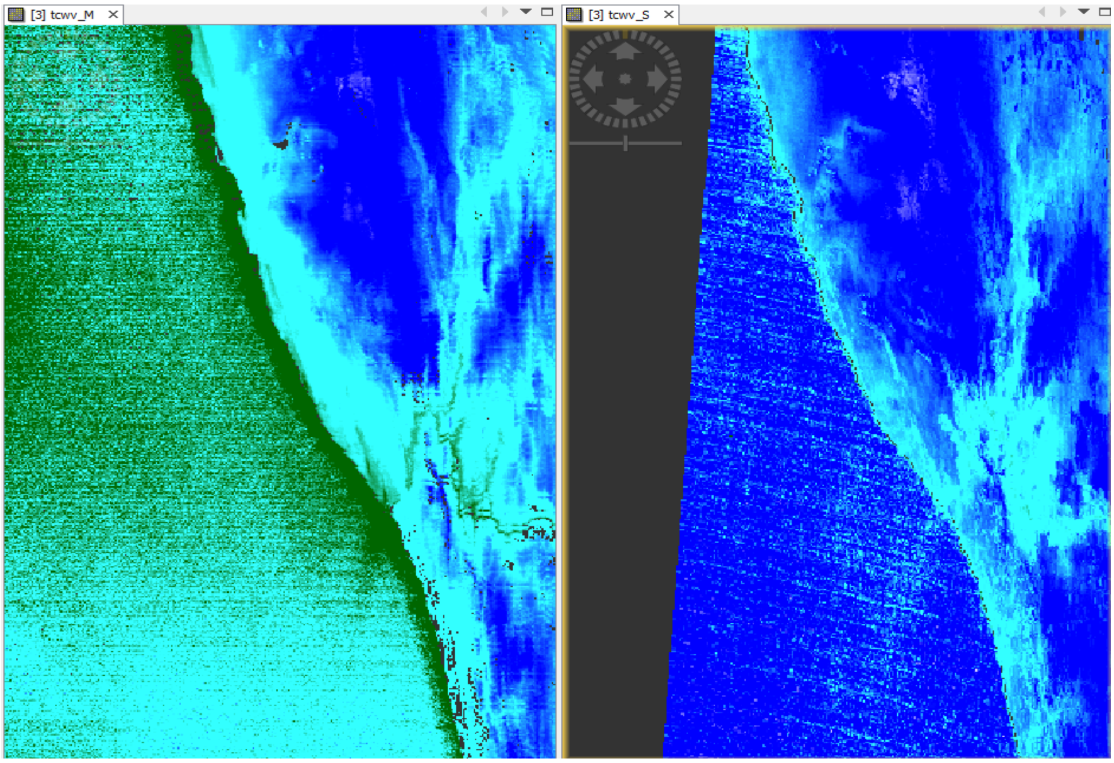


Figure 3-8: Comparison of MERIS and MODIS TCWV L2 retrieval: same as Figure 3-7, but obtained from recent 'Dataset 2' after algorithm improvements.

Figure 3-6 to Figure 3-8 illustrate the evolution and quality improvements of the TCWV retrievals during the WV_cci project Figure 3-6 shows the TCWV L2 retrieval from MODIS along Southwest Africa for product MOD021KM.A2011196.0930. The data were taken from the first test processing cycle ('Dataset 1'). The land part is totally cloud free in this case. The TCWV image shows a significant difference of relatively low values all over the land compared to higher values over water. A direct comparison with the MERIS retrieval for the same day over the same region is shown in Figure 3-7 (small subset region of product MOD021KM.A2011196.0930). There is obviously good agreement over land, whereas the MERIS retrieval does not show the increase over water at all. Moreover, the MODIS retrieval shows an even stronger increase in a narrow stripe on the water side of the coastline.

The TCWV differences in MERIS and MODIS retrievals observed in 'Dataset 1' were regarded as a critical issue and were subject to further investigation. For the next test processing cycle ('Dataset 2'), several fixes were applied in the algorithm, including the use of recompiled lookup tables for all sensors. The same result as in in Figure 3-7 but now obtained from 'Dataset 2' is shown in Figure 3-8. The alignment over water has been significantly improved, the TCWV values from MODIS are just slightly higher. The TCWV values over land are now nearly the same for MERIS and MODIS.

Figure 3-9 shows the TCWV L2 retrieval for the OLCI RR product S3A_OL_1_ERR____20160715T100617. The upper row shows RGB and IdePix cloud classification, the lower row shows TCWV prior (left) and retrieval from algorithm (right). The product was taken from 'Dataset 2'. It can be seen that the cloud detection works again quite well for most parts of the image. However, there are obviously some very thin, high clouds over parts of the Bay of Biscay and the Mediterranean Sea which are not detected and which lead to reduced values in TCWV retrieval. On the other hand, the order of magnitude of the retrieved values is in good agreement with the prior values.

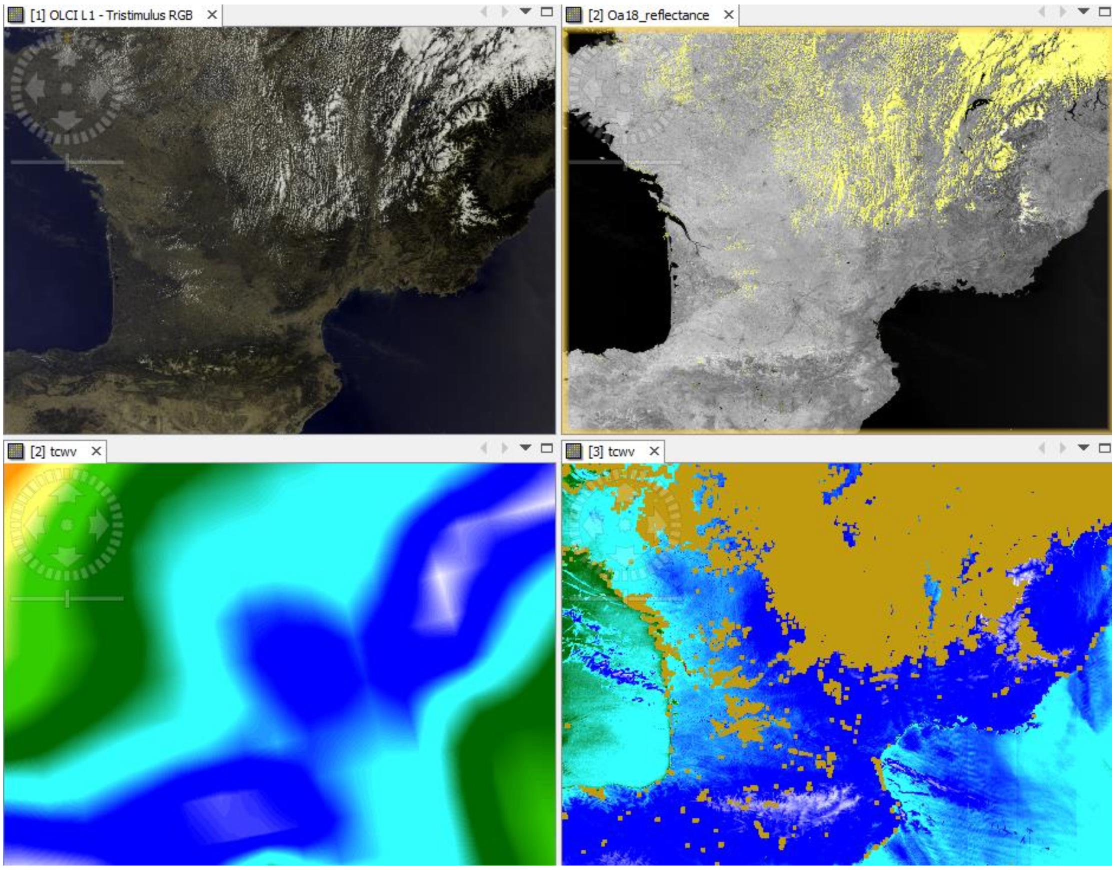


Figure 3-9: TCWV L2 retrieval for the OLCI RR product S3A_OL_1_ERR____20160715T100617. RGB (upper left), pixels flagged as cloud by IdePix (upper right), TCWV prior (lower left), and TCWV retrieval (lower right).

A more detailed quantitative analysis of the TCWV values from the different sensors is subject to the validation activities.

3.1.3 TCWV L2 file transfer JASMIN → Calvalus

An essential preparatory step for the TCWV L3 processing and the merge of products from the different sensors is the transfer of MODIS TCVW L2 products from the JASMIN cluster to Calvalus where the MERIS products are generated, the OLCI products will be generated, and the CM SAF HOAPS TCWV provided by DWD are ingested. This data transfer is done via rsync command, using the hpxfer high speed data transfer service provided by CEDA. As simple verification steps for a correct transfer, the number of files (e.g. per one month of MODIS TCWV L2) is counted on the outgoing and incoming sides, and the corresponding MD5 checksums are generated and compared.

3.1.4 TCWV L3 processing

For all sensors MERIS, OLCI and MODIS Terra, the TCWV L3 processing is executed on the Calvalus processing system using the PMonitor Python framework.

As an example, this is again illustrated for one month (31 days) of MERIS TCWV L3 processing on Calvalus. This task consists of 31 'per day' jobs, each of these jobs subsequently processes the single preprocessed products (see previous section) of one day.

3.1.4.1 Code sanity checks

A number of code sanity checks have been introduced into the TCWV L3 processing code. The most important ones are:

- the TCWV L3 basic processing (a daily spatial and/or a monthly temporal aggregation) is not initiated if the corresponding L2 input products are not completely available as expected and specified in the startup script;
- the TCWV L3 post processing (i.e. generating the final CF- and CCI-compliant NetCDF4 products) is not initiated if the corresponding basic processing did not terminate successfully;

- any of the L3 processing may fail due to a variety of conditions checked by the underlying Calvalus L3 core software. A detailed technical description of Calvalus is beyond the scope of this document, more information can be found in [7].

3.1.4.2 Unit-level testing

A large number of unit level tests are performed in the Calvalus L3 core software. Again, see [7] for more details. This software fully provides the means needed for our purposes, so there are no further specific L3 unit level tests for WV_cci.

3.1.4.3 Monitoring

The monitoring of the L3 processing on Calvalus follows the same principles as for the MERIS L2 processing described in Section 3.1.2.3.

The TCWV L3 processing also uses the PMonitor Python framework. As an example, this is illustrated for one month (31 days) of MODIS TCWV L3 with 0.5-degree target resolution, which consists of the spatial aggregation per day of all L2 products and their reprojection onto the 0.5-degree Plate Carrée global grid, and as second step the monthly temporal aggregation of the single days[3]. This task consists of 64 separate jobs:

- 31 spatial aggregations and reprojection for every day;
- 31 generations of daily final CF- and CCI compliant NetCDF products (this will be described in more detail in Section 3.1.5.1);
- 1 monthly temporal aggregation;
- 1 generation of a monthly final CF- and CCI compliant NetCDF product.

Again, when the processing is initiated, a status file is generated which is monitoring the status of running, finished, and queued jobs. Here it looks like this:

```
Every 10.0s: less tcwv_modis_terra_chain.status
Thu May  2 18:32:00 2019

64 created, 31 running, 33 backlog, 0 processed, 0 failed


Every 10.0s: less tcwv_modis_terra_chain.status
Thu May  2 18:39:46 2019

64 created, 25 running, 24 backlog, 15 processed, 0 failed
```

[3] *The monthly L3 products are not part of the initially agreed products to be delivered, but have been added as they were identified as a very useful add-on.*

```
Every 10.0s: less tcwv_modis_terra_chain.status
Thu May  2 19:23:24 2019

64 created, 0 running, 0 backlog, 64 processed, 0 failed
```

Initially, 31 (of max. 32) jobs are running. After a while, a bunch of jobs (15), and finally all created jobs were processed successfully. The conditions for sequential processing are:

- the jobs for final NetCDF4 generation have to wait until the corresponding aggregation job has successfully finished;
- the job for the monthly aggregation has to wait until all daily aggregations have successfully finished.

As for the L2 processing, for each job a dedicated log file is written, and in case of successful processing an entry is generated in a report file.

As final check, we need to verify again the number of generated TCWV L3 products. Here we simply must have 31 daily aggregated products, 31 corresponding final products, 1 monthly aggregated product and 1 corresponding final product:

```
>ls /calvalus/projects/wvcci/tcwv/modis_terra/l3-daily/05/2011/07/*/*.nc |wc -l
>31

>ls /calvalus/projects/wvcci/tcwv/modis_terra/l3-daily-final-nc/05/2011/07/*/*.nc |wc -l
>31


>ls /calvalus/projects/wvcci/tcwv/modis_terra/l3-monthly/05/2011/07/*.nc |wc -l
>1


>ls /calvalus/projects/wvcci/tcwv/modis_terra/l3-monthly-final-nc/05/2011/07/*.nc |wc -l
>1
```

In case the numbers are not as expected, the missing final products (e.g. due to corrupt inputs) are identified by appropriate comparisons of directory listings, and a reprocessing is done if possible.

In case of CM SAF HOAPS TCWV data, the daily aggregation is not required, as these products were delivered by DWD already as daily L3 products on both 0.5-degree and 0.05-degree Plate Carrée global grid. Also, a monthly aggregation of

HOAPS TCWV products is not foreseen. Therefore these products can be taken 'as they are' and directly be used for the sensor merge (see Section 3.1.4.5).

3.1.4.4 Visual inspection

A number of final TCWV L3 products from MERIS, OLCI and MODIS have been visually inspected. At verification stage the checks are:

- the products must contain all bands as specified in the PSD;
- the TCWV values should overall be in an expected range;
- there should be TCWV values only over land, coastal regions over ocean, and sea ice (if identified)
- 'num_obs' (the number of TCWV L2 retrievals contributing to L3 grid cell) should have reasonable values. For example, highest values are expected over cloud-free deserts, values should overall be higher in monthly than in daily products
- the TCWV images should be given on a 0.5-degree or 0.05-degree Plate Carrée global grid;
- the TCWV images should not contain significant artefacts (unless the specific reason is known).

Figure 3-10 to Figure 3-14 show MERIS July 2011 examples for daily and monthly aggregates for both TCWV and TCWV counts for 0.05-degree resolution. The products were taken from the second test processing cycle ('Dataset 2'). Overall we observe what is expected. We have most TCWV L2 retrievals over the deserts. As expected in July, we have samples in the Arctic, but no samples in the Antarctic. There are also no samples over the oceans and also no samples over sea ice, as for the MERIS data period (2002–2012) there was no sea ice information available yet from the CM SAF L3 mask (see PUG [9]). The number of observations is much higher in the monthly than in the daily product. Also, we have no data regions within the swaths due to the exclusion of cloudy pixels. Overall, the global data coverage for a single day is rather poor from MERIS only. In the monthly aggregate, this coverage is much better, however, there are data gaps remaining as we have some regions with continuous cloud cover (e.g. in tropical convergence zone). This is also true for Greenland, as probably many bright pixels over clear ice are misinterpreted as clouds. The data quality is overall good, as indicated by the TCWV quality flag (Figure 3-12), which shows only a few greid cells with an enhanced cost function value.

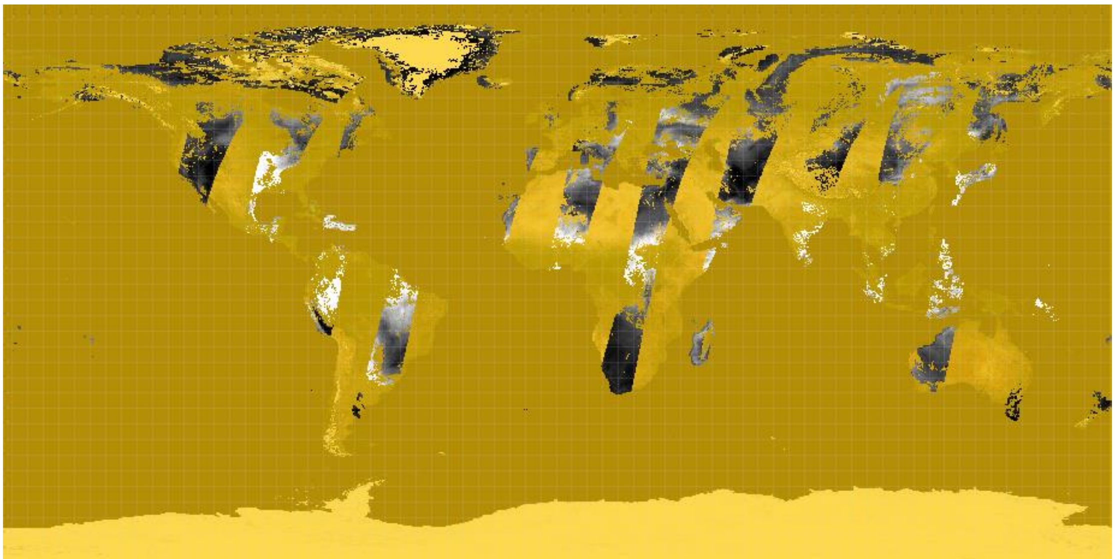


Figure 3-10: MERIS TCWV L3 daily aggregate for July 15th, 2011 (greyscale, 0–70 kg m^{-2}). Yellow indicates no data.

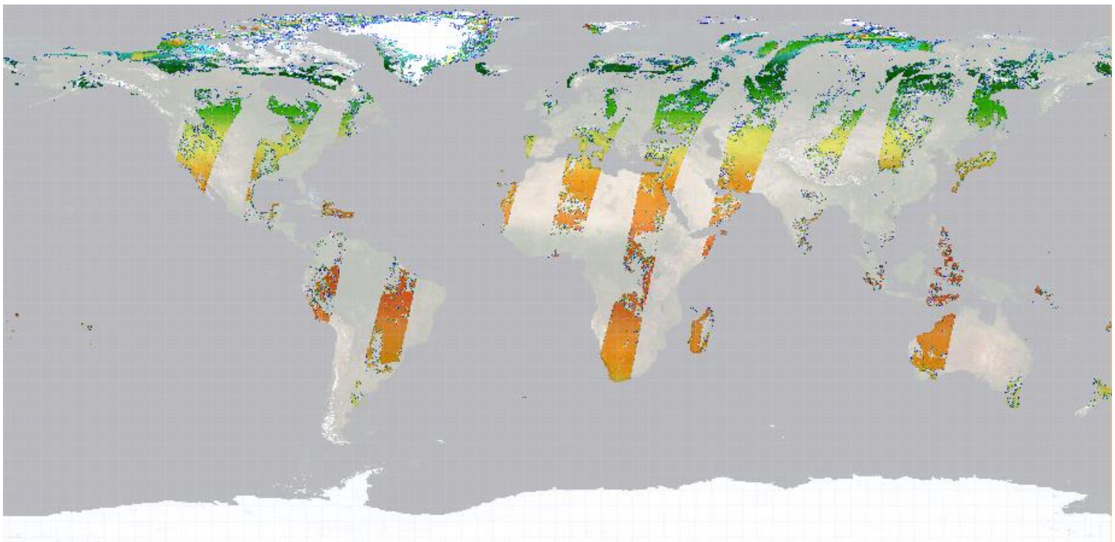


Figure 3-11: Number of MERIS TCWV L2 retrievals in L3 daily aggregate for July 15th, 2011. Colour scale is 0–250.

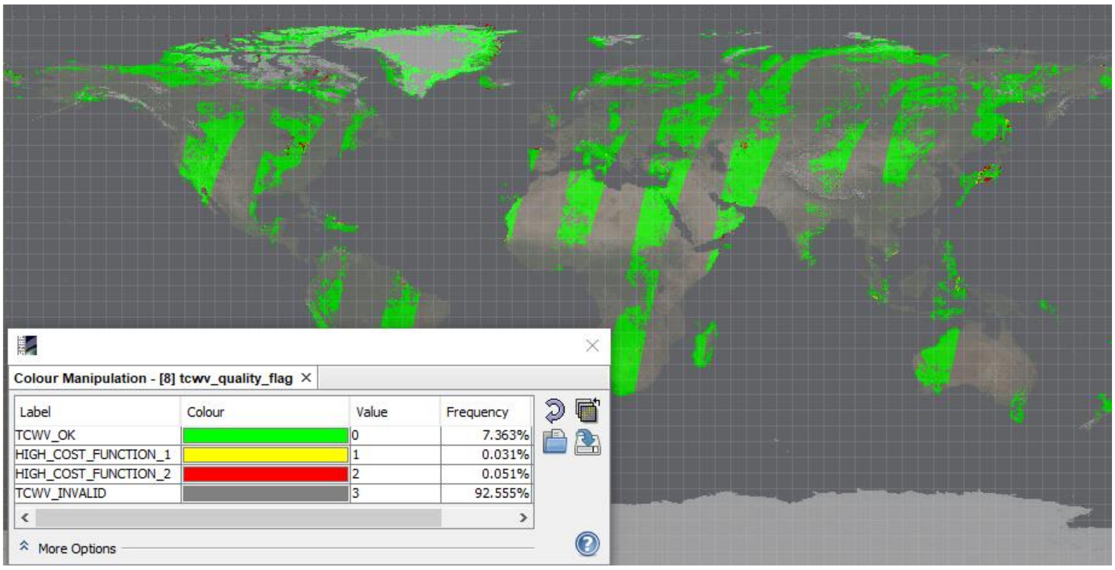


Figure 3-12: MERIS TCWV quality flag in L3 daily aggregate for July 15th, 2011. Colours are explained in the table.

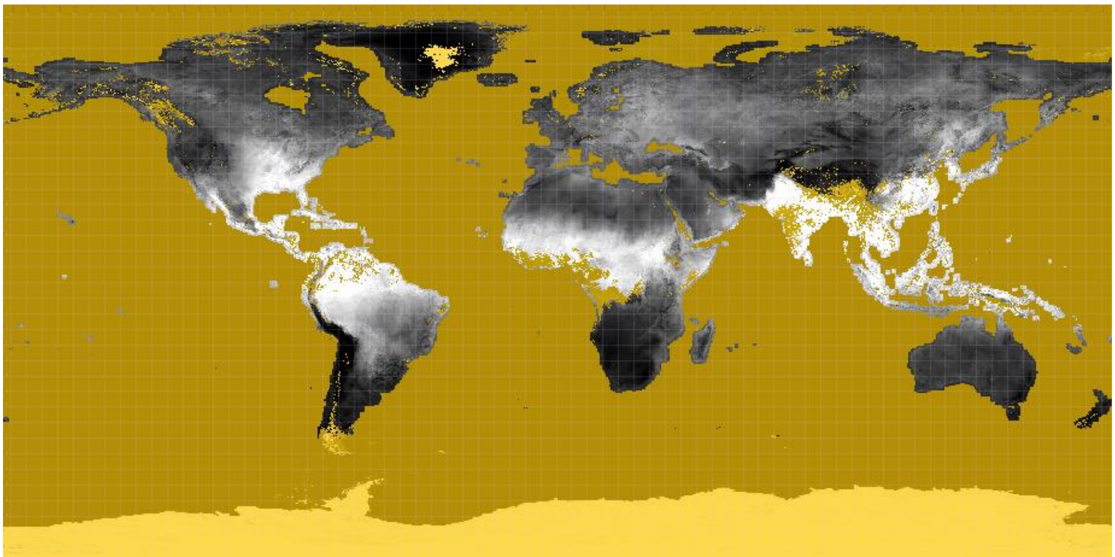


Figure 3-13: MERIS TCWV L3 monthly aggregate for July 2011 (greyscale, 0–70 kg m^{-2}).

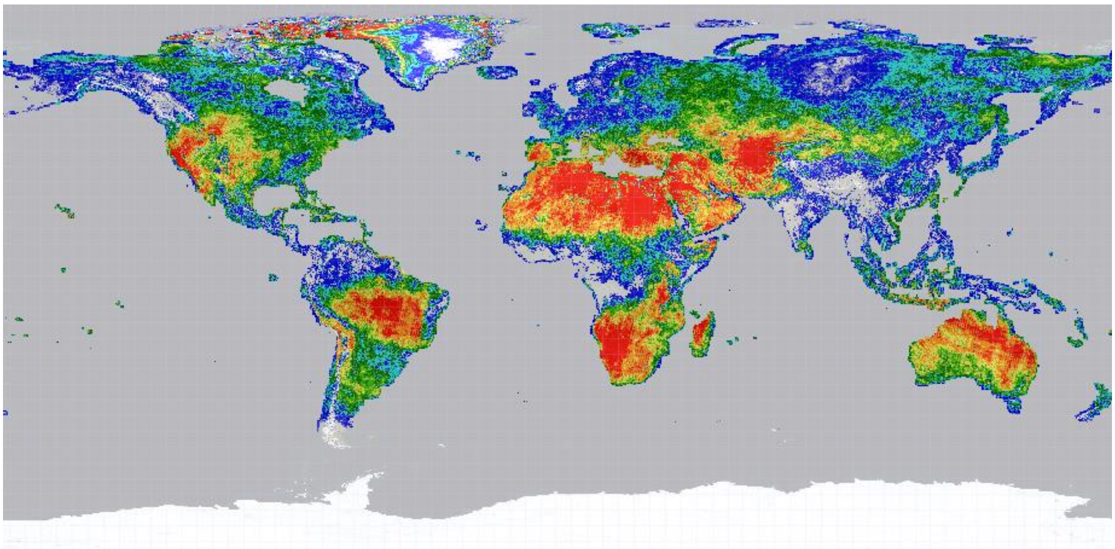


Figure 3-14: Number of MERIS TCWV L2 retrievals in L3 monthly aggregate for July 2011. Colour scale is 0–2500.

Figure 3-15 to Figure 3-19 show the same as Figure 3-10 to Figure 3-14, but for MODIS. The coverage for a single day as well as for the month is better than for MERIS. Although the majority of TCWV values have a good quality, we have more grid cells with an enhanced cost function here (Figure 3-17). Again, we have remaining gaps of data even in the monthly product due to permanent cloud coverage. In contrast to MERIS, we have now sea ice pixels included (information available for 2016), as well as a high number of pixels over Greenland. Obviously, the MODIS cloud mask does not misinterpret ice as clouds here, but the other way round. The distinction of clouds and snow/ice is a difficult task and a well known issue in most cloud mask schemes. In the number of L2 retrievals in the daily product (Figure 3-16), we can see that the numbers decrease from the centre of the swaths to the edges, which is caused by the bow-tie effect from the MODIS scan geometry [10].

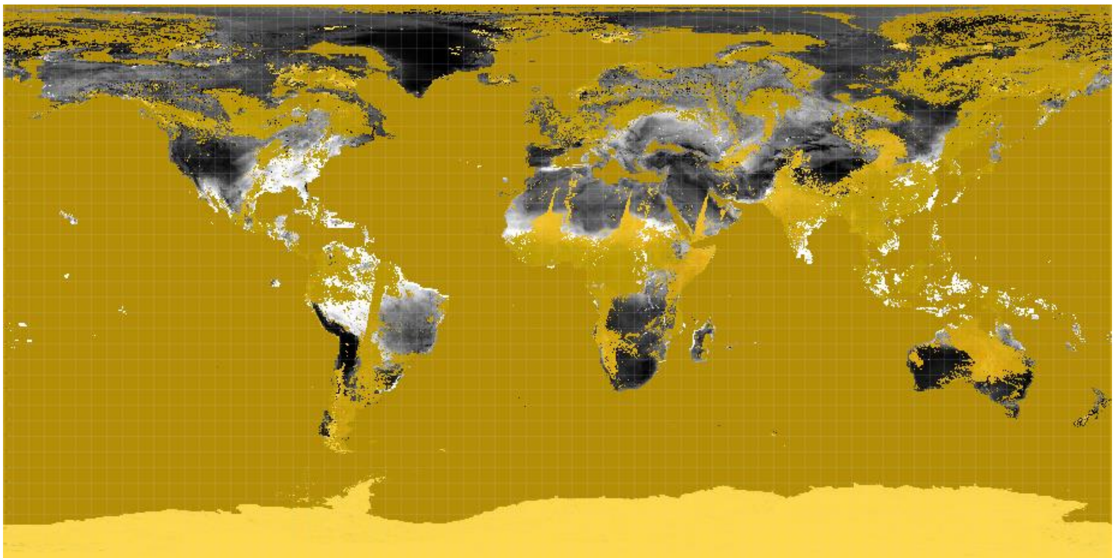


Figure 3-15: MODIS TCWV L3 daily aggregate for July 13th, 2016 (greyscale, 0–70 kg m^{-2}).

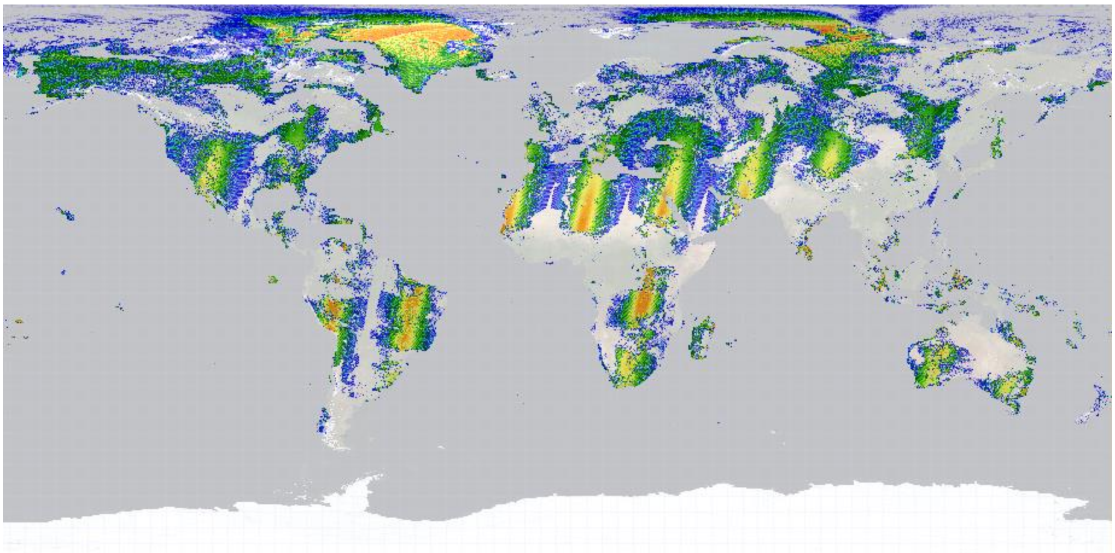


Figure 3-16: Number of MODIS TCWV L2 retrievals in L3 daily aggregate for July 13th, 2016. Colour scale is 0–250.

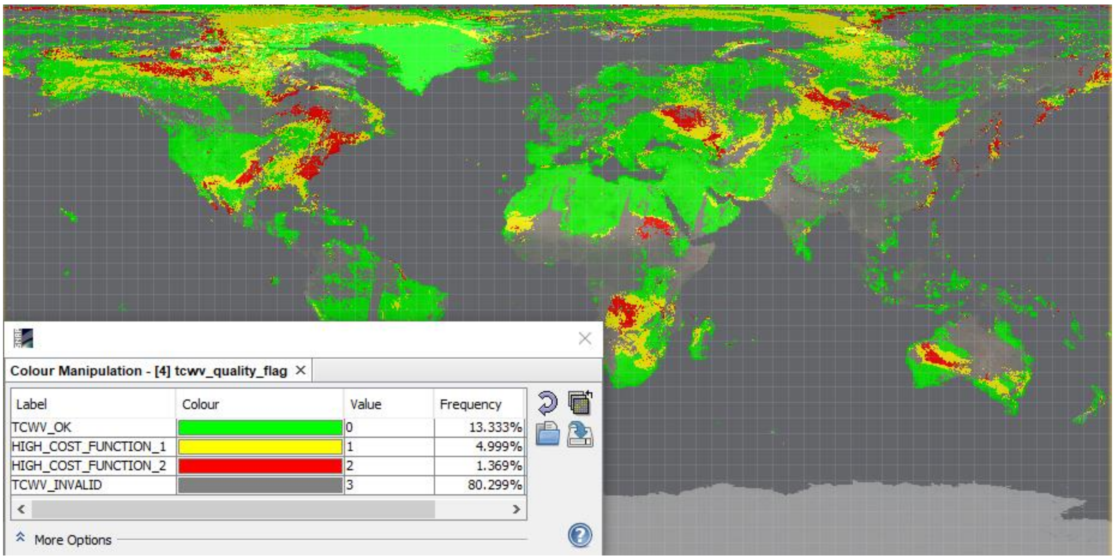


Figure 3-17: MODIS TCWV quality flag in L3 daily aggregate for July 13th, 2016. Colours are explained in the table.

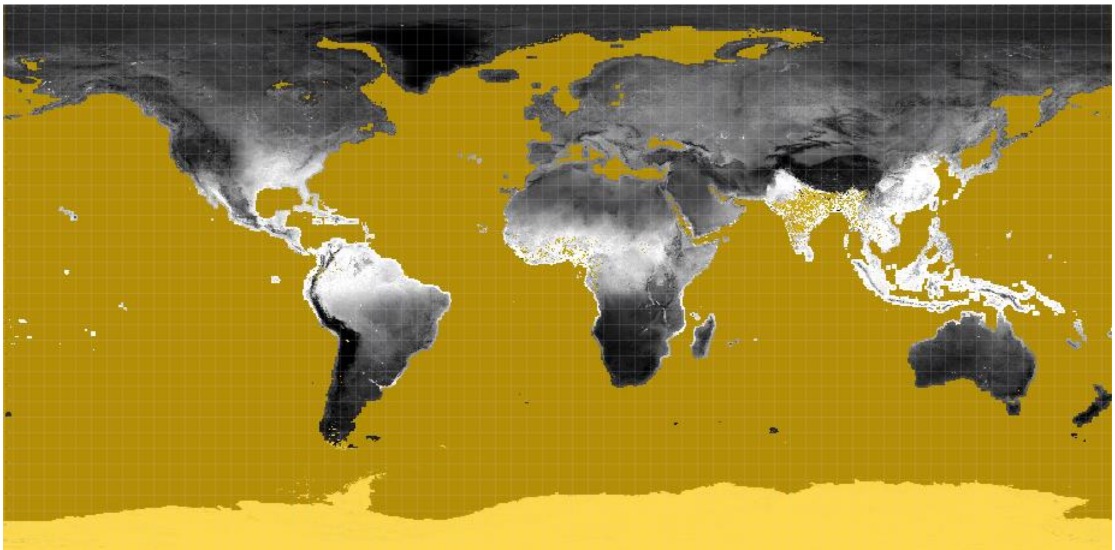


Figure 3-18: MODIS TCWV L3 monthly aggregate for July 2016 (greyscale, 0–70 kg m^{-2}).

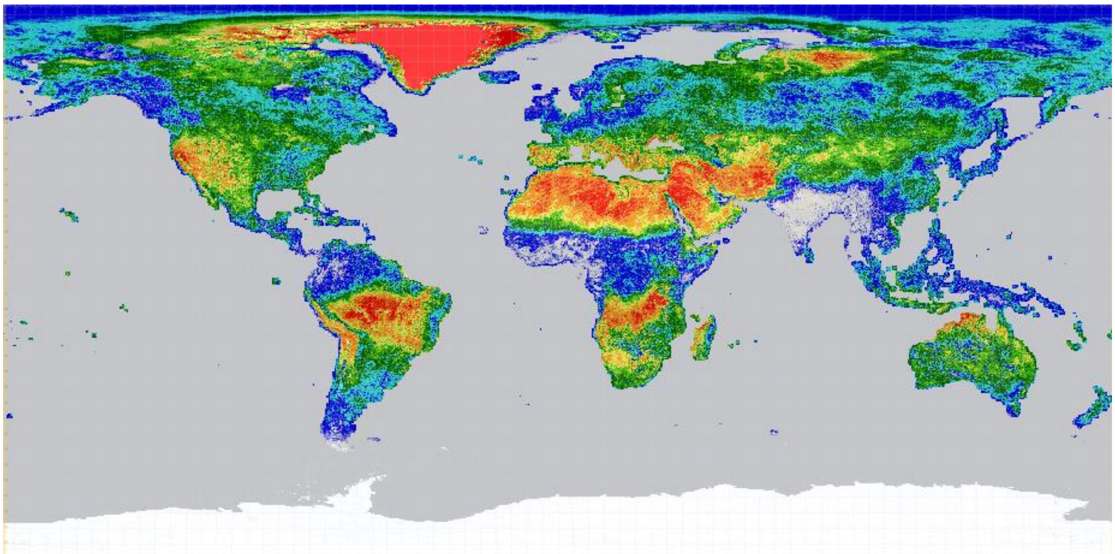


Figure 3-19: Number of MODIS TCWV L2 retrievals in L3 monthly aggregate for July 2016. Colour scale is 0–5000.

Figure 3-20 to Figure 3-24 show the same as the previous figures for MERIS and MODIS, but now for OLCI. The coverage for a single day as well as for the month is similar to MERIS. Compared with MERIS, we have many more TCWV samples over Greenland for OLCI. The reason is that the distinction of clouds and snow/ice in the IdePix classification could be significantly improved within the activities in the ESA SICE and S3-Snow projects [11]. As for MODIS, sea ice is included for OLCI. However, the TCWV values over sea ice in very high latitudes are obviously higher than the ones from MODIS. They correspond with enhanced cost function values (Figure 3-22), thus, these retrievals may be doubtful. The reason for this is not yet clear. Otherwise, the data quality is overall good. Again, we have overall remaining gaps of data even in the monthly product due to permanent cloud coverage.

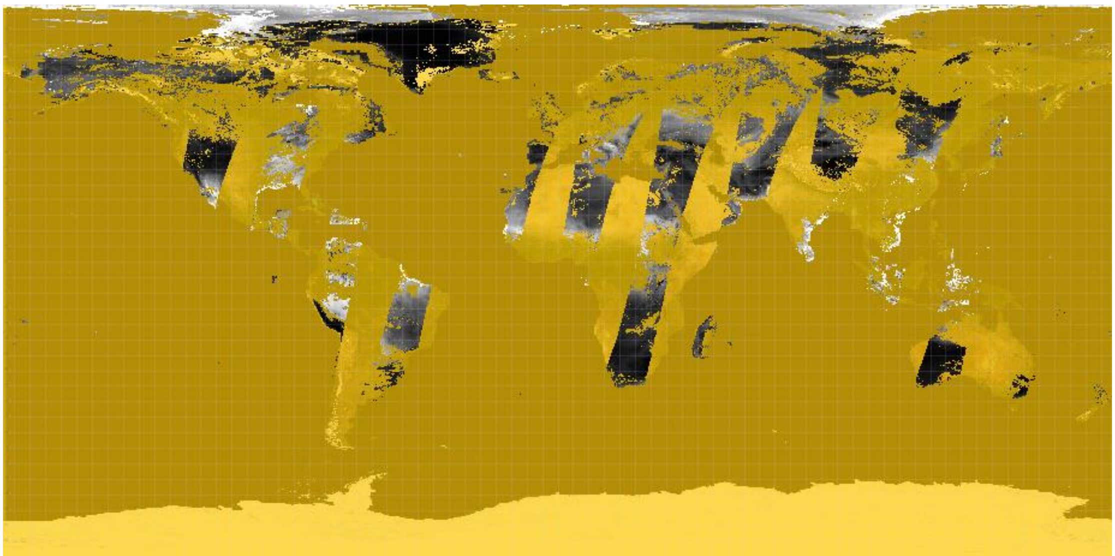


Figure 3-20: OLCI TCWV L3 daily aggregate for July 13th, 2016 (greyscale, 0–70 kg m^{-2}). Yellow indicates no data.

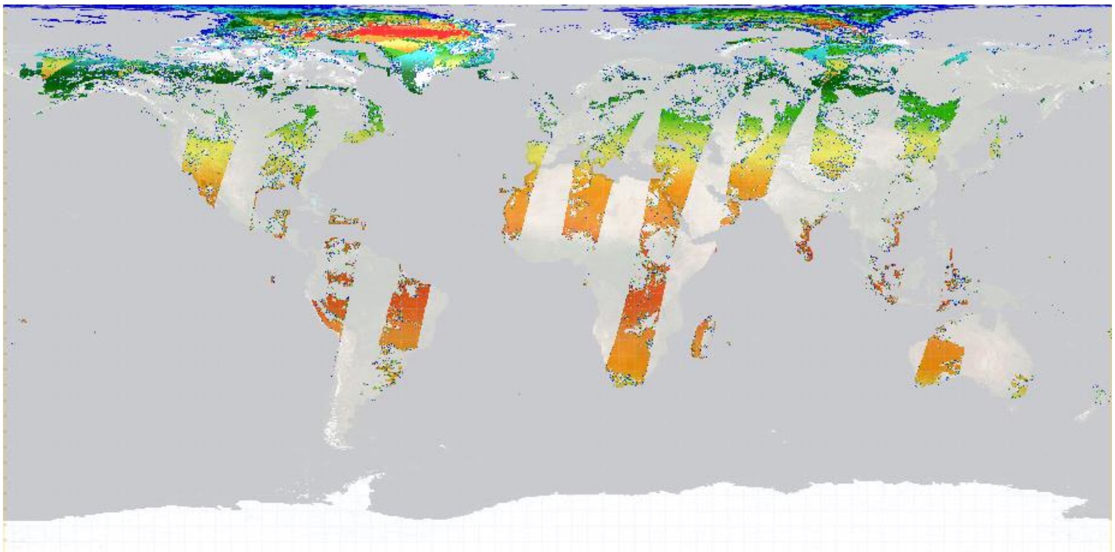


Figure 3-21: Number of OLCI TCWV L2 retrievals in L3 daily aggregate for July 13th, 2016. Colour scale is 0–250.

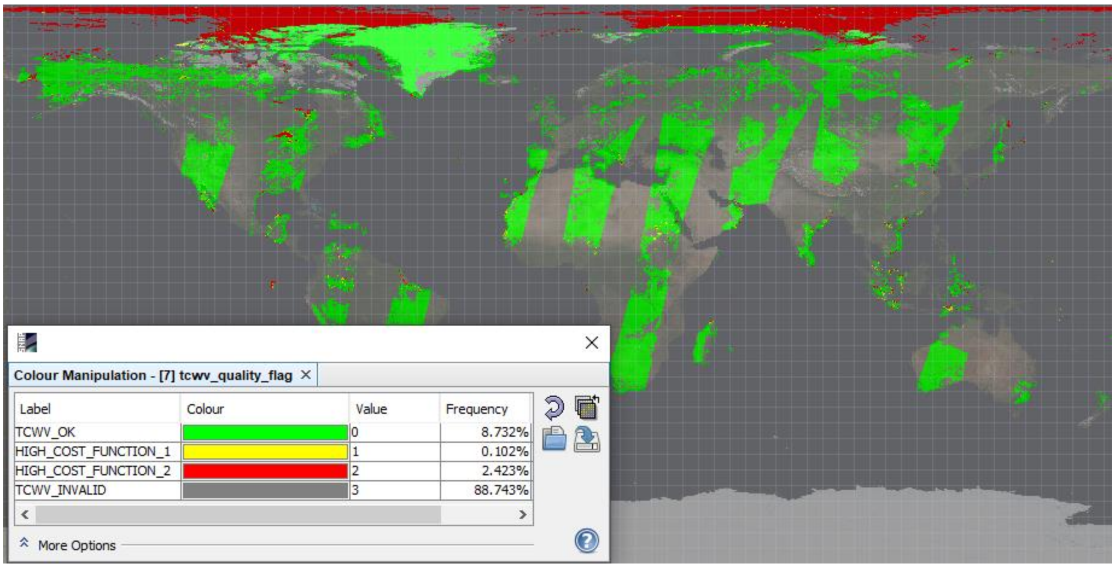


Figure 3-22: OLCI TCWV quality flag in L3 daily aggregate for July 13th, 2016. Colours are explained in the table.

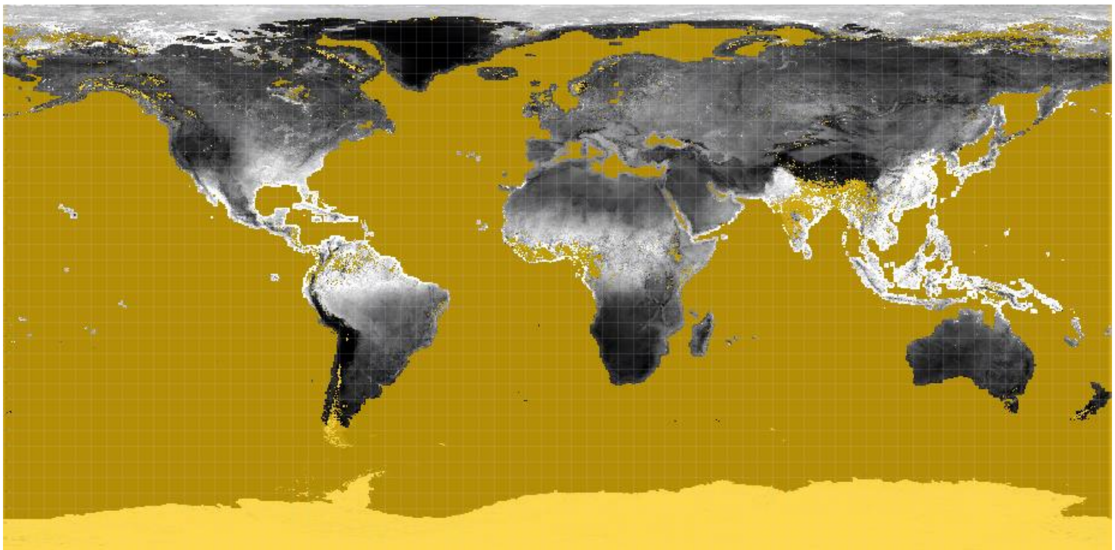


Figure 3-23: OLCI TCWV L3 monthly aggregate for July 2016 (greyscale, 0–70 kg m^{-2}).

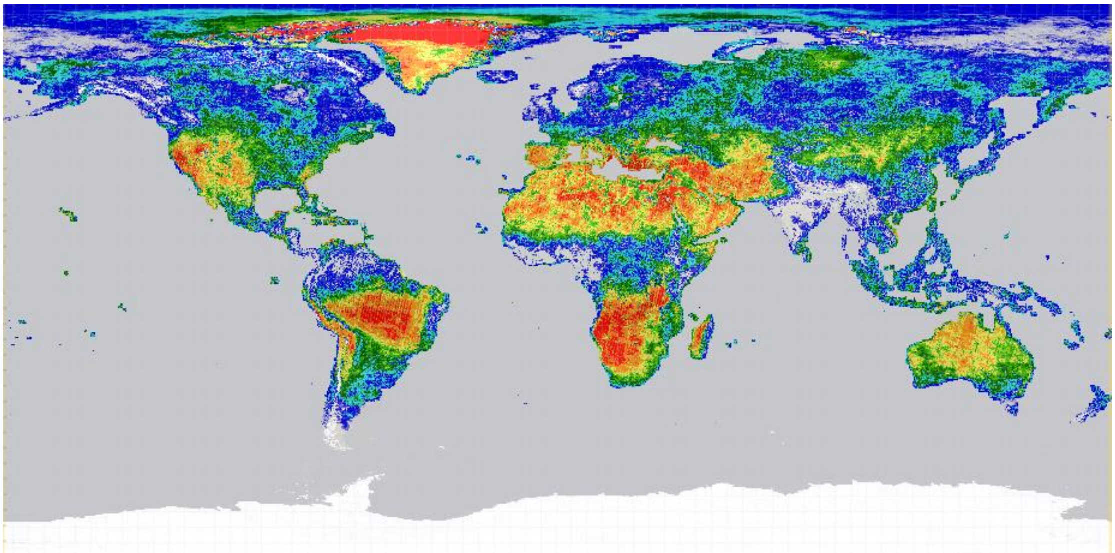


Figure 3-24: Number of OLCI TCWV L2 retrievals in L3 monthly aggregate for July 2016. Colour scale is 0–2500.

Figure 3-25 shows the TCWV from the HOAPS TCWV L3 product for July 15th, 2011. As specified, TCWV is provided over ocean only, but clouds are not excluded here. An introduction and more details on the retrieval and generation of these products can be found in [8].

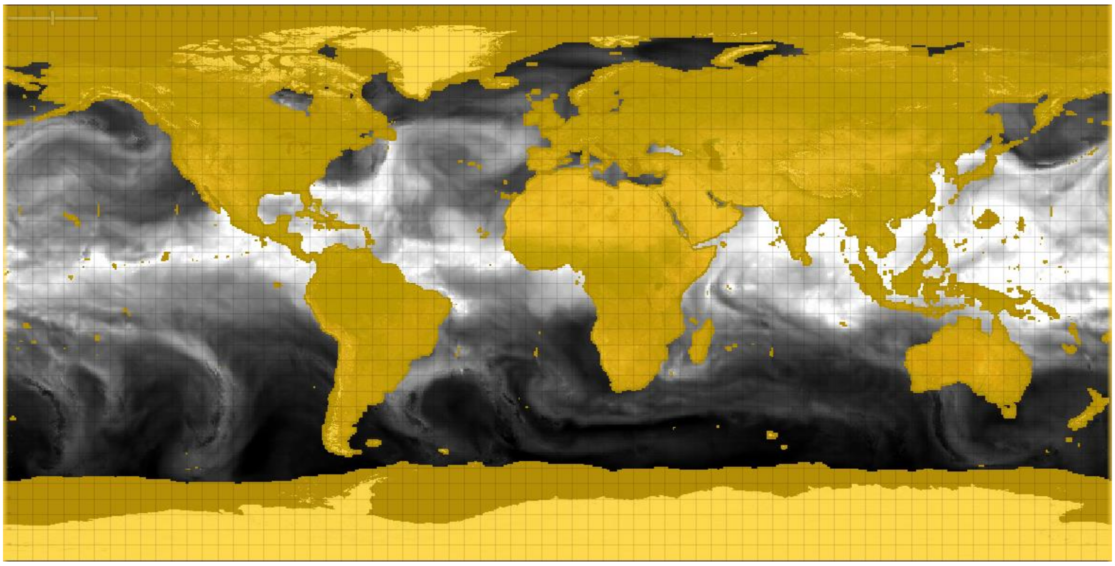


Figure 3-25: HOAPS TCWV L3 daily product for July 15th, 2011 (greyscale, 0–70 kg m^{-2}).

3.1.4.5 Product Merging

One of the main goals of the TCWV retrieval part within WV_cci is the merging of the TCWV data of the available sensors in order to fill all data gaps in global products as far as possible, and to identify regional, temporal, or systematic differences and discontinuities when bringing the data together. This should ideally result in an improvement of the underlying algorithms algorithms and the elimination of existing problems.

3.1.4.6 Code sanity checks

A number of code sanity checks have been introduced into the TCWV L3 product merging processing code. TCWV L3 product merging fails if:

- the number of input products is not equal to 2;
- one of the input products does not fulfill the expected specifications;
- the input products are not given on the same grid or have a different size;
- the time range (i.e. the reference day) of the input products is different.

3.1.4.7 Unit-level testing

The merging processing module also contains a couple of unit-level tests, which check in particular the correct application of the merging rules. The report is again given below in a summary format.

```
[INFO] -------------------------------------------------------
[INFO]  T E S T S
[INFO] -------------------------------------------------------
[INFO] Running org.esa.snap.wvcci.tcwv.l3.L3MergeSensorsOpTest
[INFO] Tests run: 3, Failures: 0, Errors: 0, Skipped: 0, Time
elapsed: 0.159 s
[INFO] Results:
[INFO] Tests run: 3, Failures: 0, Errors: 0, Skipped: 0
```

3.1.4.8 Monitoring

The monitoring of the TCWV L3 product merging on Calvalus follows the same principles as for the MERIS/OLCI L2 and the L3 processing described earlier.

The merging is performed for the sensor combinations:

- MERIS + MODIS
- OLCI + MODIS
- MERIS + CM SAF HOAPS
- MODIS + CM SAF HOAPS
- (MERIS + MODIS) + CM SAF HOAPS
- OLCI + CM SAF HOAPS
- MODIS + CM SAF HOAPS
- (OLCI + MODIS) + CM SAF HOAPS

The TCWV L3 product merging also uses the PMonitor Python framework. As an example, this is again illustrated for one month (31 days) of MERIS and MODIS TCWV L3 with 0.5-degree target resolution.. This task again consists of 64 separate jobs:

- 31 merges per day of all MERIS with all MODIS L3 daily products;
- 31 generations of daily final CF- and CCI compliant NetCDF products;
- 1 monthly aggregation of the daily merged products;
- 1 generation of a monthly final CF- and CCI compliant NetCDF product.

Again, when the processing is initiated, a status file is generated which is monitoring the status of running, finished, and queued jobs. Here it looks like this:

```
Every 10.0s: less tcwv_merge.status
Fri May  3 14:12:03 2019

64 created, 31 running, 33 backlog, 0 processed, 0 failed


Every 10.0s: less tcwv_merge.status
Fri May  3 14:16:24 2019

64 created, 21 running, 20 backlog, 23 processed, 0 failed


Every 10.0s: less tcwv_merge.status
Fri May  3 14:22:51 2019

64 created, 0 running, 0 backlog, 64 processed, 0 failed
```

Initially, 31 (of max. 32) jobs are running. After a while, a bunch of jobs (23), and finally all created jobs were processed successfully The conditions for sequential processing are:

- the jobs for final NetCDF4 generation have to wait until the corresponding merge job has successfully finished;
- the job for the monthly aggregation has to wait until all daily merge jobs have successfully finished.

As for the previously described processing steps, for each job a dedicated log file is written, and in case of successful processing an entry is generated in a report file.

As a final check, we need to verify again the number of generated merged TCWV L3 products. Here we simply must have 31 daily merged products, 31 corresponding final products, 1 monthly aggregated product and 1 corresponding final product:

```
>ls /calvalus/projects/wvcci/tcwv/meris-modis_terra/l3-
daily/05/2011/07/*/*.nc |wc -l
>31
>ls /calvalus/projects/wvcci/tcwv/meris-modis_terra/l3-daily-
final-nc/05/2011/07/*/*.nc |wc -l
>31


>ls /calvalus/projects/wvcci/tcwv/meris-modis_terra/l3-
monthly/05/2011/07/*.nc |wc -l
>1


>ls /calvalus/projects/wvcci/tcwv/meris-modis_terra/l3-monthly-
final-nc/05/2011/07/*.nc |wc -l
>1
```

In case the numbers are not as expected, the missing final products are identified by appropriate comparisons of directory listings, and a reprocessing must be initiated.

3.1.4.9 Visual inspection

The TCWV L3 merged products should basically look exactly like the non-merged L3 products, now with the TCWV data complementing each other. This is illustrated in Figure 3-26 to Figure 3-29, showing the merge of MERIS and MODIS TCWV L3 daily and monthly products for July 2011.

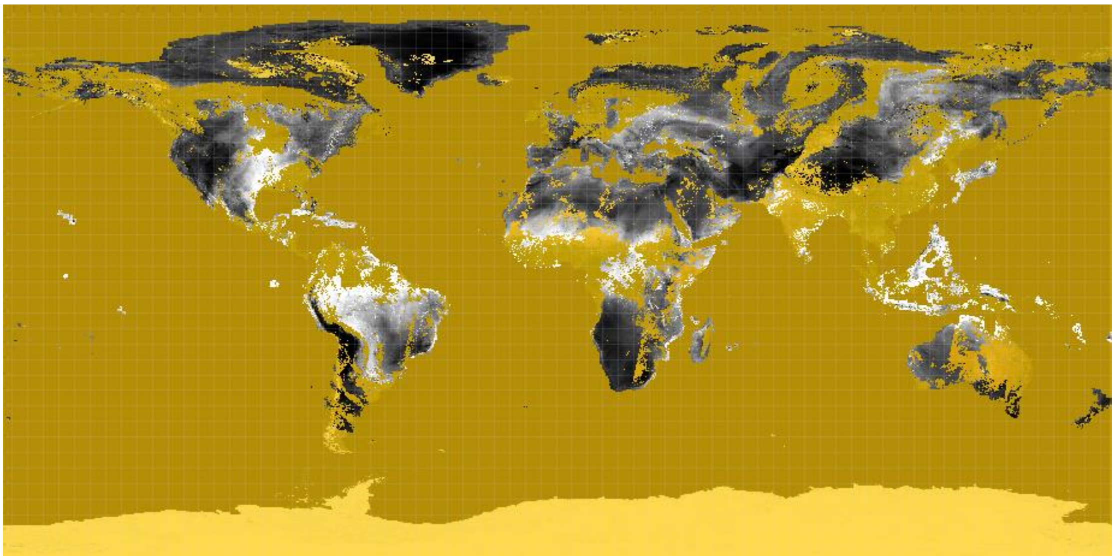


Figure 3-26: MERIS/MODIS TCWV L3 daily merge for July 15th, 2011 (greyscale, 0–70 kg m^{-2}). Yellow indicates no data.

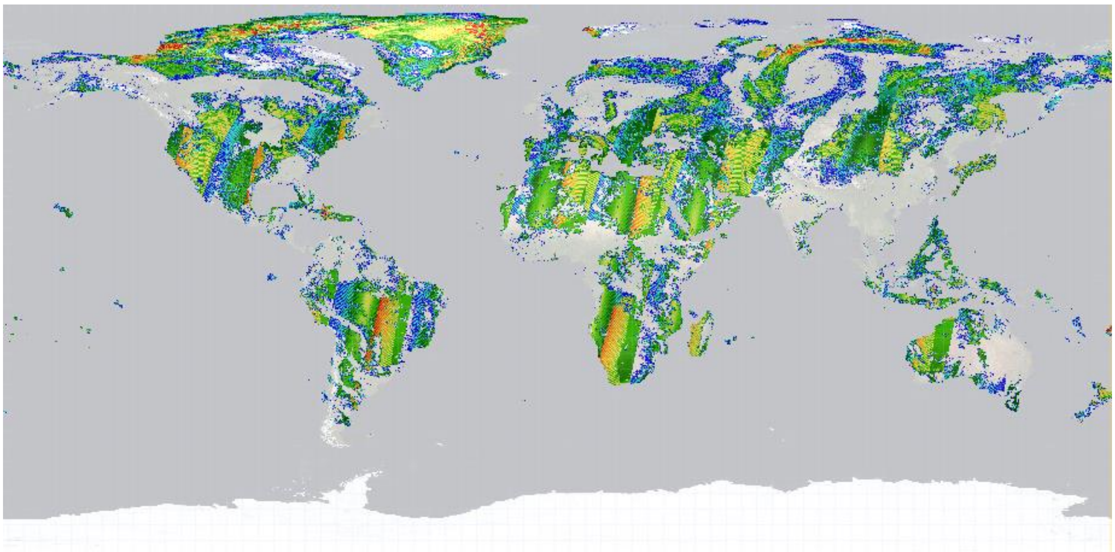


Figure 3-27: Number of MERIS+MODIS TCWV L2 retrievals in L3 daily aggregate for July 15th, 2011. Colour scale is 0–500.

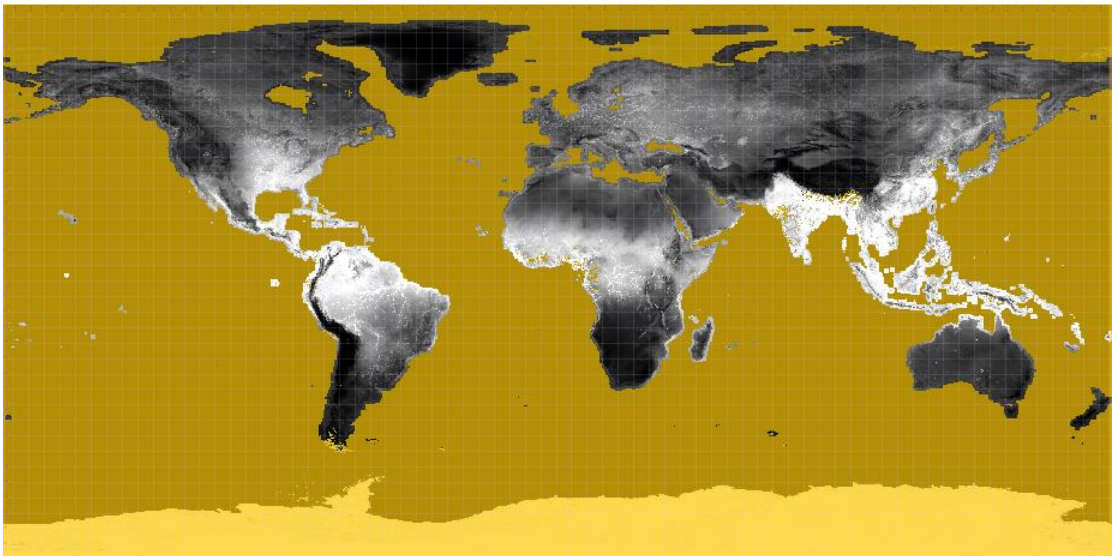


Figure 3-28: MERIS/MODIS TCWV L3 monthly merge for July 2011 (greyscale, 0–70 kg m^{-2}). Yellow indicates no data.

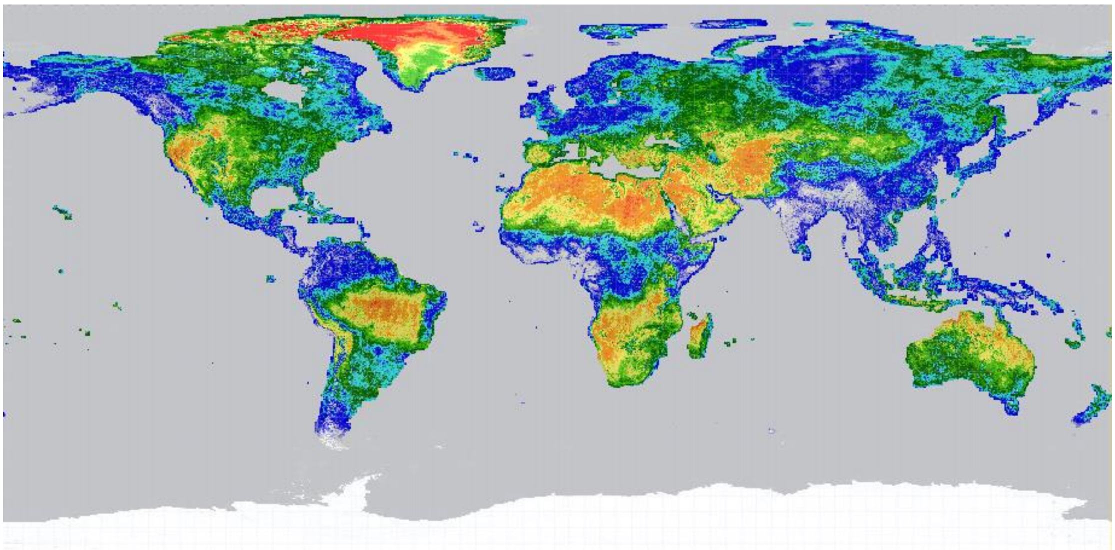


Figure 3-29: Number of MERIS+MODIS TCWV L2 retrievals in L3 daily aggregate for July 2011. Colour scale is 0–7500.

Figure 3-30 to Figure 3-33 shows the merge of the MERIS/MODIS merges with CM SAF HOAPS, so we have the combination of all three sensors. As we would expect, the TCWV pattern over water is dominated by the HOAPS data, as MERIS/MODIS retrievals are used here only for gap filling, i.e. near the coasts (Figure 3-30, Figure 3-32). The number of TCWV retrievals are in line with this (Figure 3-31, Figure 3-33).

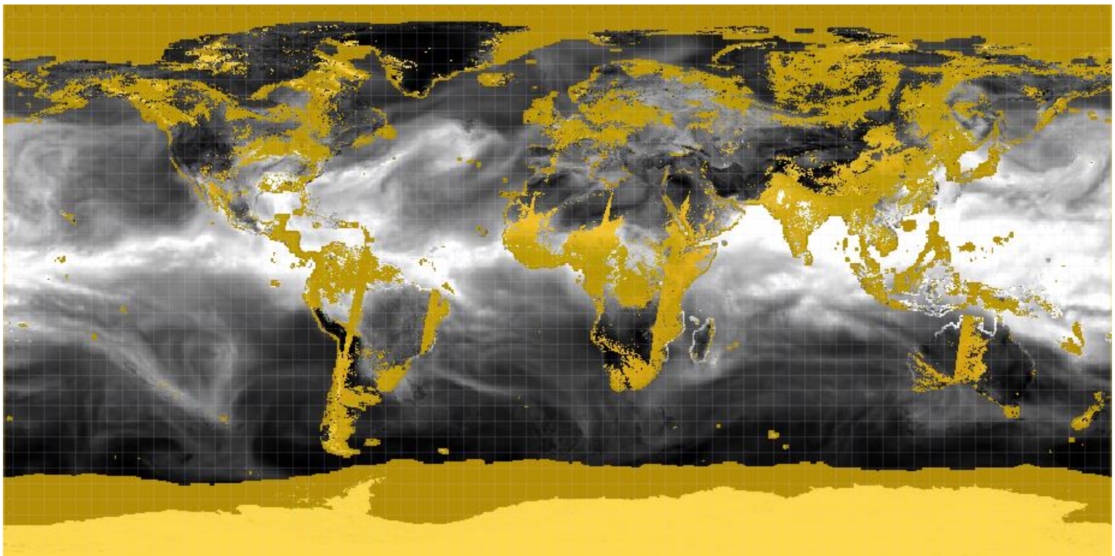


Figure 3-30: MERIS/MODIS/HOAPS TCWV L3 daily merge for July 28th, 2011 (greyscale, 0–70 kg m^{-2}). Yellow indicates no data.

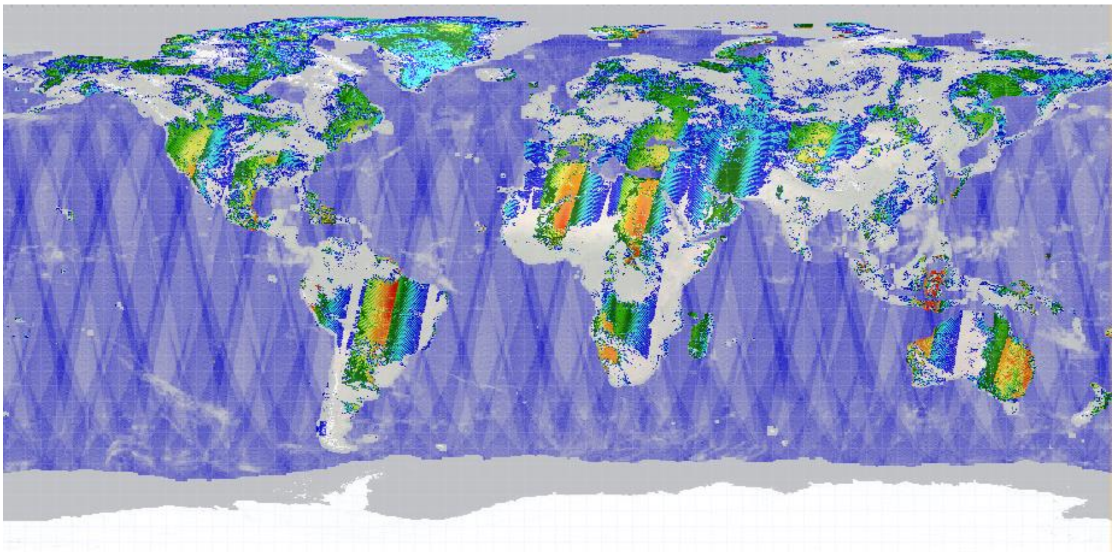


Figure 3-31: Number of MERIS+MODIS+HOAPS TCWV L2 retrievals in L3 daily aggregate for July 28th, 2011. Colour scale is 0–500.

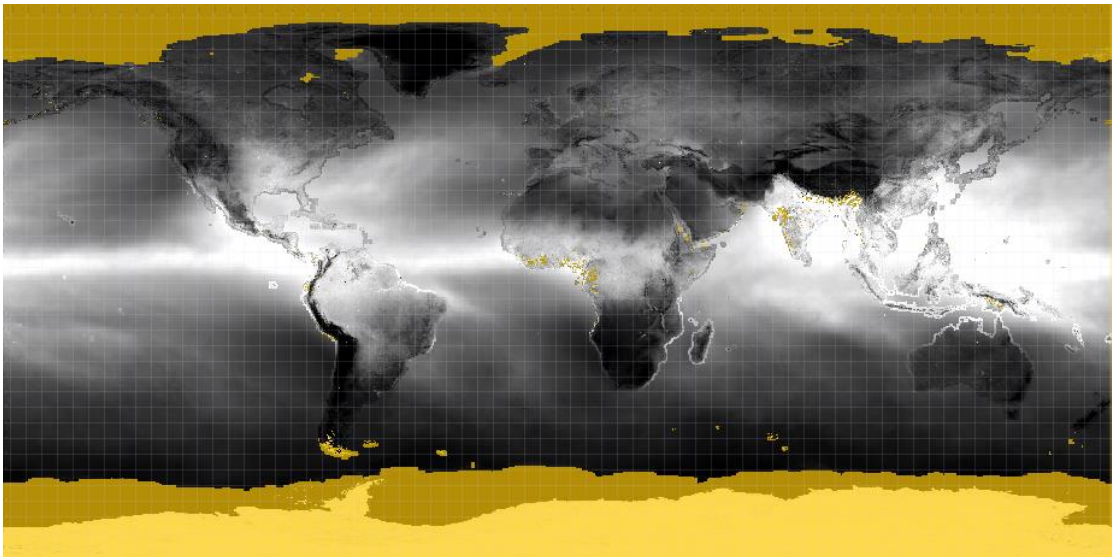


Figure 3-32: MERIS/MODIS/HOAPS TCWV L3 monthly merge for July 2011 (greyscale, 0–70 kg m^{-2}). Yellow indicates no data.

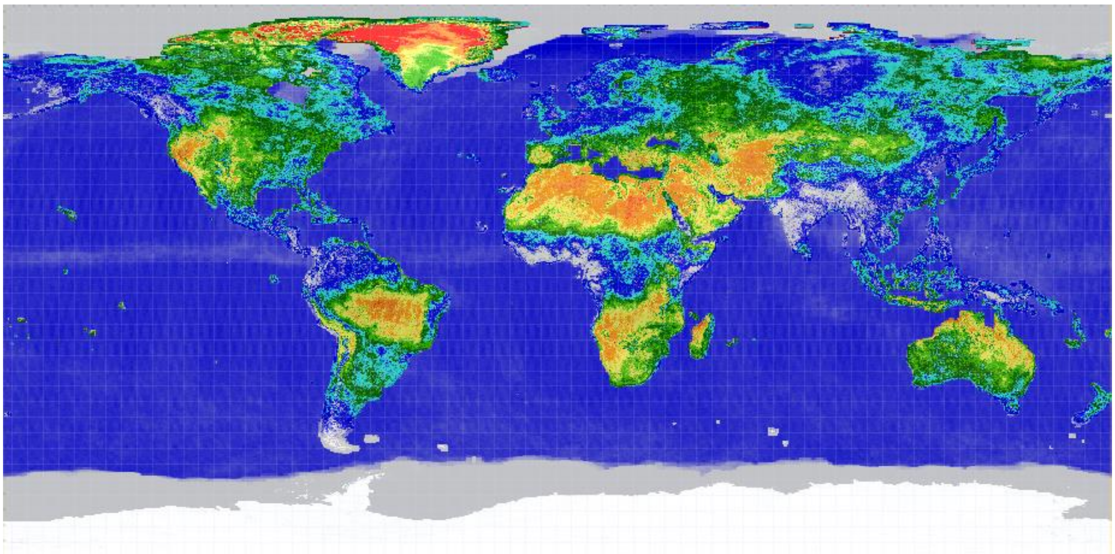


Figure 3-33: Number of MERIS+MODIS+HOAPS TCWV L2 retrievals in L3 monthly aggregate for July 2011. Colour scale is 0–7500.

Figure 3-34 to Figure 3-39 show the merge of the OLCI/MODIS merges with CM SAF HOAPS, so we have the combination of all three sensors. As we would expect, the TCWV pattern over water is again dominated by the HOAPS data, as OLCI/MODIS retrievals are used here only for gap filling, i.e. near the coasts (Figure 3-34, Figure 3-38), and over sea ice. The number of TCWV retrievals are in line with this (Figure 3-35, Figure 3-39). Figure 3-36 shows the TCWV L3 daily average (top) and retrieval (bottom) uncertainty. For both of these, the values from HOAPS over water are about an order of magnitude higher. Over water, the average uncertainty is significantly higher than retrieval uncertainty, over land it is just slightly higher. Over water, there is a discontinuity in both quantities at ~30°N. The reason for this is not clear yet. Figure 3-37 shows the L3 surface type flag. The type ‘land’ refers to clear land as identified by the pixel classification, or not identified at all (e.g. over Antarctica). The type ‘cloud over land’ means that all contributing L2 pixels were identified as cloudy (thus no TCWV retrieval), the type ‘partly cloud over land’ means that the majority but not all contributing L2 pixels were identified as cloudy (thus we have a TCWV retrieval).

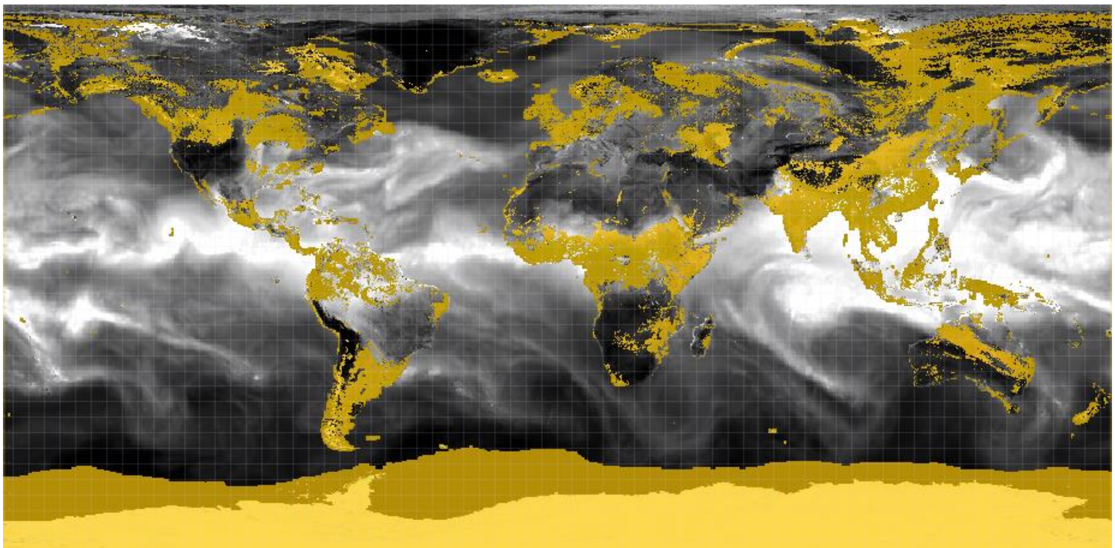


Figure 3-34: OLCI/MODIS/HOAPS TCWV L3 daily merge for July 11th, 2016 (greyscale, 0–70 kg m^{-2}). Yellow indicates no data.

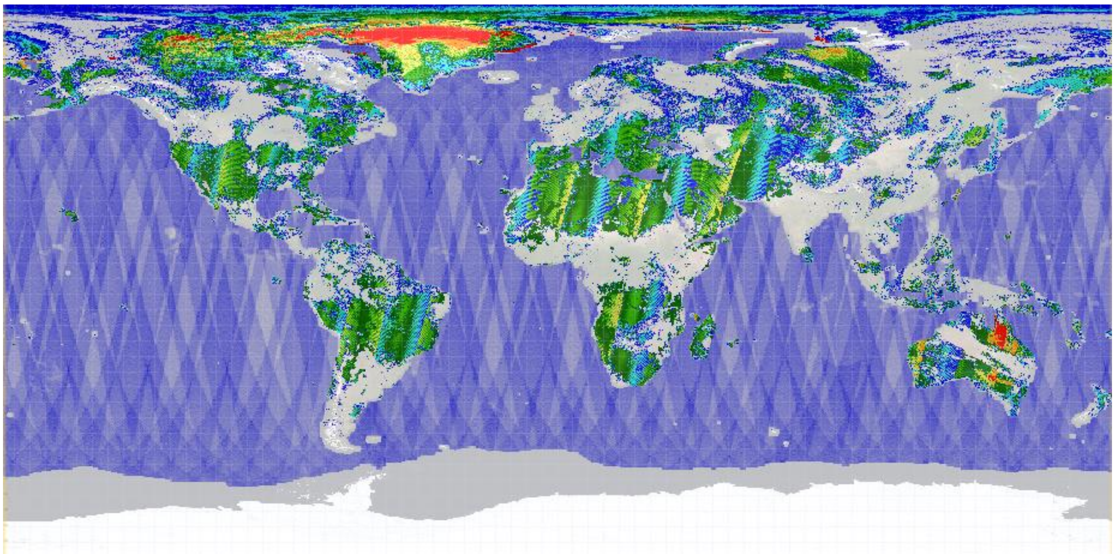


Figure 3-35: Number of OLCI+MODIS+HOAPS TCWV L2 retrievals in L3 daily aggregate for July 11th, 2016. Colour scale is 0–500.

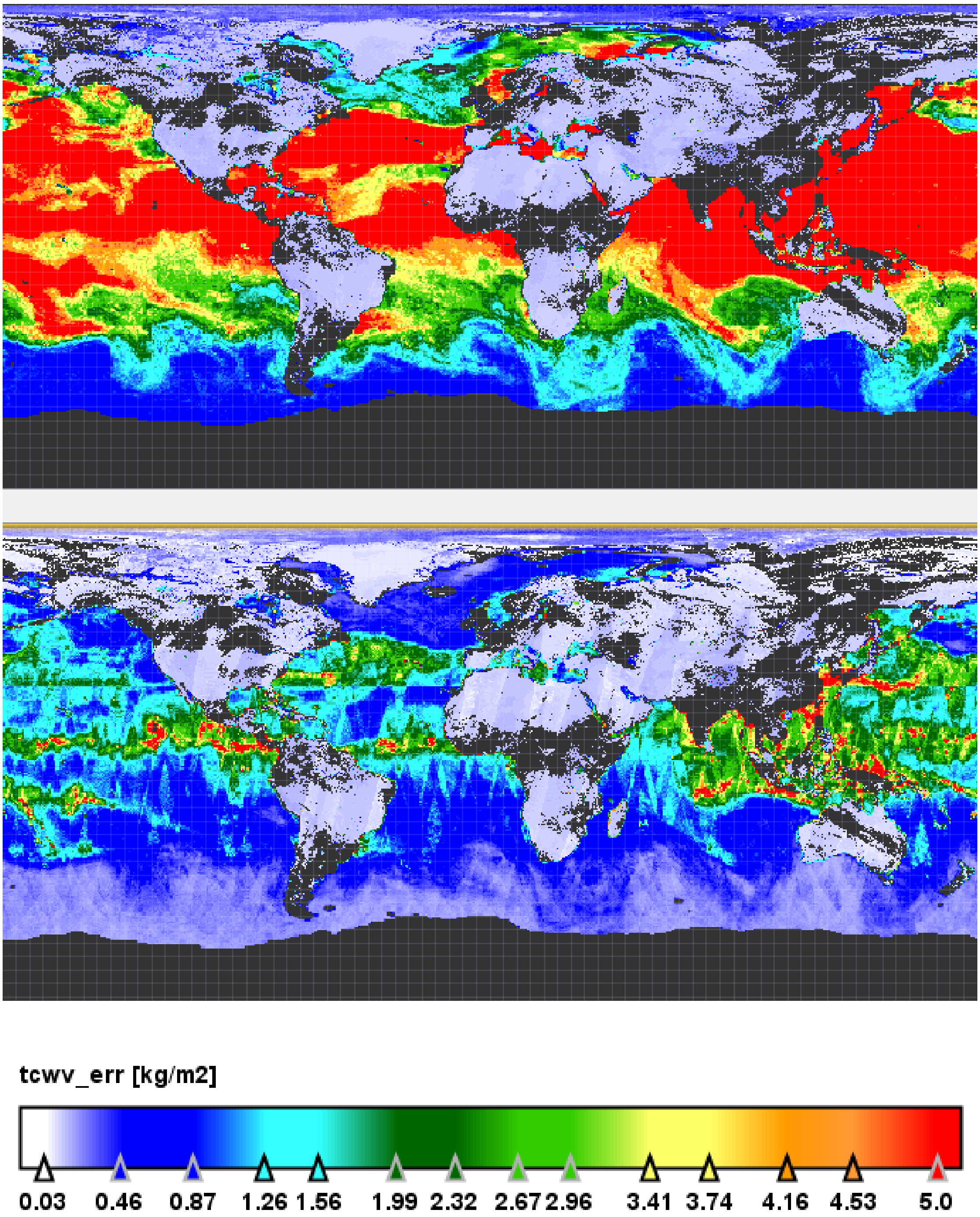
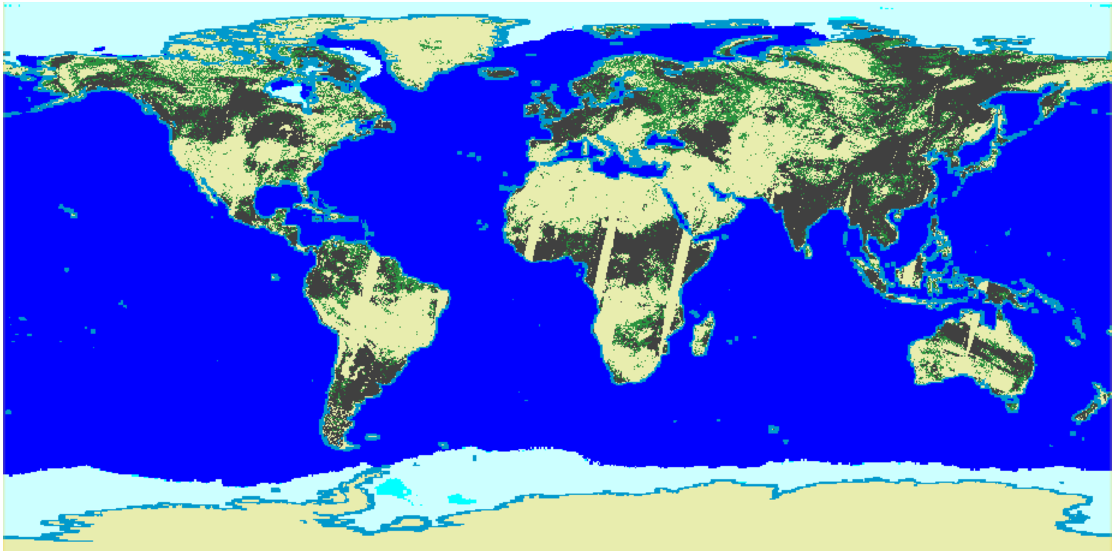
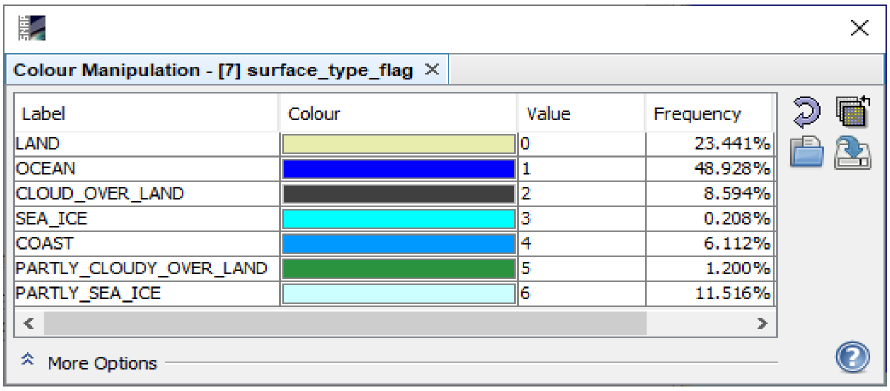


Figure 3-36: OLCI/MODIS/HOAPS TCWV L3 average (top) and retrieval (bottom) uncertainty for July 11th, 2016. Colour scale is 0–5 kg m^{-2} for both. Grey indicates no data.





Colour Manipulation - [7] surface_type_flag

Label	Colour	Value	Frequency
LAND		0	23.441%
OCEAN		1	48.928%
CLOUD_OVER_LAND		2	8.594%
SEA_ICE		3	0.208%
COAST		4	6.112%
PARTLY_CLOUDY_OVER_LAND		5	1.200%
PARTLY_SEA_ICE		6	11.516%

More Options

Figure 3-37: L3 surface type flag for July 11th, 2016. Colours are explained in the bottom figure.

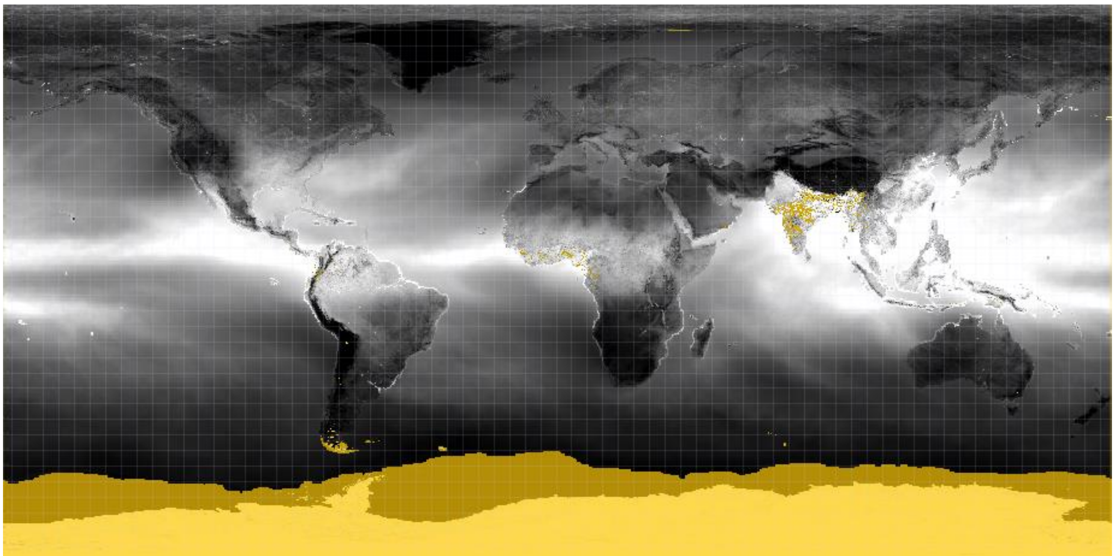


Figure 3-38: OLCI/MODIS/HOAPS TCWV L3 monthly merge for July 2016 (greyscale, 0–70 kg m^{-2}). Yellow indicates no data.

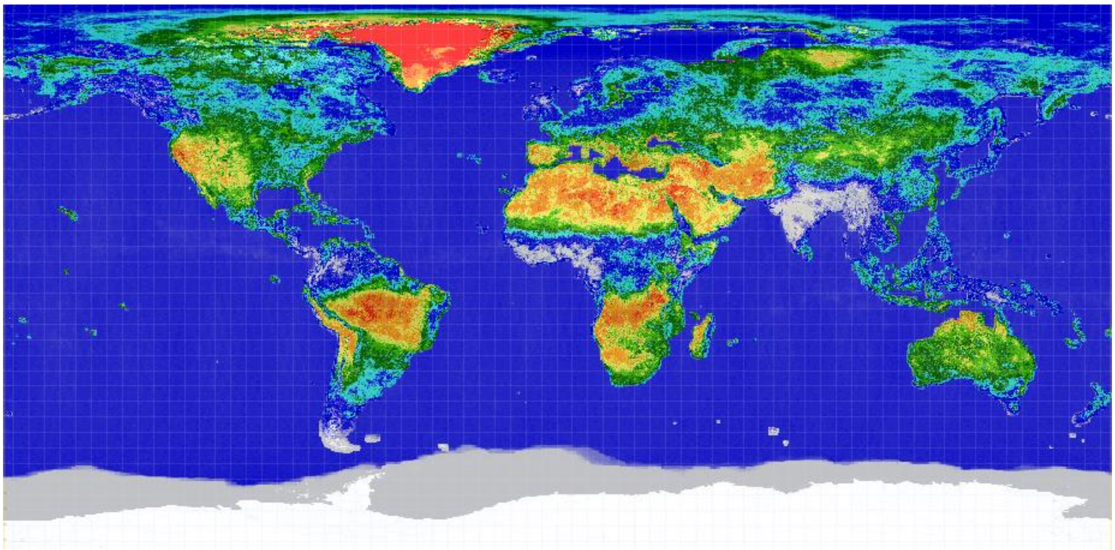


Figure 3-39: Number of OLCI+MODIS+HOAPS TCWV L2 retrievals in L3 monthly aggregate for July 2016. Colour scale is 0–7500.

3.1.5 Final Products

3.1.5.1 CF- and CCI compliance

To verify that a final TCWV product follows the NetCDF CF metadata conventions [12], it is checked by a web based compliance checker, e.g. [13]. An example for the output from this check is shown in Figure 3-40 for the MODIS monthly aggregate product for July 2011. The indicator for CF compliance is that no errors were detected in the product. The report may also provide warnings or information with recommendations to further improve the metadata format in the product.

CF-Convention Compliance Checker for NetCDF Format

Checking against CF version auto...

Check another file | NetCDF format | CF Convention.

File name: ESACCI-WATERVAPOUR-L3C-TCWV-meris-05deg-201107-fv2.2.nc

Output of CF-Checker follows...

```
/home/ros/anaconda2/lib/python2.7/site-packages/cfchecker/cfchecks.py:2435: RuntimeWarning: invalid value encountered in less
  varData=self.f.variables[varName][:]
/home/ros/anaconda2/lib/python2.7/site-packages/cfchecker/cfchecks.py:2435: RuntimeWarning: invalid value encountered in greater
  varData=self.f.variables[varName][:]
CHECKING NetCDF FILE: /tmp/3859.nc
=====================
Using CF Checker Version 3.1.1
Checking against CF Version CF-1.7
Using Standard Name Table Version 72 (2020-03-10T11:52:02Z)
Using Area Type Table Version 9 (07 August 2018)
Using Standardized Region Name Table Version 4 (18 December 2018)


------------------
Checking variable: time
------------------

------------------
Checking variable: time_bnds
------------------

------------------
Checking variable: lat_bnds
------------------

------------------
Checking variable: lon_bnds
------------------
```

```
------------------
Checking variable: lon_bnds
------------------

------------------
Checking variable: tcwv
------------------

------------------
Checking variable: num_obs
------------------

------------------
Checking variable: stdv
------------------

------------------
Checking variable: tcwv_err
------------------

------------------
Checking variable: tcwv_ran
------------------

------------------
Checking variable: lat
------------------

------------------
Checking variable: lon
------------------

------------------
Checking variable: crs
------------------

------------------
Checking variable: surface_type_flag
------------------

ERRORS detected: 0
WARNINGS given: 0
INFORMATION messages: 0
```

Figure 3-40: CF compliance check result for MERIS July 2011 monthly 0.5-degree aggregate product.

Besides the CF metadata conventions, the WV_cci products, as any other CCI products, must also follow the CCI data standards described in detail in [14]. This is checked by examination of the layout and the content of the NetCDF products with ncdump. An example ncdump output for a daily L3 TCWV product is shown in Appendix 4: NetCDF ‘ncdump’ example.

3.1.5.2 Completeness

To verify that the product set generated by a given processing task is complete, the number of the final products is simply counted by appropriate directory listings. This number depends on the available sensors for the given time period, e.g. for the month July 2011 we have all MERIS, MODIS and HOAPS available, so there should be:

- 31 products of daily merges MERIS/MODIS (CDR-1, land based TCWV, see PSD);
- 1 product of monthly merge MERIS/MODIS;
- 31 products of daily merges MERIS/MODIS/HOAPS (CDR-2, land+ocean based TCWV, see PSD);
- 1 product of monthly merge MERIS/MODIS/HOAPS.

As in Section 3.1.4, we count them like:

```
>ls /calvalus/projects/wvcci/tcwv/meris-modis_terra/l3-daily-final-nc/05/2011/07/*/*.nc |wc -l
>31


>ls /calvalus/projects/wvcci/tcwv/meris-modis_terra/l3-monthly-final-nc/05/2011/07/*.nc |wc -l
>1
>ls /calvalus/projects/wvcci/tcwv/meris-modis_terra-cmsaf_hoaps/l3-daily-final-nc/05/2011/07/*/*.nc |wc -l
>31


>ls /calvalus/projects/wvcci/tcwv/meris-modis_terra-cmsaf_hoaps/l3-monthly-final-nc/05/2011/07/*.nc |wc -l
>1
```

3.1.5.3 Dissemination to users

The details of the data dissemination to users needs to be further discussed and agreed upon at a later date with regard to the final data products. In particular, it is planned that the final products will be released by EUMETSAT CM SAF if formal approval by appropriate bodies is given. Up to now, the data products, including the final datasets, were internally provided by BC to the consortium via ftp.

3.1.6 Advanced verification

The basic verification steps described in the previous sections are essential and certainly very important and valuable. However, for a given processing task, more advanced verification steps could include, e.g.:

- do all generated products follow the naming conventions?
- can all generated products be opened and read with the appropriate tools? → This step has been implemented in a bash script on JASMIN and Calvalus and has been applied for previous product versions from ‘Dataset 2’ onwards, and also for the final products.
- do all products have the expected product version?
- do all products contain all specified variables?
- are the data of all variables in their expected value range?

As these questions consider all products of a generated dataset, an appropriate verification framework needs to be implemented. This has been done in the frame of the SST_cci (T. Block, Brockmann Consult, personal communication), and first results indicate that this kind of framework would be very valuable also for WV_cci. This will be further investigated and addressed in a potential future phase of WV_cci if appropriate.

3.2 VRWV Processing System

3.2.1 VRWV CDR-3

3.2.1.1 Preprocessing

The preprocessing steps applied in the CDR-3 VRWV retrieval are:

- Ingestion of single instrument L3 SPARC Data Initiative climatologies
- Quality control of the ingested climatologies
- Ingestion of chemistry–climate model (CCM) WV fields
- Temporal and spatial interpolation of CCM input onto 2-dimensional L3 product grid
- Quality control of the interpolated CCM anxillary product.

3.2.1.1.1 Code sanity checks

A number of code sanity checks have been introduced into the pre-processing code:

- the preprocessing fails if the files of a certain instrument cannot be found/loaded;
- the preprocessing fails if the instrument files contain WV in the wrong unit;
- the preprocessing fails if the model files contain WV units that are inconsistent with the units of the observed WV products.

The processing stops as soon as one of the above code sanity checks fails. An example of the code is provided in Figure 3-41.

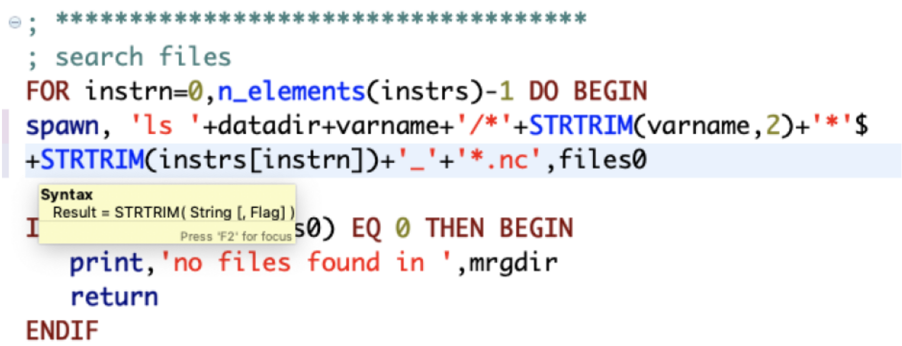


Figure 3-41: Example IDL code that generates an error message if a given instrument's file ingestion fails.

3.2.1.1.2 Unit-level testing

The unit-level testing includes the preprocessing tests on each instrument's input data set. The tests are performed on the RACC cluster with the IDL ingestion code. All invalid input data are excluded in the processing system and returned to the data provider for improvement.

3.2.1.1.3 Visual inspection

A number of each limb sounder's input files of VRWV SPARC Data Initiative L3 products has been visually inspected using the Panoply tool. At verification stage the checks are:

- the products must contain all variables as specified in the PSD;
- the VRWV values must be in an expected range.
- the VRWV units must be correct (in ppmv);

- the VRWV cross-sections should not contain significant artefacts;
- the VRWV cross-sections should show typical stratospheric water vapour distributions.

Figure 3-42 shows example VRWV L3 input data fields from SMR and Aura-MLS for January 2005. Both datasets show typical stratospheric tropopause distributions with isopleths following approximately the tropopause shape, although other details can differ substantially.

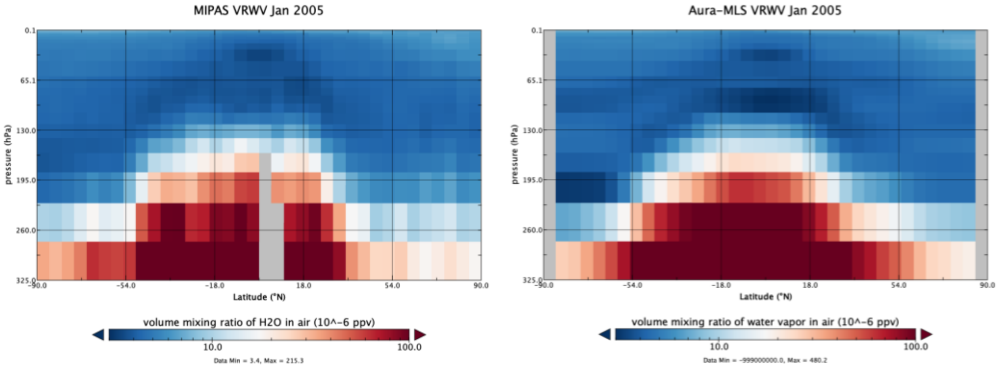


Figure 3-42: MIPAS and Aura-MLS crossections for January 2005.

3.2.1.2 VRWV bias correction processing

The main goal relating to CDR-3 within WV_cci is to obtain a merged zonal mean, monthly mean vertically resolved WV product from all currently available Limb sounders in the stratosphere. The processing of the data thereby focuses on an improved bias-correction of the input data of different Limb sounders so to improve the homogeneity of the resulting CDR.

The steps involved in the bias correction processing step applied to the CDR-3 VRWV are specified in the ATBD.

3.2.1.2.1 Code sanity checks

A number of code sanity checks have been introduced into the VRWV bias correction processing code of CDR-3. The most important ones are:

- the bias correction processing fails if the input data is regarded invalid;
- the bias correction processing fails if ancillary CCM data cannot be loaded;

- the bias correction fails if geo-spatial variables and time periods cannot be matched.
- The bias correction flags data points that exhibit exceptionally large biases against the CCM relative to other limb sounder's data.

3.2.1.2.2 Unit-level testing

The unit-level testing applies to each SPARC Data Initiative instrument's input data after bias-correction. The tests are performed on the RACC cluster within the IDL bias correction code. All files have to pass the test after they have been bias-corrected, otherwise the files are not passed onwards to the merging step.

3.2.1.2.3 Visual inspection

The relatively small volume of SPARC Data Initiative L3 input data allows for a visual inspection that is not very time-consuming (1–2 days). This step is crucial to find missed outliers and is supporting the unit level testing step. One test is the QSUM-plot (accumulated sum of the differences between observations and model) that allows for identification of specific drifts in either the instrument (when multiple instruments disagree with each other) or the validity of the model simulation (when QSUMs between instruments agree on the tendency away from the model) (see Figure 3-43).

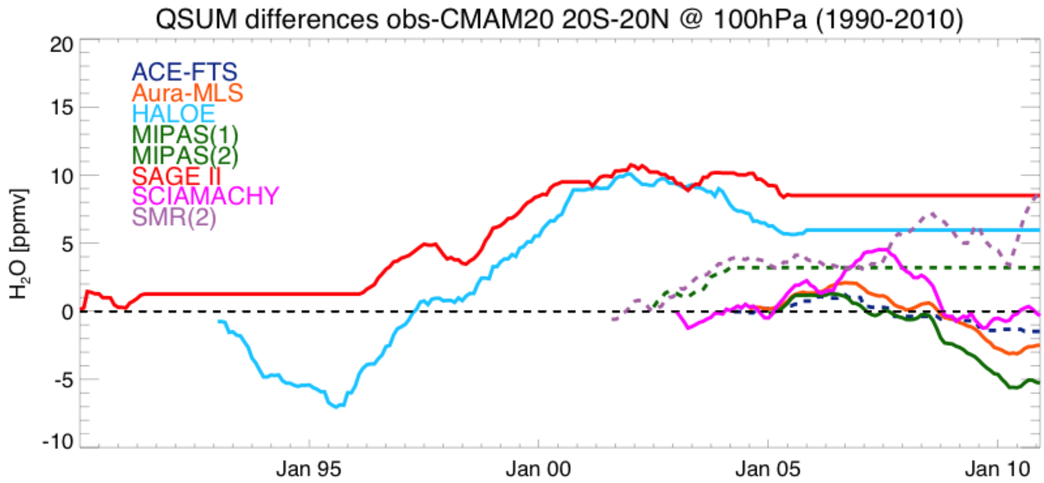


Figure 3-43: QSUMs for differences between each observational dataset and the CCM CMAM.

3.2.1.3 VRWV seasonal correction processing

Homogenisation of multiple datasets from different instruments into one climate data record can be challenging due to differences in instrument sensitivities, horizontal and vertical resolutions, as well as spatial and temporal sampling patterns. The more instruments that are considered in a climate data record, the more challenging homogenisation across these instrument characteristics can become. The observational characteristics of the different instruments considered within the WV_cci have been found to lead to inhomogeneities in the seasonal cycle amplitude of CDR-3, with the severity of this issue being latitude and altitude dependent. An example is shown in Figure 3-44. However, no satisfying approach to the seasonal bias correction processing has been found by this point of the project due to the strong latitude and altitude dependence of the issue and the lack of a reliable reference (both CCMs and SWOOSH may have their own inhomogeneities).

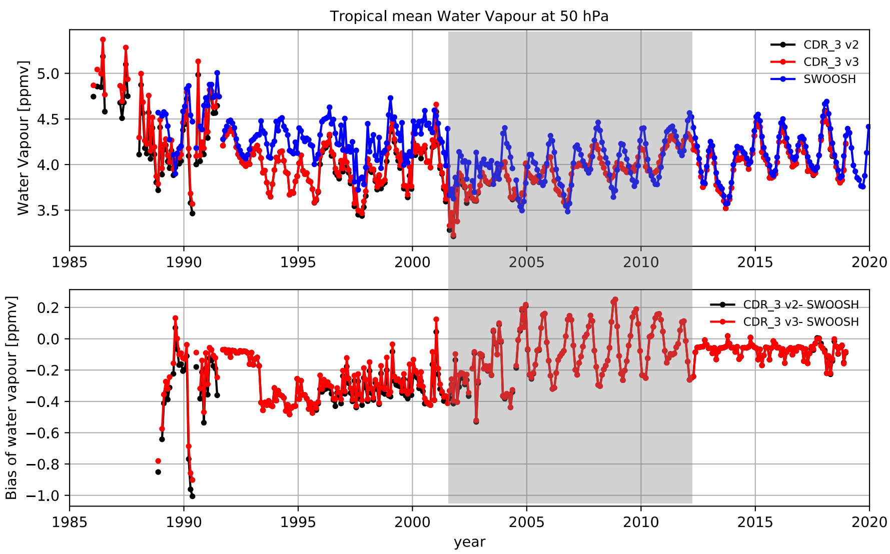


Figure 3-44: Differences between the water vapour time series at 50 hPa and for the tropics from CDR-3 and SWOOSH. The differences between the datasets are time-dependent and thus indicating an inhomogeneity in the representation of the seasonal cycle amplitude, especially during the ENVISAT period (grey shading). Ideally, the differences between the two climate data records would show a similar pattern over the whole time period.

3.2.1.4 VRWV product merging

The product merging is performed on the L3 products obtained from Section 3.2.1.3. The VRWV L3 product merging uses IDL code on the RACC. The merged product has a horizontal resolution of 5 degrees with the pressure levels specified by the SPARC Data Initiative (see DARD [15]).

3.2.1.4.1 Code sanity checks

A number of code sanity checks have been introduced into the VRVW L3 product merging processing code. VRVW L3 product merging fails if:

- one of the input products does not fulfil the expected specifications;
- the input products are not given on the same grid or have a different size;
- the time period (i.e. the reference month) of the input products is different.

3.2.1.4.2 Visual Inspection

The final VRWV CDR-3 merged product has undergone visual inspection. Figure 3-45 shows an example for January 2005. When compared to Figure 3-42 one can see that the minimum in Aura-MLS in the Southern hemisphere lower stratosphere has been corrected towards higher values.

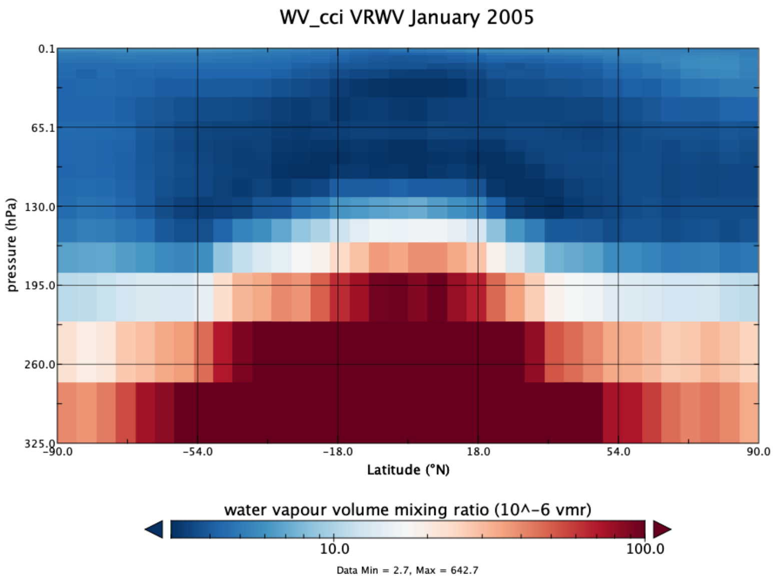


Figure 3-45: Merged ESA WV_cci product for January 2005.

3.2.2 VRWV CDR-4

3.2.2.1 Preprocessing/ingestion

The preprocessing steps applied in the CDR-4 VRWV retrieval are:

- Ingestion of input satellite data and balloon-borne hygrometer profiles;
- Correspondence of VRWV retrieval from all input satellite data and the derived meteorological products;
- Quality control on the input satellite data and balloon-borne hygrometer profiles and also the corresponding derived meteorological products.

3.2.2.1.1 Code sanity checks

Checks on the code sanity are applied into the preprocessing steps, including the following:

- the preprocessing fails if the input data is invalid;
- the preprocessing fails if the input data has no corresponding derived meteorological products (e.g. geopotental height and tropopause height);
- the preprocessing fails if the derived meteorological products have multiple thermal tropopauses.

3.2.2.1.2 Unit-level testing

The processing system check includes the preprocessing tests on each input data shown in Table 3-1. The tests are performed on the RACC cluster with Python code and all the tests passed successfully. After the preprocessing test, all the invalid input data (e.g. missing corresponding DMP or without ‘reasonable’ DMP) are excluded in the processing system.

For extending to the full length of the CDR-4 (2010–2014), the preprocessing on IMS is duplicated to JASMIN system with approporiate updates on the Python codes. As the JASMIN system has access to the RAL IMS archive on CEDA, the DMP data was uploaded to JASMIN once obtained from NASA JPL. Tests on the preprocessing on both systems (JASMIN and RACC) provided consistent results.

Table 3-1: Test data used by the preprocessing tests

Test Data	Test Data Description
Aura MLS L2 v4.2	Two months of Aura MLS L2 data for January/July 2010
MIPAS L2	Two months of MIPAS L2 data for January/July 2010
IMS L2	Two months of IMS L2 data for January/July 2010
FPH/CFH	Balloon-borne hygrometer VRWV profiles from 2000–2017
Auxiliary Data	DMP data for all above input data

3.2.2.2 VRWV bias correction processing

3.2.2.2.1 Code sanity checks

A number of code sanity checks have been introduced into the VRWV bias correction processing code. The most important ones are:

- the bias correction processing fails if the input data is regarded invalid;
- the bias correction processing fails if ancillary data cannot be loaded;
- the bias correction processing fails if FPH/CFH data is regarded invalid;
- the bias correction processing fails if geo-information is missing
- if multiple thermal tropopauses are provided by the DMP, only the lowest tropopause is used in the bias correction processing

3.2.2.2.2 Unit-level testing

With all input data shown in Table 3-1 after the preprocessing tests, the bias correction processing is applied to all input data. The processing module for the bias correction is developed using Python according to the algorithm detailed in the ATBD. The bias correction process applied biases to each L2 VRWV profile individually for each input instrument with respect to the FPH/CFH profiles in tropopause-based coordinates. All the tests must pass successfully in order to build and deploy the codes into the processing system.

Table 3-2 gives an overview of all test cases performed. All these tests are run in RACC cluster with one core. The third column indicates the summary of the test result for each test case. All the test cases need to pass successfully.

As a final check, the number of VRWV profiles after bias correction processing need to be the same number of original VRWV profiles for each input data. After the bias correction processing, all the L2 VRWV profiles are linearly interpolated in the vertical onto the pressure levels given for the CDR-4 product using the logarithm of pressure. Tests are carried out to check that the linear interpolation processing did not introduce anomalous errors into the VRWV profiles.

For extending to the full length of the CDR-4 (2010–2014), the processing on IMS is replicated to the JASMIN system with appropriate updates on the Python codes. Tests on these two systems provided consistent results and the processing speed for submitted jobs on JASMIN is about twice quicker than those on the RACC. This system change was a very important step in the system development to allow for the faster processing of the large size IMS data.

Table 3-2: Test cases and test results for the bias correction processing

Test Case	Summary Test Case Description	Summary Test Result
Aura MLS bias correction processing	Test the bias correction processing of two month of Aura MLS data	Pass Processing time for two months: 20 minutes
MIPAS bias correction processing	Test the bias correction processing of two month of MIPAS data	Pass Processing time for two months: 15 minutes
IMS bias correction processing	Test the bias correction processing of two month of IMS data	Pass Processing time for two months: 2 hours

3.2.2.3 VRWV L3 processing

The VRWV L3 processing of all input VRWV L2 profiles, both before and after bias correction processing, is carried out on the RACC with Python code. The L3 processing consists of the spatial aggregation and monthly temporal aggregation of L2 VRWV profiles with 5-degrees by 5-degrees horizontal resolution of latitude and longitude on each given pressure levels. The further processing of the RAL IMS for the full coverage of CDR-4 is performed on JASMIN instead.

All L3 processing test cases for Aura MLS, MIPAS, and RAL IMS input data are passed successfully on RACC for the two months January/July 2010 in the UTLS region. The final VRWV L3 products for each input data have been visually inspected through Panoply **Error! Reference source not found.**to **Error! Reference source not found.**show examples of monthly WV from Aura MLS before bias-correction in January/July 2010 on 150 hPa and 250 hPa, respectively. Figure 3-50 to Figure 3-53 show corresponding Aura MLS WV horizontal distributions after bias correction processing. Overall, as expected, the bias corrrection processing keeps the pattern of horizontal distributions but increases the VRWV value (note the different colour bar ranges in the corresponding figures). Further reports on the evaluations of the VRWV L3 product are provided in the PVIR.

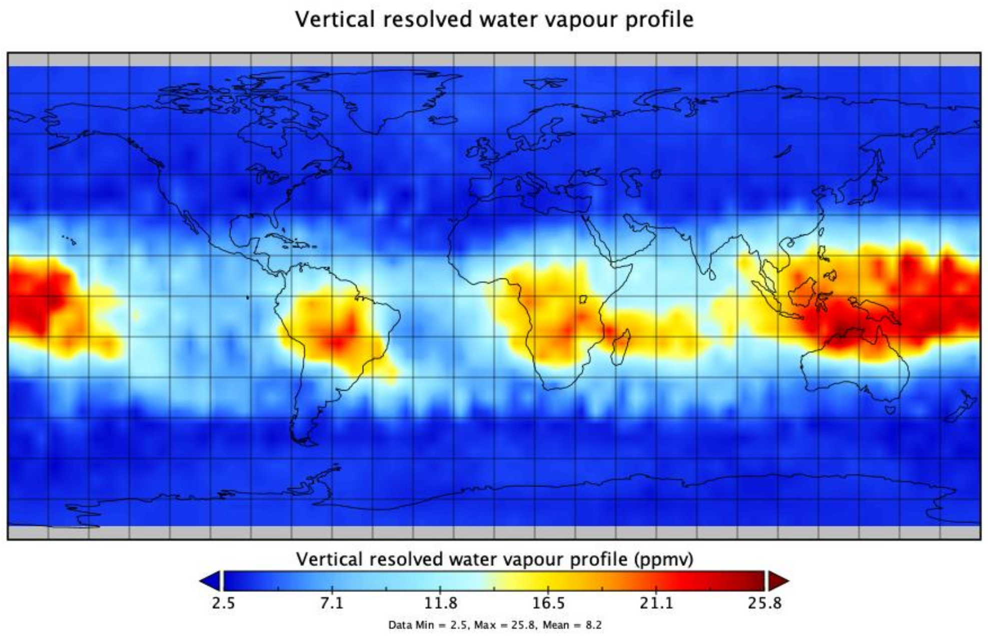


Figure 3-46: Aura MLS L3 monthly VRWV before bias correction on 150 hPa in January 2010.

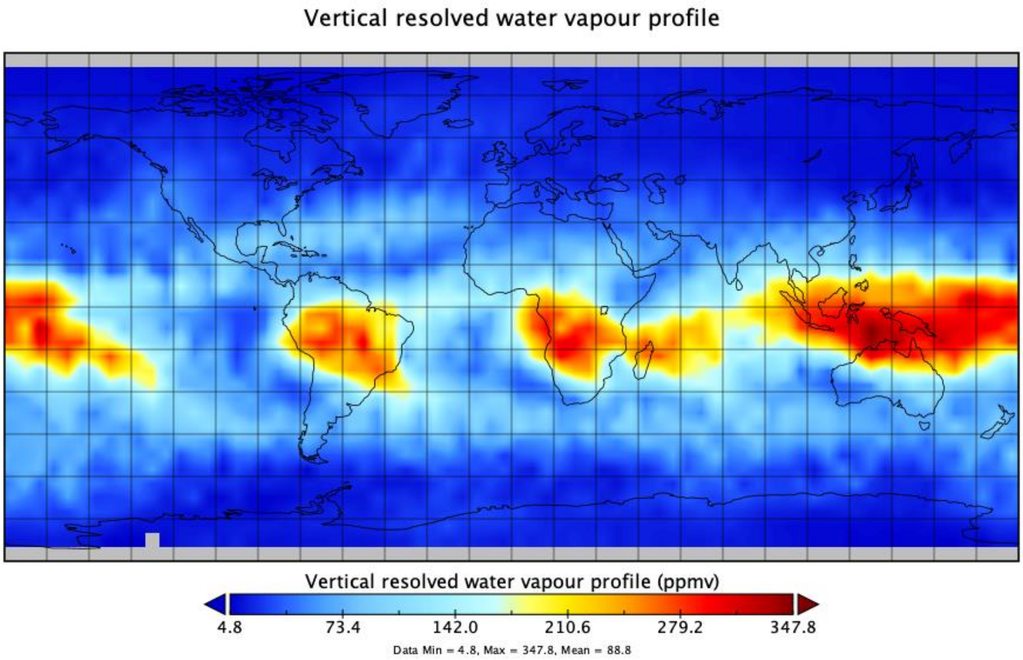


Figure 3-47: Aura MLS L3 monthly VRWV before bias correction on 250 hPa in January 2010.

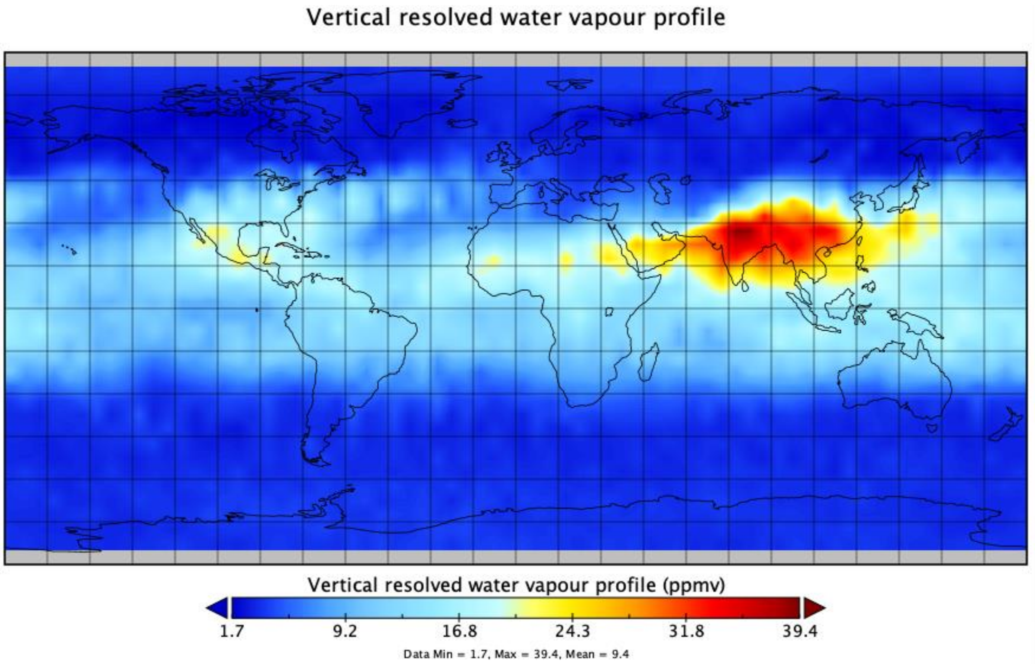


Figure 3-48: Aura MLS L3 monthly VRWV before bias correction on 150 hPa in July 2010.

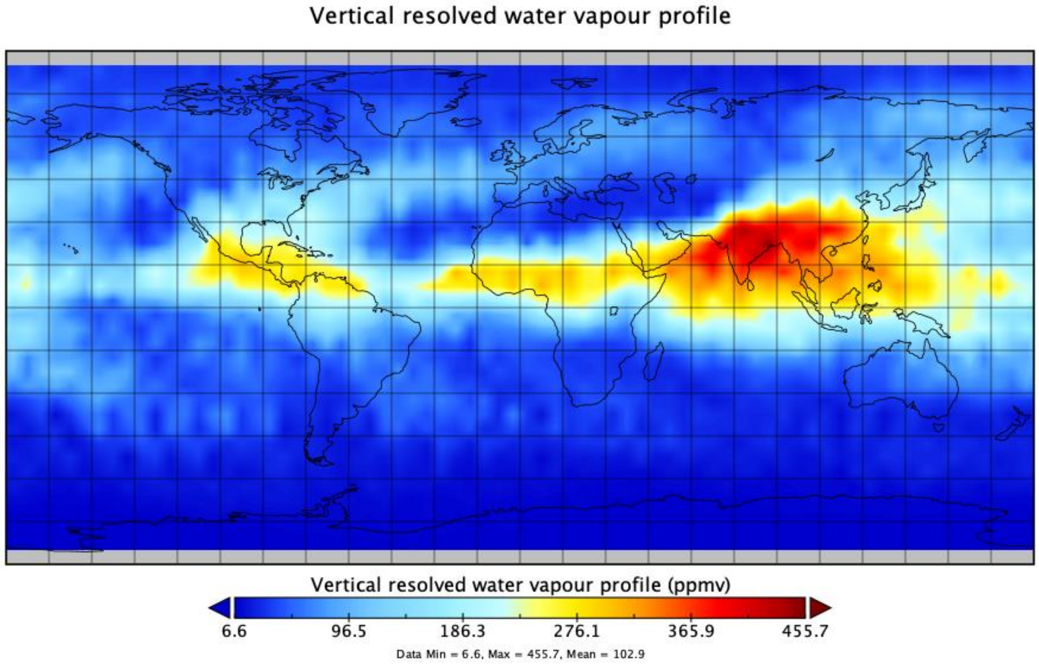


Figure 3-49: Aura MLS L3 monthly VRWV before bias correction on 250 hPa in July 2010.

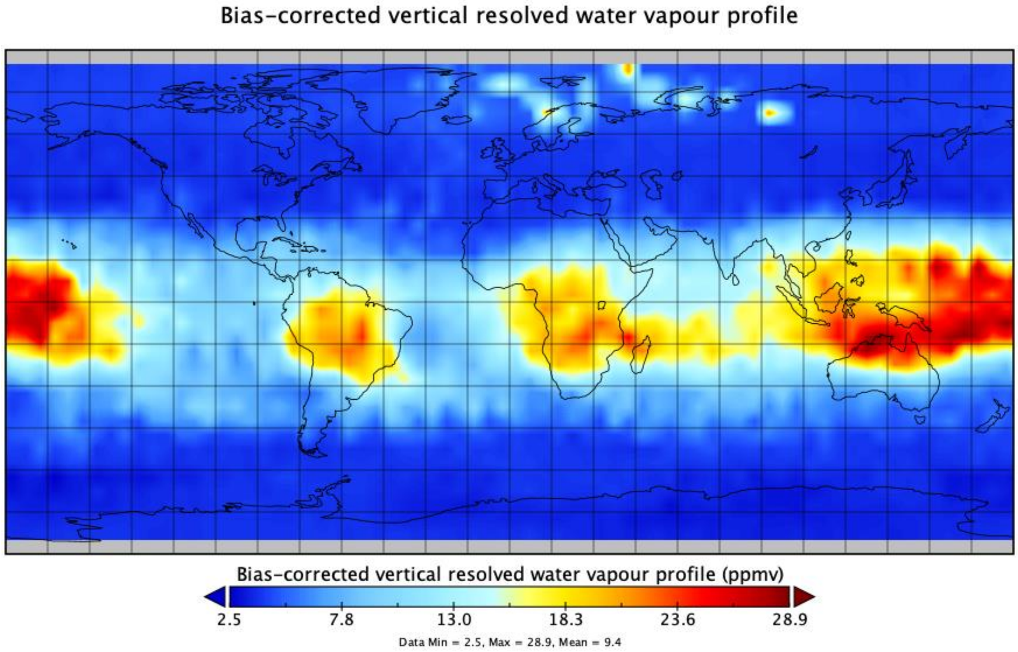


Figure 3-50: Aura MLS L3 monthly VRWV after bias correction on 150 hPa in January 2010.

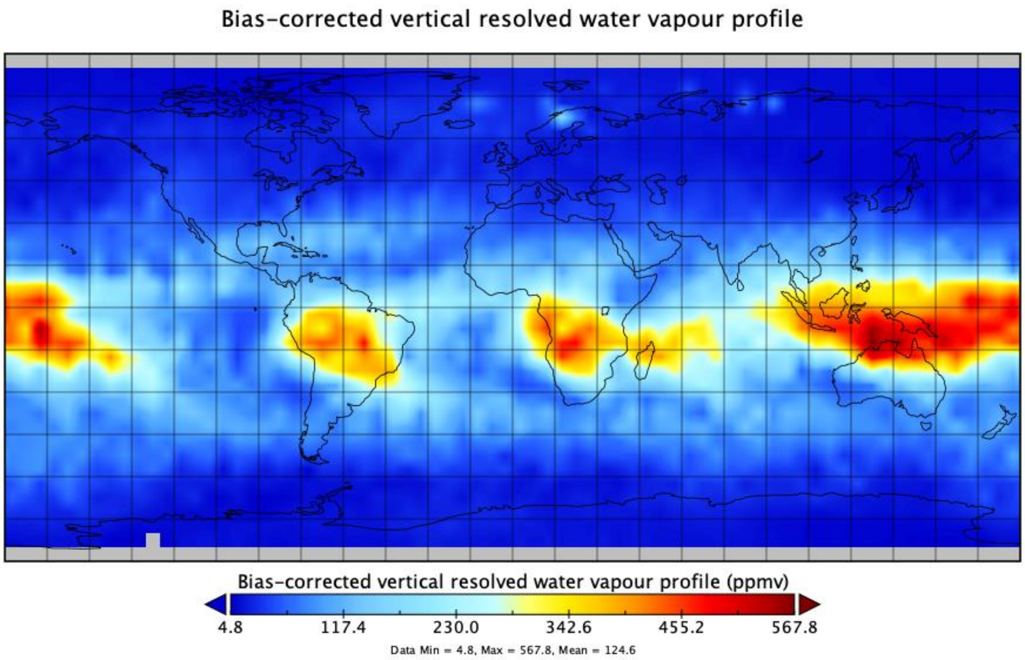


Figure 3-51: Aura MLS L3 monthly VRWV after bias correction on 250 hPa in January 2010.

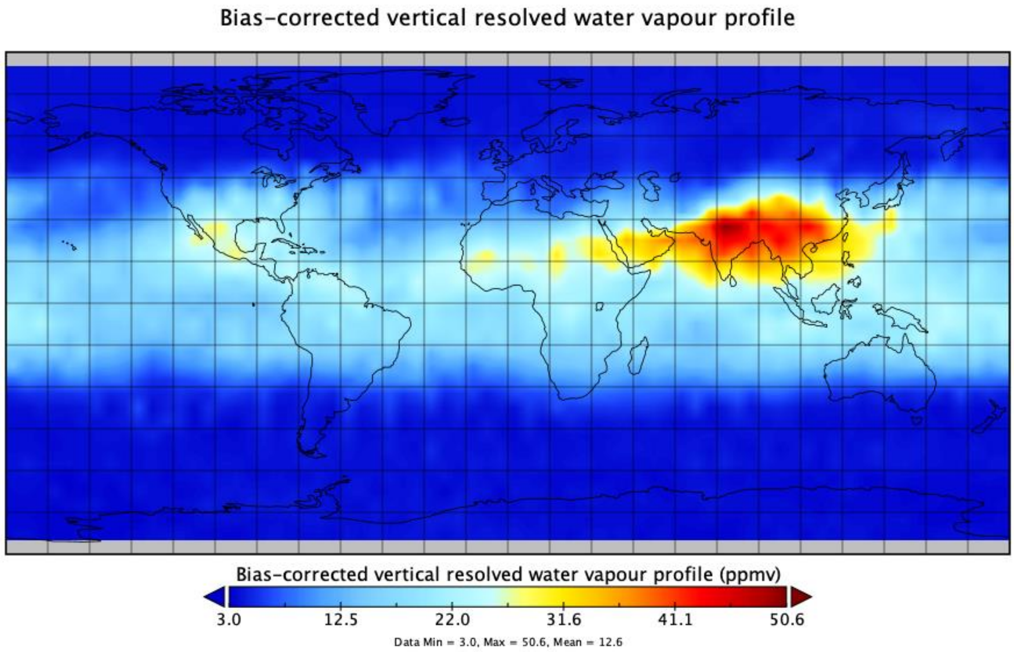


Figure 3-52: Aura MLS L3 monthly VRWV after bias correction on 150 hPa in July 2010.

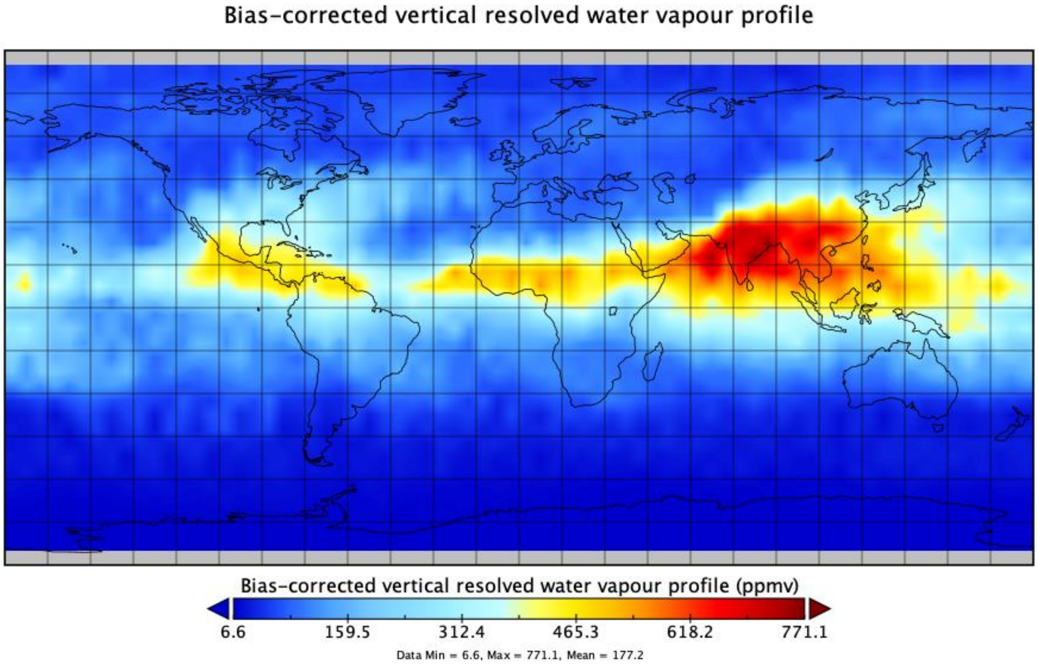


Figure 3-53: Aura MLS L3 monthly VRWV after bias correction on 250 hPa in July 2010.

3.2.2.4 VRWV product merging

One of the main goals of the VRWV data production within WV_cci is to obtain a merged monthly 3-D vertically resolved CDR-4 prototype product from both Limb and Nadir sounders in the UTLS and tropospheric regions. This should ideally result in an improvement of the global VRWV profiles and an attempt on a merging methodology on VRWV observations from these two different types of satellite instruments.

The product merging processing is carried out on the L3 products from all input data obtained from Section 3.2.2.3. The merging module is coded with Python and executed on the RACC system. The merged product has a horizontal resolution of 5 degrees by 5 degrees in latitude and longitude on **Error! Reference source not found.** given pressure levels. The final VRWV merged product has been visually inspected through Panoply to show examples of the merged L3 VRWV after bia -correction in January/July 2010 on 150 hPa and 250 hPa, respectively. Overall, the merged products followed our expectations and further documentation of the evaluation of the merged product is provided in the PVIR.

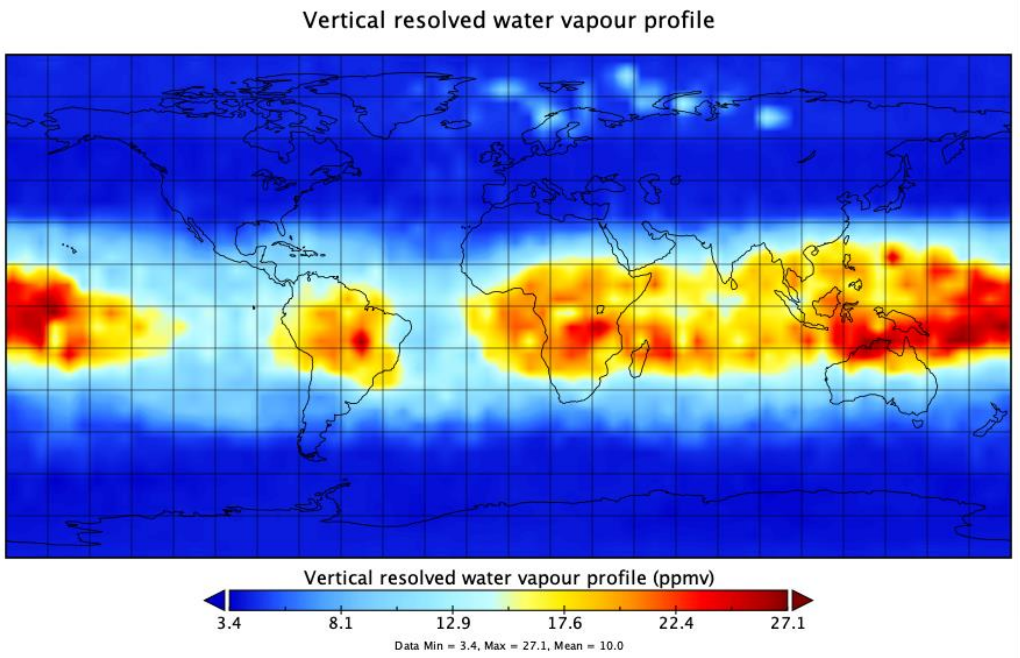


Figure 3-54: L3 monthly VRWV merged product on 150 hPa in January 2010.

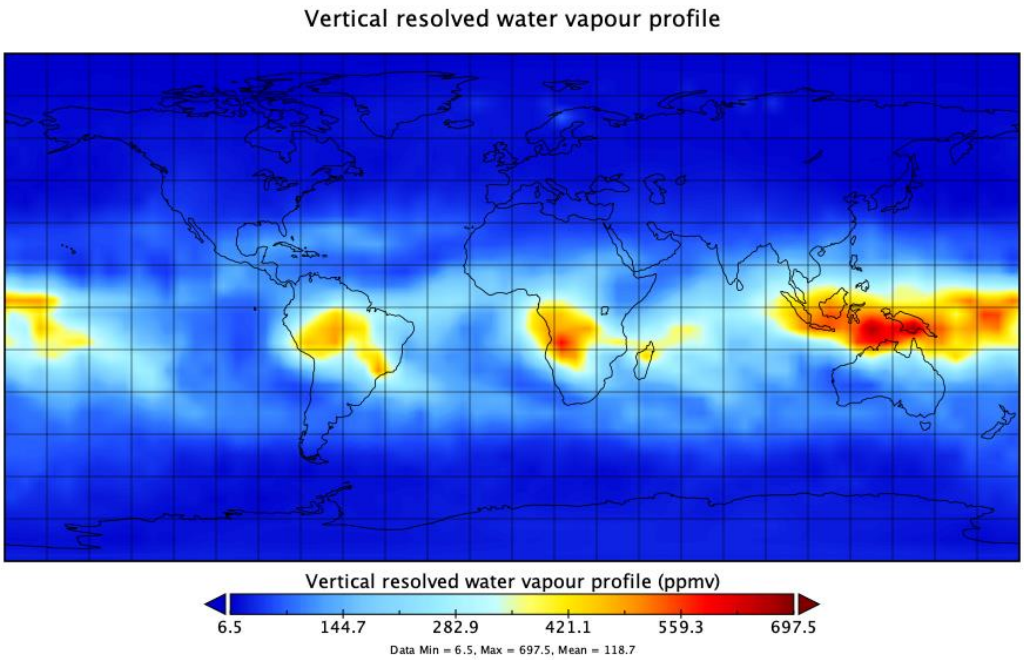


Figure 3-55: L3 monthly VRWV merged product on 250 hPa in January 2010.

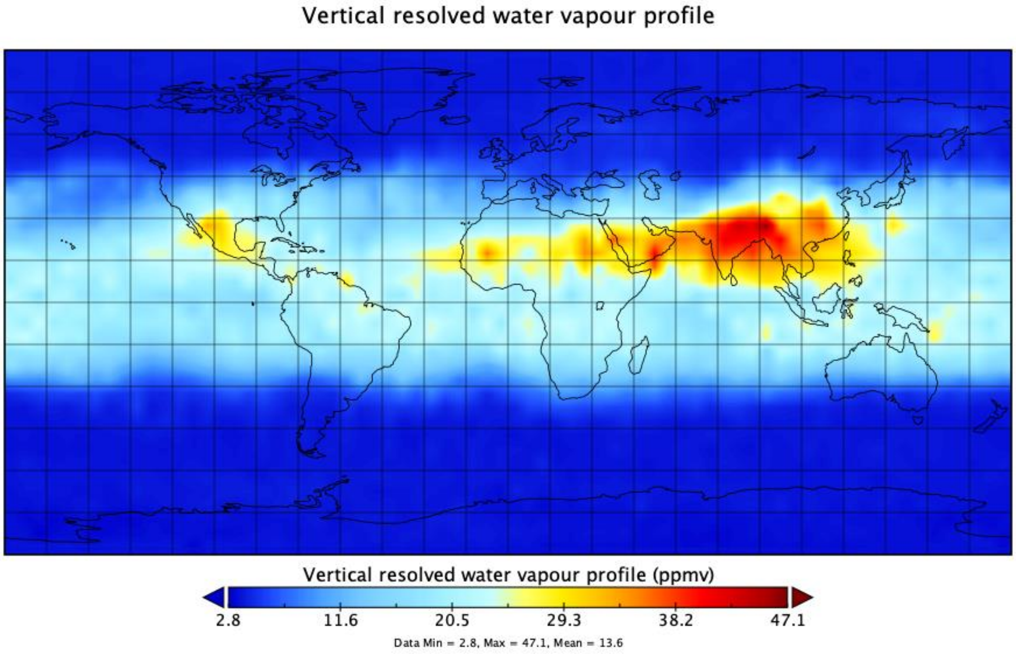


Figure 3-56: L3 monthly VRWV merged product on 150 hPa in July 2010.

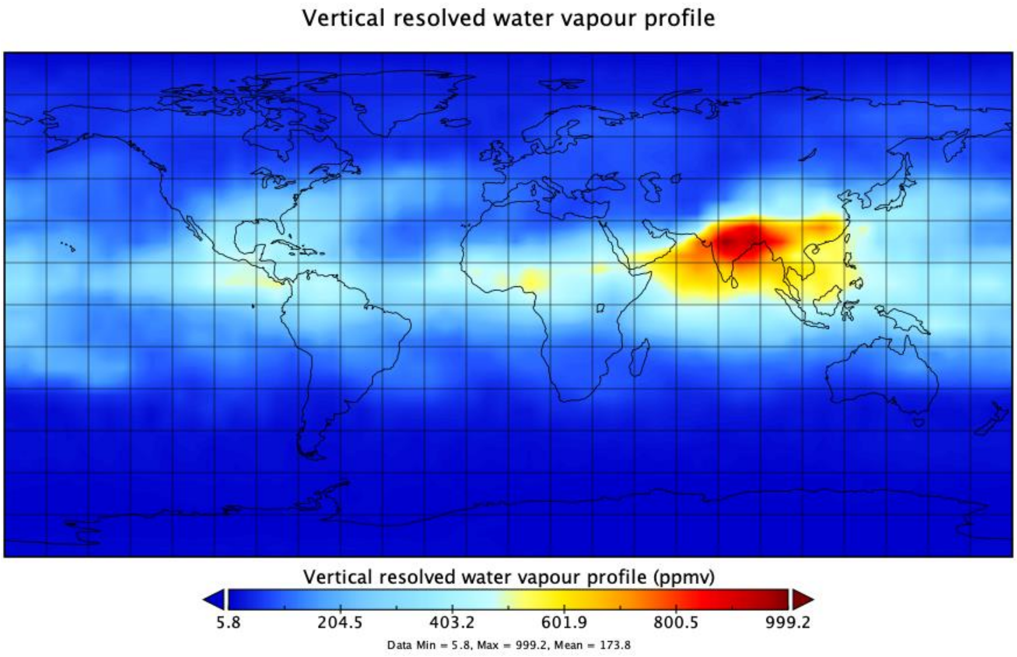


Figure 3-57: L3 monthly VRWV merged product on 250 hPa in July 2010.

3.2.3 Final products

3.2.3.1 CF- and CCI compliance

To verify that the final VRWV CDR-3 and CDR-4 products follows the NetCDF CF metadata conventions [12], it is checked by a web-based CF-convention compliance checker for NetCDF Format (e.g. https://pumatest.nerc.ac.uk/cgi-bin/cf-checker.pl). An example for the output from this check is shown in for an early version of the WV_cci CDR-3 product which has not passed the test (Figure 3-58). The indicator for CF compliance is that no errors were detected in the product. The report may also provide warnings or information with recommendations to further improve the metadata format in the product. Figure 3-59 shows the same output from the check of the final CDR-3 v3.0 product.

Besides the CF metadata conventions, the WV_cci products, as any other CCI products, must also follow the CCI data standards described in detail in [14]. This is checked by examination of the layout and the content of the NetCDF products with ncdump. An example ncdump for the monthly CDR-4 product is shown in Figure 3-60.

File name: H2O_1985-2020_v0.nc

Output of CF-Checker follows...

```
CHECKING NetCDF FILE: /tmp/13135.nc
=====================
Using CF Checker Version 3.1.1
Checking against CF Version CF-1.6
Using Standard Name Table Version 73 (2020-06-23T11:50:17Z)
Using Area Type Table Version 10 (23 June 2020)
Using Standardized Region Name Table Version 4 (18 December 2018)

WARN: (2.6.1): No 'Conventions' attribute present

------------------
Checking variable: h2ovmr
------------------
ERROR: (3.3): Invalid standard_name: water
ERROR: (3.3): Invalid standard_name modifier: vapour
ERROR: (3.1): Invalid units: vmr

------------------
Checking variable: h2ovmr_std
------------------
ERROR: (3.1): Invalid units: vmr

------------------
Checking variable: lat
------------------
ERROR: (4): axis attribute inconsistent with coordinate type as deduced from units and/or positive

------------------
Checking variable: pressure
------------------
ERROR: (3.3): Invalid standard_name: pressure

------------------
Checking variable: time
------------------
WARN: (4.4): The unit 'year', defined by udunits to be exactly 365.242198781 days, should be used with caution. It is not a calendar year.
WARN: (4.4.1): Use of the calendar and/or month_lengths attributes is recommended for time coordinate variables
ERROR: (4.4): Invalid units and/or reference time

ERRORS detected: 7
WARNINGS given: 3
INFORMATION messages: 0
```

Figure 3-58: Report from CF-compliance checker for an early version of WV_cci CDR-3 that failed several checks.

Output of CF-Checker follows...

```
CHECKING NetCDF FILE: /tmp/27024.nc
=====================
Using CF Checker Version 3.1.1
Checking against CF Version CF-1.7
Using Standard Name Table Version 77 (2021-01-19T13:38:50Z)
Using Area Type Table Version 10 (23 June 2020)
Using Standardized Region Name Table Version 4 (18 December 2018)


------------------
Checking variable: lat
------------------

------------------
Checking variable: plev
------------------

------------------
Checking variable: time
------------------

------------------
Checking variable: lat_bnds
------------------

------------------
Checking variable: time_bnds
------------------

------------------
Checking variable: zmh2o
------------------

------------------
Checking variable: zmh2o_stdv
------------------

------------------
Checking variable: zmh2o_uncertainty
------------------

------------------
Checking variable: quality_flag
------------------

ERRORS detected: 0
WARNINGS given: 0
INFORMATION messages: 0
```

Figure 3-59: Report from CF-compliance checker for the final CDR-3 v3.0 file, now with previous problems resolved and of final CDR-3 v3.0 fully compliant.

```
netcdf ESACCI-WATERVAPOUR-L3S-LP-NP-MERGED-MLS-MIPAS-IMS-5deg-2010-2014-v3 {
dimensions:
        lon = 72 ;
        lat = 36 ;
        plev = 25 ;
        time = UNLIMITED ; // (60 currently)
        nv = 2 ;
variables:
        float lon(lon) ;
                lon:long_name = "Longitude" ;
                lon:units = "degrees_east" ;
                lon:standard_name = "longitude" ;
                lon:reference_datum = "geographical coordinates, WGS84 projection" ;
                lon:axis = "X" ;
                lon:bounds = "lon_bnds" ;
        float lon_bnds(lon, nv) ;
                lon_bnds:long_name = "Longitude cell boundaries" ;
                lon_bnds:valid_range = 0.f, 360.f ;
                lon_bnds:reference_datum = "geographical coordinates, WGS84 projection" ;
                lon_bnds:comment = "Contains the eastern and western boundaries of the grid cells." ;
        float lat(lat) ;
                lat:long_name = "Latitude" ;
                lat:units = "degrees_north" ;
                lat:standard_name = "latitude" ;
                lat:valid_range = -90.f, 90.f ;
                lat:reference_datum = "geographical coordinates, WGS84 projection" ;
                lat:axis = "Y" ;
                lat:bounds = "lat_bnds" ;
        float lat_bnds(lat, nv) ;
                lat_bnds:long_name = "Latitude cell boundaries" ;
                lat_bnds:valid_range = -90.f, 90.f ;
                lat_bnds:reference_datum = "geographical coordinates, WGS84 projection" ;
                lat_bnds:comment = "Contains the northern and southern boundaries of the grid cells." ;
        float plev(plev) ;
                plev:long_name = "Plev" ;
                plev:units = "hPa" ;
                plev:standard_name = "level" ;
                plev:valid_range = 30.f, 1000.f ;
                plev:axis = "Z" ;
        int time(time) ;
                time:long_name = "Product dataset time given as days since 1970-01-01" ;
                time:standard_name = "time" ;
                time:units = "days since 1970-01-01" ;
                time:calendar = "gregorian" ;
                time:axis = "T" ;
                time:bounds = "time_bnds" ;
        int time_bnds(time, nv) ;
                time_bnds:long_name = "Time cell boundaries" ;
                time_bnds:comment = "Contains the start and end times for the time period the data represent." ;
        float vmrh2o(time, plev, lat, lon) ;
                vmrh2o:long_name = "Vertical resolved water vapour profile" ;
                vmrh2o:units = "ppmv" ;
                vmrh2o:standard_name = "atmosphere water vapour profile" ;
        float vmrh2o_uncertainty(time, plev, lat, lon) ;
                vmrh2o_uncertainty:long_name = "Uncertainty of Water Vapor Volume Mixing Ratio" ;
                vmrh2o_uncertainty:units = "ppmv" ;


// global attributes:
                :_NCProperties = "version=1|netcdflibversion=4.6.1|hdf5libversion=1.10.2" ;
                :title = "Water Vapour CCI vertical resolved profile of Water Vapour Product" ;
                :institution = "University of Reading" ;
                :source = "MLS L2 v5; MIPAS IMK v7; IMS L2 v2.1" ;
                :references = "XXXX" ;
                :tracking_id = "04b04d7b-21b6-41a5-9bb6-368bc3bbc1e5" ;
                :Conventions = "CF-1.7" ;
                :product_version = "v3" ;
                :format_version = "CCI Data Standards v2.2" ;
                :product_doi = "10.5285/48ecd3eb90434cc3a843544065c69d47" ;
                :summary = "Water Vapour CCI VRWV Dataset 4 (2010-2014)" ;
                :keywords = "EARTH SCIENCE > ATMOSPHERE > ATMOSPHERIC WATER VAPOR > WATER VAPOR" ;
                :id = "ESACCI-WATERVAPOUR-L3S-LP-NP-MERGED-MLS-MIPAS-IMS-5deg-2010-2014-v3.nc" ;
                :keywords_vocabulary = "GCMD Science Keywords, Version 8.1" ;
                :cdm_data_type = "grid" ;
                :comment = "These data were produced in the frame of the Water Vapour ECV (Water_Vapour_cci) of the ESA Climate Change Initiative Extension (CCI+) Phase 1." ;
                :creator_name = "University of Reading Department of Meteorology" ;
                :creator_url = "https://www.reading.ac.uk/met/" ;
                :project = "Climate Change Initiative - European Space Agency" ;
                :geospatial_lat_min = "-90.0" ;
                :geospatial_lat_max = "90.0" ;
                :geospatial_lon_min = "0.0" ;
                :geospatial_lon_max = "360.0" ;
                :geospatial_vertical_min = "30 hPa" ;
                :geospatial_vertical_max = "1000 hPa" ;
                :time_coverage_duration = "P1M" ;
                :time_coverage_resolution = "P1M" ;
                :time_coverage_start = "2010-01-01 00:00:00 UTC" ;
                :time_coverage_end = "2014-12-31 23:59:59 UTC" ;
                :standard_name_vocabulary = "NetCDF Climate and Forecast (CF) Metadata Convention version 67" ;
                :license = "ESA CCI Data Policy: free and open access." ;
                :platform = "NASA\'s EOS Aura; ENVISAT;Eumetsat Metop" ;
                :sensor = "MLS; MIPAS; IASI/MHS/AMSU" ;
                :spatial_resolution = "556km at Equator" ;
                :geospatial_lat_units = "degrees_north" ;
                :geospatial_lon_units = "degrees_east" ;
                :geospatial_lat_resolution = "5" ;
                :geospatial_lon_resolution = "5" ;
                :key_variables = "vmrh2o" ;
```

Figure 3-60: ncdump example to check content of final CDR-4 v3.0 data file is consistent with CCI information. CDR-3 v3.0 data file exhibits the same information content.

APPENDIX 1: REFERENCES

[1]: Space engineering – Verification, ECSS-E-ST-10-02C, ESA ESTEC, Noordwijk, The Netherlands, 6 Mar 2009.

[2]: ESA CCI Water Vapour: ESA CCI Water Vapour: Product Specification Document. M. Schröder, M. Hegglin, H. Ye, U. Falk, O.Danne, J. Fischer, A. Laeng, R. Siddans, C. Sioris, G. Stiller, and K. Walker. Issue 3.2, 27 March 2021.

[3]: ESA CCI Water Vapour: System Specification Document. O.Danne, M. Hegglin and H. Ye, Issue 3.0, 19 August 2021.

[4]: ESA CCI Water Vapour: Technical Proposal. M. Hegglin and the WV ECV Consortium. Ref: KPT90658, 27 October 2017.

[5]: Berrisford, P, Dee, D.P., Poli, P, Brugge, R, Fielding, M, Fuentes, M, Kållberg, P.W., Kobayashi, S, Uppala, S, and A. Simmons, 2011: The ERA-Interim archive Version 2.0. ERA Report Series, Document Number 1, ECMWF, Reading, UK.

[6]: Ackerman S. A., and R. Frey, 2015: MODIS Atmosphere L2 Cloud Mask Product (35_L2). NASA MODIS Adaptive Processing System, Goddard Space Flight Center. Dataset DOI: http://dx.doi.org/10.5067/MODIS/MOD35_L2.006.

[7]: Cal/Val and User Services – Calvalus Technical Specification, ESA SME-LET AoO 2009, Brockmann Consult. Issue 1.2, 21 March 2011.

[8]: CM SAF: Ocean Surface Fluxes and Atmospheric Parameters. EUMETSAT CM SAF Climate Monitoring, April 2019.
https://www.cmsaf.eu/EN/Overview/OurProducts/Hoaps/Hoaps_node.html

[9]: ESA CCI Water Vapour: Product User Guide. O. Danne, M. Hegglin, H. Ye, M. Schröder, R. Preusker, J. Fischer, C. Brockmann, Issue 2.0, 11 August 2021.

[10]: Sayer, A.M., Hsu, A.C., and C. Bettenhausen: Implications of MODIS bowtie distortion on aerosol optical depth retrievals, and techniques for mitigation. Atmos. Meas. Tech. Discuss., 8, 8727–8752, 2015. doi:10.5194/amtd-8-8727-2015.

[11]: The ESA SICE project. J.E. Box et al. http://snow.geus.dk/index.php/about/

[12]: Eaton, B., Gregory, J., Drach, B., Taylor, K., Hankin, S., Blower, J., Caron, J., Signell, R., Bentley, P., Rappa, G., Höck, H., Pamment, A., Juckes, M., and M. Raspaud, 2017: NetCDF Climate and Forecast (CF) Metadata Conventions. Version 1.7. http://cfconventions.org/

[13]: NERC CF-Convention Compliance Checker for NetCDF Format. Version 3.1.0.
http://pumatest.nerc.ac.uk/cgi-bin/cf-checker-3.1.0.pl

[14]: ESA Climate Office: CCI Data Standards. Reference CCI-PRGM-EOPS-TN-13-0009, Issue 2.1, 2 August 2019.

[15]: ESA CCI Water Vapour: Data Access Requirement Document. M. Schröder, M. Hegglin, U. Falk, H. Brogniez, J. Fischer, S. Hubert, A. Laeng, R. Siddans, C. Sioris, G. Stiller. T. Trent, K. Walker, O. Danne, , H. Ye, Issue 3.2, 27 July 2021.

APPENDIX 2: GLOSSARY

Term	Definition
ACE-FTS	Atmospheric Chemistry Experiment Fourier Transform Spectrometer
ACE-MAESTRO	Atmospheric Chemistry Experiment Measurements of Aerosol Extinction in the Stratosphere and Troposphere Retrieved by Occultation
ATBD	Algorithm Theoretical Baseline Document
BC	Brockmann Consult
Calvalus	Cal/Val and User Services
CCI	Climat Change Initiative
CCM	Chemistry-Climate Model
CDR	Climate Data Records
CMAM	Canadian Middle Atmosphere Model
CEDA	Centre for Environmental Data Analysis
CM SAF	Satellite Application Facility on Climate Monitoring
DARD	Data Access Requirement Document
DMP	Derived Meteorological Product
DMSP	Defense Meteorological Satellite Program
DWD	Deutscher Wetterdienst (German Weather Service)
ECSAT	European Centre for Space Applications and Telecommunications
ECV	Essential Climate Variable
ERA	European Re-Analysis
ESA	European Space Agency
GMCD	Global Change Master Directory
GRUAN	GCOS Reference Upper-Air Network
HOAPS	Hamburg Ocean Atmosphere Parameters and Fluxes
IMS	Infrared Microwave Sounding
JASMIN	Joint Analysis System Meeting Infrastructure
KO	Kick-off
MD5	Message Digest (version 5)
MERIS	Medium Resolution Imaging Spectrometer

Term	**Definition**
MIPAS	Michelson Interferometer for Passive Atmospheric Sounding
MLS	Microwave Limb Sounder
MODIS	Moderate Resolution Imaging Spectroradiometer
NEODC	NERC Earth Observation Data Centre
NERC	National Environment Research Council
NetCDF	Network Common Data Form
OLCI	Ocean and Land Colour Instrument
PSD	Product Specification Document
PVIR	Product Validation and Intercomparison Report
RACC	Reading Academic Computing Cluster
RR	Reduced Resolution
SE	Spectral Earth
SMR	Submillimeter wave Radiometer
SNAP	Sentinel Application Platform
SPARC	Stratosphere-troposphere Processes And their Role in Climate
SSMI	Special Sensor Microwave Imager
SSMIS	Special Sensor Microwave Imager/Sounder
TCWV	Total Column of Water Vapour
UoR	University of Reading
UTLS	Upper Troposphere and Lower Stratosphere
VRWV	Vertically Resolved Water Vapour

APPENDIX 3: PROCESSING LOGS AND REPORTS

MERIS Preprocessing on Calvalus: Report file (just 4 of 62 entries shown):

```
template.py l2-idepix-MERIS.xml sensor MERIS minDate 2011-07-01
maxDate 2011-07-31 year 2011 month 07 day 10 input
/calvalus/eodata/MER_RR__1P/r03/2011/07/10 erainterim
/calvalus/projects/wvcci/era-interim/meris/2011/07/10 output
/calvalus/projects/wvcci/idepix/meris/2011/07/10 l1b l2_idepix-
meris-2011-07-10

template.py l2-idepix-MERIS.xml sensor MERIS minDate 2011-07-01
maxDate 2011-07-31 year 2011 month 07 day 13 input
/calvalus/eodata/MER_RR__1P/r03/2011/07/13 erainterim
/calvalus/projects/wvcci/era-interim/meris/2011/07/13 output
/calvalus/projects/wvcci/idepix/meris/2011/07/13 l1b l2_idepix-
meris-2011-07-13

template.py l2-idepix-erainterim-MERIS.xml sensor MERIS minDate
2011-07-01 maxDate 2011-07-31 year 2011 month 07 day 10 idepix
/calvalus/projects/wvcci/idepix/meris/2011/07/10 input
/calvalus/projects/wvcci/era-interim/meris/2011/07/10 output
/calvalus/projects/wvcci/idepix-era-interim/meris/2011/07/10
l2_idepix-meris-2011-07-10 l2_idepix-erainterim-meris-2011-07-10

template.py l2-idepix-erainterim-MERIS.xml sensor MERIS minDate
2011-07-01 maxDate 2011-07-31 year 2011 month 07 day 03 idepix
/calvalus/projects/wvcci/idepix/meris/2011/07/03 input
/calvalus/projects/wvcci/era-interim/meris/2011/07/03 output
/calvalus/projects/wvcci/idepix-era-interim/meris/2011/07/03
l2_idepix-meris-2011-07-03 l2_idepix-erainterim-meris-2011-07-03
```

MERIS Preprocessing on Calvalus: Part of log file of Pixel Classification job:

```
SLF4J: Class path contains multiple SLF4J bindings.

SLF4J: Found binding in [jar:file:/opt/hadoop-
2.7.3/share/hadoop/common/lib/slf4j-log4j12-
1.7.10.jar!/org/slf4j/impl/StaticLoggerBinder.class]

SLF4J: Found binding in
[jar:file:/ssd1/yarn/local/usercache/olaf/filecache/614/snap-
all.jar!/org/slf4j/impl/StaticLoggerBinder.class]

SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for
an explanation.

SLF4J: Actual binding is of type
[org.slf4j.impl.Log4jLoggerFactory]

2019-04-29 15:30:03,572 INFO com.bc.calvalus:
PendingJobAttemptsPath
hdfs://calvalus/calvalus/projects/wvcci/idepix/meris/2011/08/20/_t
emporary_job_1553782106366_32444

2019-04-29 15:30:03,576 INFO com.bc.calvalus:
PendingTaskAttemptPath
hdfs://calvalus/calvalus/projects/wvcci/idepix/meris/2011/08/20/_t
```

```
emporary_job_1553782106366_32444/attempt_1553782106366_32444_m_000
002_0

2019-04-29 15:30:03,927 INFO com.bc.calvalus: Set 'snap.userdir',
'snap.home', 'snap.pythonModuleDir' to CWD:
/ssd1/yarn/local/usercache/olaf/appcache/application_1553782106366
_32444/container_1553782106366_32444_01_000008/.

2019-04-29 15:30:03,928 INFO com.bc.calvalus: Setting system
property: beam.reader.tileWidth=*

2019-04-29 15:30:03,929 INFO com.bc.calvalus: Setting system
property: snap.dataio.reader.tileWidth=*

2019-04-29 15:30:03,929 INFO com.bc.calvalus: Setting system
property: snap.dataio.reader.tileHeight=64

2019-04-29 15:30:03,929 INFO com.bc.calvalus: Setting system
property: beam.reader.tileHeight=16

2019-04-29 15:30:15,901 INFO
org.esa.snap.core.gpf.operators.tooladapter.ToolAdapterIO:
Initializing external tool adapters

2019-04-29 15:30:16,022 WARNING org.esa.snap: Not able to check
for new SNAP version. Local version could not be retrieved.

2019-04-29 15:30:16,023 INFO com.bc.calvalus: processing input
hdfs://calvalus/calvalus/eodata/MER_RR__1P/r03/2011/08/20/MER_RR__
1PRACR20110820_041347_000026253105_00378_49530_0000.N1 ...

2019-04-29 15:30:16,023 INFO com.bc.calvalus: testing whether
target product exists ...

2019-04-29 15:30:16,155 INFO com.bc.calvalus: target product does
not exist

2019-04-29 15:30:16,158 INFO com.bc.calvalus: computed
inputRectangle = null

2019-04-29 15:30:18,337 INFO com.bc.calvalus: Copying file to
local:
hdfs://calvalus/calvalus/eodata/MER_RR__1P/r03/2011/08/20/MER_RR__
1PRACR20110820_041347_000026253105_00378_49530_0000.N1 -->
./MER_RR__1PRACR20110820_041347_000026253105_00378_49530_0000.N1

2019-04-29 15:31:19,449 INFO com.bc.calvalus: Opened product width
= 1121 height = 14929

2019-04-29 15:31:19,449 INFO com.bc.calvalus: Tiling: width = 1121
height = 64

2019-04-29 15:31:19,449 INFO com.bc.calvalus: ProductReader:
org.esa.snap.dataio.envisat.EnvisatProductReader[input=./MER_RR__1
PRACR20110820_041347_000026253105_00378_49530_0000.N1]

2019-04-29 15:31:19,449 INFO com.bc.calvalus: ProductReaderPlugin:
org.esa.snap.dataio.envisat.EnvisatProductReaderPlugIn@76b305e1

2019-04-29 15:31:19,449 INFO com.bc.calvalus: GeoCoding:
org.esa.snap.core.datamodel.TiePointGeoCoding@29cf1c8e

2019-04-29 15:31:19,450 INFO com.bc.calvalus: computed
inputRectangle = null

2019-04-29 15:31:19,450 INFO com.bc.calvalus: computed
inputRectangle = null
```

```
INFO: org.esa.s3tbx.meris.l2auxdata.AuxDatabase: loading ancillary
resources...

2019-04-29 15:31:20,682 INFO
org.esa.s3tbx.meris.l2auxdata.AuxDatabase: loading ancillary
resources...

WARNING: org.esa.s3tbx.meris.l2auxdata.AuxDatabase: missing
resource (might be OK): index_f.txa

2019-04-29 15:31:20,696 WARNING
org.esa.s3tbx.meris.l2auxdata.AuxDatabase: missing resource (might
be OK): index_f.txa

WARNING: org.esa.s3tbx.meris.l2auxdata.AuxDatabase: missing
resource (might be OK): index_g.txa

2019-04-29 15:31:20,696 WARNING
org.esa.s3tbx.meris.l2auxdata.AuxDatabase: missing resource (might
be OK): index_g.txa

WARNING: org.esa.s3tbx.meris.l2auxdata.AuxDatabase: missing
resource (might be OK): index_i.txa

2019-04-29 15:31:20,697 WARNING
org.esa.s3tbx.meris.l2auxdata.AuxDatabase: missing resource (might
be OK): index_i.txa

WARNING: org.esa.s3tbx.meris.l2auxdata.AuxDatabase: missing
resource (might be OK): index_j.txa

2019-04-29 15:31:20,698 WARNING
org.esa.s3tbx.meris.l2auxdata.AuxDatabase: missing resource (might
be OK): index_j.txa

WARNING: org.esa.s3tbx.meris.l2auxdata.AuxDatabase: missing
resource (might be OK): index_k.txa

2019-04-29 15:31:20,698 WARNING
org.esa.s3tbx.meris.l2auxdata.AuxDatabase: missing resource (might
be OK): index_k.txa

WARNING: org.esa.s3tbx.meris.l2auxdata.AuxDatabase: missing
resource (might be OK): index_m.txa

2019-04-29 15:31:20,699 WARNING
org.esa.s3tbx.meris.l2auxdata.AuxDatabase: missing resource (might
be OK): index_m.txa

WARNING: org.esa.s3tbx.meris.l2auxdata.AuxDatabase: missing
resource (might be OK): index_y.txa

2019-04-29 15:31:20,706 WARNING
org.esa.s3tbx.meris.l2auxdata.AuxDatabase: missing resource (might
be OK): index_y.txa

WARNING: org.esa.s3tbx.meris.l2auxdata.AuxDatabase: missing
resource (might be OK): index_z.txa

2019-04-29 15:31:20,706 WARNING
org.esa.s3tbx.meris.l2auxdata.AuxDatabase: missing resource (might
be OK): index_z.txa

INFO: org.esa.s3tbx.meris.l2auxdata.AuxDatabase: ancillary
resources loaded
```

```
2019-04-29 15:31:20,707 INFO
org.esa.s3tbx.meris.l2auxdata.AuxDatabase: ancillary resources
loaded

INFO: org.esa.s3tbx.meris.l2auxdata.AuxFile: opening ancillary
database file
'/ssd1/yarn/local/usercache/olaf/appcache/application_155378210636
6_32444/container_1553782106366_32444_01_000008/./auxdata/meris_l2
/lv2conf/lv2conf.41.08.prd'

2019-04-29 15:31:20,707 INFO
org.esa.s3tbx.meris.l2auxdata.AuxDatabase: opening ancillary
database file
'/ssd1/yarn/local/usercache/olaf/appcache/application_155378210636
6_32444/container_1553782106366_32444_01_000008/./auxdata/meris_l2
/lv2conf/lv2conf.41.08.prd'

WARNING: org.esa.s3tbx.meris.l2auxdata.AuxFile: file
'lv2conf.41.08.prd': variable 'O112': data conversion required
from storage data type (ascii) to memory data type (uint32)

2019-04-29 15:31:20,707 WARNING
org.esa.s3tbx.meris.l2auxdata.AuxDatabase: file
'lv2conf.41.08.prd': variable 'O112': data conversion required
from storage data type (ascii) to memory data type (uint32)

INFO: org.esa.s3tbx.meris.l2auxdata.AuxFile: opening ancillary
database file
'/ssd1/yarn/local/usercache/olaf/appcache/application_155378210636
6_32444/container_1553782106366_32444_01_000008/./auxdata/meris_l2
/atmosphere/atmosphere.32.04.prd'

2019-04-29 15:31:20,708 INFO
org.esa.s3tbx.meris.l2auxdata.AuxDatabase: opening ancillary
database file
'/ssd1/yarn/local/usercache/olaf/appcache/application_155378210636
6_32444/container_1553782106366_32444_01_000008/./auxdata/meris_l2
/atmosphere/atmosphere.32.04.prd'

WARNING: org.esa.s3tbx.meris.l2auxdata.AuxFile: file
'atmosphere.32.04.prd': variable 'P11Y': data conversion required
from storage data type (ascii) to memory data type (uint32)

2019-04-29 15:31:20,708 WARNING
org.esa.s3tbx.meris.l2auxdata.AuxDatabase: file
'atmosphere.32.04.prd': variable 'P11Y': data conversion required
from storage data type (ascii) to memory data type (uint32)

INFO: org.esa.s3tbx.meris.l2auxdata.AuxFile: opening ancillary
database file
'/ssd1/yarn/local/usercache/olaf/appcache/application_155378210636
6_32444/container_1553782106366_32444_01_000008/./auxdata/meris_l2
/landaero/landaero.42.02.prd'

2019-04-29 15:31:20,708 INFO
org.esa.s3tbx.meris.l2auxdata.AuxDatabase: opening ancillary
database file
'/ssd1/yarn/local/usercache/olaf/appcache/application_155378210636
6_32444/container_1553782106366_32444_01_000008/./auxdata/meris_l2
/landaero/landaero.42.02.prd'
```

```
WARNING: org.esa.s3tbx.meris.l2auxdata.AuxFile: file
'landaero.42.02.prd': variable 'S10M': data conversion required
from storage data type (ascii) to memory data type (uint32)

2019-04-29 15:31:20,708 WARNING
org.esa.s3tbx.meris.l2auxdata.AuxDatabase: file
'landaero.42.02.prd': variable 'S10M': data conversion required
from storage data type (ascii) to memory data type (uint32)

WARNING: org.esa.s3tbx.meris.l2auxdata.AuxFile: file
'landaero.42.02.prd': variable 'S10U': data conversion required
from storage data type (ascii) to memory data type (uint32)

...

... many more lines

...

2019-04-29 15:31:21,467 WARNING
org.esa.s3tbx.meris.l2auxdata.AuxDatabase: file 'cloud.40.03.prd':
variable 'V200': data conversion required from storage data type
(int32) to memory data type (float64)

WARNING: org.esa.s3tbx.meris.l2auxdata.AuxFile: file
'cloud.40.03.prd': variable 'V201': data conversion required from
storage data type (int32) to memory data type (float64)

2019-04-29 15:31:21,468 WARNING
org.esa.s3tbx.meris.l2auxdata.AuxDatabase: file 'cloud.40.03.prd':
variable 'V201': data conversion required from storage data type
(int32) to memory data type (float64)

WARNING: org.esa.s3tbx.meris.l2auxdata.AuxFile: file
'cloud.40.03.prd': variable 'V206': data conversion required from
storage data type (float64) to memory data type (float32)

2019-04-29 15:31:21,468 WARNING
org.esa.s3tbx.meris.l2auxdata.AuxDatabase: file 'cloud.40.03.prd':
variable 'V206': data conversion required from storage data type
(float64) to memory data type (float32)

INFO: org.esa.s3tbx.meris.l2auxdata.AuxFile: file
'cloud.40.03.prd': variable 'V300': about to allocate "huge" data
buffer (24 M)

2019-04-29 15:31:21,468 INFO
org.esa.s3tbx.meris.l2auxdata.AuxDatabase: file 'cloud.40.03.prd':
variable 'V300': about to allocate "huge" data buffer (24 M)

WARNING: org.esa.s3tbx.meris.l2auxdata.AuxFile: file
'cloud.40.03.prd': variable 'V300': data conversion required from
storage data type (uint8) to memory data type (float32)

2019-04-29 15:31:21,472 WARNING
org.esa.s3tbx.meris.l2auxdata.AuxDatabase: file 'cloud.40.03.prd':
variable 'V300': data conversion required from storage data type
(uint8) to memory data type (float32)

INFO: org.esa.s3tbx.meris.l2auxdata.AuxFile: closed ancillary
database file
'/ssd1/yarn/local/usercache/olaf/appcache/application_155378210636
6_32444/container_1553782106366_32444_01_000008/./auxdata/meris_l2
/cloud/cloud.40.03.prd'
```

```
2019-04-29 15:31:21,512 INFO
org.esa.s3tbx.meris.l2auxdata.AuxDatabase: closed ancillary
database file
'/ssd1/yarn/local/usercache/olaf/appcache/application_155378210636
6_32444/container_1553782106366_32444_01_000008/./auxdata/meris_l2
/cloud/cloud.40.03.prd'

2019-04-29 15:31:22,370 INFO com.bc.calvalus: Processed product
width = 1121 height = 14929

2019-04-29 15:31:22,370 INFO com.bc.calvalus:
attempt_1553782106366_32444_m_000002_0 target product created

2019-04-29 15:31:23,474 INFO com.bc.calvalus:  written header

2019-04-29 15:31:23,474 INFO com.bc.calvalus: Writing bands of the
same size
```

MERIS Preprocessing on Calvalus: Part of log file of ERA Interim ingestion job:

```
Log Type: stderr


Log Upload Time: Mon Apr 29 15:38:36 +0200 2019


Log Length: 12113


Showing 4096 bytes of 12113 total. Click here for the full log.


nap.core.datamodel.PixelGeoCoding2@61096fbd

2019-04-29 15:38:01,554 INFO com.bc.calvalus: input subsetting of
split hdfs://calvalus/calvalus/projects/wvcci/era-
interim/meris/2011/08/11/MER_RR__1PRACR20110811_012222_00002626310
5_00247_49399_0000_era-interim.nc

2019-04-29 15:38:01,554 INFO com.bc.calvalus: computed
inputRectangle = null

2019-04-29 15:38:01,563 SEVERE org.esa.snap: Error attempting to
read
PathConfiguration{path=hdfs://calvalus/calvalus/projects/wvcci/ide
pix/meris/2011/08/11/L2_of_MER_RR__1PRACR20110811_012222_000026263
105_00247_49399_0000.seq} with plugin reader
org.esa.s2tbx.dataio.spot6.Spot6ProductReaderPlugin@698ac187: null

2019-04-29 15:38:01,563 SEVERE org.esa.snap: Error attempting to
read
PathConfiguration{path=hdfs://calvalus/calvalus/projects/wvcci/ide
pix/meris/2011/08/11/L2_of_MER_RR__1PRACR20110811_012222_000026263
105_00247_49399_0000.seq} with plugin reader
org.esa.s2tbx.dataio.spot.SpotDimapProductReaderPlugin@334b392d:
null

2019-04-29 15:38:01,563 SEVERE org.esa.snap: Error attempting to
read
PathConfiguration{path=hdfs://calvalus/calvalus/projects/wvcci/ide
pix/meris/2011/08/11/L2_of_MER_RR__1PRACR20110811_012222_000026263
105_00247_49399_0000.seq} with plugin reader
```

```
org.esa.s2tbx.dataio.rapideye.RapidEyeL1ReaderPlugin@7c9372ed:
null

2019-04-29 15:38:02,842 INFO com.bc.calvalus: Opened product width
= 1121 height = 14929

2019-04-29 15:38:02,842 INFO com.bc.calvalus: Tiling: width = 1121
height = 64

2019-04-29 15:38:02,842 INFO com.bc.calvalus: ProductReader:
com.bc.calvalus.processing.beam.StreamingProductReader[input=PathC
onfiguration{path=hdfs://calvalus/calvalus/projects/wvcci/idepix/m
eris/2011/08/11/L2_of_MER_RR__1PRACR20110811_012222_000026263105_0
0247_49399_0000.seq}]

2019-04-29 15:38:02,842 INFO com.bc.calvalus: ProductReaderPlugin:
com.bc.calvalus.processing.beam.StreamingProductPlugin@b841713

2019-04-29 15:38:02,842 INFO com.bc.calvalus: GeoCoding:
org.esa.snap.core.datamodel.TiePointGeoCoding@36a66b54

2019-04-29 15:38:03,952 INFO com.bc.calvalus: Processed product
width = 1121 height = 14929

2019-04-29 15:38:03,952 INFO com.bc.calvalus:
attempt_1553782106366_32459_m_000000_0 target product created

2019-04-29 15:38:04,705 INFO com.bc.calvalus:  written header

2019-04-29 15:38:04,705 INFO com.bc.calvalus: Writing bands of the
same size

2019-04-29 15:38:20,507 INFO com.bc.calvalus: Merging data from
FileStatus{path=hdfs://calvalus/calvalus/projects/wvcci/idepix-
era-
interim/meris/2011/08/11/_temporary_job_1553782106366_32459/attemp
t_1553782106366_32459_m_000000_0; isDirectory=true;
modification_time=1556545099354; access_time=0; owner=olaf;
group=bc; permission=rwxrwxr-x; isSymlink=false} to
hdfs://calvalus/calvalus/projects/wvcci/idepix-era-
interim/meris/2011/08/11

2019-04-29 15:38:20,515 INFO com.bc.calvalus: Merging data from
FileStatus{path=hdfs://calvalus/calvalus/projects/wvcci/idepix-
era-
interim/meris/2011/08/11/_temporary_job_1553782106366_32459/attemp
t_1553782106366_32459_m_000000_0/L2_of_MER_RR__1PRACR20110811_0122
22_000026263105_00247_49399_0000_era-interim.seq;
isDirectory=false; length=131584122; replication=1;
blocksize=2147483136; modification_time=1556545099325;
access_time=1556545085681; owner=olaf; group=bc; permission=rw-rw-
r--; isSymlink=false} to
hdfs://calvalus/calvalus/projects/wvcci/idepix-era-
interim/meris/2011/08/11/L2_of_MER_RR__1PRACR20110811_012222_00002
6263105_00247_49399_0000_era-interim.seq

2019-04-29 15:38:20,524 INFO com.bc.calvalus: Merging data from
FileStatus{path=hdfs://calvalus/calvalus/projects/wvcci/idepix-
era-
interim/meris/2011/08/11/_temporary_job_1553782106366_32459/attemp
t_1553782106366_32459_m_000000_0/L2_of_MER_RR__1PRACR20110811_0122
22_000026263105_00247_49399_0000_era-interim.seq.index;
```

isDirectory=false; length=73201; replication=1; blocksize=2147483136; modification_time=1556545101839; access_time=1556545099354; owner=olaf; group=bc; permission=rw-rw-r--; isSymlink=false} to hdfs://calvalus/calvalus/projects/wvcci/idepix-era-interim/meris/2011/08/11/L2_of_MER_RR__1PRACR20110811_012222_000026263105_00247_49399_0000_era-interim.seq.index

Log Type: stdout

Log Upload Time: Mon Apr 29 15:38:36 +0200 2019

Log Length: 1732

graphAsText =

```
&lt;graph id="wvcci-idepix-erainterim"&gt;
    &lt;version&gt;1.0&lt;/version&gt;

    &lt;header&gt;
      &lt;target refid="idepix-erainterim" /&gt;
      &lt;source name="erainterim"&gt;hdfs://calvalus/calvalus/projects/wvcci/era-interim/meris/2011/08/11/MER_RR__1PRACR20110811_012222_000026263105_00247_49399_0000_era-interim.nc&lt;/source&gt;
      &lt;source name="idepix"&gt;hdfs://calvalus/calvalus/projects/wvcci/idepix/meris/2011/08/11/L2_of_MER_RR__1PRACR20110811_012222_000026263105_00247_49399_0000.seq&lt;/source&gt;
    &lt;/header&gt;

    &lt;node id="idepix-erainterim"&gt;
        &lt;operator&gt;ESACCI.MergeIdepixEraInterim&lt;/operator&gt;
        &lt;sources&gt;
            &lt;eraInterimProduct&gt;erainterim&lt;/eraInterimProduct&gt;
            &lt;idepixProduct&gt;idepix&lt;/idepixProduct&gt;
        &lt;/sources&gt;
        &lt;parameters&gt;
            &lt;sensor&gt;MERIS&lt;/sensor&gt;
        &lt;/parameters&gt;
    &lt;/node&gt;
```

```
</graph>

fileLocation =
./MER_RR__1PRACR20110811_012222_000026263105_00247_49399_0000_era-
interim.nc

Node = idepix-erainterim Operator = MergeIdepixEraInterimOp

Product: mergedClassif

Bands:

org.esa.snap.core.datamodel.Band[pixel_classif_flags,int16,1121,14
929,-1,0.0,0.0,0.0]

org.esa.snap.core.datamodel.Band[reflectance_13,int16,1121,14929,1
2,864.87604,20.047,933.7583]

org.esa.snap.core.datamodel.Band[reflectance_14,int16,1121,14929,1
3,884.94403,10.018001,905.94464]

org.esa.snap.core.datamodel.Band[reflectance_15,int16,1121,14929,1
4,900.00006,10.02,872.44965]

     org.esa.snap.core.datamodel.Band[l1_flags,uint8,1121,14929,-
1,0.0,0.0,0.0]

     org.esa.snap.core.datamodel.Band[t2m,int16,1121,14929,-
1,0.0,0.0,0.0]

     org.esa.snap.core.datamodel.Band[msl,int16,1121,14929,-
1,0.0,0.0,0.0]

     org.esa.snap.core.datamodel.Band[tcwv,int16,1121,14929,-
1,0.0,0.0,0.0]



Log Type: syslog

Log Upload Time: Mon Apr 29 15:38:36 +0200 2019

Log Length: 2556

2019-04-29 15:37:25,790 WARN [main]
org.apache.hadoop.metrics2.impl.MetricsConfig: Cannot locate
configuration: tried hadoop-metrics2-maptask.properties,hadoop-
metrics2.properties

2019-04-29 15:37:25,890 INFO [main]
org.apache.hadoop.metrics2.impl.MetricsSystemImpl: Scheduled
snapshot period at 10 second(s).
```

```
2019-04-29 15:37:25,890 INFO [main]
org.apache.hadoop.metrics2.impl.MetricsSystemImpl: MapTask metrics
system started

2019-04-29 15:37:25,900 INFO [main]
org.apache.hadoop.mapred.YarnChild: Executing with tokens:

2019-04-29 15:37:26,123 INFO [main]
org.apache.hadoop.mapred.YarnChild: Kind: mapreduce.job, Service:
job_1553782106366_32459, Ident:
(org.apache.hadoop.mapreduce.security.token.JobTokenIdentifier@5d0
a1059)

2019-04-29 15:37:26,189 INFO [main]
org.apache.hadoop.mapred.YarnChild: Sleeping for 0ms before
retrying again. Got null now.

2019-04-29 15:37:26,719 INFO [main]
org.apache.hadoop.mapred.YarnChild: mapreduce.cluster.local.dir
for child:
/ssd1/yarn/local/usercache/olaf/appcache/application_1553782106366
_32459

2019-04-29 15:37:27,134 INFO [main]
org.apache.hadoop.conf.Configuration.deprecation: session.id is
deprecated. Instead, use dfs.metrics.session-id

2019-04-29 15:37:27,920 INFO [main]
org.apache.hadoop.mapreduce.lib.output.FileOutputCommitter: File
Output Committer Algorithm version is 1

2019-04-29 15:37:28,014 INFO [main] org.apache.hadoop.mapred.Task:
Using ResourceCalculatorProcessTree : [ ]

2019-04-29 15:37:28,449 INFO [main]
org.apache.hadoop.mapred.MapTask: Processing split:
hdfs://calvalus/calvalus/projects/wvcci/era-
interim/meris/2011/08/11/MER_RR__1PRACR20110811_012222_00002626310
5_00247_49399_0000_era-interim.nc:0+836811540

2019-04-29 15:38:02,789 INFO [main]
org.apache.hadoop.io.compress.zlib.ZlibFactory: Successfully
loaded & initialized native-zlib library

2019-04-29 15:38:02,790 INFO [main]
org.apache.hadoop.io.compress.CodecPool: Got brand-new
decompressor [.deflate]

2019-04-29 15:38:03,772 INFO [main]
hsqldb.db.HSQLDB4AD417742A.ENGINE: dataFileCache open start

2019-04-29 15:38:18,006 INFO [main]
org.apache.hadoop.io.compress.CodecPool: Got brand-new compressor
[.deflate]

2019-04-29 15:38:20,489 INFO [main] org.apache.hadoop.mapred.Task:
Task:attempt_1553782106366_32459_m_000000_0 is done. And is in the
process of committing

2019-04-29 15:38:20,502 INFO [main] org.apache.hadoop.mapred.Task:
Task attempt_1553782106366_32459_m_000000_0 is allowed to commit
now

2019-04-29 15:38:20,542 INFO [main] org.apache.hadoop.mapred.Task:
Task 'attempt_1553782106366_32459_m_000000_0' done.
```

APPENDIX 4: NETCDF ‘NCDUMP’ EXAMPLE

```
netcdf ESACCI-WATERVAPOUR-L3C-TCWV-olci-05deg-20160715-fv2.2 {
      dimensions:
              lat = 360 ;
              lon = 720 ;
              time = UNLIMITED ; // (1 currently)
              nv = 2 ;

      variables:
              int time(time) ;
                      time:long_name = "Product dataset time given as
                                  days since 1970-01-01" ;
                      time:standard_name = "time" ;
                      time:units = "days since 1970-01-01" ;
                      time:calendar = "gregorian" ;
                      time:axis = "T" ;
                      time:bounds = "time_bnds" ;
              int time_bnds(time, nv) ;
                      time_bnds:long_name = "Time cell boundaries" ;
                      time_bnds:comment = "Contains the start and end
                      times for the time period the data represent." ;
              float lat_bnds(lat, nv) ;
                      lat_bnds:long_name = "Latitude cell boundaries" ;
                      lat_bnds:valid_range = -90.f, 90.f ;
                      lat_bnds:reference_datum = "geographical
                              coordinates, WGS84 projection" ;
                      lat_bnds:comment = "Contains the northern and
                              southern boundaries of the grid cells." ;
              float lon_bnds(lon, nv) ;
                      lon_bnds:long_name = "Longitude cell boundaries" ;
                      lon_bnds:valid_range = -180.f, 180.f ;
                      lon_bnds:reference_datum = "geographical
                                  coordinates, WGS84 projection" ;
                      lon_bnds:comment = "Contains the eastern and
                               western boundaries of the grid cells." ;
              int num_obs(lat, lon) ;
                      num_obs:_FillValue = -1 ;
                      num_obs:coordinates = "lat lon" ;
```

```
        num_obs:long_name = "Number of Total Column of
                 Water Vapour retrievals contributing
                 to L3 grid cell" ;
        num_obs:units = " " ;
float tcwv(lat, lon) ;
        tcwv:_FillValue = NaNf ;
        string tcwv:coordinates = "lat lon" ;
        tcwv:long_name = "Total Column of Water" ;
        tcwv:units = "kg/m2" ;
        tcwv:standard_name =
                "atmosphere_water_vapor_content " ;
        tcwv:ancillary_variables = "stdv num_obs" ;
        tcwv:actual_range = 0.725221f, 550.1542f ;
        tcwv:valid_range = 0.f, 80.f ;
float stdv(lat, lon) ;
        stdv:_FillValue = NaNf ;
        string stdv:coordinates = "lat lon" ;
        stdv:long_name = "Standard deviation of
                       Total Column of Water Vapour" ;
        stdv:units = "kg/m2" ;
float tcwv_err(lat, lon) ;
        tcwv_err:_FillValue = NaNf ;
        string tcwv_err:coordinates = "lat lon" ;
        tcwv_err:long_name = "Average retrieval
                             uncertainty" ;
        tcwv_err:units = "kg/m2" ;
float tcwv_ran(lat, lon) ;
        tcwv_ran:_FillValue = NaNf ;
        string tcwv_ran:coordinates = "lat lon" ;
        tcwv_ran:long_name = "Propagated retrieval
                             uncertainty" ;
        tcwv_ran:units = "kg/m2" ;
double lat(lat) ;
        lat:long_name = "Latitude" ;
        lat:units = "degrees_north " ;
        lat:standard_name = "latitude" ;
        lat:valid_range = -90.f, 90.f ;
        lat:reference_datum = "geographical coordinates,
                              WGS84 projection" ;
        lat:axis = "Y" ;
        lat:bounds = "lat_bnds" ;
```

```
double lon(lon) ;
        lon:long_name = "Longitude" ;
        lon:units = "degrees_east" ;
        lon:standard_name = "longitude" ;
        lon:valid_range = -180.f, 180.f ;
        lon:reference_datum = "geographical coordinates,
                              WGS84 projection" ;
        lon:axis = "X" ;
        lon:bounds = "lon_bnds" ;
int crs ;
        string crs:wkt = "GEOGCS[\"WGS84(DD)\", \n
           DATUM[\"WGS84\", \n    SPHEROID[\"WGS84\",
           6378137.0, 298.257223563]], \n
           PRIMEM[\"Greenwich\", 0.0], \n
           UNIT[\"degree\", 0.017453292519943295], \n
           AXIS[\"Geodetic longitude\", EAST], \n
           AXIS[\"Geodetic latitude\", NORTH]]" ;
        string crs:i2m = "0.5,0.0,0.0,-0.5,-180.0,90.0" ;
        crs:long_name = "Coordinate Reference System " ;
        crs:comment = "A coordinate reference system (CRS)
                       defines how the georeferenced
                       spatial data relates to real
                     locations on the Earth\'s surface " ;
byte tcwv_quality_flag(lat, lon) ;
        tcwv_quality_flag:long_name = "Quality flag of
                          Total Column of Water Vapour" ;
        tcwv_quality_flag:units = " " ;
        tcwv_quality_flag:standard_name = "status_flag " ;
        tcwv_quality_flag:_FillValue = -128b ;
        tcwv_quality_flag:valid_range = 0b, 3b ;
        tcwv_quality_flag:flag_values = 0b, 1b, 2b, 3b ;
        tcwv_quality_flag:flag_meanings = "TCWV_OK
                              HIGH_COST_FUNCTION_1
                              HIGH_COST_FUNCTION_2
                              TCWV_INVALID" ;
byte surface_type_flag(lat, lon) ;
        surface_type_flag:long_name = "Surface type flag"
        surface_type_flag:units = " " ;
        surface_type_flag:standard_name = "status_flag " ;
        surface_type_flag:_FillValue = -128b ;
        surface_type_flag:valid_range = 0b, 6b ;
```

```
                surface_type_flag:flag_values = 0b, 1b, 2b, 3b,
                                                4b, 5b, 6b ;
                surface_type_flag:flag_meanings = "LAND OCEAN
                                   CLOUD_OVER_LAND SEA_ICE COAST
                                   PARTLY_CLOUDY_OVER_LAND
                                   PARTLY_SEA_ICE" ;


// global attributes:
                :title = "Water Vapour CCI Total Column of Water
                          Vapour Product" ;
                :institution = "Brockmann Consult GmbH;
                               EUMETSAT/CMSAF" ;
                :source = "MERIS RR L1B 3rd Reprocessing; MODIS
                           MOD021KM L1B; HOAPS-S version 4.0
                           released by CM SAF" ;
                :history = "python nc-compliance-py-process.py
                     l3_tcwv_olci_05deg_2016-07-15_2016-07-15.nc" ;
                :references = "WV_cci D2.2: ATBD Part 1 - MERIS-
                               MODIS-OLCI L2 Products, Issue 1.1,
                               3 April 2019; WV_cci D4.2: CRDP
                               Issue 1.0, 13 June 2019 " ;
                :tracking_id = "0b353c3e-738b-11ea-b0f5-
                                0cc47a9d1a69" ;
                :Conventions = "CF-1.7" ;
                :product_version = "2.2" ;
                :format_version = "CCI Data Standards v2.0" ;
                :summary = "Water Vapour CCI TCWV Dataset 1
                            (2010-2012)" ;
                :keywords = "EARTH SCIENCE > ATMOSPHERE >
                             ATMOSPHERIC WATER VAPOR > WATER
                             VAPOR,EARTH SCIENCE > ATMOSPHERE >
                             ATMOSPHERIC WATER VAPOR >
                             PRECIPITABLE WATER" ;
                :id = "ESACCI-WATERVAPOUR-L3C-TCWV-olci-05deg-
                       20160715-fv2.2.nc" ;
                :naming-authority = "brockmann-consult.de" ;
                :keywords-vocabulary = "GCMD Science Keywords,
                                        Version 8.1" ;
                :cdm_data_type = "grid" ;
                :comment = "These data were produced in the frame
                            of the Water Vapour ECV
                            (Water_Vapour_cci) of the ESA Climate
                            Change Initiative Extension (CCI+)
                            Phase 1. They include CM SAF products
                            over the ocean." ;
                :date_created = "2020-03-31 20:06:01 UTC" ;
```

```
                      :creator_name = "Brockmann Consult GmbH;
                                       EUMETSAT/CMSAF" ;
                      :creator_url = "www.brockmann-consult.de;
                                      http://www.cmsaf.eu" ;
                      :creator_email = "info@brockmann-consult.de;
                                        contact.cmsaf@dwd.de" ;
                      :project = "Climate Change Initiative - European
                                  Space Agency" ;
                      :geospatial_lat_min = "-90.0" ;
                      :geospatial_lat_max = "90.0" ;
                      :geospatial_lon_min = "-180.0" ;
                      :geospatial_lon_max = "180.0" ;
                      :geospatial_vertical_min = "0.0" ;
                      :geospatial_vertical_max = "0.0" ;
                      :time_coverage_duration = "P1D" ;
                      :time_coverage_resolution = "P1D" ;
                      :time_coverage_start = "20160715 00:00:00 UTC" ;
                      :time_coverage_end = "20160715 23:59:59 UTC" ;
                      :standard_name_vocabulary = "NetCDF Climate and
                                                   Forecast (CF) Metadata
                                                   Convention version 67" ;
                      :license = "ESA CCI Data Policy: free and open
                                  access. Products containing CM SAF
                                  data are made available under the CM
                                  SAF data policy." ;
                      :platform = "Envisat, Terra, DMSP-F16, DMSP-F17,
                                   DMSP-F18" ;
                      :sensor = "MERIS, MODIS, SSMIS" ;
                      :spatial_resolution = "56km at Equator" ;
                      :geospatial_lat_units = "degrees_north" ;
                      :geospatial_lon_units = "degrees_east" ;
                      :geospatial_lat_resolution = "0.5" ;
                      :geospatial_lon_resolution = "0.5" ;
                      :key_variables = "tcwv" ;
}
```

End of Document