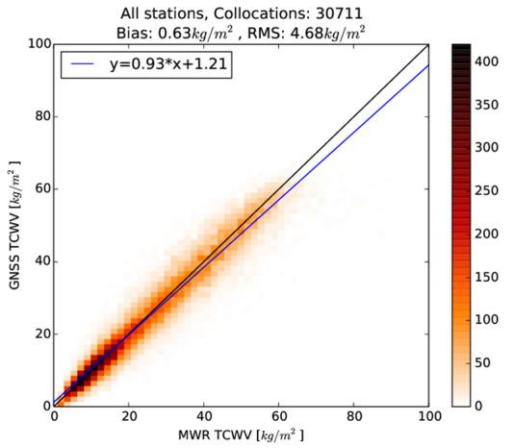

EMiR TCWV compared to GNSS TCWV. The least square fit is shown in blue.

ESA contract

4000109537/13/I-AM

ERS/Envisat MWR Recalibration and Water Vapour TDR Generation (EMiR)

EMiR Validation Report

(DLV-EXT-08)

Berlin, 20. February 2017

Document control sheet		
Project	ERS/Envisat MWR recalibration (EMiR)	
Client	European Space Agency, ESRIN Via Galileo Galilei, Casella Postale 64 00044 Frascati, Italy http://www.esa.int	
Main contractor	Informus GmbH Belziger Str. 44 10823 Berlin, Germany http://www.informus.de	
Consortium partners	Collecte Location Satellites	
	Deutscher Wetterdienst	
	Swedish Meteorological and Hydrological Institute	
Title of document	EMiR Validation Report	
Authors of document	Frank Fell, Ralf Bennartz, Stefano Casadio, Hannes Diedrich, Heidrun Höschen, Maarit Lockhoff, Bruno Picard, Marc Schröder, Ulrika Willén	
Status	Released	
Distribution	EMiR consortium, ESA	
Version control	2016-11-30 0.90 Draft 2017-01-17 0.99 Revised draft 2017-02-20 1.00 Final editorial revision, initial version	

Table of contents

List of tables

List of figures

Executive summary

A new data record (termed EMiR) of the total column water vapour (TCWV) and derived wet tropospheric correction (WTC) was generated from observations of the Microwave Radiometer (MWR) instruments flown on board the satellites ERS-1, ERS-2, and Envisat using optimal interpolation retrieval techniques. To achieve consistent time series from these three instruments, a new method for intercalibrating the brightness temperatures from succeeding MWRs was developed and successfully applied.

The EMiR data record covers the period from 1992/10/23 to 2012/04/08. While the end of the EMiR data record coincides with the loss of Envisat and the related end of MWR observations, EMiR does not cover the first ca. 15 months of ERS-1 operations due to the lack of specifically pre-processed MWR L1B brightness temperatures used as input to the EMiR processing. A number of additional gaps occur, most of them lasting between one and ten days. Despite these gaps, EMiR data availability is very good after 1994/04/10.

EMiR products have been compared against ground-based (GNSS) satellite-based (different approaches), and climate model based (EC Earth) reference data:

- *EMiR brightness temperatures were validated against EC-Earth simulations. As expected, some channel dependent biases were observed, but no temporal trends.*
- *EMiR TCWV compares well with the reference data, with absolute bias values mostly smaller reference data sets ranging from than 0.5 kg/m^2, and statistically insignificant temporal trends.*
- *The wet tropospheric correction (WTC) derived from the EMiR TCWV is almost as accurate as is ESA's operational WTC, even though the latter incorporates additional information on the sea surface roughness, which has not yet been introduced into EMiR processing.*

Concluding, the EMiR data record is deemed mature and accurate enough to be used for climatological and oceanographic applications. A number of suggestions were made to further enhance the quality and user friendliness of the EMiR data record.

1 Introduction

1.1 Purpose of this document

This document summarises the efforts made to assess the quality of the *total column water vapour* (TCWV) and *wet tropospheric correction* (WTC) products derived in the context of the project *ERS/Envisat MWR Recalibration and Water Vapour Thematic Data Record Generation* (EMiR) from observations of the *Microwave Radiometer* (MWR) flown on board the satellites ERS-1, ERS-2, and Envisat. Both products cover the global oceans for the period 1992 to 2012.

The EMiR project was funded by the *European Space Agency* (ESA) through its *Long-Term Data Preservation* (LTDP) programme (https://earth.esa.int/web/gscb/ltdp) and lasted from November 2013 to October 2016. Further information about EMiR can be found at http://esa-mwr.org.

Based on the findings of the validation and verification activities presented in this report, improvements are suggested towards further enhancing quality and usefulness of the EMiR TCWV and WTC data records.

1.2 The Microwave Radiometer series

The MWRs are nadir-pointing passive instruments measuring the *top of atmosphere* (TOA) brightness temperatures in two spectral channels. They have been part of the payload of the ERS-1, ERS-2 and Envisat satellites (see Table 1 for more details). Similar instruments are part of the configuration of the Sentinel-3 series of satellites such that the continuity of MWR observations is assured for next two decades.

Table 1: Main characteristics of the Microwave Radiometer series[1].

Absolute accuracy	Brightness temperature: ca. 2.6 K
3-dB beam width	1.5 degrees
Spatial resolution	20 km (from an altitude of approx. 780 km)
Swath width	20 km
Frequencies	23.8 GHz and 36.5 GHz
MWR lifetime (see also Table 2)	ERS-1: 07/1991 - 06/1996 ERS-2: 04/1995 - 07/2011 Envisat: 03/2002 - 04/2012

The main purpose of the MWRs is to provide the observations required to assess the tropospheric path delay of concomitant altimeter observations caused by tropospheric water vapour as well as to provide an estimate of the attenuation of the altimeter signal by the cloud *liquid water path* (LWP).

[1] Source: Envisat/RA-2/MWR Product Handbook, http://earth.esa.int/pub/ESA_DOC/ENVISAT/RA2-MWR/ra2-mwr.ProductHandbook.2_2.pdf

This is achieved by measuring the *brightness temperatures* (Tb) at 23.8 GHz and 36.5 Ghz, which are sensitive to tropospheric water vapour and liquid water, respectively. Combined with the altimeter backscattering coefficient (sigma_0), the Tbs measured by the MWR instruments therefore allow for the determination of the *wet tropospheric correction* (WTC), quantifying the path delays of altimeter observations caused by atmospheric water vapour. Due to its large spatial and temporal variability, WTC is the most critical correction in the altimeter error budget.

Beyond their use in altimetry, global time series on TCWV and cloud liquid water have a high scientific value per se due to the paramount importance of clouds and water vapour in the climate system.

The MWR instruments have been flown in different orbital configurations and underwent a number of instrumental problems, which are likely responsible for some of the issues observed when validating EMiR data products. Table 2 sums up the main events during the lifetimes of MWR on board ERS-1 and ERS-2.

Table 2: Life and operation cycles of the MWR instruments on-board ERS-1, ERS-2, and Envisat.

Time	MWR on ERS-1	MWR on ERS-2	MWR on Envisat
1991/07/17	ERS-1 launch		
1991/07/17 – 1991/12/28	35-day period		
1991/12/28 – 1992/03/30	3-day period		
1992/03/30 – 1993/12/24	35-day period		
1993/12/24 – 1994/04/10	3-day period		
1994/04/10 – 1995/03/21	168-day period		
1995/03/21 – 1995/05/15	35-day period		
1995/04/21		ERS-2 launched	
1995/05/15 - 1996/06/02	ERS-1 and ERS-2 on identical orbits (35 days) with a 1-day shift		
1996/06/02	ERS-1 decommissioned		
1996/06/26		Gain drop on 23.8 GHz, channel, which starts drifting (1)	
2000/03/31	Retired		
2002/03/01			Envisat launched
2003/06/22		Tape recorder incident (2)	
2011/07/06		ERS-2 decommissioned	
2012/04/08			Connection to Envisat lost

(1) After pass number 650 in cycle 12, a gain drop in the 23.8 GHz channel occurred, probably due to an amplifier break down. From this date, a drift in the 23.8 GHz observations is detected.

(2) Tape recorder "A" stopped functioning on 22 June 2003. Only observations with the spacecraft in the line of sight of a receiving station could be recorded after that date. Due to the existence of several such stations in Europe and Canada, continuous ERS-2 coverage of these regions including the North Atlantic could be achieved.

1.3 Assessed dataset

The main outcomes of the EMiR project consist of two *fundamental data records* (FDRs) of satellite inter-calibrated TOA brightness temperatures at 23.8 GHz and 36.5 GHz, as well as a *thematic data record* (TDR) on atmospheric TCWV over the global ice-free oceans derived from those.

Consistent TOA brightness temperatures were obtained by intercalibration of the three different MWR instruments, using constraints on liquid water path (LWP) and TCWV. A one-dimensional variational (1D-VAR) approach has been applied to derive TCWV from the intercalibrated brightness temperatures and ERA-Interim background information. The EMiR TCWV has subsequently been used to calculate the WTC, which is an important input for generating accurate sea level maps from the altimeter observations on-board ERS-1/2 and Envisat. Further results comprise the cloud LWP as well as ancillary retrieval parameters (e.g. cost function values). Details on the applied retrieval method can be found in the EMiR *Algorithm Theoretical Basis Document* (ATBD) *[EMiR DLV-EXT-07, 2016]* and references listed therein, as well as in *[Bennartz et al., 2016]*.

The quality assessment presented herein refers to the EMiR dataset in version 1.0, available free of charge under DOI 10.5676/DWD_EMIR/V001. On the DOI landing page, further information on EMiR data formats, file contents, etc. can be obtained.

The EMiR data record covers most of, but not the entire, MWR observation period. The main data gaps and their causes (as far as known) are listed in Table 3. Where not explicitly specified, data gaps may result from a number of causes[2], including, but not limited to instrumental issues, orbital manoeuvres, downlink interruptions, ground-segment processing failures.

Table 3: Data gaps ≥3 days in the EMiR Level-2 data record as well as their causes, as far as known to the EMiR consortium.

Date / Period	Mission			Explanation	Can it be fixed?
1991/07/17 – 1991/12/28	ERS-1			Commissioning phase, no altimeter time tag	Yes
1991/12/28 – 1992/10/22	ERS-1			No altimeter time tag	Yes
1992/10/23	ERS-1			First EMiR L2 product from ERS-1	
1993/12/21 – 1994/04/09	ERS-1			No altimeter time tag	Yes
1995/10/02		ERS-2		First EMiR L2 product from ERS-2	
1996/06/02	ERS-1			Last EMiR L2 product from ERS-1 (MWR switched off)	
2000/07/01 – 2000/07/10		ERS-2			

[2] See e. g. ESA's quality control reports for MWR on Envisat: the https://earth.esa.int/web/sppa/sppa/mission-performance/esa-missions/envisat/mwr/quality-control-reports

Date / Period	Mission			Explanation	Can it be fixed?
2001/01/18 – 2001/02/08		ERS-2			
2002/03/09 – 2002/03/19		ERS-2			
2002/05/14			Envisat	First EMiR L2 product from Envisat	
2002/05/18 – 2002/05/23			Envisat		
2002/05/27 – 2002/06/10			Envisat		
2002/04/11 – 2002/11/30		ERS-2			
2003/06/22		ERS-2		Last EMiR L2 product from ERS-2 (tape recorder incident)	
2003/12/07 – 2003/12/09			Envisat		
2006/12/13 – 2006/12/15			Envisat		
2010/10/23 – 2010/10/25			Envisat		
2012/04/08			**Envisat**	**Last EMiR L2 product from Envisat (contact to Envisat lost)**	

In addition to the gaps in the EMiR data record listed in Table 3, short term data gaps of one or two days do also occur (Table 4). The total number of dates missing within the data record amounts to 128 for ERS-1, 85 for ERS-2, and 44 for Envisat. In general, all EMiR data in temporal vicinity to data gaps should be treated cautiously. It is recommended that future reprocessed EMiR versions aim at closing the temporal gaps in the data records to the best possible extent.

Table 4: Full list of gaps within the EMiR Level-2 data record, i.e. in between the first day (1992/10/23) and last day (2012/04/08) of product availability. Gaps lasting three days or longer are highlighted in bold.

Mission	Year	Data gaps	Days missing
ERS-1	1993	03/22, 05/03-05/04, 05/07, 05/10, 08/07, 08/10, 08/15, 08/17, 08/23, 08/26, **12/21-12/31**	22
ERS-1	1994	**01/01-04/09**	99
ERS-1	1995	03/22-03/23, 06/19, 11/29, 12/06, 12/14, 12/17	7
ERS-2	1996	01/25	1
ERS-2	1998	06/04-06/05	2
ERS-2	1999	02/06	1
ERS-2	2000	01/01, 02/08-02/09, **07/01-07/10**, 10/08-10/09	15
ERS-2	2001	**01/18-02/08**, 05/22-05/23, 11/18	25

Mission	Year	Data gaps	Days missing
ERS-2	2002	**03/09-03/19, 11/04-11/30**	38
ERS-2	2003	05/17-05/18	2
Envisat	2002	**05/18-05/23, 05/27-06/10,** 09/09	18
Envisat	2003	01/26, 02/21, 03/16, 09/05, 12/04, **12/07-12/09**	8
Envisat	2006	04/07, 05/27-05/28, 09/08-09/09, **12/13-12/15**	8
Envisat	2007	02/17-02/18, 03/10-03/11, 07/01, 09/25-09/26	7
Envisat	2010	**10/23-10/25**	3

1.4 Validation approach

According to the definition given by *Justice et al. [2000],* subsequently adopted by the *Committee on Earth Observation Satellites* (CEOS), validation refers to "*the process of assessing, by independent means, the quality of the data products derived from system outputs*".

Where the product to be quality controlled cannot be directly validated against independent measurements, consistency checks and plausibility considerations should be attempted to get insight into product performance and quality.

Considering the availability of reference data for the EMiR products, the validation strategy summarised in Table 5 was applied. LWP is considered a retrieval by-product in the context of this study, and therefore has not been validated herein.

Table 5: Specific validation approaches for the main EMiR products Tb, TCWV, and WTC.

Validation approach	Tb	TCWV	WTC
Plausibility considerations	Applied to identify potential problems	Applied to identify potential problems	Applied to identify potential problems
Simulations	EMiR L3 vs. EC-Earth Tb (➔ 2.4)	EMiR L3 vs. EC-Earth TCWV (➔ 2.4)	EMiR WTC vs. ERA-Int. derived WTC (➔ 2.6)
Space-borne retrievals	Not applied	EMiR L2 vs. TCWV from – MERIS CAWA (➔2.2) – AIRWAVE (➔ 2.3) – G-VAP (➔ 2.5)	EMiR WTC vs. operational MWR WTC (➔ 2.6)
Ground-based retrievals	Not available	EMiR L2 vs. TCWV from GNSS (➔ 2.1)	Not applied
Direct observations	Not available	Not available	Not available

1.4.1 Brightness temperature

Due to the specific characteristics of the MWR channels (e. g. operating frequencies, observation geometry), their large satellite footprints, and the general lack of knowledge on the emissivity of the Earth surface in the considered spectral range, neither direct observations nor other satellite-retrieved reference information is available to validate the EMiR brightness temperature product based on a pixel collocation approach.

In this case, simulations provide a way to assess the quality of the EMIR Tb data record. While based on many assumptions, they can provide indications on areas or periods where EMiR retrievals are potentially affected by quality issues. In this report, the EMiR brightness temperature records have been compared against simulations produced with the climate model EC-Earth version 3 (section 2.4).

1.4.2 Total column water vapour

The availability of reference data is better for TCWV than it is for the brightness temperature.

Water vapour retrievals obtained from the *Global Navigation Satellite System* (GNSS) have been identified as being best suited for EMiR TCWV validation, as they combine a totally independent retrieval approach characterised by high accuracies on the order of 1-2 kg/m^2 with good data availability in coastal areas and on islands *[Gendt et al, 2004]*. Comparison against GNSS-based TCWV is therefore at the core of the EMiR validation (section 2.1).

Within the frame of the EMiR project, TCWV validation was also performed against the MERIS-based *Advanced Clouds, Aerosols and Water Vapour Products* (CAWA) products (section 2.2), as well as simulations obtained with the EC-Earth version 3 climate model (section 2.4).

Further TCWV validation took place in the context of external activities not funded through EMiR, most importantly against the *Advanced Infra-Red Water Vapour Estimator* (AIRWAVE) data record carried out by ESA (section 2.3) as well as a comparison against more than twenty satellite and re-analysis based TCWV data records in the context of the *GEWEX Water Vapor Assessment* (G-VAP, section 2.5).

Radiometer-based TCWV retrievals, e.g. from the *Aerosol Robotic Network*[3] (AERONET), were not used for EMiR TCWV validation due to the difficulty in obtaining a sufficient number of observations representing open ocean conditions. Please note that MWR observations closer than 100 km to land are not part of EMiR due to potential side lobe contamination from land surfaces (section 2.1.1).

Similar limitations apply to radiosonde TCWV retrievals. In addition, the various global radiosonde datasets exhibit a number of incompatibilities and are considered to provide less accurate TCWV values than does GNSS. Radiosonde TCWVs were therefore also not used for EMiR TCWV validation.

The different approaches to validate EMiR TCWV are described in more detail in sections 2.1 to 2.6. While the strategies differ among the various validation activities, they altogether provide a good characterisation of the strengths and weaknesses of the EMiR TCWV data record.

1.4.3 Wet tropospheric correction

Altimetry satellites basically determine the distance from the satellite to a target surface by measuring the satellite-to-surface round-trip time of a radar pulse. The principle is that the altimeter emits a radar wave and analyses the return signal that bounces back from the surface. Sea surface height (SSH) is then the difference between the satellite's altitude (i.e. the height above the chosen reference

[3] Aerosol Robotic Network: https://aeronet.gsfc.nasa.gov/ (accessed 2017/01/17)

ellipsoid) and the satellite-to-surface range, inferred from the time it takes for the signal to make the round trip *[EMiR DLV-INT-12, 2016]*.

The tropospheric path delay, i.e. the deceleration of electromagnetic radiation in the atmosphere, has the largest impact on the accuracy of altimetry-based *sea surface height* (SSH) retrievals, requiring corrections of up to ca. 2.5 m *[Fu and Cazenave, 2000]*. More than 80% of the tropospheric path delay relates to the dry atmosphere, while up to 20% is due to atmosphere moisture. However, atmospheric moisture is highly variable and may therefore cause significant errors in the retrieved SSH if not properly accounted for.

The so-called *wet tropospheric correction* (WTC) accounts for the path delay in the radar altimeter return signal due to atmospheric moisture. Over the ocean, it is usually derived from microwave radiometer observations taken concomitantly to the actual altimeter observations. Over land surfaces, the wet tropospheric correction is usually derived from atmospheric models.

In the context of EMiR, a WTC product has been derived using the EMiR TCWV product as the most important input parameter *[Bennartz et al., 2016]*. The performance of the EMiR WTC has been assessed by comparing it against a re-analysis (ERA-Interim) based WTC as well as the WTC operationally applied by ESA for processing the altimeter observations from ERS-1, ERS-2, and Envisat (section 2.6). The comparison is based on the SSH cross-over approach, i.e. SSH differences between ascending and descending nodes with a maximum temporal difference of 10 days are analysed. Assuming that the SSH variability is negligible over this short period *[Legeais et al., 2014]*, the best WTC will then result in the smallest SSH variance at cross-overs.

1.5 Acronyms and abbreviations

Acronym	Description
(A)ATSR	(Advanced) Along Track Scanning Radiometer
AERONET	Aerosol Robotic Network
AIRWAVE	Advanced Infra-Red Water Vapour Estimator
AMIP	Atmospheric Model Intercomparison Project
ARSA	Analyzed Radio Soundings Archive
ASL	Above sea level
ATBD	Algorithm Theoretical Basis Document
CAWA	Advanced Clouds, Aerosols and Water Vapour Products
CEOS	Committee on Earth Observation Satellites
CCI	Climate Change Initiative
CLS	Collecte Localisation Satellites
EC-Earth (ECE)	European Consortium Earth System Model
ECMWF	European Centre for Medium-Range Weather Forecasts
EMiR	ERS/Envisat MWR Recalibration and Water Vapour FCDR Generation
Envisat	Environmental Satellite
EO	Earth observation
ERA-Interim	Global atmospheric reanalysis from 1979 to present by ECMWF
ERS	European Remote Sensing satellite
ESA	European Space Agency

Acronym	Description
FDR	Fundamental data record
GDAP	GEWEX Data and Assessments Panel
GEWEX	Global Energy and Water Cycle Exchanges Project
GNSS	Global Navigation Satellite System
GPS	Global Positioning System
GTOPO30	Global 30 Arc-Second Elevation Data
G-VAP	GEWEX Water Vapor Assessment
HIRS	High-resolution Infrared Radiation Sounder
HOAPS	Hamburg Ocean Atmosphere Parameters and Fluxes from Satellite
IFS	Integrated Forecast System
ITCZ	Intertropical convergence zone
LTDP	Long-Term Data Preservation
LWP	Liquid water path
L1	Level 1 (processing or data)
L2	Level 2 (processing or data)
L3	Level 3 (processing or data)
MERIS	Medium Resolution Imaging Spectrometer
MWR	Microwave Radiometer
NCAR	National Center for Atmospheric Research
NIR	Near infrared
RMSD	Root mean square difference
RTTOV	Radiative transfer for TOVS
SDD	Standard deviation of the difference
SSH	Sea surface height
SSM/I	Special Sensor Microwave Imager
Tb	Brightness temperature
TCWV	Total column water vapour
TDR	Thematic data record
TOA	Top of atmosphere
TOVS	TIROS Operational Vertical Sounder
TIR	Thermal infrared
USGS	United States Geological Survey
VIS	Visible (part of the spectrum)
WTC	Wet tropospheric correction
1D-VAR	One-dimensional variational (retrieval)

2 Validation of the EMiR total column water vapour product

2.1 Comparison against GNSS

2.1.1 Method

The velocity of light in the atmosphere is determined by the atmospheric refractive index, depending itself mainly on the vertical distribution of air pressure, air temperature and total amount of atmospheric water vapour. Ground-based receiving stations of the *Global Positioning System* (GPS) provide the observations to estimate the atmospheric delay with great accuracy, i.e. the delay relative to electromagnetic radiation propagating in a vacuum. Combining atmospheric delay measurements with appropriate models of the dry atmosphere allows to estimate the atmospheric TCWV with good accuracy (RMSE <2 kg/m^2, *[Gendt et al, 2004]*).

Here, the *NCAR Global, 2-hourly Ground-Based GPS Precipitable Water*[4] dataset in version 721.1, containing TCWV data at 997 stations covering the period 1995 to 2014 with a temporal resolution of two hours, has been used as reference for the validation of the EMiR TCWV. More information on the NCAR GNSS TCWV dataset and the applied retrieval scheme can be found in *[Wang et al., 2007]* and *[Wang and Zhang, 2009]*.

The following criteria were applied to collocate MWR- and GNSS-derived TCWV values:

- Minimum of 100 km spatial distance between MWR observations and the closest coastline to avoid side lobe contaminations from land surfaces.
- Less than 150 km spatial distance between MWR observations and GNSS sites to ensure the best spatial consistency considering the required distance to land (see previous criterion).
- Less than one hour time difference between MWR and GNSS observations to ensure good temporal consistency.
- EMiR TCWV retrievals with corresponding cost function values ≤ 5, identical to the threshold chosen for the calculation of the EMiR Level 3 products *[EMiR DLV-EXT-07, 2016]*.
- Only GNSS sites at altitudes <500 m *above sea level* (ASL) to avoid systematic errors due to decreasing atmospheric water vapour with altitude *[Ross and Elliot, 1996]*.

In case several MWR observations match the collocation criteria above for an individual GNSS observation, the one closest to the GNSS station is selected (nearest neighbour).

The full collocation dataset comprising more than 30.000 entries (Figure 1) has been used to assess bias, bias corrected *root mean square difference* (RMSD), and stability of the EMiR TCWV. While bias and RMSD were calculated from individual collocations, the stability was assessed from monthly

[4] Earth Observing Laboratory/National Center for Atmospheric Research/University Corporation for Atmospheric Research (2011), NCAR Global, 2-hourly Ground-Based GPS Precip. Water, http://rda.ucar.edu/datasets/ds721.1/, Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory, Boulder, Colorado, US.

means, the latter being calculated as arithmetic average of all collocations available within a month, i.e. no geographic subsampling was applied. The statistical analysis is based on linear least-square regression[5].

2.1.2 Results

More than 100 sites contained in the NCAR GNSS dataset provide TCWV observations meeting the collocation criteria defined in section 2.1.1. Most of these sites are located along the coastlines of Europe and North America, but there are also several sites far away from continental influence offering large numbers of collocations (Figure 1). While not fully representative of the conditions over the global oceans, the collocation data set is deemed appropriate to assess the performance of the EMiR TCWV product at global scale.

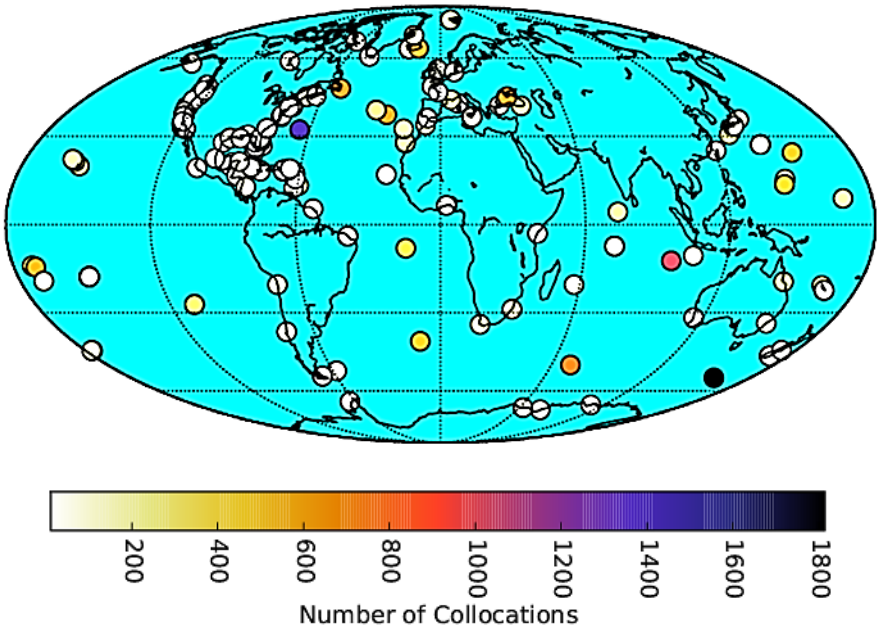

Figure 1: Stations contained in the NCAR GNSS dataset with number of EMiR collocations meeting the criteria defined in section 2.1.1.

Figure 2 summarises the comparison of EMiR against the NCAR GNSS TCWV values. The top left panel plots EMiR against GNSS TCWV for the full collocation dataset. While the bias is very small (+0.63 kg/m^2, i.e. GNSS TCWV slightly larger on average), the RMSD amounts to a non-negligible 4.68 kg/m^2. The slope (+0.93) of the regression indicates that EMiR adopts larger TCWV values than does NCAR GNSS with increasing TCWV.

A number of reasons possibly contribute to the observed behaviour:

- TCWV decreases with increasing surface altitude. This translates into a systematic TCWV reduction at GNSS sites significantly located above sea level.

[5] http://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.linregress.html

- Heavily precipitating cases may not have been entirely excluded from the analysis, adding noise to EMIR TCWV retrievals.
- The collocation uncertainty caused by the "generous" spatial collocation criteria additionally contributes to the comparably large RMSD.
- GNSS-based TCWV retrievals at sites situated along the continental coastlines encompass atmospheric conditions over both land surfaces and the ocean, depending on the position of the GPS satellite relative to the ground station. This may lead to systematic differences with EMiR TCWV retrievals always representing ocean conditions.

The relative differences between GNSS and EMiR TCWV monthly mean values are depicted in Figure 2, top right panel. While there is a certain scatter for the early ERS-1/2 years, followed by a systematic reduction starting in 2003 (overlap period of ERS-2 and Envisat), the overall temporal stability is high, showing only a very small trend +0.68 % per decade (i.e. EMiR TCWVs increasing as compared to NCAR GNSS), which is not significantly different from zero ($p \leq 0.05$).

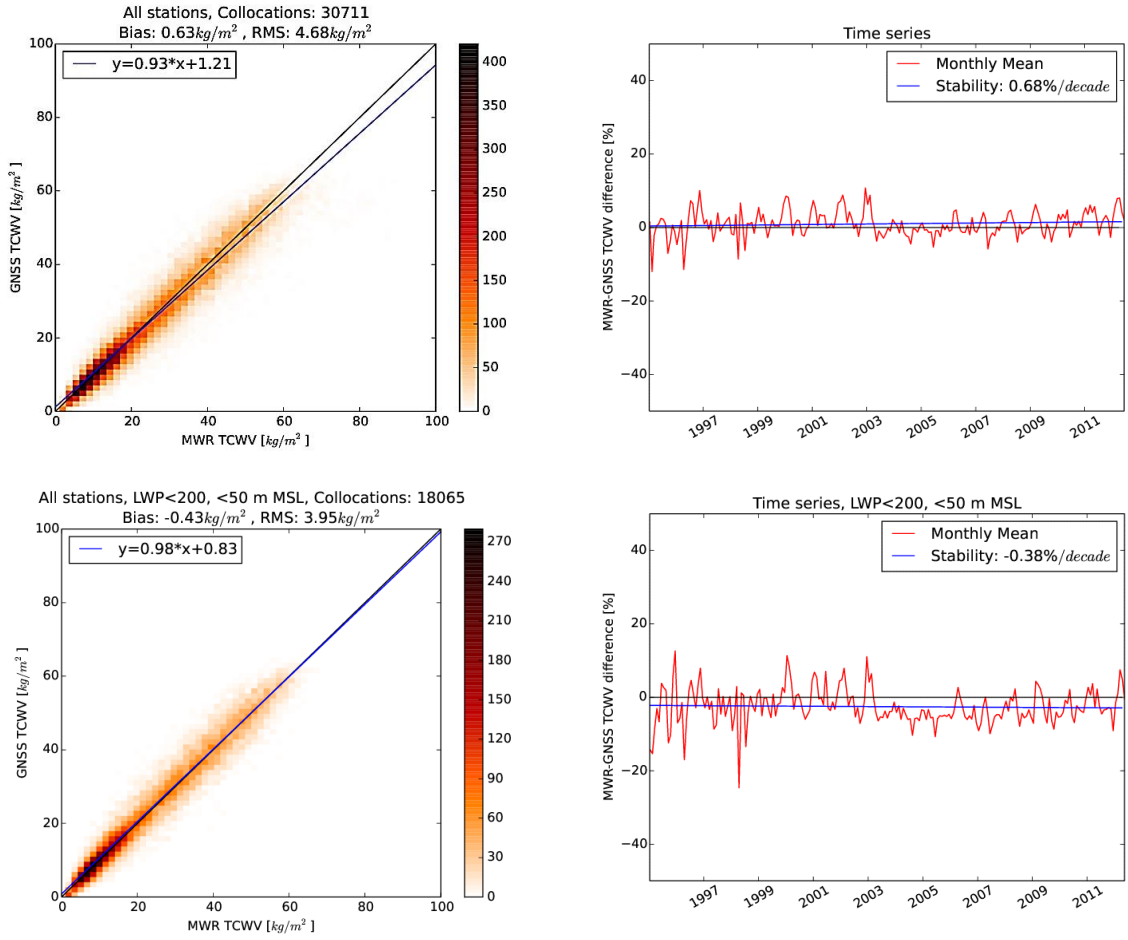

Figure 2. Left column: EMiR vs. NCAR GNSS TCWV for individual collocations. The least square fit is shown in blue and the 1:1 line in black. Right column: time series of relative TCWV differences (EMiR-GNSS)/GNSS derived from monthly mean values using stations with more than ten collocations. Top row: all collocations; bottom row: subset of collocations with LWP < 200 kg/m^2 and site altitude < 50 m ASL.

In order to assess the impact of two of the above-mentioned error sources in more detail, the analysis was repeated for a subset of the collocation dataset, where GNSS sites located at altitudes > 50 m ASL as well as MWR observations likely affected by precipitation (LWP >200 g/m^2) were discarded (Figure 2, bottom left). This reduces the number of collocations to 18,065. Bias, RMSD, and regression slope now adopt values of -0.43 kg/m^2, 3.95 kg/m^2, and +0.98, respectively. This indicates that GNSS stations in the altitude range between 50 to 500 m ASL do indeed show on average systematically reduced TCWV values as compared to those at sea level, especially at high absolute TCWV levels.

The temporal stability of the filtered dataset (Figure 2, bottom right) is higher (-0.38 %/decade), however, the "scatter" of the monthly mean values is enhanced, especially for the early years. The reasons for this are not fully clear yet, however, it is suspected that the reduced number of collocations has a role in this case.

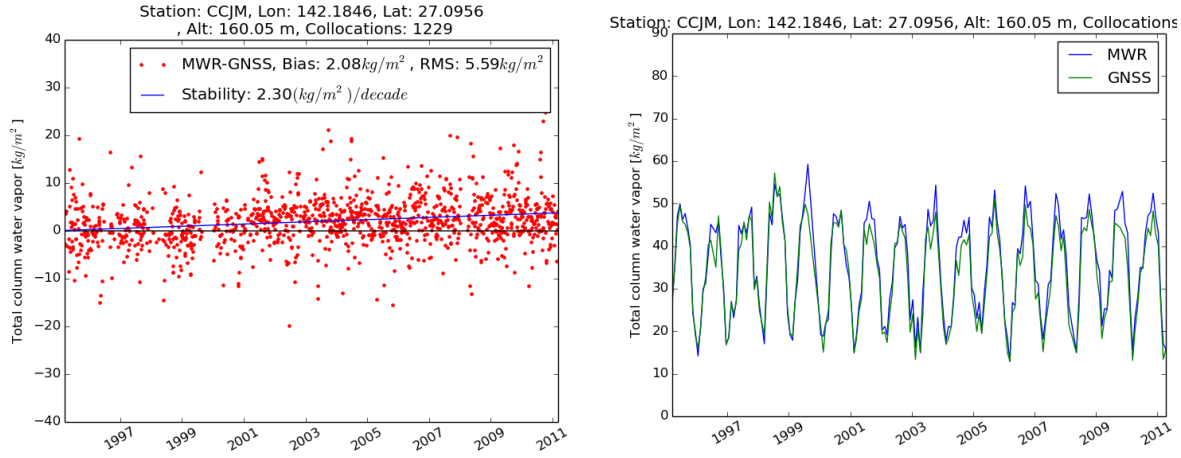

Figure 3: Time series at GNSS site CCJM (Pacific island Chichijima). Left: TCWV difference EMiR-GNSS. Right: TCWV monthly mean values for both EMiR and GNSS.

Time series for one individual GNSS site with a large number of collocations (N=1229) are exemplarily shown for site CCJN (Figure 3), located on the Japanese island of Chichijima about 1000 km south of Japan's main island of Honshu. The left panel shows the temporal evolution of the TCWV difference (EMiR – GNSS). As expected, higher scatter is observed for the individual site (RMSD = 5.59 kg/m^2) as compared to the entire collocation data set. EMiR TCWV values are on average higher (bias: +2.08 kg/m^2), to which the non-negligible altitude of the GNSS site located at 160 m ASL likely contributes. The right panel shows time series of EMiR and GNSS monthly mean TCWV values. While good agreement is observed in winter, EMiR TCWV is systematically higher in the summer months. This seasonal pattern can partly be explained by the fact that the impact of station height on TCWV is larger in the humid summer months. Local geographic or station-specific effects may further contribute to the observed differences.

2.1.3 Conclusions

EMiR and NCAR GNSS-derived TCWV values agree well on the global scale. Differences and scatter are larger at individual stations.

If only considering EMiR retrievals for dry or at most mildly precipitating conditions, and GNSS retrievals from sites situated at altitudes ≤ 50 m ASL, a small bias of -0.43 kg/m^2 is observed.

EMiR TCWV shows a trend of -0.38 % per decade relative to NCAR GNSS TCWV, which is statistically not significantly different from zero and very close to the GCOS requirement of ±0.3 %/decade. The observed RMSD of 3.95 kg/m^2 is mainly explained by the "soft" collocation criteria resulting from the necessity to consider only EMiR TCWV values with a minimum distance of 100 km from the coastline (and thus from the closest GNSS site).

2.2 Comparison against MERIS CAWA

2.2.1 Method

In the context of the ESA-funded SEOM-S3 project *Advanced clouds, aerosols and water vapour products* (CAWA)[6], a generic TCWV retrieval has been established [*Lindstrot, 2012*]. It is based on a 1D-VAR procedure that minimizes the difference between measured and simulated TOA radiances in the water vapour absorption band around 900 nm with the aid of optimal estimation, and provides TCWV values and uncertainty estimates on a per-pixel basis. The retrieval is limited to cloud-free day time scenes. While applicable above land and ocean, the uncertainties generally increase over ocean surfaces due to the lower signal-to-noise ratio.

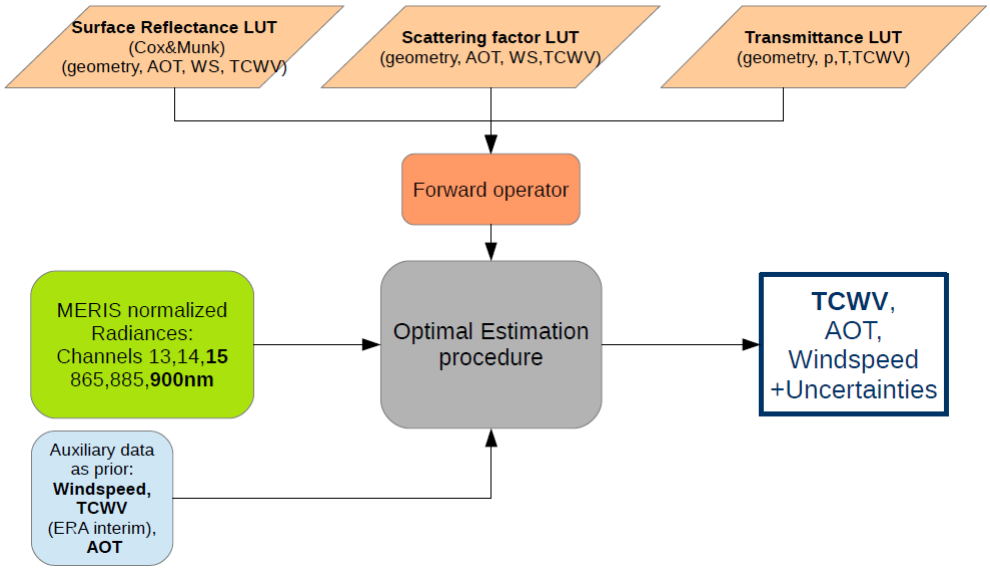

Figure 4: Structure of the MERIS CAWA retrieval scheme.

CAWA has been specifically implemented for the *Medium Resolution Imaging Spectrometer* (MERIS). CAWA retrieved TCWV is fundamentally suitable for a cross-comparison with EMiR TCWV, since it is based on a physically differing retrieval mechanism (VIS/NIR *vs.* microwave observations), and is

[6] CAWA project website: http://esa-seom-cawa.net/ (accessed 2016/11/18)

available for large number of collocations as MERIS and MWR have both been operated on the same satellite, Envisat.

The general functionality of the MERIS CAWA processor is shown in Figure 4. TCWV, *aerosol optical thickness* (AOT) and wind-speed (as a proxy of the sea surface roughness) are the main factors contributing to the variability of the TOA reflectance. Consequently, these variables are part of the state vector that is to be retrieved.

For the comparison against EMiR, MERIS CAWA TCWV was retrieved for ocean pixels for the period 01/2003 to 04/2012. EMiR Level 2 TCWV data from individual orbits were used for comparison. Spatial collocation was done by identifying all valid CAWA TCWV values within a radius of 50 km around each valid EMiR TCWV pixel. The following statistical measures were derived: mean TCWV and standard deviation as well as cloud fraction. Temporal collocation is not required due to the concomitant nature of MERIS and MWR observations. Despite the reduced availability of CAWA TCWV values above dark ocean surfaces due to insufficient signal levels, more than 2.5 Mio. EMiR vs. CAWA TCWV collocations were available for the comparison.

2.2.2 Results

Of all available collocations, only those with CAWA cloud fractions below 10% were selected to minimise the impact of undetected clouds. While the correlation between the two data sets is very high (r^2= 0.97), CAWA shows a significant wet bias of ca. +3.8 kg/m^2 as compared to EMiR, which is rather constant over the entire observed TCWV range (Figure 5).

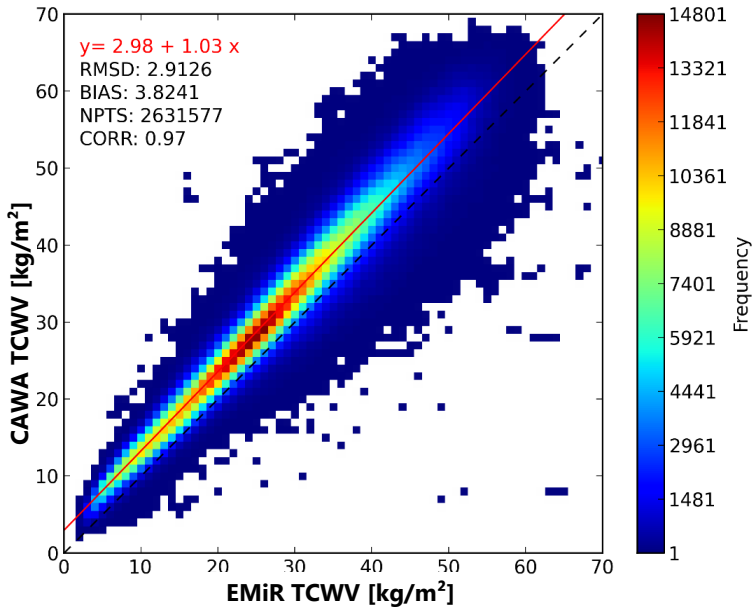

Figure 5: Scatterplot of EMiR *vs.* MERIS CAWA TCWV.

Figure 6 shows the global monthly mean TCWV values for both EMiR and CAWA data sets as well as their difference. Although seasonal variations are observed for both the individual data sets and their difference, no obvious long-term trends are observed.

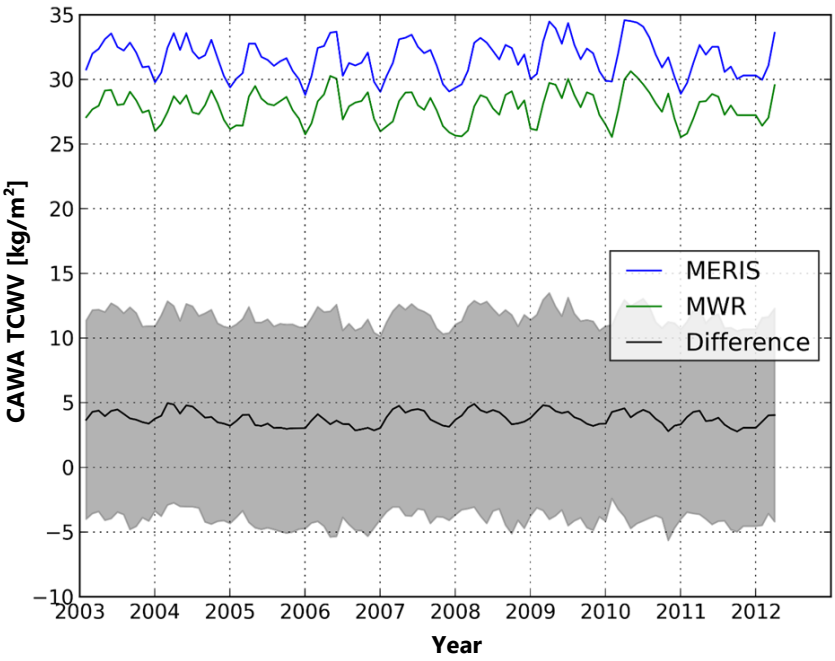

Figure 6: Monthly mean TCWV values [kg/m^2] for EMiR (green) and CAWA (blue), as well as their difference (black) and the corresponding standard deviation (grey).

2.2.3 Conclusions

The evaluation has revealed a wet bias of the MERIS CAWA TCWV on the order of +4 kg/m^2 as compared to the EMiR TCWV. Since a similar bias has been observed when comparing CAWA TCWV against radiosondes from the *Analyzed Radio Soundings Archive* (ARSA), we assume that CAWA systematically overestimates TCWV, which the developers have tentatively linked to deficiencies in the CAWA ocean forward operator (personal communication).

Still, both data sets agree well in their temporal characteristics. No obvious trend in global TCWV is observed in either dataset, which confirms the high stability of EMiR TCWV observed in the other validation exercises. The RMSD of CAWA is reduced as compared to EMiR (2.91 kg/m^2 *vs.* 3.95 kg/m^2), which is likely due to the sharper collocation criteria of the two datasets.

2.3 Comparison against AIRWAVE

2.3.1 Method

[The comparison of EMiR against AIRWAVE TCWV has been established by S. Casadio (SERCO/ESA) outside the frame of the EMiR project. It is included herein to complement the EMiR validation activities.]

The *Advanced Infra-Red Water Vapour Estimator approach* (AIRWAVE) provides TCWV retrievals from observations of the two the *thermal infrared* (TIR) channels of the *Along Track Scanning Radiometer* (ATSR) instrument series on board ESA's ERS-1, ERS-2, and Envisat satellites *[Casadio et al., 2016]*.

The use of the dual view capability of the ATSR-type instruments allows for accurate and precise, day-time and night-time retrievals of TCWV over cloud free oceans. The impact from sea surface

temperature and atmospheric aerosols is minimised through the combination of nadir and forward view retrievals, applying relative factors estimated through radiative transfer simulations.

The estimated accuracy of the AIRWAVE TCWV ranges from about 4% (AATSR and ATSR-2) to 15% (ATSR-1) when compared to independently derived SSM/I and ERA-Interim TCWV, i.e. when aggregated to a significantly coarser resolution. The estimated precision at the native AIRWAVE resolution of 1×1 km^2 ranges from ca. 12% for ATSR-2 and AATSR to ca. 30% for ATSR-1.

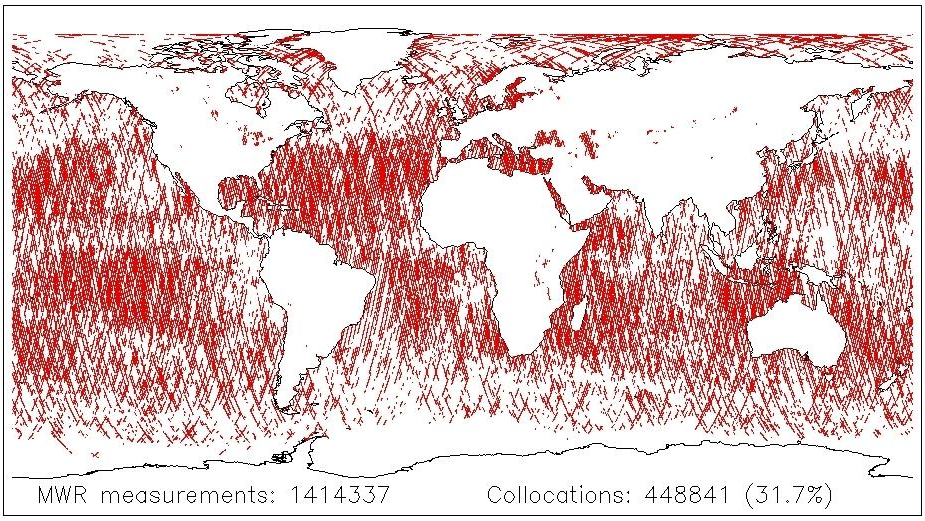

Figure 7: AIRWAVE and EMiR collocations for the month of August 2002 (Envisat). AIRWAVE TCWV retrievals are available for ca. 450.000 of the more than 1.4 Mio EMiR TCWV retrievals.

AIRWAVE TCWV has a good potential for the evaluation of EMiR TCWV retrievals:

- AIRWAVE is based on an independent measurement principle (TIR vs. microwave).
- AIRWAVE retrievals are available concomitantly to the entire EMiR TCWV time, resulting in a high number of match-ups (Figure 7).

Here, we compare AIRWAVE V1.0[7] TCWV against EMiR Level 2 TCWV retrievals from individual orbits. The following criteria have been applied to identify suitable collocations:

- Value of the EMiR TCWV retrieval cost function ≤ 1, i.e. a good signal-to-noise ratio for EMiR retrievals is required.
- Spatial coverage ≥ 20%, i.e. MWR field-of-view sufficiently well covered by (A)ATSR cloud free data.
- Standard deviation of AIRWAVE TCWV ≤ 50%, i.e. AIRWAVE retrievals less likely impacted by cloud-masking errors.
- Considered geographical areas limited to the latitude range 50°S - 50°N, i.e. no retrievals from areas with known AIRWAVE dry bias *[Casadio et al., 2016]*.

[7] A revised version 2.0 of AIRWAVE is currently under preparation that removes some of its known deficiencies.

As AIRWAVE only provides TCWV retrievals under cloud-free conditions (but day and night), the actual number of match-ups is reduced, but still very substantial. Applying the above criteria results in ca. 20 Mio. collocations for Envisat, 13 Mio. collocations for ERS-2, and 2.3 Mio collocations for ERS-1. Note that the applied geographic limitation removes a lot of dry (polar) cases from the intercomparison, and shifts upward the average TCWV of the considered collocations. The comparison of EMiR against AIRWAVE TCWV shown herein is thus not fully conclusive for dry atmospheres.

The actual collocation procedure is depicted in Figure 8.

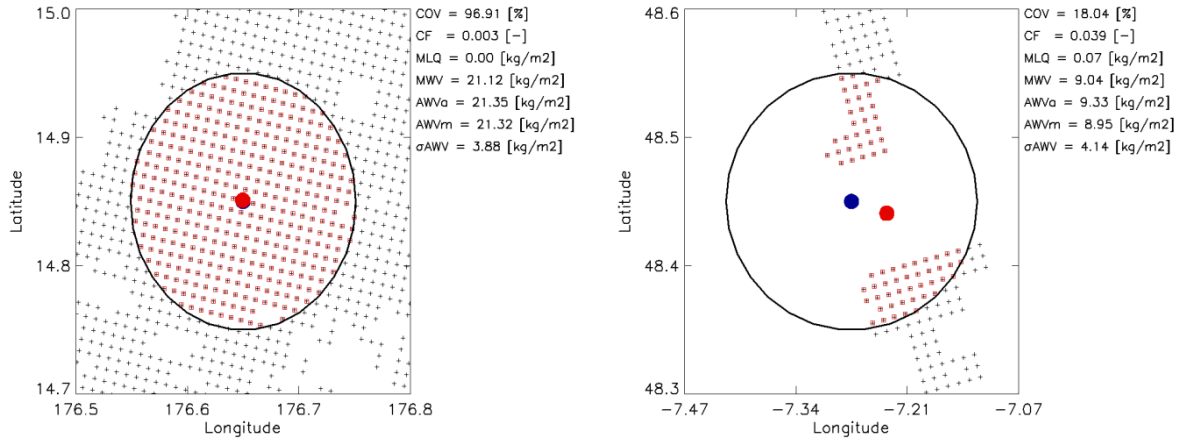

Figure 8: Collocation of AIRWAVE with EMiR TCWV retrievals for one site in the Central Pacific (left) and one in the Celtic Sea (right). The solid black circle represents the EMiR footprint, while the small red dots indicate the AIRWAVE retrievals used for the comparison. The large dots indicate the effective position of the EMiR (blue) and the effective AIRWAVE (red) retrievals.

Further explanations:
COV: percentage of MWR pixel covered by (A)TSR data; CF: EMIR TCWV cost function; MLQ: EMiR LWP [kg/m²]; MWV: EMIR TCWV [kg/m²]; AWVa: AIRWAVE TCWV average [kg/m²]; AWVm: AIRWAVE TCWV median [kg/m²]; σAWV: AIRWAVE TCWV standard deviation [kg/m²].

2.3.2 Results

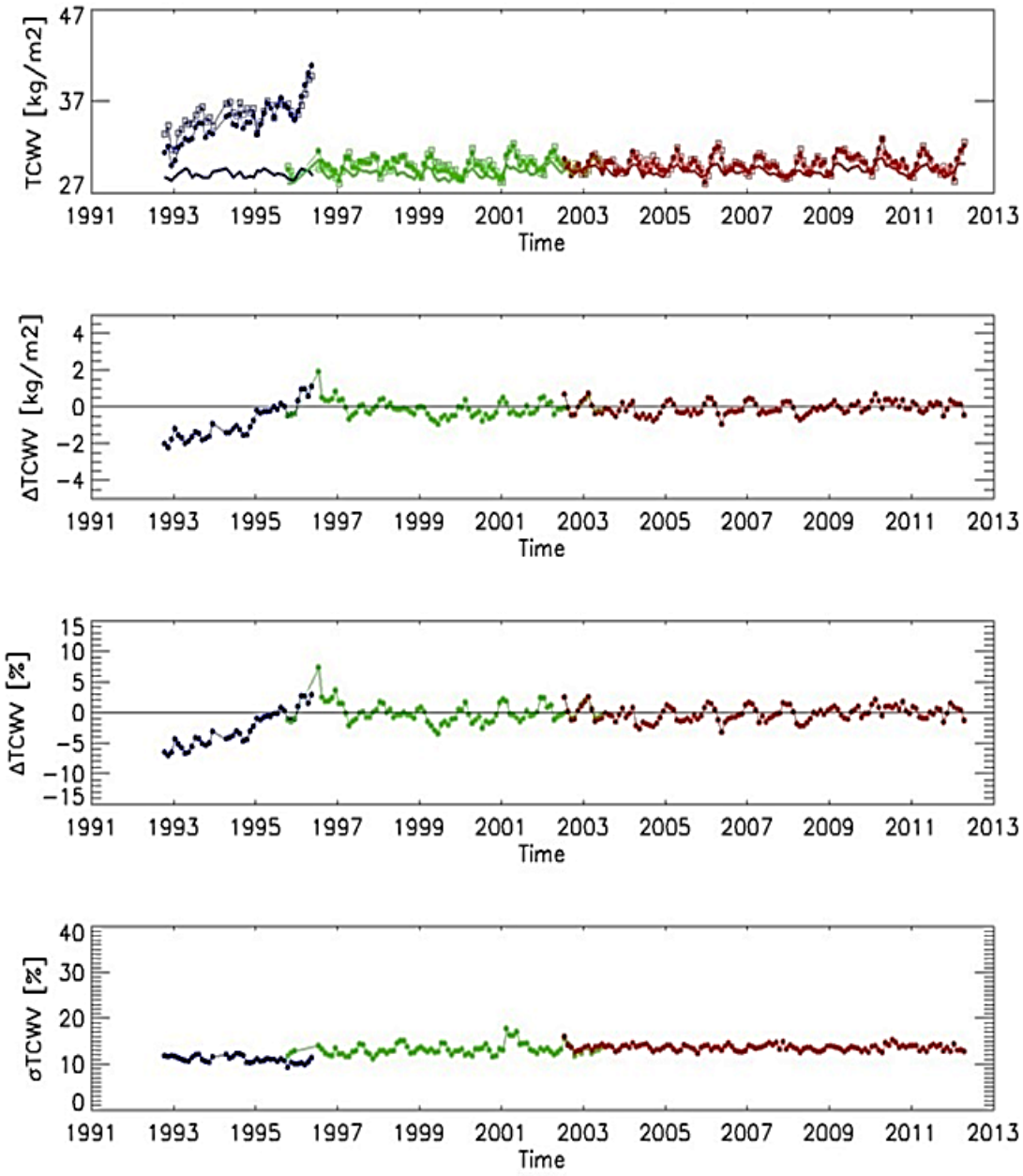

Figure 9: Comparison of AIRWAVE against EMiR TCWV monthly averages. From top to bottom: Absolute TCWV values, absolute and relative differences (EMiR-AIRWAVE), standard deviation. Blue: ERS-1; green: ERS-2; red: Envisat. Solid line: EMiR (all data); filled circle: EMiR (collocated data); square: AIRWAVE (collocated data).

Time series of collocated EMiR and AIRWAVE TCWV monthly mean values are shown in Figure 9. The agreement is very good for ERS-2 and Envisat, both in terms of absolute TCWV values as well as absolute and relative deviations, and no obvious long-term trends are observed. The apparent overestimation of the EMiR TCWV in 1996 is most likely due to instrumental issues of the MWR 23.8 GHz channel (see Table 2).

The situation is more complex for ERS-1: both AIRWAVE and EMiR TCWV show a significant temporal increase for the collocated data. We assume a temporally changing geographic distribution of the collocations as the main reason behind the observed increase. Since the average of all (i.e. not limited to the collocations) EMiR TCWV values from ERS-1 (top row, blue solid line) is temporally stable, an additional effect must be responsible for residual trend in AIRWAVE ΔTCWV, which is likely related to the ATSR instrument on-board ERS-1. Table 6 summarises the comparison between EMiR and AIRWAVE TCWV by providing average and trend values, specified individually for ERS-1, ERS-2, and Envisat.

Table 6: Comparison of EMiR against AIRWAVE TCWV retrievals. The trend values shown in red indicate problematic values, likely caused by trends in the geographical distribution of the collocations.

Parameter and mission	Average	Trend
	TCWV [kg/m²]	[% per year]
EMiR ERS-1 (all)	*29.04±0.39*	*0.22±0.19*
EMiR ERS-1 (collocations)	34.78±2.29	5.20±0.47
AIRWAVE ERS-1 (collocations)	35.63±1.63	3.00±0.46
EMiR ERS-2 (all)	*29.26±0.54*	*0.30±0.09*
EMiR ERS-2 (collocations)	29.98±0.82	0.15±0.14
AIRWAVE ERS-2 (collocations)	30.08±0.91	0.24±0.15
EMiR Envisat (all)	*29.48±0.46*	*0.05±0.05*
EMiR Envisat (collocations)	30.15±0.88	0.16±0.09
AIRWAVE Envisat (collocations)	30.23±0.94	0.08±0.10

2.3.3 Conclusions

Careful quality screening of data used for any comparison is mandatory to arrive at meaningful conclusions. However, it should be kept in mind that such screening may introduce biases in absolute values and trends.

The analysis presented herein has shown that AIRWAVE V1.0 and EMiR TCWV values agree well for ERS-2 and Envisat observations, neither significant biases nor trends are observed.

In the case of the ERS-1, some peculiarities are observed which are likely caused by a combination of unreliable cloud screening (AIRWAVE), temporal trends in the geographical distribution of the collocations (AIRWAVE and EMiR), and issues of the ATSR-1 instrument (AIRWAVE).

Working practically with the EMiR data product has led to a number of suggestions for further enhancing its utility and user friendliness:

- Provide information on the actual satellite orbit in the EMiR Level 2 data product to facilitate test case analysis and quality checks.
- Provide a day-night flag in the EMiR Level 2 data product, e.g. to allow automated comparison of day time *vs.* night time retrievals.

- Provide pixel-based TCWV uncertainty estimates in the EMiR Level-2 data product, e.g. to allow uncertainty-based pre-selection of data for subsequent comparisons.
- Provide information on the TCWV 1D-VAR prior in the EMiR Level 2 data product to allow for an evaluation of its impact on the final results, especially in "complex" situations characterised by high cost function values.
- Provide a cloud flag in the EMiR Level 2 data product, e.g. to support validation studies involving "clear sky" TCWV datasets.

2.4 Comparison against EC-Earth

2.4.1 Method

Climate models can be used as another source for evaluating EO-derived climate data records. Here, we report on the comparison of EMiR brightness temperature and TCWV time series against simulations obtained from the global climate model EC-Earth *[Hazeleger et al. 2010]*, run in an *Atmospheric Model Intercomparison Project* (AMIP) configuration with the lower boundary constrained by sea surface temperature and sea ice fields prescribed from observations. The atmospheric part of EC-Earth is based on the *Integrated Forecast System* (IFS) of the *European Centre for Medium Range Weather Forecasts* (ECMWF). The latest EC-Earth version *3* (ECEv3) based on IFS cycle cy36r4 has been used for the comparison presented herein. It was run at 1°×1° horizontal resolution with 91 vertical levels.

The full archive of EMiR Level-3 monthly fields of brightness temperatures at 23.8 GHz (Tb1) and 36.5 GHz (Tb2) and TCWV has been compared against ECEv3 simulations. While TCWV is directly provided by ECEv3, the brightness temperatures had to be calculated off-line using ECEv3 model results[8] as input to the code *Radiative transfer for TOVS* (RTTOV) in version 11.3[9]. To facilitate the comparison, simulation output was resampled onto the EMiR 2°×2° latitude-longitude grid using bilinear interpolation.

The evaluation is based on the comparison of maps and time series of mean values, biases and the *standard deviation of the difference* (SDD). To study annual variability and to detect potential jumps in the satellite data, deseasonalised anomaly time series have been derived by subtracting the 20 year mean values from each month of the year.

2.4.2 Results

EMiR and ECEv3 mean TCWV and brightness temperatures are shown in Figure 10 for the entire EMiR period 1992-2012. The climatologies visually agree well on a global scale, with large values over the tropical oceans and low values towards higher latitudes. The difference plots ECEv3- EMiR in Figure 10

[8] 2D fields: sea ice, surface pressure, skin temperature, 2 m temperature, total cloud cover, TCWV, longitudinal and latitudinal wind at 10 m; 3D fields: air temperature, water vapour mixing ratio, and cloud LWP.

[9] http://nwpsaf.eu/site/software/rttov/rttov-v11/ (accessed 2016/11/21)

show that ECEv3 is wetter (ca. +3 kg/m^2) and warmer (Tb1: ca. +3 K; Tb2: ca. +2 K) along the *intertropical convergence zone* (ITCZ) and off the East coast of Africa as compared to EMiR. Marked TCWV differences of ca. 2 kg/m^2 are also observed for the North Pacific and the Arctic Ocean, while Tb1 and Tb2 differences are less pronounced in these areas. The differences in the tropics are likely caused by EC-Earth, since other CMIP5 models do not exhibit this pattern (not shown). In contrast, the high latitude northern differences appear to be an EMiR issue since they are more pronounced for the derived TCWV as for the brightness temperatures.

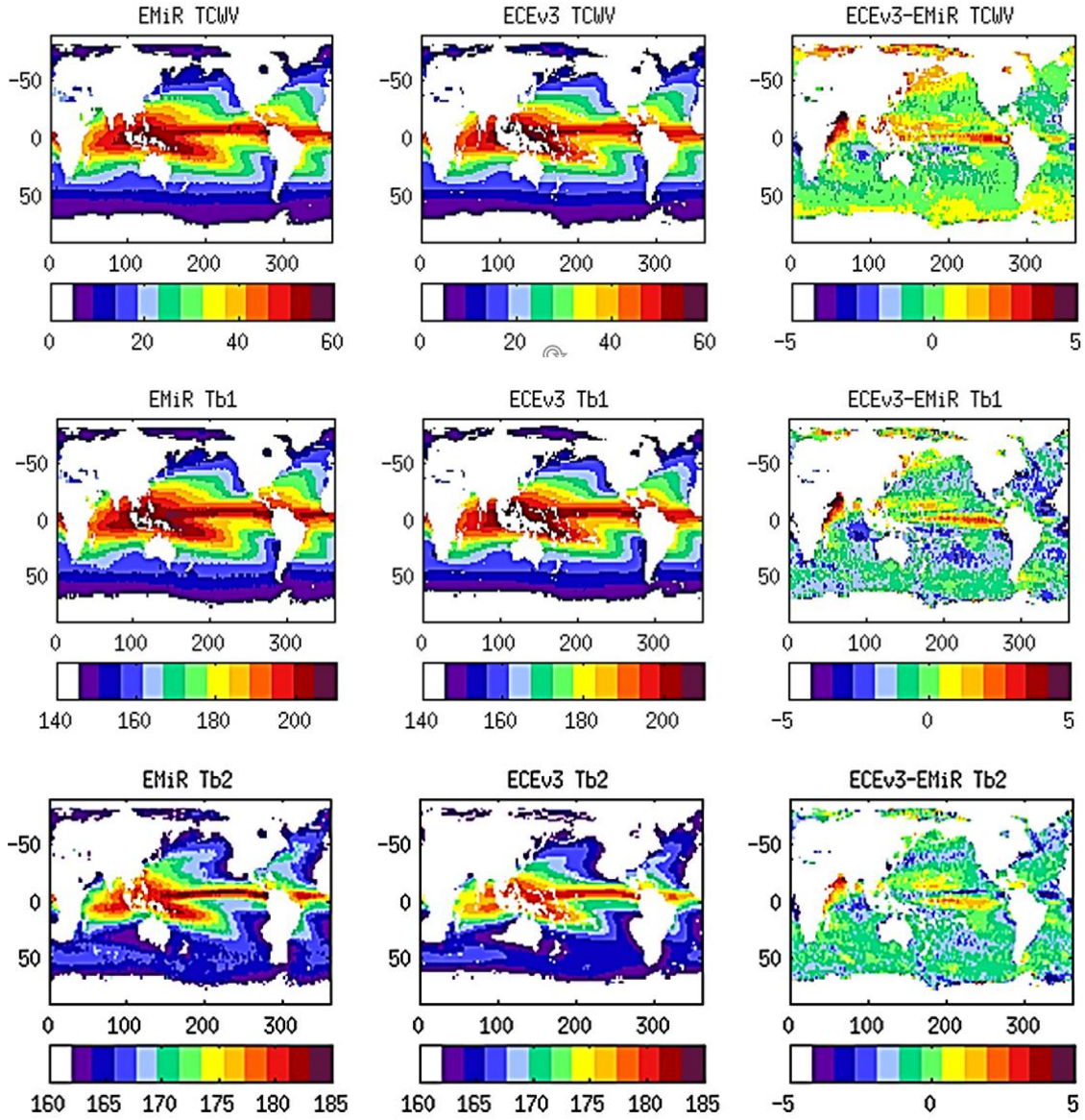

Figure 10: Mean TCWV and brightness temperatures for EMiR and ECEv3 for the period 1992-2012. Top to bottom: TCWV [kg/m^2], Tb1 [K], Tb2 [K]. Left to right: EMiR, ECEv3, ECEv3-EMiR.

Time series of the monthly mean values spatially averaged between 60° S and 60° N (Figure 11, left column) show that the ECEv3 (and ERA-Interim) long-term TCWV mean is close to the corresponding EMiR value, but that the models exhibit larger annual amplitudes. The reasons for this are not clear at

this point. Systematic seasonal deviations are observed for the brightness temperatures, with EMiR brightness temperatures being globally ca. 1-2 K warmer in the winter months than the corresponding temperatures from ECEv3. An interesting feature is observed for Tb2 (Figure 11, bottom left): While EMiR values are systematically higher than ECEv3 values before the year 2003, they agree well afterwards. This peculiarity temporally coincides with EMiR switching from ERS-2 to Envisat and may hint to some remaining intercalibration issues of the EMiR Tb2 time series.

The deseasonalised anomaly time series (Figure 11, right column) show that both EMiR and the models are well able to capture the global variability, showing e.g. large positive anomalies for the strong El Niño events in 1998 and 2010. A strong anomaly only observed for EMiR Tb1 in 1996, translating into an equivalent EMiR TCWV anomaly, is most likely being caused by a specific issue ("gain drop") of the MWR instrument on-board ERS-2 (Table 2).

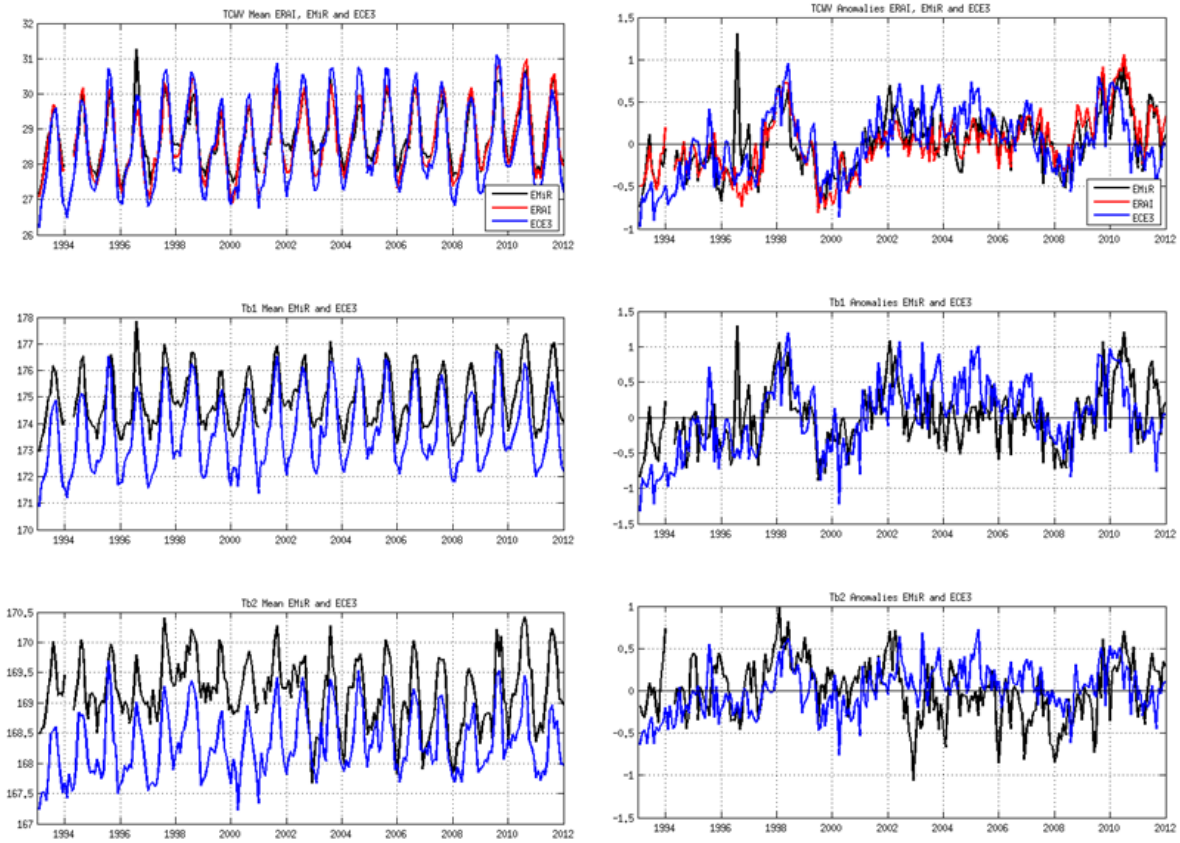

Figure 11: Time series of monthly mean values (left column) and anomalies (right column). Top to bottom: TCWV, Tb1, and Tb2 for EMiR (black), EC-Earth v3 (blue), and ERA-Interim (red). Time series have been spatially averaged for the area 60° S to 60° N.

Time series of the SDD between EMIR and model-based TCWV values are shown in Figure 12. As expected, ERA-Interim, which assimilates observations, is closer to EMiR with a mean SDD of about 2.2 kg/m^2, while ECEv3 shows a mean SDD of about 3.2 kg/m^2. A number of anomalies can be observed in the SDD time series, most prominently between 01/1998 and 04/1998. In cases where such anomalies

coincide with MWR known instrumental issues, e.g. in 12/1993, 12/1994, 02/2001, or 03/2002 (see Table 2, Table 3, Table 4), they are possibly triggered by anomalies in the EMiR TCWV time series.

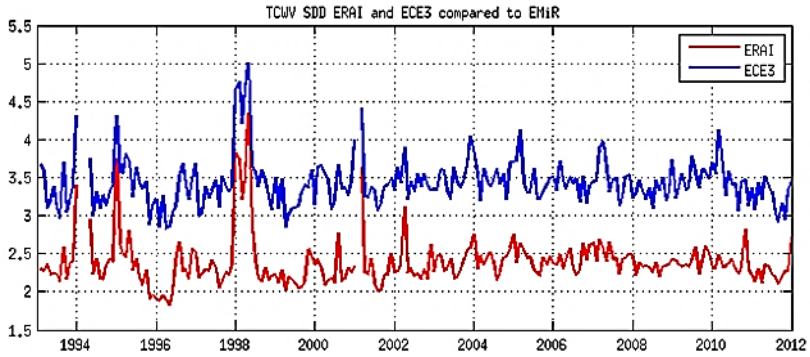

Figure 12: Time series of the TCWV SSD for ECEv3 (blue) and ERA-Interim (red) as compared to EMiR. Spatial averaging between 60° S and 60° N.

2.4.3 Conclusions

The comparison of EMiR data products against EC-Earth and ERA-Interim leads to the following general conclusions:

- The EMiR L3 brightness temperatures are consistent over time as compared to EC-Earth. Small shifts and trends are possibly linked to the intercalibration of the different MWRs.
- More specifically, a shift of ca. 1.0 K is observed at 36.5 GHz between the EMiR and EC-Earth brightness temperatures at the transition from ERS-2 to Envisat (Figure 11, bottom left).
- The EMiR L3 TCWV climatology is consistent over time and agrees well with both EC-Earth and ERA-Interim and therefore appears suitable for climate model evaluation.

There are some potentially questionable EMiR TCWV values in temporal vicinity to data gaps. Prominently flagging the concerned periods (e.g. on the DOI landing page) would be beneficial for users.

2.5 Participation to G-VAP

2.5.1 Method

To date, a large variety of satellite based water vapour data records are available (see e.g. http://gewex-vap.org/?page_id=309 or http://ecv-inventory.com). Without proper background information and understanding of their limitations, these data may be incorrectly utilised or misinterpreted. In this respect, the *GEWEX Data and Assessments Panel* (GDAP)[10] has initiated the *GEWEX Water Vapor Assessment* (G-VAP)[11] to quantify the current state of the art in water vapour

[10] GDAP: http://www.gewex.org/panels/gewex-data-and-assessments-panel/ (accessed 2017/01/25)

[11] G-VAP: http://gewex-vap.org/ (accessed 2017/01/25)

products being constructed for climate applications and to support the selection process of water vapour products by GDAP for its production of globally consistent water and energy cycle products.

The EMiR TCWV data record has been submitted to G-VAP and constitutes one of the 22 participating *"short-term"* (common period 01/2003 to 12/2008) data records, of which sixteen are satellite- and six reanalysis-based. These 22 data records have been analysed on the basis of monthly mean values on a regular latitude/longitude grid of two degrees' resolution.

The intercomparison methods include:

- Bias and standard deviations relative to the ensemble means.
- Weather type analysis (*"clear sky"*, *"cloudy sky"*, *"all sky"*, see below).

Further information on the G-VAP intercomparison of TCWV data records will become available through the G-VAP Report currently under review by GDAP *[Schroeder et al., 2016]*.

2.5.2 Results

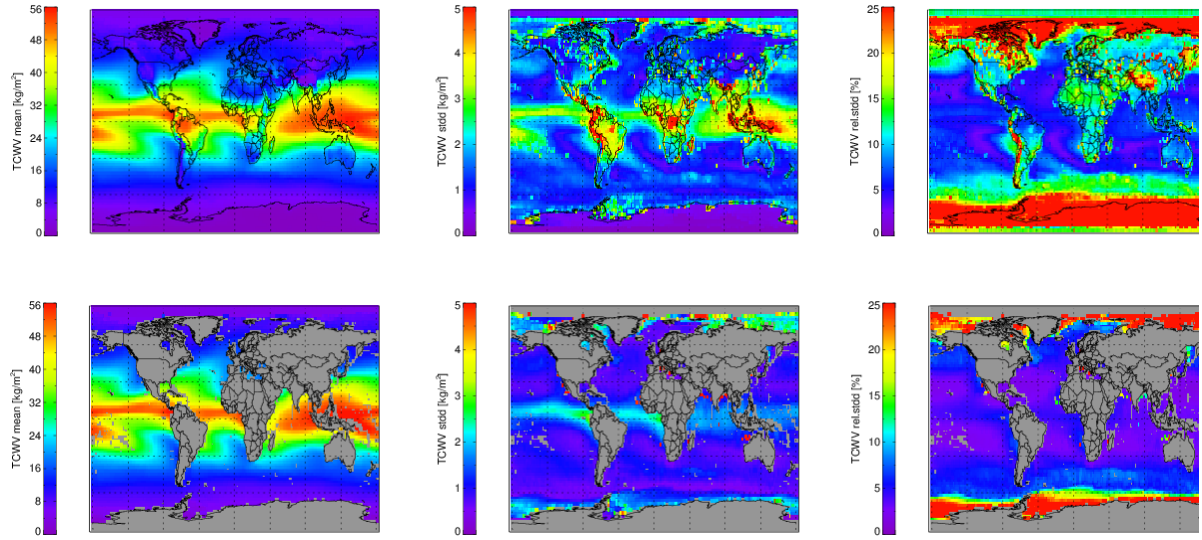

Figure 13: TCWV ensemble mean (left), absolute (middle) and relative (right) standard deviation for all 22 available data sets (top), as well as the 11 microwave-derived cloudy sky data sets (bottom). Note that the number of available data records differs regionally.

Figure 13 (top) shows the TCWV ensemble mean as well as the corresponding absolute and relative standard deviations of all 22 short-term TCWV data records. As not all data records provide global coverage, the number of available data records differs regionally. High absolute TCWV standard deviations are mostly found in areas of high TCWV, while high relative standard deviations are bound to dry atmospheres mostly occurring in mountainous areas and polar regions.

Figure 13 (bottom) depicts the same parameters, but limited to the cloudy sky data records, consisting of satellite-based products partly or entirely based on passive microwave observations with retrieval schemes limited to ice-free ocean areas. These cloudy sky data records agree generally well, both

among themselves and with the ensemble mean, except for the polar (ocean) regions where their (internal) relative standard deviation exceeds 25 %.

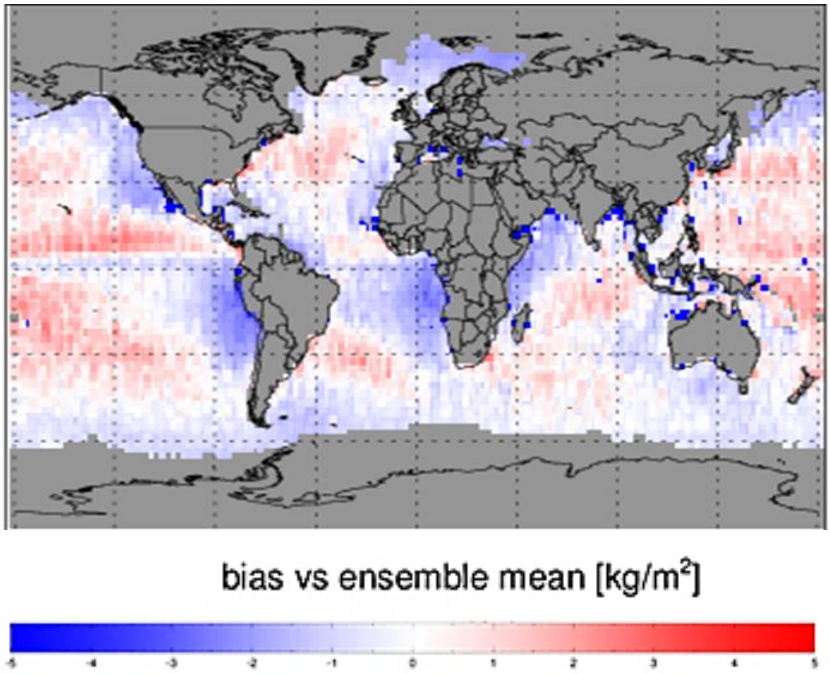

Figure 14: EMiR TCWV bias relative to the ensemble mean for all 22 G-VAP "short-term" ensemble members.

The bias of EMiR TCWV against the mean of all 22 short-term ensemble members is depicted in Figure 14. Generally, the deviations from the ensemble are limited to the range from -2.0 kg/m^2 to 2.0 kg/m^2. Areas of negative EMiR TCWV bias are located in the ocean upwelling regions off the Eastern continental margins, especially west of North- and South-America and Southern Africa. These regions are characterised by frequent atmospheric inversion occurrence, which is problematic for certain TCWV retrieval approaches, especially those based on infrared sounding. EMiR TCWV may thus be closer to the truth than the ensemble mean in these areas.

Close to the coast, there are a number of areas where EMiR TCWV shows strong negative bias. The reasons for this behaviour are not fully clear at this point, however, insufficient side lobe correction of the underlying Level-1B MWR data could be one reason for the observed deviations. Underestimation may party also be related to high aerosol loads *[Ge et al., 2008]*.

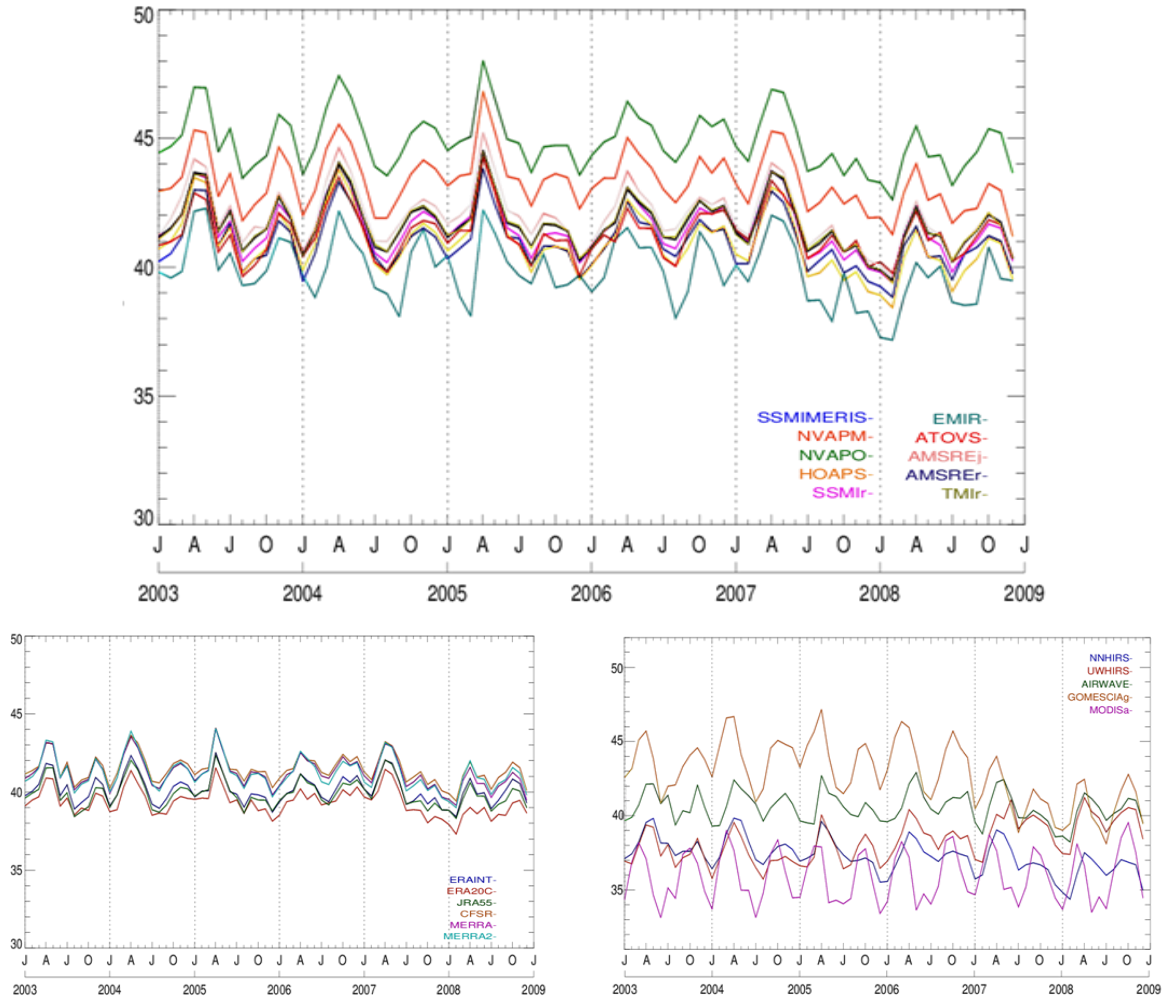

Figure 15: Time series (01/2003 – 12/2008) of the mean TCWV over the ocean for the tropics [20°S to 20°N]. Top: cloudy sky; bottom left: all sky; bottom right: clear sky.

The temporal evolution of the monthly mean TCWV over the tropics (20°S to 20°N) is shown in Figure 15 for each individual ensemble member. While EMiR TCWV shows on average the lowest TCWV values among the cloudy sky data records, it matches the all sky (i.e. re-analyses) data records and fits well into the range covered by the clear sky (VIS/NIR/TIR based) data records.

2.5.3 Conclusions

EMiR has participated to the G-VAP intercomparison of 22 short-term TCWV data records. Generally, EMiR TCWV is very close to the ensemble mean, both with respect to its geographical and temporal distribution. While this does not provide any direct assessment on EMiR's overall quality (the ensemble mean does not necessarily represent the truth), it further fosters confidence into the general layout of the EMiR brightness temperature intercalibration and TCWV retrieval schemes. Some of the observed EMiR biases may point to underlying problems and should be investigated in more detail prior to a potential reprocessing of the EMiR data record.

2.6 Evaluation of the EMiR wet tropospheric correction

2.6.1 Method

A *wet tropospheric correction* (WTC) product has been derived in the context of EMiR by combining the EMiR TCWV product with atmospheric profiles from the ERA-Interim reanalyses (see section 1.4.3). Details on the method applied can be found in *[Bennartz et al., 2016]*.

The evaluation of the EMiR WTC makes use of legacy work from the ESA Sea Level *Climate Change Initiative* (CCI)[12], where several methods have been implemented for evaluating the quality of WTC products. In the present report, we focus on the SSH cross-over approach, i.e. we analyse SSH differences between ascending and descending nodes with a maximum temporal difference of 10 days. Assuming that SSH changes are negligible over such time span *[Legeais et al., 2014]*, the best WTC will then result in the smallest SSH variance at cross-overs.[13]

In the following, the SSH computed with the EMiR WTC (herein always referred to as *"study"*) is compared to different *"reference"* SSHs based on other WTC approaches. An improvement of the altimetry system performance is observed when $\Delta VAR_{SSH} \equiv VAR_{SSH}(study) - VAR_{SSH}(reference)$ is negative, i.e. if the study SSH shows lower variance than the reference SSH.

2.6.2 Results

A first assessment of the EMiR WTC was done by comparing it to a WTC product derived from ERA Interim humidity and temperature profiles by applying linear interpolation between the two temporally closest analyses (which are provided in 6 hour intervals) and bilinear spatial interpolation of the gridded WTC onto the altimeter track. Overall, the EMiR WTC shows improved performance over the entire EMiR period (1992-2012) by $\Delta VAR_{SSH} \approx -1.5\ cm^2$ as compared to the ERA Interim WTC (see Figure 16, left column). This is typical when comparing the WTC derived from a concomitantly operated radiometer against a product derived entirely from re-analyses.

The EMiR WTC was then compared against ESA's operational WTC. For ERS-1 and ERS-2, the latter is derived from a log-linear approach using both brightness temperature channels as well as the altimeter derived wind speed *[Eymard et al., 1996]*. In the case of Envisat, the updated v2.1b WTC has been used (as recommended by ESA), representing the state-of-the-art available to users. The differences between EMiR and ESA's WTC schemes are shown in Figure 16, right column: EMiR provides smaller values at higher latitudes (dry atmosphere) and larger values in the tropics and subtropics (humid atmosphere). Overall, the performance of the EMiR WTC is slightly inferior as compared to ESA's operational WTC. The average differences of the SSH variances at the cross-overs amount to $\Delta VAR_{SSH} \approx +0.09\ cm^2$ for ERS-1 (small deterioration), $\Delta VAR_{SSH} \approx -0.04\ cm^2$ for ERS-2 (very small improvement), and $\Delta VAR_{SSH} \approx +0.30\ cm^2$ (small, but noticeable deterioration) for Envisat.

[12] http://www.esa-sealevel-cci.org/

[13] Obviously, the cross-over method does not provide information on a potential TCWV bias.

Comparing study against reference WTC over the entire EMiR period, a number of interesting patterns can be observed:

- **ERS-1**: A significant underperformance (ΔVAR_{SSH} between +0.4 cm^2 and +1.0 cm^2) is observed for the EMiR WTC for most of the year 1994.
- **ERS-2**: The EMiR WTC performs significantly better in the first quarter of each year, most notably for the years 2001, 2002, and 2003 (ΔVAR_{SSH} between -1.0 cm^2 and -2.0 cm^2). ESA's operational MWR WTC provides slightly better results in summer and autumn.
- **Envisat**: While the EMiR WTC performance is slightly inferior for the entire observation period, the deviations seem to be larger at the beginning of the Envisat period (2002-2004).

The observed patterns need to be investigated in more detail and may be helpful in identifying deficiencies of either WTC method.

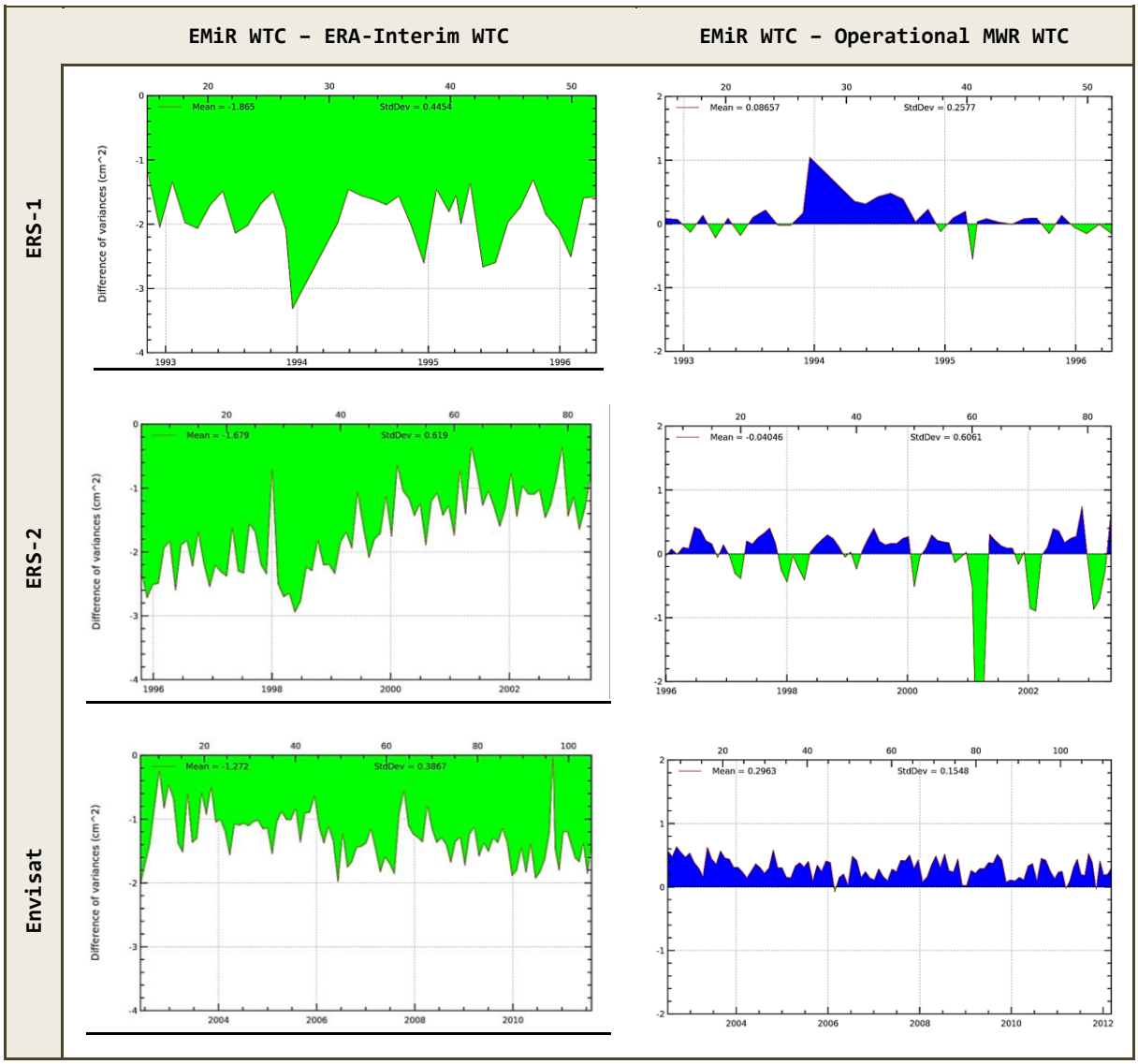

Figure 16: Left: Temporal course of the spatially averaged SSH variance differences (ΔVAR_{SSH}) between EMiR and ERA Interim WTC schemes for ERS-1 (top), ERS-2 (middle), and Envisat (bottom). Right: Same as left, but EMiR WTC compared to ESA's operational WTC.

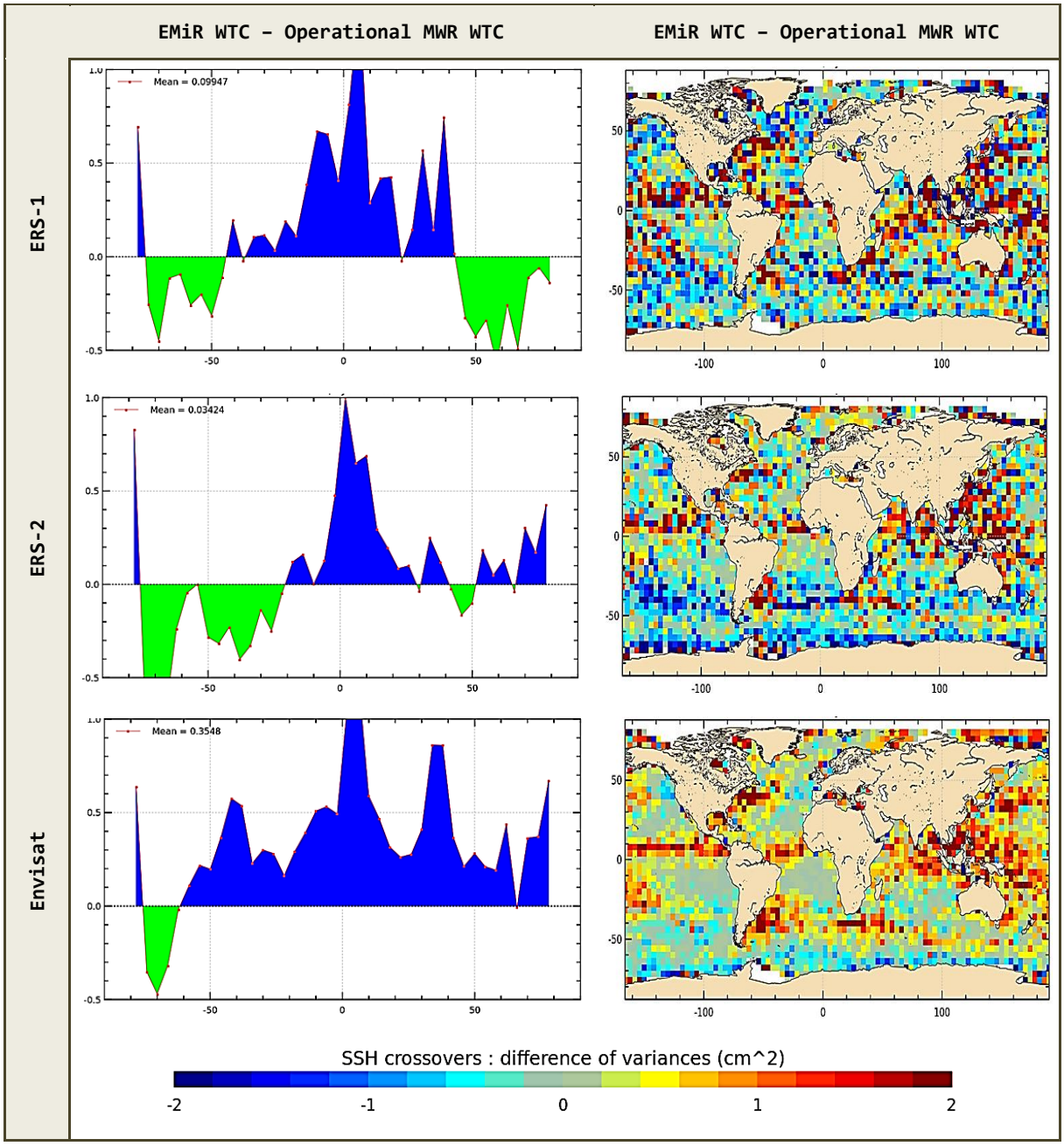

Figure 17. Left: Latitudinal distribution of the temporally averaged SSH variance differences (ΔVAR_{SSH}) between EMiR and ESA's operational WTC schemes for ERS-1 (top), ERS-2 (middle), and Envisat (bottom). Right: Same as left, but at full spatial resolution.

Zonally, larger differences between EMiR and ESA's operational WTC schemes are observed (see Figure 17, left):

- **ERS-1**: EMIR WTC is slightly superior ($\Delta VAR_{SSH} \approx -0.3\ cm^2$) to ESA's operational WTC for latitudes beyond 50° N/S, and slightly inferior (ΔVAR_{SSH} between +0.3 cm^2 and +1.0 cm^2) for latitudes between 20° N and 20° S.

- **ERS-2**: EMiR WTC provides slightly superior results ($\Delta VAR_{SSH} \approx -0.35\ cm^2$) for latitudes between 20° S and 70° S. In contrast to that, a small deterioration (ΔVAR_{SSH} between +0.2 cm² and +1.0 cm²) is observed for latitudes between 10° S and 20° N.
- **Envisat**: EMiR WTC provides slightly superior results ($\Delta VAR_{SSH} \approx -0.3\ cm^2$) for latitudes between 55° S and 65° S. A small deterioration (ΔVAR_{SSH} between +0.2 cm² and +1.0 cm²) is observed at all other latitudes.

Global maps of the SSH variance differences between EMiR WTC and ESA's operational WTC may provide hints on the reasons for the observed differences (see Figure 17, right column). In the case of Envisat, for example, the better performance of the EMiR WTC south of ca. 50° S is clearly visible, while the operational WTC provides better results in most other areas, especially in the ITCZ. Some of the observed differences may relate to upwelling regions (e.g. Oregon upwelling) or large surface currents (e.g. Gulf Stream, Brazil Current). The picture is not so clear for ERS-1 and ERS-2 where the differences between the two WTC methods are generally smaller. Interestingly, local differences between the two methods are significantly larger for ERS-1 and ERS-2, where ΔVAR_{SSH} between neighbouring grid cells may reach values of up to 3.0 cm².

In order to better grasp the relevance of the above-mentioned differences, note that the improvement expected for the next version v3.0 of the Envisat operational MWR WTC is assumed to be on the order of $\Delta VAR_{SSH} \approx -0.5\ cm^2$.

2.6.3 Conclusions

EMiR paves the way to developing a promising alternative method for WTC: The EMiR WTC has proven superior to an ERA Interim derived WTC, and is almost at equal with the performance of ESA's operational MWR WTC product.

The 1D-VAR approach applied for TCWV retrieval may eventually solve some of the issues encountered by current operational algorithms when dealing with non-average atmospheric situations, such as e.g. temperature inversions over ocean upwelling regions.

Cloudy situations are an issue for any retrieval approach. It would thus be interesting to sharpen the comparison between the EMiR WTC and the operational radiometer WTC by distinguishing between cloudy and non-cloudy situations. Such comparison would be supported by a dedicated cloud mask integrated into the EMiR product.

The accuracy of the EMiR WTC may be further improved by a number of measures, most importantly to make use of information on sea surface roughness which is available from the concomitantly operated altimeters on board ERS-1, ERS-2, and Envisat.

3 Suggested improvements

A number of measures have been proposed to improve the EMiR data record (see Table 7) in terms of user friendliness and product quality. Implementing such measures would be highly beneficial, especially when considering the extension of the EMiR data record into the Sentinel-3 era *[EMiR DLV-EXT-09, 2017]*.

Table 7: Suggested improvements to the EMiR data record. List alphabetically ordered by the ID.

No.	ID	Type	Suggestion
1	PE_Anomaly	Product enhancement	Further investigate the observed anomalies and systematic differences of EMiR Tb and TCWV as compared to the references and aim at identifying and removing their causes, in case they are EMiR-related.
2	PE_CloudMask	Product enhancement	Provide a cloud flag in the EMiR Level 2 data product, e.g. to support validation studies involving "clear sky" TCWV datasets.
3	PE_DayNight	Product enhancement	Provide a day-night flag in the EMiR L2 data product, e.g. to allow automated comparison of day *vs.* night time retrievals.
4	PE_FillGaps	Product enhancement	Fill the temporal gaps in the EMiR data record, especially those lasting for one week or more (see Table 4). This will require the provision of updated L1B data records, which are the basis for EMiR processing.
5	PE_FillOrb	Product enhancement	It has been reported that ERS-1 / ERS-2 descending orbits are frequently incomplete. If confirmed and correctable, implement remedial action.
6	PE_Orbit	Product enhancement	Provide information on the satellite orbit in the EMiR L2 data product, e.g. to facilitate test case analysis and quality checks.
7	PE_PrecFlag	Product enhancement	Devise and implement improved flagging of meteorological situations characterised by heavy precipitation.
8	PE_Prior	Product enhancement	Provide information on the TCWV prior in the EMiR L2 data product, e.g. to allow for an evaluation of its impact on the final results, especially in “complex” situations characterised by high cost function values.
9	PE_Sigma0	Product enhancement	Make synergetic use of the sigma_0 information from the radar altimeters to better characterise the state of the sea surface and to such improve TCWV retrieval accuracy.
10	PE_SLCorr	Product enhancement	Reduce land contamination in coastal areas. This will require the implementation of an improved method for the side-lobe correction of MWR observations in the underlying L1B processing.
11	PE_Uncert	Product enhancement	Provide pixel based TCWV uncertainty estimates in the EMiR L2 data product, e.g. to support the selection of EMiR subsets for sensitivity analyses.

No.	ID	Type	Suggestion
12	UF_Gaps	User friendliness	Prominently inform users about the data gaps and the potentially associated product quality issues. This could be achieved by providing such information on the DOI landing page.
13	UF_SWOT	User friendliness	Communicate strengths and weaknesses of the EMiR data record more clearly to interested users. This could be achieved by providing such information on the DOI landing page.

4 Summary and conclusions

A new data record (termed EMiR) of the total column water vapour (TCWV) and derived wet tropospheric correction (WTC) was generated from observations of the Microwave Radiometer (MWR) instruments flown on board the satellites ERS-1, ERS-2, and Envisat using optimal interpolation retrieval techniques. To achieve consistent time series from these three instruments, a new method for intercalibrating the brightness temperatures from succeeding MWRs was developed and successfully applied.

The EMiR data record covers the period from 1992/10/23 to 2012/04/08. While the end of the EMiR data record coincides with the loss of Envisat and the related end of MWR observations, EMiR does not cover the first ca. 15 months of ERS-1 operations due to the lack of specifically pre-processed MWR L1B brightness temperatures used as input to the EMiR processing. A number of additional gaps occur, most of them lasting between one and ten days. Despite these gaps, EMiR data availability is very good after 1994/04/10.

EMiR products have been compared against ground-based (GNSS) satellite-based (different approaches), and climate model based (EC Earth) reference data (Table 8):

- EMiR brightness temperatures were validated against EC-Earth simulations. As expected, some channel dependent biases were observed[14], but no temporal trends.
- EMiR TCWV compares well with the reference data, with absolute bias values mostly smaller reference data sets ranging from than 0.5 kg/m^2, and statistically insignificant temporal trends. Seemingly larger deviations against AIRWAVE TCWV from ERS-1 and MERIS CAWA TCWV from Envisat can be attributed to retrieval issues for the latter and geographical sampling effects for the former.
- The wet tropospheric correction (WTC) derived from the EMiR TCWV is almost as accurate as is ESA's operational WTC, even though the latter incorporates additional information on the sea surface roughness, which has not yet been introduced into EMiR processing.

Table 8: Summary of EMiR validation. "ALL" refers to the sum of collocations from ERS-1, ERS-2, and Envisat (ENV).

Reference data	Number of collocations	TCWV bias (EMiR – REF) [kg/m²]	TCWV stability (EMiR – REF) [% per decade]	Remarks
GNSS	ALL: 18,065	ALL: -0.43	ALL: -0.38	Period: 1995-2012
MERIS CAWA	ENV: 2.6 Mio.	ENV: -3.8	[no apparent trend]	Period: 2003-2012 CAWA TCWV values likely too high.
AIRWAVE	ERS-1: 2.3 Mio	ERS-1: -0.85	ERS-1 n/a	Period: 1992-2012

[14] Due to the lack of an absolute reference, it is not possible to determine whether EMiR or EC-Earth brightness temperatures are closer to the truth.

Reference data	Number of collocations	TCWV bias (EMiR – REF) [kg/m²]	TCWV stability (EMiR – REF) [% per decade]	Remarks
	ERS-2: 13 Mio ENV: 20 Mio	ERS-2: -0.10 ENV: -0.08	ERS-2: -0.90 ENV: +0.80	ERS-1: trend artefacts likely caused by regionalised match-ups.
EC Earth	No match-up approach, comparison based on zonal means for the period 1993-2012. Generally good agreement between EMiR and ECEv3. Some questionable EMiR Tb and TCWV values have been identified in temporal vicinity to data gaps.			
G-VAP	No match-up approach, comparison based on zonal means for the period 2003-2008. EMiR TCWV is on the drier end of other microwave-based TCWV products, but is very close to the ensemble mean of all 22 investigated TCWV data records. Some questionable EMiR TCWV values have been identified along coastlines.			
WTC Intercomparison	Indirect validation approach by comparing WTC derived from EMiR TCWV with state-of-the-art operational WTC. Similar results are obtained for all instruments, indicating the already good quality of EMiR TCWV, especially for ERS-1 and ERS-2.			

Concluding, the EMiR data record is deemed mature and accurate enough to be used for climatological and oceanographic applications. A number of suggestions were made to further enhance the quality and user friendliness of the EMiR data record.

5 References

Casadio, S., E. Castelli, E. Papandrea, B. M. Dinelli, G. Pisacane, A. Burini, B. R. Bojkov (2016): Total column water vapour from along track scanning radiometer series using thermal infrared dual view ocean cloud free measurements: The Advanced Infra-Red WAter Vapour Estimator (AIRWAVE) algorithm, *Remote Sensing of Environment*, Vol. 172, pp 1-14. DOI: 10.1016/j.rse.2015.10.037.

EMiR DLV-EXT-07 (2016): EMiR Level-2 and -3 Product Generation, Algorithm Theoretical Basis Document, V2.11, 36 pp., available at http://esa-mwr.org/?attachment_id=324.

EMiR DLV-EXT-09 (2017): ERS/Envisat MWR Recalibration and Water Vapour FDR Generation (EMiR): Guidance for Sentinel-3, V1.02, 11 pp., available at http://esa-mwr.org/?attachment_id=323.

EMiR DLV-INT-12 (2016): EMiR wet tropospheric correction: performance analysis, V1.01, 20 pp., available at http://esa-mwr.org/?attachment_id=325.

Eymard, L., L. Tabary, E. Gérard, S.-A, Boukabara, and A. Le Cornec (1996): The Microwave Radiometer Aboard ERS-1: Part II-Validation of the Geophysical Products, *IEEE Transactions on Geoscience and Remote Sensing*, 34 (2). DOI: 10.1109/36.485108.

Fu, L.-L., and A. Cazenave, Editors (2000): Satellite Altimetry and Earth Sciences: A Handbook of Techniques and Applications, *International Geophysics*, Vol. 69, Academic Press, 463 pages. E-book ISBN: 9780080516585.

Ge J., J. Huang, F. Weng, and W. Sun (2008): Effects of dust storms on microwave radiation based on satellite observation and model simulation over the Taklamakan Desert. *Atmospheric Chemistry and Physics*, 8, 4903-4909. DOI: 10.5194/acp-8-4903-2008.

Gendt, G., G. Dick, C. Reigber, M. Y. L. Tomassini, and M. Ramatschi (2004): Near Real Time GPS Water Vapor Monitoring for Numerical Weather Prediction in Germany, *J. Meteorol. Soc. Jpn*., 82, 361–370. DOI: 10.2151/jmsj.2004.361.

Hazeleger, W., and 31 co-authors (2010): EC-Earth—a seamless earth system prediction approach in action. *Bull Am Meteorol Soc*, 91:1357–1363. DOI: 10.1175/2010BAMS2877.1.

Justice, C., A. Belward, J. Morisette, P. Lewis, J. Privette, and F. Baret (2000): Developments in the 'validation' of satellite sensor products for the study of land surface. *International Journal of Remote Sensing*, 21(17), 3383 – 3390. DOI: 10.1080/014311600750020000.

Legeais, J.-F., M. Ablain, and S. Thao (2014): Evaluation of wet troposphere path delays from atmospheric reanalyses and radiometers and their impact on the altimeter sea level, *Ocean Sci.*, 10, 893–905. DOI:10.5194/os-10-893-2014.

Lindstrot, R. (2012): Algorithm Theoretical Basis of the Level 2 MERIS total column water vapour product for the ESA DUE GlobVapour project, issue 3, revision 0, .15 pp. Available at: http://www.globvapour.info/download/GlobVapour_D16_ATBD_L2_MERIS_v3.0.pdf.

Ross, R.J., and W.P. Elliot (1996): Tropospheric Water Vapour Climatology and Trends over North America: 1973-93, *J. Climate*, 9, 3561-3574. DOI: 10.1175/1520-0442(1996)009<3561:TWVCAT>2.0.CO;2.

Schroeder, M., and 26 co-authors (2016): GEWEX water vapor assessment (G-VAP) - Final Report, 214 pp., DWD, Offenbach, Germany *[under review by GDAP]*.

Wang, J., and L. Zhang (2009): Climate applications of a global, 2-hourly atmospheric precipitable water dataset from IGS ground-based GPS measurements. *J. of Geodesy*, 83, 209-217. DOI: 10.1007/s00190-008-0238-5.

Wang, J., L. Zhang, A. Dai, T. Van Hove, and J. Van Baelen (2007): A near-global, 8-year, 2-hourly atmospheric precipitable water dataset from ground-based GPS measurements. *J. Geophys. Res.*, 112, D11107. DOI: 10.1029/2006JD007529.