

Water Vapour Climate Change Initiative (WV_cci) - CCI+ Phase 2



Product Validation and Intercomparison Report (PVIR) - Part 1: CDR-1 & CDR-2

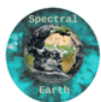
Ref: D4.1

Date: 25 June 2025

Issue: 5.0

For: ESA / ECSAT

Ref: CCIWV.REP.016



aeronomie.be



UNIVERSITY OF
TORONTO



UNIVERSITY OF
LEICESTER

UNIVERSITÉ DE
VERSAILLES
SAINT-QUENTIN-EN-YVELINES



Science & Technology Facilities Council
Rutherford Appleton Laboratory

Universidade de Vigo

This Page is Intentionally Blank

Project : Water Vapour Climate Change Initiative (WV_cci) - CCI+ Phase 22

Document Title: Product Validation and Intercomparison Report (PVIR) - Part 1: CDR-1 & CDR-2

Reference : D4.1

Issued : 25 June 2025

Issue : 5.0

Client: ESA / ECSAT

Author(s) : Johannes Bärlin (DWD), Hannes Konrad (DWD), Anja Niedorf (DWD), Rene Preusker (SE), Tim Trent (UoL), Jürgen Fischer (SE), Marc Schröder (DWD)

Copyright : Water_Vapour_cci Consortium and ESA

Document Change Log

Issue/ Revision	Date	Comment
0.9	28 July 2020	Initial issue submitted to project coordination for review
1.0	31 July 2020	First issue for submission to ESA
1.1_draft	14 Sep 2020	Feedback from co-authors to RID's implemented
1.1	13 October 2020	RID's from ESA implemented
1.2	21 May 2021	Correction to equation 4.8
2.0	21 August 2021	Version submitted to ESA prior to AR3
2.1	30 September 2021	Feedback from AR3 implemented, RIDs from ESA answered
3.0	15 April 2024	First version of phase 2, submitted to ESA for review
4.0	31 January 2025	Second version of phase 2, submitted to ESA for review
5.0	25.06.2025	Final version of phase 3, submitted to ESA for review: • Minor updates to all sections • Major updates to sections 4, 7 and 8, to include CDR-2 analyses. • Deletions in sections 4.3 (contrasting of different sensor combinations for the high-resolution products), 5 (consistency of CDR-1 against AIRS), and 6 (clear-sky bias based on GRUAN observations)

TABLE OF CONTENTS

1. INTRODUCTION	11
1.1 Purpose	11
1.2 Scope	11
1.3 The ESA Water_Vapour_cci project	12
1.4 WV_cci datasets	12
2. VALIDATION STRATEGY	14
2.1 Datasets	14
2.2 Methodology	14
3. VALIDATION OF NIR LEVEL 2 ALGORITHM	15
4. VALIDATION AND INTERCOMPARISON OF WV_CCI PRODUCTS	18
4.1 Validation over global land surfaces (CDR-1)	20
4.1.1 Validation using station-based data	23
4.1.2 Intercomparison with long-term data records	31
4.2 Validation over all surfaces (CDR-2)	39
4.2.1 Intercomparison with long-term data records	39
4.2.1.1 Results over coasts	45
4.2.1.2 Results over ocean	49
4.3 Validation of the regional, high resolution product	52
5. UNCERTAINTY ASSESSMENT	57
5.1 General approach	57
5.2 Refined approach using stratified data	59
6. CLEAR-SKY BIAS ESTIMATION USING SUOMINET	63
7. OVERALL STATISTICS AND COMPLIANCE WITH REQUIREMENTS	67
8. SUMMARY	72
APPENDIX 1: REFERENCES	74
APPENDIX 2: GLOSSARY	76

INDEX OF TABLES

Table 4-1: Name conventions in GRUAN files for RS-92 and RS-41 radiosonde datasets, see also
the GRUAN technical note 13

(https://www.gruan.org/gruan/editor/documents/gruan/GRUAN-TN-13_RS41-GDP_UserGuide_v1.0_2022-11-21.pdf).	24
Table 4-2: Basic statistical measures are shown for the data in Figure 4-4 with TCWV values in GRUAN subtracted from TCWV values in CDR-1, abbreviated as “CDR1 – GRUAN”. The relative quantities are obtained by dividing their absolute counter parts through the mean TCWV in collocated GRUAN data. <i>Used CDR-1 dataset: 0.05° x 0.05° x daily.</i>	26
Table 4-3: This table and Figure 4-6 are related to each other in a similar way as Table 4-2 and Figure 4-4. The panels in Figure 4-6 and rows in this tables can be identified by the respective headers. The relative quantities are obtained by dividing their absolute counter parts through the mean TCWV in collocated SuomiNet data. <i>Used CDR-1 dataset: 0.05° x 0.05° x daily.</i>	30
Table 4-4: Basic statistical measures derived from the time series in the panels a), b) and c) of Figure 4-7. The relative values are taken with respect to the mean TCWV of the reference. The given uncertainties are the standard deviations of the timeseries. The slopes are obtained from the regression curves in Figure 4-8. <i>Used CDR-1 datasets: 0.5° x 0.5° x monthly.</i>	33
Table 4-5: Basic statistical measures for the timeseries in Figure 4-11 are shown. <i>Used CDR-2 datasets: 0.5° x 0.5° x daily.</i>	39
Table 4-6: Basic statistical measures for the timeseries in Figure 4-12 and Figure 4-13 are shown, i.e. global, all surfaces. <i>Used CDR-2 datasets: 0.5° x 0.5° x daily.</i>	41
Table 4-7: Basic statistical measures for the timeseries in Figure 4-17 and Figure 4-18 are shown. <i>Used CDR-2 datasets: 0.5° x 0.5° x daily and 0.05° x 0.05° x daily.</i>	48
Table 4-8: Basic statistical measures for the timeseries in Figure 4-19 and Figure 4-20 are shown. <i>Used CDR-2 datasets: 0.5° x 0.5° x monthly.</i>	52
Table 4-9: Specifications of the high-resolution product.	52
Table 4-10: This table shows the MD and cRMSD of high-resolution CDR-2 to the reference AIRS for all three regions of interest separately. <i>Used CDR datasets: 0.01° x 0.01° x monthly regridded to 0.5° x 0.5° x monthly.</i>	55
Table 4-11: This table shows MD and cRMSD of high-resolution CDR-2 to the reference ERA5 for all three regions of interest separately. <i>Used CDR datasets: 0.01° x 0.01° x monthly regridded to 0.5° x 0.5° x monthly.</i>	56
Table 5-1: Consistency assessment of CDR-1 using GRUAN. <i>Used CDR-1 dataset: 0.05° x 0.05° x daily.</i>	58
Table 6-1: The global, latitude-weighted means of each climatology obtained in the course to estimate CDR-1’s clear sky bias as well as the CSB itself are shown. The unit is kg/m ² in all columns. At coasts, the sample size is small. <i>Used CDR-1 dataset: 0.05° x 0.05° x daily.</i>	63
Table 7-1: User requirements on CDR-1 as given in Table 6.1 of URD v4.1.	68
Table 7-2: Expected product requirements on CDR-1 as given in Table 3.2 of PSD v4.2.	68

Table 7-3: The table shows the overall statistics for CDR-1 compared to the different references. For GRUAN and SuomiNet the data is the same as in Table 4-2 and the second to last row of Table 4-3 ("actual over passes", clear-sky). The metrics for AIRS, ERA5 and CDR_pre are taken from Table 4-4. Consistencies are the same as in Table 5-1. *Used CDR-1 dataset: $0.05^\circ \times 0.05^\circ \times \text{daily}$.* 69

Table 7-4: The table shows the overall statistics for CDR-2 compared to the references AIRS, ERA5 and CDR_pre. The metrics are taken from Table 4-6. *Used CDR-2 dataset: $0.05^\circ \times 0.05^\circ \times \text{daily}$.* 70

INDEX OF FIGURES

Figure 3-1: Comparison between MERIS and ARM-microwave based TCWV retrievals. 15

Figure 3-2: Comparison between MODIS and ARM-microwave based TCWV retrievals for the morning overpath (Terra upper plots) and the afternoon overpath (Aqua lower plots). 16

Figure 3-3: Comparison between OLCI (A upper, B lower) and SGP-ARM MWR based TCWV. . 17

Figure 4-1: a) is a global map of the temporal mean of valid TCWV values with all monthly data points equally weighted, b) shows the corresponding standard deviation. c) displays the sample size, or number of months with a valid TCWV value, per grid. *Used CDR-2 dataset: $0.5^\circ \times 0.5^\circ \times \text{monthly}$.* 18

Figure 4-2: a) is a time series of the global spatial mean of valid TCWV values with all monthly data points equally weighted, b) shows the corresponding standard deviation. c) displays the sample size, or number of months with a valid TCWV value, per grid. *Used CDR-2 dataset: $0.5^\circ \times 0.5^\circ \times \text{monthly}$.* 19

Figure 4-3: a) is a time series of the spatial mean of valid TCWV values over global land with all monthly data points equally weighted, b) shows the corresponding standard deviation. c) displays the sample size, or number of months with a valid TCWV value, per grid. *Used CDR-1 dataset: $0.5^\circ \times 0.5^\circ \times \text{monthly}$.* 22

Figure 4-4: The figure shows two scatter plots of collocated TCWV values from GRUAN and CDR-1. Only means obtained from measurements close to actual over pass enter on the GRUAN side. All CCI-data is filtered by allowing only "0" for "tcwv_quality_flag" and by applying the "atmospheric_conditions_flag" with value equal to "1" which corresponds to cloud-free, or clear sky, situations. *Used CDR-1 dataset: $0.05^\circ \times 0.05^\circ \times \text{daily}$.* 25

Figure 4-5: The histogram with bin size 1 mm shows the distribution of all, i.e. not collocated with CDR-1, positive PWV values in the SuomiNet dataset. The median is at 16.2 mm with a maximum at 666.7 mm. The met_flag is used like above. 27

Figure 4-6: The figure shows various scatter plots of collocated TCWV values from SuomiNet and CDR-1. Panels a and b show the comparison against SuomiNet daily mean TCWV values. Panels c and d show the comparison of "all over pass" means (see also the GRUAN comparison). The results in case of the most precise temporal collocation, the "actual over pass" mean, follow in panels e and f. All CCI-data is filtered by allowing only "0" for "tcwv_quality_flag". In panels a, c, and e, CCI-data

is additionally filtered by applying the “atmospheric_conditions_flag” with value equal to “1” which corresponds to cloud-free, or clear sky, situations. In panels b, d, and f, both clear sky and partly clouded sky situations enter. *Used CDR-1 dataset: $0.05^\circ \times 0.05^\circ \times \text{daily}$.* 29

Figure 4-7: The time series’ show a comparison of CDR-1 with various references and data records used for comparison. The differences of collocated data records are reduced in their spatial components to yield MD (a), cRMSD (b), mean TCWV (c), sample size (d). Used CDR-1 datasets: $0.5^\circ \times 0.5^\circ \times \text{monthly}$ 32

Figure 4-8: The blue curves in the panels show monthly anomalies of the mean differences between CCI and the references AIRS in a) and ERA5 in b) (X_{des} in Eq. (4-1)). The regression curves (X_{reg}) are displayed in orange. The fit parameters with their uncertainties as well as the decadal slope, the overall trend, and a signal-to-noise ratio are given in the box and below the panels. Consult the main text for an explanation of the regression method. *Used CDR-1 datasets: $0.5^\circ \times 0.5^\circ \times \text{monthly}$.* 35

Figure 4-9: The blue curves in the panels show monthly anomalies of the cRMSD between CCI and the references AIRS in a) and ERA5 in b) (X_{des}). The regression curves are displayed in orange (X_{reg}). The fit parameters with their uncertainties as well as the decadal slope, the overall trend, and a signal-to-noise ratio are given in the box and below the panels. Consult the main text for an explanation of the regression method. *Used CDR-1 datasets: $0.5^\circ \times 0.5^\circ \times \text{monthly}$.* 36

Figure 4-10: The blue curves in the panels show monthly anomalies of global mean TCWV in CDR-1 in a) and monthly anomalies of the corresponding standard deviation b) (X_{des}). The regression curves are displayed in orange (X_{reg}). The fit parameters with their uncertainties as well as the decadal slope, the overall trend, and a signal-to-noise ratio are given in the box and below the panels. Consult the main text for an explanation of the regression method. *Used CDR-1 datasets: $0.5^\circ \times 0.5^\circ \times \text{monthly}$.* 36

Figure 4-11: Timeseries comparing CDR-2 to various references on a daily basis are shown. Only “clear-sky” collocated data points over land enter, i.e. the data is filtered by applying CDR-2’s “atmospheric_conditions_flag” with value “1” which means no clouds are detected in the observation window. Furthermore, the “surface_type_flag” is set equal to “0” (“land”) in order to include only data points that are also present in CDR-1 (this is ignoring inland water bodies and permanent wetlands). *Used CDR-2 datasets: $0.5^\circ \times 0.5^\circ \times \text{daily}$.* 38

Figure 4-12: The time series’ show a comparison of CDR-2 with various references and data records used for comparison. The differences of collocated data records are reduced in their spatial components to yield MD, or bias, (a), cRMSD (b), mean TCWV (c), sample size (d) over all surfaces and the entire globe. *Used CDR-2 datasets: $0.5^\circ \times 0.5^\circ \times \text{monthly}$.* 41

Figure 4-13: The blue curves in the panels show monthly anomalies of the mean differences between CCI CDR-2 (global, all surfaces) and the references AIRS in a) and ERA5 in b) (X_{des} in Eq. (4-1)). The regression curves are displayed in orange (X_{reg}). The fit parameters with their uncertainties as well as the decadal slope, the overall trend, and a signal-to-noise ratio are given in the box and below the panels. Consult the main text (section 4.1.2) for an explanation of the regression method. *Used CDR-2 datasets: $0.5^\circ \times 0.5^\circ \times \text{monthly}$.* 42

- Figure 4-14:** Global maps showing the mean differences of monthly means of CDR-2 to various data records as temporal average (no weights used in the calculations). *Used CDR-2 dataset: $0.5^\circ \times 0.5^\circ \times \text{monthly}$* 43
- Figure 4-15:** Global maps showing the cRMSD of monthly means of CDR-1 to various data records as temporal average (no weights used in the calculations). *Used CDR-1 dataset: $0.5^\circ \times 0.5^\circ \times \text{monthly}$* 44
- Figure 4-16:** Global maps showing the sample sizes of various collocated data records. Here sample size corresponds to number of months with valid data in both data records in question. *Used CDR-2 dataset: $0.5^\circ \times 0.5^\circ \times \text{monthly}$* 45
- Figure 4-17: The time series' show a comparison of CDR-2 with various references and data records used for comparison over coasts, i.e. the "surface_type_flag" is equal to "5". The differences of collocated data records are reduced in their spatial components to yield MD, or bias, (a), cRMSD (b), mean TCWV (c), sample size (d). *Used CDR-2 datasets: $0.5^\circ \times 0.5^\circ \times \text{monthly}$* 47
- Figure 4-18: The figure shows scatter plots of collocated TCWV values at coasts from GRUAN and CDR-2 in panel a) and from SuomiNet and CDR-2 in panel b). The reference data has been processed like in section 4.1.1 and only "actual over pass means" enter. The "tcwv_quality_flag" is set to "0" ("best quality"), the "atmospheric_conditions_flag" is equal to "1" ("cloud-free"), and the "surface_type_flag" always has the value "5" for "coast". *Used CDR-2 dataset: $0.05^\circ \times 0.05^\circ \times \text{daily}$* 48
- Figure 4-19: The time series' show a comparison of CDR-2 with various references and data records used for comparison over ocean, i.e. the "surface_type_flag" is equal to "1". The differences of collocated data records are reduced in their spatial components to yield MD, or bias, (a), cRMSD (b), mean TCWV (c), sample size (d). *Used CDR-2 datasets: $0.5^\circ \times 0.5^\circ \times \text{monthly}$* . Note that the vertical lines indicating changes in the constellation are not relevant over ocean. 50
- Figure 4-20: The blue curves in the panels show monthly anomalies of the mean differences between CCI and the references AIRS in a) and ERA5 in b) over ocean (X_{des} in Eq. (4-1)). The regression curves are displayed in orange (X_{reg}). The fit parameters with their uncertainties as well as the decadal slope, the overall trend, and a signal-to-noise ratio are given in the brownish box or below the panels. Consult the main text for an explanation of the regression method. *Used CDR-2 datasets: $0.5^\circ \times 0.5^\circ \times \text{monthly}$* . Note that the vertical lines indicating changes in the constellation are not relevant over ocean. 51
- Figure 4-21: The blue curves in the panels show monthly anomalies of global mean TCWV in CDR-2 in a) and monthly anomalies of the corresponding standard deviation b) over ocean (X_{des} in Eq. (4-1)). The regression curves are displayed in orange (X_{reg}). The fit parameters with their uncertainties as well as the decadal slope, the overall trend, and a signal-to-noise ratio are given in the brownish box or below the panels. Consult the main text for an explanation of the regression method. *Used CDR-1 datasets: $0.5^\circ \times 0.5^\circ \times \text{monthly}$* . Note that the vertical lines indicating changes in the constellation are not relevant over ocean. 51
- Figure 4-22: Regions where the high-resolution product is available. 53

- Figure 4-23: The bias of the regional higher resolution data to the standard CDR-2 product is shown for roi_1 as well as the relative bias with respect to standard CDR-2. *Used CDR datasets: 0.01° x 0.01° x monthly regridded to 0.05° x 0.05° x monthly, 0.05° x 0.05° x monthly (CDR-2).* 54
- Figure 4-24: The bias of the regional higher resolution data to the standard CDR-2 product is shown for roi_2 as well as the relative bias with respect to standard CDR-2. *Used CDR datasets: 0.01° x 0.01° x monthly regridded to 0.05° x 0.05° x monthly, 0.05° x 0.05° x monthly (CDR-2).* 55
- Figure 4-25: The bias of the regional higher resolution data to the standard CDR-2 product is shown for roi_3 as well as the relative bias with respect to standard CDR-2. *Used CDR datasets: 0.01° x 0.01° x monthly regridded to 0.05° x 0.05° x monthly, 0.05° x 0.05° x monthly (CDR-2).* 55
- Figure 5-1:** 2D-histogram of total expected uncertainties and absolute differences of CDR-1 and GRUAN. *Used CDR-1 dataset: 0.05° x 0.05° x daily.* 59
- Figure 5-2:** Uncertainty assessment of CDR-1 and GRUAN obtained from **Figure 5-1** following Sayer et al. (2020). The applied procedure and meaning of each curve are explained in the main text. *Used CDR-1 dataset: 0.05° x 0.05° x daily.* 61
- Figure 5-3:** Uncertainty assessment of CDR-1 and GRUAN with CDR-1 error correlation coefficient equal to 0.1. *Used CDR-1 dataset: 0.05° x 0.05° x daily.* 61
- Figure 5-4:** Uncertainty assessment of CDR-1 and GRUAN with CDR-1 error correlation coefficient equal to 0.2. *Used CDR-1 dataset: 0.05° x 0.05° x daily.* 62
- Figure 6-1:** Distribution of the clear-sky bias estimate using SuomiNet (GNSS). Cloud mask was constructed from CDR-1. *Used CDR-1 dataset: 0.05° x 0.05° x daily.* 65
- Figure 6-2: Distribution of the clear-sky bias estimate using ECMWF's ERA5 reanalysis (coloured map) and SuomiNet (coloured circles). In both cases, the cloud mask was constructed from CDR-1 in the same way. *Used CDR-1 dataset: 0.05° x 0.05° x daily.* 66
- Figure 6-3: Boxplot showing the distribution of differences in the clear-sky estimates from SuomiNet and ERA5 using the same data as in Figure 6-2. 66

1. INTRODUCTION

1.1 Purpose

This document is the Product Validation and Intercomparison Report (PVIR) for the ESA Water Vapour Climate Change Initiative project (WV_cci) Part 1. It contains the validation and intercomparison results for CDR-1, the high-resolution product and CDR-2, i.e. the atmospheric total column water vapour (TCWV) data records produced by WV_cci. This final version of the PVIR utilises CDR-1 (version 4.2), the no-modis high-resolution product (version 4.2) and CDR-2 (version 4.2). CDR-1 covers the period March 2000 to December 2023 while the high-resolution products are available as follows (ROI: region of interest):

- ROI-1 60E-80E / 20N-40N, Jun-Sep 2017-2022
- ROI-2 15W-30E / 10S-20N, Jun-Sep 2017-2022
- ROI-3 70W-40W / 0S-25S, Jun-Sep 2017-2022

CDR-2 covers the period July 2002 to December 2022. Ground-based, *in situ*, and satellite-based as well as reanalysis data records are used as reference and for comparison and are compared to the WV_cci data records as described in detail in the Product Validation Plan (PVP v5.1).

The PVIR documents are 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.

This document is PVIR Part 1.

1.2 Scope

The scope of the product validation and intercomparison efforts within the ESA WV_cci project is to assess and validate TCWV datasets as well as attached uncertainties of each product, in order to prepare datasets for the climate research community. By exploiting synergies between microwave and NIR sensors from Earth Observation missions, time series of global coverage TCWV are produced. The evaluation of TCWV and uncertainties derived from NIR sensors for land surfaces (CDR-1) is assessed prior to the assessment of the combined product of microwave and NIR TCWV data (CDR-2). Finally, the compliance with requirements is assessed.

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 2022 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 Product Validation Plan (PVP v5.1):

- 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 March 2000 to December 2023, and is a merged product of data from MERIS, MODIS and OLCI. An illustration is given in Figure 4-1 (land surfaces only!) and Figure 4-3.
- 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 2022, and combines WV_cci CDR-1 with the EUMETSAT CM SAF HOAPS (microwave imager based) data record. An illustration is given in Figure 4-1 and Figure 4-2.

- CDR-3 contains the vertically resolved water vapour ECV in units of ppmv (volume mixing ratio) and is provided as zonal monthly means on the SPARC Data Initiative latitude/pressure level grid (SPARC, 2017; Hegglin et al., 2013). It covers the vertical range between 250 hPa and 1 hPa, and the time period 1978 to the end of 2023.
- CDR-4 consists of three-dimensional vertically resolved water vapour data, covering the troposphere and lower stratosphere from 2007 to 2023. The global product includes mixing ratios in units of ppmv (mixing ratios).

2. VALIDATION STRATEGY

2.1 Datasets

An extensive evaluation of TCWV data products and their associated uncertainties has been carried out based on reference data sets and intercomparison data sets as outlined in the Product Validation Plan (PVP v5.1).

Results from the validation of Level 2 TCWV data over land using SuomiNet data as reference is discussed in section 3.

The remaining part of the PVIR focuses on the validation of the gridded Level 3 products from WV_cci, TCWV over land surfaces (CDR-1), regions of interest where high resolution data is available, as well as over coasts and globally (CDR-2). The validation is, thus, subdivided into subsections accordingly.

The validation is carried out against the reference and comparison data records:

- ERA5, AIRS, CDR-1 precursor, SuomiNet and GRUAN for global land surfaces and coasts.
- ERA5, AIRS, and CDR-1 precursor for global surfaces

2.2 Methodology

The WV_cci data products of TCWV are validated against reference datasets declared in the PVP v5.1. The methodology and metric underlying this Product Validation and Intercomparison Report (PVIR) is also defined in the PVP v5.1.

The evaluation is divided into the following tasks:

1. Validation of NIR Level 2 data records against *in situ* observations from ARM MWR data.
2. Spatial comparison between CDR-2 (and implicitly CDR-1) and comparison data records.
3. Comparison of time series between CDR-1 and CDR-2 and reference as well as comparison data records.
4. Assessment of consistency.
5. Compliance with product and user requirements.

A summary of validation statistics is given in Table 7-3 and Table 7-4.

3. VALIDATION OF NIR LEVEL 2 ALGORITHM

The quantification of uncertainties based on comparisons with ground-based measurements are demonstrated with MERIS; MODIS and OLCI Level 2 results of the WV_cci TCWV algorithm that were used as input for CDR vf4.2.

For the validation of the TCWV products, estimated from MERIS and MODIS measurements within the timeframe 2002 to 2011, we used the ARM microwave based TCWV retrievals, because of the high accuracy of this product. The cloud detection is very strict in order to exclude incorrect estimations due to undetected clouds. Only MERIS and MODIS measurements are taken which are cloud-free in an area of about 10 x 10 km² around the ARM stations. For the cloud detection, the standard L2 cloud-mask has been applied (including the cloud ambiguous and cloud margin flags). Due to the strict requirements for the comparisons between satellite and ground-based measurements the number of cases is quite limited. The comparison of MERIS and MODIS (Terra and Aqua) and ARM TCWV retrievals show a very high agreement. No wet- or dry bias is found as detected in the TCWV standard product. The correlation between the satellite retrievals and the ground truth is 0.98-0.99. The root-mean-squared-differences (RMSD) are 1.1 - 1.3 kg/m² (see Figure 3-1 and Figure 3-2).

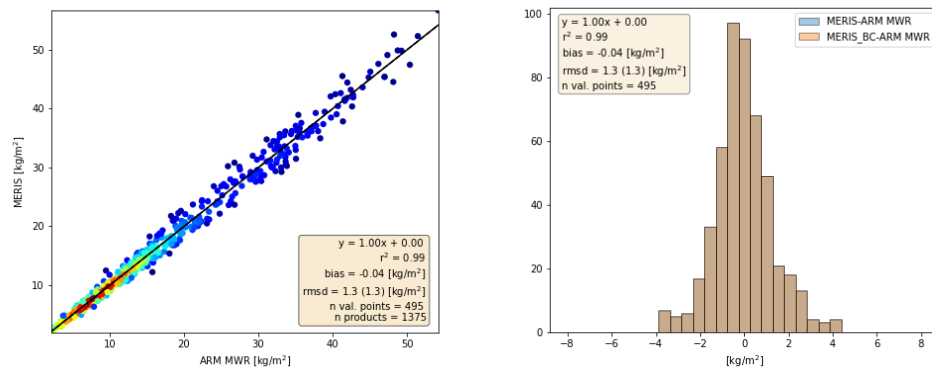


Figure 3-1: Comparison between MERIS and ARM-microwave based TCWV retrievals.

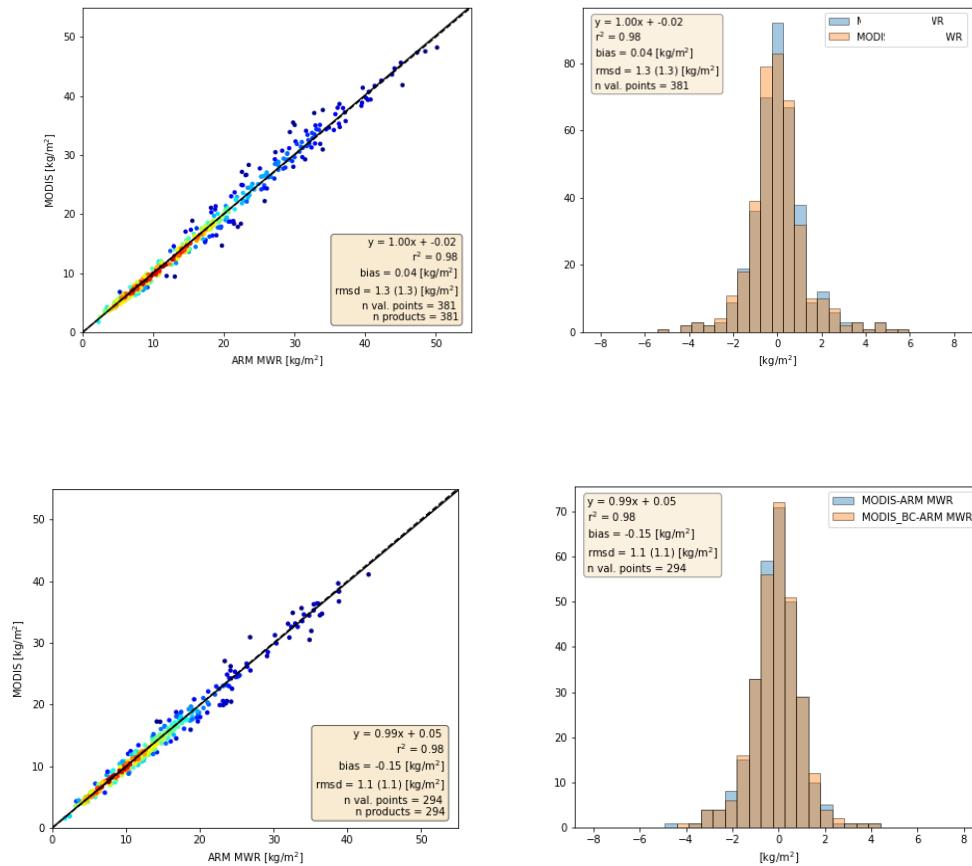


Figure 3-2: Comparison between MODIS and ARM-microwave based TCWV retrievals for the morning overpath (Terra upper plots) and the afternoon overpath (Aqua lower plots).

For OLCI A/B we took approximately 1500 matchups between January 2019 and June 2022. From these 750 matchups per satellite eventually 230-240 showed valid conditions. The correlation between the satellite product and the ground truth is 0.99. The root-mean-squared-difference is slightly better than for MERIS and MODIS 1.1 kg/m^2 and 1.2 kg/m^2 . There is no systematic overestimation contrary to 13% by OLCI as found for the standard ground-processor (see Figure 3-3).

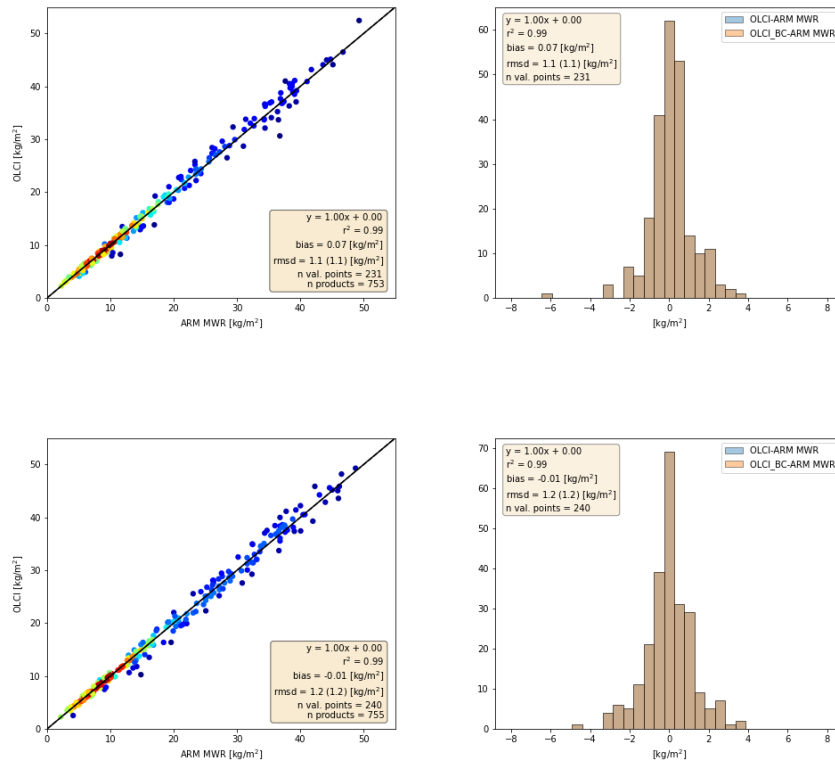


Figure 3-3: Comparison between OLCI (A upper, B lower) and SGP-ARM MWR based TCWV.

4. VALIDATION AND INTERCOMPARISON OF WV_CCI PRODUCTS

This section presents the WV_cci validation results for the TCWV product CDR-1, CDR-2 and the regional high-resolution product within the project's framework. The final versions of these datasets is version 4.2. The version is based on latest version from WV_cci phase 2 but has still room for improvements. Before starting with the presentation of the validation and intercomparison results a brief overview of the TCWV content in CDR-2's monthly data is given in the form of time series, maps and some basic overall statistics.

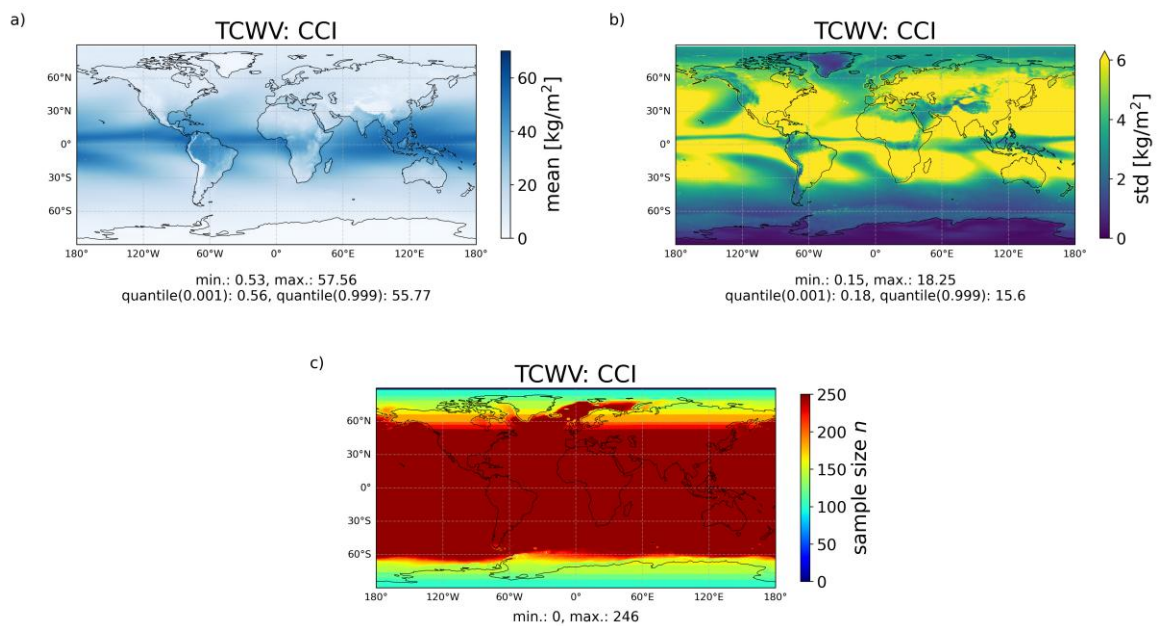


Figure 4-1: a) is a global map of the temporal mean of valid TCWV values with all monthly data points equally weighted, b) shows the corresponding standard deviation. c) displays the sample size, or number of months with a valid TCWV value, per grid. *Used CDR-2 dataset: 0.5° x 0.5° x monthly.*

The map in Figure 4-1 a) shows the distribution of TCWV as temporal mean deduced from the monthly means in CDR-2. As expected, the highest water vapour content can be found near the equator and lower latitudes in general. Towards the poles the mean TCWV decreases. On a regional scale, TCWV as recorded in CDR-2 depends on the topography and/or climatological conditions. In Figure 4-1 b) the corresponding standard deviation is shown. It is high in various places on the globe. Deseasonalising has shown that fluctuations in higher latitudes are mainly due to seasonal changes while the variation in lower latitudes is inter-annual. Panel c) in Figure 4-1 shows the number of valid data points that feed into the statistical quantities of the other panels at each location. Here, this number is the number of months with a valid monthly mean within the dataset's time range at a

given grid cell. Towards the poles the sample size for grid cells on land is the lowest due to the inapplicability of the near-infrared retrieval algorithm in the absence of sunlight during the polar night.

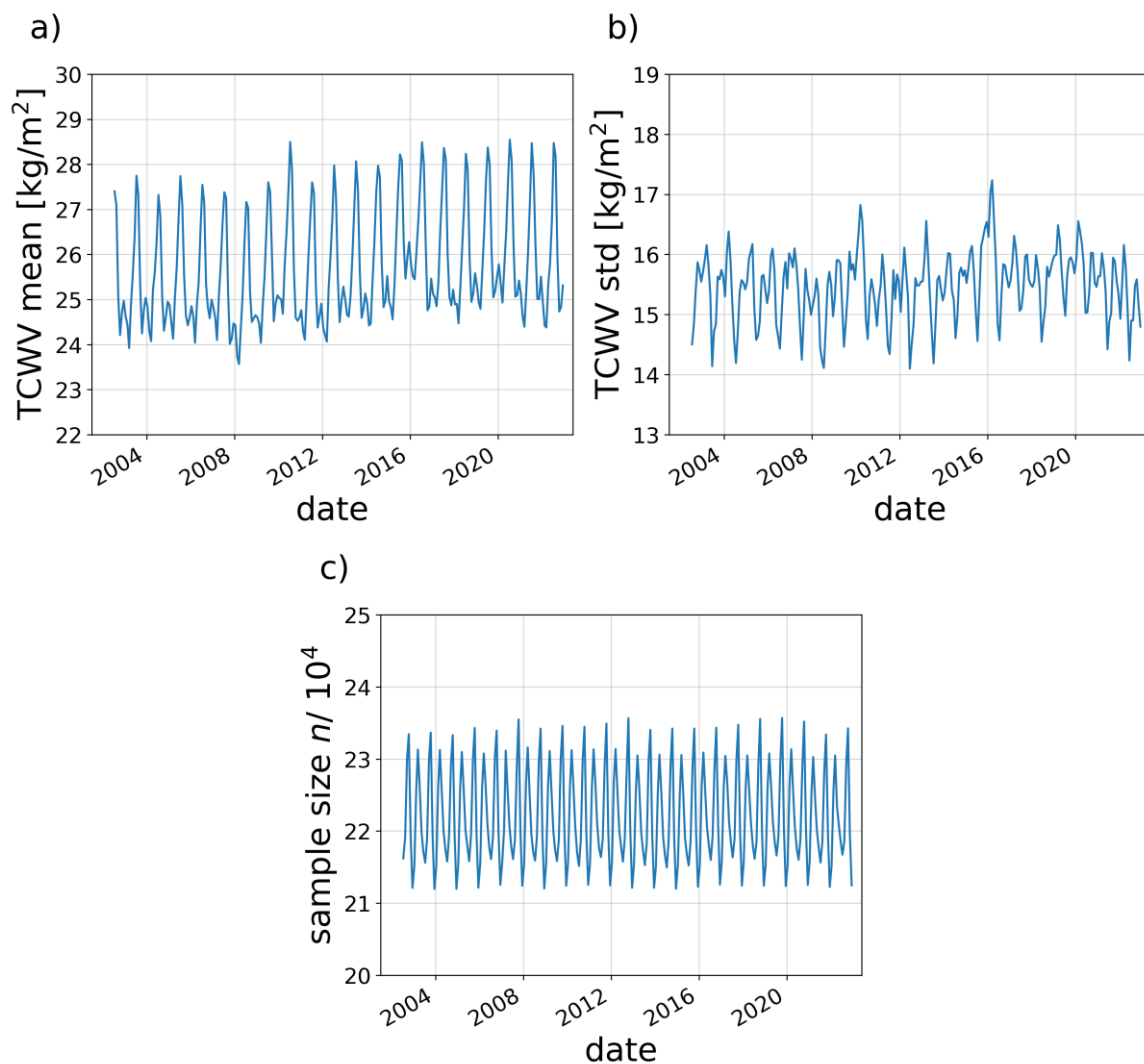


Figure 4-2: a) is a time series of the global spatial mean of valid TCWV values with all monthly data points equally weighted, b) shows the corresponding standard deviation. c) displays the sample size, or number of months with a valid TCWV value, per grid. *Used CDR-2 dataset: 0.5° x 0.5° x monthly.*

The time series in Figure 4-2 show the temporal evolution of global, latitude-weighted mean TCWV in a) and of the corresponding standard deviation in b), as well as the course of the sample size consisting of the number grid cells with a valid TCWV value over time. Again, all data from CDR-2 dataset of monthly mean TCWV on a regular 0.5° grid enter this analysis. All curves show a strong seasonality. Two factors that likely contribute to this seasonality are i) the seasonally varying spatial coverage since the NIR-retrieval requires

the presence of solar insolation (see Figure 4-2 c)), and ii) characteristic seasonal variations in the predominant atmospheric conditions at specific locations (for example the transitions from winter to summer in higher latitudes, and the monsoon season in, e.g., India and Pakistan).

The mean TCWV obtained from the time series in Figure 4-2 a) is $25.68 \pm 1.2 \text{ kg/m}^2$.

4.1 Validation over global land surfaces (CDR-1)

The validation of the first preliminary level 3 (L3) data record over global land surfaces CDR-1 (v4.2) was carried out against data records from the global observation networks GRUAN and SuomiNet. The CDR-1 product is also compared to AIRS and ERA5 data records for global land surfaces as well as to the precursor version of CDR-1 from WV_cci Phase 1.

The GRUAN dataset consists of radiosonde observations measuring profiles of atmospheric variables that allow for the calculation of TCWV. Data measured by two radiosonde types, RS-92 and its successor RS-41, is used. From amongst the datasets used for validation and comparison in this report the GRUAN dataset is the only one that can be considered fiducial. Unfortunately, the spatial and temporal coverage of this network is rather low. Analysis of valid collocations between the radiosonde observation sites and the CDR-1 gridded data show only few collocations per day and even per month. The respective statistical measures of individual daily and monthly statistics are limited in their applicability to the full CDR. Thus, only scatter plots of associated results are shown in Figure 4-4. Each GRUAN station covers a time period lying in-between May 2006 and December 2023 depending on the station in question.

The SuomiNet network consists of globally distributed permanent stations of GPS receivers conducting real-time atmospheric precipitable water vapour measurements. The number of stations is substantially higher than in the GRUAN dataset with locations predominantly situated in North America. Furthermore, measurements are carried out every 30 minutes. Since model data partly based on satellite observations feeds into the calculation of the TCWV values, the SuomiNet dataset cannot be considered as completely independent of satellite data. Scatter plots are shown in Figure 4-6. Again, the time period covered by an individual station depends on the choice of station with earliest available measurements taken in January 2005 while the latest measurements were taken in December 2021.

Note that time series per SuomiNet station are frequently affected by data gaps (see the Phase I PVIR). Such variability might hamper the interpretation of results. We chose to omit showing a time series comparing SuomiNet and CDR-1 because the highly fluctuating sample size of collocated data points does not permit a stability analysis of CDR-1.

The AIRS global data record, retrieved from data of NASA's AQUA satellite mission, is a level 3 data product providing information on the total column of water vapour on a $1^\circ \times 1^\circ$ grid on daily and monthly basis. Because the combined product AIRS/AMSU is only supported up to mid 2016 the AIRS-only products AIRS3STD (v7) and AIRS3STM (v7) are used. Uncertainties are included in the form of standard deviations. With data available from September 2002 to today it covers almost the complete time range of CDR-1.

ERA5 is ECMWF's atmospheric reanalyses on a spatial $0.25^\circ \times 0.25^\circ$ grid with hourly temporal resolution for various physical variables including TCWV. It does not come with uncertainties by itself. They can be computed from a gridded 10-member ensemble. This is not carried out here due to the low number of ensemble members. The ERA5 data record covers the complete time range of CDR-1.

The precursor version of CDR-1 was produced in the previous phase of the WV_cci project and provides information on TCWV including various uncertainties. It is available from July 2002 to December 2017 on $0.05^\circ \times 0.05^\circ$ and $0.5^\circ \times 0.5^\circ$ grids with data representing daily and monthly means.

The comparison to all non-station-based data records include global maps and time series of bias and cRMSD along with a stability analysis using the bias timeseries. Furthermore, trends are analysed in the comparison time series as well as in time series of CCI data alone.

As for terminology when comparing datasets, sometimes the name "reference TCWV" or "comparison TCWV" is used to refer to the TCWV value coming from the reference or comparison dataset, and the name "CCI-TCWV" is used when referring to a TCWV value from a CCI-dataset.

The validation using station-based data and the comparison with the global data records is carried out for the CDR-1 product in the next sections 4.1.1 and 4.1.2. Consistency assessments, a clear-sky bias analysis and compliance checks are carried out in sections 5, 6 and 7, respectively.

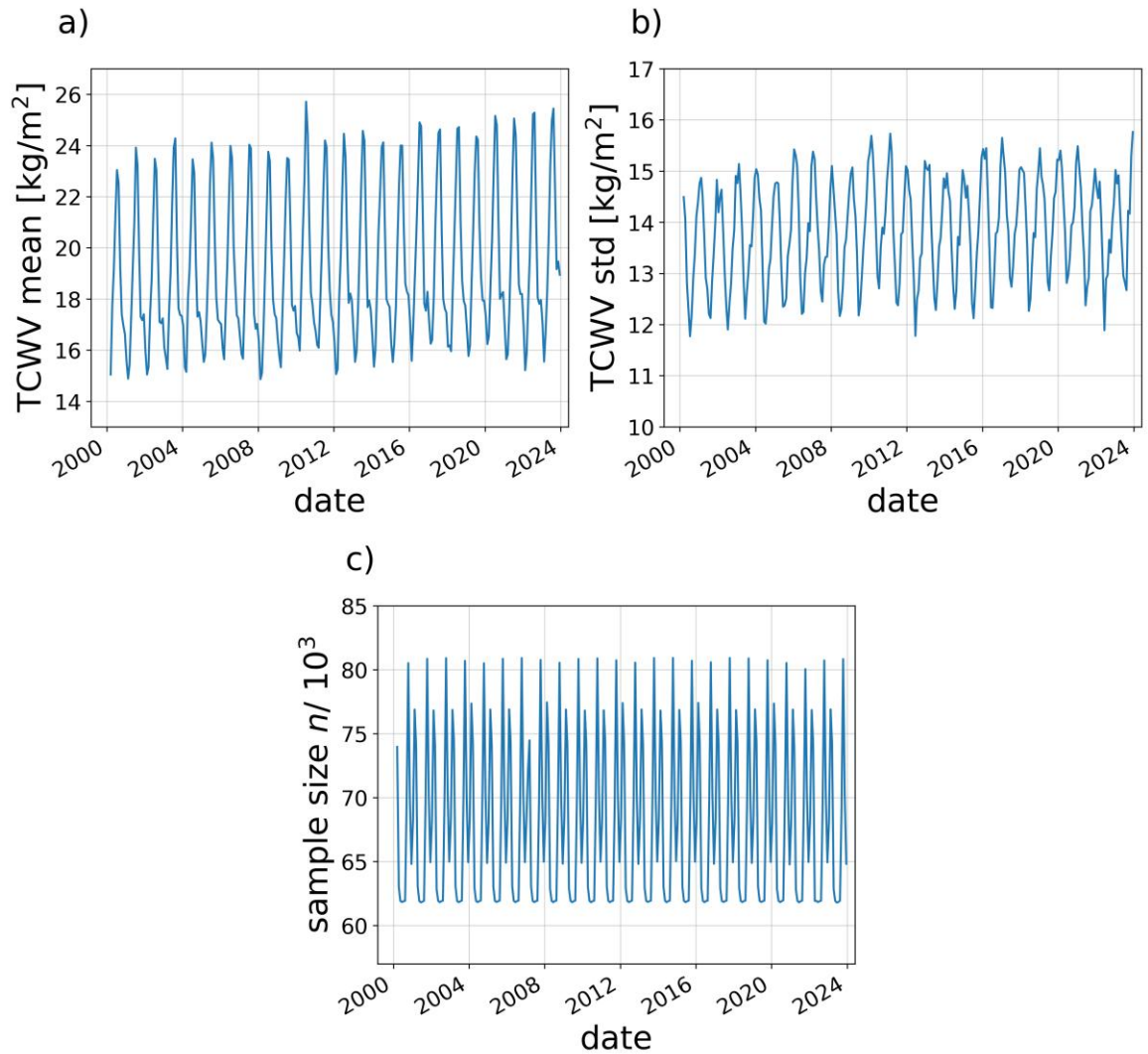


Figure 4-3: a) is a time series of the spatial mean of valid TCWV values over global land with all monthly data points equally weighted, b) shows the corresponding standard deviation. c) displays the sample size, or number of months with a valid TCWV value, per grid. *Used CDR-1 dataset: $0.5^\circ \times 0.5^\circ \times$ monthly.*

Figure 4-3 complements the previous display of CCI-CDRs by time series obtained from data over land alone, as is the case for valid data points in CDR-1. Comparing the time series for mean TCWV, panel a) in Figure 4-2 and Figure 4-3, one finds that considering grid cells on land alone yields in general a smaller global mean TCWV than when taking data from all surfaces into account.

The mean TCWV obtained from the time series in Figure 4-3 a) is 19.2 ± 2.9 kg/m².

4.1.1 Validation using station-based data

In this section the validation of CDR-1 using the station-based datasets is reported. Both, GRUAN and Suominet, were collocated with the CDR-1 of resolution $0.05^\circ \times 0.05^\circ \times$ daily. The application of filters is mentioned separately for each reference dataset.

Before presenting the validation results, the temporal collocation is discussed since it is the same for both datasets GRUAN and SuomiNet. As reference TCWV value for a station at a given UTC day, three differently constructed means are used throughout this validation and compared to CCI-data on the given day and location:

1. There is a “daily mean” for which all measurements on a given UTC-day feed into the final reference TCWV value.
2. Furthermore, there is a mean, “all overpass mean”, where those measurements enter the TCWV computation which lie within a certain time window of (UTC-)overpass times of those platforms that have fed into a corresponding CDR-1 daily file. The ESA-CCI filename conventions is used to determine which platforms are contributing at a given day, anywhere on the globe. Whether a considered platform has *actually* passed a station is not taken into account. In local time the overpass time, or equator crossing times, of MERIS, OLCI-A, OLCI-B is 10:00 a.m., the overpass time of MODIS-TERRA is 10:30 a.m., and the overpass of MODIS-AQUA occurs at 1:30 p.m.. These times are approximate values since the true overpass times vary due to eccentricity of the Earth and shifts within the orbits. These facts may result in an increase of the temporal collocation error. The overpass times are converted to UTC-time-of-day on the given UTC-day at each location.
3. Finally, the last construction of a reference TCWV mean is refined and further localized by using the “num_obs_*” variables that contain the number of observations per instrument contributing to the mean CCI-TCWV value in a given grid cell. If an instrument’s number of observations in a given grid cell is non-zero then its overpass time is considered in the temporal collocation with reference measurements. Hence, compared to the previous construction, an overpass time is dropped if the corresponding instrument did not feed into the grid cell’s CCI-TCWV value.

Validation using GRUAN:

Each file in the GRUAN dataset contains a single ascent at a single station. Most information used for the following validation is being stored in the global attributes as

strings. Only the measurement position of a RS-92 ascent is taken directly from the sonde's trajectory. A name glossary is provided in Table 4-1.

Table 4-1: Name conventions in GRUAN files for RS-92 and RS-41 radiosonde datasets, see also the GRUAN technical note 13 (https://www.gruan.org/gruan/editor/documents/gruan/GRUAN-TN-13_RS41-GDP_UserGuide_v1.0_2022-11-21.pdf).

Sonde type	RS-92	RS-41
Time	g.Ascent.StartTime	g.Measurement.StartTime
Station position	g.MeasuringSystem.Latitude g.MeasuringSystem.Longitude	g.Site.Latitude g.Site.Longitude
Measurement position	First non-NaN value in "lat" and "lon" variable of NetCDF-File	g.MeasurementSystem.Latitude g.MeasurementSystem.Longitude
TCWV [kg/m ²]	g.Ascent.PrecipitableWaterColumn	g.Measurement.PrecipitableWaterColumn
TCWV uncertainty [kg/m ²]	g.Ascent.PrecipitableWaterColumnU	g.Measurement.PrecipitableWaterColumnUc

The collocation procedure of GRUAN with CDR-1 of resolution 0.05° x 0.05° x daily is the following: The two datasets are specially collocated by searching the 0.05° grid cell containing the GRUAN site in question. Any measurement whose sonde's starting point lies within 10 km of the station's position is considered for the comparison statistics.

The time window for the temporal collocation of GRUAN measurements and a selected overpass time of platforms feeding into CCI-data is 90 minutes. Only results where GRUAN data is averaged as an "actual over pass" mean are analysed here.

The precision of the above collocation procedure might be influenced by the following: A GRUAN measurement is not an instant point measurement of TCWV but rather a path integral of water vapour along the sonde's trajectory. Hence, measurements at different times and over different geolocations, possibly spanning over multiple grid cells, enter into GRUAN TCWV values. Furthermore, (changes in) local weather conditions, in particular wind strength and direction, interfere with the spatial and temporal collocation of GRUAN data with CCI-data due to, e.g., the drift of the balloon during its ascent. Because most of

the atmosphere's water vapour is contained in the lower atmosphere, TCWV recorded in GRUAN files is nonetheless interpreted as an instant point measurement over the station in question and believed to be precise enough to serve as a reference for validation purposes.

In summary, collocation criteria are chosen in such a way that the GRUAN measurements and the satellite observations on which the CDR-1 data record is based are close in space and time.

a) actual over passes, clear sky

b) actual over passes, partly cloudy

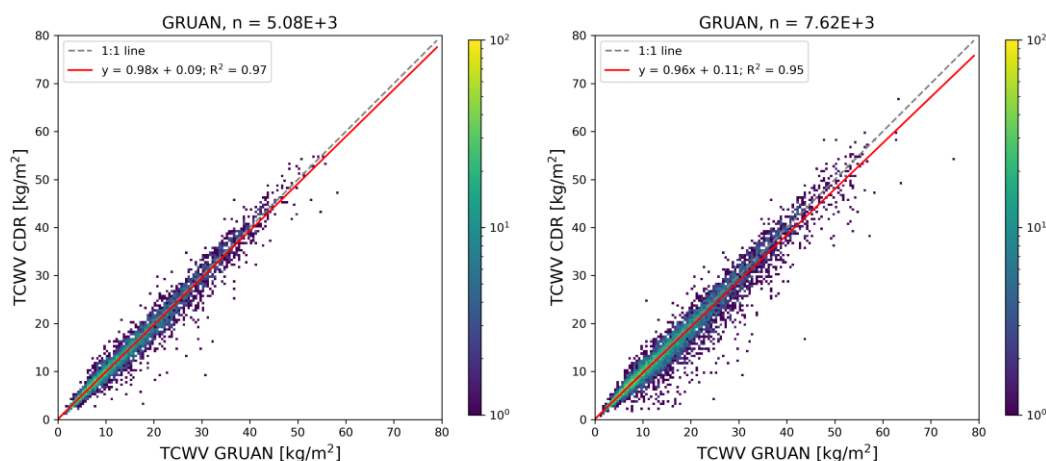


Figure 4-4: The figure shows two scatter plots of collocated TCWV values from GRUAN and CDR-1. Only means obtained from measurements close to actual over pass enter on the GRUAN side. All CCI-data is filtered by allowing only “0” for “tcwv_quality_flag”. In the left panel CCI-data is additionally filtered by applying the “atmospheric_conditions_flag” with value equal to “1” which corresponds to cloud-free, or clear sky, situations. In the right panel both clear sky and partly cloudy sky situations enter. *Used CDR-1 dataset: 0.05° x 0.05° x daily.*

The scatter plots in Figure 4-4 show the relation of TCWV in CDR-1 to TCWV in GRUAN at collocated data points and a generally good quality of CDR-1 in the sense that it reproduces the GRUAN data well. In addition to the GRUAN “tcwv_quality_flag” being equal to “0”, the data is filtered by demanding the qualitative CDR-1 variable “atmospheric_conditions_flag” to be equal to “1”, which corresponds to a cloud-free satellite observation with best possible quality. Comparing the scatter plot in panel a) to the one in panel b) one finds that the application of a “clear-sky” flag (panel a) leads to a better agreement of GRUAN and CDR-1 data which is manifested in the slope of the regression line being equal to 0.98 in panel a) instead of 0.96 in the partly cloudy case of panel b). This finding is in principle expected and hence confirms the applicability of the atmospheric

conditions flag. At the same time, the clear-sky filter reduces the sample size of collocated data points from approximately 7600 in panel b) to 5000 in panel a). Note that coasts are not part of CDR-1 (but see the scatter plots in Figure 4-18).

Table 4-2: Basic statistical measures are shown for the data in Figure 4-4 with TCWV values in GRUAN subtracted from TCWV values in CDR-1, abbreviated as “CDR1 – GRUAN”. The relative quantities are obtained by dividing their absolute counter parts through the mean TCWV in collocated GRUAN data. *Used CDR-1 dataset: 0.05° x 0.05° x daily.*

CDR1 – GRUAN		Bias [kg/m ²]	Bias_rel (absolute) [%]	cRMSD [kg/m ²]	cRMSD_rel [%]	TCWV ref. [kg/m ²]	Sample size
Actual over passes	Clear sky	-0.22	1.4	1.78	11.1	16.07	5083
	Partly clouded	-0.61	3.6	2.32	13.6	17.05	7623

Table 4-2 compares the TCWV content in CDR-1 to the one in GRUAN at collocated data points and reduces the data in Figure 4-4 to basic statistical measures. The better agreement of CDR-1 with GRUAN when applying the clear-sky mask is reflected in the decrease of (absolute) relative bias from 3.6% to 1.4% for mean GRUAN data obtained at actual over passes. Activating the “atmospheric_conditions_flag” with value “1” also results in a smaller variability as measured by the cRMSD but the latter remains above 10% of the mean TCWV content in collocated GRUAN data exceeding the relative bias, regardless of flags being used. Also note the drop in overall bias in Table 4-2 when moving to clear-sky conditions only. This hints towards an underestimation of TCWV in CCI data in partly clouded situations.

Validation using SuomiNet:

A SuomiNet file contains for each day the measurements of all stations registered in the network at the chosen date. There are at most 48 measurements per day and station starting at 0:00 AM local time with a duration of 30 minutes. TCWV values are stored in the variable “pwv” as a function of “station” and “time_offset” with units [mm] and with an uncertainty estimate named “pwv_err”.

The preparation of the SuomiNet dataset and the collocation procedure are as follows: A SuomiNet measurement is considered if

- $0 < PWV < 80$ mm;

- “met_flag” is neither “U” nor “X”.

Concerning the criteria for “met_flag”, preliminary analysis has shown that for many stations and on many days predating 2014 either a value assigned to “met_flag” does not follow its documentation or no value is assigned to “met_flag” at all. The above described filtering method does not take care of these cases, and a data point is dropped if and only if the “met_flag” is correctly indicating questionable data quality.

The choice of an upper threshold at 80 mm is made to avoid unrealistic high PWV values in SuomiNet data. It is derived from the histogram showing the distribution of valid PWV entries in the SuomiNet dataset (see Figure 4-5).

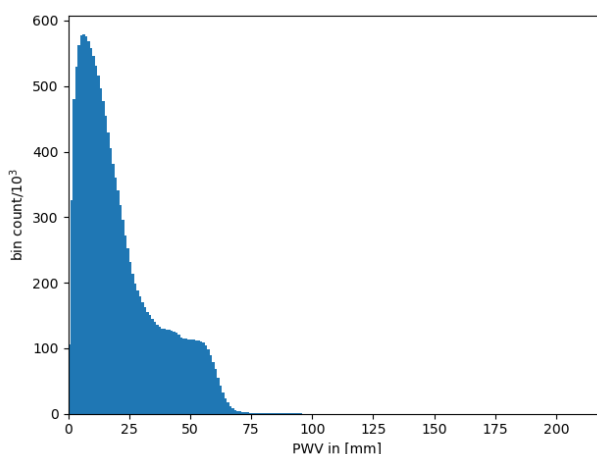


Figure 4-5: The histogram with bin size 1 mm shows the distribution of all, i.e. not collocated with CDR-1, positive PWV values in the SuomiNet dataset. The median is at 16.2 mm with a maximum at 666.7 mm. The met_flag is used like above.

Data preparation is finalized by the calculation of the various means described in the beginning of this section: a daily mean, an “all over pass” mean and an “actual over pass” mean. The “all over pass” mean is included here, too, since the clear-sky-bias analysis which follows in chapter 6 below, is based on this temporal sampling. The collocation time window is 60 minutes for matching over passes and SuomiNet measurements. It is smaller than the time window for GRUAN since the measurement frequency in SuomiNet data is much higher (every 30 minutes compared to one or only a few radiosonde ascents in GRUAN per day and station).

The spaial collocation procedure of SuomiNet data and CDR-1 of resolution $0.05^\circ \times 0.05^\circ \times$ daily is the following: On a given day the daily averaged data of a SuomiNet station is collocated with a CDR-1 data point at the same day if the station under consideration lies in the $0.05^\circ \times 0.05^\circ$ grid centred at the CDR-1’s data point geolocation.

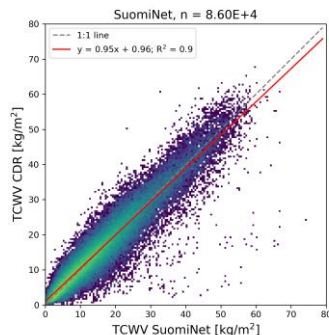
Some issues influence the representatives of the collocation procedure applied when comparing CDR-1 and SuomiNet: The data in SuomiNet is derived from temporal delays in GNSS-signals exchanged between the station and various nearby satellites. On the one hand the path of these signals is rarely straight up from the station's location, on the other hand it is, in a sense, an average derived from measurements in different directions depending on the satellites' locations and orbits. This means a measurement entering into SuomiNet data is again not a point measurement. Furthermore, recorded values are 30-minute averages which might increase uncertainties in temporal collocations.

The quality of TCWV data in SuomiNet is also affected by uncertainties in meteorological quantities feeding into the derivation of TCWV from signal delays. The associated "met_flag" is, as discussed above, not always present or reliable.

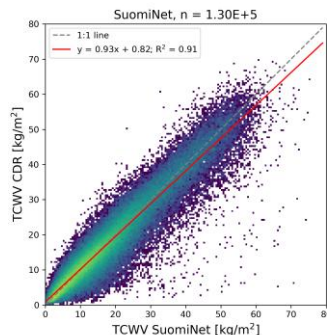
Finally, the surrounding topography of a station is another source of uncertainty in the comparison between GNSS-TCWV and TCWV derived from satellite measurements (Roman et al. 2016).

Still, it is a station-based dataset with TCWV values obtained by an established and well-studied method. Hence, it is fit for the purpose of being a rather reliable source of comparison TCWV values for the validation of TCWV data in CDR-1.

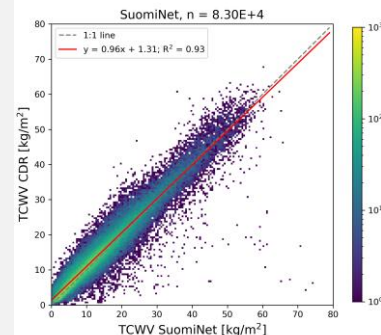
a) daily mean, clear sky



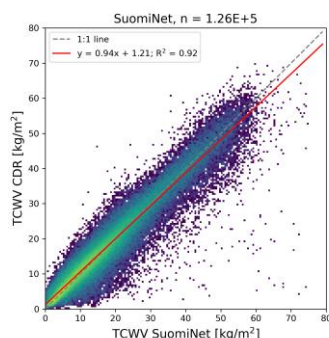
b) daily mean, partly cloudy



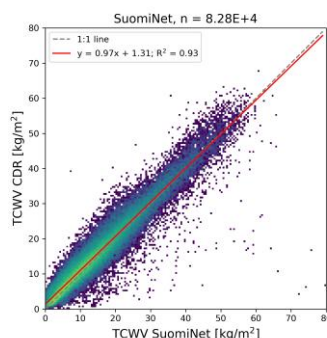
c) all over passes, clear sky



d) all over passes, partly cloudy



e) actual over passes, clear sky



f) actual over passes, partly cloudy

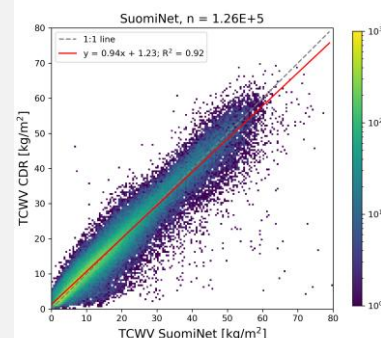


Figure 4-6: The figure shows various scatter plots of collocated TCWV values from SuomiNet and CDR-1. Panels a and b show the comparison against SuomiNet daily mean TCWV values. Panels c and d show the comparison of “all over pass” means (see also the GRUAN comparison). The results in case of the most precise temporal collocation, the “actual over pass” mean, follow in panels e and f. All CCI-data is filtered by allowing only “0” for “tcwv_quality_flag”. In panels a, c, and e, CCI-data is additionally filtered by applying the “atmospheric_conditions_flag” with value equal to “1” which corresponds to cloud-free, or clear sky, situations. In panels b, d, and f, both clear sky and partly cloudy sky situations enter. *Used CDR-1 dataset: 0.05° x 0.05° x daily.*

In Figure 4-6 CDR-1 is compared to SuomiNet at collocated data points. The scatter plots show rather good agreement between TCWV values of the two datasets with visible improvements if CDR-1’s “atmospheric_conditions_flag” is used which is the case for panels a, c, and e. This fact can also be deduced from the decrease in slopes, clear-sky to partly-clouded, of the regression lines which are 0.95 to 0.92 from panel a to b, 0.96 to 0.94 from panel c to d, and 0.97 to 0.94 from panel e to f, respectively. With the exception of the panels a and b, there is also an improved goodness-of-fit (R^2). In all scatter plots there are some outliers most of them correlating with high SuomiNet-TCWV.

The low bias and the relatively high cRMSD in Table 4-3 confirm the visual impression of a “cloud” of data following the 1:1 line whose width decrease with better temporal collocation and higher CCI-data quality. This is reflected in the decreasing values for cRMSD in Table 4-3. Furthermore, the decreasing cRMSD from 3.65 kg/m² to 2.75 kg/m² hints towards a reduction of the random component in differences between SuomiNet and CDR-1 if the temporal collocation between SuomiNet measurements and CCI data is improved, and CCI data of presumably highest quality is selected. On the same route, the systematic component in differences between SuomiNet and CDR-1, the bias, increase from its lowest value at -0.27 kg/m² in the “daily mean and partly clouded” comparison to its highest value at 0.87 kg/m² in the “actual over pass and clear sky” comparison, as shown in Table 4-3. This increase can be taken as a hint that, on the one hand and similar to the validation using GRUAN, in partly clouded situations the CCI-data underestimates the presumably high TCWV (high since some clouds are present, see the chapter on the clear sky bias assessment), and, on the other hand, the over pass times around noon do not encompass those times with the highest diurnal TCWV values.

Finally, note that sample sizes shown for SuomiNet in Table 4-3 are substantially higher compared to those in the validation using GRUAN.

Table 4-3: This table and Figure 4-6 are related to each other in a similar way as Table 4-2 and Figure 4-4. The panels in Figure 4-6 and rows in this tables can be identified by the respective headers. The relative quantities are obtained by dividing their absolute counter parts through the mean TCWV in collocated SuomiNet data. Used CDR-1 dataset: 0.05° x 0.05° x daily.

CDR1 – SuomiNet		Bias [kg/m ²]	Bias_rel (absolute) [%]	cRMSD [kg/m ²]	cRMSD_rel [%]	TCWV ref. [kg/m ²]	Sample size
Daily mean	Clear sky	0.18	1.2	3.22	21.7	14.87	85983
	Partly clouded	-0.27	1.6	3.65	22.1	16.51	130146
All over passes	Clear sky	0.8	5.6	2.77	19.4	14.27	83018
	Partly clouded	0.24	1.5	3.3	20.6	16.04	126007
	Clear sky	0.87	6.1	2.75	19.4	14.21	82837

Actual over passes	Partly clouded	0.33	2.1	3.28	20.5	15.96	125621
--------------------	----------------	------	-----	------	------	-------	--------

4.1.2 Intercomparison with long-term data records

This section contains the intercomparison of CDR-1 with spatial resolution at 0.5° with the long-term data records AIRS and ERA5. A comparison with the precursor version of CDR-1 is included (labelled as “CCI_pre”), too. ERA5 is averaged from $0.25^\circ \times 0.25^\circ$ to the CDR-1 $0.5^\circ \times 0.5^\circ$ grid while AIRS is oversampled from $1^\circ \times 1^\circ$ to the CDR-1 grid. In all cases monthly means of TCWV retrievals are used if not otherwise stated.

Regridding routines:

ERA5's TCWV product that enters into the intercomparison with CDR1 in this section was download as monthly means on a $0.25^\circ \times 0.25^\circ$ grid. The product was regridded to a 0.5° regular grid with the TCWV value of each grid representing a spatial mean taken in a $0.5^\circ \times 0.5^\circ$ grid cell centred at the grid point's location. The spatial means were obtained by averaging over all data points included in a chosen grid cell with appropriate weights applied to those points lying inside the new grid cell (1.0), on edges (0.5) or on corners (0.25) such that each cell of the original grid contributes to the regridded version to the same degree regardless of its situation truly inside a coarser grid cell or on the boundary of several of these.

In the same way the ERA5 TCWV hourly product with regular 0.25° grid was regridded to $0.5^\circ \times 0.5^\circ \times$ daily. Besides spatial regridding also temporal averaging was carried out.

The daily and monthly AIRS was oversampled in a straightforward way from its original 1° resolution to 0.5° where the TCWV value in an original 1° grid cell is passed to the four 0.5° sub-grid cells without change. The ascending orbit of AIRS, corresponding to the noon over pass, is used to allow for a closer temporal collocation with CDR-1 over land.

Intercomparison results for monthly data:

The time series in Figure 4-7 show the temporal evolution of various quantities derived from differences per grid cell between CDR-1 and the above-mentioned monthly data records. Weights according to latitude are applied.

Although the comparison data records cannot be taken as a reference, the temporal mean of those time series concerned with spatial mean differences, short MD, will be referred to as “bias”.

Time series of the mean differences can be found in Figure 4-7 a). The temporal evolution of cRMSD is displayed in panel b). Panels c) and d) contain time series of global mean TCWV and number of collocated data points (grid cells), i.e. the sample size, respectively. Basic statistical measures associated to the time series in Figure 4-7 can be found in Table 4-4. The dashed vertical grey lines mark changes in the used sensor combination in CDR-1.

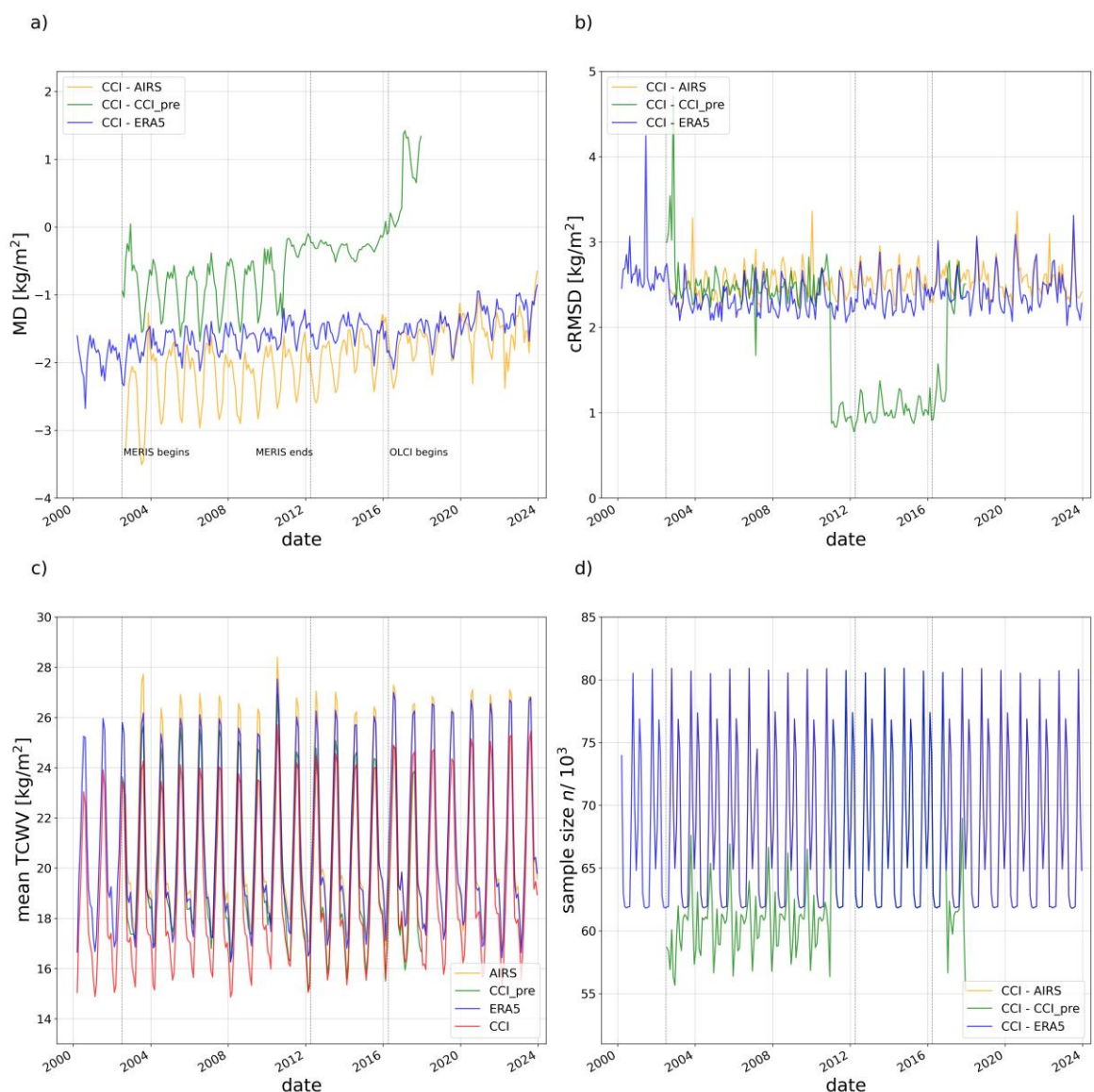


Figure 4-7: The time series' show a comparison of CDR-1 with various references and data records used for comparison. The differences of collocated data records are reduced in

their spatial components to yield MD (a), cRMSD (b), mean TCWV (c), sample size (d).
Used CDR-1 datasets: $0.5^\circ \times 0.5^\circ \times$ monthly.

There is a negative bias when comparing CDR-1 to AIRS and ERA5 with mean biases of $-1.9 \pm 0.5 \text{ kg/m}^2$ and $-1.6 \pm 0.3 \text{ kg/m}^2$, respectively.

CDR-1 of the project's current phase is based on MERIS (until 2012), MODIS (full time series) and OLCI (starting 2016) with changes in utilised sensors in CDR-1 being marked in the panels of Figure 4-7. It was shown in the validation report of the precursor version (PVIR v2.1) that the monthly timeseries is affected by a breakpoint when MODIS enters if the reference data is not restricted to clear-sky data. It further showed a break with OLCI data coming in. The issues in the precursor version show up as seeming breakpoint in the CDR-CDR_pre timeseries. In view of the high stability relative to AIRS and ERA5 an improvement in stability can be noted for CDR-1 from WV_cci phase 2, relative to its precursor version. Basic statistics from the comparison in Figure 4-7 are contained in Table 4-4.

Table 4-4: Basic statistical measures derived from the time series in the panels a), b) and c) of Figure 4-7. The relative values are taken with respect to the mean TCWV of the reference. The given uncertainties are the standard deviations of the timeseries. The slopes are obtained from the regression curves in Figure 4-8. *Used CDR-1 datasets: $0.5^\circ \times 0.5^\circ \times$ monthly.*

Reference	Mean MD [kg/m ²]	Mean MD_rel (absolute) [%]	Slope (bias, linear fit) [kg/m ² /dec]	Slope_rel [%/dec]	Mean cRMSD [kg/m ²]	Mean cRMSD_rel [%]	Mean TCWV [kg/m ²]
ERA5	-1.6 ± 0.3	7.6 ± 1.3	0.3	1.2	2.4 ± 0.2	11.5 ± 1.1	20.8 ± 3.0
AIRS	-1.9 ± 0.5	9.2 ± 2.2	0.5	2.4	2.5 ± 0.2	11.9 ± 1.0	21.2 ± 3.2
CCI_pre	-0.5 ± 0.6	2.5 ± 3.0			1.9 ± 0.8	9.8 ± 3.8	19.9 ± 2.9

Trend analysis for monthly data:

In the following, trends in the timeseries shown above are analysed. The fit procedure to obtain the regression curves is the following:

Given a time series $X = X(t)$ with monthly spaced data, monthly anomalies are obtained yielding a new deseasonalised time series X_{des} . Then a simple model is regressed to the

time series of monthly anomalies using ordinary least squares regression yielding X_{reg} . The model consists of an intercept with coefficient c_{int} , a linear contribution with slope c_{slo} , and an ENSO index X_{MEIV2} , here MEIV2 as provided by NOAA, with coefficient c_{MEIV2} together with its derivative $\frac{d}{dt}X_{MEIV2}$ with coefficient c_{MEIV2_d} . The differences between the data and the fit $X_{noise} = X_{des} - X_{reg}$ are the residuals and interpreted as “noise”. The model formula is

$X_{des}(t) = \underbrace{c_{int} + c_{slo} \cdot t + c_{MEIV2} \cdot X_{MEIV2}(t) + c_{MEIV2_d} \cdot X_{MEIV2_d}(t)}_{X_{reg}(t)} + X_{noise}(t)$	4-1
---	-----

For each individual curve, a shift $shift_{MEIV2}$ is determined for the ENSO index by maximizing the absolute value of the correlation cor_{MEIV2} between X_{des} and a shifted version of the index time series X_{MEIV2} . Shifts are always carried out backwards in time, ($t \rightarrow X_{MEIV2}(t - shift_{MEIV2})$) and only assume positive integer values (i.e., full months). In this manner, possibly delayed effects of ENSO on TCWV are taken into account. The derivative of the ENSO index is included in the regression model to detect effects depending on changes in the ENSO signal. Both, the ENSO index and its derivative, are normalized with respect to each curve’s absolute maximum.

The result of the regression method is then summarized in plots like in Figure 4-8. In each panel, a box is included showing the goodness-of-fit R^2 , the value of the coefficients together with uncertainty estimates obtained from the covariance matrix of the OLS-regression, as well as the shift applied to the ENSO signal and the correlation of the deseasonalised data and the shifted ENSO index. Below each panel the decadal slope $s_{dec} = \frac{c_{slo}}{12 \cdot 10}$, the overall trend, or gain, $trend = c_{slo} \cdot \Delta t$, where Δt is the length of the time series, and a signal-to-noise ratio $stnr = \frac{trend}{RMSD}$ with $RMSD$ equal to the root mean square of the residuals, the differences of original data and the regression curve, that is RMS of X_{noise} . The statistics are also summarized in Table 4-4.

As per the slopes of the regression curves of the mean differences in Figure 4-8, a positive trend can be detected in the MD monthly anomalies of CDR-1 against both references ERA5 and AIRS. Relative to the mean TCWV of the reference the fits have relative slopes of 2.4 %/dec for AIRS and 1.2 %/dec for ERA5. Hence the MD drift only slightly and CDR-1 is relatively stable when compared to AIRS and ERA5. An application of the regression method to CDR-1 only (i.e., not on differences between CDR-1 and a reference dataset, but on absolute CDR-1 values, see Figure 4-10) yields a decadal slope of 0.6 kg/m²/dec which is more than twice as large as the decadal slope of 0.25 kg/m²/dec in the ERA5 comparison (compare Figure 4-10 a) and Figure 4-8 b)). Furthermore, in Figure 4-10 a),

one finds that the slope coefficient is relatively large when compared to its (model) uncertainty which is also reflected in the large signal-to-noise-ratio. This suggests that the trend in TCWV in CDR-1 is significant and that this trend is seemingly larger than the trends in AIRS and ERA5..

The monthly anomalies of the cRMSD as shown in Figure 4-9 appear to be very stable for both comparison datasets AIRS and ERA5 with decadal slopes close to zero.

Note that since no auto-correlation has been taken into account, uncertainty estimates, or variances, are possibly underestimated. As a consequence, the signal-to-noise ratio might be overestimated.

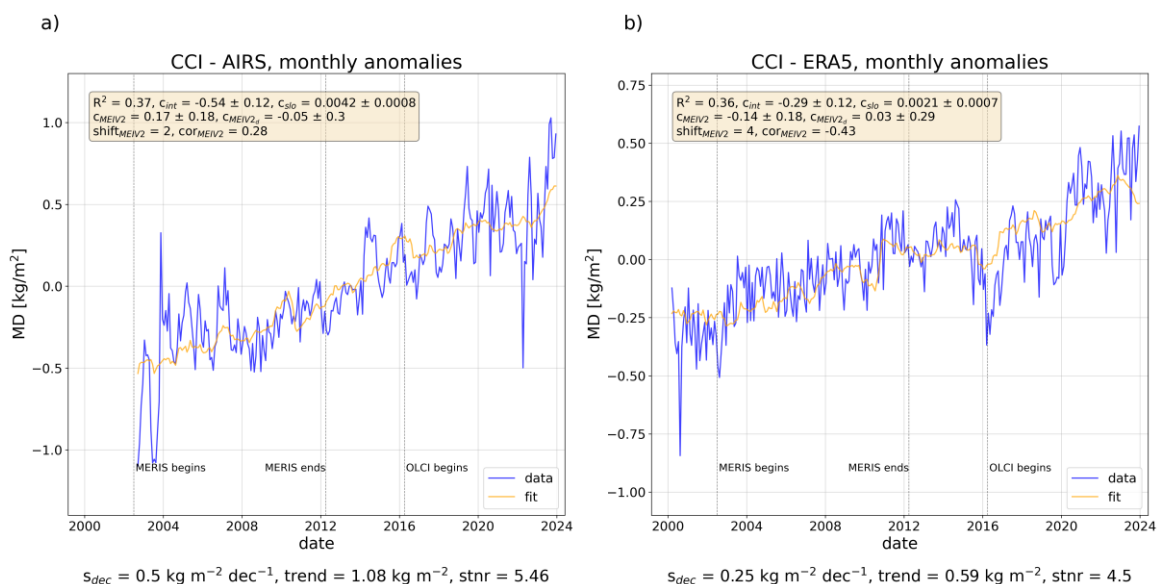


Figure 4-8: The blue curves in the panels show monthly anomalies of the mean differences between CCI and the references AIRS in a) and ERA5 in b) (X_{des} in Eq. (4-1)). The regression curves (X_{reg}) are displayed in orange. The fit parameters with their uncertainties as well as the decadal slope, the overall trend, and a signal-to-noise ratio are given in the box and below the panels. Consult the main text for an explanation of the regression method. Used CDR-1 datasets: $0.5^\circ \times 0.5^\circ \times$ monthly.

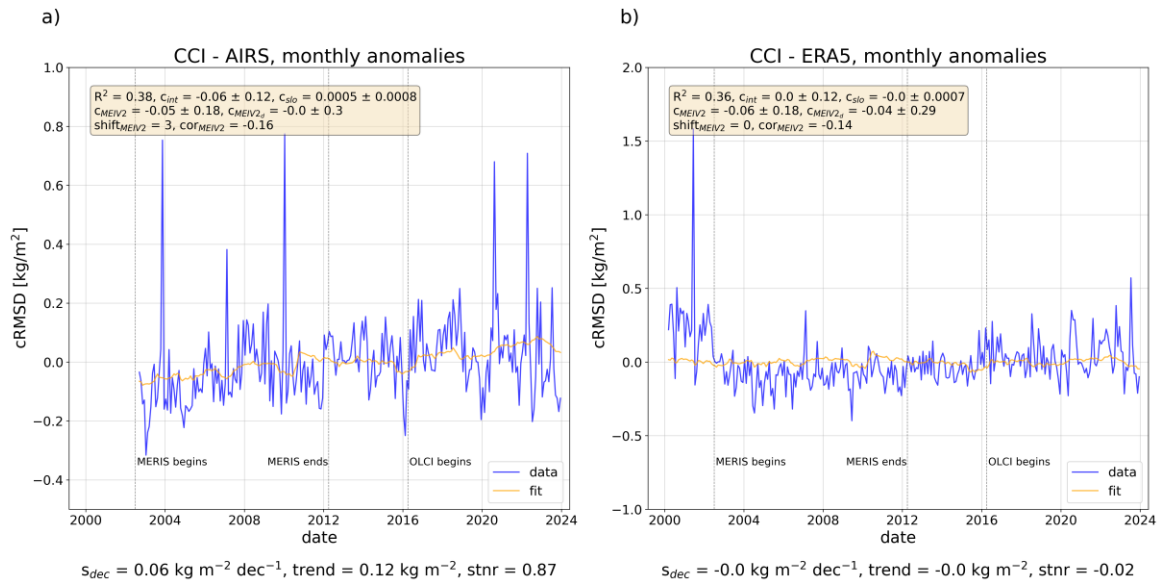


Figure 4-9: The blue curves in the panels show monthly anomalies of the cRMSD between CCI and the references AIRS in a) and ERA5 in b) (X_{des}). The regression curves are displayed in orange (X_{reg}). The fit parameters with their uncertainties as well as the decadal slope, the overall trend, and a signal-to-noise ratio are given in the box and below the panels. Consult the main text for an explanation of the regression method. *Used CDR-1 datasets: $0.5^\circ \times 0.5^\circ \times$ monthly.*

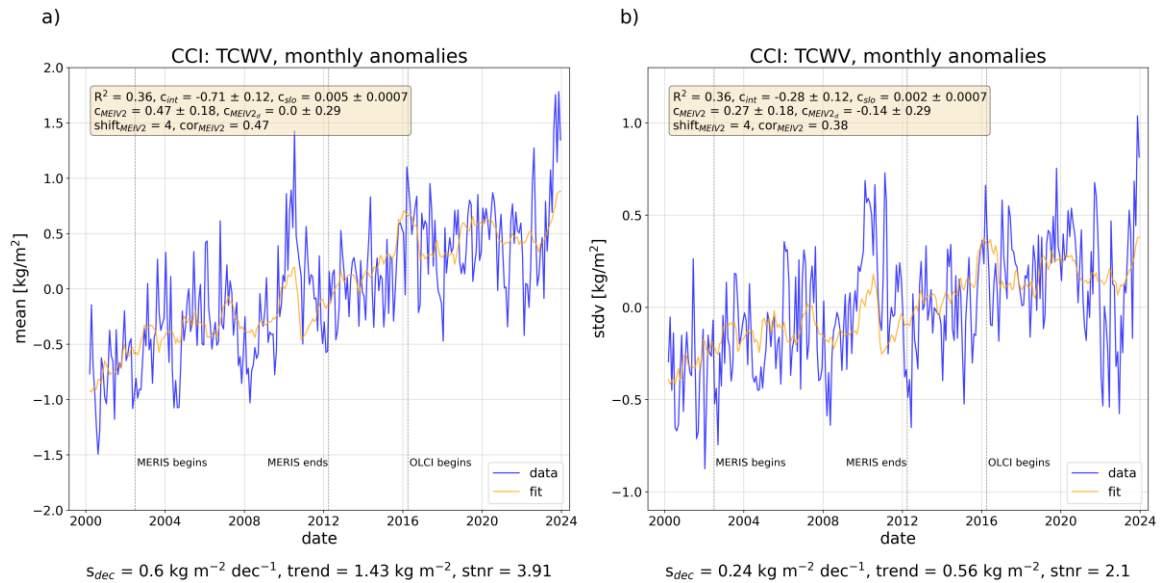


Figure 4-10: The blue curves in the panels show monthly anomalies of global mean TCWV in CDR-1 in a) and monthly anomalies of the corresponding standard deviation b) (X_{des}). The regression curves are displayed in orange (X_{reg}). The fit parameters with their uncertainties as well as the decadal slope, the overall trend, and a signal-to-noise ratio are given in the box and below the panels. Consult the main text for an explanation of the regression method. *Used CDR-1 datasets: $0.5^\circ \times 0.5^\circ \times$ monthly.*

Intercomparison results for daily data: To shed some light on the biases between CDR-1 and the references a study using daily data records with “cloud-free”, or “clear-sky”,

collocations is shown in Figure 4-11 with corresponding statistical measures in Table 4-5. Quite remarkable changes occur for the time series of MD: The biases, that is the mean MD, drop in absolute and relative value for all considered references when considering clear-sky conditions only. The reductions are quite large when compared to the results obtained in the previous analysis of monthly data records, see above. For ERA5 the bias diminishes from $1.6 \pm 0.3 \text{ kg/m}^2$ to $-0.2 \pm 0.3 \text{ kg/m}^2$, the latter being a relative mean bias of $1.5 \pm 1.8 \%$. This can be explained by the presence of a clear-sky bias (see also CAR v3.1 and references therein): the clear-sky bias is present if all-sky data (e.g. ERA5 monthly) are compared with clear-sky-only data (e.g. CDR-1), with the clear-sky data being smaller than the all-sky data. While the Level 2 are not affected by a clear-sky bias an increase in the level of aggregation leads to a stronger impact of the clear-sky bias. This is observed here.

Note also the obvious steps in MD between CDR-1 and ERA5 around 2015 and 2020 (see the blue curve in Figure 4-11 a)) which can also be detected in the monthly timeseries in Figure 4-8. Similar steps are not observed in the MD between CDR-1 and AIRS (see the orange curve Figure 4-8 and Figure 4-11 a)) which can be seen as a hint that the discussed steps are a feature of ERA5 and not of CDR-1. Some of the fluctuations in the quantities in Figure 4-11, in particular the cRMSD and the sample size, can be associated with the sensor combination contributing to CCI data on a given day

In the following section concerned with CDR-2 maps showing the spatial distribution of MD, cRMSD and sample size are shown (see Figure 4-14, Figure 4-15 and Figure 4-16). Since over land the maps resulting from CDR-2 are equal to maps obtained from CDR-1 data, no separate maps for CDR-1 are included here.

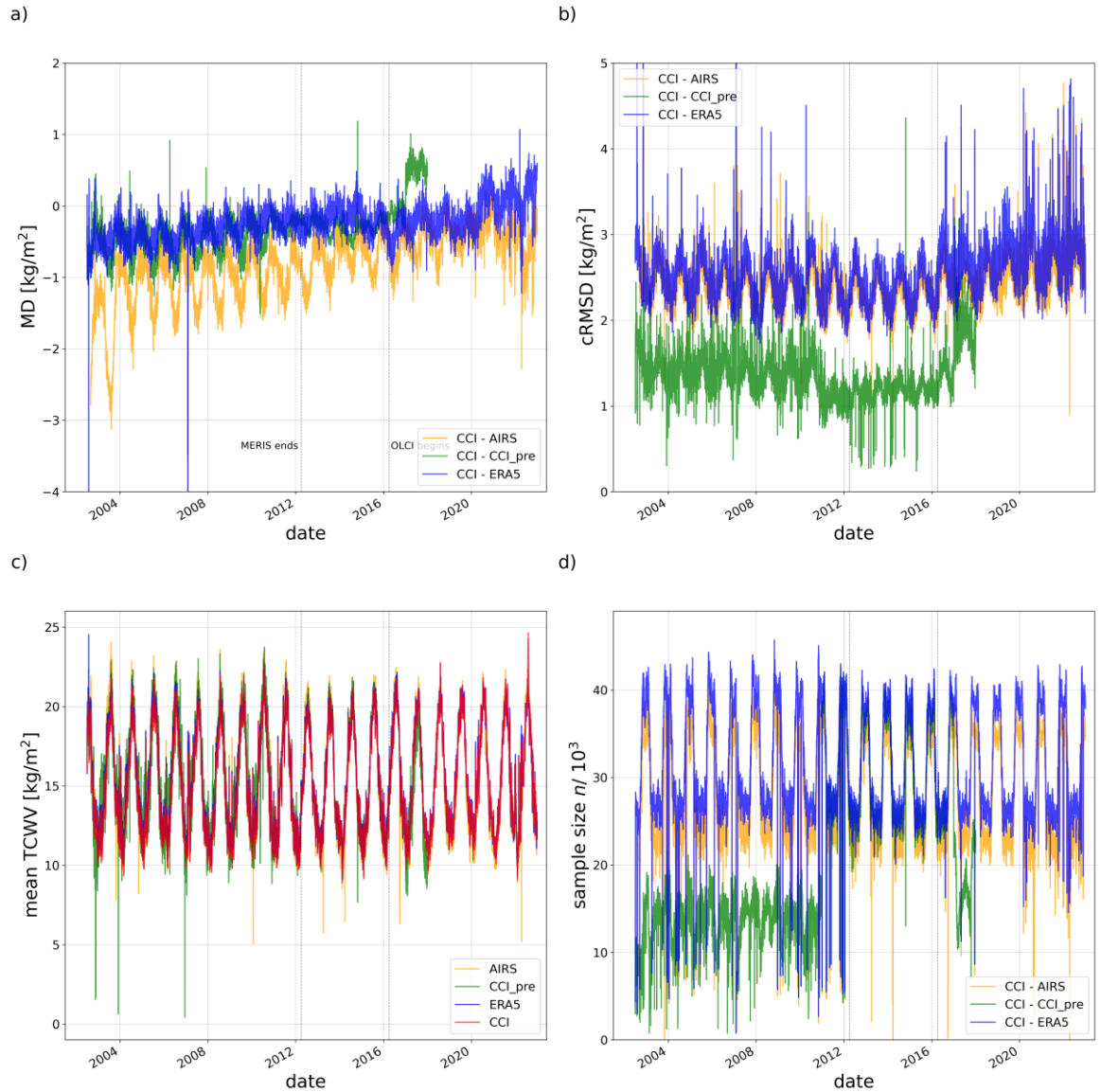


Figure 4-11: Timeseries comparing CDR-2 to various references on a daily basis are shown. Only “clear-sky” collocated data points over land enter, i.e. the data is filtered by applying CDR-2’s “atmospheric_conditions_flag” with value “1” which means no clouds are detected in the observation window. Furthermore, the “surface_type_flag” is set equal to “0” (“land”) in order to include only data points that are also present in CDR-1 (this is ignoring inland water bodies and permanent wetlands). . *Used CDR-2 datasets: 0.5° x 0.5° x daily*

Table 4-5: Basic statistical measures for the timeseries in Figure 4-11 are shown. *Used CDR-2 datasets: 0.5° x 0.5° x daily.*

Reference	Mean MD [kg/m ²]	Mean MD_rel (absolute) [%]	Mean cRMSD [kg/m ²]	Mean cRMSD_rel [%]	Mean TCWV [kg/m ²]
ERA5	-0.2 ± 0.3	1.5 ± 1.8	2.5 ± 0.4	16.5 ± 2.5	15.4 ± 3.1
AIRS	-0.7 ± 0.5	5.3 ± 3.1	2.4 ± 0.3	15.6 ± 2.1	15.3 ± 3.6
CCI_pre	-0.2 ± 0.3	2.1 ± 2.1	1.4 ± 0.3	8.8 ± 2.1	15.4 ± 3.3

4.2 Validation over all surfaces (CDR-2)

In this section validation and intercomparison activities are carried out for CDR-2. While the validation part only involves scatter plots for data at coasts, the main part is the intercomparison of CDR-2 with long-term reference datasets over all surfaces, i.e., globally. Since land surfaces have already been considered in section 4.1 they are not separately discussed here.

4.2.1 Intercomparison with long-term data records

In Figure 4-12, CDR-2 is compared to the long-term data records ERA5 und AIRS, as well as the precursor CDR-2 version CCI_pre produced in WV_cci Phase 1. Corresponding overall statistics are summarised in Table 4-6.

In panel a) a reasonable stability of CDR-2 is found when compared to ERA5 (blue curve) which is confirmed by the small decadal slope of 0.09 kg/m², which is a relative 0.3%/dec (see Table 4-6 and Figure 4-13). The trend analysis leading to Figure 4-13 is following the approach outlined in 4.1.2. Again, a step between 2016 and 2020 can be detected like in the intercomparison activities for CDR-1 in section 4.1.2. The absence of a trend in the CDR-2 to ERA5 comparison contrasts the increase found in the MD of CDR-1 and ERA5 (see Table 4-4 and Figure 4-8). The explanation can be found in the section 4.2.1.2 on intercomparisons over ocean surfaces: No trend is detected there, too. Since the majority of the Earth is covered by oceans the increase in TCWV differences between CDR-1 and ERA5 is suppressed due to less grid points over land than over ocean.

The increase in bias between CDR-2 and ERA5 from well below zero in the CDR-1 case to close to zero, $-1.6 \pm 0.3 \text{ kg/m}^2$ to $-0.1 \pm 0.1 \text{ kg/m}^2$ (see Table 4-4 and Table 4-6) can be explained by the positive bias over ocean surfaces (see Table 4-8 and Figure 4-14). A similar comment applies to the comparisons to AIRS and CCI_pre.

The AIRS comparison exhibits quite some instability which can be attributed to AIRS since the ERA5 and CCI_pre comparisons do not show the same behaviour. Most of the presumable AIRS instability is also occurring in the comparison of differences over ocean surfaces (see section 4.2.1.2, Figure 4-19).

The higher stability in the green MD curve, CDR-2 – CCI_pre, (Figure 4-12 a) when compared to the CDR-1 results (Figure 4-7 a) is due to a high consistency of the two CDR-2 versions over ocean surfaces, i.e. where relying on HOAPS (see also Figure 4-19).

The cRMSD (Figure 4-12 b) drops for all curves when comparing CDR-2 to the other datasets instead of CDR-1 (Figure 4-7 b; see also Table 4-4 and Table 4-6). For example, the cRMSD in the comparison to ERA5 drops from $2.4 \pm 0.2 \text{ kg/m}^2$ to $2.0 \pm 0.1 \text{ kg/m}^2$. This again is explained by consulting the results over ocean surfaces with generally lower cRMSD.

The stability of the cRMSD curves in panel b) of Figure 4-12 for the comparison with ERA5 and AIRS is generally good. The specific behaviour of the curves is dominated by those differences taken over land (see panel b) in Figure 4-7). The spikes in the yellow (CCI-AIRS) and green (CCI – CCI_pre) can be attributed to the comparison dataset since they appear at the specific time only in one of the comparison curves.

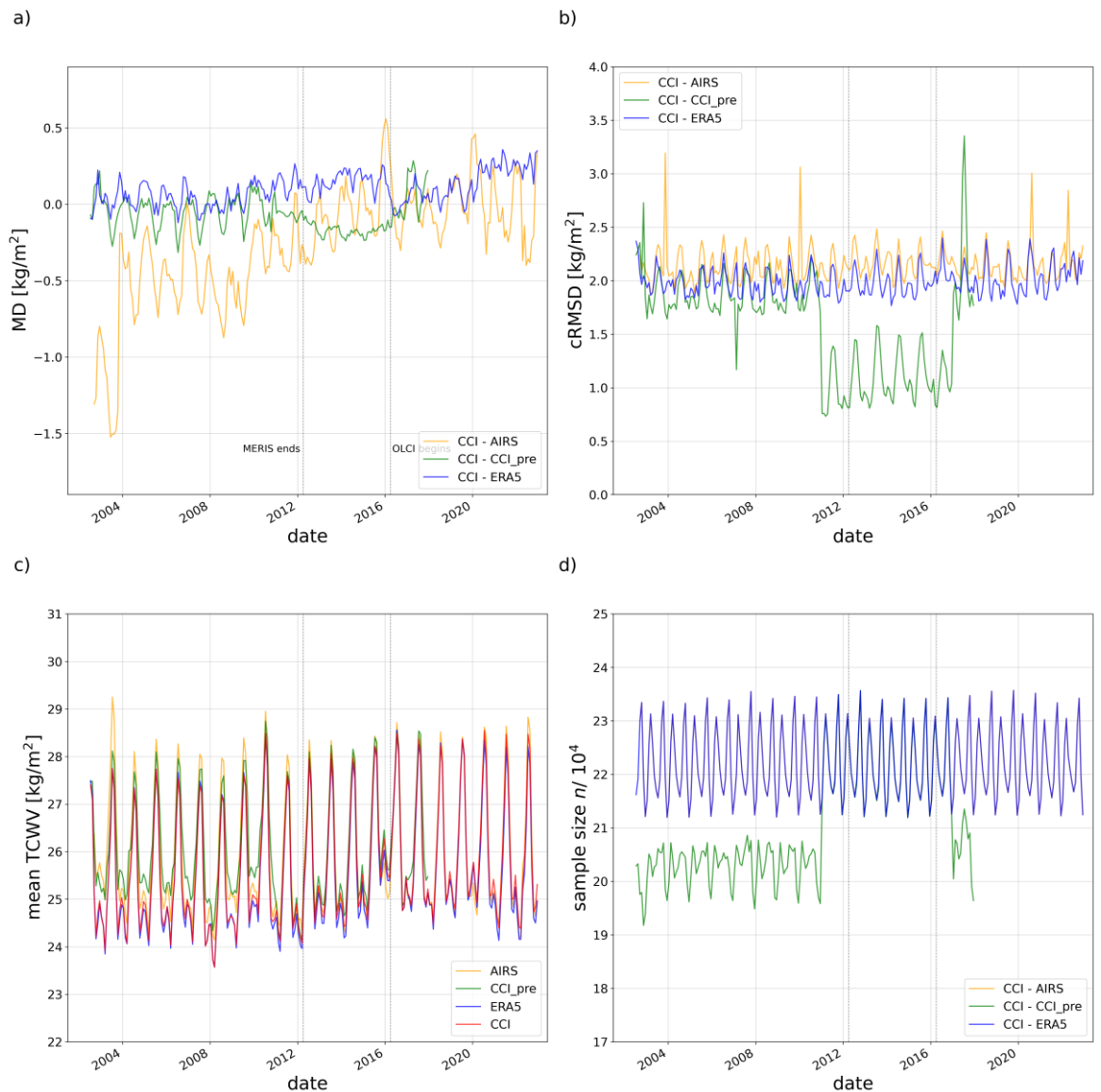


Figure 4-12: The time series' show a comparison of CDR-2 with various references and data records used for comparison. The differences of collocated data records are reduced in their spatial components to yield MD, or bias, (a), cRMSD (b), mean TCWV (c), sample size (d) over all surfaces and the entire globe. *Used CDR-2 datasets: 0.5° x 0.5° x monthly.*

Table 4-6: Basic statistical measures for the timeseries in Figure 4-12 and Figure 4-13 are shown, i.e. global, all surfaces. *Used CDR-2 datasets: 0.5° x 0.5° x daily.*

Reference	Mean MD [kg/m²]	Mean MD_rel (absolute) [%]	Slope (MD, fit) [kg/m²/dec]	Slope_rel [%/dec]	Mean cRMSD [kg/m²]	Mean cRMSD_rel [%]	Mean TCWV [kg/m²]
ERA5	-0.1 ± 0.1	0.4 ± 0.4	0.1	0.3	2.0 ± 0.1	7.8 ± 0.5	25.6 ± 1.3

AIRS	-0.3 ± 0.4	1.0 ± 1.4	0.5	1.9	2.2 ± 0.2	8.3 ± 0.7	25.9 ± 1.3
CCI_pre	-0.1 ± 0.1	0.3 ± 0.4			1.6 ± 0.5	6.1 ± 1.8	26.1 ± 1.1

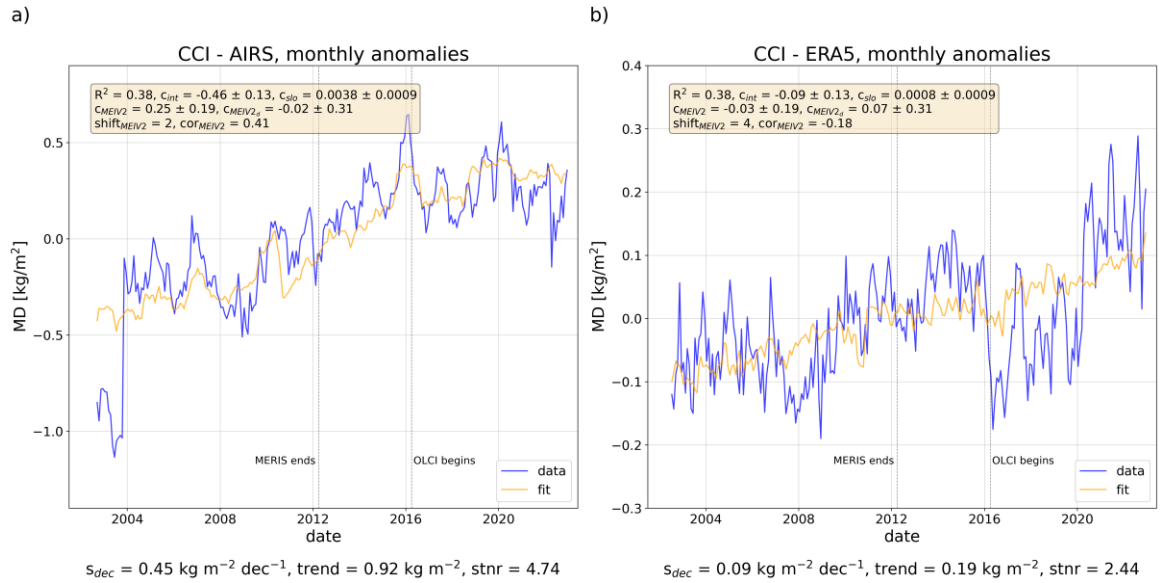


Figure 4-13: The blue curves in the panels show monthly anomalies of the mean differences between CCI CDR-2 (global, all surfaces) and the references AIRS in a) and ERA5 in b) (X_{des} in Eq. (4-1)). The regression curves are displayed in orange (X_{reg}). The fit parameters with their uncertainties as well as the decadal slope, the overall trend, and a signal-to-noise ratio are given in the box and below the panels. Consult the main text (section 4.1.2) for an explanation of the regression method. *Used CDR-2 datasets: $0.5^\circ \times 0.5^\circ \times$ monthly.*

The global maps in Figure 4-14, Figure 4-15 and Figure 4-16 illustrate the spatial distribution of MD, cRMSD and sample size along the temporal dimension for monthly data records.

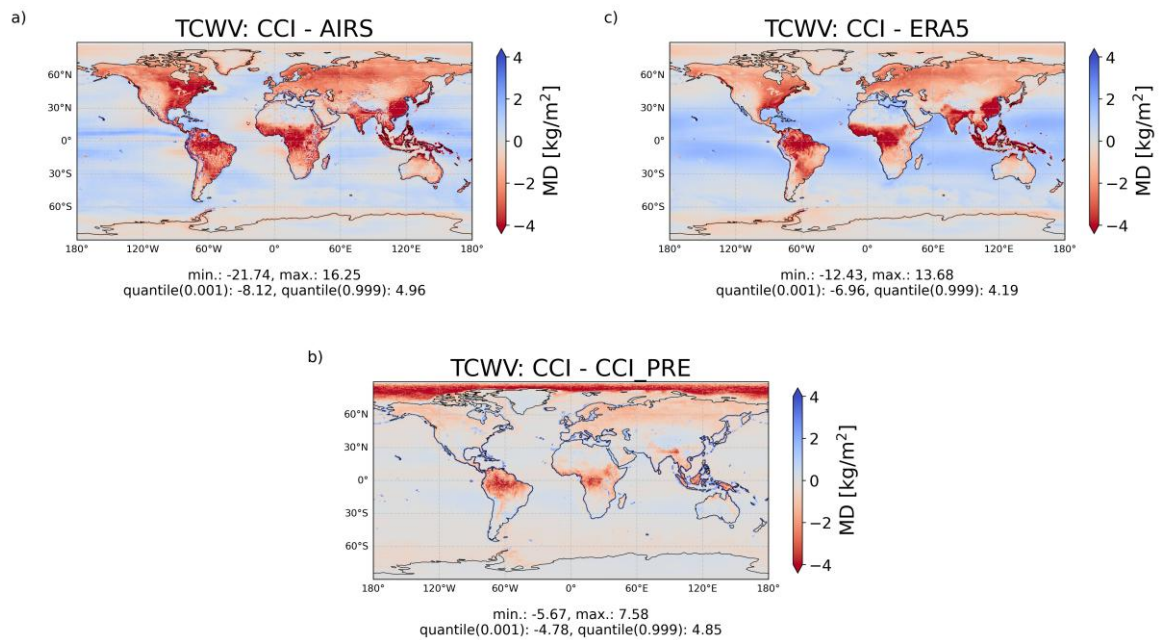


Figure 4-14: Global maps showing the mean differences of monthly means of CDR-2 to various data records as temporal average (no weights used in the calculations). *Used CDR-2 dataset: 0.5° x 0.5° x monthly.*

In Figure 4-14 the previously found characteristic of CDR-2 being overall drier over land than the other data records is apparent again (compare to the time series in Figure 4-7 and Table 4-4). The sign of mean differences varies for different locations. In a few regions, most prominently the Saharan Desert and the Tibetan Plateau, MD is positive. In most other places, a negative bias is present. Furthermore, larger features are similarly distributed for all data records, in particular when comparing the MD maps for AIRS and ERA5.

Over ocean on the other hand a moist bias is dominant (see Table 4-8) with smaller absolute values and a regional depending sign.

The high negative values for MD in northern high latitudes in panel c), CCI – CCI_pre, of Figure 4-14 can be explained by a revised NIR-retrieval over sea ice (see the PUG).

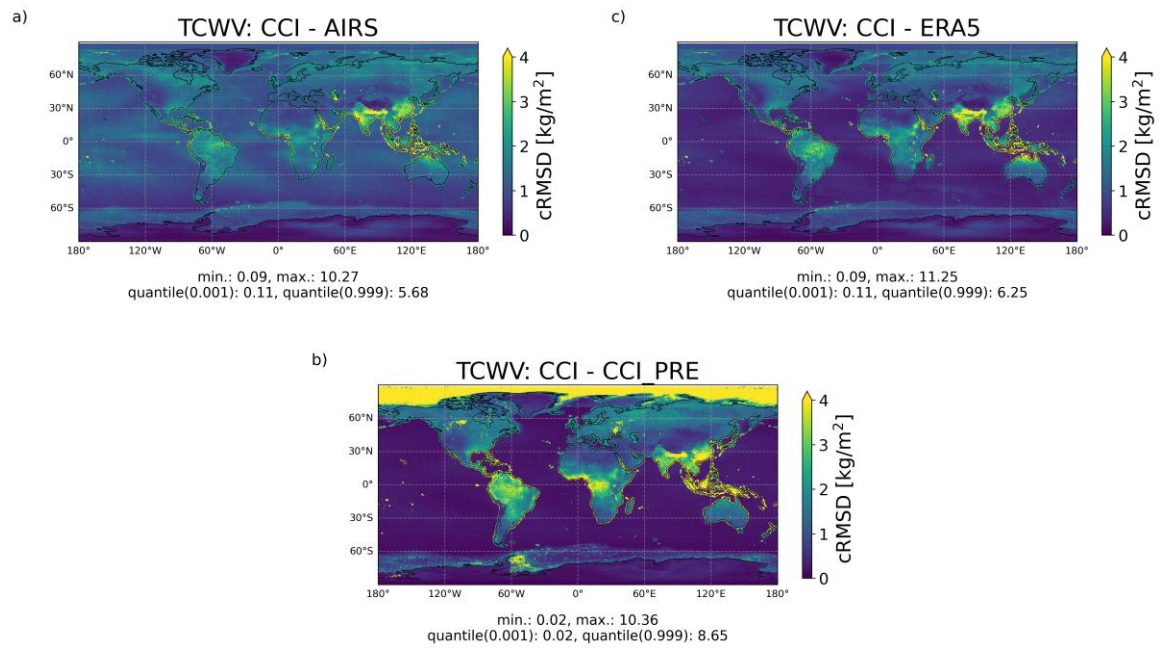


Figure 4-15: Global maps showing the cRMSD of monthly means of CDR-1 to various data records as temporal average (no weights used in the calculations). *Used CDR-1 dataset: $0.5^\circ \times 0.5^\circ \times$ monthly.*

The global maps in Figure 4-15 show that most regions with a higher cRMSD are found close to the equator where the average levels of TCWV are in general higher (see Figure 4-1). The highest values larger than 4 kg/m^2 appear near the Ganges river/Himalaya mountain range. Over ocean the cRMSD is generally smaller than over land. Note again that in northern latitudes in panel c) a higher discrepancy between CDR-2 and its precursor can be observed.

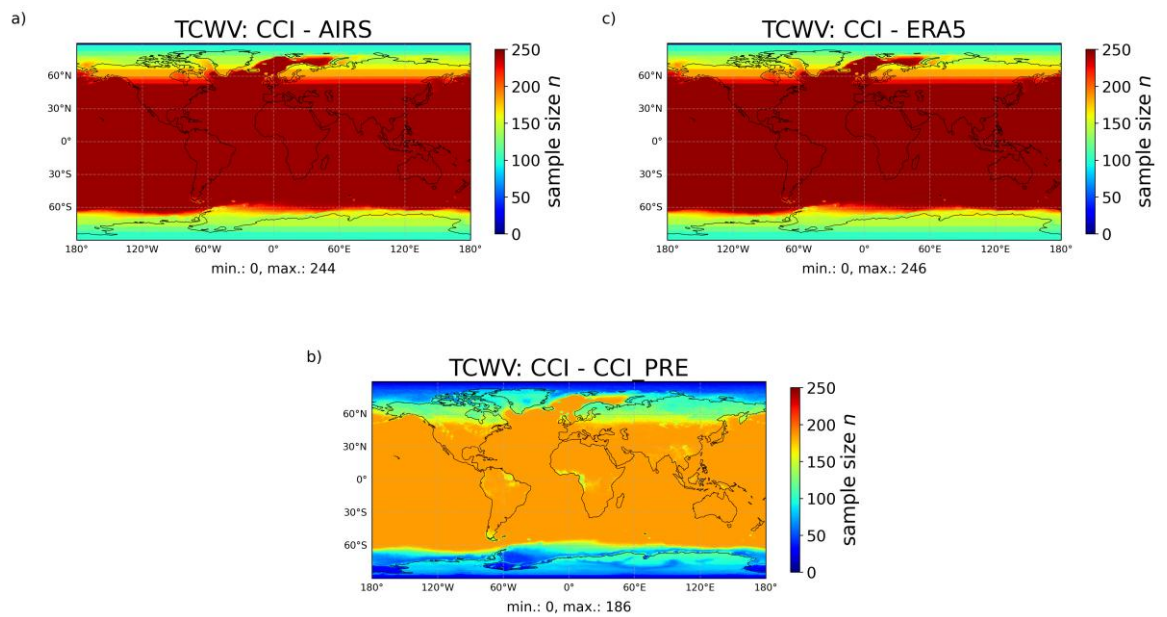


Figure 4-16: Global maps showing the sample sizes of various collocated data records. Here sample size corresponds to number of months with valid data in both data records in question. Used CDR-2 dataset: $0.5^\circ \times 0.5^\circ \times$ monthly.

As per **Figure 4-16**, the sample size is highest near the equator and decreases outside the subtropics. It is affected by the seasonal change of possible observation time of instruments measuring near infrared.

4.2.1.1 Results over coasts

In this section, the previous comparison with other long-term data records is repeated for coastal regions only. The time series in Figure 4-17 are less stable and more noisy (see the higher cRMSD in Table 4-7 compared to, e.g., Table 4-6) than those in the previous intercomparisons (see e.g. Figure 4-12), but this is an expected behaviour due to the increased difficulty of retrieving TCWV from NIR-observations over coasts and the lower sample size. In the comparison to AIRS and ERA5 small biases are found (see Table 4-7).

The algorithm improvements in the NIR-retrieval over coasts can be clearly seen in the behaviour of the green curve in panels a) and b) of Figure 4-17. Furthermore, changing sensor combinations influence the curves (see the orange and blue curves in panel a) and b) and the steps coinciding with the indicated constellation changes).

Judging from the comparisons to AIRS and ERA5 the CCI-TCWV values are more reliable than in previous dataset versions.

As a supplement, scatter plots comparing CDR-2 at coasts to station measurements of GRUAN and SuomiNet are displayed in Figure 4-18. An overall good agreement can be found between CCI and the references as reflected by the slopes being equal to 0.97 for the scatter plot comparing to GRUAN data in panel a), and equal to 0.94 when comparing to SuomiNet in panel b). In both cases the intercept is positive, suggesting, just like the positive biases to GRUAN and SuomiNet in Table 4-7, that over coast TCWV is overestimated in CCI-data. The cRMSD of 3.0 kg/m³ and 4.9 kg/m² for coastal regions is larger than over land surfaces (see Table 4-2 and the last row in Table 4-3 concerned with “actual over pass” means).

Reducing the red curve in Figure 4-17 c) showing the evolution mean CCI-TCWV yields an overall mean TCWV of 28.5 ± 2.4 kg/m² at coasts.

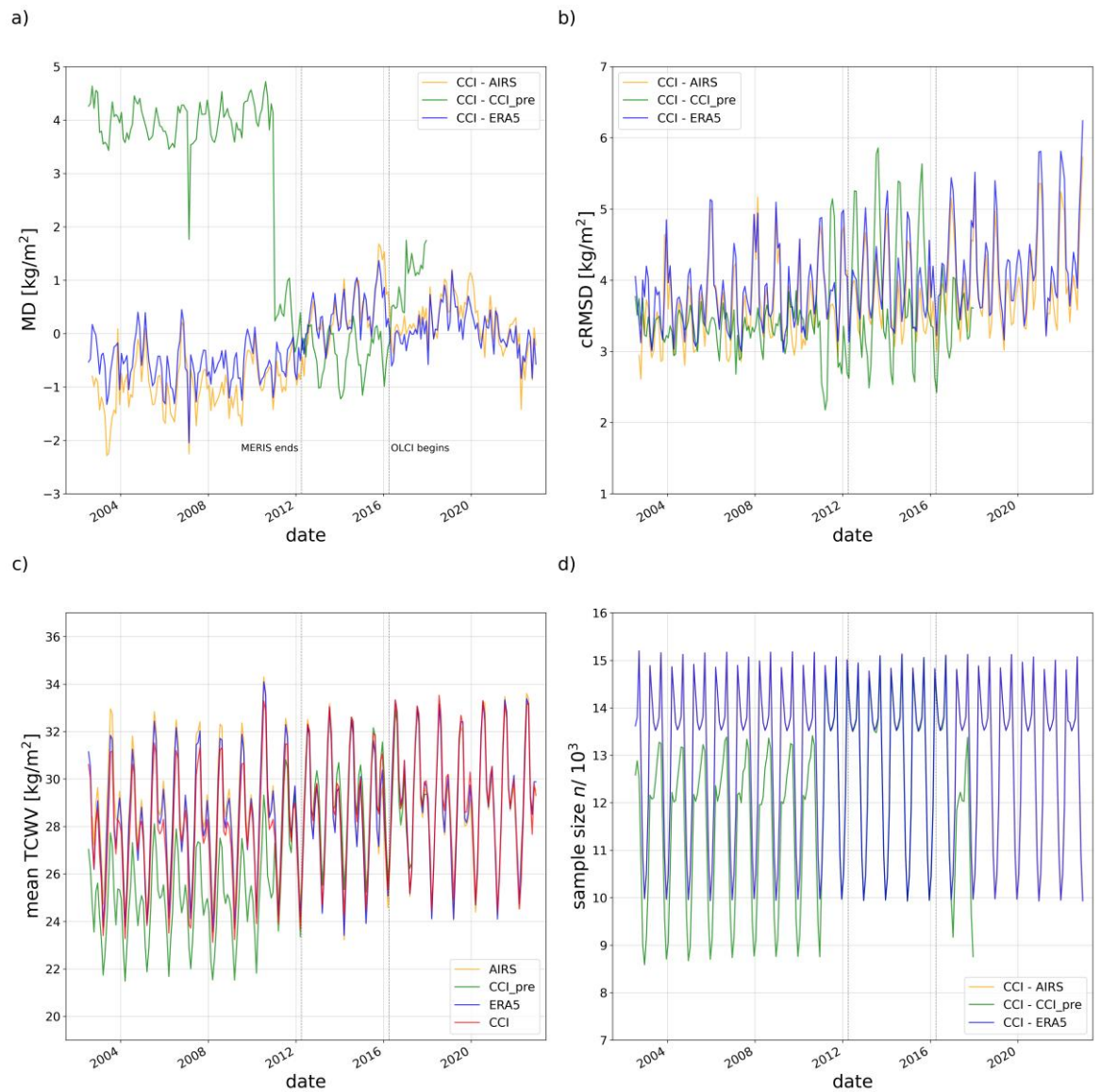


Figure 4-17: The time series' show a comparison of CDR-2 with various references and data records used for comparison over coasts, i.e. the "surface_type_flag" is equal to "5". The differences of collocated data records are reduced in their spatial components to yield MD, or bias, (a), cRMSD (b), mean TCWV (c), sample size (d). *Used CDR-2 datasets: $0.5^\circ \times 0.5^\circ \times$ monthly.*

- a) GRUAN, actual over passes, clear sky b) SuomiNet, actual over passes, clear sky

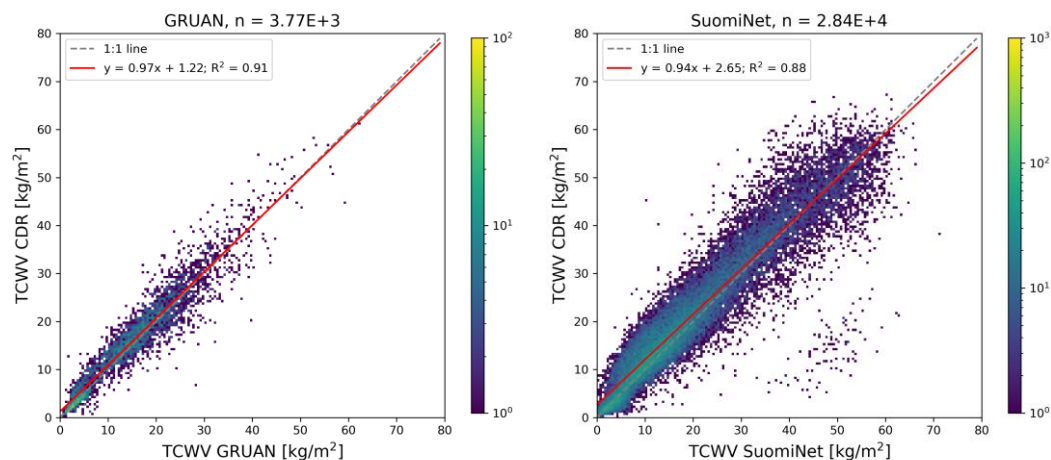


Figure 4-18: The figure shows scatter plots of collocated TCWV values at coasts from GRUAN and CDR-2 in panel a) and from SuomiNet and CDR-2 in panel b). The reference data has been processed like in section 4.1.1 and only “actual over pass means” enter. The “tcwv_quality_flag” is set to “0” (“best quality”), the “atmospheric_conditions_flag” is equal to “1” (“cloud-free”), and the “surface_type_flag” always has the value “5” for “coast”. Used CDR-2 dataset: $0.05^\circ \times 0.05^\circ \times \text{daily}$.

Table 4-7: Basic statistical measures for the timeseries in Figure 4-17 and Figure 4-18 are shown. Used CDR-2 datasets: $0.5^\circ \times 0.5^\circ \times \text{daily}$ and $0.05^\circ \times 0.05^\circ \times \text{daily}$.

Reference	Mean MD [kg/m ²]	Mean MD_rel (absolute) [%]	Mean cRMSD [kg/m ²]	Mean cRMSD_rel [%]	Mean TCWV [kg/m ²]
ERA5	-0.2 ± 0.6	0.6 ± 2.0	4.0 ± 0.6	13.9 ± 2.1	28.6 ± 2.4
AIRS	-0.3 ± 0.8	1.1 ± 2.7	3.8 ± 0.6	13.1 ± 2.1	28.7 ± 2.4
CCI_pre	-2.2 ± 2.0	8.3 ± 7.6	3.5 ± 0.6	13.1 ± 2.4	26.7 ± 2.8
GRUAN	0.8	5.1	3.0	19.4	15.5
SuomiNet	1.5	7.4	4.9	24.1	20.3

4.2.1.2 Results over ocean

The time series in Figure 4-19 compare CDR-2 to AIRS, ERA5 and CCI_pre over ocean surfaces. Since CDR-2 is a merged product obtained from NIR observations over land and MW observations over ocean, the data of CDR-2 analysed in this section is obtained by different methods than in the previous discussion, namely only from MW observations as per CM SAF's HOAPS activities. HOAPS is an all-day and all-sky (with the exception of heavy rain) product which is different from the NIR-product available only in the presence of direct sunlight. Reducing the red curve in Figure 4-19 c) showing the evolution mean CCI-TCWV yields an overall mean TCWV of $29.0 \pm 0.5 \text{ kg/m}^2$ over ocean.

The trend analysis leading to Figure 4-20 and Figure 4-21 is following the approach outlined in 4.1.2.

The MD curves in Figure 4-19 a) show at most times positive mean differences between CDR-2 and AIRS and ERA5 over ocean. While the AIRS-curve (orange) is affected by high fluctuations that likely originate from the AIRS dataset itself (see discussions above), the curve showing the comparison against ERA5 (blue) displays a relatively high stability with a bias of $0.9 \pm 0.1 \text{ kg/m}^2$ (see Table 4-8). Note the steps in the anomaly curves in Figure 4-20 b) around 2008 and between 2016 and 2020. The green curve comparing CDR-2 with its earlier version is effectively comparing different HOAPS versions.

No trend is detected in the ERA5 bias curves (see Figure 4-20 and Table 4-8) in contrast to the assessment for CDR-1 in Figure 4-8 where visual inspection and a decadal slope of $0.25 \text{ kg/m}^2/\text{dec}$ suggest a small trend.

By visual inspection, there is a slight downwards drift in the green curve (CCI-CCI_pre, Figure 4-19 a) that can be explained by a new intercalibration of SSMIS L1 data feeding into the HOAPS version used for the current CDR-2.

The cRMSD curves in Figure 4-19 b) show a high stability with occasional spikes in the comparison to AIRS. Since these outliers are only present in the AIRS comparison curve they are attributed to AIRS and not to CDR-2, see also the above discussion of such features in the AIRS dataset. Comparing Table 4-8 with Table 4-5, Table 4-6 and Table 4-7 one finds that the mean cRMSD is lowest when restricting to data over ocean.

Finally, consulting Figure 4-21 a), where a regression curve for monthly anomalies of TCWV over ocean in CDR-2 is shown, a high correlation coefficient of 0.84 between ocean CCI_TCWV and a three-month shifted ENSO index MEIV2 can be found, as also noted in,

e.g., Mieruch et al. (2014). It seems that the ENSO signal accounts for most of the larger fluctuations in deseasonalised monthly TCWV in CDR-2 over ocean. While for the corresponding time series of CDR-2 data over land in Figure 4-10 a) a decadal slope of 0.6 kg/m²/dec can be found, a smaller linear trend with decadal slope of 0.37 kg/m²/dec is present in CDR-2 data over ocean as shown in Figure 4-21. Note that in both cases the slope coefficient of the fit is a multiple of its uncertainty (see the boxes in Figure 4-10 a) and Figure 4-21 a)). This is also the case for the overall trend and the RMSs of residuals as reflected by their quotients, the signal-to-noise ratio, of 3.91 and 3.58, respectively.

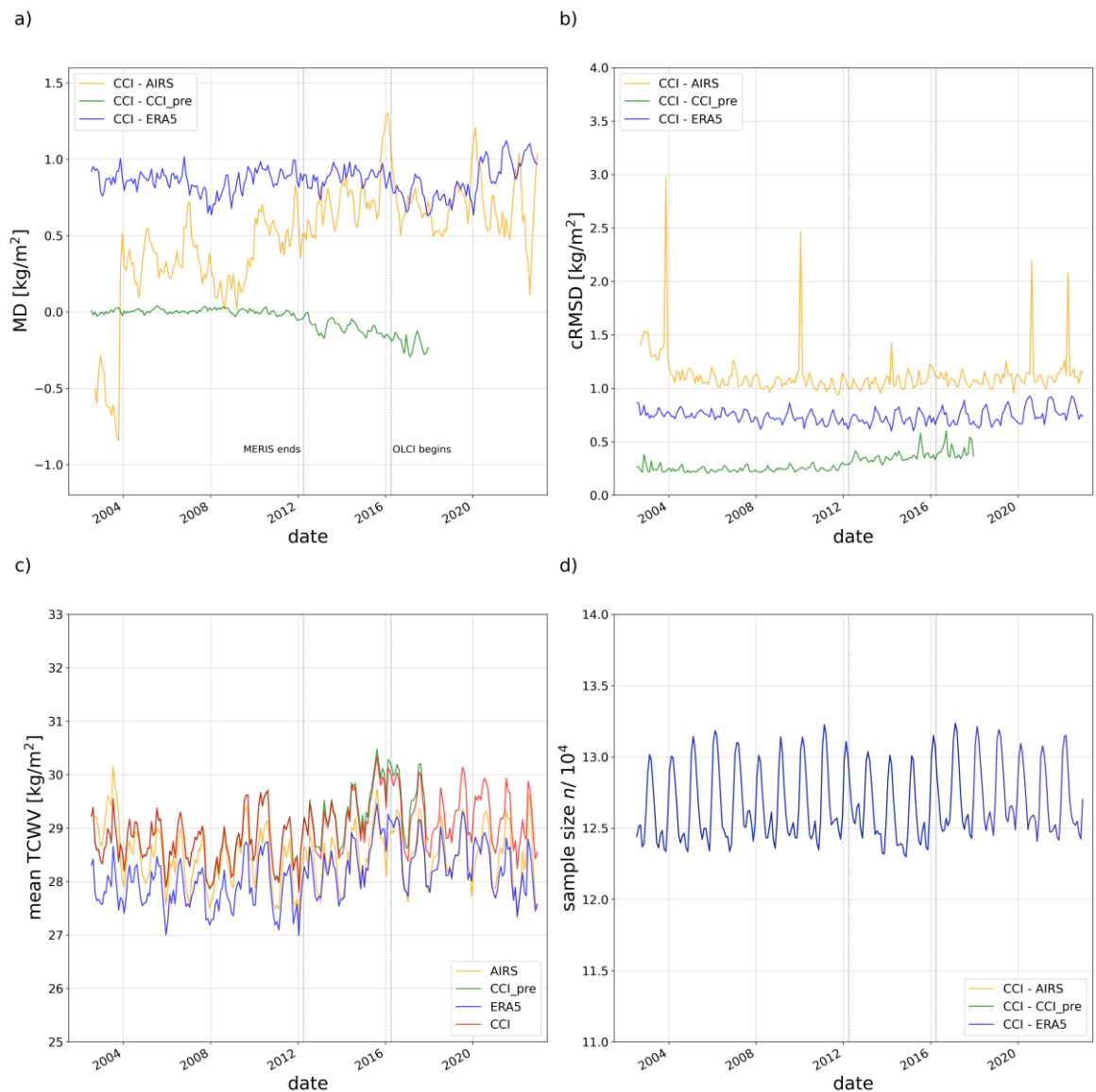


Figure 4-19: The time series' show a comparison of CDR-2 with various references and data records used for comparison over ocean, i.e. the "surface_type_flag" is equal to "1". The differences of collocated data records are reduced in their spatial components to yield MD, or bias, (a), cRMSD (b), mean TCWV (c), sample size (d). Used CDR-2 datasets: 0.5°

$\times 0.5^\circ \times \text{monthly}$. Note that the vertical lines indicating changes in the constellation are not relevant over ocean.

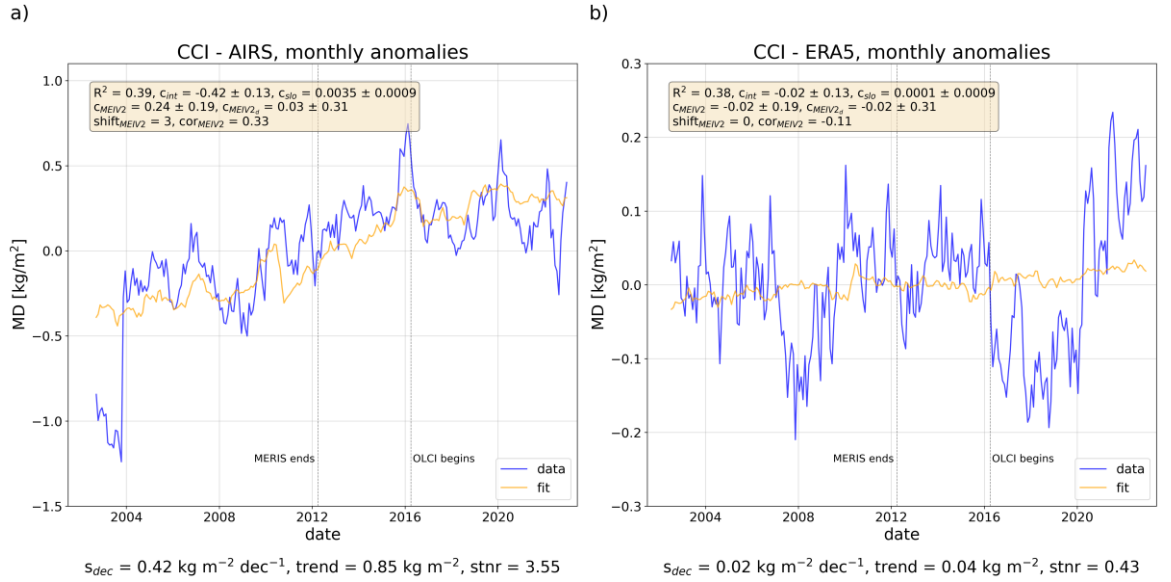


Figure 4-20: The blue curves in the panels show monthly anomalies of the mean differences between CCI and the references AIRS in a) and ERA5 in b) over ocean (X_{des} in Eq. (4-1)). The regression curves are displayed in orange (X_{reg}). The fit parameters with their uncertainties as well as the decadal slope, the overall trend, and a signal-to-noise ratio are given in the brownish box or below the panels. Consult the main text for an explanation of the regression method. Used CDR-2 datasets: $0.5^\circ \times 0.5^\circ \times \text{monthly}$. Note that the vertical lines indicating changes in the constellation are not relevant over ocean.

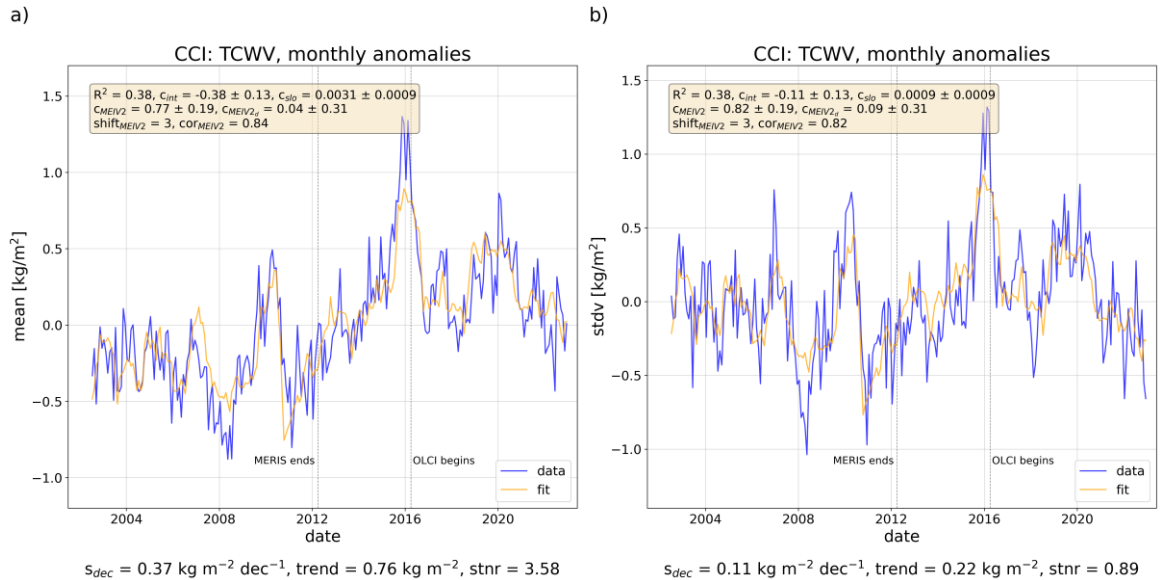


Figure 4-21: The blue curves in the panels show monthly anomalies of global mean TCWV in CDR-2 in a) and monthly anomalies of the corresponding standard deviation b) over ocean (X_{des} in Eq. (4-1)). The regression curves are displayed in orange (X_{reg}). The fit

parameters with their uncertainties as well as the decadal slope, the overall trend, and a signal-to-noise ratio are given in the brownish box or below the panels. Consult the main text for an explanation of the regression method. *Used CDR-1 datasets: $0.5^\circ \times 0.5^\circ \times$ monthly.* Note that the vertical lines indicating changes in the constellation are not relevant over ocean.

Table 4-8: Basic statistical measures for the timeseries in Figure 4-19 and Figure 4-20 are shown. *Used CDR-2 datasets: $0.5^\circ \times 0.5^\circ \times$ monthly.*

Reference	Mean bias [kg/m ²]	Mean bias_rel (absolute) [%]	Slope (bias, regression) [kg/m ² /dec]	Slope_rel [%/dec]	Mean cRMSD [kg/m ²]	Mean cRMSD_rel [%]	Mean TCWV [kg/m ²]
ERA5	0.9 ± 0.1	0.3 ± 0.3	0.2	0.7	0.7 ± 0.2	2.6 ± 0.1	28.1 ± 0.5
AIRS	0.5 ± 0.4	1.7 ± 1.3	0.4	1.4	1.1 ± 0.7	3.9 ± 0.7	28.5 ± 0.5
CCI_pre	-0.1 ± 0.1	0.2 ± 0.3			0.3 ± 0.3	1.0 ± 0.3	27.0 ± 0.6

4.3 Validation of the regional, high resolution product

For the three previously selected regions the high-resolution product with 0.01° regular grid and based on NIR-measurements from OLCI alone, over land and ocean (in contrast du CDR-1 and - 2, where the NIR-based retrieval is restricted to land or land and coasts, respectively), is compared to the standard CDR-2 product via a global bias map and to the references AIRS and ERA5 by computing the bias and cRMSD. The latter quantites are given separately for each region.

The high-resolution data was regridded to a resolution of 0.05° and monthly means are compared to the standard CDR-2 product on a regular 0.05° grid. For the comparison with ERA5 and AIRS the monthly high-resolution product's grid was coarsed to a resolution of 0.5° . Latitude depending weights are used in the regridding procedure for high-resolution CCI-TCWV.

Because the high-resolution data spans only from June to September, no time series are shown.

The three regions are (see also Figure 4-22):

Table 4-9: Specifications of the high-resolution product.

Region	Latitude [°]	Longitude [°]	Time span
ROI-1	20N - 40N	60E-80E	Jun - Sept 2017 - 2022

ROI-2	10S - 20N	15W - 30E	Jun - Sept 2017 - 2022
ROI-3	25S - 0S	70W - 40W	Jun - Sept 2017 - 2022

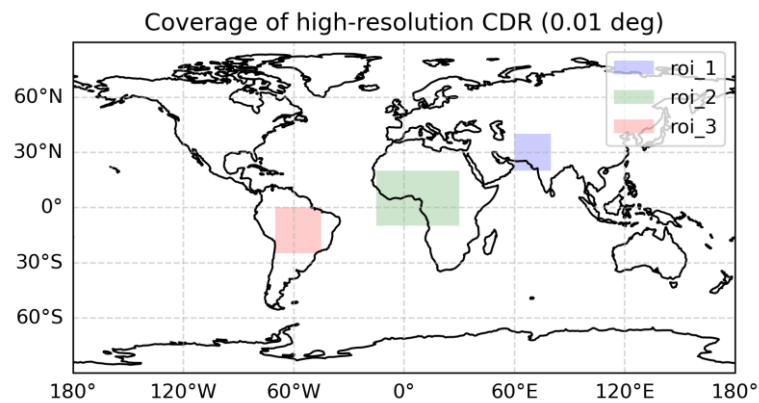


Figure 4-22: Regions where the high-resolution product is available.

In Figure 4-23, Figure 4-24 and Figure 4-25 the high-resolution product is compared to the standard CDR-1 product via regional maps showing MD obtained along the temporal dimension. Geographical features are clearly visible. These likely originate from differences between the differently resolved products on a submonthly scale. When each of the products involved is aggregated over time to monthly means, these differences become systematic and visible in the comparison of monthly TCWV data. The high-resolution product likely resolves the topography in a more accurate way. Since TCWV retrievals are sensitive to the topographic altitude (see e.g. Roman et al. 2016), differences appear when comparing finer and coarser spatial resolution, due to considering a different number of locations at different heights in the averaging process.

The coarser structures over ocean consisting of larger grid cells, with finer variations within each grid cells, correspond to the coarser resolution of 0.5° of CDR-2 over ocean, i.e. HOAPS. Since, over ocean, differences between nearby grid cells in the (regrided) high-resolution products are smaller than the differences between nearby ocean grid cells in the standard CDR-2 product a coarser structure consisting of the boundaries of the larger 0.5° cells is visible.

In ROI 1, Figure 4-23, negative MD dominates with largest absolute and relative values over the Himalayan mountain ranges and the Indus rivers, as well as over ocean.

In ROI 2, Figure 4-24, the picture is more diverse. There still is a predominantly negative MD but over the Congo river and tributaries, it is positive which is a sharp contrast to the behaviour of MD in region 1. Again, the relative values are largest in regions with heterogeneous topography.

ROI 3, Figure 4-25, shows the most balanced maps with parts of negative MD, mostly in the Andes mountain ranges and the contained sections of the Rio de la Plata tributaries, and parts of larger positive MD, mostly over the Amazon river and tributaries.

Summarizing Table 4-10 and Table 4-11, the MD is negative in all ROIs. ROI 1 exhibits the largest mean differences and cRMSD when compared to AIRS and ERA5 on a 0.5° grid while ROI 3 is the region which is overall closest to the comparison datasets when taking MD and cRMSD as measures. Note that the last column on mean TCWV shows that over all three regions the water vapour content in the atmosphere is relatively high (compare with, e.g., the last columns in Table 4-4 and Table 4-7, or mean values of CCI_TCWV at the end of the introduction of section 4).

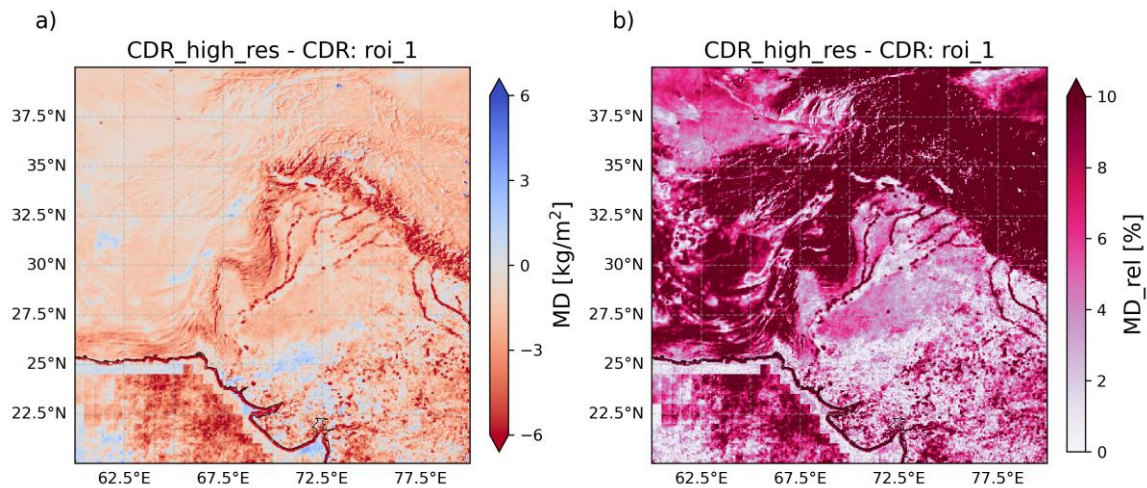


Figure 4-23: The bias of the regional higher resolution data to the standard CDR-2 product is shown for roi_1 as well as the relative bias with respect to standard CDR-2. *Used CDR datasets: $0.01^\circ \times 0.01^\circ \times$ monthly regridded to $0.05^\circ \times 0.05^\circ \times$ monthly, $0.05^\circ \times 0.05^\circ \times$ monthly (CDR-2).*

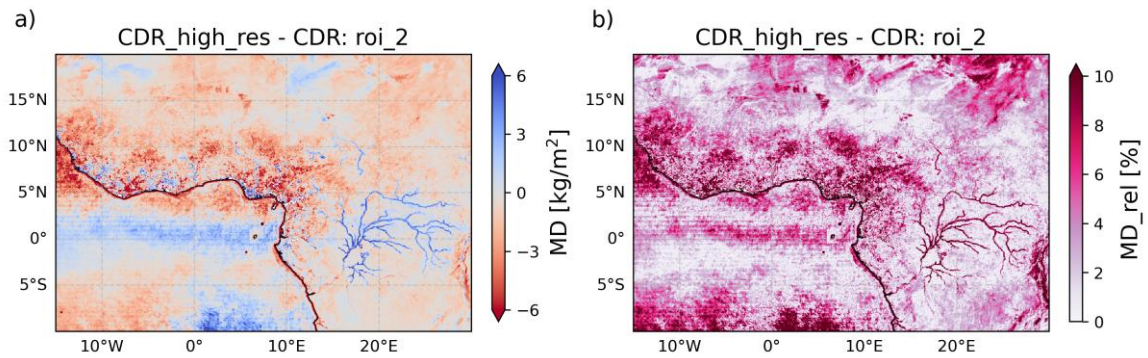


Figure 4-24: The bias of the regional higher resolution data to the standard CDR-2 product is shown for roi_2 as well as the relative bias with respect to standard CDR-2. *Used CDR datasets: $0.01^\circ \times 0.01^\circ \times$ monthly regridded to $0.05^\circ \times 0.05^\circ \times$ monthly, $0.05^\circ \times 0.05^\circ \times$ monthly (CDR-2).*

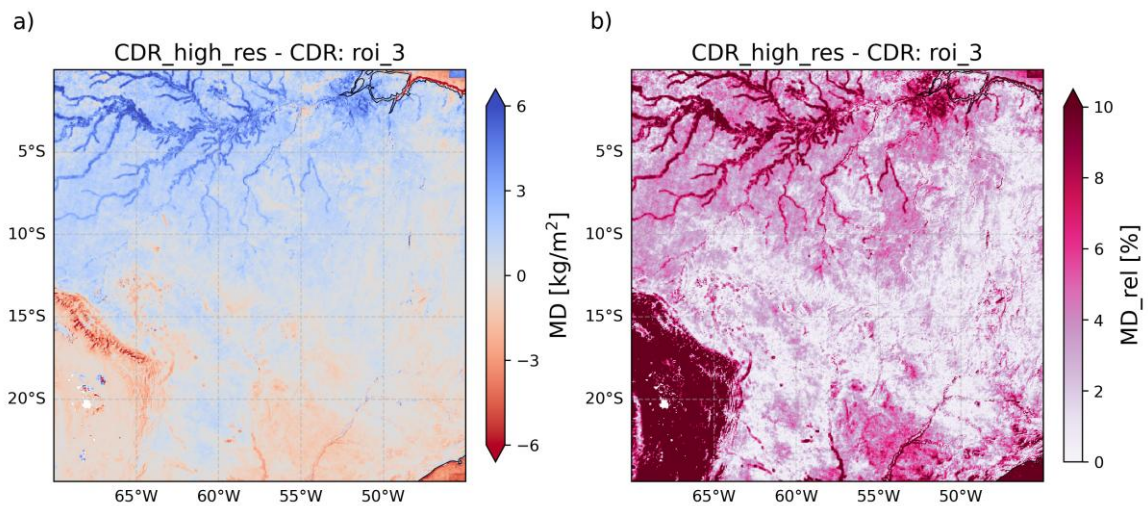


Figure 4-25: The bias of the regional higher resolution data to the standard CDR-2 product is shown for roi_3 as well as the relative bias with respect to standard CDR-2. *Used CDR datasets: $0.01^\circ \times 0.01^\circ \times$ monthly regridded to $0.05^\circ \times 0.05^\circ \times$ monthly, $0.05^\circ \times 0.05^\circ \times$ monthly (CDR-2).*

Table 4-10: This table shows the MD and cRMSD of high-resolution CDR-2 to the reference AIRS for all three regions of interest separately. *Used CDR datasets: $0.01^\circ \times 0.01^\circ \times$ monthly regridded to $0.5^\circ \times 0.5^\circ \times$ monthly.*

CCI - AIRS	MD [kg/m ²]	MD_rel (absolute) [%]	cRMSD [kg/m ²]	cRMSD_rel [%]	TCWV [kg/m ²]
Roi_1	-4.0	12.5	5.1	16.2	31.7
Roi_2	-2.6	6.5	4.2	10.7	39.6
Roi_3	-1.2	3.8	3.0	9.5	31.6

Table 4-11: This table shows MD and cRMSD of high-resolution CDR-2 to the reference ERA5 for all three regions of interest separately. *Used CDR datasets: 0.01° x 0.01° x monthly regridded to 0.5° x 0.5° x monthly.*

CCI - ERA5	MD [kg/m ²]	MD_rel (absolute) [%]	cRMSD [kg/m ²]	cRMSD_rel [%]	TCWV [kg/m ²]
Roi_1	-4.0	12.7	5.3	13.7	31.8
Roi_2	-2.2	5.7	4.4	11.2	39.3
Roi_3	-0.4	1.4	3.0	9.6	30.8

5. UNCERTAINTY ASSESSMENT

5.1 General approach

In this section the consistency analysis of CDR-1 and the reference dataset GRUAN is presented. As GRUAN climatology the “actual over pass” means (see section 4.1.1) are used for both TCWV and measurement errors. The consistency assessment consists of pointwise comparison of the TCWV differences between CDR-1 and GRUAN and the combined uncertainty. The latter will also be called “total expected uncertainty” in the following and is given by

$u_{total_i} = \sqrt{u_{x_i}^2 + u_{y_i}^2}.$	5-1
---	-----

For $k = 1, 2, 3$ a collocated data point satisfies the “consistency inequality” (see Immler et al. (2010)) if

$ x_i - y_i < k \sqrt{u_{x_i}^2 + u_{y_i}^2}.$	5-2
---	-----

For $k = 1, 2, 3$ the expected percentage of collocated data points fulfilling inequality 5-2 is at least 68%, 95%, 98%, respectively. Otherwise, this is an indication that the total uncertainty is underestimated or overestimated. Here, x_i represents CDR-1 and y_i a reference data record, with i being an index to collocated data points.

For GRUAN the collocation criteria of section 4.1.1 are applied. The data is filtered by demanding CDR-1’s “atmospheric_conditions_flag” to be equal to “1” and “tcwv_quality_flag” to be equal to “0” (“clear-sky” and “best-quality”).

Refining the approaches from WV_cci phase 1, each satellite observation is as an independent observation. Whether a satellite contributed to given grid cell’s TCWV value is determined by the associated “number of observations” variable being non-zero. The number of contributing satellites is then taken as number of independent observations N . The degree of correlation of uncertainties in CDR-1 is given by the parameter c . If the quantities N (number of independent observations, see above) and c are inserted in the approaches defined by Stengel et al. (2017), for CDR-1, the squared uncertainty at a valid collocated data point x_i is “stdv” squared, scaled by N , plus squared mean measurement error “tcwv_err” times correlation factor c :

$$u_{x_i}^2 = \frac{\text{"stdv"}_{x_i}^2}{N} + c \text{"tcwv_err"}_{x_i}^2.$$

5-3

In the following “c_CCI” is used for the correlation coefficient in Eq. 5-3 when referring to uncertainties of CCI-TCWV. For the reference GRUAN, formula Eq. 5-3 is used, too. In this case, N is the number of measurements feeding the mean obtained as daily comparison TCWV value, here the “actual over pass mean”, and c is set to 0, 0.1, and 0.2. “stdv” in Eq. 5-3 is now the standard deviation of the reference measurements used on a given day. The error term “tcwv_err” in Eq. 5-3 is the RMS of uncertainties as given for each considered measurement in the global attributes of the corresponding GRUAN files (see Table 4-1). This approach to assess the expected uncertainties is more complex than the approaches in Phase I but remain restricted.

Not considered in the consistency inequality is the representativeness, σ . Because mainly gridded data records are validated/compared the representativeness error is assumed to be small, i.e. $\sigma=0$ kg/m². When using station-based data instead of gridded data records this assumption may lead to an underestimation of the total uncertainty and to an underestimation of the expectation (but see the study on different temporal means per day in the validation using SuomiNet, Figure 4-6 and Table 4-3).

Finally, it needs to be mentioned, that outliers were not filtered and that the PDF of the differences is not perfectly Gaussian. This will affect the consistency assessment as well. Depending on the skewness it can lead to over- and underestimations of the expectation. E.g., long tails of positive and negative differences will lead to a seeming underestimation of total uncertainty.

The results of the consistency analysis are shown in Table 5-1.

Table 5-1: Consistency assessment of CDR-1 using GRUAN. *Used CDR-1 dataset: 0.05° x 0.05° x daily.*

Reference GRUAN	Consistency [%]			Sample size
	k=1	k=2	k=3	
c_CCI = 0	55.1	82.0	92.1	5083
c_CCI = 0.1	57.1	82.9	92.5	5083

c_CCI = 0.2	57.9	83.3	92.9	5083
-------------	------	------	------	------

These results show that uncertainties in CDR-1 are underestimated when assuming the availability of valid uncertainty estimates for GRUAN and a negligible representativeness uncertainty. Increasing the correlation coefficient c_{CCI} yields slightly better agreements of the determined consistencies with their required, quantile-based values.

5.2 Refined approach using stratified data

In this section a refined uncertainty assessment of CDR-1 uncertainties against GRUAN is carried out using a method described by Sayer et al. (2020). This method is based on the above assessment but instead of applying inequality 5-2 on all valid data points the data is aggregated in bins along the total expected uncertainties which was defined in Equation 5-1. The same data as in Section 5.1 was used to obtain following results.

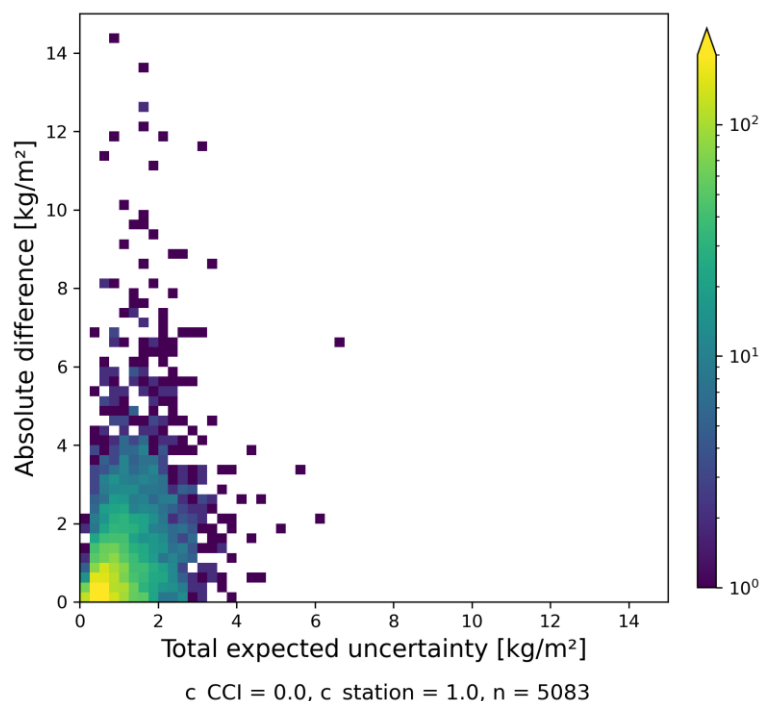


Figure 5-1: 2D-histogram of total expected uncertainties and absolute differences of CDR-1 and GRUAN. Used CDR-1 dataset: $0.05^\circ \times 0.05^\circ \times$ daily.

The uncertainty assessment is the study of a joint distribution of two quantities, namely the total expected uncertainties (TEU) and the absolute differences (AD), see Figure 5-1. The

general approach in Section 5.1 can suffer from the fact that this joint distribution is not a bivariate normal distribution. The approach by Sayer et al tries to account for this possible defect by binning, or stratifying, the joint distribution of TEU and AD along TEU. In this way the objective is to find that the population of differences contained in each bin is approximately normally distributed with a variance estimated, in the present case, by the median of TEU values in the bin under consideration. In other words, the idea is that the conditional probability density function of the differences is given by a Gaussian distribution with conditional variance determined by TEU.

The procedure is the following: First sort the bivariate frequency distribution of TEU and AD as shown in Figure 5-1 into ten equally populated bins along the TEU-axis. Then find in each of the bins the 38th, 68th and 95th quantile of the binned AD populations as well as the medians of the TEU values, and display the results in a plot like in Figure 5-2. The expectation is that the quantiles in the above order correspond to 0.5σ , 1σ and 2σ where σ is the median of the binned TEU values.

In this way one obtains Figure 5-2 from the distribution shown in Figure 5-1. The differently coloured points connected by dashed lines indicate how the respective AD quantiles depend on the TEU. The straight coloured lines have slopes equal to 0.5, 1, and 2, hence representing the respective ideal multiple-of- σ relationships, and the TEU bin bounds are indicated by the vertical dashed grey lines. If in each bin the differences were indeed normally distributed with variance equal to the median of the total expected uncertainties within the bin, then each point should lie on the accordingly coloured straight line.

In Figure 5-2 all AD quantiles follow their ideal courses but lie in all cases above their expected values. This indicates that the total expected uncertainties underestimate the actual uncertainties. Probably due to small sample sizes, the 95th quantile is rather far off the expected line and fluctuates while the 38th and 68th quantiles stick closer to their sought-after behaviour.

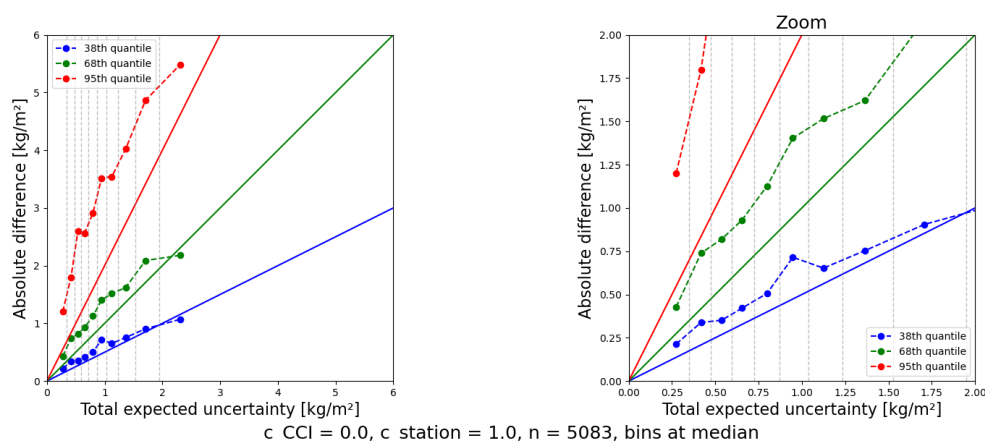


Figure 5-2: Uncertainty assessment of CDR-1 and GRUAN obtained from Figure 5-1 following Sayer et al. (2020). The applied procedure and meaning of each curve are explained in the main text. *Used CDR-1 dataset: $0.05^\circ \times 0.05^\circ \times$ daily.*

At this point correlations CCI-uncertainties have, more or less, been neglected by setting c_CCI equal to zero. By increasing the coefficient c_CCI (see the subtitles of the figures in this section) one allows for some correlation in the NIR-measurement errors that constitute the CDR-1 dataset. Arithmetically, according to a formula in Stengel et al. (2017) (see the previous section 5.1), this means that the uncertainty u of CDR-1 also includes the measurement errors as given in the variable “tcwv_err” which contains the mean measurement errors within a grid cell. The result is shown in Figure 5-3 for $c_CCI = 0.1$, and in Figure 5-4 for $c_CCI = 0.2$.

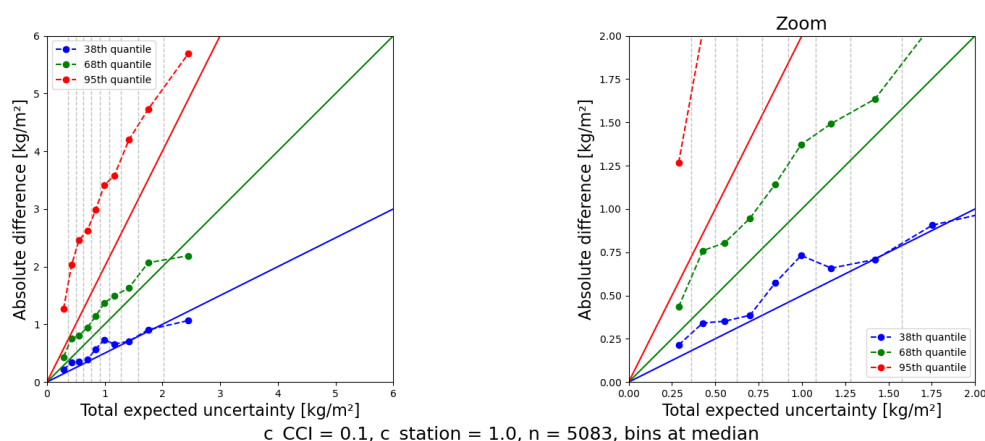


Figure 5-3: Uncertainty assessment of CDR-1 and GRUAN with CDR-1 error correlation coefficient equal to 0.1. *Used CDR-1 dataset: $0.05^\circ \times 0.05^\circ \times$ daily.*

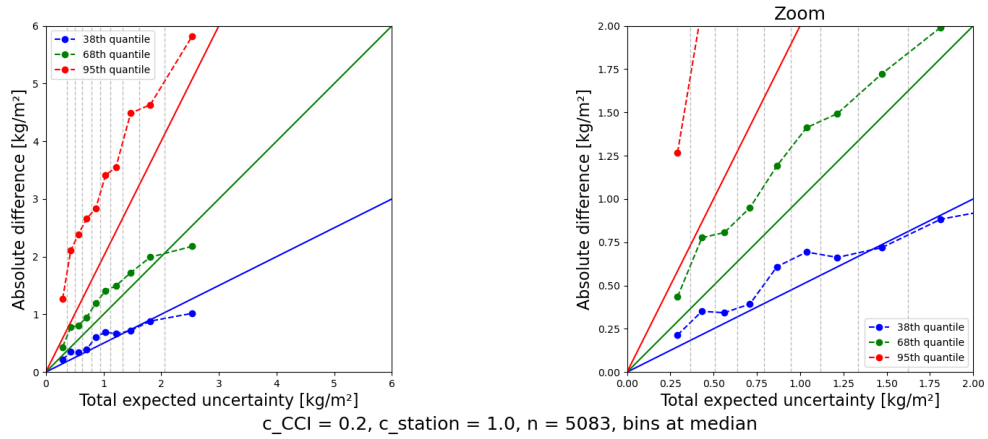


Figure 5-4: Uncertainty assessment of CDR-1 and GRUAN with CDR-1 error correlation coefficient equal to 0.2. Used CDR-1 dataset: $0.05^\circ \times 0.05^\circ \times \text{daily}$.

Unsurprisingly, when increasing the TEUs alone the quantile curves are stretched towards higher TEU values. Compared to Figure 5-2 more of the AD quantiles in Figure 5-3 and Figure 5-4 lie now below their respective ideal straight lines, and with this the 38th and 68th quantiles overall seem to be closer to their expected values.

6. CLEAR-SKY BIAS ESTIMATION USING SUOMINET

In this section the so-called clear-sky bias in the CDR-1 dataset is estimated using the GNSS-based measurements reported in the SuomiNet dataset. Since the NIR-based retrieval yielding the TCWV values in CDR-1 does not apply in cloudy situations, a possibly non-zero sampling bias is present after aggregation in the data when compared to all-sky measurements. This means that certain situations with likely higher humidity than usual, as indicated by the presence of clouds, do not feed into CDR-1 resulting in a dry bias.

To estimate this clear-sky bias (CSB), data obtained from GNSS measurements which work in all-sky situations is resampled on a given day in two steps: First, only those SuomiNet values are considered that lie within an hour window in the local time of a possible overpass of one of the platforms contributing to CDR-1. Since only Level-3 data is used, for which actual overpasses cannot be collocated, an overpass at the platform's equator crossing time is assumed at a certain location if the platform contributes to CDR-1 on the given day. The "all over pass times" mean (see section 4.1.1) is computed using all collocated SuomiNet measurements yielding the daily values feeding the full climatology used in the following analysis. In a second step, this climatology of daily values is resampled using CDR-1's "atmospheric_conditions_flag". The largest such subsample is the "all-sky climatology" where the "atmospheric_condition_flag" assumes any value in the set {1, 2, 3}, i.e. {"clear_sky", "partly_clouded", "clouded"}. This means all daily SuomiNet values are considered whenever any of the platforms passed a given station, regardless of a valid TCWV data point. The "all-sky climatology" is further refined by demanding the "atmospheric_condition_flag" to be equal to "1" or "2". In this way the "partly-clouded climatology" is defined. Finally, the "clear-sky" climatology is constructed from those data points where the "atmospheric_conditions_flag" is equal to "1" indicating that no clouds have been detected by any instrument.

The clear-sky bias is then defined as the difference of the respective mean climatologies:

$$bias_{clear_sky} = mean(TCWV_{clear_sky}) - mean(TCWV_{all_sky})$$

In order to have an idea of the clear-sky bias at coasts, too, the actually used CCI-dataset is CDR-2 restricted to land, i.e. CDR-1, and coasts.

Table 6-1: The global, latitude-weighted means of each climatology obtained in the course to estimate CDR-1's clear sky bias as well as the CSB itself are shown. The unit is kg/m²

in all columns. At coasts, the sample size is small. *Used CDR-1 dataset: $0.05^\circ \times 0.05^\circ \times$ daily.*

SuomiNet	Land			Coast
Climatology	all	$ \text{lat.} < 30^\circ$	$ \text{lat.} > 30^\circ$	all
Complete	18.9	29.8	14.0	26.9
All-sky	19.2	29.8	14.4	27.2
Part. clouded	17.2	26.1	13.2	25.4
Clear-sky	15.9	23.6	12.6	23.7
CSB	-3.3	-6.2	-1.8	-3.5

Table 6-1 shows that in each step of the reduction of the SuomiNet data towards an clear-sky climatology reduces the mean global TCWV value suggesting that those situations are discarded that are originally above average in water vapour content. In particular the CSB is negative with an overall value of -3.3 kg/m^2 , implying a dry bias of clear-sky conditions w.r.t. all-sky conditions. Furthermore, its magnitude depends on the latitudinal zone with a larger absolute CSB in low latitudes.

The drop of 1.3 kg/m^2 from 17.2 kg/m^2 in the global partly clouded climatology to a clear-sky value of 15.9 kg/m^2 suggests that discarding partly clouded situations excludes above-average TCWV values. In the results obtained from the intercomparison of CDR-1 with ERA5 in section 4.1.2 the bias is raised by 1.3 kg/m^2 , when moving from monthly data in Figure 4-7, which in the end corresponds to data obtained from partly clouded and cloud-free observations, to data obtained from cloud-free observations alone, as shown in Figure 4-11. Both findings together confirm the impression that CCI-TCWV observations in partly clouded situations underestimate the true TCWV by missing observations over regions with higher TCWV but covered by clouds.

At coasts the clear-sky bias is larger than the one over all land surfaces with generally higher mean TCWV values for the respective climatologies. Note that while some stations at the coast contribute to the all-sky climatology they do not enter in the clear-sky climatology since no measurements are recorded under clear-sky conditions at the affected stations.

Besides global means the spatial distribution has been estimated as well and is shown in Figure 6-1. Most stations show a negative clear-sky bias, and this bias is strongest in the lower latitudes, confirming the impression gained in the assessment of zonal means.

Finally, taken from an assessment with an interim version of CCI-data and obtained by a slightly different method, Figure 6-2 shows how the distribution of clear-sky estimates using SuomiNet data might extend away from the stations by applying the same procedure as used for SuomiNet to ERA5 reanalysis data. As shown in the supplementary boxplot in Figure 6-3 the CSB estimates from SuomiNet and ERA5 agree quite well at collocated locations with a rather narrow distribution and a median difference of -0.17 kg/m^2 and a mean difference of -0.07 kg/m^2 .

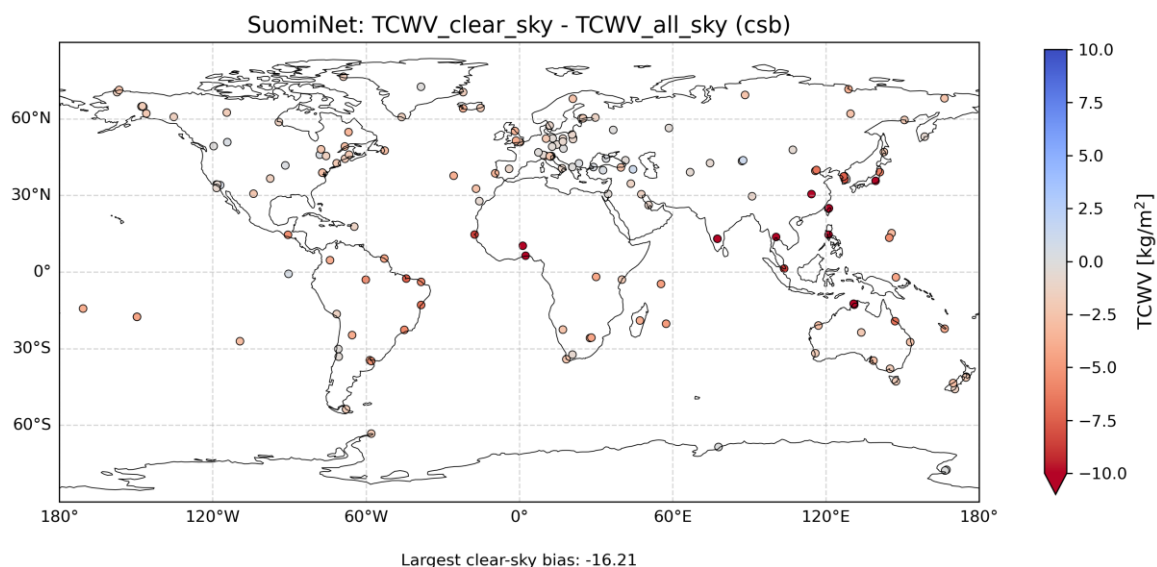


Figure 6-1: Distribution of the clear-sky bias estimate using SuomiNet (GNSS). Cloud mask was constructed from CDR-1. *Used CDR-1 dataset: $0.05^\circ \times 0.05^\circ \times$ daily.*

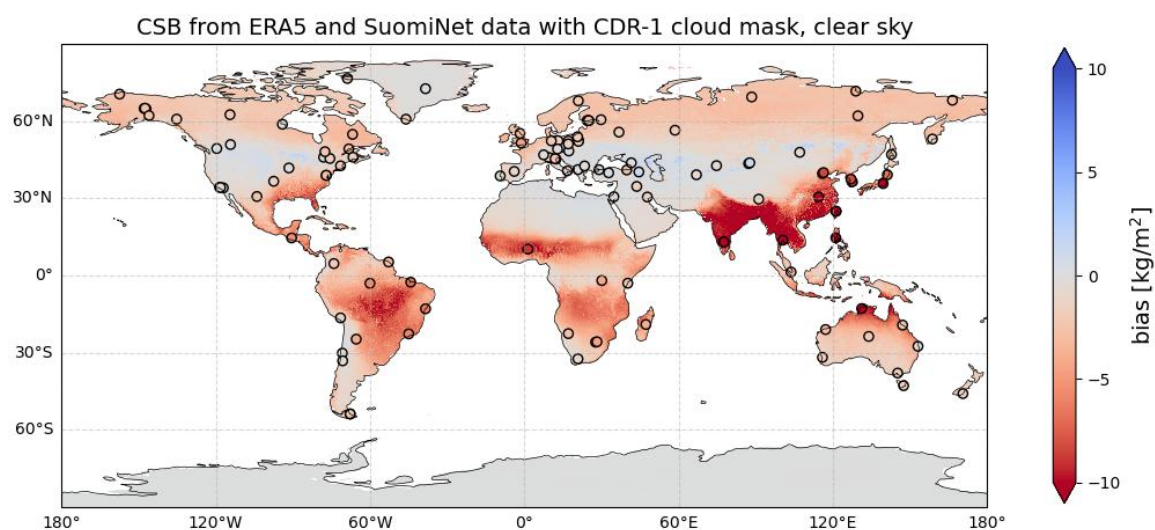


Figure 6-2: Distribution of the clear-sky bias estimate using ECMWF's ERA5 reanalysis (coloured map) and SuomiNet (coloured circles). In both cases, the cloud mask was constructed from CDR-1 in the same way. *Used CDR-1 dataset: $0.05^\circ \times 0.05^\circ \times$ daily.*

Boxplot: CSB_SuomiNet - CSB_ERA5

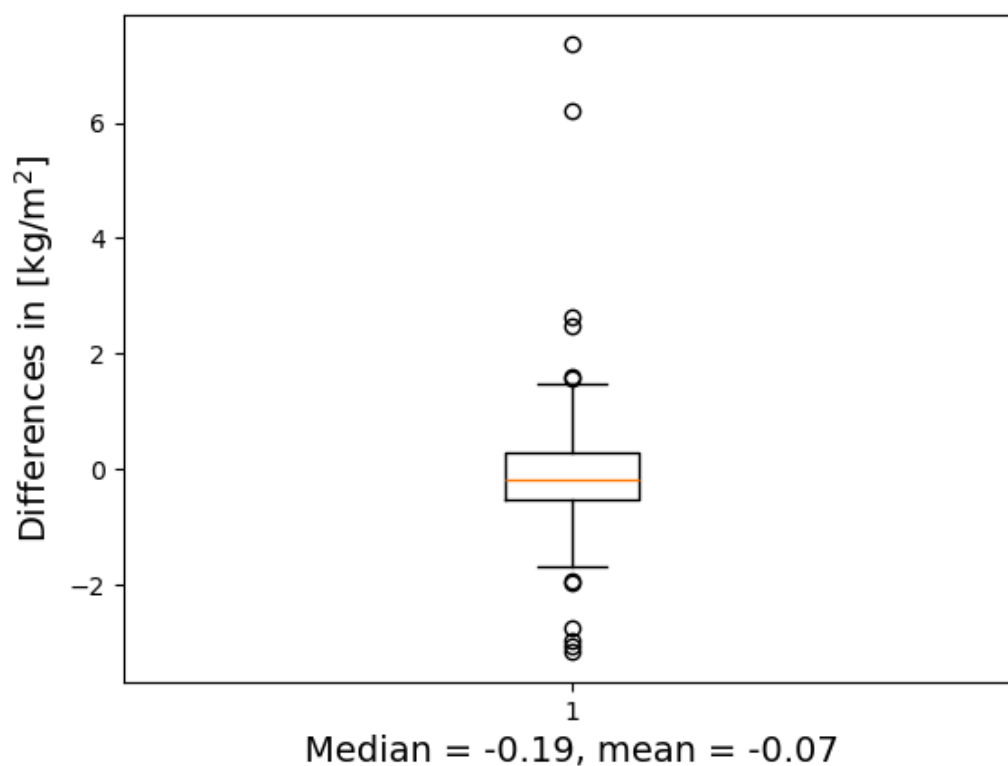


Figure 6-3: Boxplot showing the distribution of differences in the clear-sky estimates from SuomiNet and ERA5 using the same data as in Figure 6-2.

7. OVERALL STATISTICS AND COMPLIANCE WITH REQUIREMENTS

In this section overall statistics are given and their compliance with requirements is discussed.

Reducing the CDR-1 product, $0.5^\circ \times 0.5^\circ \times$ monthly, by taking latitude-weighted averages in all dimensions, one finds a mean TCWV of $19.2 \pm 2.9 \text{ kg/m}^2$.

Reducing the CDR-2 product, $0.5^\circ \times 0.5^\circ \times$ monthly, by taking latitude-weighted averages in all dimensions, one finds a mean TCWV of $25.7 \pm 1.2 \text{ kg/m}^2$. Restricting CDR-2 to coastal grid cells, one finds a mean TCWV of $28.5 \pm 2.4 \text{ kg/m}^2$, and over ocean surfaces there is a mean TCWV of $29.0 \pm 0.54 \text{ kg/m}^2$.

The compliance with requirements is carried out under three perspectives: i) relative to user requirements as specified in URD v4.1 (see Table 7-1). Only relative values are of importance here. ii) The compliance with the product requirements as stated in PSD 6.0 (Table 7-2) is carried out by comparing absolute quantities. The results of the assessments for items i) and ii) are given in Table 7-3 (CDR-1) and Table 7-4 (CDR-2). iii) Finally, the fulfilment of GCOS requirements is discussed.

The overall statistics of the validation and intercomparison efforts for CDR-1 can be found in Table 7-3 below, which is a union of Table 4-2 and the second to last row of Table 4-3 for validation against GRUAN and SuomiNet, of Table 4-4 for CDR_pre, AIRS and ERA5, and Table 5-1 containing the consistency assessment. For GRUAN, SuomiNet and the consistencies the data is filtered by demanding the "tcwv_quality_flag" to be equal to "0" and the "atmospheric_conditions_flag" to be equal to "1". Note that for GRUAN and SuomiNet the values are total quantities, i.e. reductions in all dimensions of the complete considered sample, and not derived from timeseries. Furthermore, no weights are used in the calculations involving station-based data.

The overall statistics of the validation and intercomparison efforts for CDR-2 can be found in Table 7-4 which is just a copy of Table 4-6 but with those fields coloured where some requirement is met (see below).

Fields in Table 7-3 and Table 7-4 are filled in green if their entries meet the target user requirement of the corresponding quantity, and in blue if this is true for the product requirement. Consistencies close to the expected percentage range are marked green as well (none here, see Section 5).

Table 7-1: User requirements on CDR-1 as given in Table 6.1 of URD v4.1.

User requirement	Accuracy component [%]		Stability [%/dec]
	Systematic	Random	
Threshold	25	25	5
Target	5	5	< 1
Goal	< 1	< 1	< 0.5

Table 7-2: Expected product requirements on CDR-1 as given in Table 3.2 of PSD v4.2.

Product requirement	Accuracy component [kg/m ²]		Stability [kg/m ² /dec]
	Systematic	Random	
Target	0.75	2.75	0.5

Table 7-3: The table shows the overall statistics for CDR-1 compared to the different references. For GRUAN and SuomiNet the data is the same as in Table 4-2 and the second to last row of Table 4-3 (“actual over passes”, clear-sky). The metrics for AIRS, ERA5 and CDR_pre are taken from **Table 4-4**. Consistencies are the same as in Table 5-1. *Used CDR-1 dataset: 0.05° x 0.05° x daily.*

Reference	Mean bias [kg/m ²]	Mean bias_rel (absolute) [%]	Mean cRMSD [kg/m ²]	Mean cRMSD_rel [%]	Mean TCWV [kg/m ²]	Stability [kg/m ² /dec]	Relative stability [%/dec]	Consistency [%] c_CCI = 0.1		
								k=1	k=2	k=3
GRUAN	-0.22	1.4	1.8	11.1	16.7			57.1	82.9	91.7
SuomiNet	0.87	6.1	2.75	19.4	14.2					
ERA5	-1.6 ± 0.3	7.6 ± 1.3	2.4 ± 0.2	11.5 ± 1.1	20.8 ± 3.0	0.3	1.2			
AIRS	-1.9 ± 0.5	9.2 ± 2.2	2.5 ± 0.2	11.9 ± 1.0	21.2 ± 3.2	0.5	2.4			
CCI_pre	-0.5 ± 0.6	2.5 ± 3.0	1.9 ± 0.8	9.8 ± 3.8	19.9 ± 2.9					

The CDR-1 bias satisfies both target user and target product requirements against reference datasets GRUAN and CCI_pre (Table 7-3), but violates the target requirements for references SuomiNet, ERA5 and AIRS; in these latter cases, the threshold user requirement of 25% is comfortably met. The cRMSD against all references fulfils the product requirements. The target user requirements on the random uncertainty component are not met, but the threshold requirement of 25% is met. The target product requirement for the stability is met against both ERA5 and AIRS, as is the threshold user requirement (5%/dec), while the target user requirement is not met.

Table 7-4: The table shows the overall statistics for CDR-2 compared to the references AIRS, ERA5 and CDR_pre. The metrics are taken from Table 4-6. Used CDR-2 dataset: 0.05° x 0.05° x daily.

Reference	Mean bias [kg/m ²]	Mean bias_rel (absolute) [%]	Mean cRMSD [kg/m ²]	Mean cRMSD_rel [%]	Mean TCWV [kg/m ²]	Stability [kg/m ² /dec]	Relative stability [%/dec]
ERA5	-0.1 ± 0.1	0.4 ± 0.4	2.0 ± 0.1	7.8 ± 0.5	25.6 ± 1.3	0.1	0.3
AIRS	-0.3 ± 0.4	1.0 ± 1.4	2.2 ± 0.2	8.3 ± 0.7	25.9 ± 1.3	0.5	1.9
CCI_pre	-0.1 ± 0.1	0.3 ± 0.4	1.6 ± 0.5	6.1 ± 1.8	6.1 ± 1.8		

The CDR-2 bias satisfies both the target and the goal user requirements of 5 % and 1 %, respectively, against all datasets (Table 7-4). Likewise, the respective product requirement for bias of 0.75 kg/m² is met in all scenarios. The product requirement of 2.75 kg/m² for cRMSD is also met in all comparisons, but only the threshold user requirement of 20% is fulfilled in all cases. The product requirement for stability of 0.5 kg/m²/dec is fulfilled in both comparisons (against ERA5 and AIRS), while the goal user requirement of 0.5%/dec is met in the comparison against ERA5 and the threshold user requirement of 5% is met in the comparison against AIRS. The comparison against AIRS data has shown stronger trends throughout (section 4), so it is very likely that the higher stability value here is a result of instabilities in the AIRS record. The GCOS-245 threshold requirements are 1 kg/m² (2-

sigma uncertainty) and 0.5 kg/m²/dec (stability, dependent on latitude and season). Both CDR-1 and CDR-2 are within the threshold GCOS-245 requirement on stability (relative to ERA5 but not relative to AIRS in the case of CDR-2) but does not meet the GCOS-245 uncertainty threshold ("Mean cRMSD" columns in Table 7-3 and Table 7-4).

We recall comments that were made on the GCOS-245 requirements. Given project internal results on satellite-based TCWV validation the cRMSD is hardly smaller than 1 kg/m², i.e., in almost all cases the cRMSD is larger than the GCOS requirement. Exceptions are only observed if the compared measurements were not fully independent. Also, when looking into the literature on this subject, the observed results from comparison and validation are in the vast majority of cases larger than 1 kg/m². It can be argued that the cRMSD and the uncertainty requirements cannot be compared directly. The cRMSD includes uncertainties from the CDR, the reference and a representativeness uncertainty, while the GCOS-245 uncertainty is defined as a 2-sigma uncertainty. To conclude it will be very challenging / hardly possible to demonstrate a valid measurement uncertainty of better than 1 kg/m². Given results from a user requirements survey from ESA WV_cci (URD, v3) a fairly large range of requirements was received. As for the URD exercise, and in GCOS context, it might be adequate to consider the requirements as PDF and define the threshold requirement as an inclusive interpretation of this PDF, i.e., pick e.g. a 95% percentile or similar of the received requirements. In the WV_cci exercise the demand for higher quality data records was then expressed in the optimal requirement by looking at the 5%-percentile of received requirements.

Finally, uncertainties as assigned in CDR-1 and hence CDR-2 are likely underestimating the true uncertainty (section 5).

8. SUMMARY

In this document the WV_cci project's datasets CDR-1 and CDR-2 (version 4.2) are validated and compared to datasets GRUAN, SuomiNet, AIRS, ERA5 and its precursor version, referred to as CDR_pre. In addition, the over-land NIR-based Level-2 data are validated against ARM MW-based observations.

The validation of CDR-1 and implicitly CDR-2 over land is carried out against the station-based datasets GRUAN and SuomiNet. If "clear-sky" and "best-quality" CDR-1 daily data is considered, then CDR-1 has a small bias of opposite signs for the two references. The bias is relatively small (0.02 kg/m² and 0.2 kg/m² which correspond to 0.1% and 1.5%) and fulfills the target user and product requirements in the comparison against GRUAN. For cRMSD, the product requirement in the validation is met. Consistency of the datasets including uncertainties are not satisfied for both references. This suggests that uncertainties might be underestimated.

CDR-1 and CDR-2 are compared to the data records AIRS, ERA5 and its precursor version, both on daily and monthly basis. In all cases, the bias is negative and decreases when extending the comparison from land-only (CDR-1) to the entire globe (CDR-2) in which the case goal user and the product requirements are met.. With values between 1.9 kg/m² and 2.5 kg/m², the product requirements on cRMSD are also met.

The timeseries of the bias of CDR-1 and CDR-2 has a positive trend relative to AIRS and ERA5. The stabilities against AIRS and ERA5 are 0.5 kg/m² and 0.3 kg/m² per decade, respectively, for CDR-1 and 0.5 kg/m² and 0.1 kg/m² per decade, respectively, for CDR-2, hence within the product requirements.

When considering monthly means of CDR-1/2 and AIRS and ERA5, the regions with the highest absolute bias are found in northern South-America, central Africa, India, southern China and at the East coast of North-America. The highest cRMSD values appear in southern Asia. In general, over oceans (CDR-2), the mean biases and cRMSD are closer to zero than over land, except over sea ice areas.

A study comparing regional high-resolution data with the CDR-1 standard product shows that monthly means vary for the differently resolved datasets, and geographical structures like large rivers and mountain ranges showing biases of region-dependent signs.

For each region the overall bias between monthly means of the high-resolution data and the references AIRS and ERA5 is negative. The cRMSD is highest in region roi_1 which agrees with the findings from the study of the low-resolution CDR-1 product.

Finally, the analysis of the clear-sky bias in the SuomiNet dataset reveals an expected overall clear-sky bias of -3.3 kg/m^2 , which is unevenly distributed with more negative values in low latitudes.

APPENDIX 1: REFERENCES

ATBD (2021) Algorithm Theoretical Basis Document, Part 1 - MERIS-MODIS-OLCI L2 Products. ESA/ECSAT, WV_cci Phase One, Ref: CCIWV.REP.004, version 2.1, 21 January 2021.

Calbet, X., Peinado-Galan, N., Rípodas, P., Trent, T., Dirksen, R., and Sommer, M.: Consistency between GRUAN sondes, LBLRTM and IASI, *Atmos. Meas. Tech.*, 10, 2323–2335, <https://doi.org/10.5194/amt-10-2323-2017>, 2017.

CAR v3, Climate Assessment Report. ESA/ECSAT, WV_cci Phase One, Ref: CCIWV.REP.018, v3.0, 11 February 2021.

GCOS, The Global Climate Observing System for Climate: Implementation Plan. WMO GCOS-200, 341 pp., 2016. Available at https://library.wmo.int/doc_num.php?explnum_id=3417 [accessed 15 January, 2021].

Ghent, D., Veal, K., Trent, T., Dodd, E., Sembhi, H. and Remedios, J., 2019. A new approach to defining uncertainties for MODIS land surface temperature. *Remote Sensing*, 11(9), p.1021.

Hegglin, M. I., et al. (2013), SPARC Data Initiative: Comparison of water vapor climatologies from international satellite limb sounders, *J. Geophys. Res. Atmos.*, 118, 11,824–11,846, doi:10.1002/jgrd.50752.

Iglewicz, B. and Hoaglin, D.C., 1993. How to detect and handle outliers (Vol. 16). Asq Press.

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

Mears, C. A., Smith, D. K., Ricciardulli, L., Wang, J., Huelsing, H., & Wentz, F. J. (2018). Construction and uncertainty estimation of a satellite-derived total precipitable water data record over the world's oceans. *Earth and Space Science*, 5(5), 197–210.

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

Roman, J., R. Knuteson, T. August, T. Hultberg, S. Ackerman, and H. Revercomb (2016): A global assessment of NASA AIRS v6 and EUMETSAT IASI v6 precipitable water vapor using ground-based GPS SuomiNet stations, *J. Geophys. Res. Atmos.*, 121, 8925–8948.

PSD v4.2, Product Specification Document. ESA/ECSAT, WV_cci Phase Two, Ref: CCIWV.REP.002, version 4.2, 14 December 2023.

PVIR (2021): Product Validation and Intercomparison Report Part 1: CDR-1 & CDR-2. ESA/ECSAT, WV_cci Phase One, Ref: CCIWV.REP.016, version 2.1, 30 September 2021.

PVP v4.0: Product Validation Plan. ESA/ECSAT, WV_cci Phase Two, Ref: CCIWV.REP.008, version 4.0, 26 May 2023.

Sayer, A. M., Govaerts, Y., Kolmonen, P., Lipponen, A., Luffarelli, M., Mielonen, T., Patadia, F., Popp, T., Povey, A. C., Stebel, K., and Witek, M. L.: A review and framework for the evaluation of pixel-level uncertainty estimates in satellite aerosol remote sensing, *Atmos. Meas. Tech.*, 13, 373–404, <https://doi.org/10.5194/amt-13-373-2020>, 2020.

Schröder, M., M. Lockhoff, J. Forsythe, H. Cronk, T. H. Vonder Haar, R. Bennartz, 2016: 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>.

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, Q. Yang, 2019: 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>.

Sohn, B. J., & Bennartz, R., 2008: Contribution of water vapor to observational estimates of longwave cloud radiative forcing. *Journal of Geophysical Research: Atmospheres*, 113(D20).

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/ethza-010863911. (available at www.sparc-climate.org/publications/sparc-reports/sparc-reportno8/).

Stengel, M., Stapelberg, S., Sus, O., Schlundt, C., Poulsen, C., Thomas, G., Christensen, M., Henken, CC., Preusker, R., Fischer, J., Devasthale, A., Wille, U., Karlsson, KG., McGarragh, GR., Proud, S., Povey, AC., Meirink, JF., Feofilov, A., Bennartz, R., Bojanowski, JS. & Hollmann, R. (2017). Cloud property datasets retrieved from AVHRR, MODIS, AATSR and MERIS in the framework of the Cloud_cci project. *Earth System Science Data*, **9(2)**, 881-904. URD v4.1 User Requirement Document, ESA/ECSAT, WV_cci Phase Two, Ref: CCI WV.REP.001, version 4.1, 12 April 2023.

Wang, X.L. (2008a): Penalized maximal F test for detecting undocumented mean shift without trend change. *J. Atmos. Ocean. Technol.* 25, 368–384.

Wang, X.L. (2008b): 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.

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 Radiometer
BC	Brockmann Consult
CARIBIC	Civil Aircraft for the Regular Investigation of the atmosphere Based on an Instrument Container
CCI	Climate Change 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
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	Deutsches Zentrum 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	ERA5 - reduced resolution ten member ensemble
EMiR	ERS/Envisat MWR Recalibration and Water Vapour Thematic Data Record Generation
Envisat	Environmental Satellite

Term	Definition
ERA5	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
GRUAN	GCOS Reference Upper-Air Network
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
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
MD	Mean difference
MERIS	Medium Resolution Imaging Spectrometer Instrument
MERRA-2	Modern-Era Retrospective analysis for Research and Applications, Version 2

Term	Definition
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
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
SE	Spectral Earth
SMILES	Solar wind Magnetosphere Ionosphere Link Explorer
SMR	Software Modification Report
SNR	Signal-to-noise ratio
SOFIE	Solar Occultation For Ice Experiment
SPARC	Stratosphere-troposphere Processes And their Role in Climate

Term	Definition
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