D5_3 Product user guide – UTH CDRs from microwave sounders

Theresa Lang, Imke Hans, Martin Burgdorf and Stefan Bühler

University of Hamburg

1/29/2019

FIDUCEO has received funding from the European Union's Horizon 2020 Programme for Research and Innovation, under Grant Agreement no. 638822

1 Contents

2 Introduction

2.1 Scope

This document describes the Upper Tropospheric Humidity (UTH) CDR (version 1.2) data files uploaded to CEMS in February 2019. The released data record is based on the Microwave "easy" FCDR version 4.1 (see Microwave FCDR PUG). The instantaneous observations from the FCDR are used to derive a spatio-temporal averaged data record, which contains monthly mean UTH and brightness temperature mapped to a regular latitude/longitude grid covering the tropical region with a spatial resolution of 1° x 1°. It covers all mission years of SSMT2 on F11, F12, F14, F15, AMSU-B on NOAA15-17 and MHS missions (NOAA18, NOAA19, MetopA, Metop-B). This product user guide gives:

1. An overview of the specifications of the data record;
2. A description of the implementation of the retrieval processing chain;
3. Information on limitations of this current version of the data record;
4. Technical details on the format and on how to access the data.

2.2 Document Version Control

Version	Reason	Reviewer	Date of Issue
1.0	Initial version	RPhipps	19/8/2019

2.3 Applicable and Reference Documents

- FIDUCEO website, http://www.fiduceo.eu/ or https://research.reading.ac.uk/fiduceo/
- D2.4-c, Uncertainty report for Upper Tropospheric Humidity (UTH) CDR from HIRS and microwave sounders (available on website)
- CF-standards version 1.7, http://cfconventions.org/Data/cf-conventions/cf-conventions-1.7/cf-conventions.html
- Microwave FCDR PUG, Product user guide – Microwave FCDR release 4.1 (available on website)
- Buehler, S.A. and V.O. John (2005), A simple method to relate microwave radiances to upper tropospheric humidity, Journal of Geophysical Research, Vol. 110, D02110, doi:10.1029/2004JD005111
- Eresmaa, R. and A.P. McNally (2014), Diverse profile datasets from the ECMWF 137-level short-range forecasts, doi: 10.13140/2.1.4476.8963
- Soden, B.J. and F. P. Bretherton (1993), Upper Tropospheric Relative Humidity From the GOES 6.7 µm Channel: Method and Climatology for July 1987, Journal of Geophysical Research, Vol. 98, D9, 16669 – 16688, doi: 10.1029/93JD01283

2.4 Glossary

BT	Brightness Temperature
CDR	Climate Data Record
CEDA	Centre for Environmental Data Archiving
FCDR	Fundamental Climate Data Record
FIDUCEO	Fidelity and Uncertainty in Climate data records for Earth Observation
IWV	Integrated Water Vapour/ Water vapour column

MHS	Microwave Humidity Sounder
NOAA	National Oceanic and Atmospheric Administration
RH	Relative Humidity
UTH	Upper Tropospheric Humidity

3 CDR overview characteristics

General	CDR name	FIDUCEO UTH CDR from microwave sounders
	CDR reference	D5_3 v1_2
	CDR digital identifier(s)	10.5285/ 2083b33b5c3d4cf0acb9a49226789caa
	CDR description	Climate Data Record containing monthly averages of Upper Tropospheric Humidity (UTH) and brightness temperature.
	CDR type	Level 3 UTH CDR
	CDR period	Encompasses all years between 1994-2018
	CDR satellites	• F11 • F12 • F14 • F15 • NOAA-15 • NOAA-16 • NOAA-17 • NOAA-18 • NOAA-19 • MetOp-A • MetOp-B All the satellites are in a sun-synchronous near-polar orbit.
	CDR content	Monthly averaged upper tropospheric humidity and 183.31±1 GHz brightness temperatures on a regular latitude/ longitude grid covering the tropical region, with uncertainties propagated from Microwave FCDRS.
Instrument	Instrument name	Advanced Microwave Sounding Unit-B (AMSU-B), Microwave Humidity Sounder (MHS), Special Sensor Microwave Water Vapor Profiler (SSMT-2)
	Instrument description	AMSU-B, MHS and SSMT-2 are scanning radiometers. They scan the Earth in 90 Earth views (28 for SSMT-2) per scan line in five spectral channels. Only the 183.31±1 GHz channel is used to retrieve UTH. For every scan line, the instrument gets a new calibration based on 4 views of a warm target and 4 views of a cold target.
Data	Input data	• Input data are the Microwave “easy” FCDR version 4.1 • Parameters used to scale brightness temperature to UTH were obtained by regression using a set of training atmospheres created by Eresmaa and McNally (2014)

	Output data	CDR including monthly averages (from ascending and descending overpasses) of tropical upper tropospheric humidity and 183.31±1 GHz brightness temperatures as well as uncertainties split into independent, structured and common components Each UTH CDR file of this version contains: • Longitudes and Latitudes of grid cell centres and grid cell boundaries • Observation and satellite overpass counts for ascending and descending branches; • Earliest and latest time of day of pixel contribution to the monthly average, for ascending and descending branches; • Monthly average and standard deviation of 183.31±1 GHz brightness temperatures that were used to derive UTH (excluding cloudy pixels), for ascending and descending branches; • Monthly average of all 183.31±1 GHz brightness temperatures (including cloudy pixels), for ascending and descending branches; • Monthly average and standard deviation of UTH (derived from channel 3 brightness temperatures); • Independent, structured and common uncertainty for monthly averages of brightness temperatures and UTH, for ascending and descending branches; The dimensions of the variables in the NetCDF file are x and y, indicating east-west and north-south direction, respectively.
	Format	The data are provided in NetCDF4 format.
Access	CEDA	The data are hosted by CEDA. This version of the CDR will be in the format MW/SATELLITE/YEAR
	Delivery	Available through CEDA
Resolution	Horizontal	Latitude/Longitude grid has 1° x 1° resolution.
	Vertical	-
	Temporal	Monthly averages
Physical Content	CDR physical quantity	The core physical quantities are monthly averages of Upper Tropospheric Humidity and Planck Brightness Temperatures for each grid cell. Associated information stored in the same files is latitudes and longitudes as well as independent, structured and common uncertainties.
	CDR physical description	One UTH CDR file has a size of about 4.3 MB and contains two data fields for each of the core quantities, corresponding to monthly averages from ascending and descending satellite overpasses, respectively.

Uncertainty target	Accuracy	Metrologically traceable uncertainties provided for each grid cell average.
	Precision	Brightness temperatures and the corresponding uncertainties are stored with a precision of 0.01 K, UTH and the corresponding uncertainties are stored with a precision of 0.01 %.
	Stability	UTH and brightness temperatures from the different instruments agree within about 2 % and 0.5 K, respectively. The good agreement results in consistent time series without major jumps.
	Known problems	The uncertainties included in the UTH CDR only those propagated from the MW FCDR, i.e. those associated with the measurement process. There are additional sources of uncertainty at CDR level, which are not explicitly included in the CDR but the user should be aware of. An overview is given in Section 8.
Data record characteristics	For every quantity two monthly averages are provided. One is the average from ascending satellite overpasses, the other one is from descending overpasses. Uncertainties are split into three different classes: • The independent uncertainty describes the uncertainty in the brightness temperature due to purely random effects that generate a completely independent uncertainty from pixel to pixel. These independent effects are the noise on the Earth view counts and the random fluctuations of the angle of the Earth view. • The structured uncertainties encompass effects that have a correlation scale below the time/ space scales of one orbit. • The common uncertainties represent effects that have a correlation scale larger than one orbit. Usually the underlying parameters of the effect are constant over the mission.	

4 Description of AMSUB and MHS

SSMT-2, AMSU-B and MHS are a passive microwave radiometer with five channels measuring in the microwave spectrum. The 183.31±1 GHz channel is used to derive UTH. Note that this channel is named differently among the instruments: channel 2 for SSMT-2, channel 18 for AMSU-B and channel 3 for MHS.

SSMT-2

Note the different numbering of the channels compared to MHS.

Channel	Wavelength [GHz]	Notes
1	183.31±3.0	
2	183.31±1.0	
3	183.31±7.0	
4	91.655±1.250	
5	150.00±1.25	

AMSU-B

Note that the AMSU-B channels are counted including the AMSU-A channels (1-15).

Channel	Wavelength [GHz]	Notes
16	89.0±0.9	
17	150.0±0.9	
18	183.31±1.0	
19	183.31±3.0	
20	183.31±7.0	

MHS

Channel	Wavelength [GHz]	Notes
1	89.0	
2	157.0	
3	183.31±1.0	
4	183.31±3.0	
5	190.31	

The UTH CDR is based on FCDR data, which is available for the following satellite missions and time periods:

Instrument	Satellite	Start	End
SSMT-2	DMSP F11	1994-07-05	1995-04-02
SSMT-2	DMSP F12	1994-10-13	2001-01-08
SSMT-2	DMSP F14	1997-04-28	2005-01-10
SSMT-2	DMSP F15	2000-01-24	2005-01-02
AMSU-B	NOAA-15/K	1999-01-01	2011-03-28
AMSU-B	NOAA-16/L	2001-03-20	2014-04-30
AMSU-B	NOAA-17/M	2002-10-15	2013-04-10
MHS	NOAA-18/N	2005-08-30	2017-12-31
MHS	NOAA-19/N’	2009-11-01	2017-12-31

MHS	MetOp-A	2007-06-01	2017-12-31
MHS	MetOp-B	2013-01-29	2017-12-31

Note that the final CDR does not cover exactly the same time periods as the FCDR, since only measurements from the 183.31±1 GHz channel are used to derive UTH. Whenever there is a data gap in this channel in the FCDR, no CDR can be produced. The following Table lists the time satellite missions and time periods covered by the UTH CDR:

Instrument	Satellite	Start	End
SSMT-2	DMSP F11	07/1994	04/1995
SSMT-2	DMSP F12	10/1994	01/2001
SSMT-2	DMSP F14	04/1997	01/2005
SSMT-2	DMSP F15	01/2000	01/2005
AMSU-B	NOAA-15/K	01/1999	09/2010
AMSU-B	NOAA-16/L	01/2001	05/2011
AMSU-B	NOAA-17/M	10/2002	12/2009
MHS	NOAA-18/N	08/2005	12/2017
MHS	NOAA-19/N'	11/2009	12/2017
MHS	MetOp-A	06/2007	12/2017
MHS	MetOp-B	01/2013	12/2017

5 Differences with existing products

The FIDUCEO UTH CDR differs from existing UTH CDRs in the following points:

- A new definition of UTH is used in the retrieval, which will allow to combine microwave and infrared observations to one consistent long-term data record of UTH in the future.
- Uncertainties are given for all physical quantities. They are propagated from the FCDRs and divided into three different classes: independent, structured and common uncertainties

6 UTH retrieval

This section provides an overview of the method used to retrieve UTH for this product.

6.1 BT transformation method

Generally, UTH is a quantity that helps to interpret brightness temperatures (BTs) observed in the water vapour channels of microwave and infrared radiometers by scaling it to a more intuitive unit. The concept of UTH is based on the fact that the BTs in the water vapour channels are mainly controlled by the relative humidity (RH) in the atmosphere. Soden and Bretherton (1993) first made use of this relation to infer RH in the upper troposphere from satellite measurements of BT. They derived a simple linear relation between BT in the 6.7 µm water vapour channel and the natural logarithm of UTH. This relation is also valid for the BT in the 183.31±1 GHz channel of microwave instruments such as SSMT2, AMSU-B and MHS (e.g. Buehler and John, 2005) and is used to derive UTH from BT in this CDR:

$$\ln(UTH_\theta) = a_\theta + b_\theta \, BT_\theta \qquad \textit{Equation 6-1}$$

Where θ is the satellite viewing angle and a and b are constants. UTH is a weighted vertical mean of the RH in the upper troposphere, which roughly corresponds to the average RH in a broad layer between 200 and 500 hPa. Equation 6-1 will hereafter be referred to as the BT transformation method. Note that an absolute change of +1 K in BT leads to a relative change of e^b (≈0.9) in UTH. More details on the exact definition of UTH and the scaling parameters a and b are given in the following sections.

6.2 Definition of UTH

Traditionally, UTH is defined as the RH in the upper troposphere, weighted with the humidity Jabobian of the instrument channel that is used to measure BT. Generally, this roughly corresponds to the mean RH in a broad layer between 200 and 500 hPa. However, the exact position and shape of the Jacobian depend on the amount of water vapour in the atmosphere as well as the specific instrument, channel and satellite viewing angle.

This established definition of UTH complicates the comparison of UTH derived from infrared and microwave measurements, because their Jacobians are slightly different even if the atmospheric conditions are identical. Another disadvantage of this method arises when data from climate models are compared with satellite measurements. To obtain the UTH of a model atmosphere, a detour via a radiative transfer simulation has to be made in order to get either humidity Jacobians or BT.

To overcome these shortcomings, we developed a new method to define UTH for this CDR. The new definition is identical for the 183.31±1.0 GHz channels of the microwave sounders and the 6.72 µm channel of the HIRS/2 instrument. Thus, it will be possible to combine this UTH CDR (from microwave sounders) with a UTH CDR based on HIRS/2 observations to get a homogeneous data record covering a time range of more than 40 years. The general concept of UTH stays the same with the new definition. UTH is still roughly the average RH in a layer between 200 and 500 hPa and the exact altitude range depends on the atmospheric conditions. It is the way of averaging the RH profile that changes. This is illustrated in Figure 1. Instead of a Jacobian weighted average, UTH is now the RH averaged over a certain altitude layer, hereafter called UTH layer. This UTH layer is bounded by two characteristic water vapour overburdens. In other words, the boundaries of the UTH layer are the two altitude levels, at which the integrated water vapour (IWV) above have certain values/ exceed certain thresholds IWV1 and IWV2, which depend on the satellite viewing angle θ:

$$UTH_\theta = \frac{1}{z(\mathrm{IWV2}_\theta) - z(\mathrm{IWV1}_\theta)} \int_{z(\mathrm{IWV1}_\theta)}^{z(\mathrm{IWV2}_\theta)} RH(z)\, dz \qquad \textit{Equation 6-2}$$

Thus, the UTH definition itself depends on θ. The following section explains how the thresholds IWV1 and IWV2 are determined.

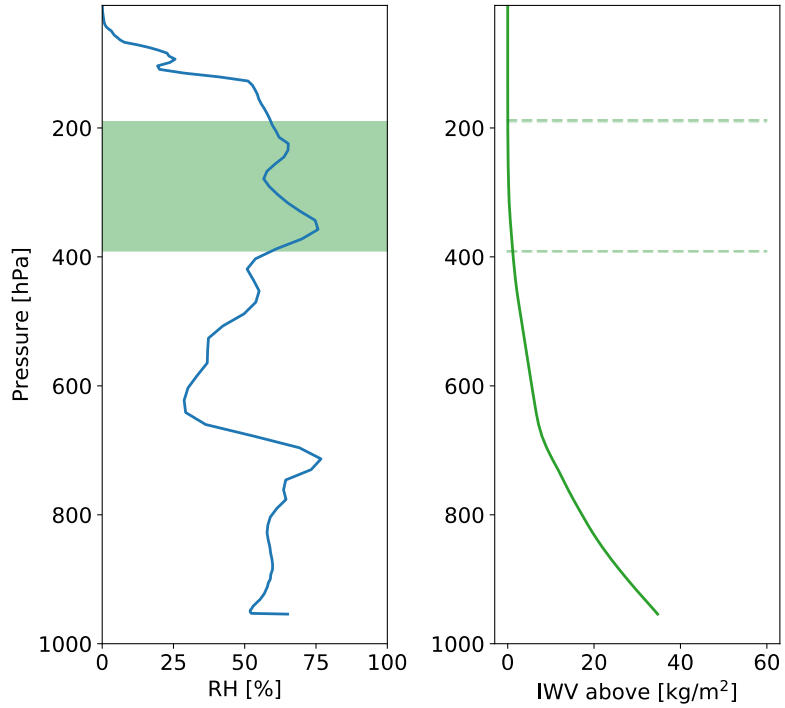


Figure 1 Illustration of the newly introduced UTH definition. The relative humidity of an atmospheric profile (left panel) is averaged in a layer (green shading), which is bounded by two altitude levels, at which the integrated water vapour above (right panel) reaches a characteristic value (dashed lines).

6.3 Determination of scaling parameters and thresholds

As explained above, UTH in this CDR is defined as the mean RH in an atmospheric layer bounded by two altitudes at which the IWV above exceeds certain thresholds (Equation 6-2). This section explains how these thresholds and the corresponding UTH scaling parameters a and b (Equation 6-1) are determined.

For this purpose, we use a training data set developed by Eresmaa and McNally (2014). It consists of diverse atmospheric profiles, selected from the 137-level operational short-range forecast of the European Centre for Medium-range Weather Forecasts (ECMWF). The complete data set includes five subsets, each selected to maximise the variability of a particular atmospheric parameter. The subset chosen for this study consists of 5000 atmospheric profiles and is sampled such that the variability of specific humidity is covered as well as possible. Since the UTH CDR will only cover the tropical region, only tropical atmospheres are selected from the data set, leaving a set of 2812 atmospheres. For these atmospheres, BTs as measured by AMSU-B and MHS are simulated with the Atmospheric Radiative Transfer Simulator (ARTS), described in detail by Buehler et al. (2005) and Eriksson et al. (2011). Under extremely dry conditions the measured BTs can be influenced by the surface emissivity. Profiles where this is the case are discarded. They are identified with the method described by Buehler and John (2005), using the difference in AMSU-B channel 20 and channel 18 BTs.

The most suitable IWV thresholds are determined by an optimization procedure. We start with an arbitrary pair of thresholds for the IWV. For all profiles of the training data set, we determine the two altitude levels at which these thresholds are exceeded. The altitude range between these two levels is the UTH layer. UTH is calculated for every profile by averaging RH within this layer. A linear regression is performed for the logarithm of UTH versus the simulated brightness temperatures of all training profiles and the root mean square deviation of the data points from the regression line is calculated. The whole procedure is repeated several times with different pairs of IWV thresholds. The pair that results in the smallest deviation from the linear relationship is chosen to be the most suitable one.

Figure 3 shows the IWV thresholds that were determined for AMSU-B, MHS and the HIRS/2 instrument. The thresholds for infrared and microwave instruments are very similar, especially at near-Nadir viewing angles. As indicated by the red line, averaged thresholds can be used as a compromise. UTH is recalculated using these compromise thresholds for all three instruments and the regression parameters a and b are determined from the relation $\ln(UTH) = a + b\,BT$ (Figure 4). This approach ensures that the UTH layer is defined identically for both microwave and infrared instrument types.

The thresholds depend on the satellite viewing angle (Figure 3) because the measured signal originates from higher atmospheric layers as the viewing angle moves away from Nadir. A higher altitude corresponds to a lower water vapour overburden and hence a lower threshold of IWV at off-nadir angles. Consequently, the UTH scaling parameters a and b have to be calculated separately for every instrument and viewing angle. They are listed in Table 2 for the 13 innermost viewing angles of AMSU-B and MHS.

For AMSU-B and MHS only these innermost 14 pixels on both sides of the Nadir view are used for the production of this CDR. For SSMT-2 the corresponding pixels are the innermost 5 ones. The advantage of this is that we do not mix information from different altitude layers in the final monthly averages, because for all these viewing angles the thresholds determining the UTH layer are almost constant (Figure 1). This is important, especially when the UTH from satellite observations is compared to UTH from model simulations. However, this comes at the expense of spatial and temporal coverage.

Table 2 Regression coefficients for the innermost 13 viewing angles of AMSU-B and MHS, which are used in the CDR processing.

AMSU-B			MHS		
Viewing angle	**a**	**b**	**Viewing angle**	**a**	**b**
179.45	22.4780	-0.0949	179.4444	22.4859	-0.0950
178.35	22.4780	-0.0949	178.3333	22.4860	-0.0950
177.25	22.4782	-0.0950	177.2222	22.4862	-0.0950
176.15	22.4782	-0.0950	176.1111	22.4863	-0.0950
175.05	22.4785	-0.0950	175.0000	22.4869	-0.0950
173.95	22.4793	-0.0950	173.8889	22.4874	-0.0950
172.85	22.4803	-0.0950	172.7778	22.4884	-0.0950
171.75	22.4821	-0.0950	171.6667	22.4904	-0.0951
170.65	22.4848	-0.0950	170.5556	22.4932	-0.0951
169.55	22.4865	-0.0951	169.4445	22.4947	-0.0951
168.45	22.4879	-0.0951	168.3333	22.4964	-0.0951
167.35	22.4894	-0.0951	167.2222	22.4977	-0.0952
166.25	22.4899	-0.0952	166.1111	22.4984	-0.0952

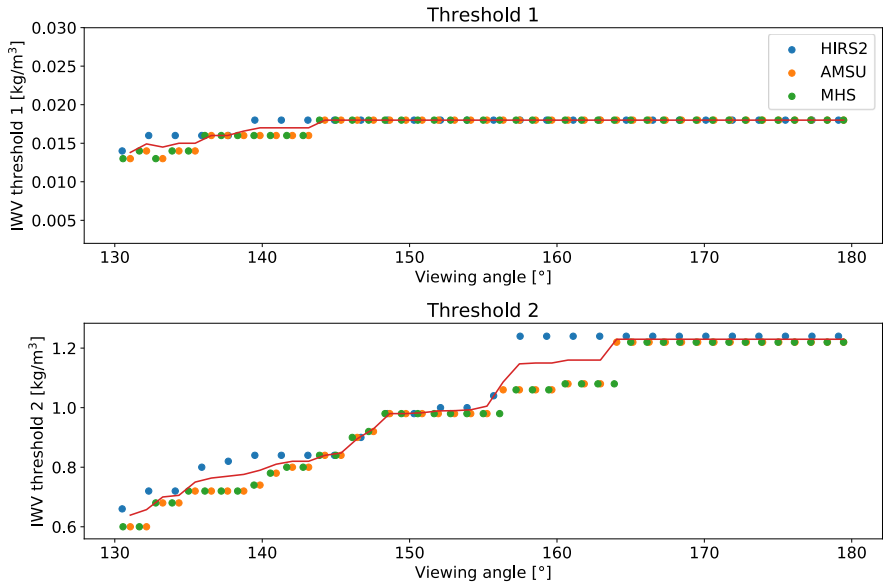


Figure 3 Most suitable thresholds for integrated water vapour determined for AMSU-B (orange), MHS (green) and HIRS2 (blue). Threshold 1 (upper panel) corresponds to the upper boundary of the UTH layer, threshold 2 (lower panel) corresponds to the lower boundary. Averaged compromise thresholds are indicated by the red line.

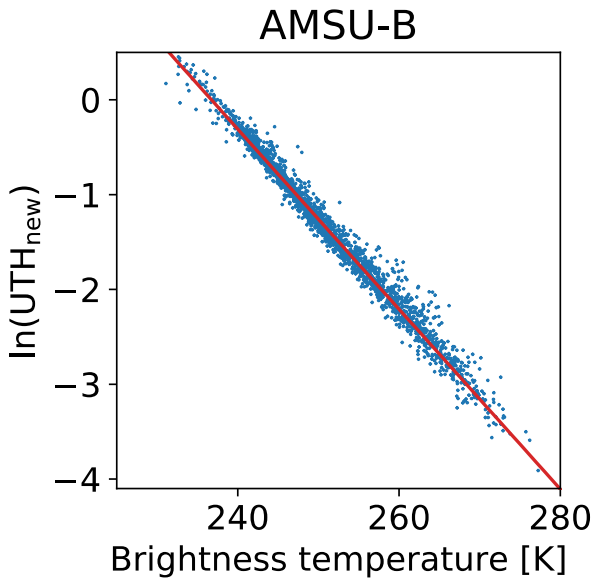


Figure 4 Logarithm of Nadir UTH of the training atmospheres versus simulated Nadir brightness temperature for AMSU-B channel 18.

6.4 Retrieval Error analyses

To assess the performance of the UTH retrieval, the retrieved or "fitted" UTH (UTH_{fitted}) is plotted against the true UTH (UTH_{true}) of the training atmospheres. The latter is calculated directly from the atmospheric profiles of humidity and temperature using the new UTH definition (Section 6.2). For AMSU-B this is shown in Figure 5. For comparison, the same is done for the traditional UTH definition, using the fractional Volume Mixing Ratio Jacobian to weight the RH profile. The scatter of the data points seems to be approximately equal in both panels of Figure 5, indicating that the retrieval performances with the new definition and the

traditional definition are more or less equal. For more detailed retrieval statistics, the data points are aggregated in 10% RH bins of the true UTH values. For each UTH bin, retrieval bias and standard deviation are calculated in absolute and relative units. Here, the retrieval bias is defined as $\overline{\Delta UTH}$, where

$$\Delta UTH = UTH_{\text{fitted}} - UTH_{\text{true}}\,.$$

The retrieval standard deviation $\sigma_{\Delta UTH}$ is the standard deviation of ΔUTH. Similarly, the relative retrieval bias and relative retrieval standard deviation are defined based on the relative difference between fitted and true UTH:

$$\Delta UTH_{\text{rel}} = \frac{UTH_{\text{fitted}} - UTH_{\text{true}}}{UTH_{\text{true}}}.$$

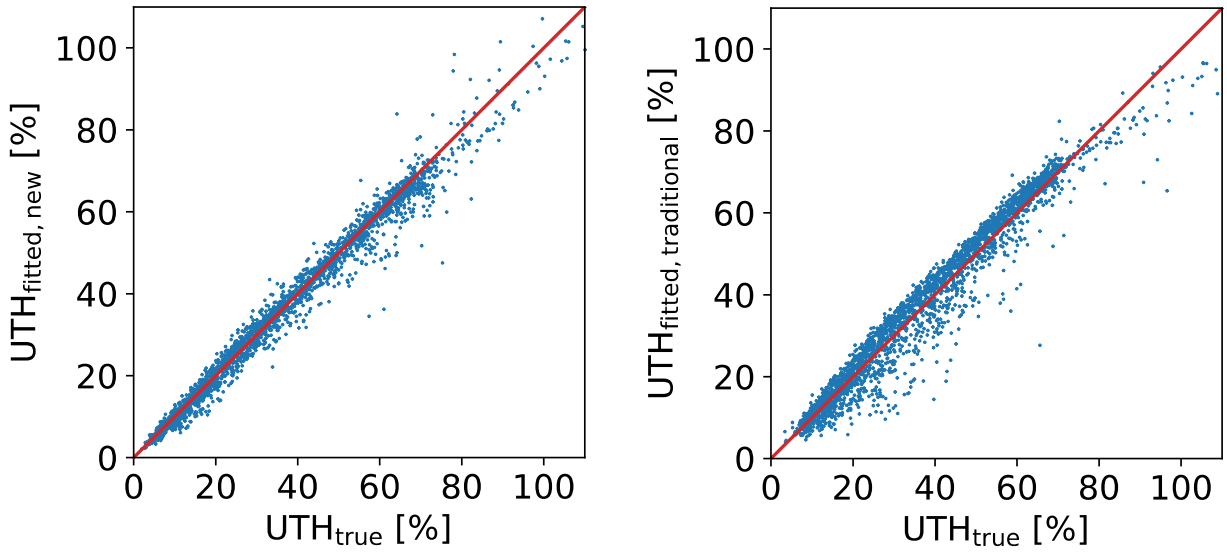


Figure 5 Fitted nadir UTH (from AMSU-B channel 18 brightness temperature) versus true nadir UTH of the training atmospheres defined with the new method (left panel) and the traditional method (right panel).

The results are shown in Figure 6 for both traditional and new UTH definition for the Nadir view of the AMSU-B instrument. Note that UTH values above 80% are very unlikely and the retrieval performance in the corresponding bins is of less importance. Using the new definition, the absolute retrieval bias is close to 0 for UTH values below 40% and increases to about -4% RH for UTH values between 40% and 80%. The relative bias is between -5% and 0.5% in all UTH bins below 80%. These biases have opposite sign but similar magnitudes as the biases obtained with the traditional UTH definition. Only at high UTH values above 60% the traditional definition performs slightly better regarding the bias.

The absolute retrieval standard deviation of the retrieval with the new UTH is approximately 1% RH for 5% UTH and continuously increases to approximately 6% RH for 80% UTH. In relative units, the retrieval standard deviation decreases from about 12-13% for the UTH values below 20% to about 6-8% for UTH values above 40%. Compared to the retrieval with the traditional UTH definition, standard deviations are lower for UTH below 60% and slightly higher for UTH above 60%.

It can be summarized that the retrieval with the new UTH definition performs approximately equally well as the retrieval with the traditional definition. The overall retrieval bias in the Nadir view of AMSU-B is -0.58%

RH in absolute units and -1.4% in relative units. The overall standard deviation is 3.2% RH in absolute and 9.6% in relative units.

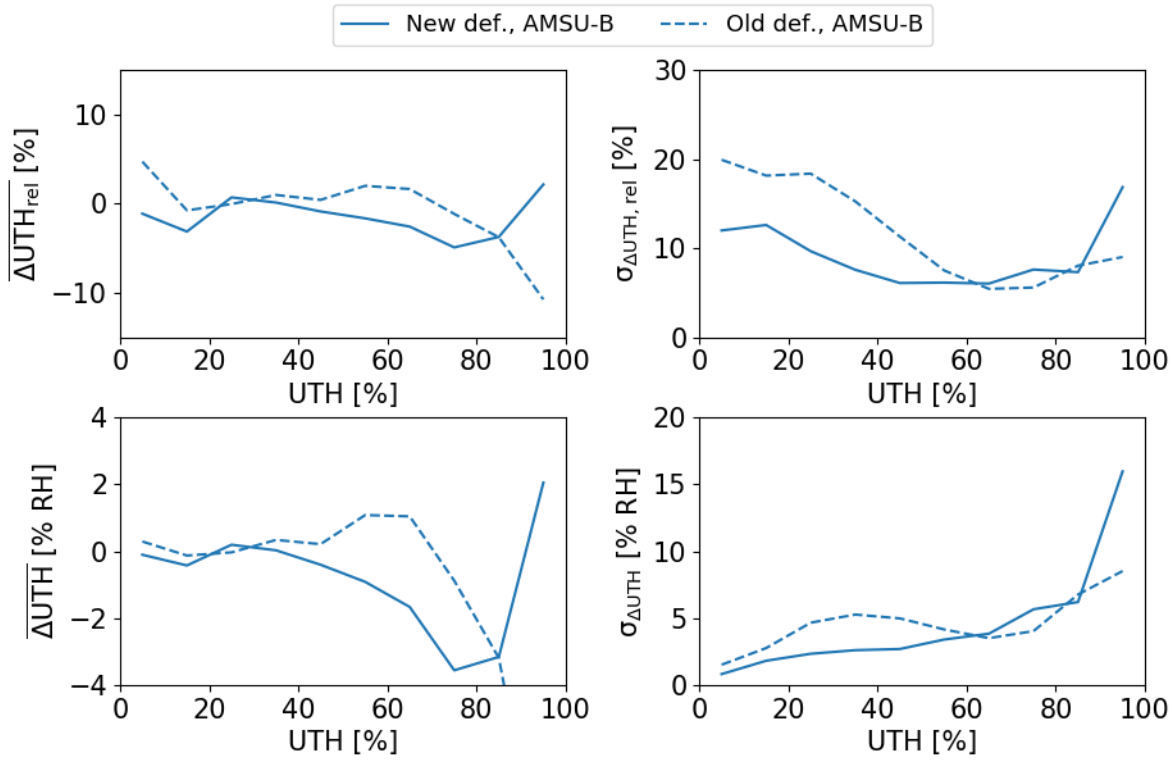


Figure 6 Retrieval bias (left panels) and standard deviations (right panels) in relative units (upper panels) and absolute units (lower panels) in 10% UTH bins obtained with traditional definition (dashed lines) and new definition (solid lines) of UTH.

6.5 Retrieval measurement function

Following equation Equation 6-1, UTH at pixel level (UTH_e) is derived from the 183.31±1.0 GHz brightness temperatures at pixel-level (BT_e), which are provided by the FCDR:

$$UTH_e = e^{\,a+b\,BT_e} + 0 \qquad \textit{Equation 6-3}$$

where

UTH_e	is the Upper Tropospheric Humidity for a given pixel e,
BT_e	is the brightness temperature in the 183.31 ± 1.0 GHz channel of SSMT-2, MHS or AMSU-B, respectively,
a, b	are scaling coefficients,
0	represents the assumption that the form of the equation is right. In particular it assumes that there is a perfectly linear relationship between $\ln(UTH)$ and BT.

Note that all quantities and constants in Equation 6-3 depend on the viewing angle θ, but we have omitted the index θ for better clarity. A measurement-function centred uncertainty diagram for the UTH CDR is given in Figure 7. The specific sources of uncertainty are discussed in D2.4-c.

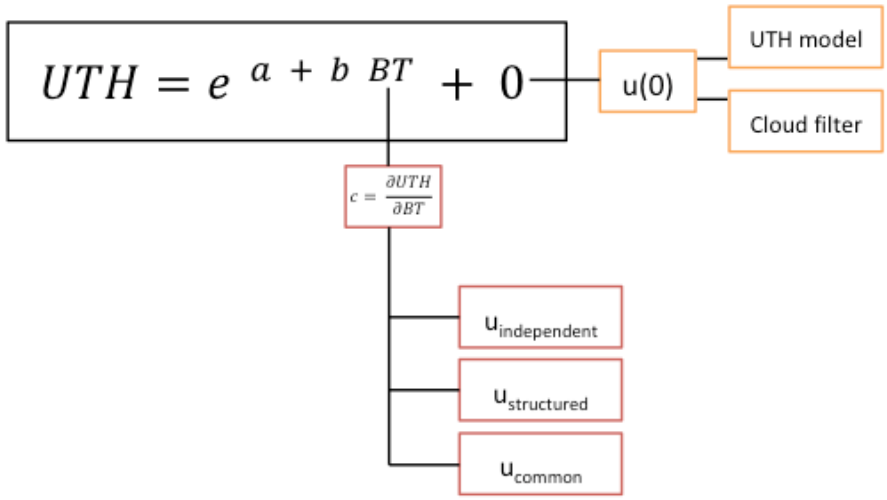


Figure 7 Retrieval measurement function for pixel-level UTH and the corresponding uncertainties.

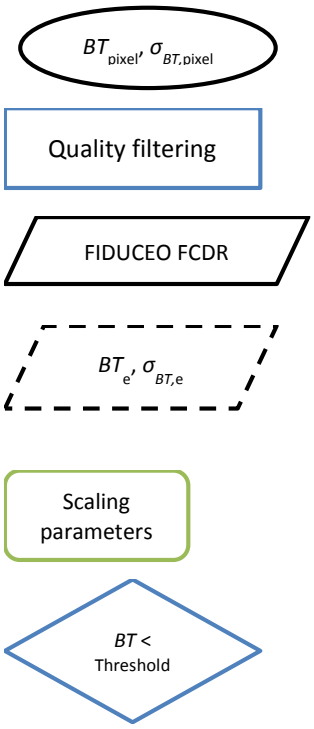
7 Processing chain

7.1 Overview of the processing chain

This section gives an overview of all steps involved in the processing of the UTH CDR, starting with Level 0 (satellite data). The processing chain is shown schematically in Figure 8.

We use the following symbols in the processing chain diagram:

- Input variable
- Processing step
- Input/ output product (available to the user)
- Intermediate product (not available to the user)
- Auxiliary input
- Decision



In a first step, the FIDUCEO FCDR processing, Level 1B data (satellite data) is transformed into Level 1C data, which contains calibrated brightness temperatures BT. This step involves removing overlaps between

adjacent files ("equator-to-equator-framing") as well as calibration and derivation of uncertainties. For more details, see the Microwave FCDR product user guide (Microwave FCDR PUG).

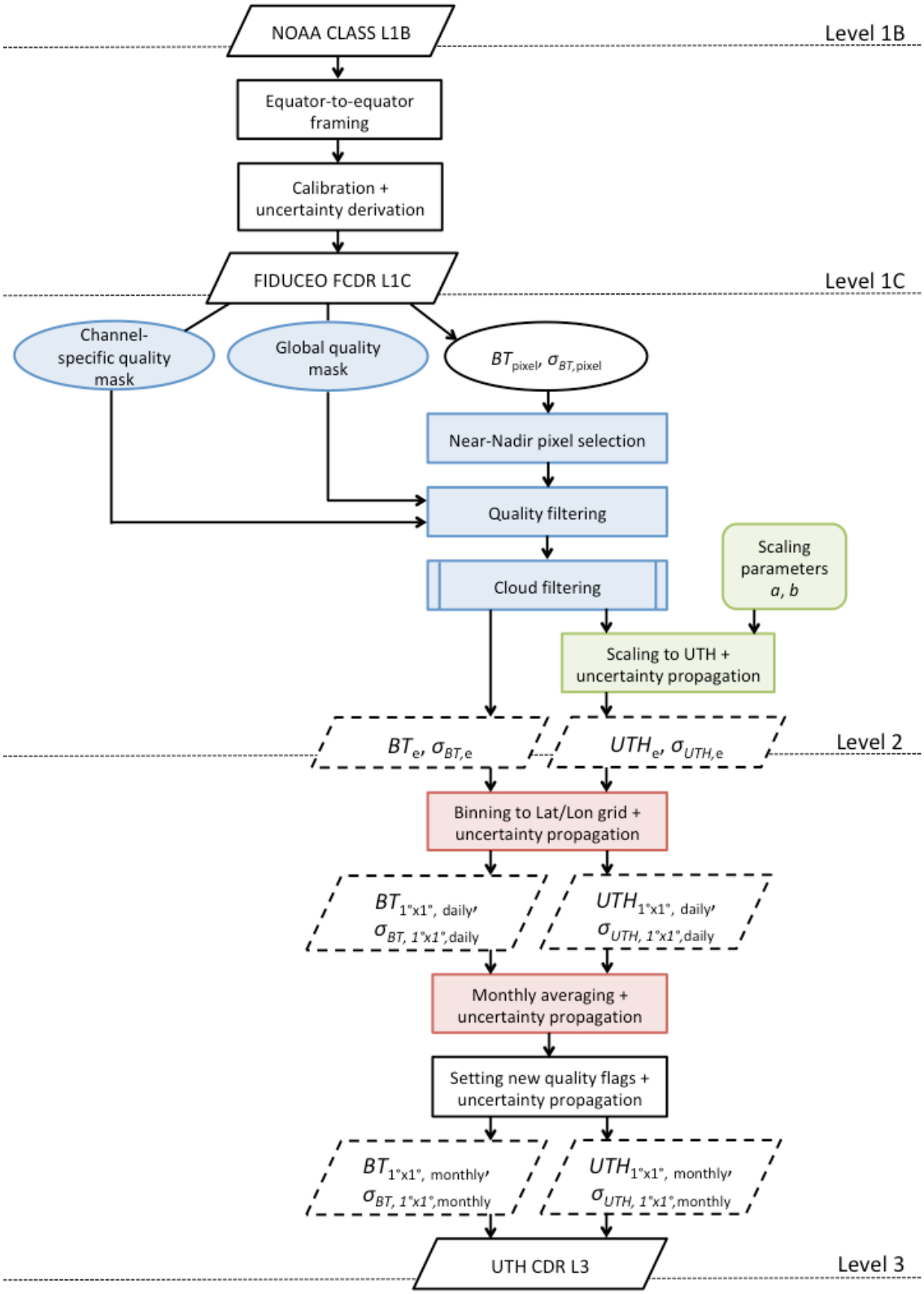


Figure 8 Processing chain of UTH CDR including pre-screening (blue), UTH scaling (green) and gridding/averaging (red).

The Level 2 processing involves two main processing steps: The pre-screening and the conversion of BT to UTH. During the pre-screening, a subset of pixels is selected from the pixel level brightness temperatures BT_{pixel} for further processing. Only pixels located close to the Nadir view of the satellite, which meet certain quality criteria and are not contaminated by clouds are selected. A more detailed explanation of the selection of pixels and the cloud screening is given in Section 7.2. Note that the final CDR product does also contain monthly BTs that still include all cloudy pixels, because this may be more useful for some applications than the cloud-cleared BTs. The processing of these BTs will not be explained separately here since the only difference is that no cloud filter is used.

After the pre-screening the resulting filtered pixel level brightness temperatures BT_e are converted to pixel level Upper Tropospheric Humidity UTH_e using the scaling parameters a and b (see Equation 6-3). At the same time, uncertainties of BT_e are propagated to UTH_e. The propagation of uncertainties is explained in more detail in Section 8.1. Note that the resulting level 2 product is only an intermediate product and not provided to the user.

The remaining processing steps involve gridding and averaging and are performed equally for BT and UTH. All available pixel values (BT_e and UTH_e) of one day are aggregated and averaged in grid cells of a regular latitude/ longitude grid covering the tropical region from 30° North to 30° South with a resolution of 1°. These daily averages ($BT_{1°x1°,\text{daily}}$ and $UTH_{1°x1°,\text{daily}}$) are then combined to monthly averages of each grid cell. A more detailed description of pixel aggregation and averaging is given in Section 7.37.3. Uncertainties of the pixel values are propagated to the monthly grid cell averages, taking into account the correlations between them (see Section 8.1 for more details). The Level 3 product contains both monthly averages of brightness temperature $BT_{1°x1°,\text{monthly}}$ and Upper Tropospheric Humidity $UTH_{1°x1°,\text{monthly}}$ as well as the corresponding uncertainties.

7.2 Details of Level 1 to Level 2 pre-screening

7.2.1 Selection of pixels

Only 32 pixels around the satellite Nadir view are selected from the FCDR data. An advantage of this selection is that the thresholds of integrated water vapour defining the boundaries of the UTH layer are approximately constant for these viewing angles (see Section 6.3). This simplifies the comparison with data from climate models.

The Microwave FCDR contains information about the quality of each pixel in the form of an overall quality bitmask and a channel-specific bitmask. Pixels that are marked as “invalid” by the overall quality bitmask are rejected. Pixels are also discarded if the calibration of the water vapour channel used to retrieve UTH was not possible or there was bad data from the Earth views.

7.2.2 Cloud screening

A general advantage of microwave observations of UTH is that they are less affected by clouds than infrared observations. Brightness temperatures in the 183.31 ± 1.0 GHz channels of AMSU-B, MHS and SSMT2 are only contaminated by high clouds with high ice content. To filter out these clouds, a method developed by Buehler et al. (2007) is used. It combines two criteria: a viewing angle dependent threshold on the 183.31 ± 1.0 GHz brightness temperature and a threshold on the difference between the brightness temperatures in the 183.31 ± 3.0 GHz channel and the 183.31 ± 1.0 GHz channel. The method is illustrated in Figure 9.

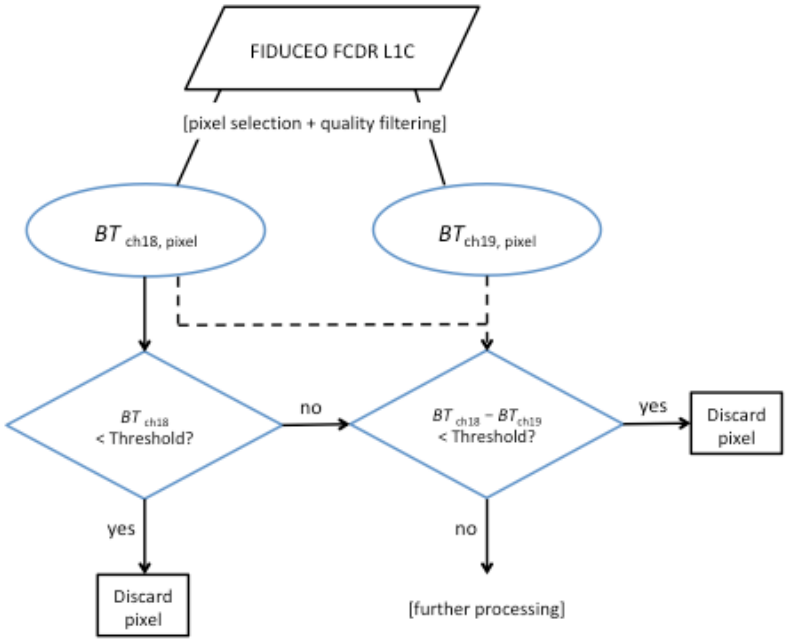


Figure 9 Illustration of the method used to filter out cloudy pixels.

7.3 Level 3 processing (pixel aggregation)

UTH is first determined for each pixel, as described in Section 6.5. Following this, data from different pixels are combined in a two-step process. In the following, the pixel aggregation is described for UTH values, but the same procedure is applied to BT values. In the first step a simple aggregation is made averaging all pixels from observations on one day within a 1° × 1° region. Thus

$$UTH_{1°\text{x}1°,\text{daily}} = \frac{1}{N}\sum_{e=1}^{N} UTH_e + 0 \qquad \textit{Equation 7-1}$$

Here, UTH_e is the UTH within a single pixel, and N is the number of pixels for which UTH is retrieved within the 1° × 1° region (discounting those that were rejected by cloud masking or due to quality issues). Note that this is often considered the "level 2" product, but here we explicitly consider it a level 3 step as it combines data from different pixels.

A second step creates monthly means. Similarly, the equation is also a simple average

$$UTH_{1°\text{x}1°,\text{monthly}} = \frac{1}{N_t}\sum_{t=1}^{N_t} UTH_{1°\text{x}1°,t} + 0 \qquad \textit{Equation 7-2}$$

Here, N_t is the number of daily averages that are combined to one monthly average. Because of the simplicity of these equations, a measurement-function-centred diagram is not provided for this step. The sources of uncertainty are given in Table 1.

Table 1 Sources of uncertainty at Level 2

Measurement function term	Source of uncertainty	Sensitivity coefficient	Comment
$\boldsymbol{UTH_e}$	Propagated uncertainty from L2, separately for independent, structured and common uncertainties	$\frac{1}{N}$	There are N such terms
$+\mathbf{0}$	Representativeness of measured pixels within the area	1	Due to too few data either because of clouds, quality issues or few satellite overpasses (see Section 0 for more details)

8 Uncertainty information

For all grid cell values of UTH and BT in the CDR, three different classes of uncertainty are provided. These uncertainties are propagated from the Microwave FCDR and are described in more detail in Section 8.1. However, there are additional sources of uncertainty at CDR level, which are not explicitly covered in the CDR but the user should be aware of. An overview of these uncertainties is given in Section 8.2. More detailed descriptions of the uncertainties and their propagation are given in D2.4-c.

8.1 Uncertainties propagated from the FCDRs

Pixel-level BT from the Microwave FCDR is the main input variable for the UTH CDR. During the FCDR production, uncertainties associated with all effects that lead to errors in the measurement are propagated through the measurement function to the BT. For more details on this procedure, see the Microwave FCDR PUG. The uncertainties of BT are divided into three different classes, depending on the correlation structure of the underlying effects:

- **Uncertainties due to independent effects** (hereafter independent uncertainties): uncertainties due to purely random effects that generate a completely independent uncertainty from pixel to pixel
- **Uncertainties due to structured effects** (hereafter structured uncertainties): encompass effects that have a correlation scale below the time/ space scales of one orbit (+/- six scanlines for the microwave sounders).
- **Uncertainties due to common effects** (hereafter common uncertainties): represent effects that have a correlation scale larger than one orbit.

All three types of uncertainties are propagated separately to the UTH CDR, using the Law of the Propagation of Uncertainties.

The propagation is done in two steps: From Level 1 to Level 2 and from Level 2 to Level 3. In the first step, pixel-level uncertainties of BT are propagated to pixel-level UTH:

$$u(UTH_e) = \frac{\partial UTH_e}{\partial BT_e} u(BT_e) = b \exp(a + bBT_e)\, u(BT_e) = b\, UTH_e\, u(BT_e)$$

where u is the uncertainty and the index the index *e* indicates values at pixel level.

The second step includes the aggregation and averaging of pixel values. In this step, error correlations have to be taken into account, because they determine whether uncertainties are reduced in an averaging process or not. This becomes evident when the Law of the Propagation of Uncertainty is applied to an average of N numbers (here: pixels):

$$u(UTH_{1^\circ x1^\circ}) = \frac{1}{N} \sqrt{\sum_e u(UTH_e)^2 + 2 \sum_{e>e'} u(UTH_e)\, u(UTH_{e'})\, r(e,e')}$$

Where $u(UTH_{1^\circ x1^\circ})$ is the uncertainty of a grid cell average of UTH and $r(e,e')$ is the correlation coefficient of pixels e and e'. The second term in the square root shows that correlations between uncertainties at pixel level increase the uncertainties of the grid cell average. For independent uncertainties, for example, the correlation term is zero. Thus, the independent uncertainty of a grid cell average is reduced by a factor of

$1/\sqrt{N}$. The common uncertainty of a grid cell average is not reduced compared to the pixel-level uncertainties because of the underlying errors are fully correlated.

The fact that the errors of individual pixel values are correlated also leads to correlations between the individual grid cell averages. These "inter-grid cell correlations" must be taken into account whenever further averaging is performed by the CDR user. For independent and common errors, it is easy to determine these correlations: Independent errors are not correlated at pixel level, so they are also not correlated at grid cell level. Common errors of pixel values are fully correlated over infinite length and time scales and have the same characteristics at grid cell level.

For structured uncertainties, the inter-grid cell error correlations cannot be easily determined. The most accurate way would be to explicitly propagate them to the CDR, but this requires huge correlation matrices. Therefore, we estimate the correlation length scales by performing an example calculation: Correlations are explicitly propagated from pixel values to gridded, monthly averaged values for a small grid with 15 x 15 cells and 1° resolution for one month of data from MHS on Metop-B (January 2016). Figure 10 shows how the error of brightness temperature in the centre grid cell is correlated with those in other grid cells. Correlations occur over a maximum range of +/- 4 grid cells in zonal direction and +/- 2 grid cells in meridional direction. The correlation pattern differs between ascending and descending satellite overpasses because the satellite ground path is tilted in different directions with respect to the meridians. Inter-grid cell correlations are strongest in cross-track direction because for the structured effects, pixels in the same scanline are fully correlated, whereas the correlation is weaker between pixels in neighbouring scanlines.

Assuming that one wants to calculate the average brightness temperature or UTH of the 15 x 15 grid, one has to propagate the structured uncertainties to the average value. There are two options to do that. The first, complex option is to assume a correlation pattern based on Figure 10 for each grid cell and perform a full uncertainty propagation using covariance matrices. However, the effort associated with this approach is somewhat disproportionate, given the fact that the structured uncertainties are generally small compared to the other two classes of uncertainty in the UTH CDR. The second, simpler option is to assume the worst case and treat the structured uncertainty the same as the common uncertainty. This provides an upper limit for the structured uncertainty of the average.

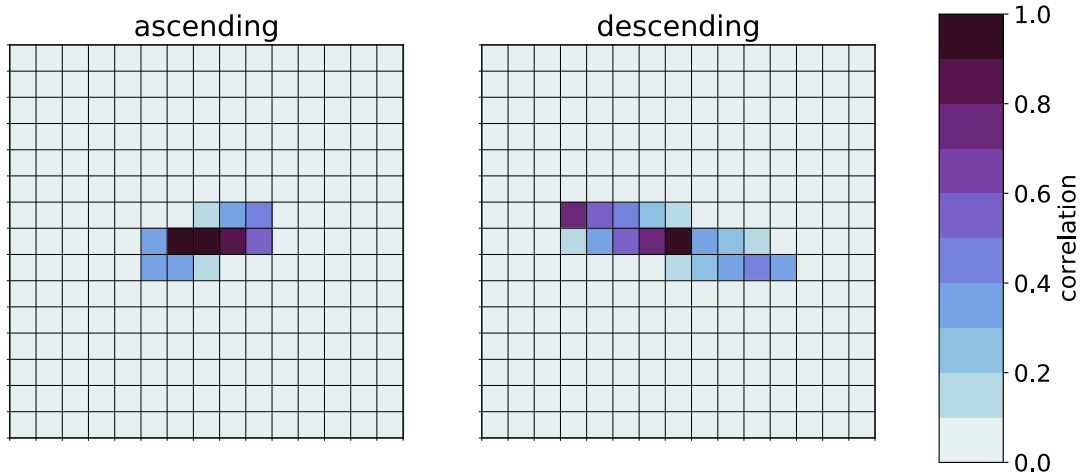


Figure 10 Example of correlations between grid cells in a 15° x 15° grid located at the equator from one month (January 2016) of data from MHS on Metop-A. The correlations of the monthly averaged brightness temperature of the centre grid cell with those in the remaining grid cells are shown for ascending (left) and descending (right) satellite overpasses.

8.2 Uncertainties that have not been covered

Apart from the three types of uncertainty explained above, there are additional uncertainties that are not contained in the CDR. Some of them have a significant effect on the final UTH climatology and should be considered when analysing the CDR data. An overview of these uncertainties is given in the following. See also D2.4-c for more detailed explanations.

8.2.1 Additional uncertainties at Level 2

The assumptions made during the Level 2 processing result in additional uncertainties. They are represented by the "+0" term in the measurement function (Equation 6-3) and summarised in Table 2.

The largest uncertainty at level 2 results from the fact that the linear relationship between brightness temperature and the logarithm of UTH is not perfect (see Section FIXME). This uncertainty is not contained in the CDR, but we advise the reader to take it into account. One possible way to do that is by adding an uncertainty that equals the retrieval standard deviation (i.e. the standard deviation of $\Delta UTH = UTH_{\text{fitted}} - UTH_{\text{true}}$), which was given in Section 6.4 for the 137-level ECMWF data set. This retrieval standard deviation depends on UTH. The values are given in Table 3 (and shown in Figure 6) for UTH bins of 10%. Note that this is a simplified approach, since the grid cell averages of UTH given in the CDR are a combination of many instantaneous UTH values, which all have different uncertainties. A future version of the CDR should include a metrologically traceable uncertainty associated with the UTH model.

Table 2 Additional sources of uncertainty at Level 2

Source of uncertainty and measurement function term affected	Description	Likely sensitivity of output to this	
		On small scales	On large scales
"linear UTH model", + 0	In reality, there are deviations from the linear relationship between brightness temperature and $\ln(UTH)$ (Figure 5). Furthermore, the brightness temperature is also influenced by other factors than relative humidity. One important factor is the pressure in the UTH layer, which controls the width of the water vapour absorption lines. In a warming climate, the UTH layer shifts to lower pressures (higher altitudes). This can lead to an increase in brightness temperature and hence to a decrease in UTH even though the relative humidity does not change. This effect is small for the 183.31+/-1 GHz channel, but significant for infrared water vapour channels.	Overall retrieval standard deviation: ~ 3.2% RH (see Section 6.4) Overall retrieval bias: ~ -0.58% RH (see Section 6.4)	
scaling parameters, a and b	The UTH scaling parameters a and b are determined from a linear regression of the logarithm of UTH of a training data set versus BT simulated with a radiative transfer model. The scaling parameters might be subject to uncertainties due to biases in the simulated BT and due to the fact that the training data set might not sufficiently represent the variability of the real atmosphere. We assume that this uncertainty is negligibly small compared to the "linear UTH model" uncertainty, since ARTS is well validated, and we use a big set of training atmospheres.	Low	Low
Cloud mask, + 0	UTH can only be retrieved from pixels that are not contaminated by clouds. Uncertainties in the cloud masking can lead to erroneous rejection of clear-sky pixels or utilization of cloud-contaminated pixels.	High	Low

Table 3: UTH retrieval standard deviations based on the 137-level ECMWF dataset, which can be used to estimate the uncertainty of UTH that is associated with the exponential UTH model.

UTH bins	0-10%	10-20%	20-30%	30-40%	40-50%	50-60%	60-70%	70-80%	80-90%	90-100%
retrieval standard deviation [% RH]	0.822	1.82	2.34	2.61	2.70	3.41	3.84	5.67	6.21	15.98

8.2.2 Additional uncertainties at Level 3

Additional uncertainties at level 3 arise mainly from the fact that the monthly averages of UTH and BT are obtained from a few satellite observations only and are therefore not equal to the true monthly average. These uncertainties are represented by the "+0" term in Equation 7-1 and Equation 7-2 and summarised in Table 4.

One problem that the user should be particularly aware of is the dry bias in UTH that results from the filtering of clouds. Since clouds are associated with high values of relative humidity, removing cloud contaminated data introduces a dry bias or clear-sky bias in the retrieved UTH climatology. Buehler et al. (2007) estimate this dry bias to 2-3% RH in monthly averages of UTH in regions with deep convective clouds. Figure 11 shows the difference between tropical, monthly mean BTs that are cloud cleared (BT_{filtered}) and BTs that are not cloud cleared (BT_{full}) for all satellite missions covered by the CDR. For most of the instruments BT_{filtered} is about 0.2 K warmer than BT_{full}, which corresponds to a fractional bias of about -2% in the tropical mean UTH. However, for some instruments the difference between BT_{filtered} and BT_{full} is larger. Affected instruments are SSMT-2 on F14 (from 2002 on), AMSU-B on NOAA15 (before 2000 and after 2009), AMSU-B on NOAA16 (from 2007 on) and MHS on NOAA19 (whole mission). As shown by Hans et al. (2017), these instruments are characterized by a poor noise performance. Strong instrument noise leads to a broader distribution of measured brightness temperatures. Cutting brightness temperatures below a fixed threshold at the cold end of the distribution leads to a larger change in the mean brightness temperature than for instruments with less noise. The consequence is an increased dry bias in the monthly averages. For AMSU-B on NOAA15, BT_{full} becomes warmer than BT_{filtered} after 2009. The reason is that the noise in the 183.31+/-3 GHz channel, which is also used for the cloud filtering, becomes very strong. This leads to a moist bias in UTH. Thus, the filtered brightness temperatures and the derived UTH of the above-named satellite missions should be used with caution in the affected time periods.

Another point that the user should bear in mind is the fact that the satellites considered here always pass a given point on Earth at the same local time, because they are in a sun-synchronous orbit. This means that they always observe the same phase of the diurnal cycle. Thus, the derived monthly averages are actually only valid for a certain time of day. To get a better estimate of the true monthly average, the averages from ascending and descending passes should always be combined. Moreover, when monthly averages from multiple satellite missions with different equator-crossing times are available, they should also be combined in order to get a better sampling of the diurnal cycle.

Connected to the uncertainty associated with the sampling of the diurnal cycle is the uncertainty that results from a potential orbit drift of the satellite. With the exception of Metop-A and Metop-B, which are actively stabilized, all satellites included in the UTH CDR are subject to an orbit drift (Figure 12). This means that the observed phase of the diurnal cycle changes over time, which can lead to artificial trends when long time scales are analysed. The user is advised to always compare trends derived from ascending and descending

satellite overpasses and from overlapping satellite missions in consideration of the respective equator-crossing times in order to identify trends that result from an orbit drift.

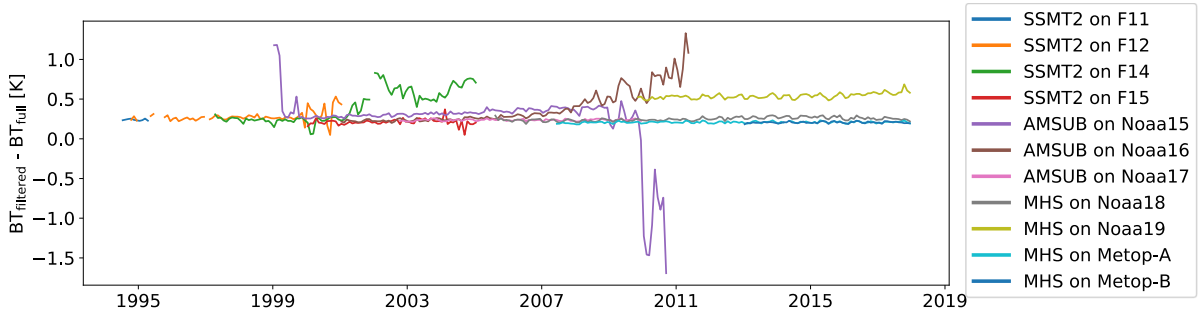


Figure 11: Difference between tropical mean cloud-cleared brightness temperatures (BT_{full}) and unfiltered brightness temperatures ($BT_{filtered}$) for all satellite missions included in the Microwave UTH CDR.

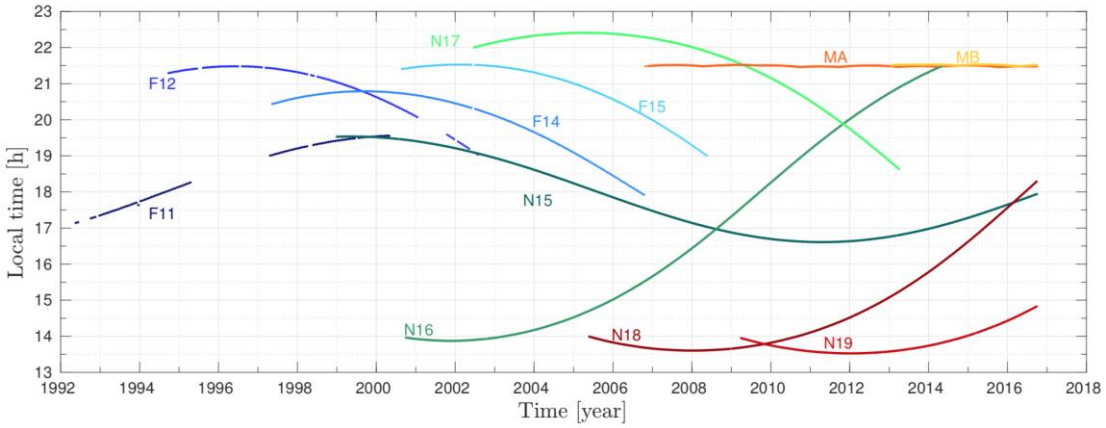


Figure 12: Equator crossing times of ascending passes of all satellites included in this CDR.

Table 4 Additional sources of uncertainty at Level 3

Source of uncertainty and measurement function term affected	Description	Likely sensitivity of output to this	
		On small scales	On large scales
Sparse sampling and diurnal cycle +0	The monthly averages of BT and UTH are estimated from only a few satellite overpasses per month. Furthermore, a satellite always passes a given point on Earth at the same local time and therefore always observes the same phase of the diurnal cycle. Thus, the monthly averages are actually only valid for a certain time of day.	Estimated uncertainty due to diurnal cycle: - Dry subsidence regions with weak diurnal cycle: < 1% RH - Convective regions with strong diurnal cycle: up to 10% RH (See D2.4-c)	
Orbit drift	Orbit drift leads to a shift in observation time. This means that the observed phase of the diurnal cycle of UTH changes over time, which can lead to artificial trends when long time scales are analysed.	low	high
Dry bias/ Clear-sky bias	The preferential sampling of drier, lower UTH cloud-free scenes leads to a dry bias or clear-sky bias. This effect becomes particularly large for instruments with a poor noise performance in the 183.31+/-1 GHz channel and the 183.31+/-3 GHz channel, which are used for cloud filtering (SSMT-2 on F14, AMSU-B on NOAA16, MHS on NOAA19, AMSU-B on NOAA15).	Dry-bias in monthly averages of UTH: 2-3% RH in regions of precipitating and deep convective clouds (Buehler et al., 2007), but larger for instruments with poor noise performance.	

9 Product definition

9.1 Product contents

In this CDR, each file contains monthly averages of UTH and BT on a regular latitude/ longitude grid that covers the tropical region (30° S to 30°N) with 1° x 1° resolution. The monthly averages are obtained from all available FCDR data in this month. Each CDR file has a global attribute `source` that contains a list of all FCDR files used to generate the CDR file. The monthly averages are split into two parts, one average from ascending satellite overpasses and one average from descending overpasses. Each CDR file contains:

- Longitudes and Latitudes of grid cell centres and grid cell boundaries;
- Observation and satellite overpass counts for ascending and descending branches;
- Earliest and latest time of day of pixel contribution to the monthly average, for ascending and descending branches;
- Monthly average and standard deviation of all channel 3 brightness temperatures used to derive UTH (excluding cloudy pixels), for ascending and descending branches;
- Monthly average of all channel 3 brightness temperatures (including cloudy pixels), for ascending and descending branches;
- Monthly average and standard deviation of UTH (derived from channel 3 brightness temperatures);
- Independent, structured and common uncertainty for monthly averages of brightness temperatures and UTH, for ascending and descending branches;

The data may have the following dimensions:

- x – East-west dimension (size: 360)
- y – North-south dimension (size: 61)
- bounds – Dimension defining lower and upper bounds (size: 2)

A full list of variables and their descriptions can be found in Table 5. Note that there is also a variable `quality_pixel_bitmask` in the CDR, which does not contain any information, i.e. no flags are raised and all values are zero. It was originally created to store quality information on pixel level in a Level 2 product and may be used in other FIDUCEO CDRs, but is not suitable for Level 3 products like the UTH CDR.

Table 5 List of all variables in the UTH CDR

Variable Name	Description	Dimensions	Unit
lat	Geographical latitudes of grid cell centers	y	Degree north
lat_bnds	Geographical latitudes of grid cell boundaries	y, bounds	Degree north
lon	Geographical longitudes of grid cell centers	x	Degree east
lon_bnds	Geographical longitudes of grid cell boundaries	x, bounds	Degree east
time_ranges_ascend	Minimum and maximum seconds of day pixel contribution time for ascending nodes	x, y	s

time_ranges_descend	Minimum and maximum seconds of day pixel contribution time for descending nodes	x, y	s
observation_count_ascend	Number of UTH/brightness temperature observations (= number of pixels contributing to the average) in a grid box for ascending passes	x, y	1
observation_count_descend	Number of UTH/brightness temperature observations (= number of pixels contributing to the average) in a grid box for descending passes	x, y	1
observation_count_all_ascend	Number of brightness temperature observations in a grid box when no cloud filtering is done for ascending passes	x, y	1
observation_count_all_descend	Number of brightness temperature observations in a grid box when no cloud filtering is done for descending passes	x, y	1
overpass_count_ascend	Number of satellite overpasses in a grid box for ascending passes	x, y	1
overpass_count_descend	Number of satellite overpasses in a grid box for descending passes	x, y	1
BT_ascend	Monthly average of all brightness temperatures which were used to retrieve UTH in a grid box for ascending passes (calculated from daily averages)	x, y	K
BT_descend	Monthly average of all brightness temperatures which were used to retrieve UTH in a grid box for descending passes (calculated from daily averages)	x, y	K
BT_inhomogeneity_ascend	Standard deviation of all daily mean brightness temperatures which were used to retrieve UTH in a grid box for ascending passes	x, y	K
BT_inhomogeneity_descend	Standard deviation of all daily mean brightness temperatures which were used to retrieve UTH in a grid box for descending passes	x, y	K
u_independent_BT_ascend	Uncertainty of monthly mean brightness temperature due to independent effects for ascending passes	x, y	K
u_independent_BT_descend	Uncertainty of monthly mean brightness temperature due to independent effects for descending passes	x, y	K

u_structured_BT_ascend	Uncertainty of monthly mean brightness temperature due to structured effects for ascending passes	x, y	K
u_structured_BT_descend	Uncertainty of monthly mean brightness temperature due to structured effects for descending passes	x, y	K
u_common_BT_ascend	Uncertainty of monthly mean brightness temperature due to common effects for ascending passes	x, y	K
u_common_BT_descend	Uncertainty of monthly mean brightness temperature due to common effects for descending passes	x, y	K
BT_full_ascend	Monthly average of all daily mean brightness temperatures – including cloudy pixels - in a grid box for ascending passes	x, y	K
BT_full_descend	Monthly average of all daily mean brightness temperatures – including cloudy pixels - in a grid box for descending passes	x, y	K
BT_full_inhomogeneity_ascend	Standard deviation of all daily mean brightness temperatures – including cloudy pixels - in a grid box for ascending passes	x, y	K
BT_full_inhomogeneity_descend	Standard deviation of all daily mean brightness temperatures – including cloudy pixels - which were used to retrieve UTH in a grid box for descending passes	x, y	K
u_independent_BT_full_ascend	Uncertainty of monthly mean brightness temperature including cloudy pixels due to independent effects for ascending passes	x, y	K
u_independent_BT_full_descend	Uncertainty of monthly mean brightness temperature including cloudy pixels due to independent effects for descending passes	x, y	K
u_structured_BT_full_ascend	Uncertainty of monthly mean brightness temperature including cloudy pixels due to structured effects for ascending passes	x, y	K
u_structured_BT_full_descend	Uncertainty of monthly mean brightness temperature including cloudy pixels due to structured effects for descending passes	x, y	K
u_common_BT_full_ascend	Uncertainty of monthly mean brightness temperature including cloudy pixels due to common effects for ascending passes	x, y	K

u_common_BT_full_descend	Uncertainty of monthly mean brightness temperature including cloudy pixels due to common effects for descending passes	x, y	K
uth_ascend	Monthly average of all UTH retrievals in a grid box for ascending passes	x, y	%
uth_descend	Monthly average of all UTH retrievals in a grid box for descending passes	x, y	%
uth_inhomogeneity_ascend	Standard deviation of all daily UTH averages that were used to calculate the monthly UTH average in a grid box for ascending passes	x, y	%
uth_inhomogeneity_descend	Standard deviation of all daily UTH averages that were used to calculate the monthly UTH average in a grid box for descending passes	x, y	%
u_independent_uth_ascend	Uncertainty of monthly mean UTH due to independent effects for ascending passes	x, y	%
u_independent_uth_descend	Uncertainty of monthly mean UTH due to independent effects for descending passes	x, y	%
u_structured_uth_ascend	Uncertainty of monthly mean UTH due to structured effects for ascending passes	x, y	%
u_structured_uth_descend	Uncertainty of monthly mean UTH due to structured effects for descending passes	x, y	%
u_common_uth_ascend	Uncertainty of monthly mean UTH due to common effects for ascending passes	x, y	%
u_common_uth_descend	Uncertainty of monthly mean UTH due to common effects for descending passes	x, y	%

9.2 File format

Files are provided in NetCDF-4 and adhere to the CF Conventions v1.6 where possible. All data fields are internally compressed using parameters chosen based on the dynamic range of meaningful values. Filenames follow the FIDUCEO standard. The filenames have the following structure:

FIDUCEO_CDR_UTH_{INSTRUMENT}_{SATELLITE} _{STARTTIME}_{ENDTIME}_L3_v1.1_fv2.0.0.nc

where {INSTRUMENT} can be either AMSUB or MHS or SSMT-2. {SATELLITE} can be any of the satellites these instruments are flying on. {STARTTIME} is the first second of the first day in the month, with the format {YEAR}{MONTH}{DAY}{HOUR}{MINUTE}{SECOND}, {ENDTIME} is the last second of the last day in the month in UTC.

The rest of the filename is constant throughout the present version of the UTH CDR format. An example filename for MHS on NOAA18 would be:

FIDUCEO_CDR_UTH_MHS_NOAA18_20151201000000_20151231235959_L3_v1.1_fv2.2.0.nc

The NetCDF format is self-documenting. Each file contains global attributes with general information, and a set of data variables. The names of data variables follow standard names from the CF Conventions for those cases where a standard name exists. Where no standard name exists, the FIDUCEO team has introduced a name not included in the standard. All data variables are stored as compressed scaled integers. Data variable attributes describe each variable and its scaling. The appendix contains an example of the headers for a particular file.

9.3 File sizes

A typical UTH CDR file is around 4.3 MB. In total, the Microwave UTH CDR record has a size of about 3.3 GB.

10 Example contents

The figures below show a few examples of the contents of the UTH CDR. All of them show data from MHS on NOAA18 for July 2012.

Figure 13 shows the monthly mean BT (in the 183.31±1.0 GHz channel) and the corresponding inhomogeneity (=standard deviation of daily averages) from ascending passes. The retrieved UTH is shown in Figure 14. Evidently, there are patterns in the data that are not natural, but result from the sparse sampling and/or aliasing of the diurnal cycle. To some extend this can be reduced by averaging the BT/ UTH from ascending and descending passes. However, the user should keep in mind that the sampling induces a significant uncertainty on the monthly averages of BT and UTH.

Note that the standard deviation given here is the sample standard deviation defined as

$$\sigma_x = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \bar{x})^2}{N-1}}$$

where N is the number of daily averages x_i contributing to the monthly average $\bar{x}$. For $N = 1$ (corresponding to one satellite overpass per month), the standard deviation is not defined. This is the case for the grid cells depicted in white in the standard deviations in Figure 13 and Figure 14.

Figure 13 also shows the three uncertainty classes for BT. The magnitudes of the individual uncertainties will not be discussed here, since they differ among the satellite missions. The fact that they all have quite similar magnitudes in the case shown here is a coincidence. Thus, only the basic patterns that emerge in the uncertainties will be discussed. By comparing the independent and structured uncertainties to the satellite overpass counts shown in Figure 15, it becomes clear that a small number of satellite overpasses results in high independent and structured uncertainties. The reason is that these types of uncertainties are reduced by a factor of $1/\sqrt{N}$ in an averaging process with N samples. The more satellite overpasses, the smaller the independent and structured uncertainties. Common uncertainties, by contrast, are not reduced by averaging. Thus, they show a different pattern than independent and structured uncertainties. The magnitude of the common uncertainties seems to depend mainly on the brightness temperature. This mainly results from the uncertainty due to the polarisation correction (see MW FCDR PUG). Note that for later years of AMSU-B on NOAA-15 and on NOAA-16, the strong instrumental degradation causes an increase in the common uncertainties (because of applied radio frequency interference-correction, see MW FCDR PUG).

In the uncertainties of UTH (Figure 14) the same structures are generally visible. However, there is another pattern superimposed, which resembles the UTH itself. As mentioned in Section 6.1, a certain absolute change in BT leads to a large change in UTH when UTH is large and vice versa. Consequently, when uncertainties are propagated from BT to UTH, they have to be weighted with UTH. This explains why the magnitude of the uncertainties of UTH is primarily determined by the magnitude of UTH itself.

Figure 15 shows the different count variables available in the CDR. As the name suggests, the variable `overpass_count` counts the number of satellite overpasses in each grid cell. Every overpass that yields one or more pixels for the grid cell counts. Thus, the variable is most suitable for identifying large data gaps, e.g. the lack of a whole orbit of data. The observation count counts the number of pixels contributing to the monthly average. Two different observation counts are given: `observation_count` only includes pixels that were used to derive UTH, i.e. cloudy pixels are not counted, whereas `observation_count_all` counts all pixels, including cloudy pixels and therefore provides the number of pixels averaged in `BT_full`. Thus, the difference between these two variables yields the number of cloudy pixels that have been discarded during the processing.

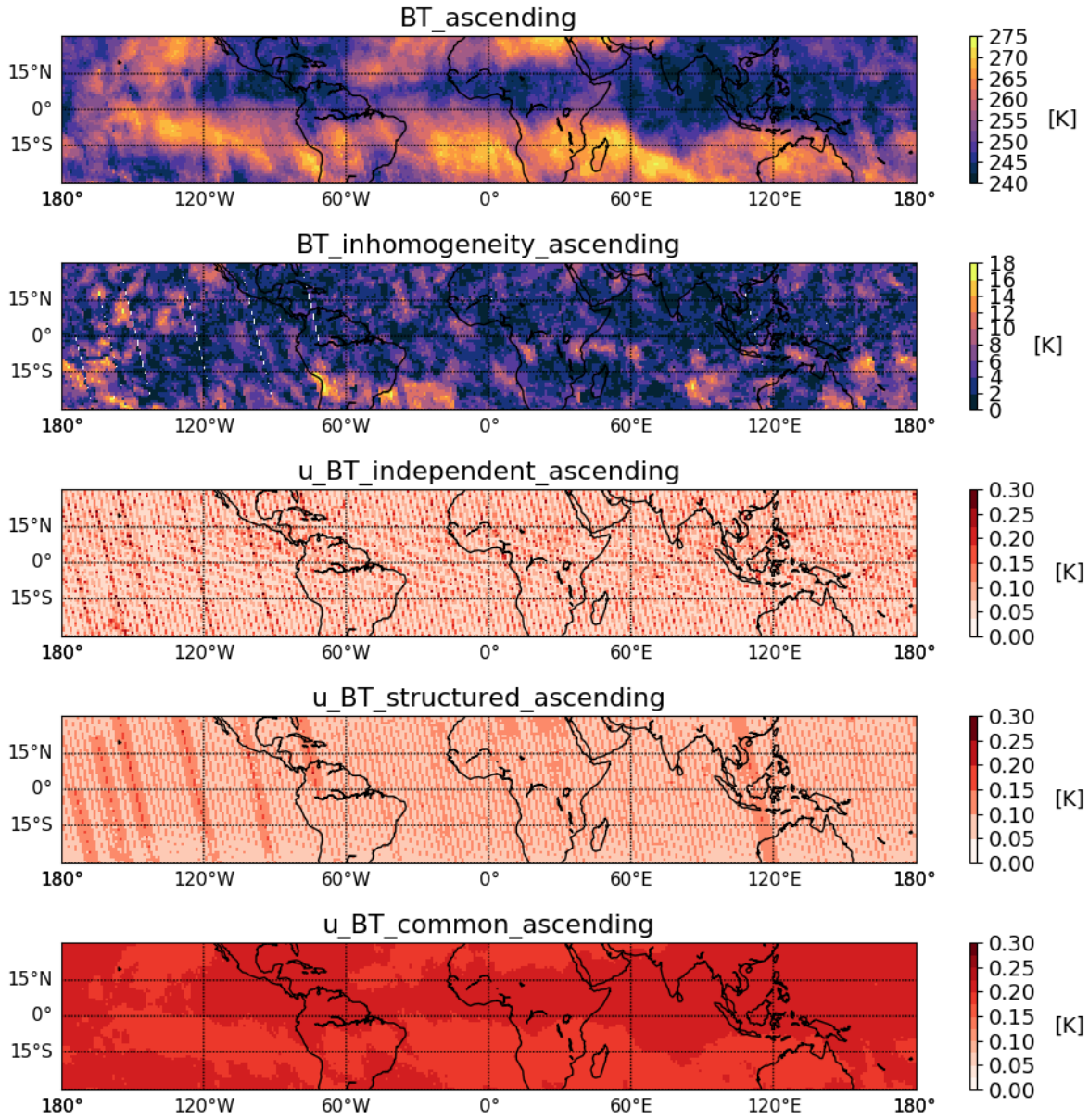


Figure 13 Monthly mean brightness temperature (first panel) for July 2012 from ascending passes of MHS on NOAA18 ass well as standard deviation of daily brightness temperature averages (second panel) and uncertainties of brightness temperature split into three classes: independent, structured and common uncertainties (third to fifth panel).

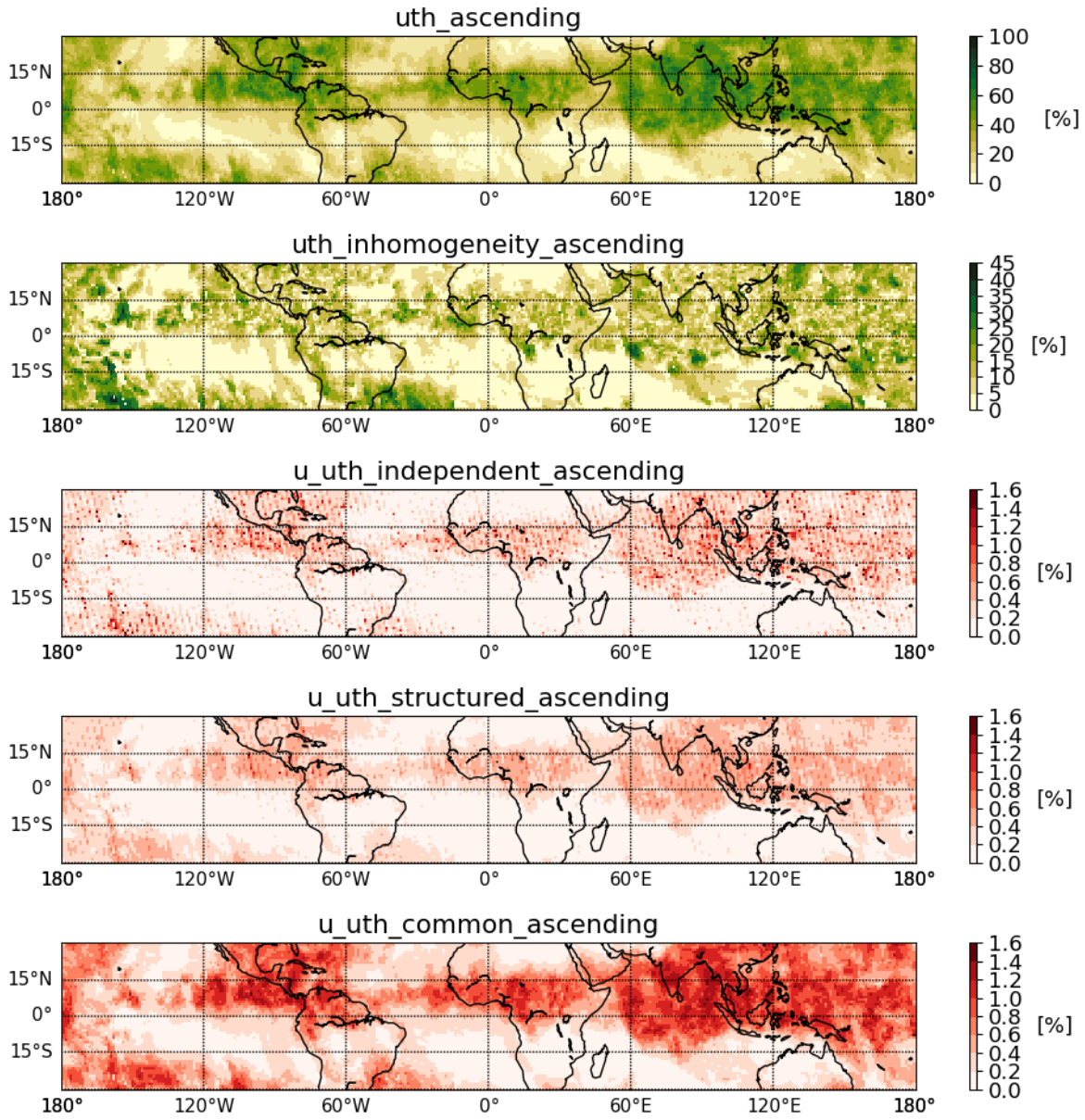


Figure 14 Monthly average of UTH (first panel) for July 2012 from ascending passes of MHS on NOAA18 as well as standard deviation of daily UTH means (second panel) and uncertainties of UTH split into three different classes: independent, structured and common uncertainties (third to fifth panel).

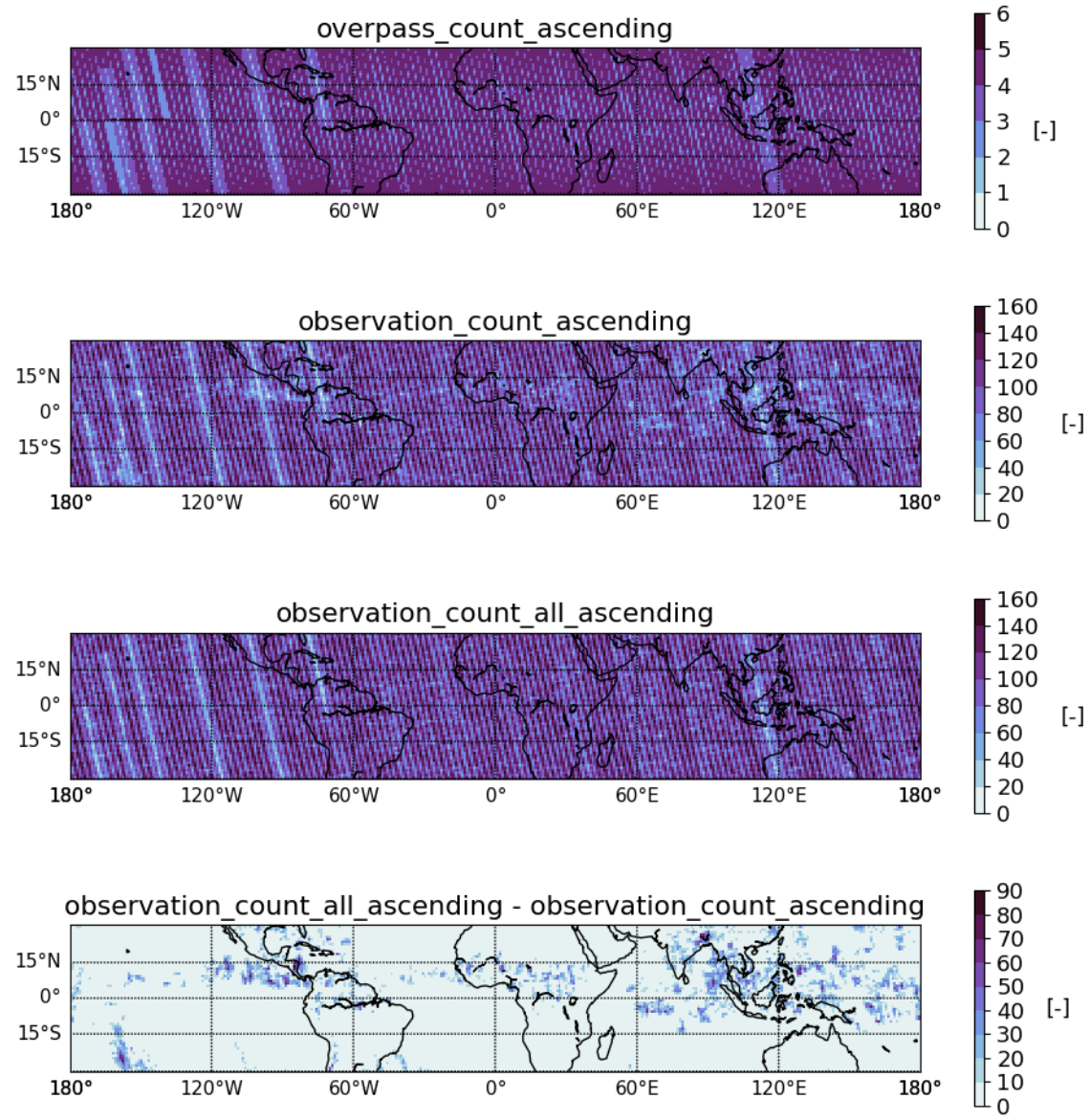


Figure 15 Different kinds of count variables in the CDR: count of satellite overpasses (first panel), count of pixels – excluding cloudy pixels – contributing to the monthly averages (observation_count, second panel) and counts of pixels including cloudy pixels (observation_count_all, third panel). The difference between observation_count and observation_count_all yields the number of cloudy pixels that have been discarded during the processing.

A. Example header

The extract below shows the header for the FIDUCEO CDR L3 file of NOAA17, AMSU-B, containing monthly averages for October 2008, data version 1.1, format version 2.0.0.

```
netcdf FIDUCEO_CDR_UTH_AMSUB_NOAA17_20081001000000_20081031235959_L3_v1_1_fv2.0.0 {
dimensions:
        y = 61 ;
        bounds = 2 ;
        x = 360 ;
variables:
        double lat(y) ;
                lat:_FillValue = NaN ;
                lat:standard_name = "latitude" ;
                lat:long_name = "latitude" ;
                lat:bounds = "lat_bnds" ;
                lat:units = "degrees_north" ;
        double lat_bnds(y, bounds) ;
                lat_bnds:_FillValue = NaN ;
                lat_bnds:long_name = "latitude cell boundaries" ;
                lat_bnds:units = "degrees_north" ;
        double lon(x) ;
                lon:_FillValue = NaN ;
                lon:standard_name = "longitude" ;
                lon:long_name = "longitude" ;
                lon:units = "degrees_east" ;
                lon:bounds = "lon_bnds" ;
        double lon_bnds(x, bounds) ;
                lon_bnds:_FillValue = NaN ;
                lon_bnds:units = "degrees_east" ;
                lon_bnds:long_name = "longitude cell boundaries" ;
        ubyte quality_pixel_bitmask(y, x) ;
```

```
            quality_pixel_bitmask:standard_name = "status_flag" ;

            quality_pixel_bitmask:coordinates = "longitude latitude" ;

            quality_pixel_bitmask:flag_masks = "1, 2, 4, 8, 16, 32, 64, 128" ;

            quality_pixel_bitmask:flag_meanings = "invalid use_with_caution
invalid_input invalid_geoloc invalid_time sensor_error padded_data
incomplete_channel_data" ;

      double time_ranges_ascend(bounds, y, x) ;

            time_ranges_ascend:_FillValue = 4294967295. ;

            time_ranges_ascend:units = "s" ;

            time_ranges_ascend:description = "Minimum and maximum seconds of day
pixel contribution time, ascending nodes" ;

            time_ranges_ascend:coordinates = "lon lat" ;

      double time_ranges_descend(bounds, y, x) ;

            time_ranges_descend:_FillValue = 4294967295. ;

            time_ranges_descend:units = "s" ;

            time_ranges_descend:description = "Minimum and maximum seconds of day
pixel contribution time, descending nodes" ;

            time_ranges_descend:coordinates = "lon lat" ;

      double observation_count_ascend(y, x) ;

            observation_count_ascend:_FillValue = -32767. ;

            observation_count_ascend:description = "Number of UTH/brightness
temperature observations in a grid box for ascending passes" ;

            observation_count_ascend:coordinates = "lon lat" ;

      double observation_count_descend(y, x) ;

            observation_count_descend:_FillValue = -32767. ;

            observation_count_descend:description = "Number of UTH/brightness
temperature observations in a grid box for descending passes" ;

            observation_count_descend:coordinates = "lon lat" ;

      double overpass_count_ascend(y, x) ;

            overpass_count_ascend:_FillValue = 255. ;

            overpass_count_ascend:description = "Number of satellite overpasses
in a grid box for ascending passes" ;

            overpass_count_ascend:coordinates = "lon lat" ;
```

```
      double overpass_count_descend(y, x) ;

            overpass_count_descend:_FillValue = 255. ;

            overpass_count_descend:description = "Number of satellite overpasses
in a grid box for descending passes" ;

            overpass_count_descend:coordinates = "lon lat" ;

      double uth_ascend(y, x) ;

            uth_ascend:_FillValue = NaN ;

            uth_ascend:coordinates = "lon lat" ;

            uth_ascend:long_name = "upper_tropospheric_humidity" ;

            uth_ascend:units = "%" ;

            uth_ascend:description = "Monthly average of all UTH retrievals in a
grid box for ascending passes (calculated from daily averages)" ;

      double uth_descend(y, x) ;

            uth_descend:_FillValue = NaN ;

            uth_descend:coordinates = "lon lat" ;

            uth_descend:long_name = "upper_tropospheric_humidity" ;

            uth_descend:units = "%" ;

            uth_descend:description = "Monthly average of all UTH retrievals in
a grid box for descending passes (calculated from daily averages)" ;

      double u_independent_uth_ascend(y, x) ;

            u_independent_uth_ascend:_FillValue = NaN ;

            u_independent_uth_ascend:description = "Uncertainty of UTH due to
independent effects for ascending passes" ;

            u_independent_uth_ascend:units = "%" ;

            u_independent_uth_ascend:coordinates = "lon lat" ;

      double u_independent_uth_descend(y, x) ;

            u_independent_uth_descend:_FillValue = NaN ;

            u_independent_uth_descend:description = "Uncertainty of UTH due to
independent effects for descending passes" ;

            u_independent_uth_descend:units = "%" ;

            u_independent_uth_descend:coordinates = "lon lat" ;

      double u_structured_uth_ascend(y, x) ;
```

```
            u_structured_uth_ascend:_FillValue = NaN ;

            u_structured_uth_ascend:description = "Uncertainty of UTH due to
structured effects for ascending passes" ;

            u_structured_uth_ascend:units = "%" ;

            u_structured_uth_ascend:coordinates = "lon lat" ;

      double u_structured_uth_descend(y, x) ;

            u_structured_uth_descend:_FillValue = NaN ;

            u_structured_uth_descend:description = "Uncertainty of UTH due to
structured effects for descending passes" ;

            u_structured_uth_descend:units = "%" ;

            u_structured_uth_descend:coordinates = "lon lat" ;

      double u_common_uth_ascend(y, x) ;

            u_common_uth_ascend:_FillValue = NaN ;

            u_common_uth_ascend:description = "Uncertainty of UTH due to common
effects for ascending passes" ;

            u_common_uth_ascend:units = "%" ;

            u_common_uth_ascend:coordinates = "lon lat" ;

      double u_common_uth_descend(y, x) ;

            u_common_uth_descend:_FillValue = NaN ;

            u_common_uth_descend:description = "Uncertainty of UTH due to common
effects for descending passes" ;

            u_common_uth_descend:units = "%" ;

            u_common_uth_descend:coordinates = "lon lat" ;

      double uth_inhomogeneity_ascend(y, x) ;

            uth_inhomogeneity_ascend:_FillValue = NaN ;

            uth_inhomogeneity_ascend:description = "Standard deviation of all
daily UTH averages which were used to calculate the monthly UTH average in a grid
box for ascending passes" ;

            uth_inhomogeneity_ascend:units = "%" ;

            uth_inhomogeneity_ascend:coordinates = "lon lat" ;

      double uth_inhomogeneity_descend(y, x) ;

            uth_inhomogeneity_descend:_FillValue = NaN ;
```

```
            uth_inhomogeneity_descend:description = "Standard deviation of all
daily UTH averages which were used to calculate the monthly UTH average in a grid
box for descending passes" ;

            uth_inhomogeneity_descend:units = "%" ;

            uth_inhomogeneity_descend:coordinates = "lon lat" ;

      double BT_ascend(y, x) ;

            BT_ascend:_FillValue = NaN ;

            BT_ascend:coordinates = "lon lat" ;

            BT_ascend:standard_name = "toa_brightness_temperature" ;

            BT_ascend:units = "K" ;

            BT_ascend:description = "Monthly average of all brightness
temperatures which were used to retrieve UTH in a grid box for ascending passes
(calculated from daily averages)" ;

      double BT_descend(y, x) ;

            BT_descend:_FillValue = NaN ;

            BT_descend:coordinates = "lon lat" ;

            BT_descend:standard_name = "toa_brightness_temperature" ;

            BT_descend:units = "K" ;

            BT_descend:description = "Monthly average of all brightness
temperatures which were used to retrieve UTH in a grid box for descending passes
(calculated from daily averages)" ;

      double BT_full_ascend(y, x) ;

            BT_full_ascend:_FillValue = NaN ;

            BT_full_ascend:coordinates = "lon lat" ;

            BT_full_ascend:standard_name = "toa_brightness_temperature" ;

            BT_full_ascend:units = "K" ;

            BT_full_ascend:description = "Monthly average of all brightness
temperatures - including cloudy pixels - for ascending branches (calculated from
daily averages)" ;

      double BT_full_descend(y, x) ;

            BT_full_descend:_FillValue = NaN ;

            BT_full_descend:coordinates = "lon lat" ;

            BT_full_descend:standard_name = "toa_brightness_temperature" ;

            BT_full_descend:units = "K" ;
```

```
            BT_full_descend:description = "Monthly average of all brightness
temperatures - including cloudy pixels - for descending branches (calculated from
daily averages)" ;

      double u_independent_BT_ascend(y, x) ;

            u_independent_BT_ascend:_FillValue = NaN ;

            u_independent_BT_ascend:description = "Uncertainty of brightness
temperature due to independent effects for ascending passes" ;

            u_independent_BT_ascend:units = "K" ;

            u_independent_BT_ascend:coordinates = "lon lat" ;

      double u_independent_BT_descend(y, x) ;

            u_independent_BT_descend:_FillValue = NaN ;

            u_independent_BT_descend:description = "Uncertainty of brightness
temperature due to independent effects for descending passes" ;

            u_independent_BT_descend:units = "K" ;

            u_independent_BT_descend:coordinates = "lon lat" ;

      double u_independent_BT_full_ascend(y, x) ;

            u_independent_BT_full_ascend:_FillValue = NaN ;

            u_independent_BT_full_ascend:description = "Uncertainty of brightness
temperature including cloudy pixels due to independent effects for ascending
branches" ;

            u_independent_BT_full_ascend:units = "K" ;

            u_independent_BT_full_ascend:coordinates = "lon lat" ;

      double u_independent_BT_full_descend(y, x) ;

            u_independent_BT_full_descend:_FillValue = NaN ;

            u_independent_BT_full_descend:description = "Uncertainty of
brightness temperature including cloudy pixels due to independent effects for
descending branches" ;

            u_independent_BT_full_descend:units = "K" ;

            u_independent_BT_full_descend:coordinates = "lon lat" ;

      double u_structured_BT_ascend(y, x) ;

            u_structured_BT_ascend:_FillValue = NaN ;

            u_structured_BT_ascend:description = "Uncertainty of brightness
temperature due to structured effects for ascending passes" ;

            u_structured_BT_ascend:units = "K" ;
```

```
            u_structured_BT_ascend:coordinates = "lon lat" ;

      double u_structured_BT_descend(y, x) ;

            u_structured_BT_descend:_FillValue = NaN ;

            u_structured_BT_descend:description = "Uncertainty of brightness
temperature due to structured effects for descending passes" ;

            u_structured_BT_descend:units = "K" ;

            u_structured_BT_descend:coordinates = "lon lat" ;

      double u_structured_BT_full_ascend(y, x) ;

            u_structured_BT_full_ascend:_FillValue = NaN ;

            u_structured_BT_full_ascend:description = "Uncertainty of brightness
temperature including cloudy pixels due to structured effects for ascending
branches" ;

            u_structured_BT_full_ascend:units = "K" ;

            u_structured_BT_full_ascend:coordinates = "lon lat" ;

      double u_structured_BT_full_descend(y, x) ;

            u_structured_BT_full_descend:_FillValue = NaN ;

            u_structured_BT_full_descend:description = "Uncertainty of brightness
temperature including cloudy pixels due to structured effects for descending
branches" ;

            u_structured_BT_full_descend:units = "K" ;

            u_structured_BT_full_descend:coordinates = "lon lat" ;

      double u_common_BT_ascend(y, x) ;

            u_common_BT_ascend:_FillValue = NaN ;

            u_common_BT_ascend:description = "Uncertainty of brightness
temperature due to common effects for ascending passes" ;

            u_common_BT_ascend:units = "K" ;

            u_common_BT_ascend:coordinates = "lon lat" ;

      double u_common_BT_descend(y, x) ;

            u_common_BT_descend:_FillValue = NaN ;

            u_common_BT_descend:description = "Uncertainty of brightness
temperature due to common effects for descending passes" ;

            u_common_BT_descend:units = "K" ;

            u_common_BT_descend:coordinates = "lon lat" ;
```

```
      double u_common_BT_full_ascend(y, x) ;

            u_common_BT_full_ascend:_FillValue = NaN ;

            u_common_BT_full_ascend:description = "Uncertainty of brightness
temperature including cloudy pixels due to common effects for ascending branches"
;

            u_common_BT_full_ascend:units = "K" ;

            u_common_BT_full_ascend:coordinates = "lon lat" ;

      double u_common_BT_full_descend(y, x) ;

            u_common_BT_full_descend:_FillValue = NaN ;

            u_common_BT_full_descend:description = "Uncertainty of brightness
temperature including cloudy pixels due to common effects for descending branches"
;

            u_common_BT_full_descend:units = "K" ;

            u_common_BT_full_descend:coordinates = "lon lat" ;

      double BT_inhomogeneity_ascend(y, x) ;

            BT_inhomogeneity_ascend:_FillValue = NaN ;

            BT_inhomogeneity_ascend:description = "Standard deviation of all
daily brightness temperature averages which were used to calculate the monthly
(cloud filtered) brightness temperature average for ascending passes" ;

            BT_inhomogeneity_ascend:units = "K" ;

            BT_inhomogeneity_ascend:coordinates = "lon lat" ;

      double BT_inhomogeneity_descend(y, x) ;

            BT_inhomogeneity_descend:_FillValue = NaN ;

            BT_inhomogeneity_descend:description = "Standard deviation of all
daily brightness temperature averages which were used to calculate the monthly
(cloud filtered) brightness temperature average for descending passes" ;

            BT_inhomogeneity_descend:units = "K" ;

            BT_inhomogeneity_descend:coordinates = "lon lat" ;

      double observation_count_all_ascend(y, x) ;

            observation_count_all_ascend:_FillValue = -32767. ;

            observation_count_all_ascend:description = "Number of all
observations in a grid box for ascending passes - no filtering done" ;

            observation_count_all_ascend:coordinates = "lon lat" ;

      double observation_count_all_descend(y, x) ;
```

```
            observation_count_all_descend:_FillValue = -32767. ;

            observation_count_all_descend:description = "Number of all
observations in a grid box for descending passes - no filtering done" ;

            observation_count_all_descend:coordinates = "lon lat" ;
```

B. Potential further work

- Creation of a UTH CDR based on the harmonized FIDUCEO HIRS FCDR, which could be used to expand this data record to earlier years.
- The original form of the BT transformation method contained a scaled reference pressure (Soden and Bretherton, 1993). Re-introducing this pressure parameter could help to ensure that shifts in the pressure level of the UTH layer do not lead to artificial trends in UTH. This would be particularly important for the UTH CDR based on HIRS observations, since BTs in the infrared water vapour channel are more affected by pressure changes than in the 183.31+/-1 GHz channel.
- Include additional uncertainties that emerge from the BT transformation method, the sparse sampling, orbit drifts and cloud filtering.

C. Known problems

Not all uncertainties that we are aware of are also quantified in the UTH CDR. Apart from the uncertainties of the instantaneous brightness temperatures, which are quantified in the FCDR and propagated to the CDR, additional uncertainties emerge at Level 2 and Level 3. One important example is the uncertainty due to the fact that the model used to scale BT to UTH is not perfect. Another uncertainty is associated with drifts of the satellite orbits and the resulting change in local observation time. This can lead to artificial trends when long time series are analysed, but it is not relevant for a single monthly average, since the orbit drift happens over much longer time scales. Another example is the dry bias that results from the cloud filtering, particularly for instruments with a poor noise performance. A full list of uncertainties that are not covered in this version of the CDR is given in Section 8.2. For the most important ones, we try to provide an estimation. However, we have not included these uncertainties yet since a full understanding and quantification of these uncertainties requires thorough assessments, which were not possible in the time period available to create the CDR and should be part of future work.

Since this CDR is created from the Microwave “easy” FCDR version 4.1, all problems associated with this FCDR version also concern this CDR (see the Microwave CDR PUG for more information).