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

Product Validation and Intercomparison Report (PVIR) - Part 2: CDR-3 & CDR-4

Ref: D4.1

Date: 01 June 2022

Issue: 2.0

For: ESA / ECSAT

Ref: CCIWV.REP.023

This Page is Intentionally Blank

Project	**:**	**Water Vapour Climate Change Initiative (WV_cci) - CCI+ Phase 1**
Document Title:		**Product Validation and Intercomparison Report (PVIR) - Part 2: CDR-3 & CDR-4**
Reference	**:**	**D4.1**
Issued	**:**	**1 June 2022**
Issue :		**2.0**
Client:		**ESA / ECSAT**

Author(s)	**:**	H. Ye (UoR), M. Hegglin (UoR), Marc Schröder (DWD), Anja Niedorf (DWD)

Copyright	**:**	Water_Vapour_cci Consortium and ESA

Document Change Log

Issue/ Revision	Date	Comment
1.0	26 August 2020	First issue for submission to ESA
1.1	13 October 2020	Update according to v1.0 RIDs
2.0	1 June 2022	Final submission to ESA

TABLE OF CONTENTS

INDEX OF FIGURES

INDEX OF TABLES

1. INTRODUCTION

1.1 Purpose

This document is the Product Validation and Intercomparison Report (PVIR) for ESA WV_cci Part 2. This part includes the validation and intercomparison results for CDR-3 and CDR-4, i.e. the monthly vertically resolved water vapour data (VRWV) in the stratosphere and UTLS. This document covers the CDR-3 version 3.3 for 1985–2019 and CDR-4 prototype version 3 for 2010–2014. For validation purposes, the water vapour data from chemistry–climate models, other similarly merged satellite instrument databases (SWOOSH), ground-based observations, aircraft measurements, and reanalyses (ERA-5 and MERRA-2) have been used as reference datasets as described in detail in the Product Validation Plan [RD-1]. A comparison of CDR-3 and CDR-4 with their respective input datasets is also included in this document.

This document is Part 2 of the PVIR, which is split into two parts:

- Part 1 contains the validation and intercomparison of CDR-1 and CDR-2,

- Part 2 contains the validation and intercomparison of CDR-3 and CDR-4.

1.2 Scope

The scope of this product validation and intercomparison report within the ESA WV_cci project is to evaluate and validate the monthly vertically resolved water vapour (VRWV) products with corresponding uncertainties and qualities for users in the climate modelling community. The PVIR thereby provides the validation and intercomparison of monthly VRWV ECV products following the validation protocol and methodology described in the WV_cci PVP [RD-1]. The document also discusses the compliance of the datasets with user requirements formulated by GCOS and the CRG in the URD [RD-2].

1.3 The ESA Water_Vapour_cci project

Water vapour is the single most important natural greenhouse gas in the atmosphere, thereby constraining the Earth's energy balance, and it is also a key component of the water cycle. Due to its importance, the WMO's Global Climate Observing System (GCOS) program has highlighted water vapour as an Essential Climate Variable (ECV) in the GCOS 2016 Implementation Plan. There is consequently the need to consolidate

our knowledge of natural variability and past changes in water vapour and to establish climate data records of both total column and vertically resolved water vapour for use in climate research. These climate data records need to be homogeneous in space and time, which bears great challenges due to changing instrument characteristics and performances. Well-characterised uncertainties are a key attribute of such climate data records in order to reduce the uncertainty in estimates of climate change and global radiative forcing.

The Climate Change Initiative (CCI) is a program of the European Space Agency (ESA), established to tackle the challenges encountered in merging climate data records of ECVs and has the goal to provide climate modellers and researchers with long-term satellite records from current and past European (and other space agencies') missions. The ESA CCI Water Vapour project generates stratospheric and tropospheric water vapour by developing novel methods to determine, merge, and estimate such water vapour data and associated uncertainties.

1.4 WV_cci datasets

The WV_cci datasets comprise four CDRs that are compared with and validated by ground-based, *in situ*, airborne, and satellite-based data records as described in detail by the PVP [RD-1].

- CDR-1: gridded monthly and daily time series of TCWV in units of kg/m^2 that cover the global land areas with a spatial resolution of 0.05° and 0.5°. It covers the period July 2002 to December 2017, and is a combined product of MERIS, MODIS and HOAPS.

- CDR-2: gridded monthly and daily time series of TCWV in units of kg/m^2 that cover the global land and ocean areas with a spatial resolution of 0.05° and 0.5°. It covers the period July 2002 to December 2017, and combines WV_cci CDR-1 with CM SAF HOAPS (microwave imager based) data records.

- CDR-3 contains the vertically resolved water vapour ECV in units of ppmv (volume mixing ratio) and will be provided as zonal monthly means on the SPARC Data Initiative latitude/pressure level grid (SPARC, 2017 [RD-3]; Hegglin et al., 2013, 2021 [RD-4], [RD-5]). It covers the vertical range between 300 hPa and 0.1 hPa, and the time period 1985 to the end of 2019.

- CDR-4 consists of three-dimensional vertically resolved monthly mean water vapour data (in ppmv) with spatial resolution of 5° by 5° in latitude and longitude, covering the troposphere and lower stratosphere. The prototype version of this data

covers the vertical range between 1000 hPa and 10 hPa, and the time period 2010 to 2014.
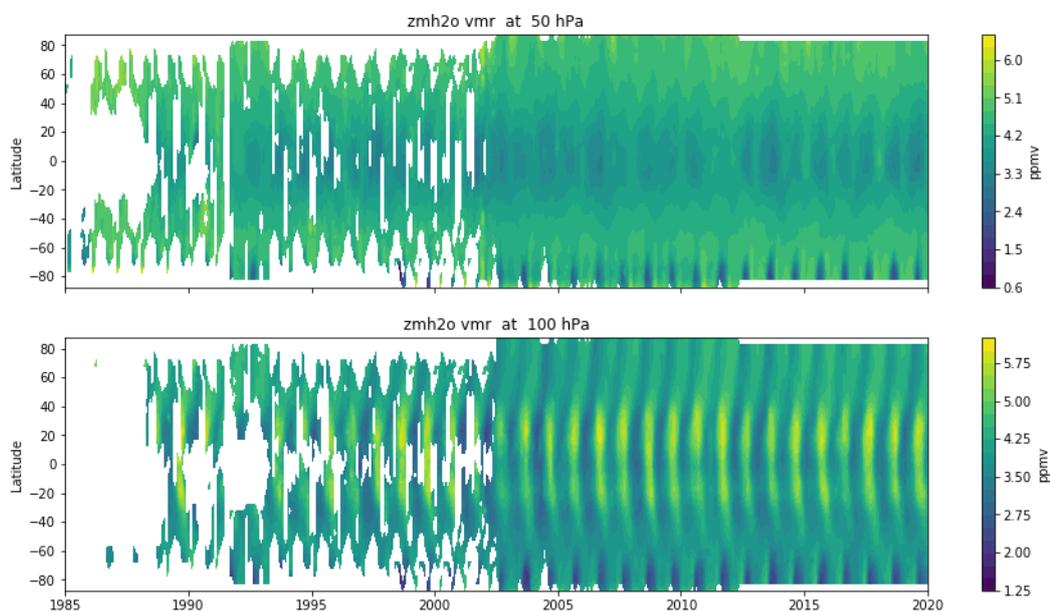
Figure 1-1: Timeseries of CDR-3 v3.3 VRWV at 50 hPa (top) and 100 hPa (bottom) from 1984 to 2019.

Figure 1-1 shows the spatial and temporal coverage of CDR-3 v3.3 VRWV at 50 hPa and 100 hPa over the period 1985–2019. In the early years (1985–2002), the spatial and temporal coverage of CDR-3 is limited due to the sampling characteristics of the instruments offering input data. After 2002, the observations from ESA ENVISAT and NASA MLS provide better spatial and temporal coverage.

2. VALIDATION STRATEGY

2.1 Data sets for validation and comparison

An extensive evaluation of VRWV data products has been carried out based on reference data sets and intercomparison data sets as compiled in the DARD [RD-1] and outlined in the PVP [RD-1], respectively. Note, as already mentioned in the PVP [RD-1], there is a serious lack of reference datasets for VRWV useable for validation, especially going back to the 1980s and 1990s, and also in the UTLS. Thus the evaluations presented here should be considered a comparison rather than classic validation exercises in their character, with knowledge of the true state (and absolute mean) of the atmosphere in terms of water vapour concentrations still highly uncertain.

For CDR-3, three kinds of reference dataset are used for comparison. First, the VRWV data from the merged satellite product SWOOSH, an alternative CDR produced by NOAA (US), is used for comparison. Second, the VRWV data from chemistry–climate models (CCMs) are used for validation and intercomparison. The advantage of the data from CCMs is the global coverage and availability of specified dynamics simulations, i.e. model fields that represent the actual meteorology and thus main driver of WV in the stratosphere. Third, a comparison with available reanalysis datasets is provided.

For CDR-4, the VRWV profile observations from global networks like balloon-borne hygrometer (BBH) data are used as reference to test the product quality at the BBH sites. It is noteworthy that the bias-correction process for satellite observations is relative to the seasonal BBH observations within fixed latitude bins in the tropopause-based coordinate system, described in detail in the ATBD [RD-7]. The climatological distribution of CDR-4 in the equivalent latitude and potential temperature coordinates is compared with JULIA aircraft *in situ* data. Furthermore, the VRWV data from reanalyses (ERA-5 and MERRA-2) are used for comparison. It should be noted that initial comparisons with IAGOS data did not yield satisfying validation results for CDR-4 data, and new methods will need to be used to exploit this database (now planned for Phase 2 of WV_cci). In addition, SHADOZ radiosonde data are found not suitable for the validation of CDR-4 data due to the limited capability (accuracy) of these radiosonde observations in the UTLS region (altitudes above 300 hPa). New validation methods and observations for the validation of CDR-4 will be explored in the next phase of the WV_cci project.

All datasets used for validation and intercomparison are referenced and described in the Data Access Requirement Document [RD-6].

2.2 Methodology

The methodology for the product validation and comparison is described in the PVP [RD-1]. The WV_cci data products are validated with the reference datasets and other VRWV datasets (from chemistry–climate models and reanalyses) introduced in the DARD [RD-1]. For CDR-3, the validation and comparison are performed to the v3.3 data against the WV data from SWOOSH, observations from chemistry–climate models, and the WV data from reanalysis datasets. For CDR-4, the validation and comparison on prototype v3 data are performed against the *in situ* observations from BBH and reanalyses data from ERA-5 and MERRA-2.

The evaluation of VRWV data products is divided into the following parts:

1. Comparisons of bias-correction procedure for CDR-3 and CDR-4.
2. Comparisons of CDR-3 v3.3 against WV data from SWOOSH data.
3. Comparisons of CDR-3 v3.3 against WV data from chemistry−climate models with specified dynamics settings.
4. Comparisons of CDR-3 v3.3 against WV data from reanalysis datasets.
5. Validation of CDR-4 v3 data against VRWV reference data from BBH observations.
6. Comparisons of CDR-4 v3 data against VRWV analyses data from ERA-5 and MERRA-2.

A summary of the comparisons and validation results will be given in Section 8.

3. COMPARISON WITH INPUT DATASETS

This section presents the WV_cci comparison results of CDR-3 and CDR-4 with their respective input datasets. For both CDRs, bias-correction procedures had to be applied since the different input-datasets are known to exhibit large mean biases (Hegglin et al., 2013, 2014, 2021 [RD-4], [RD-8], [RD-5] that make a straight merging between the datasets difficult. The latest and final version of CDR-3 is version 3.3 and CDR-4 is version 3. Sections 3.1 and 3.2 thus present the comparisons between the input data to the merged CDR-3 and CDR-4 products before and after bias-correction.

3.1 Comparison of bias-correction procedure for CDR-3

Figure 3-1 and Figure 3-2 show the input datsets of CDR-3 before and after bias-correction at two different pressure and latitude grid points in the Northern and Southern hemisphere, respectively.

In both cases, the original (uncorrected) input datasets show large relative differences (top panels), with 20% before 1995 and 10% after 1995 for 10 hPa at 37.5°N, and 40% and 65% before and after 2000 for the 100 hPa level at 57.5°S. The bias-correction procedure (see Hegglin et al., 2014 [RD-8] for details) against the chemistry–climate model CMAM30, removes the mean bias of each instrument, with the bias-corrected input datasets (bottom panel) now collapsing onto the CMAM30 time series with good to very good agreement. The inter-instrument biases are reduced to 10% and 2% before and after 1995 for 10 hPa at 37.5°N, and 7% and 3% before and after 2000 for the 100 hPa level at 57.5°S. The final CDR-3 time series (black line and red diamonds) are then transferred back to a reference dataset, here a mixture of Aura-MLS, ACE-FTS and MIPAS observations. The latter correction is particularly obvious in Figure 3-2, where CMAM shows a clear high-bias of 200% in water vapour concentrations at 100 hPa when compared with the instruments at this pressure and latitude grid point.
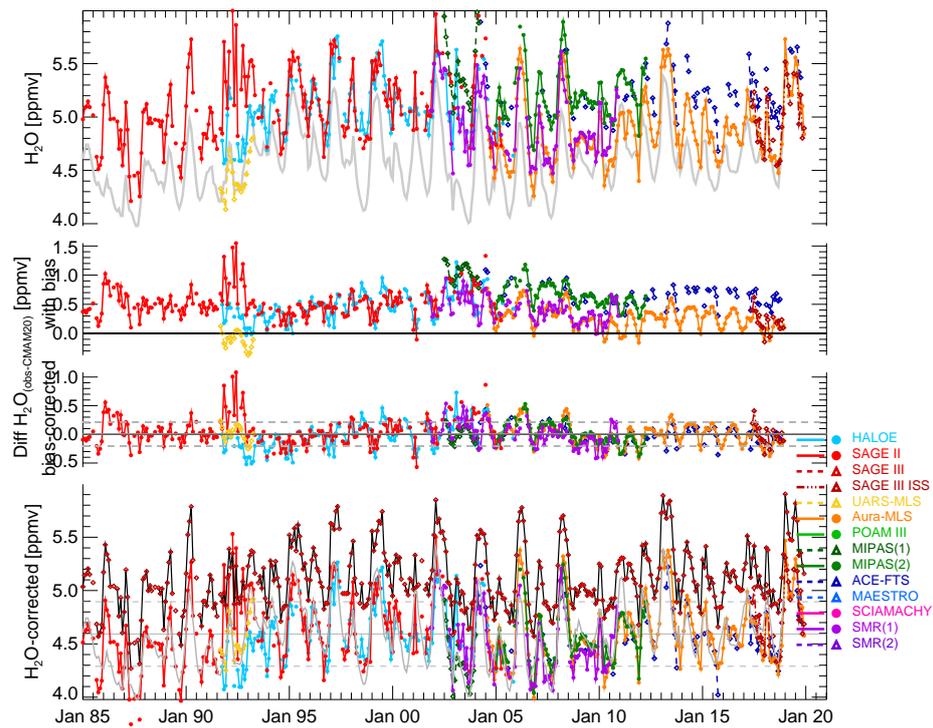
Figure 3-1: Time series of monthly zonal mean water vapour at 10 hPa and 37.5°N. Absolute mixing ratios from different instruments (colours) and CMAM30 (grey) (top panel). Initial (2nd panel) and bias-corrected (3rd panel) differences between observations and CMAM30, and bias-corrected absolute mixing ratios from observations (bottom panel). Grey solid and dashed horizontal lines indicate mean and 1-sigma (standard deviation) of the observational and bias-corrected record averaged over the whole time period. Black line with red squares indicates final merged and bias-corrected (towards reference Aura-MLS dataset) CDR-3.
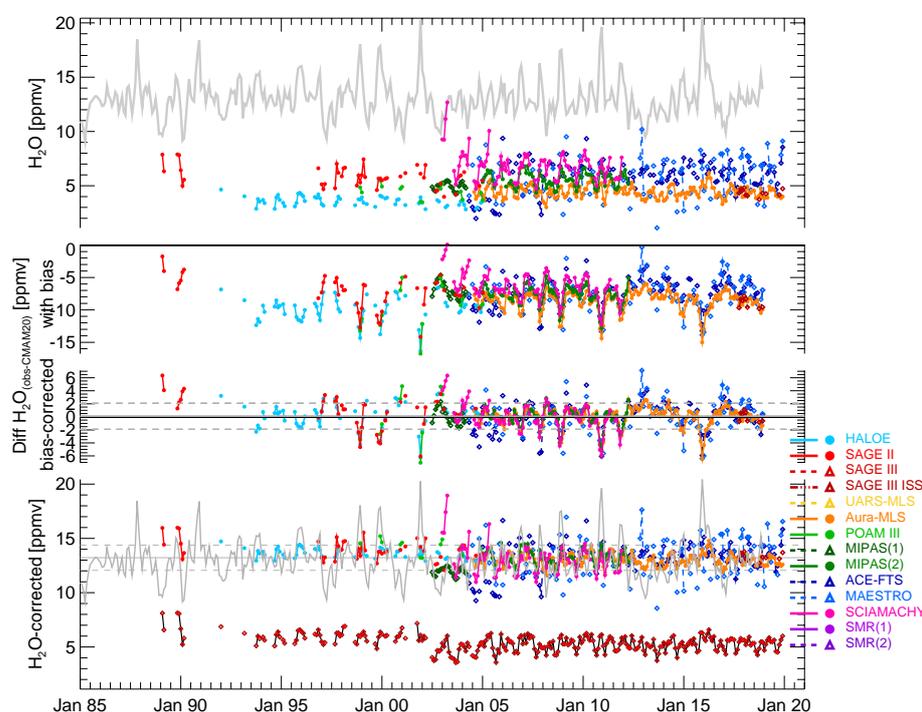
Figure 3-2: Same as Figure 3-1, but for 100 hPa and 57.5°S.

3.2 Comparison of bias-correction procedure for CDR-4

3.2.1 Comparison of bias-corrected satellite data

In the production of CDR-4 data, a bias-correction procedure based on quantile mapping (see ATBD [RD-7]) is applied to the input data before the merging procedure. The details of the merging methodology for CDR-4 are shown in the ATBD [RD-7]. The subsection represents the comparisons of input satellite data before and after applying the bias-correction relative to the *in situ* VRWV profiles from BBH observations.

Figure 3-3 shows the gridded WV from satellite observations before and after applying the bias-correction at two different pressure levels over the tropical region in two different months, respectively. The WV profiles from each satellite instrument are interpolated onto pre-defined pressure levels and then aggregated horizontally into uniform gridded data.
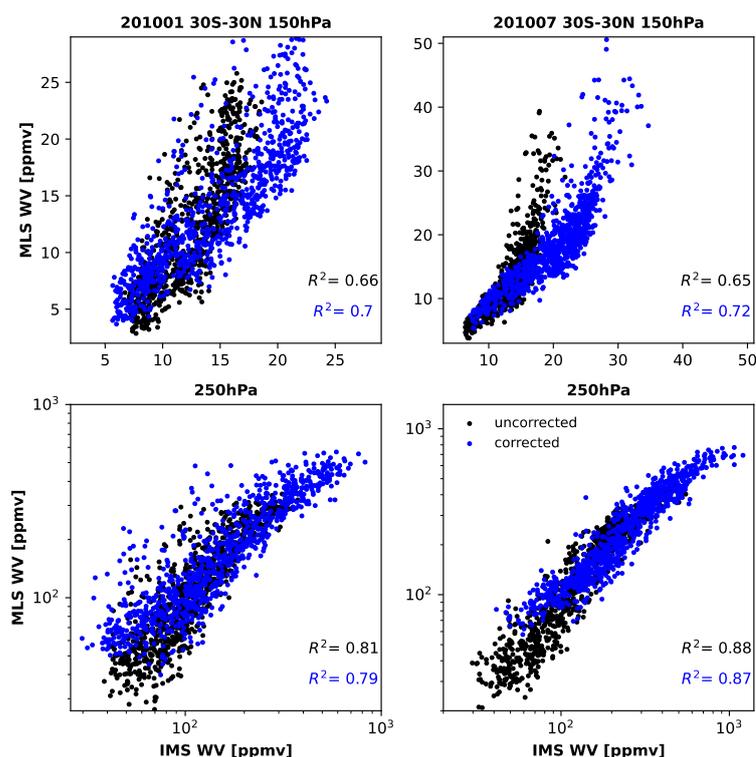
Figure 3-3: Scatterplots of gridded water vapour at 150 hPa (top panel) and 250 hPa (lower panel) from RAL IMS and Aura-MLS over the tropical region in January 2010 (left column) and July 2010 (right column) with both initial (black dots) and bias-corrected (blue dots) WV values.

At 150 hPa, the uncorrected Aura-MLS are larger than RAL IMS where the water vapour concentrations are high. On the contrary, when the water vapour values are low, the IMS, which is less sensitive to low water vapour concentrations, has higher water vapour than MLS. The bias-correction procedure relative to the *in situ* BBH VRWV profiles not only increases water vapour from both satellite observations, but also improves the agreement among satellite instruments in both months. The bias-correction is particularly obvious for IMS, largely increasing the water vapour values at the high end of the concentration range. At this level, the correlation between the two instruments (as expressed by R^2) increases from 0.66 to 0.7 and from 0.65 to 0.72 during NH winter and summer, respectively; see Figure 3-3.

At 250 hPa, MLS is less sensitive to high water vapour, thus has smaller water vapour than IMS. Similarly, the bias-correction procedure increases water vapour for both satellite observations, leading to better agreement in both months. At this pressure level, the correction shows a larger improvement of the water vapour from MLS. However, it should be noted that at this level, the correlation between the two instruments does not significantly improve (R^2 values remain very similar; see Figure 3-3).

3.2.2 Comparison of bias-corrected satellite data with JULIA data

Figure 3-4 and Figure 3-5 show the distributions of relative differences in water vapour between the individual instruments and JULIA observations before and after bias-correction, respectively. As mentioned in the Product Validation and Selection Report (PVASR) [RD-10], although the mean water vapour distribution between satellite and *in situ* observations has remarkable consistency, there are still obvious differences among satellite instruments compared to JULIA (Figure 3-4). The Aura-MLS shows a dry bias in the lowermost stratosphere for all seasons and MIPAS has a dry bias in the upper troposphere (note that due to the sampling characteristics of MIPAS, it can only measure in dry parts of this region; G. Stiller, KIT, personal communication). The bias-correction procedure clearly improves the water vapour in the lowermost stratosphere for Aura MLS and the sampling bias in the upper troposphere for MIPAS (Figure 3-5). Although the differences between satellite and *in situ* observations are still large in the upper troposphere after bias-correction, the differences show very similar distributions with good agreement among the satellite instruments. Note that the final merged product only adopts bias-correction water vapour in the UTLS between 100 hPa and 300 hPa. The initial water vapour observations from Aura-MLS and MIPAS above 100 hPa and from RAL IMS below 300 hPa are used in the merged product.

The bias-correction based on quantile mapping technique has been proven to be a practical solution to remove the differences among input satellite observations. The gridded uncorrected data from all satellite instruments are then merged into the homogenised CDR-4 product.
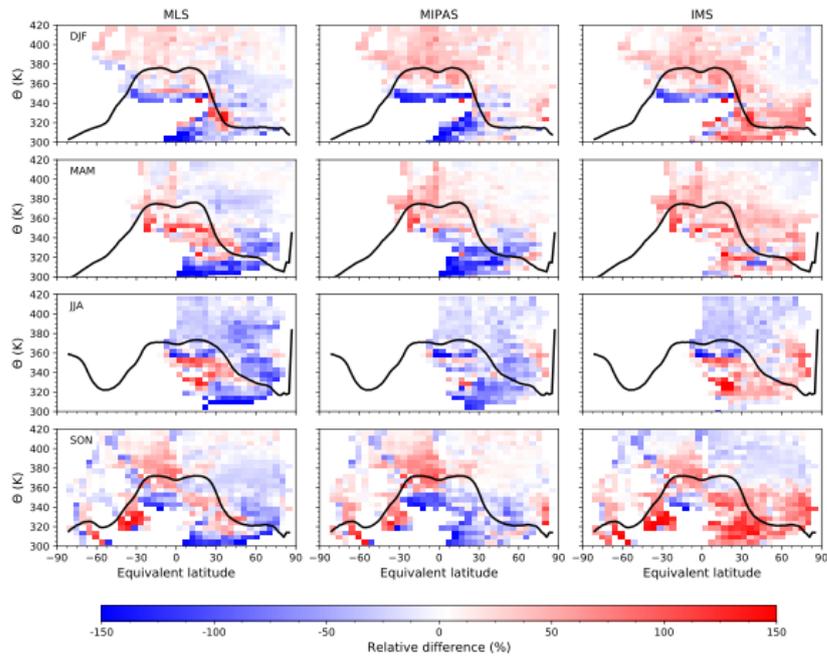
Figure 3-4: Distribution of relative differences in initial water vapour between the individual instruments and JULIA observations in equivalent latitude–potential temperature coordinates for different seasons from 2010 to 2012. From top to bottom: DJF, MAM, JJA and SON. From left to right: MLS, MIPAS and IMS. The black line denotes the location of the AVERAGE thermal tropopause in seasons derived from the MLS JETPAC information.
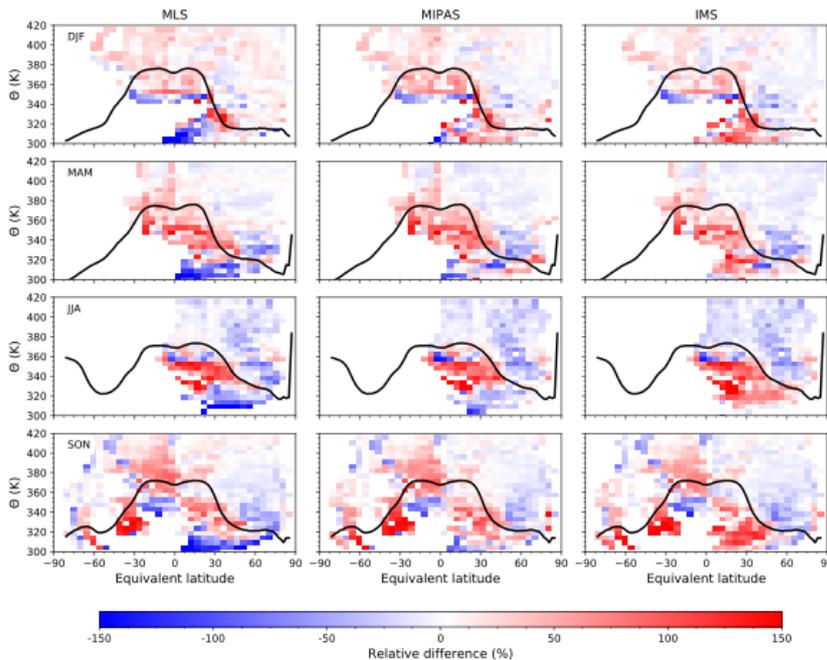
Figure 3-5: Same as Figure 3-4, but for the bias-corrected water vapour between individual instruments and JULIA observations.

4. VALIDATION AND COMPARISON OF WV_CCI PRODUCTS

This section presents the validation and comparison for the monthly VRWV CDR-3 v3.3 and CDR-4 v3 data products. The current version of CDR-3 is v3.3 from 1985 to 2019 on 28 pressure levels between 300 and 0.1 hPa and the current version of CDR-4 is v3 from 2010 to 2014 on 26 pressure levels between 1000 and 100 hPa. As mentioned above, due to the limitations in the spatial and temporal coverage of available VRWV reference datasets, one has to be aware of the limited power of the following comparisons and validation exercise to determine the uncertainty of the two monthly VRWV products.

4.1 Comparison of monthly zonal mean VRWV (CDR-3)

The validation of the final level 3 CDR-3 v3.3 is carried out against data records from SWOOSH and the IGAC/SPARC Chemistry Climate Model Initiative (CCMI) models in specified dynamics (SD) mode. The CCMI SD simulations nudge their dynamics towards observed meteorological fields from reanalyses, so to represent the observed day-to-day variations in transport and temperatures that affect the water vapour field. CDR-3 v3.3 data is also compared with reanalysis datasets. Currently, the CDR-3 data is compared to the reference datasets focussing on two aspects: water vapour mixing ratio time series and anomalies. The comparisons are carried out for time-series averaged globally and for the tropics.

4.1.1 Comparison of CDR-3 against SWOOSH data

Comparison of absolute WV mixing ratio timeseries and their differences

The SWOOSH data have zonal mean VRWV with temporal coverage from 1984 to 2019 on 31 pressure levels from 316 to 1 hPa. The spatial resolutions for SWOOSH include 2.5°, 5° and 10° intervals in latitude. Here, the comparisons are against the zonal mean SWOOSH data with 5° intervals, the same resolution as that in CDR-3.

Figure 4-1 and Figure 4-2 show the time series of zonal mean water vapour at 100 hPa for CDR-3 and SWOOSH in the global and tropical regions, respectively. The global mean water vapour mixing ratios from CDR-3 v3.3 data are close to those from SWOOSH at 100 hPa in both regions. Note that the latest v3.3 CDR-3 data has significantly increased WV mixing ratios at 100 hPa (and throughout most of the stratosphere) when compared to its v0, an earlier product of the WV_cci. This

improvement is achieved by using the model as transfer function towards a 'best' instrument reference, which now is used as the mean between Aura-MLS, ACE-FTS, and MIPAS data at a given dataset grid-point.

Consistent with the result of the similarity of absolute values, the global mean differences between CDR-3 v3.3 and SWOOSH are close to zero from 2005 onward. However, CDR-3 is clearly drier than SWOOSH before 2005. The change in the differences observed in 2005 between the two datasets is expected due to the difference in the merging algorithms between CDR-3 and SWOOSH. However, this difference will have an influence on derived trends from these two CDRs. It is expected (and evidenced hereafter for CDR-3 in comparisons with CCMs and reanalyses) that CDR-3 time series are more consistent with reanalyses than SWOOSH before 2005. After 2005, the satellite observations from MIPAS and MLS have larger spatial and temporal coverage and are more stable, leading to overall much smaller variability in the differences between the two merged CDRs.
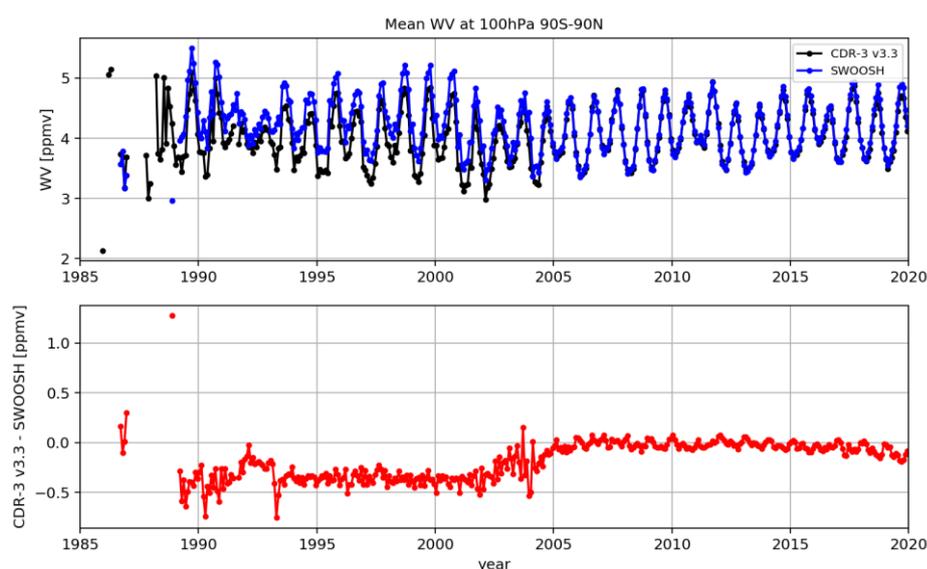
Figure 4-1: Monthly time series (top) and bias (CDR-3–SWOOSH, bottom) of global mean (90°S–90°N) VRWV of CDR-3 and SWOOSH at 100 hPa from 1984 to 2019.

The tropical mean water vapour shows smaller differences (0.25 ppmv) than the global mean (0.35 ppmv) between these two merged data before 2002. However, after 2005, the tropical mean water vapour from CDR-3 is about 0.2 ppmv higher than that from SWOOSH. Note that after 2010, both global and tropical mean water vapour from CDR-3 exhibit a slightly decreasing trend in the relative differences compared to SWOOSH. This may attribute to SWOOSH using Aura-MLS version 4.2, which showed a drift (spurious positive trend) in WV, while this drift is corrected for in Aura-MLS version 5.0 used in CDR-3.
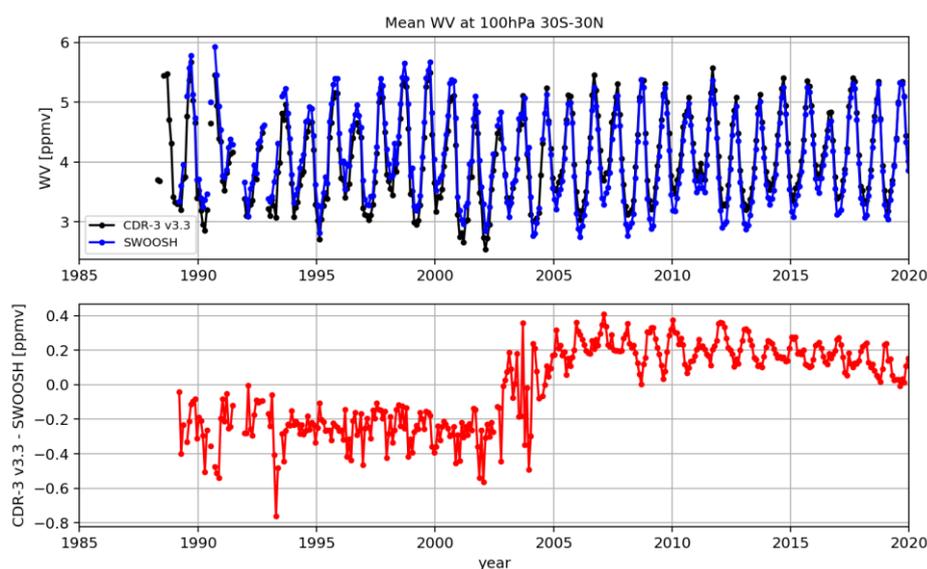
Figure 4-2: Same as Figure 4-1, but for tropical region (30°S–30°N) only.

Figure 4-3 to Figure 4-6 show the same evaluations of the time series of zonal mean water vapour at 50 and 200 hPa for global mean and tropics, respectively.

At 50 hPa (Figure 4-3 and Figure 4-4), the global and tropical mean water vapour from CDR-3 are both drier than SWOOSH and the differences are similar both in scale and variations. Note that the differences at 50 hPa reveal a strong inhomogeneity in the seasonal cycle over the period 2005–2011, which can be attributed to the differences in the seasonal cycles of ENVISAT instruments when compared to the instruments included in SWOOSH. The observed decreasing trend of water vapour after 2010 in CDR-3 compared to SWOOSH is larger than that at 100 hPa.
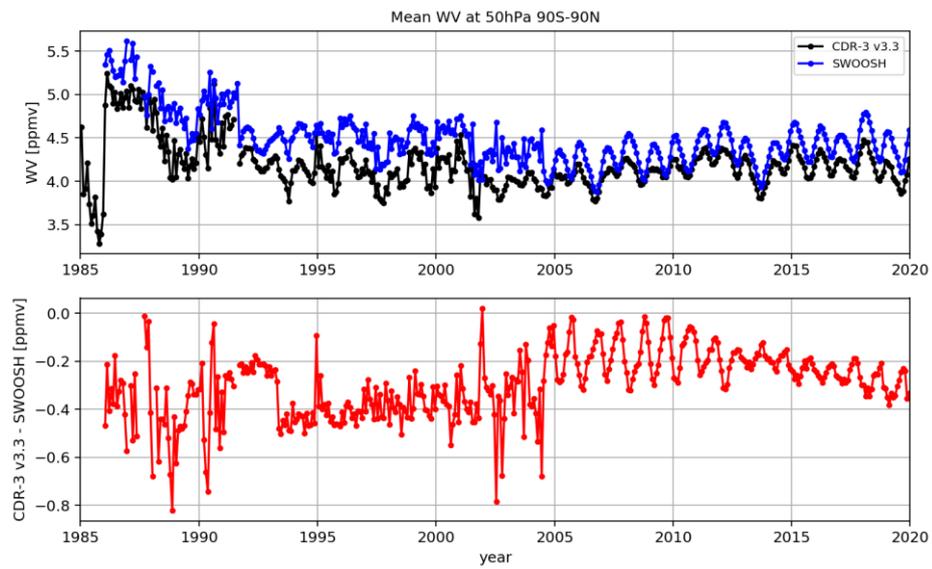
Figure 4-3: Monthly time series (top) and bias (CDR-3–SWOOSH, bottom) of global mean (90°S–90°N) VRWV of CDR-3 and SWOOSH at 50 hPa from 1984 to 2019.
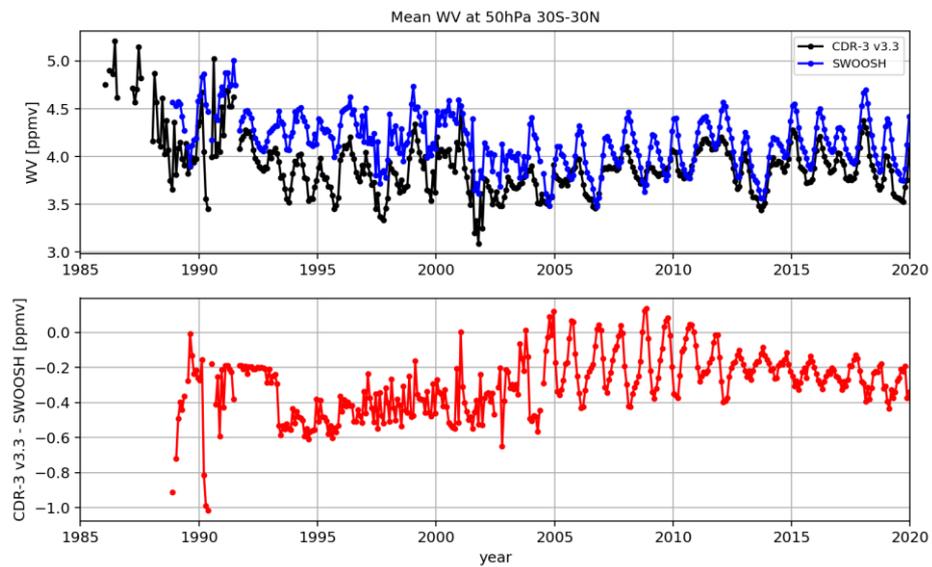
Figure 4-4: Same as Figure 4-3, but for tropical region (30°S–30°N) only.
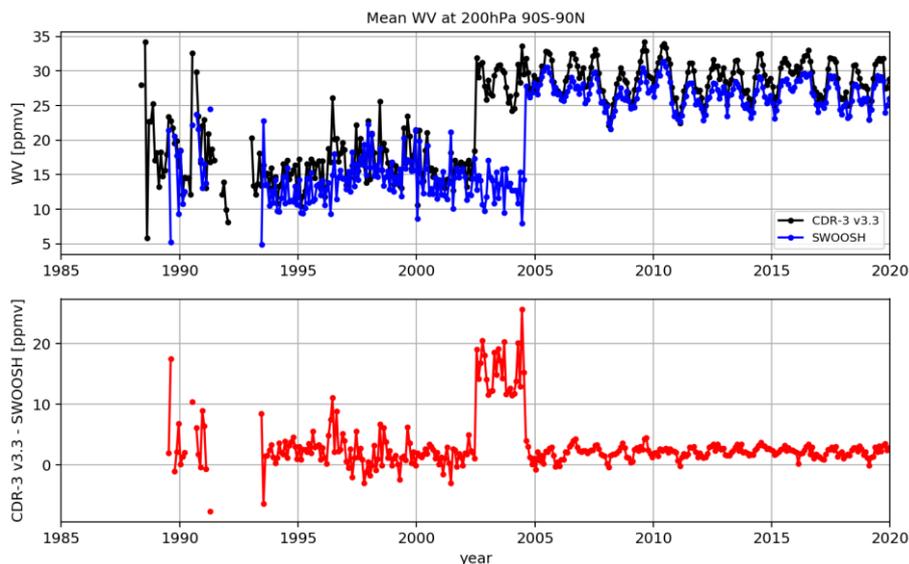
Figure 4-5: Monthly time series (top) and bias (CDR-3–SWOOSH, bottom) of global mean (90°S–90°N) VRWV of CDR-3 and SWOOSH at 200 hPa from 1984 to 2019.
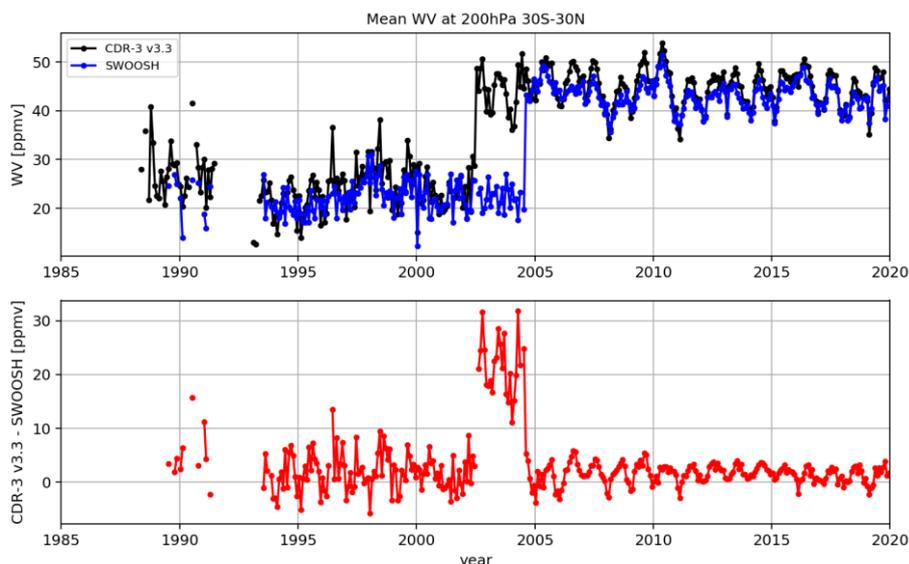
Figure 4-6: Same as Figure 4-5, but for tropical region (30°S–30°N) only.

At 200 hPa (Figure 4-5 and Figure 4-6), the water vapour differences between CDR-3 and SWOOSH exhibit similar values over the global and tropical regions. The differences before 2002 exhibit, however, larger variations and a different seasonality compared to these after 2005. At this level, the transitional period between 2003 and 2005 exhibits a large jump both in the CDR-3 and SWOOSH, with CDR-3 leading by ~2 years over SWOOSH, due to the different input data sets used. This feature is likely related to a sampling issue in the early occultation sounders available before 2005, which will be investigated in more detail in WV_cci Phase 2.

Comparison of WV mixing ratio anomaly time series and their differences

Figure 4-7 and Figure 4-8 show the time series of zonal mean water vapour anomalies relative to the seasonal cycle over the period of 2005–2011 at 100 hPa for CDR-3 v3.3 and SWOOSH in the global and tropical regions, respectively, equivalent to Figures Figure 4-1 and Figure 4-2 for absolute values. For both global and tropical time series, the zonal mean water vapour anomalies from CDR-3 are similar to those from SWOOSH, especially after 2005. Note that the differences in anomaly (CDR-3 minus SWOOSH) change from negative before 2005 to near to zero after 2005, which as mentioned above is due to the difference in the merging algorithm of these two merged data products.
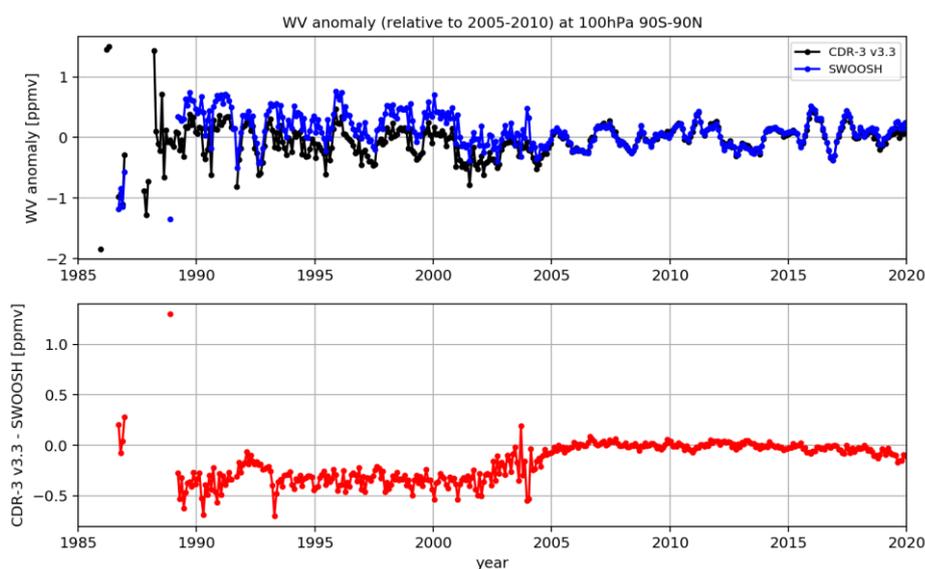
Figure 4-7: Monthly time series (top) and bias (CDR-3–SWOOSH, bottom) of global mean (90°S–90°N) water vapour anomalies of CDR-3 and SWOOSH at 100 hPa from 1985 to 2019.
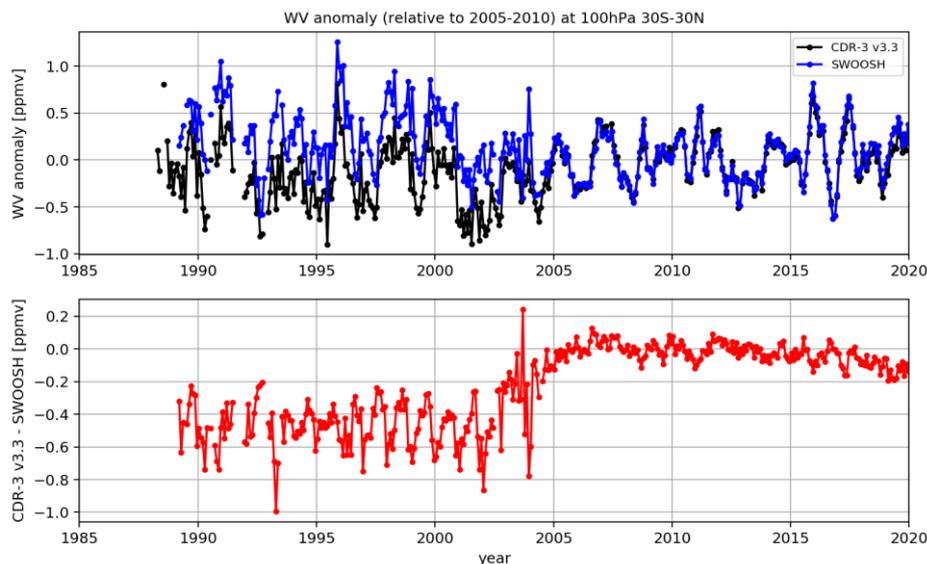
Figure 4-8: Same as Figure 4-7, but for tropical region (30°S–30°N) only.

Figure 4-9 to Figure 4-12 show the time series of water vapour deseasonalised anomalies at 50 and 200 hPa from CDR-3 and SWOOSH; equivalent to Figure 4-3 to Figure 4-6 for absolute values.
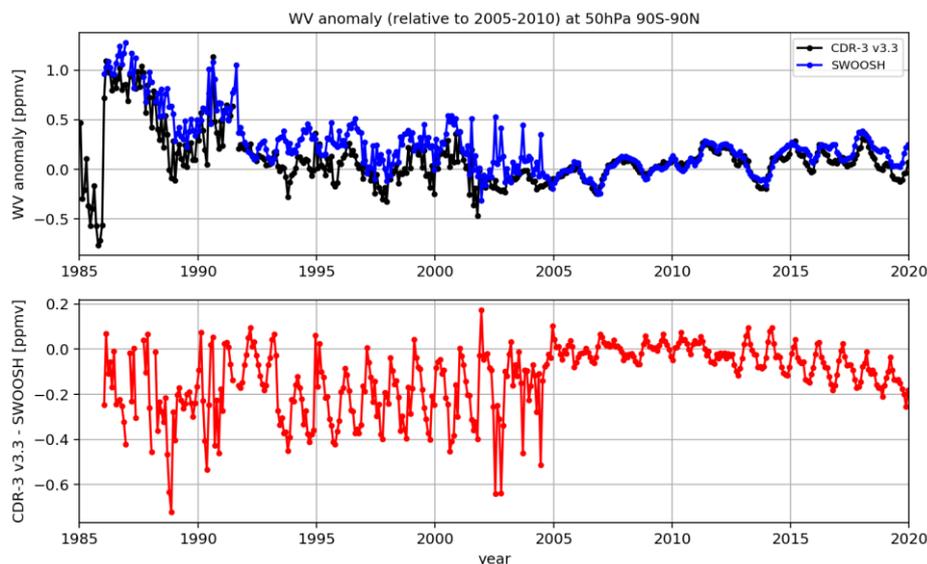
Figure 4-9: Monthly time series (top) and bias (CDR-3–SWOOSH, bottom) of global mean (90°S–90°N) water vapour anomalies of CDR-3 and SWOOSH at 50 hPa from 1985 to 2019.
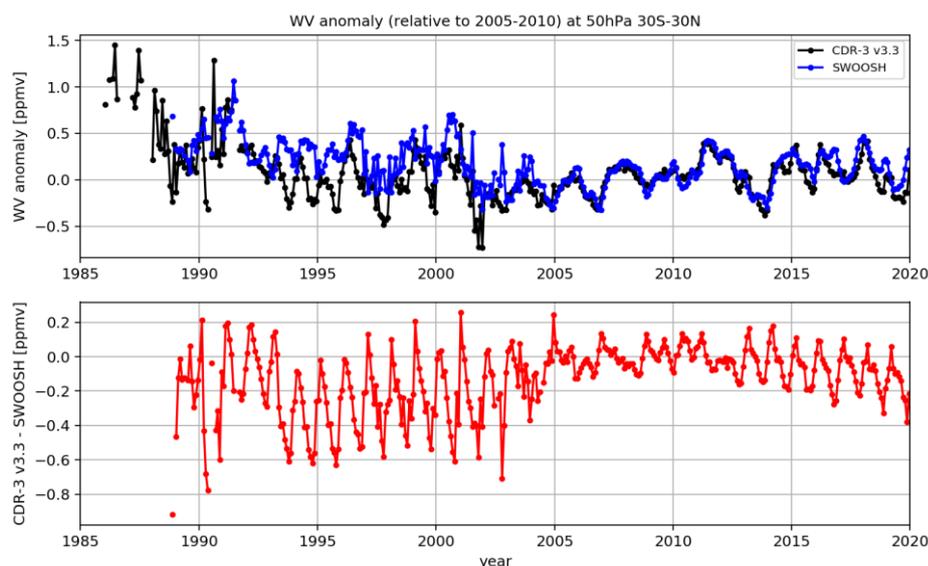
Figure 4-10: Same as Figure 4-9, but for tropical region (30°S–30°N) only.

At 50 hPa (Figure 4-9 and Figure 4-10), the differences in water vapour anomalies between CDR-3 and SWOOSH show similar features as those at 100 hPa over both global and tropical regions, indicating that the difference between CDR-3 and SWOOSH is consistent throughout the lower stratosphere. The anomaly evaluation confirms and accentuates the strong inhomogeneity in the seasonal cycle found in the evaluation of the absolute water vapour mixing ratios before and after the reference period 2005–2010 (see discussion above).

At 200 hPa (Figure 4-11 and Figure 4-12), the differences between CDR-3 and SWOOSH are close to zero for both periods before 2003 and after 2005, which is different from the anomaly differences at 50 and 100 hPa. This indicates that the difference in the two applied merging algorithms has less of an effect at this pressure. As already discussed for Figure 4-5, the two merged products exhibit a strong jump in the transitional period, with CDR-3 leading that of SWOOSH due to the difference in input datasets used. Again, this is likely the result of a sampling bias resulting from the lower coverage of satellite occultation measurements used as input before 2003 and 2005 for CDR-3 and SWOOSH, respectively.
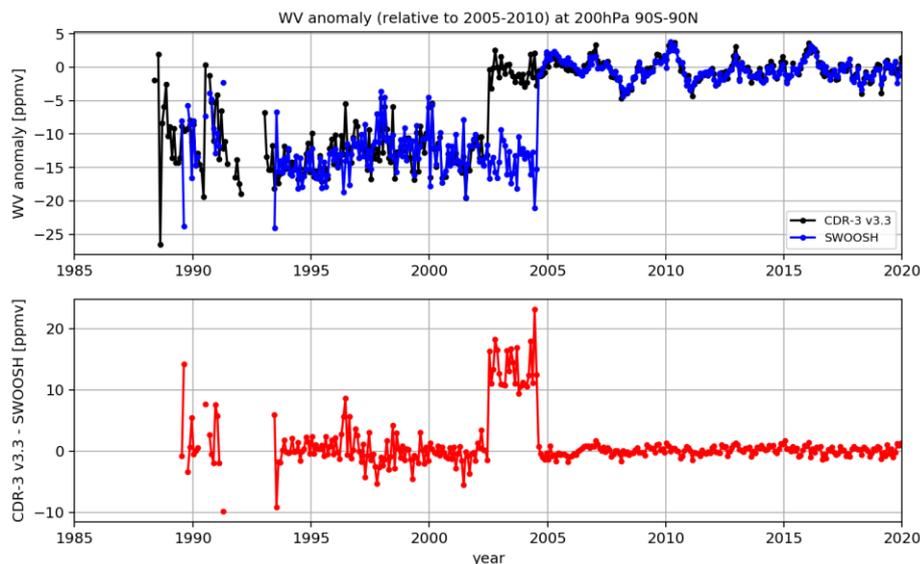
Figure 4-11: Monthly time series (top) and bias (CDR-3–SWOOSH, bottom) of global mean (90°S–90°N) water vapour anomalies of CDR-3 and SWOOSH at 200 hPa from 1985 to 2019.
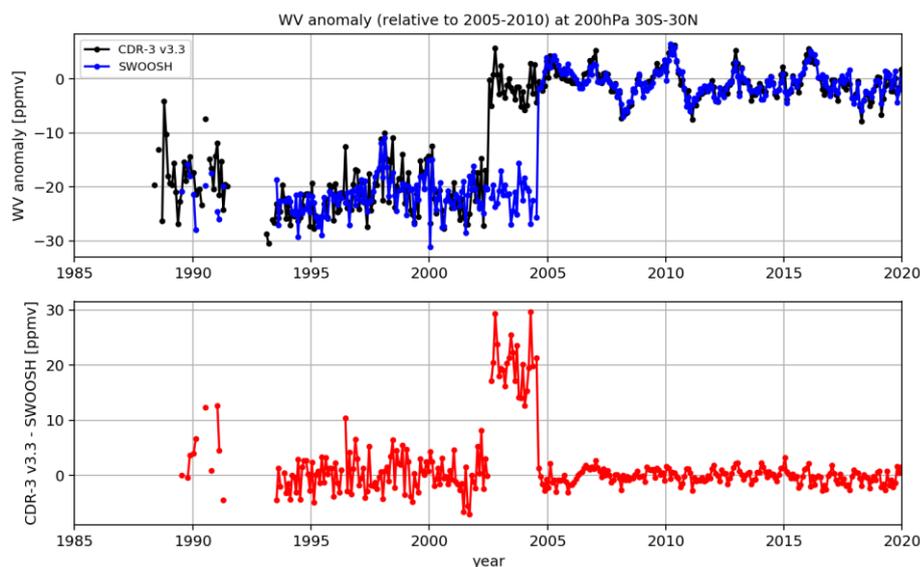
Figure 4-12: Same as Figure 4-11, but for tropical region (30°S–30°N) only.

Time series of VRWV vertical profile anomalies and their differences

Figure 4-13 and Figure 4-14 show the time series of water vapour relative differences and anomaly differences between CDR-3 and SWOOSH data over the global and tropical regions, respectively. In the stratosphere (and in the upper troposphere around 300 hPa), CDR-3 data exhibit less water vapour than SWOOSH while at around 200 hPa, it is the opposite: CDR-3 is more moist than SWOOSH. This behaviour can be explained by a low-bias in Aura-MLS v4.2 (used in SWOOSH) at the 200 hPa level,

which has been improved upon in v5.0 (used in CDR-3), with the consequence that middle stratospheric water vapour in turn decreased slightly between the two data versions due to compensating effects in the retrieval (Lucien Froidevaux, NASA JPL, personal communication).

The water vapour differences between CDR-3 and SWOOSH exhibit similar patterns in the tropics and at the global level, with somewhat stronger features found in the lower stratosphere for the tropics when compared to the global average. Here, the water vapour differences between CDR-3 and SWOOSH exhibit a rapid change in the transitional period before 2005: the large dry differences reduce suddenly. As discussed above, this inhomogeneity will have an influence on the derived trends from these two CDRs. Considering the water vapour anomalies, CDR-3 and SWOOSH are very close in the stratosphere with differences smaller than 0.3 ppmv (hence less than 8%). Similar to the water vapour relative differences, the anomaly differences also show the inhomogeneity in the two periods before and after 2005.
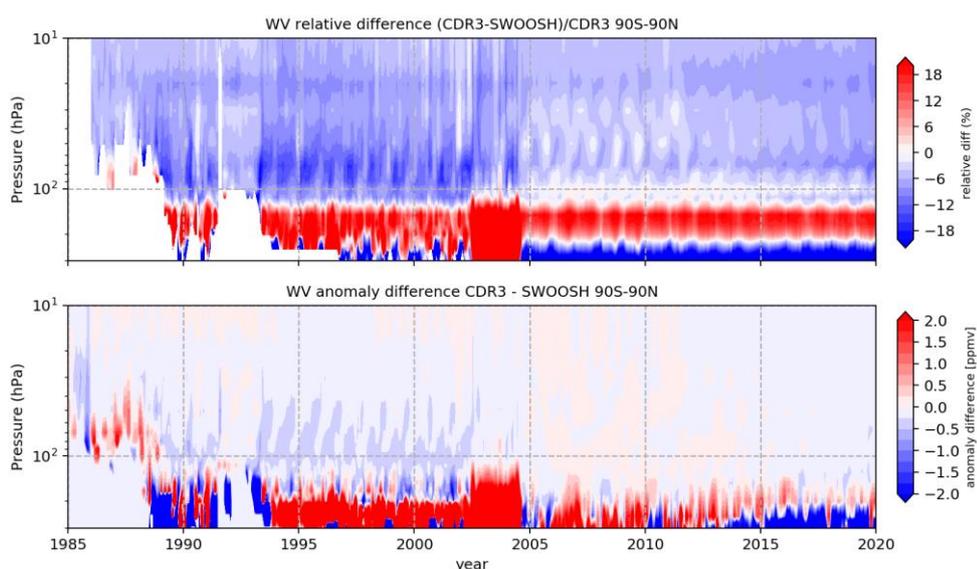
Figure 4-13: Time series of water vapour relative difference (top) and water vapour anomaly difference (bottom) between CDR-3 and SWOOSH from 1985 to 2019. The time series are global averages from 90°S to 90°N.
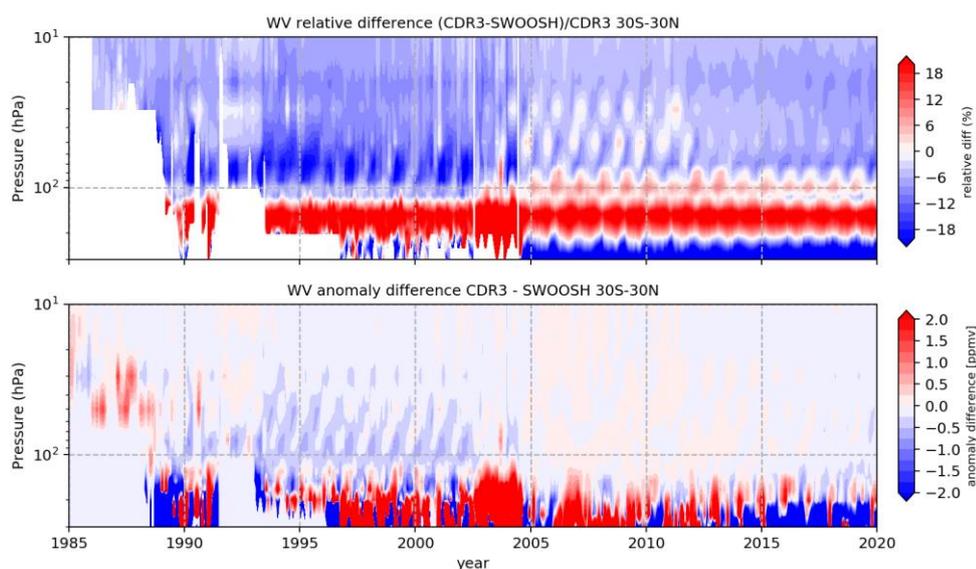
Figure 4-14: Same as Figure 4-13, but for tropical region only from 30°S to 30°N.

4.1.2 Comparison of CDR-3 against CCMI data

The Chemistry Climate Model Initiative (CCMI) is a combined activity of the International Global Atmospheric Chemistry (IGAC) and Stratosphere–troposphere Processes And their Role in Climate (SPARC) projects with multiple chemistry–climate models (CCMs). The CCMI models provide simulation experiments with state-of-knowledge historic forcings and nudged towards reanalysis datasets in the case of reference simulations with specified dynamics (REF-C1SD). These model simulations provide global coverage of water vapour over the period 1984–2014 against which to evaluate the CDR-3 data in the stratosphere. Here, we include the comparisons between CCMI model simulations from REF-C1SD experiment and CDR-3 in the lower stratosphere in two aspects: time series of global or tropical mean water vapour and corresponding deseasonalised anomalies over the period 1985–2014.

Figure 4-15 and Figure 4-16 show the mean water vapour time series and deseasonalised anomalies at 100 hPa from CDR-3 v3.3 and CCMI models over global and tropical regions, respectively. Despite the large differences among water vapour time series from CCMI models, the CDR-3 shows very good agreement with the multi-model mean over both global and tropical regions. For anomalies relative to the seasonal cycle in the period of 2005–2010, CDR-3 has even better agreement with most of the CCMI models, indicating both models and the CDR-3 data capture essential physical features of water vapour variability in the lower stratosphere. Over the global region, the water vapour anomalies from CDR-3 have larger variability in the early years (1985–1995), attributable to the limitation of the spatial coverage of the input satellite

data. After 2005, the anomalies from CDR-3 are very close to the CCMI multi-model mean, because the input data from MIPAS and MLS have large spatial coverage and very good quality.
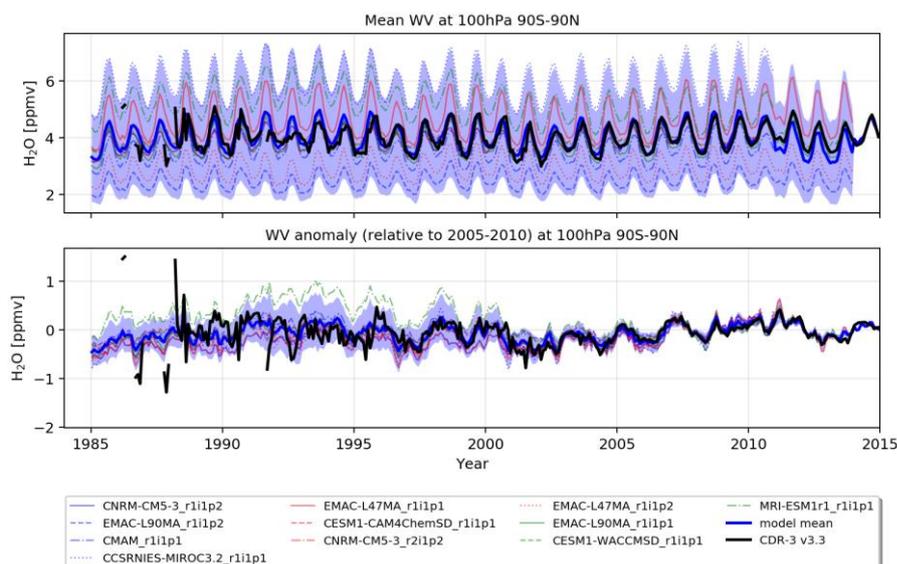
Figure 4-15: Global mean monthly water vapour time series (top) and anomalies (bottom) of CDR-3 (solid black line) and CCMI (blue line for multiple model mean) at 100 hPa from 1985 to 2014. The blue shading shows the range of two standard deviations distance to the CCMI multi-model mean. The deseasonalised anomalies are calculated relative to the seasonal cycle from 2005 to 2010.
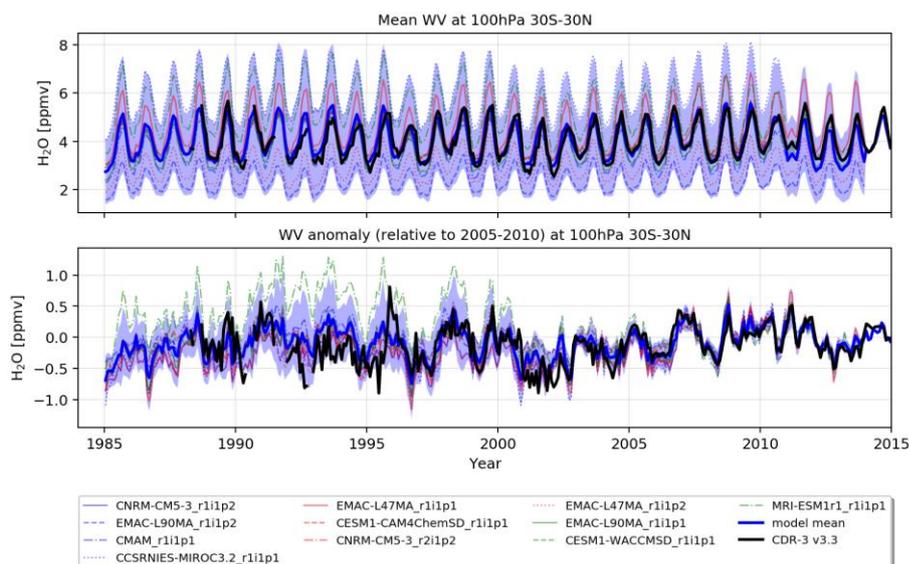
Figure 4-16: Same as Figure 4-15, but for tropical region only from 30°S to 30°N.

Figure 4-17 to Figure 4-20 show the global and tropical mean water vapour time series and deseasonalised anomalies from CDR-3 and the CCMI models at 50 and 200 hPa, respectively. Despite the large differences in mean water vapour, the deseasonalised

anomalies between CDR-3 and CCMI multi-model mean exhibit very good agreement at both 50 and 200 hPa, especially after 2003. Before 2003, the differences in deseasonalised anomalies between CDR-3 and CCMI are large, especially before 1992 at 50 hPa and before 2003 at 200 hPa. As discussed in Section 4.1.1, CDR-3 and SWOOSH have comparable deseasonalised anomalies at these pressure levels, thus the differences between CDR-3 and CCMI are likely due to the limited spatial and temporal coverage of input satellite observations in the early years.

Overall, the CDR-3 v3.3 water vapour deseasonalised anomalies show very good agreement with the CCMI model simulations after 2005, when the high quality and large coverage input data from MIPAS and MLS are incorporated in the merged products. This overall consistency indicates that the CDR-3 data exhibit the essential physical features of water vapour in the lower stratosphere. Before 2005, the quality of the CDR-3 product is limited by the spatial and temporal coverage of the input satellite observations, thus the data availability should be taken in consideration for the comparisons with model simulations.
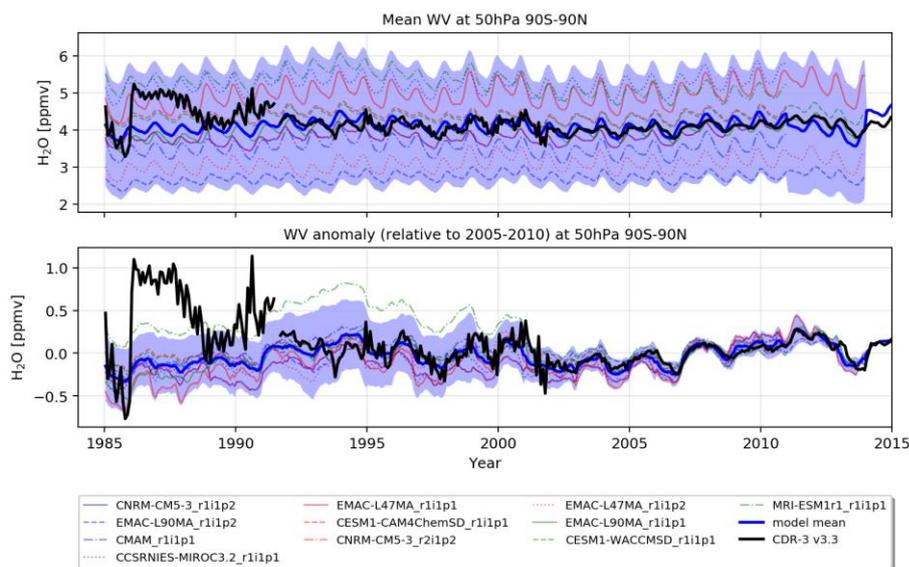
Figure 4-17: Global mean monthly water vapour time series (top) and anomalies (bottom) of CDR-3 (solid black line) and CCMI (blue line for multiple model mean) at 50 hPa from 1985 to 2014.
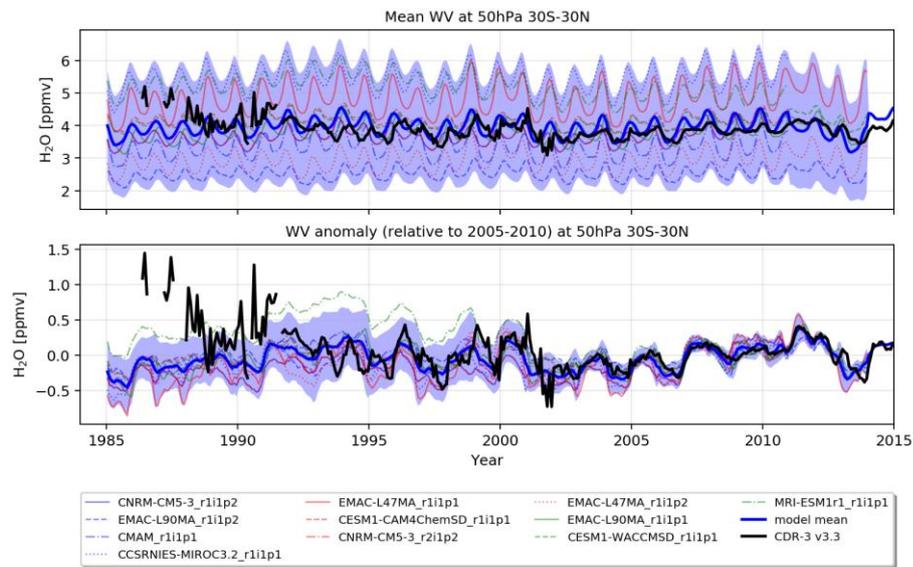
Figure 4-18: Same as Figure 4-17, but for tropical region only from 30°S to 30°N.
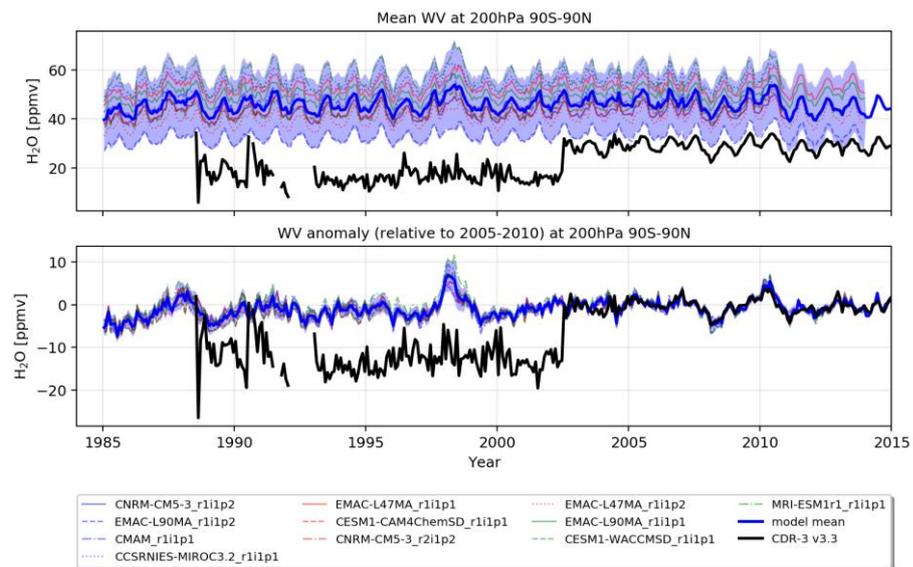
Figure 4-19: Global mean monthly water vapour time series (top) and anomalies (bottom) of CDR-3 (solid black line) and CCMI (blue line for multiple model mean) at 200 hPa from 1985 to 2014.
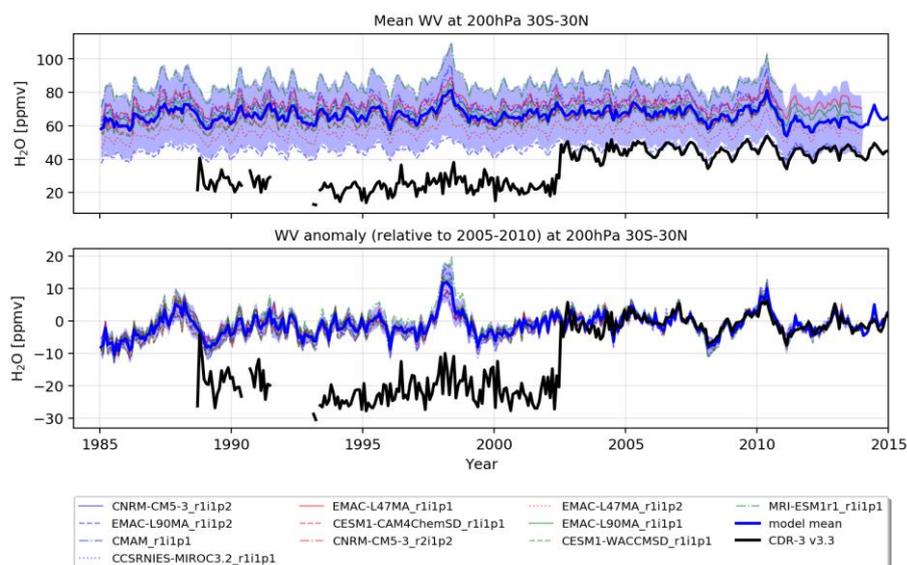
Figure 4-20: Same as Figure 4-19, but for tropical region only from 30°S to 30°N.

Table 4-1 and Table 4-2 summarise the mean and relative differences in water vapour between CDR-3 and CCMI multi-model mean, respectively. At 50 and 100 hPa, the mean difference between CDR-3 and CCMI is very small (within 0.4 ppmv or 10%) over both global and tropical regions. At 200 hPa, the water vapour from CCMI is much higher than CDR-3, with the relative difference as high as 176%. During the three periods, the differences between 2004 and 2014 are smallest, related to a strong jump in the differences around 2002 (which likely stems from a sampling issue as will be investigated in the next phase of WV_cci).

Table 4-1: Mean differences in water vapour between CDR-3 and CCMI multi-model mean (MMM–CDR-3) with one standard deviation added

Pressure (hPa)	CCMI–CDR-3 (ppmv)					
	Global			Tropical		
	1985–1992	1993–2003	2004–2014	1985–1992	1993–2003	2004–2014
50	−0.40 ± 0.45	0.05 ± 0.18	0.06 ± 0.11	−0.30 ± 0.49	0.20 ± 0.31	0.19 ± 0.18
100	0.03 ± 0.53	0.10 ± 0.22	0.08 ± 0.11	0.02 ± 0.34	0.09 ± 0.35	−0.00 ± 0.18
200	26.54 ± 5.92	28.44 ± 5.20	17.96 ± 1.24	37.73 ± 6.15	40.55 ± 8.20	22.60 ± 2.58

Table 4-2: Mean relative differences in water vapour between CDR-3 and CCMI multi-model mean (MMM–CDR-3)/CDR-3 with one standard deviation added

Pressure (hPa)	(CCMI–CDR-3)/CDR-3 (%)					
	Global			Tropical		
	1985–1992	1993–2003	2004–2014	1985–1992	1993–2003	2004–2014
50	-8.16 ± 9.52	1.42 ± 4.54	1.40 ± 2.75	-6.14 ± 10.47	5.54 ± 8.58	4.97 ± 4.83
100	2.17 ± 14.88	2.90 ± 5.62	1.93 ± 2.58	1.08 ± 8.40	2.87 ± 8.81	0.15 ± 4.24
200	175.44 ± 123.07	176.44 ± 63.08	62.05 ± 7.09	144.87 ± 42.50	173.72 ± 71.60	50.40 ± 9.11

4.1.3 Comparison of CDR-3 against reanalysis datasets

In this section, we present the comparisons of CDR-3 against reanalysis datasets in the period 1985–2014. The water vapour data from CDR-3 are compared with four reanalysis datasets: ERAi, ERA5, MERRA and MERRA-2 in two aspects: time series of mean water vapour and deseasonalised anomalies over the global and tropical regions.

Figure 4-21 and Figure 4-22 show the time series of mean water vapour and deseasonalised anomalies at 100 hPa from CDR-3 and reanalysis datasets over global and tropical regions, respectively. For mean water vapour, the CDR-3 is closer to ERAi/ERA5 than MERRA/MERRA-2, especially after 2000. However, for the tropical region, the time series of mean water vapour between CDR-3 and all the reanalysis datasets show very good agreement.

For anomalies relative to the seasonal cycle in the period 2005–2010, the differences between CDR-3 and the reanalysis datasets are consequently quite small (mostly within ±0.25 ppmv) after 2003 over both global and tropical regions. Note that after 2003, the global mean anomalies from CDR-3 are closer to ERAi/ERA5 than to MERRA/MERRA-2.

Before 2003, the CDR-3 and reanalysis datasets have much more variations over global and tropical regions. The MERRA and MERRA-2 exhibit almost the same deseasonalised anomalies over both regions. However, the ERA5 exhibits larger anomalies than CDR-3 and the other reanalysis datasets, especially in the tropics. Note that the differences between ERAi and ERA5 become larger towards the early part of the record, with the differences between ERA5 and MERRA/MERRA-2 here becoming

smaller. The strong drop in water vapour around 2001 seen in CDR-3 and investigated widely in the literature is closest in magnitude to that found in ERA5.
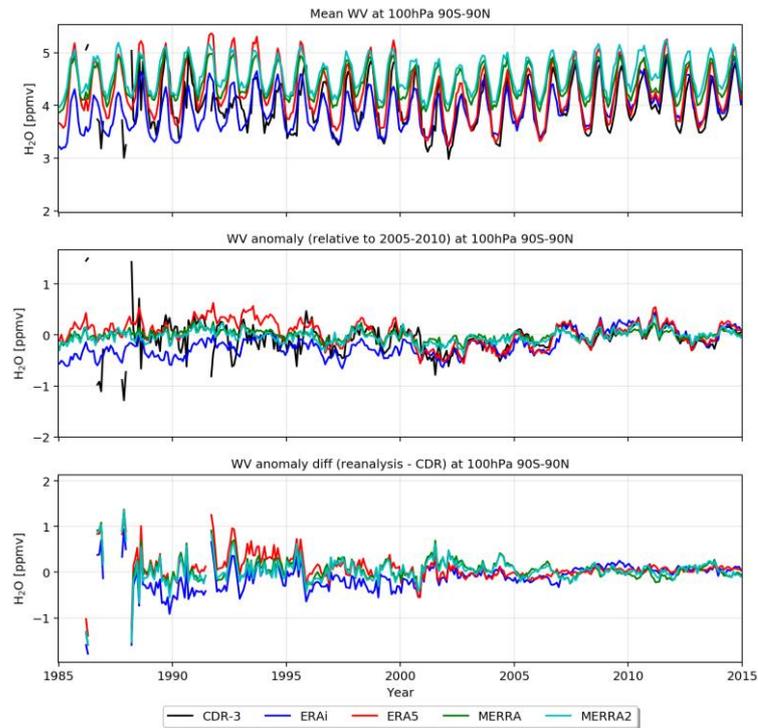
Figure 4-21: Global mean monthly water vapour time series (top), anomalies (middle) and anomaly differences (bottom) of CDR-3 (solid black line) and reanalysis datasets at 100 hPa from 1985 to 2014.
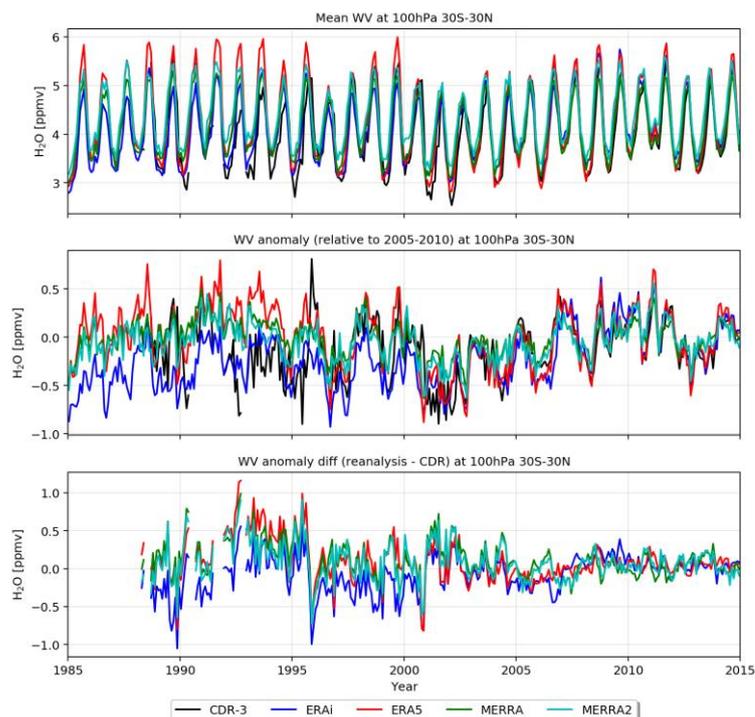
Figure 4-22: Same as Figure 4-21, but for tropical region only from 30°S to 30°N.

Figure 4-23 and Figure 4-24 show the mean water vapour time series and anomalies from CDR-3 and reanalysis datasets at 50 hPa over global and tropical regions, respectively. Similar to 100 hPa, the mean MERRA/MERRA-2 are more moist than CDR-3, ERAi, and most of ERA5 (except the early years) over both global and tropical regions. However, the anomalies from MERRA/MERRA-2 have much smaller interannual variations, especially for MERRA whose anomalies are almost zero. This is because the MERRA/MERRA-2 are relaxed to a 2-D monthly climatological moisture field derived from water vapor observations from UARS HALOE and Aura MLS in the stratosphere, thus the MERRA/MERRA-2 miss most of the physical interannual variations (Davis et al., 2017 [RD-10]).

CDR-3 mean water vapour and anomalies are both closer to ERA5 than ERAi after 2003. However, before 2003, CDR-3 mean water vapour values lie somewhere in between the strongly differing ERA5 and ERAi data. In particular, ERA5 has much higher mean water vapour and anomalies than ERAi. Similar to 100 hPa, although to a somewhat later point in time, the ERA5 shows a rapid drop in water vapour at 50 hPa around 2002, which is most similar to the drop found in CDR-3 compared to other reanalyses.
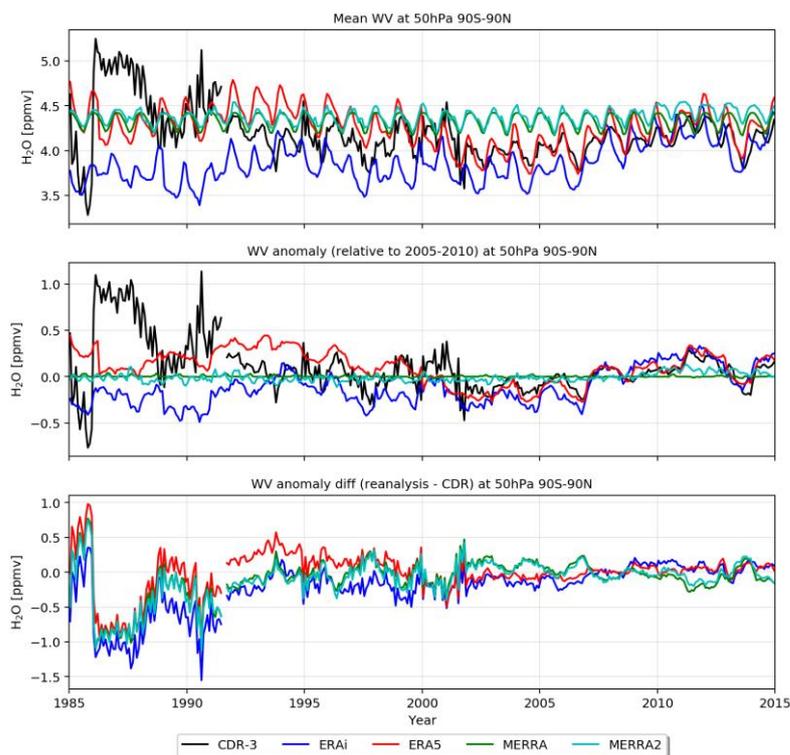
Figure 4-23: Global mean monthly water vapour time series (top), anomalies (middle) and anomaly differences (bottom) of CDR-3 (solid black line) and reanalysis datasets at 50 hPa from 1985 to 2014.
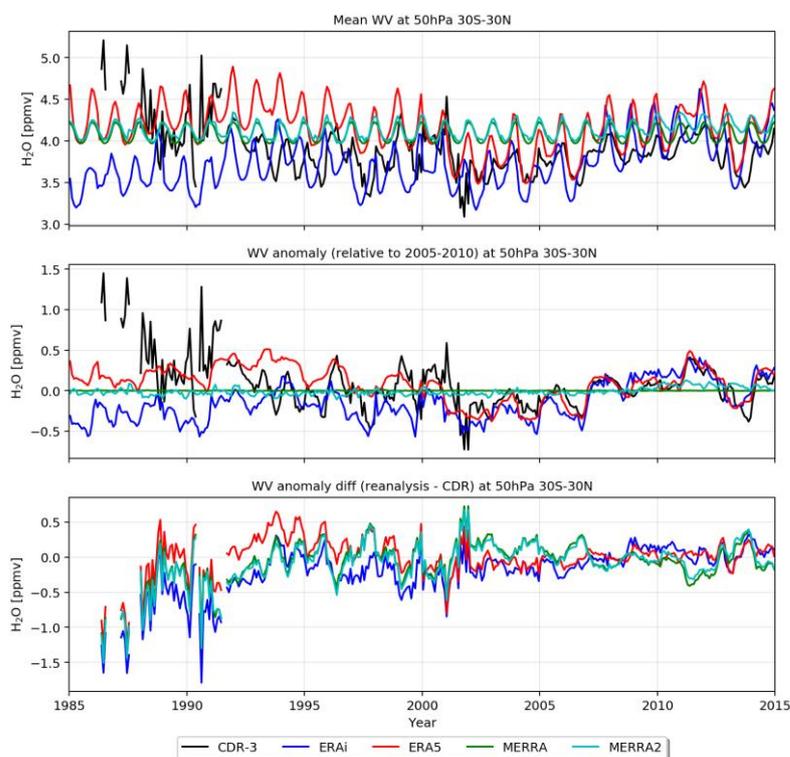
Figure 4-24: Same as Figure 4-23, but for tropical region only from 30°S to 30°N.

At 200 hPa, the comparisons between CDR-3 and reanalysis datasets (Figure 4-25 and) show that the global and tropical mean water vapour from CDR-3 are much lower than all the reanalysis datasets. However, the deseasonalised anomalies are very close between CDR-3 and the reanalysis datasets after 2003 over both global and tropical regions. Before 2003, all reanalysis datasets have very similar deseasonalised anomalies. As discussed above, the large difference between CDR-3 and reanalysis datasets in the early years at 200 hPa is likely the result of a sampling bias from the low coverage of satellite occultation measurements used as input.
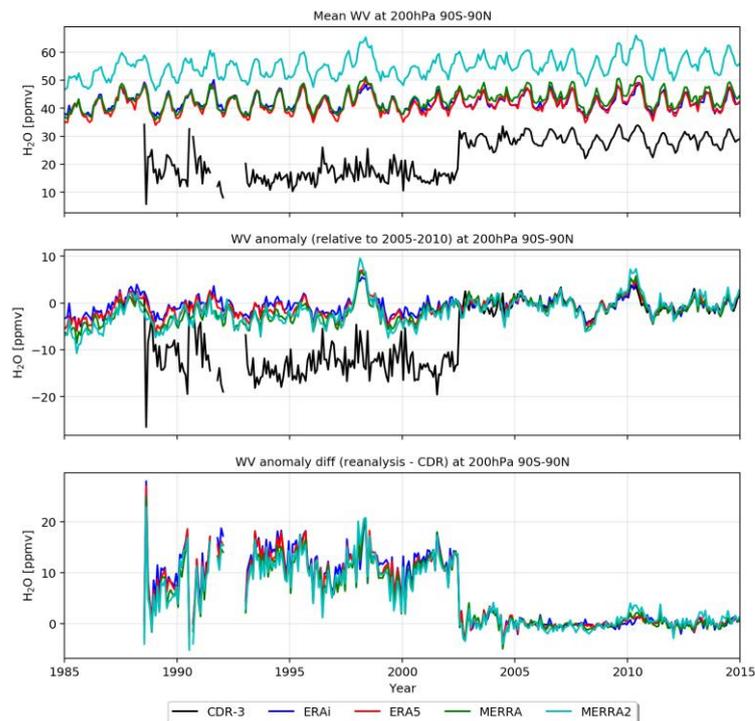
Figure 4-25: Global mean monthly water vapour time series (top), anomalies (middle) and anomaly differences (bottom) of CDR-3 (solid black line) and reanalysis datasets at 200 hPa from 1985 to 2014.
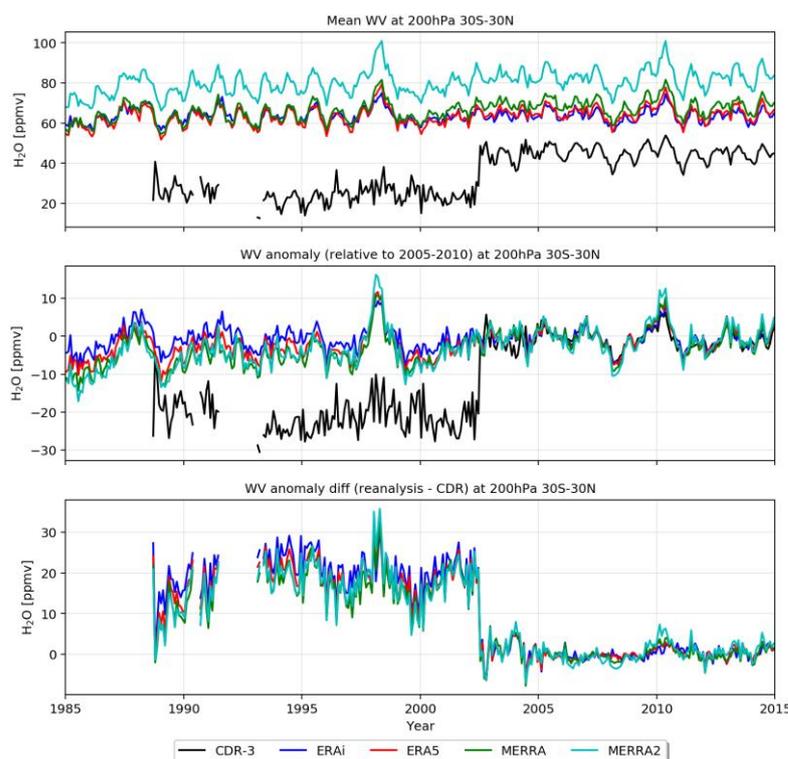
Figure 4-26: Same as Figure 4-25, but for tropical region only from 30°S to 30°N.

The mean differences and relative differences in water vapour between CDR-3 and reanalysis datasets are summarised in Table 4-3 and Table 4-4. Due to the input data availability discussed above, the comparison is divided into three periods: 1985–1992, 1993–2003 and 2004–2014. In each period, the mean differences and their standard deviations between reanalysis datasets (ERAi/ ERA5/ MERRA/ MERRA2) and CDR-3 are calculated over global and tropical regions at 50, 100, and 200 hPa.

At 50 hPa, the maximum differences in Table 4-3 are observed between ERAi and CDR-3 during 1985–1992 (-0.72 ppmv globally and -0.64 ppmv over tropics), where the ERAi is the driest reanalysis dataset (see Figure 4-23 and Figure 4-24). Almost all reanalysis datasets have differences changing from negative to positive after 1992 and the absolute differences decrease over the period 1993–2003 and 2004–2014 (0.16 ppmv to 0.09 ppmv between ERA5 and CDR-3 globally), indicating that CDR-3 has better agreement with reanalysis datasets in water vapour anomaly when more high-quality satellite observations are merged into the product and also when assimilated data into reanalyses get better and more frequent. Note that during 1985–1992, the analysis datasets and CDR-3 have small absolute differences but large standard deviations due to a compensating effect between large positive and large negative differences in these early years. The relative differences in Table 4-4 quantify these differences between reanalysis datasets and CDR-3 further, indicating best agreement

41

between CDR-3 and ERA5 in terms of smallest relative differences along with smallest standard deviations (2.07±2.67% for the global comparison and 2004–2014).

At 100 hPa, the mean differences and relative differences in water vapour between reanalysis datasets and CDR-3 are again larger during the period of 1985–1992 and decrease in the later periods, especially for ERA5 and ERAi. Unlike at 50 hPa, all reanalysis datasets, except ERAi, have more water vapour than CDR-3 globally and over the tropics, while the ERAi has smallest differences compared to CDR-3. The tropical region has smaller differences than the whole global mean, however, the corresponding standard deviations over the tropical region are larger than the global mean during the latter two periods.

Unlike the above two pressure levels, the mean differences in water vapour between reanalysis datasets and CDR-3 are much larger, with the relative differences more than 130% during the first two periods. In the period of 2004–2014, the mean differences are much smaller but still have relative differences larger than 40%. As discussed above, this change in mean and relative differences between CDR-3 and reanalysis datasets is likely the results of a sampling bias from the low coverage of satellite occultation measurements in the early years.

Table 4-3: Mean differences in water vapour between CDR-3 and reanalysis datasets (reanalysis–CDR-3) with one standard deviation added

Pressure (hPa)	Difference (*XXX*–CDR) (ppmv)	Global			Tropical		
		1985–1992	1993–2003	2004–2014	1985–1992	1993–2003	2004–2014
50	ERAi	−0.72 ± 0.45	−0.27 ± 0.21	−0.11 ± 0.18	−0.64 ± 0.50	−0.12 ± 0.34	0.05 ± 0.27
	ERA5	−0.08 ± 0.51	0.16 ± 0.22	0.09 ± 0.11	0.04 ± 0.48	0.34 ± 0.31	0.25 ± 0.18
	MERRA	−0.12 ± 0.43	0.24 ± 0.16	0.22 ± 0.13	−0.17 ± 0.40	0.30 ± 0.23	0.25 ± 0.18
	MERRA2	−0.08 ± 0.43	0.27 ± 0.16	0.30 ± 0.11	−0.13 ± 0.39	0.33 ± 0.24	0.33 ± 0.16
100	ERAi	−0.21 ± 0.51	−0.08 ± 0.26	0.08 ± 0.17	−0.07 ± 0.36	0.04 ± 0.37	0.15 ± 0.22
	ERA5	0.37 ± 0.56	0.26 ± 0.27	0.18 ± 0.12	0.34 ± 0.41	0.27 ± 0.40	0.14 ± 0.19
	MERRA	0.47 ± 0.51	0.48 ± 0.27	0.40 ± 0.20	0.24 ± 0.38	0.18 ± 0.37	0.04 ± 0.24
	MERRA2	0.58 ± 0.51	0.60 ± 0.29	0.58 ± 0.22	0.42 ± 0.43	0.38 ± 0.40	0.30 ± 0.27
200	ERAi	23.24 ± 5.88	24.19 ± 5.34	13.46 ± 1.15	35.82 ± 5.75	36.96 ± 8.46	18.92 ± 2.07
	ERA5	22.15 ± 5.94	23.52 ± 5.27	13.39 ± 1.07	34.28 ± 6.09	36.72 ± 7.94	20.37 ± 1.92

Pressure (hPa)	Difference (*XXX*–CDR) (ppmv)	Global			Tropical		
		1985–1992	1993–2003	2004–2014	1985–1992	1993–2003	2004–2014
	MERRA	22.81 ± 5.87	25.06 ± 4.97	15.47 ± 1.30	35.54 ± 5.98	39.52 ± 7.52	24.01 ± 2.09
	MERRA2	34.53 ± 6.18	36.76 ± 5.31	27.63 ± 1.82	49.64 ± 6.92	53.72 ± 8.28	38.03 ± 2.95

Table 4-4: Mean relative differences in water vapour between CDR-3 and reanalysis datasets (reanalysis–CDR-3)/CDR-3 with one standard deviation added

Pressure (hPa)	Relative Difference (%)	Global			Tropical		
		1985–1992	1993–2003	2004–2014	1985–1992	1993–2003	2004–2014
50	ERAi	−15.31 ± 9.24	−6.39 ± 5.05	−2.58 ± 4.48	−14.31 ± 10.27	−2.73 ± 9.11	1.27 ± 7.24
	ERA5	−0.76 ± 11.89	3.90 ± 5.46	2.07 ± 2.67	1.88 ± 11.00	9.31 ± 8.71	6.50 ± 4.74
	MERRA	−1.67 ± 10.25	6.11 ± 4.25	5.40 ± 3.31	−3.14 ± 8.74	8.40 ± 6.76	6.68 ± 4.97
	MERRA2	−0.77 ± 10.20	6.79 ± 4.29	7.35 ± 2.85	−2.40 ± 8.66	8.99 ± 6.87	8.72 ± 4.61
100	ERAi	−3.94 ± 13.66	−1.72 ± 6.28	2.00 ± 4.16	−0.93 ± 8.54	2.01 ± 9.38	3.94 ± 5.38
	ERA5	10.76 ± 16.86	6.80 ± 6.91	4.29 ± 2.96	8.90 ± 10.32	7.18 ± 10.06	3.14 ± 4.50
	MERRA	13.58 ± 16.03	12.93 ± 8.22	10.47 ± 5.95	7.35 ± 10.10	5.77 ± 10.18	1.63 ± 5.88
	MERRA2	16.35 ± 16.48	16.03 ± 8.77	14.77 ± 6.57	12.23 ± 12.00	11.16 ± 11.38	7.99 ± 7.23
200	ERAi	155.66 ± 117.76	151.33 ± 58.49	46.99 ± 6.23	137.61 ± 40.42	159.58 ± 69.40	42.70 ± 7.78
	ERA5	148.17 ± 113.46	146.62 ± 55.94	46.65 ± 5.10	131.41 ± 38.92	157.37 ± 65.40	45.86 ± 7.03
	MERRA	152.50 ± 116.06	155.63 ± 57.02	53.91 ± 6.24	136.10 ± 39.46	168.29 ± 66.64	53.98 ± 7.52
	MERRA2	224.29 ± 143.40	225.93 ± 74.19	96.18 ± 8.92	189.50 ± 50.74	226.68 ± 83.77	85.33 ± 9.62

When looking at the mean differences and standard deviations for the anomalies (which largely removes the mean bias between the reanalyses and CDR-3, shown in Table 4-5), a very high agreement is found between reanalyses and CDR-3. Larger differences are found in the early years, decreasing in the latter two time periods. The values among reanalyses are very close, because after moving the mean annual seasonal cycle, the absolute differences in WV anomalies are similar among the reanalysis datasets. ERAi and ERA5 agree generally better with CDR-3 than MERRA and MERRA2, except perhaps at 50 hPa during the two earlier periods.

Table 4-5: Mean absolute WV anomaly differences between CDR-3 and reanalysis datasets (reanalysis–CDR-3) and one standard deviation

Pressure (hPa)	Absolute Difference (|*XXX*–CDR|) (ppmv)	Global			Tropical		
		1985-1992	1993-2003	2004-2014	1985-1992	1993-2003	2004-2014
50	ERAi	0.33 ± 0.22	0.11 ± 0.09	0.10 ± 0.07	0.30 ± 0.19	0.15 ± 0.11	0.12 ± 0.08
	ERA5	0.40 ± 0.27	0.16 ± 0.10	0.06 ± 0.04	0.30 ± 0.21	0.19 ± 0.12	0.08 ± 0.06
	MERRA	0.33 ± 0.25	0.12 ± 0.08	0.10 ± 0.07	0.26 ± 0.20	0.16 ± 0.11	0.14 ± 0.10
	MERRA2	0.33 ± 0.24	0.26 ± 0.08	0.08 ± 0.06	0.26 ± 0.20	0.16 ± 0.11	0.13 ± 0.09
100	ERAi	0.35 ± 0.29	0.15 ± 0.11	0.10 ± 0.07	0.20 ± 0.20	0.20 ± 0.16	0.12 ± 0.10
	ERA5	0.36 ± 0.29	0.16 ± 0.13	0.06 ± 0.05	0.23 ± 0.21	0.23 ± 0.17	0.08 ± 0.07
	MERRA	0.36 ± 0.29	0.15 ± 0.11	0.10 ± 0.07	0.23 ± 0.20	0.20 ± 0.15	0.11 ± 0.08
	MERRA2	0.36 ± 0.30	0.14 ± 0.11	0.09 ± 0.07	0.25 ± 0.20	0.19 ± 0.15	0.11 ± 0.08
200	ERAi	3.76 ± 2.64	3.95 ± 3.25	0.67 ± 0.57	3.46 ± 2.75	6.08 ± 5.27	1.23 ± 0.99
	ERA5	3.76 ± 2.47	3.85 ± 3.16	0.62 ± 0.57	3.49 ± 2.81	5.81 ± 4.85	1.23 ± 1.02
	MERRA	3.70 ± 2.44	3.74 ± 2.92	0.76 ± 0.66	3.37 ± 2.76	5.50 ± 4.54	1.38 ± 1.16
	MERRA2	3.93 ± 2.45	3.89 ± 3.12	1.08 ± 0.90	3.95 ± 2.86	5.91 ± 4.89	1.89 ± 1.60

4.2 Validation and comparison of monthly VRWV profile (CDR-4)

The validation and comparison of the final level 3 CDR-4 is carried out against water vapour data records from BBH sites and reanalyses (ERA-5 and MERRA2). The

prototype CDR-4 v3 data covers global regions from 1000 to 10 hPa in the time period 2010–2014.

4.2.1 Validation of CDR-4 data at BBH sites

The BBH data records include frost-point hydrometer (FPH) data from NOAA and cryogenic frost-point hygrometer (CFH) data from NCAR, which provide water vapour profiles with high vertical resolution at the BBH sites. Due to the limited temporal (less than one profile in a month) and spatial (only 28 BBH sites distributed over the world) coverage of the BBH data, the CDR-4 data are compared with the water vapour profiles at the BBH sites. Here, CDR-4 data are chosen to be closest to the location of the BBH sites for validation. All water vapour observations from BBH profiles within a certain pressure level interval (between pressure level midpoints) are plotted along with the time series from CDR-4. Beside CDR-4, the merged data without bias-correction procedure are also plotted.

Figure 4-27 shows the time series of water vapour from CDR-4 and uncorrected merged data along with the *in situ* observations at Boulder site for different pressure levels. Compared to the uncorrected data, CDR-4 has higher water vapour concentrations during the summer season, which captures the high water vapour observed from the hygrometer soundings, especially between 150 and 250 hPa in 2010 and 2011. During the winter season, the differences between CDR-4 and uncorrected data are small, and both data are consistent with the soundings. Note that WV profiles at the Boulder site do not include similarly high water vapour values during 2012–2014 during the summer season, different from 2010–2011. As will be discussed later, this likely reflects that the BBH data are not truly representative for the mean water vapour value of a given month, and thus that the validation has to be interpreted with care.

Over tropical regions, two BBH sites (San Jose and Hilo) have a substantial number of profiles in 2010–2014. Figure 4-28 to Figure 4-29 show the time series of water vapour from CDR-4 and uncorrected data at these two BBH sites. Similar to Boulder above, the CDR-4 data significantly increases the water vapour concentration during the summer season in the UTLS region, consistent with the BBH observations at corresponding sites, especially at San Jose. Again, the differences between CDR-4 and uncorrected data are small in the winter season when the water vapour concentrations are low.

Over northern mid- to high-latitude sites in Europe (Lindenberg and Sodankylä) shown in Figure 4-30 and Figure 4-31, the CDR-4 data are close to the uncorrected data, indicating the bias-correction has little effect on the satellite observations at these sites.

Unlike at Boulder, the bias-corrected CDR-4 data miss the high WV values at the Lindenberg site in the summer season. As mentioned above, this can likely be attributed to a sampling bias from the sparse BBH observations. At Sodankylä, both the CDR-4 and uncorrected data are consistent with BBH observations, which is expected as only this site is available for the bias-correction procedure in the high latitude region.
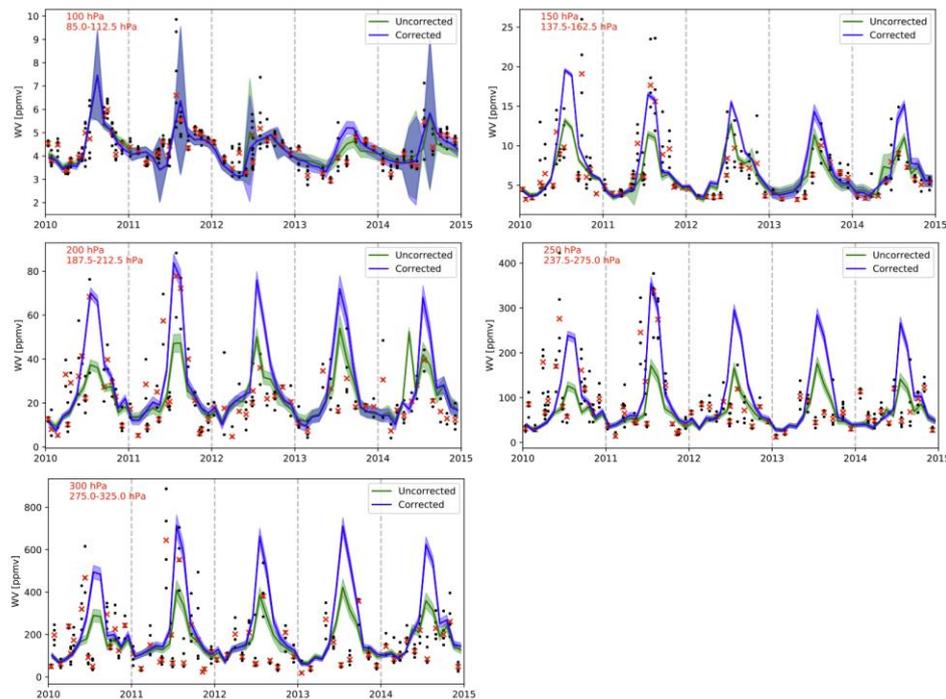
Figure 4-27: Monthly time series of water vapour with (blue) and without (green) bias-correction from CDR-4 over the period 2010–2014 at BBH site (black dots are BBH observations and red crosses median values) for Boulder, CO (40°N/105°W) at pressure levels from 100 to 300 hPa. The green and blue shadings show the range of two standard deviations distance to the initial (green) and bias-corrected (blue) WV from CDR-4, respectively.
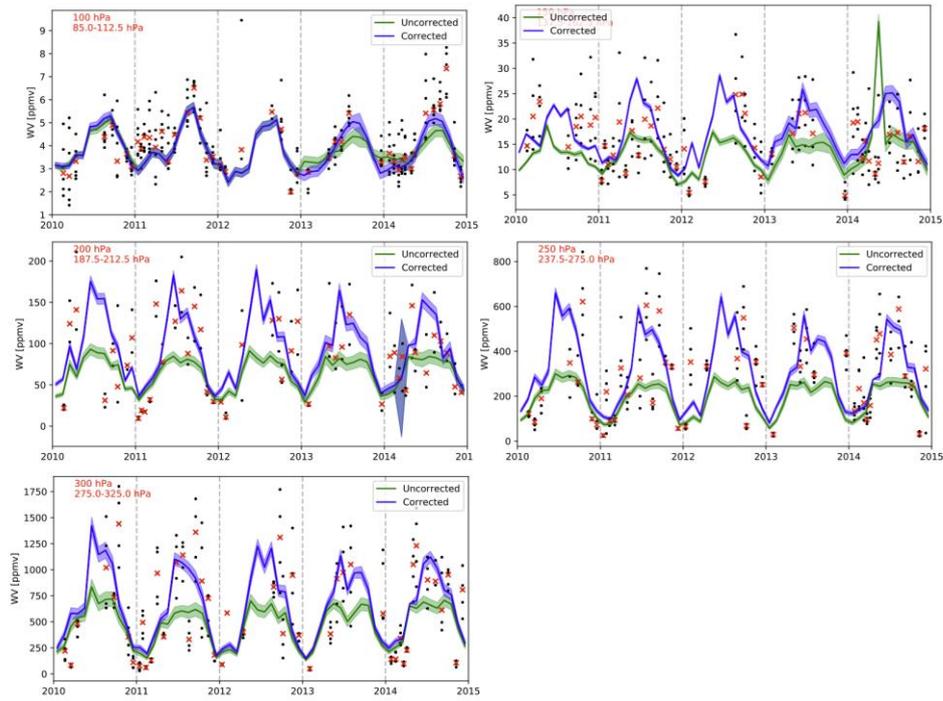
Figure 4-28: Same as Figure 4-27, but at BBH site San Jose (9°N/85°W).
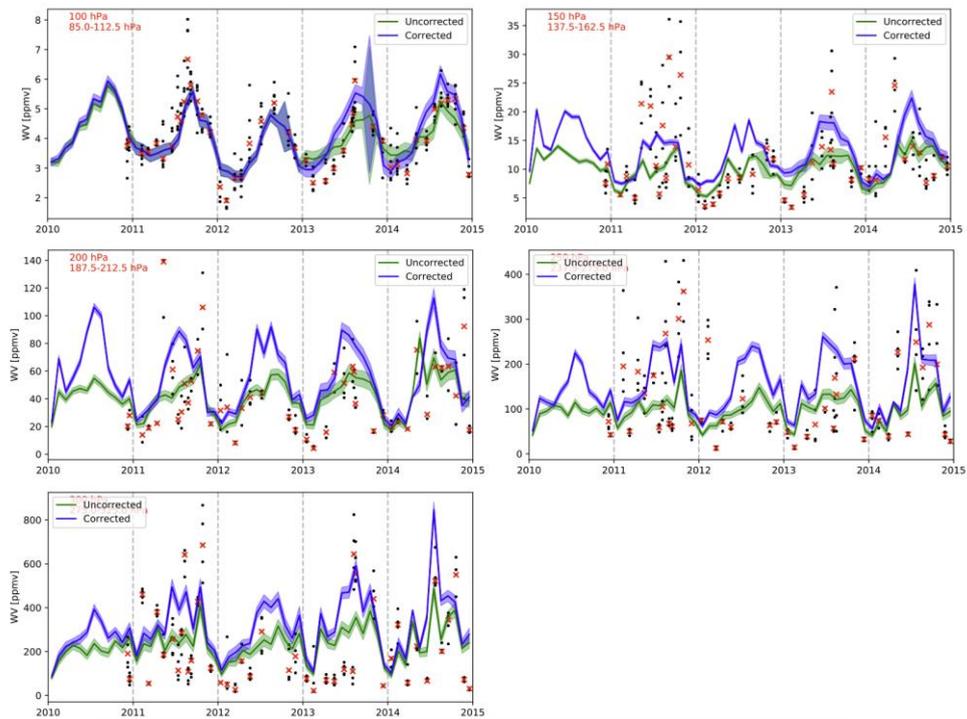
Figure 4-29: Same as Figure 4-27, but at BBH site Hilo, HI (19°N/156°W).
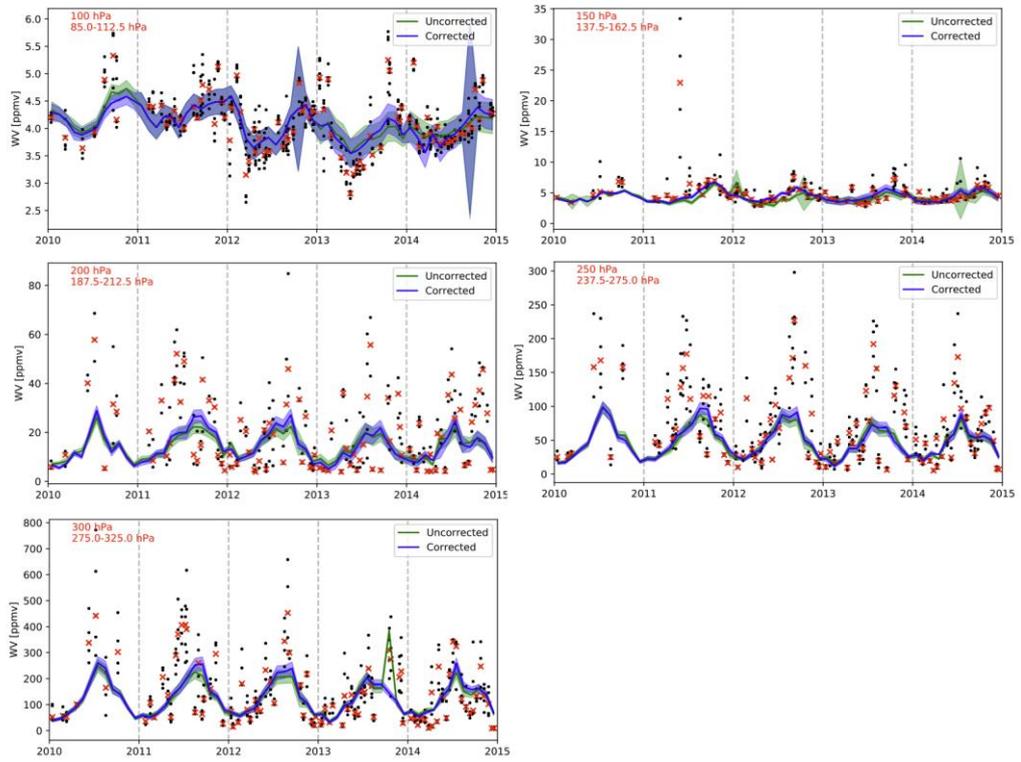
Figure 4-30: Same as Figure 4-27, but at BBH site Lindenberg, Germany (52°N/14°E).
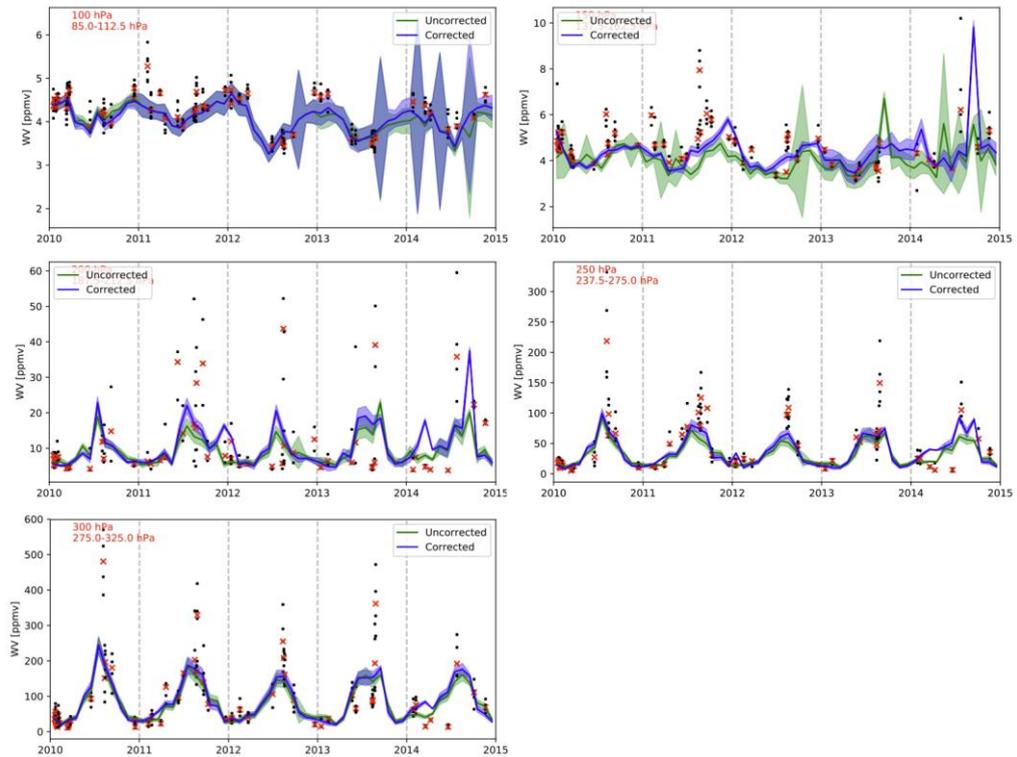
Figure 4-31: Same as Figure 4-27, but at BBH site Sodankylä, Finland (67°N/26°E).

4.2.2 Comparison of monthly CDR-4 data against reanalyses

The comparison of CDR-4 against reanalyses (ERA-5 and MERRA-2) is performed for the period 2010–2014. The water vapour from reanalysis datasets is interpolated horizontally and vertically onto the horizontal grid points and pressure levels of CDR-4. The horizontal distributions of annual mean water vapour averaged over the period 2010–2014 at different pressure levels are shown in Figure 4-32. The bias-correction method within the CDR-4 product processing increased the water vapour in the upper troposphere, which is known to be too dry in the original satellite observations according to our early validation results (see PVASR [RD-10]). Compared with reanalyses, CDR-4 shows the same geographical pattern, as the bias-correction procedure with the quantile-mapping technique is able to preserve the distribution of the water vapour observations.
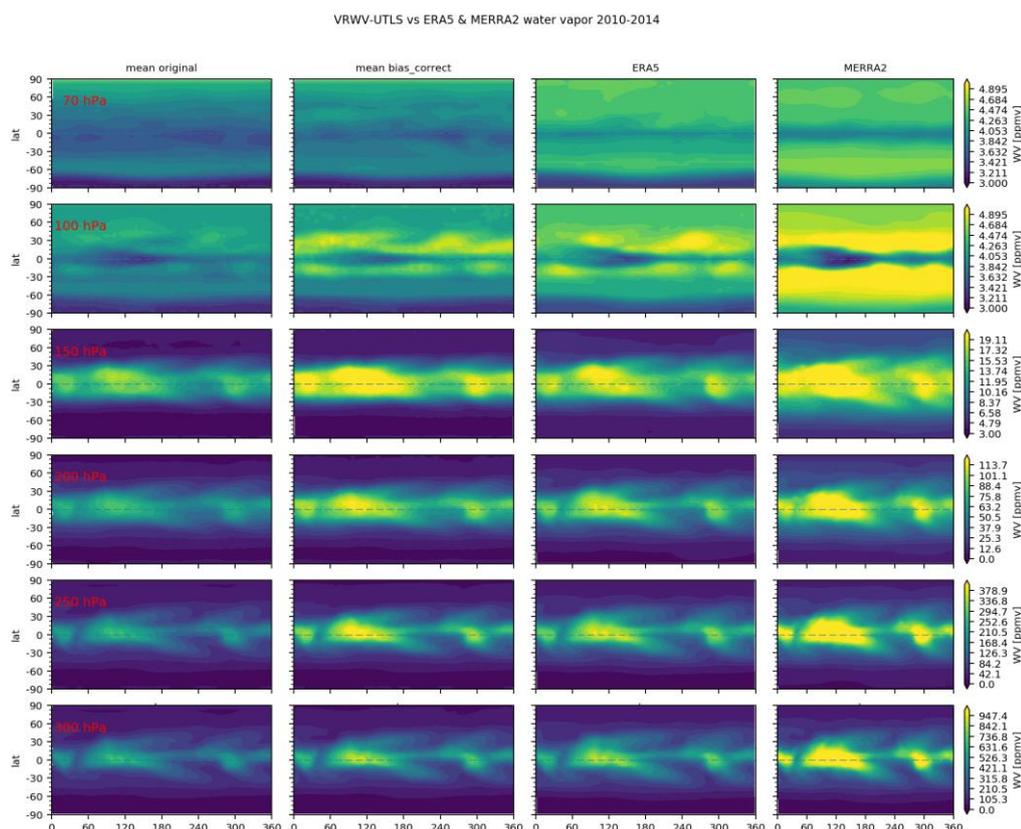
Figure 4-32: Water vapour longitude–latitude distributions at different pressure levels; from top to bottom: 70, 100, 150, 200, 250 hPa and 300 hPa. From left to right, the data are from: original satellite mean, bias-corrected satellite mean (CDR-4), ERA-5 and MERRA-2. The water vapour values are averaged from 2010 to 2014.

CDR-4 data are closer to ERA5 on all pressure levels than MERRA-2. Table 4-6 summarises the mean water vapour mixing ratio over global and tropical regions on different pressure levels. The annual mean water vapour in CDR-4 increases significantly after applying the bias-correction, bringing the values closer to ERA5. In the winter season (DJF), the change in water vapour mixing ratio from bias-correction is smaller. In the summer season (JJA), the bias-corrected CDR-4 water vapour is likely overestimated, partly due to the sampling issue from the BBH observations. Note that no water vapour observations are being assimilated in the reanalysis datasets in the UTLS above 300 hPa.

Table 4-6: Mean water vapour mixing ratio from original satellite mean, CDR-4 and two reanalysis datasets over global and tropical regions on different pressure levels. The mean values are for annual, winter and summer, averaged over the period 2010–2014

Pressure (hPa)	Mean WV (ppmv)	Global (90S–90N)			Tropical (30S–30N)		
		Annual	Winter (DJF)	Summer (JJA)	Annual	Winter (DJF)	Summer (JJA)
100	Original	3.90	3.73	4.17	3.86	3.35	4.34
	CDR-4	4.17	3.62	4.85	4.32	3.18	5.54
	ERA5	4.34	4.04	4.69	4.37	3.73	5.04
	MERRA2	4.70	4.39	5.01	4.50	3.88	5.15
150	Original	8.44	7.91	9.09	12.25	11.37	13.11
	CDR-4	10.45	9.68	13.47	15.85	14.58	20.90
	ERA5	10.46	9.69	11.39	14.75	13.70	15.81
	MERRA2	14.02	13.23	14.82	18.08	17.07	18.97
200	Original	32.49	30.15	34.97	51.06	47.70	53.83
	CDR-4	43.10	38.98	60.44	70.21	63.55	99.36
	ERA5	42.37	39.49	46.25	65.42	62.24	68.19
	MERRA2	57.20	53.32	61.55	84.84	80.22	88.10
300	Original	202.8	185.3	216.7	301.2	279.6	306.9
	CDR-4	251.7	218.5	321.0	387.9	340.8	488.5
	ERA5	279.0	266.8	304.3	413.3	406.0	424.8
	MERRA2	366.3	341.3	395.1	566.9	536.9	579.8

5.   HOMOGENEITY AND STABILITY ANALYSIS

The final CDR-4 data version will be monthly datasets in the period of 2010–2014, which will not be long enough to allow an assessment on the homogeneity and stability. Thus, only the assessment on CDR-3 will be provided here.

For CDR-3, the homogeneity and stability analysis is carried out against ERA5 water vapour time series. The assessment of homogeneity in CDR-3 includes analysis of results from the Penalised Maximal F (PMF, Wang, 2008a, b [RD-11], [RD-12]) and the Standard Normal Homogeneity (SNH, Reeves et al., 2007 [RD-13]) tests. Input to the tests are anomaly differences, i.e. the annual cycle has been removed from CDR-3 and the comparison data set and then the difference between both anomalies is used as input to the tests. Results contain information on potential break points in the anomaly difference time series. Associated information includes the time of the break point and its step size or strength. Details and cautionary notes are given in Schröder et al. (2016, 2019 [RD-14], [RD-15]). It is recalled that the uncertainty is of the order of ±3 months and it is possible that break points are missed. The results of the homogeneity test are summarised in Table 5-1. As an example, the break points are marked as vertical lines in the stability plot for the time series at 50 hPa (see Figure 5-1). In case the date of the break point coincides with known changes in the observing system or changes of the retrieval system it is mentioned in the event column, otherwise an interpretation of the breaks is attempted.
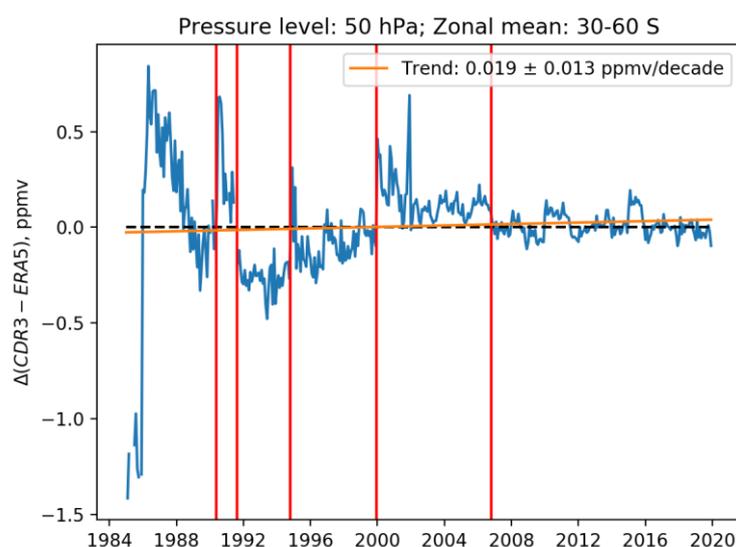
Figure 5-1: Breakpoints for the latitude band 30S–60S and the 50 hPa pressure level.

Most of the break points found seem not to be associated with a change in the measuring system; they rather indicate a gap in available input datasets or an irregularity in the ERA5 reanalysis used for the test. It seems that given the nature of reanalyses to have large discontinuities, future stability analyses may want to be made against data from CCMs.

Table 5-1: Results from homogeneity test for different pressure levels and latitude bands. Given are breakpoint, step size, and event interpretation

Pressure Level	Latitude	Break point (yyyy-mm)	Step size (ppmv)	Event
200 hPa	90N – 60N	No breaks		
	60N – 30N	No breaks		
	30N – 0	1993-04	-4.45	Restart of time series after Mt Pinatubo eruption. Unclear why this is not seen in other latitude bands.
	0 – 30S	2002-06	1.82	Likely due to sampling-bias in CDR-3 at this level that leads to a discontinuity in averaged time series. Again unclear why this is not seen in other latitude bands.
	30S – 60S	No breaks		
	60S – 90S	No breaks		
50 hPa	90N – 60N	No breaks		
	60N – 30N	No breaks		
	30N – 0	1998-11	0.097	Break in the availability of SAGE II and HALOE, leading to data gap.
	0 – 30S	No breaks		
	30S – 60S	1990-05	0.073	No interpretation of this problem.
		1991-08	-0.307	Temporary loss of SAGE II after Pinatubo eruption (until 1994-10).
		1994-10	0.114	Clear discontinuity in ERA5.
		1999-12	0.037	Strong dip in WV time series that is present in both CDR-3 and ERA5, but not well aligned with each other.
		2006-10	-0.038	
	60S – 90S	No breaks		
5 hPa	90N – 60N	No breaks		
	60N – 30N	1995-09	0.148	No clear discontinuity in CDR-3 upon closer inspection. Also, ERAi looks very similar to CDR-3!
		2002-10	-0.066	

Pressure Level	Latitude	Break point (yyyy-mm)	Step size (ppmv)	Event
	30N – 0	1995-06 2002-08	0.100 -0.043	Note, breakpoints are similar albeit earlier than those found at 60-30N, indicating propagating, true signal.
	0 – 30S	1999-11	0.236	Clear discontinuity in ERA5.
	30S – 60S	1994-07	0.211	No clear discontinuity in CDR upon closer inspection. Individual timeseries from HALOE and SAGE II show similar fluctuations, indicating a true signal.
	60S – 90S	No breaks		

6. UNCERTAINTY ASSESSMENT

This section will include an attempt to assess the quality of the uncertainty estimates of CDR-3 and CDR-4. According to the PVP [RD-1], the uncertainty assessment is carried out through the comparison of the difference of the CDR and the reference data with the associated uncertainty estimates (Immler et al., 2010 [RD-16]). As equation 3.7 in PVP (v3.2) indicates, the inequality equation should be valid for certain significance levels, i.e. 68% and 95% for k = 1 and k = 2, respectively. If not, either there is a systematic difference between the two data or the total uncertainty is underestimated or overestimated. Due to the lack of reliable reference data and given the very good agreement between CDR-3 and CCMI models (see Section 4.1.2), the assessment is carried out using deseasonalised water vapour anomalies to check the consistency between CDR-3 and the CCMI multi-model mean, with the CCMI standard deviation taken as the uncertainty in this mean. For CDR-4, as shown in Section 4.2.2, the merged bias-corrected CDR-4 data is close to ERA-5 on multiple pressure levels. Here, the consistency between the CDR-4 and EAR-5 on water vapour mixing ratio is therefore used for the uncertainty assessment.

6.1 Uncertainty assessment of monthly zonal mean VRWV product (CDR-3)

The assessment on CDR-3 is carried out during the period of 1985–2010 due to most CCMI models only providing simulations to 2010. The whole time period is then divided into three internal periods: 1985–1992, 1993–2003 and 2004–2010 due to the difference in merged satellite data in certain periods. The anomaly difference between CDR-3 and CCMI mean as a function of total uncertainty is plotted for different pressure layers over different latitude bands. In this part, the analysis is based on the anomaly difference after removing the annual seasonal cycle from the CDR-3 and CCMI mean. The total uncertainty used here is from the combination of uncertainty from CDR-3 and the standard deviation among CCMI models. Note that the outliers in CDR-3 were not filtered and the PDF of the bias is not perfectly Gaussian, which would have impacts on the assessment shown below.

Figure 6-1 to Figure 6-5 show the results from the consistency analysis between CDR-3 and CCMI. It can be seen that within all latitude ranges, the anomaly biases decrease slightly through the three periods, while the total uncertainty decreases significantly over the time. Nearly all the anomaly biases are below the k=2 line, indicating the consistency between CDR-3 and CCMI uncertainties is fairly high. As shown in the

frequency plots, some latitude ranges have long tails of large total uncertainty, partly from a small fraction of outliers in CDR-3.
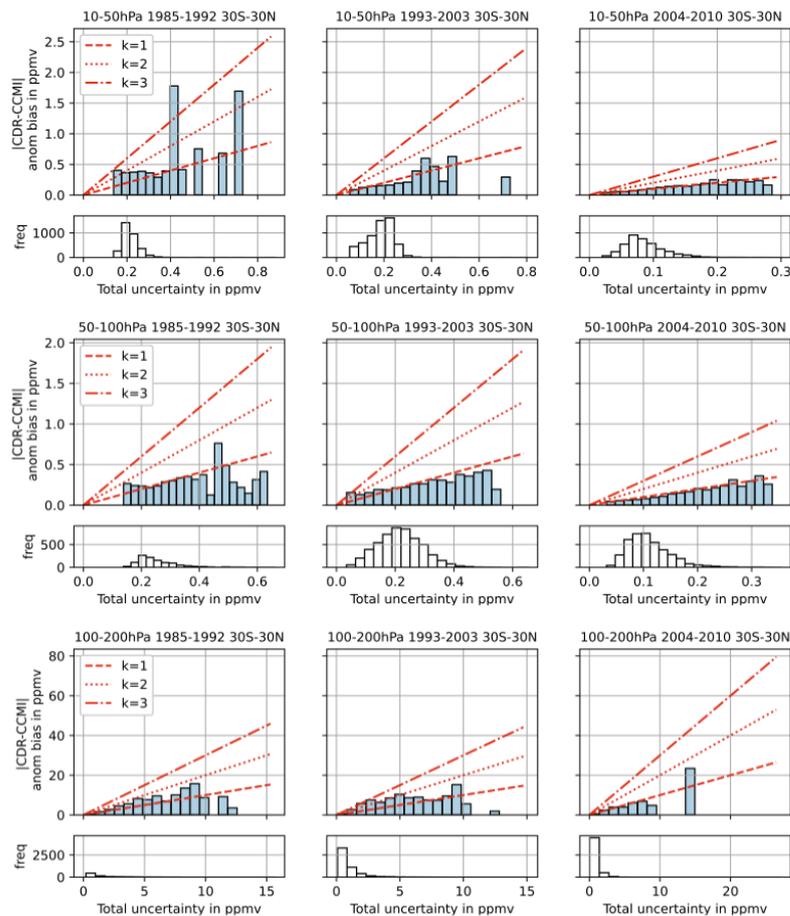
Figure 6-1: Immler inequation for the consistency analysis in water vapour deseasonalised anomaly between CDR-3 and the CCMI multi-model mean in three pressure layers: 10–50 hPa (top row), 50–100 hPa (middle row), and 100–200 hPa (bottom row) within the latitude range 30S–30N.
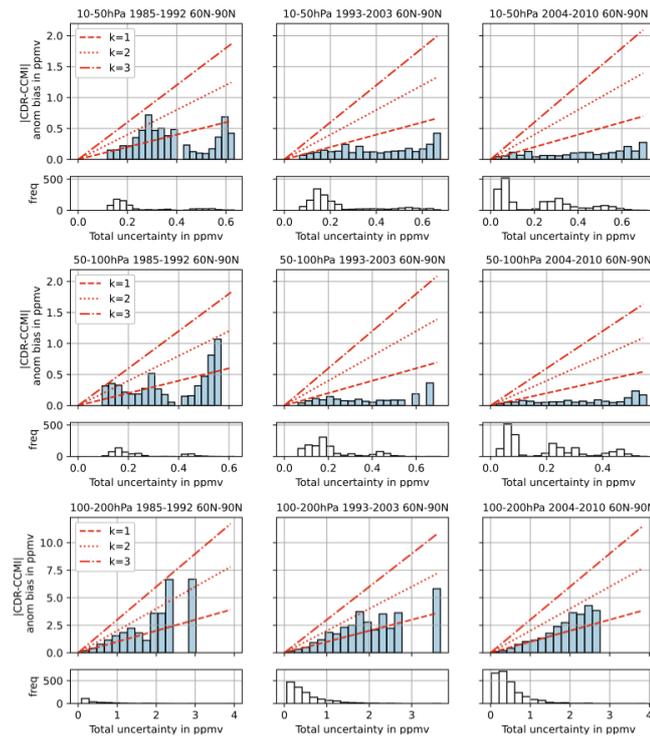
Figure 6-2: Same as Figure 6-1, but within the latitude range 60N–90N.
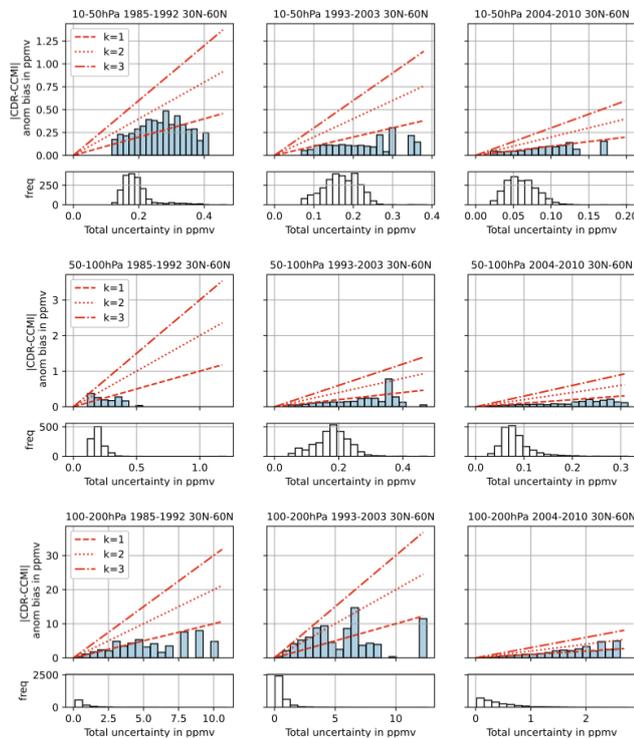
Figure 6-3: Same as Figure 6-1, but within the latitude range 30N–60N.
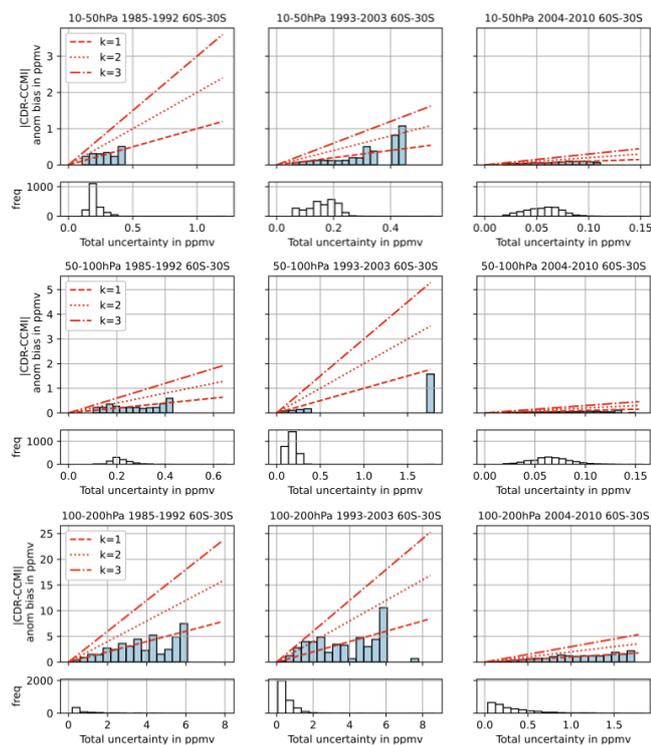
Figure 6-4: Same as Figure 6-1, but within the latitude range 60S–30S.
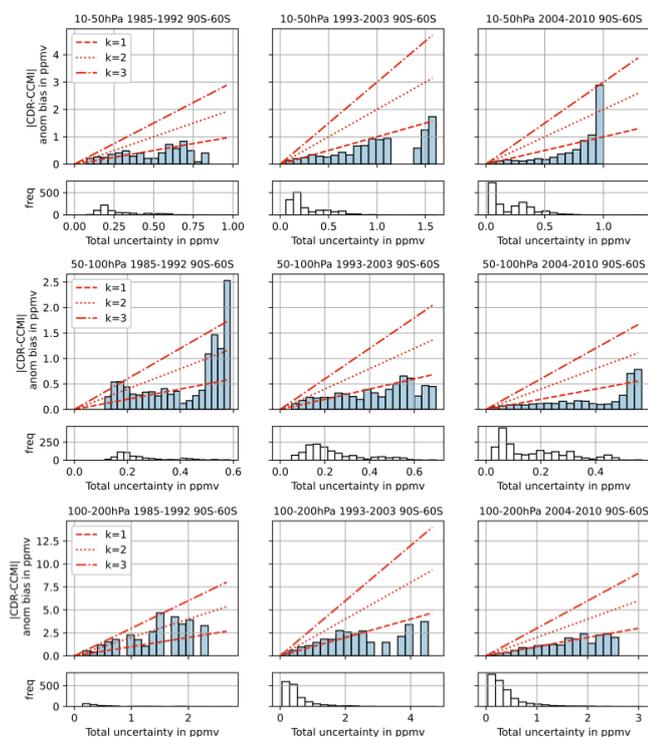
Figure 6-5: Same as Figure 6-1, but within the latitude range 60S–90S.

Table 6-1 summarises the significance levels of the consistency test between CDR-3 and CCMI multi-model mean. The probability is calculated from the inequality equation for k=1, k=2 and k=3. The probability should be close to the expectation, i.e. 68%, 95% and 99% for k=1, k=2 and k=3, respectively. Large differences indicate that the total uncertainty is underestimated or overestimated. The green values mark where they are close to the expectation. Within the pressure ranges of 10–50 hPa and 50–100 hPa, the consistency is improving towards the latter two periods, indicating a better overall quality of CDR-3 over time. However, in the 1–10 hPa pressure range, the consistency becomes worse, especially in the period 2004–2010, likely due to an underestimate of the total uncertainty from CDR-3 and CCMI, which become very small during this time period. Finally, within the pressure range of 100–200 hPa, the consistency is generally falling short of reaching expected values. This may not only be due to an underestimation of the uncertainties, but also due to true differences in the CCMI model mean and the CDR-3 data, with models commonly exhibiting problems to resolve this region of the atmosphere accurately.

Table 6-1: Summary of the consistency assessment between CDR-3 and CCMI. Values are marked green where they are close to the expected values (arbitrary differences apply; [61% to 75%] for 68%, [92% to 98%] for 95%, and [98% to 99%] for 98%)

Pressure layer (hPa)	Latitude band	Consistency (%)								
		1985–1992			1993–2003			2004–2010		
		k=1	k=2	k=3	k=1	k=2	k=3	k=1	k=2	k=3
1-10	60N-90N	66	92	99	75	93	97	76	93	98
	30N-60N	61	92	99	74	94	98	54	85	96
	30S-30N	60	91	98	70	93	98	49	83	96
	60S-30S	61	93	99	72	91	96	49	80	95
	90S-60S	69	93	99	73	91	96	73	89	97
10-50	60N-90N	65	86	94	76	96	99	78	94	99
	30N-60N	49	74	87	77	98	99	57	91	98
	30S-30N	41	64	81	66	92	98	52	88	98
	60S-30S	41	70	88	73	95	99	47	86	98
	90S-60S	51	77	90	61	86	95	66	88	96
50-100	60N-90N	51	75	86	84	97	99	84	98	99
	30N-60N	35	71	88	75	96	99	76	98	99
	30S-30N	55	87	97	58	84	94	67	94	98
	60S-30S	50	82	93	79	97	99	73	97	99
	90S-60S	43	63	74	50	80	92	72	91	97
100-200	60N-90N	48	84	92	84	98	99	65	94	98

	30N-60N	43	80	95	47	74	86	58	89	97
	30S-30N	49	82	94	43	68	83	51	80	92
	60S-30S	44	85	97	55	81	90	60	90	97
	90S-60S	26	47	76	43	77	91	61	90	96

6.2   Uncertainty assessment UTLS VRWV product (CDR-4)

Unlike CDR-3, the uncertainty assessment of CDR-4 is carried out between the consistency of water vapour mixing ratio between the merged product and ERA-5 during the period of 2010–2014. The bias between CDR-4 and ERA-5 as a function of total uncertainty is plotted on different pressure levels and latitude bands. The uncertainty of ERA-5 used here is the standard deviation of the monthly average values within the chosen latitude bands and pressure levels. The total uncertainty is calculated by combining both uncertainties from the CDR-4 and ERA-5.

Figure 6-6 to Figure 6-8 show the results from the consistency analysis between CDR-4 and ERA-5 in three latitude bands: 60S–30S, 30S–30N, 30N–60S within the UTLS region on four pressure levels: 100, 150, 200 and 300 hPa. It can be seen that the mean biases are very high relative to the total uncertainties, indicating that the overall consistency is low. When looking at the frequency plot for the total uncertainty, there are long tails of biases with high uncertainty but the majority of the data with low uncertainties have large biases. This may partly be explained by the assumption that the ERA-5 uncertainty only includes the standard deviation, leading to an underestimate of the uncertainty. However, this feature is more likely to express that the CDR-4 and ERA-5 still have significant differences in the UTLS region.

Table 6-2 summarises the significance levels of the consistency test between CDR-4 and ERA-5 on different pressure levels. Similar to Table 6-1, the probability is calculated from the inequality equation for k=1, k=2, and k=3. The table includes the results for all seasons as well as the winter and summer only. Over all time periods, the consistency is low on all pressure levels, especially on 150 hPa. The winter season exhibits a better consistency between CDR-4 and ERA-5 uncertainties than the summer season, which is consistent with the comparison shown in Table 4-4 that the water vapour are bias-corrected by too high a value in the summer season.
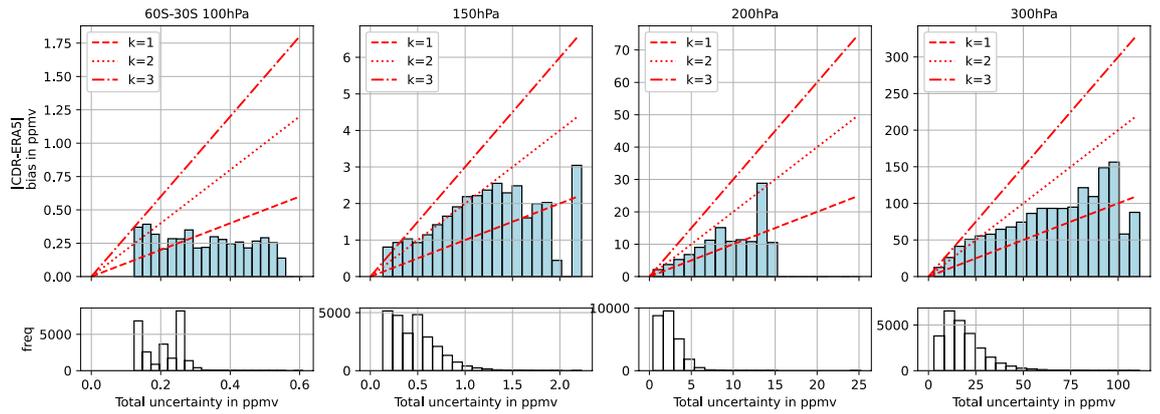
Figure 6-6: Immler inequation for the consistency analysis in water vapour deseasonalised anomaly between CDR-4 and ERA-5 on four pressure layers: 100, 150, 200 and 300 hPa within the latitude range 60S–30S.
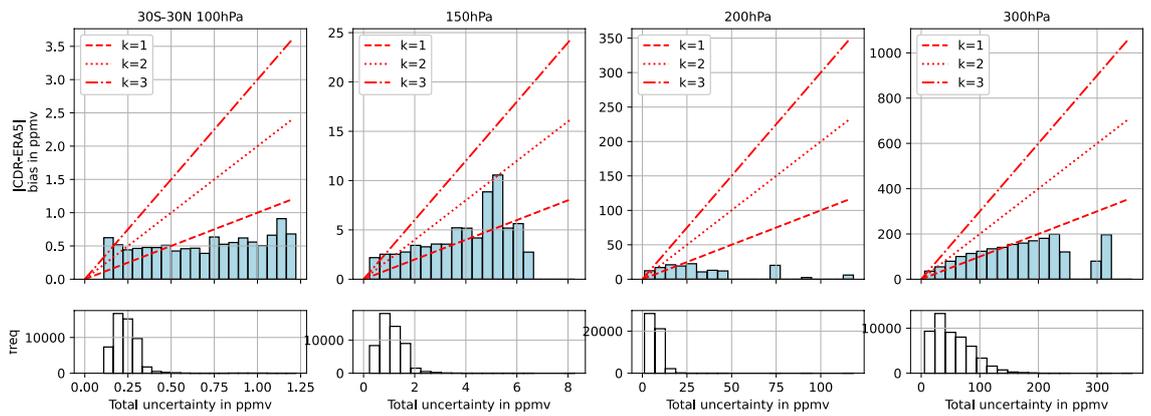
Figure 6-7: Same as Figure 6-6, but within the latitude range 30S–30N.
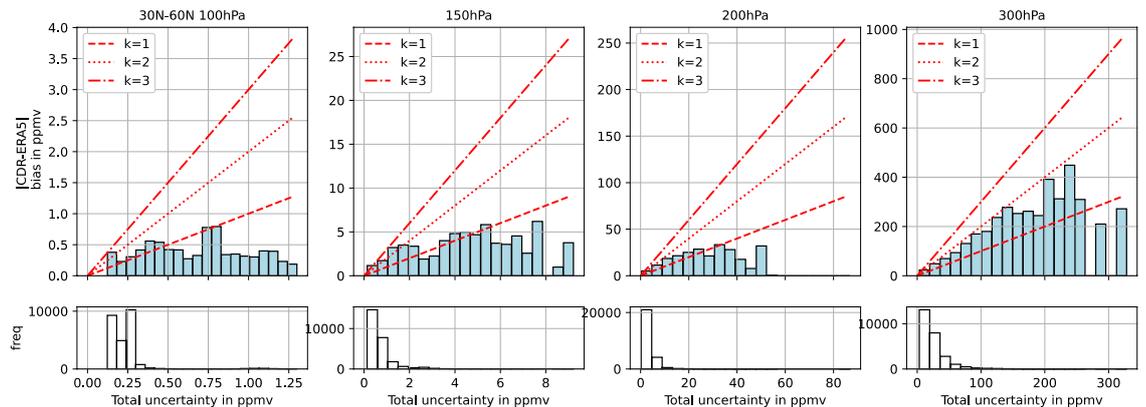
Figure 6-8: Same as Figure 6-6, but within the latitude range 30N–60N.

Table 6-2: Summary of the consistency assessment between CDR-4 and ERA-5. Values are marked green where they are close to the expected values (arbitrary differences apply; [61% to 75%] for 68%, [92% to 98%] for 95%, and [98% to 99%] for 98%)

Pressure levels (hPa)	Latitude band	Consistency (%)								
		All time			Summer (JJA)			Winter (DJF)		
		k=1	k=2	k=3	k=1	k=2	k=3	k=1	k=2	k=3
100	60N-90N	38	72	86	19	49	75	68	93	97
	30N-60N	40	69	84	13	60	79	48	81	92
	30S-30N	31	55	72	27	51	66	24	44	61
	60S-30S	36	68	85	45	77	92	31	64	82
	90S-60S	35	62	81	51	85	95	17	38	57
150	60N-90N	27	55	69	27	52	65	36	61	82
	30N-60N	9	34	55	13	37	59	4	29	47
	30S-30N	30	53	68	8	19	33	36	62	77
	60S-30S	12	39	63	12	39	60	17	45	70
	90S-60S	12	32	43	10	30	38	14	37	53
200	60N-90N	32	63	82	37	66	83	29	58	78
	30N-60N	19	55	79	18	40	63	20	64	87
	30S-30N	33	58	74	29	61	85	42	73	87
	60S-30S	35	66	85	38	71	88	40	75	91
	90S-60S	28	54	66	25	49	61	29	60	68
300	60N-90N	42	79	94	34	78	96	51	87	97
	30N-60N	27	60	83	27	58	82	31	65	86
	30S-30N	36	68	87	33	60	79	40	73	89
	60S-30S	24	49	72	28	59	80	18	48	77
	90S-60S	35	61	72	46	72	79	22	47	63

7. COMPLIANCE WITH REQUIREMENTS

We here provide a summary of the compliance of the two VRWV products CDR-3 and CDR-4 provided by the project against the user requirements as established in the URD (v3.0) [RD-2]. These requirements are based on a survey of research needs from the Climate Research Group (CRG) and on more general needs formulated by the international climate research community as summarised by monitoring bodies like GCOS.

Note, however, that limited by the lack of reliable observations as reference, both in spatial and temporal coverage, for the monthly VRWV data from CDR-3 and CDR-4, the uncertainty assessments for these two products are evaluated in terms of the consistency of the two CDRs to models and other limited observational reference datasets instead and thus have to be interpreted with care.

7.1 Compliance of monthly zonal mean VRWV product (CDR-3)

The compliance with URD requirements [RD-2] for CDR-3 can be summarised as follows. Overall, compliance is generally achieved with respect to the requirements of the climate user group for the last 17 years of the climate data record (2003–2019), while for the time series 1992–2003, the requirements are met only partially, and for the time series before 1992, most requirements are not met in a satisfactory way. It should be noted upfront that CDR-3 was not designed to meet user requirements according to GCOS, which is mostly satisfying NWP research needs and requires generally a much higher temporal and spatial resolution, as well as much higher accuracy and precision. GCOS and WMO requirements are met only by some of the single-profile instrument products as discussed in detail in the SPARC WAVAS-II activity publications (https://acp.copernicus.org/articles/special_issue830.html) and provided via the links found in the DARD [RD-6].

- *Overall uncertainty:* The accuracy (systematic component) as assessed in comparison with CCMI (as alternative to a non-existing observational reference) is compliant with the threshold requirement of 25% for all latitude bands and most pressure levels (except at altitudes equal to or below the 200 hPa pressure level). Around 100 hPa, the uncertainty improves locally even towards the breakthrough requirement of 5%. Similarly, the random component of the accuracy, which is determined from the optimal estimation method, is found to fulfil mostly the

breakthrough requirement of 5% (although not for data before 2003 and not for pressure levels below 0.5 hPa).

- *Spatial resolution:* The spatial resolution of CDR-3 (5 degrees in latitude and 1–3 km vertical resolution) is compliant with the threshold requirements of 500 km in the horizontal and 3 km in the vertical as defined by the climate user group survey. The horizontal resolution falls only short by a factor of 2–3 compared with GCOS and WMO requirements, which aim at a resolution of 100–200 km. On the other hand, the vertical resolution complies with WMO and GCOS requirements for the vertical resolution (2–3 km) over the whole time period of the record.

- *Temporal resolution:* The temporal resolution of CDR-3 (monthly) is compliant with the threshold requirement derived from the climate user group survey. GCOS and WMO requirements for the frequency can only be achieved by single instruments (for those available after 2004) and on a daily (not hourly) basis.

- *Stability:* The stability assessment focused in a first step on the identification of breakpoints. Identified breakpoints could be mostly attributed to data gaps or clear discontinuities in the ERA5 data and do not point to problems in CDR-3. A trend analysis in the biases with respect to different reanalyses then indicates that the breakthrough requirement of 1% per decade is achieved for the time period 2003 onwards for most pressure levels and latitude bands. For the time period before 2003, the stability requirement is compatible with the 2.5% per decade threshold requirement. However, the analysis indicates that this was only achieved because of strong fluctuations in the biases and thus this result is less meaningful overall.

- *Record length:* CDR-3 currently covers 35 years, thus its length in principle achieves the objective as determined by the climate user group (> 30 years). It should be noted, however, that CDR-3 should only be used with care prior to 1992, since its accuracy does not meet the threshold requirement prior to this year. Thus, the part of the CDR-3 partial time series compliant with the defined uncertainty requirements is only 27 years long and only achieves the target requirement for record length.

7.2   Compliance of UTLS product (CDR‑4)

The compliance with URD requirements [RD-2] for CDR-4 can be summarised as follows. Overall, compliance is satisfactorily achieved with respect to the requirements of the climate user group for the prototype data record for the 5 years (2010–2014), with only partial requirements being met.

- *Overall uncertainty:* The consistency assessment between CDR-4 and ERA-5 shows that the uncertainty of CDR-4 may be underestimated, especially during the summer season. However, as the lack of reliable reference datasets, the assessment of uncertainty will need investigation with additional methods than the consistency assessment provided here.

- *Spatial resolution:* The spatial resolution of CDR-4 (5 degrees in latitude and longitude and 1−3 km vertical resolution) fulfils the threshold requirements of 200 km in the horizontal and is compliant with the 2 km in the vertical as defined by the climate user group survey.

- *Temporal resolution:* The temporal resolution of CDR-4 (monthly) is compliant with the threshold requirement derived from the climate user group survey.

- *Record length:* The prototype CDR-4 currently covers 5 years, which is compliant with a breakthrough of 5 years required by the climate user group survey. Nonetheless, the record length of CDR-4 will be extended in WV_cci Phase 2 towards reaching the goal requirement of 20 years.

8. SUMMARY

This PVIR covers the monthly mean zonal mean vertically resolved water vapour climate data record CDR-3 (v3.3), **CCI WV-strato**, offering a stratospheric climate data record over the time period 1985–2019, and the three-dimensional monthly mean vertically resolved water vapour climate data record CDR-4 prototype v3, **CCI WV-UTLS**, covering the UTLS from 2010 to 2014.

It should be noted upfront that there is a clear lack of observations that can be used as a well-defined reference for validation of the vertically resolved WV climate data records, CDR-3 and CDR-4. This limits the power of the performed comparisons to yield proper quantitative constraints on the uncertainty of the two datasets and the validation/comparison activity may have to be interpreted more as 'consistency tests'. Consistency between models and CDRs, and between limited observational reference datasets and CDRs can thereby serve as an indicator for the overall quality of the CDRs produced within WV_cci.

The CDR-3 data has been compared with another merged data product (SWOOSH), CCMI model simulations in specified dynamics mode, and reanalysis datasets (ERA5, ERAi, MERRA, and MERRA2). The results show that CDR-3 data agree generally very well with other references both in time series of absolute values and anomalies, especially from 2003 onwards, better in the lower and middle stratosphere than in the UTLS or upper stratosphere, and better at lower latitudes than higher latitudes. Generally better agreement is found with CCMs than with the reanalyses, likely pointing towards inhomogeneities in the latter, a known issue in reanalyses that stems from the changing observational systems assimilated into the reanalysis models.

The CDR-3 evaluations particularly indicate that there are still some issues with data before 2003, especially between 1985 and 1992, making these data less valuable for climate research applications. On the other hand, the determined random uncertainties for CDR-3 before 2003 are larger and more consistent with the uncertainties determined from the CCMI models than those for the CDR-3 after 2003, thus indicating that the dataset is at least well-characterised despite its shortcomings. The homogeneity and stability analysis of the CDR-3 data particularly highlights potential breakpoints, although further investigation indicates that these breakpoints are likely introduced by the reference dataset (here taken to be ERA5) rather than by the CDR-3 itself.

For CDR-4, the comparison of the bias-corrected input datasets with JULIA for the climatological distribution in equivalent latitude–potential temperature coordinates indicates that the bias-correction procedure improves the consistency among input satellite observations despite the differences to JULIA being still large. Generally, the bias-correction procedure increases the water vapour mixing ratio in the UTLS region. The comparisons with BBH profiles and uncorrected data show that the bias-corrected CDR-4 matches generally better with the BBH observations at BBH locations, capturing high water vapour values during the summer season over tropical and mid-latitude regions for some years. However, the bias-correction tends to be leading to too large increases during the summer season in some of the other years.

The comparisons of CDR-4 with reanalyses (ERA-5 and MERRA-2) show that the annual mean water vapour values from CDR-4 data are brought closer to the reanalyses, with overall smaller biases between ERA-5 compared to MERRA-2, both still being too wet in the UTLS as known from the literature. The comparisons in summer season, on the other hand, indicate that CDR-4 data tend to be bias-corrected to too high values. However, the comparisons between CDR-4 and reanalysis datasets also show that the merged product has similar horizontal distribution patterns as the reanalyses and increased water vapour in the UTLS regions.

The uncertainty consistency assessment for CDR-4 with ERA-5 shows that CDR-4 uncertainties are not in agreement with the uncertainties from ERA5, however the reanalysis datasets exhibit likely a substantial wet-bias, therefore agreement is not expected. Further assessment with more reliable reference datasets should be carried out in WV_cci Phase 2. A homogeneity and stability analysis of the CDR-4 product will only be carried out in Phase 2 of WV_cci when a longer data record of CDR-4 is being produced.

Thus in summary, the results for CDR-4 suggest that the merged product with both limb and nadir satellite data can provide a reasonable dataset for water vapour in the UTLS region. However, there are still remaining problems and limitations in the CDR-4 product. Although the CDR-4 product matches well with BBH observations, the CDR-4 product still has large sampling biases due to the very limited amount of BBH profiles used as observational reference data for the bias-correction. Currently, the merged CDR-4 product includes observations from Aura-MLS, MIPAS and IMS with different spatial and temporal coverage and the ACE-FTS and ACE-MAESTRO are excluded due to the very sparse coverage. Compared with BBH, the ACE-FTS and ACE-MAESTRO have larger spatial and temporal coverage, especially over the high latitude regions. The application of ACE-FTS and ACE-MAESTRO as new reference data in

the bias-correction in place of the BBH data could thus likely improve the quality of CDR-4 data.

APPENDIX 1: REFERENCES

[RD-1] ESA CCI Water Vapour: Product Validation Plan (PVP). M. Schröder, U. Falk, M. Hegglin, H. Ye, T. Trent, J. Fischer, R. Preusker, G. Stiller, A. Laeng, and T. von Clarmann. Issue 3.2, 29 July 2021.

[RD-2] ESA CCI Water Vapour: User Requirements Document (URD). M. Hegglin, M. Schröder, H. Brogniez, D. Hubert. Issue 3.0, 11 February 2021.

[RD-3] SPARC, 2017: The SPARC Data Initiative: Assessment of stratospheric trace gas and aerosol climatologies from satellite limb sounders. By M. I. Hegglin and S. Tegtmeier (eds.), *SPARC Report No. 8*, WCRP-5/2017, doi:10.3929/ethz-a-010863911.

[RD-4] Hegglin, M.I., Tegtmeier, S., Anderson, J., Froidevaux, L., Fuller, R., Funke, B., Jones, A., Lingenfelser, G., Lumpe, J., Pendlebury, D. and Remsberg, E., SPARC Data Initiative: Comparison of water vapor climatologies from international satellite limb sounders. *J. Geophys. Res.: Atmospheres*, *118*(20), pp.11–82, 2013.

[RD-5] Hegglin, M. I., Tegtmeier, S., Anderson, J., Bourassa, A. E., Brohede, S., Degenstein, D., Froidevaux, L., Funke, B., Gille, J., Kasai, Y., Kyrłł, E. T., Lumpe, J., Murtagh, D., Neu, J. L., Prot, K., Remsberg, E. E., Rozanov, A., Toohey, M., Urban, J., von Clarmann, T., Walker, K. A., Wang, H.-J., Arosio, C., Damadeo, R., Fuller, R. A., Lingenfelser, G., McLinden, C., Pendlebury, D., Roth, C., Ryan, N. J., Sioris, C., Smith, L., and Weigel, K.: Overview and update of the SPARC Data Initiative: comparison of stratospheric composition measurements from satellite limb sounders, Earth Syst. Sci. Data, 13, 1855-1903, https://doi.org/10.5194/essd-13-1855-2021, 2021.

[RD-6] ESA CCI Water Vapour: Data Access Requirement Document (DARD). M. Schröder, M. Hegglin, U. Falk, H. Brogniez, J. Fischer, D. Hubert, A. Laeng, R. Siddans, C. Sioris, G. Stiller, T. Trent, K. Walker. O. Danne and H Ye. Issue 3.2, 27 July 2021.

[RD-7] ESA CCI Water Vapour: Algorithm Theoretical Basis Document (ATBD), Part 3 - Merged CDR-3 and CDR-4 products. M. Hegglin and H. Ye. Issue v2.0, *in progress*.

[RD-8] Hegglin, M. I., D. Plummer, J. Scinocca, T. G. Shepherd, J. Anderson, L. Froidevaux, B. Funke, D. Hurst, A. Rozanov, J. Urban, T. v. Clarmann, K. A. Walker, R. Wang, S. Tegtmeier, and K. Weigel, Vertical structure of stratospheric water vapour trends derived from merged satellite data, Nature Geoscience, doi: 10.1038/NGEO2236, 2014.

[RD-9] Davis, S. M., Hegglin, M. I., Fujiwara, M., Dragani, R., Harada, Y., Kobayashi, C., Long, C., Manney, G. L., Nash, E. R., Potter, G. L., Tegtmeier, S., Wang, T., Wargan, K., and Wright, J. S.: Assessment of upper tropospheric and stratospheric water vapor and ozone in reanalyses as part of S-RIP, Atmos. Chem. Phys., 17, 12743–12778, https://doi.org/10.5194/acp-17-12743-2017, 2017.

[RD-10] ESA CCI Water Vapour: Product Validation and Selection Report (PVASR). A. Laeng, G. Stiller, T. von Clarmann, M. Schröder, M. Hegglin, and H. Ye. Issue 3.0, 15 August 2021.

[RD-11] Wang, X.L. : Penalized maximal F test for detecting undocumented mean shift without trend change. J. Atmos. Ocean. Technol. 25, 368–384, 2008a.

[RD-12] Wang, X.L.: Accounting for autocorrelation in detecting mean shifts in climate data series using the penalized maximal t or F test. J. App. Meteor. Climatol. 47, 2423–2444, 2008b.

[RD-13] Reeves, J., Chen, J., Wang, X.L., Lund, R. and Lu, Q: A review and comparison of changepoint detection techniques for climate data. J. Appl. Meteorol. Climatol., 46, 900–915, 2007.

[RD-14] Schröder, M., M. Lockhoff, J. Forsythe, H. Cronk, T. H. Vonder Haar, and R. Bennartz: The GEWEX water vapor assessment (G-VAP) – results from the trend and homogeneity analysis. J. Applied Meteor. Clim., 1633-1649, 55 (7), https://dx.doi.org/10.1175/JAMC-D-15-0304.1, 2016.

[RD-15] Schröder, M., M. Lockhoff, L. Shi, T. August, R. Bennartz, H. Brogniez, X. Calbet, F. Fell, J. Forsythe, A. Gambacorta, S.-P. Ho, E. R. Kursinski, A. Reale, T. Trent, and Q. Yang: The GEWEX water vapor assessment of global water vapour and temperature data records from satellites and reanalyses. Rem. Sens., 11(3), 251, https://doi.org/10.3390/rs11030251, 2019.

[RD-16] Immler, F., Dykema, J., Gardiner, T., Whiteman, D., Thorne, P., & Vömel, H.: Reference quality upper-air measurements: Guidance for developing GRUAN data products. Atmospheric Measurement Techniques, 3(5), 1217–1231, 2010.

APPENDIX 2: GLOSSARY

Term	Definition
ABC(t)	Atmosphere Biosphere Climate (teledetection)
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
AMSR-E	Advanced Microwave Scanning Radiometer for EOS
AMSU	Advanced Microwave Sounding Unit
ARA	Atmospheric Radiation Analysis
ARSA	Analyzed RadioSoundings Archive
AVHRR	Advanced Very High Resolution Receiver
BBH	Balloon-Borne Hygrometer
BC	Brockmann Consult
CARIBIC	Civil Aircraft for the Regular Investigation of the atmosphere Based on an Instrument Container
CCI	Climate Change Initiative
CCMI	Chemistry Climate Model Initiative
CDO	Climate Data Operators
CDR	Climate Data Record
CDS	Copernicus Climate Data Store
CEDA	Centre for Environmental Data Analysis
CF	Conventions for Climate and Forecast
CFH	Cryogenic frost-point hygrometer
CM SAF	EUMETSAT Satellite Application Facility on Climate Monitoring
CMAM	Canadian Middle Atmosphere Model
CMIP	Coupled Model intercomparison Project
CMUG	Climate Modelling User Group
CRG	Climate Research Group
DLR	Deutschen Zentrums für Luft- und Raumfahrt
DWD	Deutscher Wetterdienst (German MetService)
ECCC	Environment and Climate Change Canada
ECMWF	European Centre for Medium-Range Weather Forecasts
ECV	Essential Climate Variable
EDA	ERA-5 - reduced resolution ten member ensemble

Term	Definition
EMiR	ERS/Envisat MWR Recalibration and Water Vapour Thematic Data Record Generation
Envisat	Environmental Satellite
ERA-5	ECMWF Re-Analysis 5
ERA-Interim	ECMWF Re-Analysis Interim
ESA	European Space Agency
EUMETSAT	European Organisation for the Exploitation of Meteorological Satellites
FOV	Field of View
FPH	Frost Point Hygrometer
GCOS	Global Climate Observing System
GEOS-5	Goddard Earth Observing System Model, Version 5
GMI	Global Precipitation Microwave Imager
GNSS	Global Navigation Satellite System
GOMOS	Global Ozone Monitoring by Occultation of Stars
GOZCARDS	Global OZone Chemistry And Related trace gas Data records for the Stratosphere
GPS	Global Positioning System
GUM	Guide to the Expression of Uncertainty in Measurement
HARMOZ	HARMonized dataset of Ozone profiles
HALOE	Halogen Occultation Experiment
HIRDLS	High Resolution Dynamics Limb Sounder
HOAPS	Hamburg Ocean Atmosphere Parameters and Fluxes from Satellite Data
IAGOS	In-service Aircraft for a Global Observing System
IASI	Infrared Atmospheric Sounder Interferometer
IGAC	International Global Atmospheric Chemistry
ILAS-II	Improved Limb Atmospheric Spectrometer-II
IMS	Infrared Microwave Sounding
IPSL-CM	Institut Pierre Simon Laplace Climate Model
IR	Infrared
LMD	Laboratoire Météorologie Dynamique
LMS	Lowermost stratosphere
LST	Land Surface Temperature
LWP	Vertically integrated liquid water
MERIS	Medium Resolution Imaging Spectrometer Instrument

Term	Definition
MERRA-2	Modern-Era Retrospective analysis for Research and Applications, Version 2
MHS	Microwave Humidity Sounder
MIPAS	Michelson Interferometer for Passive Atmospheric Sounding
MLS	Microwave Limb Sounder
MODIS	Moderate Resolution Imaging Spectrometer
MOZAIC	Measurement of OZone by Airbus In-service airCraft
MPI-M	Max-Planck Institute for Meteorology
MUDB	Match-up database
NASA	National Aeronautics and Space Administration
NCAR	National Center for Atmospheric Research
NCEO	National Centre for Earth Observation
NCEP	National Centers for Atmospheric Prediction
NDVI	Normalized Difference Vegetation Index
NIR	Near IR
NOAA	National Oceanic & Atmospheric Administration
NWP	Numerical Weather Prediction
OLCI	Ocean and Land Colour Instrument
PCs	Principle components
PMF	Penalised Maximal F
POAM	Polar Ozone and Aerosol Measurement
PSD	Product Specification Document
RAL	Rutherford Appleton Laboratory
RMS	Root mean square
RR	Reduced resolution
RTTOV	Radiative Transfer for TOVS
SAGE	Stratospheric Aerosol and Gas Experiment
SASBE	Site Atmospheric State Best Estimate
SCIAMACHY	Scanning Imaging Absorption Spectrometer for Atmospheric Cartography
SCISAT	Scientific Satellite
SD	specified dynamics
SE	Spectral Earth
SMILES	Solar wind Magnetosphere Ionosphere Link Explorer
SMR	Software Modification Report

Term	Definition
SNH	Standard Normal Homogeneity
SNR	Signal-to-noise ratio
SOFIE	Solar Occultation For Ice Experiment
SPARC	Stratosphere-troposphere Processes And their Role in Climate
SPURT	Spurenstofftransport in der Tropopausenregion, trace gas transport in the tropopause region
SSM/I	Special Sensor Microwave Imager
SSMIS	Special Sensor Microwave Imager Sounder
SST	Sea Surface Temperature
SuomiNet	Global ground based GPS network (named after Verner Suomi)
SWOOSH	Stratospheric Water and OzOne Satellite Homogenized data set
TBD	To be determined
TCWV	Total Column Water Vapour
TMI	Tropical Rainfall Measuring Mission's Microwave Imager
TOA	Top Of Atmosphere
UKMO	United Kingdom Meteorological Office
UoL	University of Leicester
UoR	University of Reading
URD	User Requirements Document
UT	Upper Troposphere
UTLS	Upper Troposphere and Lower Stratosphere
UV	Ultraviolet
vis	Visible
VMR	Volume mixing ratio
VRes	Vertically resolved
WACCM	Whole Atmosphere Community Climate Model
WAVAS-I	Water Vapour Assessment
WAVAS-II	Water Vapour Assessment 2
WCRP	World Climate Research Programme
WGS	World Geodetic System 1984
WMO	World Meteorological Organization
WV	Water Vapour
WV_cci	Water Vapour climate change initiative

End of Document