Atmos. Meas. Tech., 10, 1387–1402, 2017 www.atmos-meas-tech.net/10/1387/2017/ doi:10.5194/amt-10-1387-2017 © Author(s) 2017. CC Attribution 3.0 License.





# An intercalibrated dataset of total column water vapour and wet tropospheric correction based on MWR on board ERS-1, ERS-2, and Envisat

Ralf Bennartz<sup>1,2</sup>, Heidrun Höschen<sup>3</sup>, Bruno Picard<sup>4</sup>, Marc Schröder<sup>3</sup>, Martin Stengel<sup>3</sup>, Oliver Sus<sup>3</sup>, Bojan Bojkov<sup>5,a</sup>, Stefano Casadio<sup>5</sup>, Hannes Diedrich<sup>6,b</sup>, Salomon Eliasson<sup>7</sup>, Frank Fell<sup>8</sup>, Jürgen Fischer<sup>6</sup>, Rainer Hollmann<sup>3</sup>, Rene Preusker<sup>6</sup>, and Ulrika Willén<sup>7</sup>

<sup>1</sup>Earth and Environmental Science Department, Vanderbilt University, Nashville, TN, USA
<sup>2</sup>Space Science and Engineering Center, University of Wisconsin–Madison, USA
<sup>3</sup>German Meteorological Service (DWD), Offenbach, Germany
<sup>4</sup>Collecte Localisation Satellite (CLS), Toulouse, France
<sup>5</sup>European Space Agency (ESA-ESRIN), Frascati, Italy
<sup>6</sup>Institute for Space Sciences, Freie Universität Berlin, Germany
<sup>7</sup>Swedish Meteorological and Hydrological Institute (SMHI), Norrköping, Sweden
<sup>8</sup>Informus GmbH, Berlin, Germany
<sup>a</sup>now at: EUMETSAT, Darmstadt, Germany
<sup>b</sup>now at: Geoforschungszentrum (GFZ), Potsdam, Germany

Correspondence to: Ralf Bennartz (ralf.bennartz@vanderbilt.edu)

Received: 17 September 2016 – Discussion started: 19 October 2016 Revised: 28 January 2017 – Accepted: 20 March 2017 – Published: 12 April 2017

Abstract. The microwave radiometers (MWRs) on board the European Remote Sensing Satellites 1 and 2 (ERS-1 and ERS-2) and Envisat provide a continuous time series of brightness temperature observations between 1991 and 2012. Here we report on a new total column water vapour (TCWV) and wet tropospheric correction (WTC) dataset that builds on this time series. We use a one-dimensional variational approach to derive TCWV from MWR observations and ERA-Interim background information. A particular focus of this study lies on the intercalibration of the three different instruments, which is performed using constraints on liquid water path (LWP) and TCWV. Comparing our MWR-derived time series of TCWV against TCWV derived from Global Navigation Satellite System (GNSS) we find that the MWR-derived TCWV time series is stable over time. However, observations potentially affected by precipitation show a degraded performance compared to precipitation-free observations in terms of the accuracy of retrieved TCWV. An analysis of WTC shows further that the retrieved WTC is superior to purely ERA-Interim-derived WTC for all satellites and for the entire time series. Even compared to the European Space

Agency's (ESA) operational WTC retrievals, which incorporate in addition to MWR additional observational data, the here-described dataset shows improvements in particular for the mid-latitudes and for the two earlier satellites, ERS-1 and ERS-2. The dataset is publicly available under doi:10.5676/DWD\_EMIR/V001 (Bennartz et al., 2016).

# 1 Introduction

ESA's altimetry missions are at the heart of significant progress on oceanography. The combined coverage of highquality observations by the European Remote Sensing Satellites 1 and 2 (ERS-1 and ERS-2) and Envisat spans over more than 20 years from 1991 to 2012. During this period, improvements in instrument data processing as well as orbit and geophysical corrections allowed reaching an accuracy/sensitivity of 1 cm on instantaneous sea surface height (SSH) measurements and demonstrated the capability to observe a 3 mm yr<sup>-1</sup> sea level rise (Ablain et al., 2009). A major source of uncertainty for radar altimetry is the wet tropospheric correction (WTC). The spatial and temporal variability of water vapour is such that an instantaneous estimation of its impact is needed. To provide the observations required for the WTC is the primary role of the nadir-looking microwave radiometer (MWR) embedded into the altimetry missions on board ERS-1, ERS-2, and Envisat. In this context, requirements on accuracy, sensitivity, and long-term stability of the atmospheric water vapour observations are particularly strong since altimetry missions require a precision better than 1 cm in WTC (Eymard et al., 2005) and a temporal stability better than 1 mm yr<sup>-1</sup> (Ablain et al., 2009). Note that a total column water vapour (TCWV) contribution of 1 kg m<sup>-2</sup> is equivalent to a WTC of about 6.4 mm.

Water vapour also is a highly important climate variable in its own right. The atmospheric water vapour feedback is believed to be the strongest feedback mechanism in climate change, approximately doubling the direct warming impact of increased CO<sub>2</sub> forcing (Cess et al., 1990; Forster et al., 2007). Various groups have reported trends in the amount of columnar water vapour. In particular, over the oceans, a strong trend in TCWV has been observed (Trenberth et al., 2005). TCWV also appears to be a key factor regulating tropical precipitation (Bretherton et al., 2004).

The importance of water vapour in the climate system is recognized by the Global Energy and Water Cycle Experiment (GEWEX), which is currently performing an assessment of long-term water vapour products (the GEWEX Water Vapour Assessment, G-VAP). Of particular importance for the current study is the recent publication by Schroeder et al. (2016), who provide an overview on existing TCWV datasets and assess their long-term stability as well as issues caused, for example, by changes in the observation systems.

The MWR instruments on board ERS-1, ERS-2, and Envisat are based on very similar architectures and have measured water vapour over the ocean between 1991 and 2012 and therefore provide a dataset of 20+ years of water vapour observations. Data continuity into the future is ensured with ESA's Sentinel-3 series of satellites, the first of which was launched in February 2016. MWR's two channels located at 23.8 and 36.5 GHz allow for the simultaneous retrieval of TCWV and cloud liquid water path (LWP) as outlined later in this publication. Through the REAPER (REprocessing of Altimeter Products for ERS) project (Gilbert, 2014)<sup>1</sup>, ESA has provided significant efforts to produce an up-to-date data record of well-calibrated MWR observations.

The efforts described herein build on the REAPER dataset and address three interconnected issues. Firstly, the homogenization and intercalibration of the MWR data record is studied and an improved intercalibration is developed. Secondly, using state-of-the-art one-dimensional variational (1D-VAR) retrievals, a TCWV data product is developed and made 
 Table 1. Main characteristics of the microwave radiometer series.

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

\* Envisat/RA-2/MWR Product Handbook: http://earth.esa.int/pub/ESA\_ DOC/ENVISAT/RA2-MWR/ra2-mwr.ProductHandbook.2\_2.pdf

available to the community. A revised WTC product accompanies this dataset also. Thirdly, the dataset is validated both against Global Navigation Satellite System (GNSS) observations of TCWV as well as in terms of its mesoscale stability with respect to WTC. The so-derived dataset is made available to the community.

This publication is organized as follows: in Sect. 2 we describe the MWR brightness temperature time series as well as the methods used for retrieving TCWV and WTC. Section 3 addresses the central issue of intercalibration and Sect. 4 summarizes the validation results. The sensitivity of the retrieval with respect to the background (a priori) temperature and water vapour profiles is discussed in the Appendix.

#### 2 Datasets and methods

#### 2.1 The MWR dataset

The MWR flown on board the ERS-1, ERS-2 and Envisat missions are two-channel nadir-pointing passive microwave instruments measuring top-of-atmosphere (TOA) brightness temperatures (see Table 1 for more details). A similar instrument is also flown on ESA's current Sentinel-3 mission. The MWR is nadir-pointing only as its main use to support wet tropospheric correction for the collocated radar altimeter. ERS-1, ERS-2, and Envisat were flown in 98.4–98.5° inclination orbits at altitudes around 780 km, providing near-global MWR coverage with the exception of a narrow region near the poles.

The Envisat MWR brightness temperatures underlying the have been generated by Collecte Localisation Satellite (CLS) in 2014 in the framework of the Envisat MWR L1B Expert Support Laboratory (ESL) activities funded by ESA. It consists of a corrected dataset that removes the anomaly that has been observed in version 2.1 (Ollivier and Guibbaud, 2014). The ERS-1 and ERS-2 MWR brightness temperatures used herein are based on the REAPER project (Gilbert, 2014) but have been entirely reprocessed in the framework of the EMiR (ERS/Envisat MWR Recalibration and Water Vapour Thematic Data Record Generation) project. The so-

<sup>&</sup>lt;sup>1</sup>https://earth.esa.int/web/sppa/activities/ multi-sensors-timeseries/reaper/

called "first run" REAPER L1B data have been the basis for this reprocessing. Land measurements are discarded and no specific processing is applied in coastal areas so that contamination from land may occur above coastal waters at distances of less than ca. 50 km from land. Such potentially land-contaminated pixels are excluded from the analysis presented here.

# 2.2 GNSS dataset

MWR estimates of TCWV are evaluated against a 2-hourly dataset of TCWV measured by ground-based GPS (Wang et al., 2007), composed of the International GNSS Service (IGS), SuomiNet, and GPS Earth Observation Network (GEONET), and hosted by NCEP/NCAR. While the focus of SuomiNet (see e.g. Ware et al., 2000) is the US and Central America and GEONET (see e.g. Shoji, 2009) is based on Japanese stations, IGS (see e.g. Byun and Bar-Sever, 2009) includes about 500 stations distributed globally<sup>2</sup>. The methodology of how TCWV is derived from the measured GPS zenith path delay is described in detail in Wang et al. (2007). They also found the individual errors to be less than 1 mm and total errors to be less than 1.44 mm. The GNSS database version 721.1 includes data from 1995 to 2014. In total, 997 stations are specified. However, not all of them cover the full period. Every station is characterized by latitude, longitude and altitude and provides 2-hourly data that contain day and time (UTC), surface pressure (hPa), atmospheric weighted-mean temperature (K), TCWV, and information about zenith delay.

# 2.3 Retrieval methods

# 2.3.1 TCWV and LWP

Combined TCWV and LWP retrievals were performed here are based on a 1D-VAR scheme initially developed at ECMWF by Phalippou (1996). This scheme originally focused on microwave observations from SSMIS and AMSU. It was extended by Deblonde and English (2003) towards a stand-alone scheme applicable to SSM/I, SSMIS, and AMSU. This scheme was also used in the ESA DUE Glob-Vapour project (http://www.globvapour.info). Here, it has been modified to derive TCWV from brightness temperatures specifically from the MWR sensor family on board ERS-1/2 and Envisat over the ice-free ocean. All other components of the retrieval scheme are identical to the Glob-Vapour scheme. The underlying radiative transfer model is the TIROS Operational Vertical Sounder radiative transfer model (RTTOV, version 7) (Saunders et al., 2007; Hocking et al., 2011), which is widely used both in the operational and research community. All physical parameterizations including absorption by gases and liquid water and rough ocean surface scattering are used unchanged from RTTOV. The optimal estimation scheme follows optimal estimation theory considering the uncertainties in the required meteorological background information, forward modelling (radiative transfer simulations), and satellite observations.

The 1D-VAR scheme employed here uses daily global ERA-Interim (Dee et al., 2011) TCWV and atmospheric temperature and cloud water content profiles and various surface fields as a priori (background) and first guess information. Information on the choices of various input parameters is found in the next subsections.

For the particular case of the MWR, the observation vector consists of the two brightness temperatures at 23.8 and 36.5 GHz, from hereon referred to as 23 and 36 GHz. The 23 GHz channel is close to the 22.231 GHz water vapour line and the 36 GHz channel only experiences weak continuum absorption by gases but stronger absorption by cloud liquid water. The 1D-VAR uses a 74-element state vector including 43 temperature levels, 26 moisture levels, surface air temperature, surface specific humidity, sea surface temperature, wind speed, and cloud liquid water path. While the interdependencies especially of the vertical levels are constrained by the background error covariance, the retrieval obviously remains heavily under-constrained, so that ultimately only TCWV and LWP are constrained by the two observations at 23 and 36 GHz.

Given just these two observations, care must be taken not to overfit the solution by optimizing parameters only weakly related to the two observables (say, the atmospheric temperature profile). While the Jacobian matrix for such lessconstrained parameters is in any case small, a cleaner way of eliminating this issue is to set the Jacobian to zero for all parameters that are not desired to be retrieved. This was done for this study for the atmospheric temperature profile and all surface parameters. Therefore only the water vapour profile and cloud liquid water are actively adjusted during the 1D-VAR process, whereas all other parameters were treated as fixed background parameters, set to the values prescribed by ERA-Interim.

The cost function J(x) is defined as

$$J(\mathbf{x}) = \frac{1}{2} (\mathbf{x} - \mathbf{x}_b)^T \mathbf{S}_b^{-1} (\mathbf{x} - \mathbf{x}_b) + \frac{1}{2} (H(\mathbf{x}) - \mathbf{y})^T \mathbf{S}_o^{-1} (H(\mathbf{x}) - \mathbf{y}),$$
(1)

where the first part on the right-hand side determines the cost of the solution with respect to the background and the second part determines the cost with respect to the observations. For any given retrieval the expectation value of J(x) is equal to the number of observations. In the above formulation y is the observation vector, x the state vector,  $x_b$  the background state, H(x) the forward model,  $S_b$  the background error covariance matrix, and  $S_0$  the observation error covariance matrix. For further details on optimal estimation, see Rodgers (2000). For implementation details regarding  $S_b$  and  $S_o$ , see Appendix B.

<sup>&</sup>lt;sup>2</sup>www.igs.org/network



**Figure 1.** Example of histograms of retrieved LWP and fitted Gaussian for different bias correction values at 36 GHz. The example shows MWR on Envisat, January 2005. For the example shown here, the bias correction at 23 GHz is set to -3 K.

# 2.3.2 WTC

The derivation of the WTC follows the analytical procedure laid out in Appendix A. The value of the wet tropospheric delay depends on the MWR-derived TCWV and on the value Tm (Eq. 15), which in turn depends on the ERA-Interim temperature profile and the MWR-derived TCWV.

# 2.4 Generation of gridded data

The individual retrievals discussed above were used to calculate global fields of monthly averages of total column water vapour, liquid water path, and brightness temperatures at 23 and 36 GHz on a  $2 \times 2^{\circ}$  as well as a  $3 \times 3^{\circ}$  latitude–longitude grid.

Before calculating the monthly means, some filters were applied. A retrieved data pixel was used if a valid total column water vapour retrieval was available, meeting two additional conditions: (1) liquid water path larger than  $-1 \text{ kg m}^{-2}$  and (2) a cost function value lower than five. The last condition effectively removes heavily precipitation-contaminated pixels as well as observations with remaining sea ice or land contribution. Data of at least 20 days were required within a grid cell for monthly-mean values to be reported.

In a first step, the grid was set up according to the considered spatial resolution  $(2^{\circ} \times 2^{\circ} \text{ or } 3^{\circ} \times 3^{\circ} \text{ lat/long})$ . For each MWR footprint, the corresponding indices of the global fields were calculated from the latitude–longitude information. If all conditions were met (see above), the retrieved and auxiliary values were added to the corresponding grid box. The daily averages were then calculated as the arithmetic mean of all observations within that grid box within 1 day. If, for a given month and grid box, more than 20 days had valid observations, the arithmetic mean of those was assigned to be the monthly-mean value. The so-derived monthly-mean fields, along with the individual retrievals, are part of the published dataset. All analysis reported within this publication was, however, performed on individual retrievals.



Figure 2. The coloured contour plot in (a) shows contours of LWP bias as function of bias correction values for 23 GHz (x axis) and 36 GHz (y axis). The coloured contour plot in (b) shows biases in TCWV. In addition, labelled isolines of both TCWV bias (dashed) and LWP bias (solid) are overlaid in both plots. Data are shown for ERS-1, June 1994.

# 3 Intercalibration and bias correction

Retrieval of geophysical variables using physical models and optimal estimation procedures requires the elements of the observation vector to be on average unbiased compared to the forward model applied to the true state of the atmosphere. Comparing first guess simulations with observations, biases include contributions from the following error sources:

- representativeness of the state vector in the first guess (e.g. representation of clouds in the model);
- spatial and temporal colocation errors between first guess and MWR observations;
- calibration biases/errors of the different MWRs;
- systematic errors and uncertainties in the surface emissivity model;
- systematic errors and uncertainties in spectroscopy of liquid water absorption, dry air absorption, and water vapour absorption;
- impact of precipitation contamination and precipitation-ice scattering not accounted for in forward model,

While the first two items on this list have a significant impact on the values reported here, they only play a secondary role for the retrieval accuracy. As pointed out in Appendix B, the retrieval is sufficiently independent from the first guess to allow for valid retrievals even if the first guess and prior are relatively far away from the true state of the atmosphere.



**Figure 3.** Optimal bias correction values for the entire time series shown as coloured lines with filled circles marking each monthly value. Also shown are the temporal mean values for each channel and instrument (straight lines in slightly lighter colours than the corresponding monthly values) and a regression line (dashed lines). For ERS-2 two separate fits were performed. One corresponding to the period before the 23 GHz gain drop and another one for the period after the gain drop. The blue arrow highlights the time the day drop happened (26 June 1996). A negative bias correction value means that the observations are warmer than the simulations and thus need to be corrected downwards.



Figure 4. Overview of location and data density of GNSS stations used for TCWV analysis.

Furthermore, the relaxed background error covariance criteria formulated also in Appendix B allow for the retrieval to converge to values corresponding to the observed brightness temperatures.

The latter four items in the above list, while having smaller contributions to the overall bias, are of crucial importance to the accuracy and long-term stability of a climate data record. These are addressed in an empirical bias correction scheme as outlined below.

#### 3.1 Method

Because of the low sensitivity of the retrieval to the first guess as well as to the background state (see Appendix B), the bias correction proposed here relies on two main assumptions.

First, the globally averaged TCWV from ERA-Interim is considered reasonably accurate in terms of its absolute value to provide a reference against which to gauge the average brightness temperature biases of the MWR time series. Note that we do not make any claims about the long-term stability of the ERA-Interim TCWV or any trends and discontinuities of the dataset. The only assumption we make is that ERA-Interim TCWV on a globally and monthly averaged basis is accurate to within  $\pm 2 \text{ kg m}^{-2}$ . This assumption is justified from intercomparison efforts such as Schröder et al. (2013).

The second assumption is that histograms of instantaneous retrieved LWP must show a significant fraction of negative retrieved LWP, which corresponds to measurement noise around zero LWP for cloud-free situations. For typical bias-free optimal estimation retrievals LWP for cloud-free cases is centred around zero  $g m^{-2}$  with a standard deviation of about 30–40 g m<sup>-2</sup> (see e.g. Greenwald, 2009; Bennartz et al., 2010).

These two constraints can be used to find an optimal bias correction for both channels of each instrument in the following way: for each instrument and month a certain amount of observations out of all observations available were randomly sub-selected. We chose 4 % of the total number of observations, which presented a compromise between statistical significance and computational efficiency.



**Figure 5.** Comparison of GNSS and MWR-derived TCWV. (a) and (c) show comparisons for all collocations. (b) and (d) show comparisons for a subset of collocations where (1) MWR-retrieved LWP was smaller than  $200 \text{ g m}^{-2}$  and (2) the GNSS station height was below 50 m above MSL. (a) and (b) show data density plots. (c) and (d) show time series of monthly-mean differences between MWR and GNSS for the entire dataset. The blue lines give linear fits to the data.

For these 4 % we ran a series of retrievals with different biases at 23 and 36 GHz. We ran retrievals for bias values running from -8 to 0 K in steps of 1 K for both channels, so that in total  $9 \times 9 = 81$  retrievals were performed on the set of 4 % of each data per month.

We then found the best choice of biases for 23 and 36 GHz, i.e. the combination of biases for which the difference between background TCWV (ERA-Interim) and retrieved TCWV is smallest and the LWP shows the Gaussian behaviour around zero. This will be the optimal bias value for this particular month.

Step 3 was repeated for all months and satellites to find monthly optimal average bias values.

Figures 1 and 2 highlight some key methodological issues related to the method. In particular, Fig. 1 shows how the histogram of LWP shifts as the bias at 36 GHz varies. It also shows the histograms fitted to the retrieved LWP histograms. This fit was performed on the part of the histogram left of its peak, assuming that all values left of the peak correspond to cloud-free scenes. The super-Gaussian distribution to the right of the peak corresponds to actual clouds. In the particular case shown in Fig. 1 the best bias value for 36 GHz would be very close to -6 K, thereby centring the histogram on zero as outlined above.

Figure 2 shows isoplots of TCWV and LWP histogram biases for a full set of monthly retrievals and all combinations of 23 and 36 GHz biases. The optimal set of biases can now be inferred from this histogram as the intersect between the zero TCWV bias isoline (thick, solid) and the zero LWP histogram bias line (thick, dashed) and is located near (-4, -7 K). The same analysis was performed for all months and instruments. The results are shown in Fig. 3 and discussed in the next section.

From Fig. 2 it is also noteworthy that the sensitivity of TCWV to biases in 23 GHz brightness temperatures is roughly  $1 \text{ kg m}^{-2}$  per 1 K bias. The sensitivity of LWP to biases at 36 GHz is roughly 25 g m<sup>-2</sup> per 1 K bias. Note that both variables also exhibit sensitivity to the other frequency although the sensitivity is somewhat smaller as expected.



**Figure 6.** Validation statistics for WTC sea surface height variability. Data shown here are for Envisat only. (a) and (c) show comparisons against WTC derived purely from ERA-Interim. (b) and (d) show comparisons against the operational WTC retrieval from Envisat. (a) and (b) show globally averaged monthly statistics. (c) and (d) show long-term spatial statistics. Values smaller than zero indicate an improvement. Values larger than zero indicate a deterioration of the results relative to the comparison dataset.

Table 2. Summary of comparisons between ERA-Interim versus GNSS and MWR versus GNSS in terms of bias, RMSE, and stability of the dataset.

Comparison	Data selection	Bias (kg m <sup>-2</sup> )	RMSE (kg m <sup>-2</sup> )	Stability absolute $(kg m^{-2} decade^{-1})$	Stability relative (% decade <sup>-1</sup> )
ERA-Interim versus GNSS	All LWP<200, <i>H</i> <50	$0.78 \\ -0.03$	4.87 4.53	0.28 0.12	0.96 0.32
MWR versus GNSS	All LWP<200, <i>H</i> <50	$0.63 \\ -0.43$	4.68 3.95	0.21 -0.12	0.68 -0.38

The sensitivities of brightness temperatures to changes in LWP and TCWV together with the Gaussian plots shown in Fig. 1 also allow for a rough estimation of the total uncertainty of the brightness temperatures at 23 and 36 GHz for cloud-free cases. The width of the Gaussian distributions seen in Fig. 1 is 41 g m<sup>-2</sup>. Using the sensitivities of LWP with respect to 23 GHz ( $\sim 14$  g m<sup>-2</sup> per 1 K) and 36 GHz ( $\sim 25$  g m<sup>-2</sup> per 1 K) and Gaussian error propagation, an observation error of 1.01 K in both channels would explain the resulting width of the cloud-free Gaussian histograms seen in Fig. 1. As stated in Appendix B, in our retrieval we use a diagonal observation error covariance matrix of 1 K for both channels, which is consistent with this estimate.

#### 3.2 Bias analysis

Figure 3 shows the outcome of the above-described bias analysis for all instruments and channels. Mean bias values as well as slopes are also listed in Table 1. A negative bias correction value means that the observations are warmer than the simulations, and thus the observations need to be corrected downwards. We found the following:

- Biases correction values at 23 GHz are between about -4 K for ERS-1 and -2 K for ERS-2 with Envisat being in between these two values.
- ERS-2 23 GHz brightness temperatures show a significant decrease in bias exactly at the time of the instru-

**Table 3.** Mean optimal bias correction values and regression slopes for the all instruments and time periods. The values given here correspond to the straight lines and dashed lines in Fig. 3. A negative bias correction value means that the observations are warmer than the simulations and thus need to be corrected downwards.

Instrument	Period	Mean		Regression slope*		Regression offset*	
		23 GHz	36 GHz	23 GHz	36 GHz	23 GHz	36 GHz
		(K)	(K)	$(\mathrm{K}\mathrm{yr}^{-1})$	$(\mathrm{K}\mathrm{yr}^{-1})$	(K)	(K)
ERS-1	10/1992-06/1996	-4.42	-7.25	-0.12	-0.04	-3.86	-7.05
ERS-2	10/1995-06/1996	-2.66	-4.28	-0.57	-1.72	+0.83	+6.24
ERS-2	07/1996-06/2003	-1.93	-0.09	-0.04	-1.02	-4.14	
Envisat	05/2002-04/2012	-2.87	-5.68	+0.10	+0.06	-4.65	-6.65

\* The regression bias is calculated using bias\_corr(t) = slope \*t + offset, where t is the decimal year since 1990. For example, 2 July 1991 is day 183 in the year 1991 and therefore corresponds to a value of t = 1.5.

**Table 4.** Retrieval statistics for the four different 1D-VAR configurations described in Sect. 2.3.1. Corresponding scatterplots are shown in Fig. 8. Total number of retrievals was 35 584. Reported bias and RMSE values are for valid only for the fraction of retrievals with cost function lower than 5.

Experiment	Bias with respect to ERA-Interim (kg m <sup>2</sup> )	RMSE with respect to ERA-Interim (kg m <sup>2</sup> )	Percent retrieved (%)	Mean Tb residual after retrieval (K)	Percent with Tb residual larger than 1 K (%)
FIXED_NEW	0.88	4.80	68.5	0.23	0.02
FIXED_OLD	0.46	3.99	45.7	0.42	1.52
ERA_NEW	-0.01	5.06	97.9	0.07	0.91
ERA_OLD	0.00	3.53	87.2	0.41	6.24

ment's gain drop (indicated by the blue arrow in Fig. 3) and a downward trend in bias (slope) for the period afterwards. This strong drop prompted us to separate ERS-2 in a pre- and post-gain time period for which performed a separate analysis (listed in Table 1).

- Envisat shows a slight upward slope possibly over the first half of its lifetime or over its entire lifetime.
- Brightness temperature biases at 36 GHz are comparably stable over time for both ERS-1 and ERS-2; i.e. the regression slopes are very small.
- Envisat shows a strong annual cycle in bias at 36 GHz, which diminishes somewhat after 2008. It also shows a similar trend as it does for 23 GHz. There was no explanation for this behaviour at the time of writing.

The bias values given here are based on optimal comparison between retrieved versus background TCWV as well as constraints made on the histogram of retrieved LWP. While these constraints are physically reasonable and justifiable on average, it is not advisable to perform monthly bias corrections based on the individual values derived for that particular month. This would by example of TCWV likely result in an overfitting to the ERA-Interim time series.

However, different biases between two instruments observed in their overlap periods can be clearly attributed to the instrument calibration, as the background is identical. Similarly the mean bias values reported in Table 1 provide for a reasonable way of addressing the principal error sources outlined above. A first-order bias correction is therefore performed by subtracting the bias values in Table 1 from the observations.

We note here that biases on the order of -2 to -5 K (simulations too warm compared to observations) are also reported by ECMWF for monitoring of AMSU-A against their operational forecasting system<sup>3</sup>. It is therefore likely that ERS-2 is calibrated to within the absolute calibration accuracy of 3 K stated for the instrument. In contrast, ERS-1 and Envisat both show much larger biases, which might be indicative of remaining calibration issues.

As pointed out above, no assumptions about the long-term stability of ERA-Interim should be made in this analysis. The regressions listed in Table 1 therefore cannot conclusively be interpreted as being caused either by natural variability in TCWV or by instrument drifts.

While a conclusive statement of the origins of the regression slopes and related trends cannot be made, it is interesting to relate the slopes back to the aforementioned retrieval sensitivities. Ignoring the relatively short ERS-2 period before

<sup>&</sup>lt;sup>3</sup>http://www.ecmwf.int/en/forecasts/charts/obstat/?facets= Parameter,Allskyradiances On this website select AMSU, radiances from AMSUA (time-averaged geographical mean) channel 1, AN departure.

the gain drop, the regression slopes found in Table 1 show values between -0.12 and +0.1 K yr<sup>-1</sup>. Factoring in the sensitivities of the retrieval the retrieved TCWV, trends observed in global TCWV between the constant bias correction and the regression are expected to be different by roughly -1.2 to +1.0 kg m<sup>-2</sup> decade<sup>-1</sup>. These numbers provide bounds on how large observed TCWV trends have to be in order to be considered real, if the current dataset were being used.

# 4 Validation

# 4.1 TCWV

MWR-retrieved TCWV was validated against TCWV derived from coastal GNSS stations. Validation was performed in terms of biases and root-mean-square errors (RMSE) as well as in terms of long-term stability of the dataset. A total of 30712 valid collocations were found. The temporal collocation difference was chosen to be at maximum 1 h. Note that our dataset excludes observations close to land, so that the distance between a coastal GNSS station and the nearest MWR observation is somewhere around 100 km. The maximum collocation distance was chosen to be 150 km. Figure 4 shows the coastal GNSS stations used. Figure 5 and Table 1 summarize the validation results. For the entire dataset, we find a relatively small bias of  $0.63 \text{ kg m}^{-2}$  and an RMSE of  $4.68 \text{ kg m}^{-2}$ . If the comparisons were restricted to exclude MWR retrievals with high LWP (>  $200 \text{ g m}^{-2}$ ) and GNSS stations with altitudes higher than 50 m above mean sea level (MSL) were also excluded, the RMSE was reduced to  $3.95 \text{ kg m}^{-2}$ . LWP values larger than about  $200 \text{ g m}^{-2}$ likely represent precipitating clouds (e.g. Wentz and Spencer, 1998). We therefore reason that remaining issues with precipitation contamination might deteriorate the TCWV retrievals.

The lower panels in Fig. 5 provide time series of average relative differences between MWR and GNSS. These were calculated by first computing the relative difference for each individual observation and then averaging these individual relative differences. Note that for the same absolute difference relative differences are larger if the absolute value is smaller; e.g. for an absolute value of  $20 \text{ kg m}^{-2}$ , an absolute difference of  $2 \text{ kg m}^{-2}$  would correspond to a 10% relative difference, whereas for an absolute difference value of 60 kg m<sup>-2</sup> that same absolute difference would only correspond to a 3.3 % relative error. Therefore relative biases can be different from absolute biases as can be observed in Fig. 5. We provide the stability here in relative terms because the World Meteorological Organization's (WMO) Global Climate Observing System (GCOS) requirements for temporal stability of TCWV are formulated in relative term as well (GCOS-107, 2006). Comparing time series of MWR-GNSS, the expected stability (i.e. the slope of regression line) should be zero. The observed long-term stabilities are +0.68 % decade<sup>-1</sup> and -0.38 % for the entire dataset and for



**Figure 7.** Zonally averaged patterns of improvement (green) and deterioration (red) of our retrieval over the operational ESA retrieval.

the reduced LWP < 200 dataset respectively. Both stability values are not statistically significantly different from zero (P values for two-sided t test are 0.148 and 0.572 respectively), so that at least for the comparison against GNSS the dataset can be considered stable over time.

Several other tests including separation of daytime from night-time observations did not yield any additional insights into the uncertainties.



**Figure 8.** Scatterplots of retrievals obtained with the four different 1D-VAR configurations described in Sect. 2.3.1. In all four panels the retrieved TCWV is plotted against ERA-Interim TCWV. The green line is the 1–1 line. The red error bars show the mean and standard deviation in bins of  $5 \text{ kg m}^{-2}$ . The vertical line in the upper two plots shows the TCWV of the fixed climate background used for these plots (mid-latitude summer atmosphere). Total number of retrievals was 35 584. Reported values are for valid retrievals with cost function lower than 5. Corresponding statistics are listed in Table 2.

Table 1 further compares the validation against GNSS for between the MWR retrievals and the ERA-Interim profiles used as a background. For the less stringent screening including all GNSS-matches, the RMSE is reduced only slightly from 4.87 to  $4.68 \text{ kg m}^{-2}$ . For the more stringent screening (LWP < 200, H < 50), the RMSE for the MWR-retrieved TCWV is reduced to  $3.95 \text{ kg m}^{-2}$ , compared to  $4.53 \text{ kg m}^{-2}$ when directly compared to ERA-Interim. The retrieval biases show a less clear improvement. While the RMSE shows improvements, the assessment of the bias is inconclusive, with an improvement in bias on average for all cases but degradation in bias, if the more stringent screening criteria are applied. The long-term stability of the dataset of the differences also shows somewhat inconclusive results. It has to be noted though that none of the stability estimates are statistically significantly different from zero.

#### 4.2 Wet tropospheric delay

To assess the accuracy of WTC, we employ the "cross-over approach" outlined by Legeais et al. (2014). The idea of this approach is to find cross-over points over the same region within a relatively short time period (10 days or below). The SSH is assumed to be on average constant within this relatively short time interval. Thus, sequential SSH observations should ideally give the same answer within the expected uncertainties. Based on this analysis method, the accuracy of different WTC retrievals can be compared. In particular, the WTC retrieval showing the smallest variability in retrieved SSH will be most accurate.

Here, we compare our WTC retrieval with two independent retrievals. Following Legeais et al. (2014), we compare our WTC retrieval firstly to WTC calculated only on the basis of ERA-Interim observations. Secondly, we compare to ESA's operational retrieval, which employs a neural network for WTC retrieval and, importantly, uses the MWR brightness temperatures together with altimeter-derived information about the state of the sea surface (the altimeter backscatter coefficient). This additional information allows for a better characterization of surface emissivity and is not used in our retrieval.

Figure 6 summarizes the validation results for Envisat. Comparisons against ERA-Interim WTC show overall improvements over the entire data record consistent with similar results shown in Legeais et al. (2014) for other sensors. The lower left panel of Fig. 6 shows improvements nearly everywhere over the oceans with the exception of the Gulf Stream off the coast of North America, the Falkland current, and the confluence areas of the Agulhas and Benguela currents. All of these areas show exceptionally high variability in SSH and might therefore pose particular challenges to the validation method used here.

Comparing our retrievals with the operational Envisat retrievals (Fig. 6, upper and lower right panels), our algorithm performs slightly inferior than the operational algorithm in most areas and only shows some improvements in the Southern Ocean near the ice edge. Between the three satellites studied here, Envisat shows the strongest deterioration of results compared to the operational retrieval. In fact, ERS-1 and ERS-2 show consistent improvements over the operational retrieval in the northern and southern mid-latitudes (see Fig. 7). We note that in contrast to ESA's operational retrieval our algorithm does not use any altimeter information to constrain sea surface emissivity and solely relies on the MWR observations and ERA-Interim as background information. We therefore expect our algorithm to perform slightly worse than ESA's operational algorithm. In this context it is surprising that our results in terms of WTC are improved for significant areas of the globe at least for ERS-1 and ERS-2.

#### 5 Conclusions

We have established a new long-term (1992–2012) dataset of TCWV and WTC over the global oceans, which is readily available for users. The main difference between our dataset and earlier datasets, such as ESA's operational dataset, is that we explicitly intercalibrate the MWRs on the three different satellite platforms. Such an intercalibration is a key requirement for the generation of satellite-based climatologies. We have introduced a novel intercalibration method that allows for the adjustment of biases between the different satellites based on environmental conditions over ocean. Similar to operational data assimilation systems, we estimate calibration offsets at 23 and 36 GHz by comparing simulated and observed cloud-free brightness temperatures over the lifetime of each satellite. We only adjust for long-term, global calibration trends and offset but do not perform any regional, monthly, or seasonal bias corrections.

A potential disadvantage of our intercalibration method is that it hinges the long-term mean retrieved TCWV to the TCWV fields that were used in the simulations used in the intercalibration process (in our case ERA-Interim). While this does limit the use of our dataset to assess the global longterm mean TCWV as well as linear trends, the dataset can be used for anomaly analyses and to provide WTCs for altimeter correction. We note that limitations on absolute accuracy or linear trends are in no way unique to our dataset. For example, Schroeder et al. (2016) investigate trends and discontinuities in six different global TCWV climatologies and find significant artefacts that inhibit trend analyses in nearly all datasets.

Independent validation of our dataset was performed against GNSS radiosondes. This validation shows the TCWV dataset to be stable over time and generally yielding an improvement in RMSE over the background TCWV used in the retrievals. The dataset has also been submitted to the ongoing G-VAP (Schroeder et al., 2016) for evaluation in the context of several other TCWV datasets. A crossover analysis of altimeter corrections using the here-derived WTC provided further insights into the accuracy of the dataset. We found that for ERS-1 and ERS-2 the results of our WTC retrievals are superior to ESA's operational WTC retrieval for the midlatitudes and inferior for the tropics. For Envisat, ESA's operational WTC retrieval is generally slightly better than our retrieval. ESA's retrieval does use additional collocated altimeter backscatter information to constrain sea surface emissivity. In contrast, our algorithm currently relies only on ERA-Interim surface wind speed, so that a slight deterioration of our results compared to ESA's should be expected. A 1D-VAR retrieval of surface wind speed, water vapour, and cloud liquid water based on combined radar altimeter and MWR data could also be envisioned. This will likely have a significant positive impact on retrievals from both instruments. It would require extending the surface emissivity model as well as the 1D-VAR retrieval to include altimeter backscatter, which would also have to be intercalibrated.

A number of further limitations exist: in the current implementation, ERA-Interim fields are only used once a day at 12:00 UTC as background profiles. Rapid changes of atmospheric conditions and surface properties are not accounted for. We have studied the impact of the background state on retrieval quality (see Appendix B) and conclude that this limitation has only a marginal impact on data quality, as the algorithm is only weakly dependent on the choice of the background.

Since MWR is nadir-looking only, it does not provide any polarization information. Compared to other microwave sensors, its spectral range is also limited the two frequencies at 23 and 36 GHz. Therefore, screening observations affected by frozen hydrometeor scattering will not be possible. Thus, in cases of moderate to heavy frozen hydrometeor load, such as in deep convective cores, retrieval results will likely be degraded. Here we employ a screening based on the final value of the cost function, which has proven efficient in eliminating outliers. However, validation results show that the comparisons against GNSS are deteriorated for larger LWP, which would be indicative of remaining issues with precipitation screening.

An extension of the approach outlined here to include the MWR on board the Sentinel-3 satellites and potentially other altimeter/radiometer combinations appears highly desirable and would allow extending the current time series forward in time. Feedbacks between improved calibration efforts for single instruments and subsequent intercalibration need to be accounted for. Ideally, a processing chain would be set up to allow for re-processing the intercalibration for a single instrument if the underlying calibration for that instrument had improved.

#### **Appendix A: Calculating WTC**

# A1 General overview

The altimeter path delay along a path H is directly related to the real part of the refractive index of moist air n:

$$\Delta z = \int_{0}^{H} (n-1) \mathrm{d}z. \tag{A1}$$

Expressing this it terms of refractivity N, with N in ppm being

$$N = 10^6 (n-1), \tag{A2}$$

we get

$$\Delta z = 10^{-6} \int_{0}^{H} N \mathrm{d}z. \tag{A3}$$

Assuming H is the satellite altitude (nadir view),  $T_v$  is the virtual temperature, and using hydrostatic equilibrium we get

$$\Delta z = 10^{-6} \frac{R_{\text{AIR}}}{g} \int_{0}^{p_{\text{SFC}}} N \frac{T_{\text{v}}}{p} \mathrm{d}p.$$
(A4)

The refractivity *N* can be parameterized following references cited in Mangum (2009):

$$N = a_d \frac{p_d}{T} + a_w \frac{e}{T} + b_w \frac{e}{T^2}$$

$$a_d : 0.776890 \frac{\text{ppm} \cdot \text{K}}{\text{Pa}}$$

$$a_w : 0.712952 \frac{\text{ppm} \cdot \text{K}}{\text{Pa}}$$

$$b_w : 3754.63 \frac{\text{ppm} \cdot \text{K}^2}{\text{Pa}}.$$
(A5)

The variable p is the total pressure and pd represents the pressure of dry air, where the total pressure p = e + pd, with e being the water vapour partial pressure. With these definitions we can write the total path delay as

$$\Delta z = 10^{-6} \frac{R_{\text{AIR}}}{g} \left[ \int_{0}^{p_{\text{SFC}}} a_d \frac{p_d}{T} \frac{T_v}{p} dp + \int_{0}^{p_{\text{SFC}}} a_w \frac{e}{T} \frac{T_v}{p} dp + \int_{0}^{p_{\text{SFC}}} b_w \frac{e}{T^2} \frac{T_v}{p} dp \right].$$
(A6)

The first term in the brackets in Eq. (A7) can be split as follows:

$$\int_{0}^{p_{\text{SFC}}} \frac{p-e}{T} \frac{T_{\text{v}}}{p} dp = \int_{0}^{p_{\text{SFC}}} \frac{p}{T} \frac{T_{\text{v}}}{p} dp - \int_{0}^{p_{\text{SFC}}} \frac{e}{T} \frac{T_{\text{v}}}{p} dp, \qquad (A7)$$

so that Eq. (A7) can be expanded to become

$$\Delta z = \underbrace{10^{-6} \frac{R_{\text{AIR}}}{g} \int_{0}^{p_{\text{SFC}}} a_d \frac{p}{T} \frac{T_v}{p} dp}_{\text{Dry tropospheric delay}} + \underbrace{10^{-6} \frac{R_{\text{AIR}}}{g} \left[ \int_{0}^{p_{\text{SFC}}} (a_w - a_d) \frac{e}{T} \frac{T_v}{p} dp + \int_{0}^{p_{\text{SFC}}} b_w \frac{e}{T^2} \frac{T_v}{p} dp \right]}_{0}.$$
 (A8)

Wet tropospheric delay

Note that  $T_v/T \simeq 1$ .

# A2 Dry delay

Integrating the dry tropospheric part of Eq. (A9) yields

$$\Delta z_d = 10^{-6} \cdot \frac{R_{\text{AIR}}}{g} \cdot a_d \cdot p_{\text{SFC}}.$$
(A9)

The dry delay is on the order of 2.3 m for a straight vertical path through the atmosphere whereas the wet tropospheric delay is only on the order of 0.4 m at maximum.

# A3 Wet delay

Integrating the wet tropospheric terms in Eq. (A9) yields

$$\Delta z_w = 10^{-6} \frac{R_{\text{AIR}}}{g} \left[ \int_{0}^{p_{\text{SFC}}} (a_w - a_d) \frac{e}{T} \frac{T_v}{p} dp + \int_{0}^{p_{\text{SFC}}} b_w \frac{e}{T^2} \frac{T_v}{p} dp \right].$$
(A10)

The water vapour mass mixing ratio is defined as

$$r = \frac{R_{\rm H_2O}}{R_{\rm AIR}} \frac{e}{p}.$$
 (A11)

Replacing e/p accordingly with r into Eq. (A11) yields

$$\Delta z_w = 10^{-6} \frac{R_{\text{H}_2\text{O}}}{g} \left[ (a_w - a_d) \int_{0}^{p_{\text{SFC}}} r \,\mathrm{d}p + b_w \int_{0}^{p_{\text{SFC}}} \frac{r}{T} \,\mathrm{d}p \right].$$
(A12)

The TCWV is defined as

$$TCWV = \frac{1}{g} \int_{0}^{p_{SFC}} r dp.$$
(A13)

We further define a "water-vapour-averaged mean inverse atmospheric temperature",  $T_{\rm m}$ , as

$$T_{\rm m} = \left(\int_{0}^{p_{\rm SFC}} \frac{r}{T} \mathrm{d}p/\mathrm{TCWV}\right)^{-1}.$$
 (A14)

#### Atmos. Meas. Tech., 10, 1387-1402, 2017

#### www.atmos-meas-tech.net/10/1387/2017/

#### R. Bennartz et al.: Dataset of total column water vapour and wet tropospheric correction

With these two quantities, Eq. (A13) becomes

$$\Delta z_w = 10^{-6} R_{\text{H}_2\text{O}} \left( (a_w - a_d) + \frac{b_w}{T_{\text{m}}} \right)$$

$$\cdot \text{TCWV} = \left( A + \frac{B}{T_{\text{m}}} \right) \cdot \text{TCWV}$$

$$A : 10^{-6} \cdot R_{\text{H}_2\text{O}} \cdot (a_w - a_d) :$$

$$- 2.95077 \cdot 10^{-5} [\text{m}/(\text{kg m}^{-2})]$$

$$B : 10^{-6} \cdot R_{\text{H}_2\text{O}} \cdot b_w : 1.73276 [\text{m}/(\text{K} \cdot \text{kg m}^{-2})].$$
(A15)

The wet tropospheric delay is on the order of 0.4 m for high water vapour content.

# Appendix B: Sensitivity to a priori and background error covariance matrix

The observation and background error covariance matrices perform two major functions:

- Firstly, the trace of the observation error covariance matrix compared to the trace of the background error covariance matrix determines the relative weight of the observations compared to the background.
- Secondly, the background error covariance matrix determines the relative weight of the individual entries in the state vector with respect to each other. Similarly, the observation error covariance matrix determines the relative weight of the single observations in the observation vector with respect to each other. These "internal" weights are independent of any common multiplicative factor that can be extracted from the error covariance matrix.

Thus, by multiplying e.g. the background error covariance matrix by a scalar factor larger (smaller) than unity, the 1D-VAR can be forced to converge farther away from (closer to) the a priori. This property of the covariance matrices is of importance for example when a reanalysis background is exchanged for a climate background profile, the former likely being a much better representation of the actual state of the atmosphere than the latter.

## **B1** Sensitivity studies

Subsequently, we study the impact of fixed versus variable a priori as well as the impact of relaxation of the background error covariance matrix over the values established in earlier studies (Schröder et al., 2013), which are used as a reference. In all cases the observation error covariance matrix is diagonal and held constant at 1 K for both channels. In order to understand the sensitivity of the retrievals to different choices of background state vectors ( $x_b$ ) and background error covariance matrices ( $S_b$ ), the following series of retrieval tests

was performed on a single day of Envisat MWR observations (1 December 2011). The following four tests were run.

- 1. ERA\_OLD: ERA-Interim background with existing constraints (see below) on water vapour and liquid water background error covariance.
- 2. ERA\_NEW: ERA-Interim background but with less tight constraints on water vapour and liquid water. Constraints were relaxed by a factor of 2 in water vapour and 5 in cloud liquid water.
- 3. FIXED\_OLD: A fixed mid-latitude summer atmospheric background profile was used for all retrievals. The same existing constraints on error covariance matrices as in 1 were used.
- 4. FIXED\_NEW: The same fixed background profile as in 3 but with less tight constraints on background error co-variance (as in 2).

The fixed background was simulated using a mid-latitude standard atmosphere as first guess and background for all retrievals. Only surface wind speed and sea surface temperature were still used from ERA-Interim for the climate background. The FIXED scenario represents an extreme case of a climatological background profile in which the background is kept fixed regardless of location and season. This extreme case has been chosen because it allows study of the algorithm performance under a most restrictive scenario with a fixed background.

The modified background error covariance matrix for tests 2 and 4 was implemented by expanding the background standard deviation for cloud liquid water from 0.2 to  $1.0 \text{ kg m}^{-2}$  and by multiplying the lnQ sub-matrix of the background error covariance by a factor of 2, where lnQ is the logarithm of the water vapour mixing ratio, which is the water vapour control variable in the applied 1D-VAR retrieval scheme. For all cases we counted retrievals as valid if, after convergence, the final cost function value was lower than 5, i.e. at maximum 2.5 times larger than the expectation value of the cost function for valid retrievals.

The results of these tests are summarized in Table 2 and Fig. 8. Both FIXED retrievals show critical deficiencies. Neither the original nor the modified settings allow for a good fit using just one constant climate-like profile as background. In both FIXED cases only about 50% of the retrievals actually converge and large biases occur both at the high and low end of TCWV. These issues can be mitigated by further increasing the lnQ sub-matrix of the background error covariance by a factor of 10 instead of 2 and also by increasing the number of iterations in the minimization process from a current upper limit of 5 to 40. However, even with these newly revised parameters the number of converged profiles remains lower than for the ERA background.

An important finding from the FIXED cases is the relative insensitivity of the retrieval to the choice of the background.

As can be seen in Fig. 8. As long as the actual TCWV is less than  $5-10 \text{ kg m}^{-2}$  away from the chosen background, the retrieval will perform quite well, especially under relaxed background error covariance conditions. This is due to the high information content of the passive microwave observations with respect to both TCWV and LWP.

We note that the use of a single global background profile is not necessarily the best choice for a climatological background. A possible compromise could consist of less stringent choices of climatological backgrounds, allowing the background water vapour and temperature profile to vary with geographical position and latitude.

Compared to the FIXED cases both the ERA-NEW and ERA-OLD case show much better results. The trade-off here lies mainly between an increased RMSE (NEW) and an increased number of profiles with large remaining Tb residuals after convergence (OLD). The ERA\_NEW case allows the 1D-VAR to find low-cost solutions further away from the background water vapour profile. This will enhance the RMSE because we compare the retrieved TCWV to the background TCWV. ERA\_NEW in contrast provides tighter constraints on the background error covariance matrix and thus minimizes the RMSE better but at the cost of having a larger fraction of retrievals not converge as closely toward the observed brightness temperatures. For the particular case shown here 6.24 % of retrievals still show a residual deviation of simulated from observed Tb-s larger than 1 K. A design choice for the final retrieved time series was therefore the extent to which it adheres to the prescribed background compared to perfectly minimizing the observed brightness temperatures. We note that in an ideal world with perfect knowledge about background and observation error covariance matrix this choice could not be made and the retrieval would provide a perfect a posteriori estimate of the true state of the atmosphere accounting for correct background and observational information. However, as is always the case the actual retrieval will have to be tuned to some degree. In particular, one wants to minimize the risk of artifacts in the background data to affect the final TCWV time series. Such artifacts can include slight discontinuities in the ERA TCWV time series at time steps where new sensors are added to the reanalysis.

With these considerations in mind we have chosen the ERA\_NEW 1D-VAR set-up to be used as the basis for the full time series. The modified background error matrix allows for large deviations from the background profile; i.e. it gives stronger weight to the observations. At the same time, it provides good convergence over the entire range of variability of TCWV and allows for a high number of converged profiles, therefore providing very little sensitivity to the choice of the background profile. In particular, the choice of the background profile is uncritical as long as it represents the general conditions for the geographical region and season.

*Data availability.* Data are publicly available under doi:10.5676/DWD\_EMIR/V001 (Bennartz et al., 2016).

*Competing interests.* The authors declare that they have no conflict of interest.

Acknowledgements. This work was funded under ESA's Long-Term Data Preservation (LTDP) program. Partial funding was also provided by Vanderbilt University via a start-up grant to the first author.

Edited by: S. J. Munchak Reviewed by: two anonymous referees

#### References

- Ablain, M., Cazenave, A., Valladeau, G., and Guinehut, S.: A new assessment of the error budget of global mean sea level rate estimated by satellite altimetry over 1993–2008, Ocean Sci., 5, 193–201, doi:10.5194/os-5-193-2009, 2009.
- Bennartz, R., Watts, P., Meirink, J. F., and Roebeling, R.: Rainwater path in warm clouds derived from combined visible/nearinfrared and microwave satellite observations, J. Geophys. Res., 115, D19120, doi:10.1029/2009jd013679, 2010.
- Bennartz, R., Stengel, M., Schroeder, M., Hollmann, R., Picard, B., and Fell, F.: Total column water vapour from the Microwave Radiometer (MWR) instruments onboard ERS-1, ERS-2, and EN-VISAT, doi:10.5676/DWD\_EMIR/V001, 2016.
- Bretherton, C. S., Peters, M. E., and Back, L. E.: Relationships between water vapor path and precipitation over the tropical oceans, J. Climate, 17, 1517–1528, doi:10.1175/1520-0442(2004)017<1517:rbwvpa>2.0.co;2, 2004.
- Byun, S. H. and Bar-Sever, Y. E.: A new type of troposphere zenith path delay product of the international GNSS service, J. Geodesy, 83, 367–373, 10.1007/s00190-008-0288-8, 2009.
- Cess, R. D., Potter, G. L., Blanchet, J. P., Boer, G. J., Delgenio, A. D., Deque, M., Dymnikov, V., Galin, V., Gates, W. L., Ghan, S. J., Kiehl, J. T., Lacis, A. A., Letreut, H., Li, Z. X., Liang, X. Z., McAvaney, B. J., Meleshko, V. P., Mitchell, J. F. B., Morcrette, J. J., Randall, D. A., Rikus, L., Roeckner, E., Royer, J. F., Schlese, U., Sheinin, D. A., Slingo, A., Sokolov, A. P., Taylor, K. E., Washington, W. M., Wetherald, R. T., Yagai, I., and Zhang, M. H.: Intercomparison and interpretation of climate feedback processes in 19 atmospheric general-circulation models, J. Geophys. Res.-Atmos., 95, 16601–16615, doi:10.1029/JD095iD10p16601, 1990.
- Deblonde, G. and English, S.: One-dimensional variational retrievals from SSMIS-simulated observations, J. Appl. Meteorol., 42, 1406–1420, doi:10.1175/1520-0450(2003)042<1406:ovrfso>2.0.co;2, 2003.
- Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally,

A. P., Monge-Sanz, B. M., Morcrette, J. J., Park, B. K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J. N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Q. J. Roy. Meteor. Soc., 137, 553–597, doi:10.1002/qj.828, 2011.

- Eymard, L., Obligis, E., Tran, N., Karbou, F., and Dedieu, M.: Long-term stability of ERS-2 and TOPEX microwave radiometer in-flight calibration, IEEE T. Geosci. Remote, 43, 1144–1158, doi:10.1109/tgrs.2005.846129, 2005.
- Forster, P., Ramaswamy, P., Artaxo, P., Forster, P., and Ramaswamy, V.: Changes in Atmospheric Constituents and Radiative Forcing, Climate Change 2007: The Physical Basis. Contributions of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K. B., Tignor, M., and Miller, H. L., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2007.
- GCOS-107: GCOS-107: Systematic observation requirements for satellite-based products for climate, WMO/Global Climate Observing System (GCOS), 91, 2006.
- Gilbert, L.: REAPER Product Handbook for ERS Altimetry Reprocessed Products, University College London, London, UKREA-UG-PHB-7003, Issue 3.1, available at: https://earth.esa. int/documents/700255/2058345/Reaper-Product-Handbook-3. 1.pdf/1dee3e8d-6765-4a28-a98c-7cf498a3868a (last access: 8 April 2017), 2014.
- Greenwald, T. J.: A 2 year comparison of AMSR-E and MODIS cloud liquid water path observations, Geophys. Res. Lett., 36, L20805, doi:10.1029/2009gl040394, 2009.
- Hocking, J., Rayer, P., Saunders, R., Matricardi, M., Geer, A., and Brunel, P.: RTTOV v10 Users Guide, EUMETSAT Satellite Application Facility on Numerical Weather Prediction, 2011.
- Legeais, J.-F., Ablain, M., and Thao, S.: 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, 2014.
- Mangum, J.: Atmospheric Refractive Signal Bending and Propagation Delay, National Radio Astronomy Observatory (NRAO), Charlottesville, VA 22903-2475, 2009.
- Ollivier, A. and Guibbaud, M.: RA-2/MWR ocean data validation and cross-calibration activities, Yearly report 2013, Technical Note CLS.DOS/NT/13.290, SALP-RP-MA-EA-22293-CLS, 2014, avialable at: http://www.aviso.altimetry.fr/fileadmin/documents/calval/ validation\_report/EN/annual\_report\_en\_2013.pdf (last access: 8 April 2017), 2014.
- Phalippou, L.: Variational retrieval of humidity profile, wind speed and cloud liquid-water path with the SSM/I: Potential for numerical weather prediction, Q. J. Roy. Meteor. Soc., 122, 327–355, doi:10.1002/qj.49712253002, 1996.
- Rodgers, C. D.: Inverse Methods for Atmospheric Sounding: Theory and Practice, World Scientific, Singapore, River Edge, N.J., 2000.
- Saunders, R., Rayer, P., Brunel, P., von Engeln, A., Bormann, N., Strow, L., Hannon, S., Heilliette, S., Liu, X., Miskolczi, F., Han, Y., Masiello, G., Moncet, J. L., Uymin, G., Sherlock, V., and Turner, D. S.: A comparison of radiative transfer models for simulating Atmospheric Infrared Sounder (AIRS) radiances, J. Geophys. Res.-Atmos., 112, D01s90, 10.1029/2006jd007088, 2007.

- Schröder, M., Jonas, M., Lindau, R., Schulz, J., and Fennig, K.: The CM SAF SSM/I-based total column water vapour climate data record: methods and evaluation against re-analyses and satellite, Atmos. Meas. Tech., 6, 765–775, doi:10.5194/amt-6-765-2013, 2013.
- Schroeder, M., Lockhoff, M., Forsythe, J. M., Cronk, H. Q., Vonder Haar, T. H., and Bennartz, R.: The GEWEX water vapor assessment: Results from intercomparison, trend and homogeneity analysis of total column water vapour, J. Appl. Meteorol., 55, doi:10.1175/JAMC-D-15-0304.1, 2016.
- Shoji, Y.: A Study of Near Real-time Water Vapor Analysis Using a Nationwide Dense GPS Network of Japan, J. Meteorol. Soc. Jpn., 87, 1–18, doi:10.2151/jmsj.87.1, 2009.
- Trenberth, K. E., Fasullo, J., and Smith, L.: Trends and variability in column-integrated atmospheric water vapor, Clim. Dynam., 24, 741–758, doi:10.1007/s00382-005-0017-4, 2005.

- Wang, J. H., Zhang, L. Y., Dai, A., Van Hove, T., and Van Baelen, J.: A near-global, 2-hourly data set of atmospheric precipitable water from ground-based GPS measurements, J. Geophys. Res.-Atmos., 112, D11107, doi:10.1029/2006jd007529, 2007.
- Ware, R. H., Fulker, D. W., Stein, S. A., Anderson, D. N., Avery, S. K., Clark, R. D., Droegemeier, K. K., Kuettner, J. P., Minster, J. B., and Sorooshian, S.: SuomiNet: A real-time national GPS network for atmospheric research and education, B. Am. Meteorol. Soc., 81, 677–694, doi:10.1175/1520-0477(2000)081<0677:sarngn>2.3.co;2, 2000.
- Wentz, F. J. and Spencer, R. W.: SSM/I rain retrievals within a unified all-weather ocean algorithm, J. Atmos. Sci., 55, 1613–1627, 1998.