Retrieval of daytime total columnar water vapour from MODIS measurements over land surfaces

H. Diedrich, R. Preusker, R. Lindstrot, and J. Fischer
Institut für Weltraumwissenschaften, Freie Universität Berlin, Carl-Heinrich-Becker-Weg 6–10, 12165 Berlin, Germany

Correspondence to: H. Diedrich (hannesd@wew.fu-berlin.de)

Abstract. A retrieval of total column water vapour (TCWV) from MODIS (Moderate-resolution Imaging Spectroradiometer) measurements is presented. The algorithm is adapted from a retrieval for MERIS (Medium Resolution Imaging Spectrometer) from Lindstrot et al. (2012). It obtains the TCWV for cloud-free scenes above land at a spatial resolution of 1 km × 1 km and provides uncertainties on a pixel-by-pixel basis. The algorithm has been extended by introducing empirical correction coefficients for the transmittance calculation within the forward operator. With that, a wet bias of the MODIS algorithm against ARM microwave radiometer data has been eliminated. The validation against other ground-based measurements (GNSS water vapour stations, GUAN radiosondes, and AERONET sun photometers) on a global scale reveals a bias between −0.8 and −1.6 mm and root mean square deviations between 0.9 and 2 mm. This is an improvement in comparison to the operational TCWV Level 2 product (bias between −1.9 and −3.2 mm and root mean square deviations between 1.9 and 3.4 mm). The comparison to MERIS TCWV for an example overpass exposes a systematic dry bias.

1 Introduction

1.1 Background

The bulk of water contained in the Earth’s atmosphere exists in the form of water vapour. It is the primary greenhouse gas of the Earth-atmosphere system and plays an important role for the exchange of energy through the vertical and horizontal transport of latent heat. Moreover, the geographical distribution and movement of water vapour determines the distribution of clouds and occurrence of precipitation on Earth. Water vapour is important e.g. for driving weather systems; on long time scales, the amount of water vapour in the atmosphere is highly relevant for the evolution of the global climate. As a consequence, the Global Climate Observing System (GCOS) declared the total column of water vapour (TCWV) as an essential climate variable, with the defined goal of providing long time series of TCWV in sufficiently high resolution to enable the determination of both local and global trends (GCOS, 2010). Water vapour remote sensing over the oceans has been done since the 1980s with microwave radiometers (MWR) (e.g. with SSM/I: Schuelessel and Emery, 1990); these instruments have low spatial resolution compared to visible (VIS) and near-infrared (NIR) imagers. Trenberth et al. (2005) found a positive trend of 0.41 mm decade⁻¹ for the TCWV over the ocean between 1988 and 2003.

Detecting TCWV over land, however, is a rather challenging task because of the high heterogeneity of the (unknown) surface properties. There are some existing TCWV retrieval schemes using radiance measurements from satellites in the VIS (e.g. with GOME: Noël et al., 2002; Wagner et al., 2006; OMI: Wagner et al., 2013; or SCIAMACHY: Schrijver et al., 2009), the infrared (IR) (e.g. with IASI; Pougatchev et al., 2009; Wiegele et al., 2014), or in the NIR (e.g. with GOSAT: Frankenberg et al., 2013; or MERIS (Medium Resolution Imaging Spectrometer): Lindstrot et al., 2012). The region between 0.9 and 1 µm, called the ρστ band, is suitable for water vapour remote sensing due to the fact that all surface types are sufficiently bright (> 0.1).

Lindstrot et al. (2012) introduced a procedure to retrieve TCWV for cloud-free scenes for MERIS measurements with error estimates on a pixel-by-pixel basis. It is based on the
evaluation of the differential absorption using a band in the absorption region and one close by with only little absorption features. In order to provide a data set with global coverage and high quality over both land and ocean, a combined data set of MERIS and SSM/I retrievals was generated in the framework of the ESA Data User Element GlobVapour project (Lindstrøt et al., 2014). Unfortunately, contact with ENVISAT (ENVironmental Satellite) was lost in April 2012 and the TCWV time series applying MERIS measurements was interrupted.

Ocean and Land Color Instrument (OLCI), the follow-on to MERIS, is not going to be in space before 2015. An alternative, gap-filling TCWV data set could be the MOD05 Level2 data set from NASA (National Aeronautics and Space Administration). It uses measurements in the NIR from MODIS (MODe-rate Resolution Imaging Spectroradiometer) to retrieve TCWV (Gao and Kaufman, 2003). This operational product has been provided since 1999. Unfortunately, the accuracy of this product is limited and it has not been improved since then. Here, we validated the data with global ground-based measurements for the first time. Figure 1 shows the comparison of MOD05 TCWV values to MWR and GNSS (Global Navigation Satellite Syste) stations (see Sect. 3.4 for detailed description of the validation data sets). It reveals that the algorithm overestimates the TCWV by around 20%, which is very unsatisfactory considering the desire of determining small trends in the TCWV field.

As a consequence, we adapted the procedure of Lindstrøt et al. (2012) to the MODIS band setup (Sect. 1.2) in order to produce a precise TCWV data set that can fill the gap between MERIS and OLCI. The new retrieval has some improvements and advantages compared to the operational L2 algorithm of Gao and Kaufman (2003):

1. The temperature and humidity profile is not fixed. The atmospheric transmittance for each pixel is calculated from a mixed profile. This reduces the uncertainty of the forward operator significantly.

2. The procedure regards scattering processes on aerosols and its interaction with water vapour.

3. The TCWV is derived with the help of an inverse modelling scheme. Deviations between modelled and measured radiances in the three absorption bands are iteratively optimized with the Newton method by changing the water vapour amount. The sensitivity of each band is calculated for each retrieval step, not by fixed weighting functions.

4. Uncertainty estimates considering all error influences are given on a pixel-by-pixel basis.

However, an empirical factor in the forward operator accounting for as small wet bias is needed (Sect. 4).

1.2 The Moderate Resolution Imaging Spectroradiometer (MODIS)

MODIS is a 36-band scanning radiometer covering the spectral range between 0.4 and 14.4 μm and with a spatial resolution between 250 and 1 km depending on the band. It is mounted on both polar orbiting Earth Observing System platforms Terra (morning transit: 10:30 a.m.) and Aqua (afternoon transit: 1:30 p.m.). A two-sided paddle-wheel mirror scans in a field of view of 110° and with a swath of 2330 km. Thus, global coverage can be provided in 2 days. MODIS bands are located on four separate focal plane assemblies depending to their spectral positions and aligned in cross-track direction. Detectors of each spectral band are aligned in the along-track direction. Ten detectors, each with slight differences of their relative spectral response, scan the Earth simultaneously with a nadir spatial resolution of 1 km × 1 km per pixel in the NIR. Five bands in the NIR region between 0.8 and 1.3 μm are used for the TCWV retrieval (Table 1 and Fig. 2). The bands 2 and 5 (865 and 1240 nm) are located in regions with hardly any water vapour absorption features and are usually used for the remote sensing of vegetation and clouds. In the TCWV retrieval, these bands are used to estimate the surface reflectance in the ρστ band. Bands 17, 18, and 19 (905, 936, 940 nm) are water vapour absorption bands with different strength of absorption. Absorption is more pronounced in band 18 and is therefore still sensitive to small
Table 1. Position and widths of five NIR bands (original MODIS design specification) used for the TCWV retrieval (Xiong and Barnes, 2006).

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>865</td>
<td>40</td>
<td>201</td>
</tr>
<tr>
<td>5</td>
<td>1240</td>
<td>20</td>
<td>74</td>
</tr>
<tr>
<td>17</td>
<td>905</td>
<td>30</td>
<td>167</td>
</tr>
<tr>
<td>18</td>
<td>936</td>
<td>10</td>
<td>57</td>
</tr>
<tr>
<td>19</td>
<td>940</td>
<td>50</td>
<td>250</td>
</tr>
</tbody>
</table>

TCWV values, while the weak absorption band 17 is sensitive to high TCWV values without being saturated.

2 Physical background of the retrieval method

Water vapour has various absorption features in the solar and terrestrial spectrum which are due to a combination of the three fundamental vibration modes of the water molecule. Measurements of reflected sunlight in these absorption bands enables a determination of TCWV provided that the following conditions are given:

1. Solar radiation is available, limiting the retrieval to daytime measurements.
2. The band used is located in a sufficiently sensitive part of the spectrum and is not saturated.
3. The surface albedo is sufficiently high and can be accurately estimated, precluding dark ocean surfaces.
4. The photon paths through the atmosphere and the reflection function of the surface are known.
5. The lower troposphere, holding the main part of the TCWV, is not masked by clouds or optical thick aerosol layers.

The NIR MODIS bands (Table 1 and Fig. 2) are perfectly suited for daytime, cloud-free retrieval of TCWV over land. At this spectral range, almost all surfaces provide a sufficiently bright background. The retrieval is based on the differential absorption technique (Gao et al., 1993; Bartsch et al., 1996; Albert et al., 2001, 2005). The basic principle of the method is the comparison of the measured radiance in an absorption band to a close-by band with no or few absorption features.

Following Hansen and Travis (1974) and Fraser et al. (1992), the radiance at the top-of-atmosphere (TOA), \( L_{\text{TOA}}(\lambda) \), at a certain wavelength or band can simplified to

\[
L_{\text{TOA}}(\lambda) = E_0(\lambda) \cdot \alpha(\lambda) \cdot T_{\text{noscat}}(\lambda) \cdot \cos(\theta_s)/\pi + L_{\text{path}}(\lambda),
\]

where \( \lambda \) is the wavelength, \( E_0(\lambda) \) is the TOA solar flux (solar constant at the wavelength), \( T_{\text{noscat}}(\lambda) \) is the total atmospheric transmittance without scattering, \( \alpha(\lambda) \) is the surface reflectance, \( \theta_s \) is the solar zenith angle, and \( L_{\text{path}}(\lambda) \) is the radiance from scattering along the light path.

For monochromatic radiation, neglecting scattering processes along the photon path, the transmittance \( T \) through the atmosphere can be related to its optical depth \( \tau \) and the air mass \( \mu = 1/\cos(\theta_s) \), following the Beer–Lambert law:

\[
T = \exp(-\tau/\mu).
\]

The optical depth of a medium is a measure of mass of absorbing and scattering species on the photon path in the considered band. The depth and width of the individual water vapour absorption lines result from pressure- and temperature-dependent broadening processes. Consequently, the knowledge of the actual temperature profile and the surface pressure is necessary in order to simulate the correct atmospheric transmittance. Lindstrot et al. (2012) stated that the error of a TCWV retrieval can be significantly reduced if

1. the surface pressure reduction due to the surface elevation is accounted for, instead of taking a standard value, or
2. the surface temperature from reanalysis data is used to approximate the transmittance corresponding to the actual temperature profile by adequately mixing the pre-calculated transmittance values corresponding to the two closest standard profiles (Lindstrot and Preusker, 2012).

In Eq. (1), \( L_{\text{path}}(\lambda) \) accounts for the shortening of the photon path because of scattering from aerosols and is usually only a few percent of the direct reflected solar radiation in the NIR region. In the retrieval of TCWV a scattering factor \( f \) accounts for this effect. It is defined as

\[
f = T / T_{\text{noscat}},
\]

where \( T \) is the true atmospheric transmittance including scattering and \( T_{\text{noscat}} \) the atmospheric transmittance in case of
pure absorption only by water vapour. \( f \) is usually larger than one, as atmospheric scattering causes a shortening of the average photon path length and reduces the amount of TCWV by preventing a fraction of photons from traversing the humid lower troposphere. Additionally, \( f \) increases above dark surfaces, as the majority of the photons are reflected by atmospheric scatterers and thus do not travel through the whole vertical column of water vapour. One important parameter determining \( f \) is thus the surface reflectance (see Lindstrot et al., 2012, for detailed discussion concerning \( f \)).

3 1D-Var retrieval algorithm

Our algorithm uses the information from three absorption and two framing window bands without water vapour absorption. This leads to some changes in procedure in comparison to Lindstrot et al. (2012):

- TOA radiances are simulated for each absorption band and then compared to the measurements instead of a transmittance ratio (Sect. 3.1)
- in order to include information from all three absorption bands and to account for the measurement error in the inversion scheme, the inversion technique presented in Sect. 3.2 is used instead of applying the secant method
- the surface reflectance in each band is now determined from an interpolation between the positions of the window bands instead of an extrapolation from the 800–900 nm region
- the aerosol optical depth (AOD) is extracted from MODIS L2 data instead of taking a climatological mean value

Additionally, we regarded more error influences in the uncertainty estimation (Sect. 3.3)

3.1 Forward model

The forward model simulates the TOA radiances in the absorption bands. The introduction of the scattering factor \( f \) simplifies Eq. (1):

\[
L_{\text{TOA}}(\lambda) = E_0(\lambda) \cdot \alpha(\lambda) \cdot T_{\text{noscat}}(\lambda) \cdot f \cdot \cos(\theta_s)/\pi.
\]

As \( \theta_s \) and \( E_0(\lambda) \) are known, the following values have to be derived in the forward operator: the surface reflectance \( \alpha \), the atmospheric transmittance \( T_{\text{noscat}} \), and the scattering correction factor \( f \) at the corresponding wavelength.

The pure absorption part of the simulated transmittance \( T_{\text{noscat}} \) is derived from pre-calculated absorption coefficients using an advanced \( k \) distribution method (Bennartz and Fischer, 2000). The coefficients are calculated from the HITRAN2008 line database (Rothman et al., 2010) using the AER LBLRTM code (Clough et al., 2005). The optical depth values are stored in lookup tables on 27 different pressure levels for 6 standard profiles (McClatchey et al., 1972). The transmittance is calculated for the four lookup table grid points closest to the actual surface pressure and temperature of the considered scene, thereby assuming that the surface temperature is sufficient correlated with the actual vertical temperature profile. The actual surface temperature is taken from numerical weather prediction (NWP) reanalysis data (ERA interim 2 m temperature; http://www.ecmwf.int/research/era/do/get/era-interim). The actual surface pressure is derived from converting land elevation to pressure using the GTOPO30 digital elevation model (US Geological Survey, 1996). In order to obtain the surface reflectance in the corresponding absorption band, the two window bands are initially corrected for the influence of scattering and the small but significant influence of water vapour absorption (Fig. 2) with a lookup table approach. This atmospheric correction requires knowledge about the aerosol loading, type, and vertical distribution. However, over bright land surfaces the influence of aerosols on the retrieved surface reflectance is weak. The aerosol optical depth is extracted from MODIS L2 aerosol data (MOD04). For missing values, a climatological standard value is taken (AOD\(_{550} = 0.1\)).

Afterwards, surface reflectance of the absorption band is linearly interpolated from the window bands. This requires the assumption that the surface reflectance changes linearly with wavelength in this spectral range, which is true in most cases as shown in Gao and Kaufman (2003). This is certainly a source of error but is accounted for in the uncertainty estimation. The correction for the water vapour absorption in the window bands is done by applying the first guess (Sect. 3.2).

The scattering correction factor \( f \) has been calculated beforehand from radiative transfer simulations and stored in lookup tables presuming a continental aerosol layer (Hess et al., 1998) with an exponential increase from 1000 m to the bottom. For this reason the Matrix Operator MODel (Fischer and Grassl, 1984; Fell and Fischer, 2001; Hollstein and Fischer, 2012; Doppler et al., 2014) was used to derive TOA radiances for different TCWV, AOD, and Lambertian surface reflectance values, sun zenith angles, viewing zenith angles, and relative azimuth angles. The phase functions were calculated with the Mie code based on Wiscombe (1980).

3.2 Inversion technique

Only one state vector variable has to be found in the iterative optimization routine: the total column water vapour. Starting with the first guess, TCWV is adapted by minimizing the differences between simulated and measured radiances. The TCWV value for the next iteration step is derived by the following scheme after Rodgers (2000):

\[
G = (K^T S_c^{-1} K)^{-1} (K^T S_c^{-1} y),
\]

\[
x_{i+1} = x_i + (G(y - F_x)),
\]
where $K$ is the Jacobian matrix that contains the partial derivatives of the radiance to the TCWV value in each band, $S_x$, the measurement error covariance matrix that contains radiance scaled with the signal-to-noise ratio (SNR) (in the diagonal cells, assuming uncorrelated errors) for each band, and $y$ contains the measured and $F_i$ the modelled radiances. The first guess of TCWV $f g$ is obtained from a simple regression, relating TCWV $f g$ to a third-order polynomial of $\ln(T_{\text{noscat}})$ for band 17:

$$f g = 2/\mu \cdot c_1 - \ln(T_{17}) \cdot c_2 + (\ln(T_{17}))^2 \cdot c_3,$$

(7)

where $\mu$ is again the air mass. The regression coefficients $(c_i)$ were determined using the absorption forward operator (Table 2).

### 3.3 Uncertainty estimate

After the iteration procedure, the retrieval uncertainty is calculated, taking into account the following sources of uncertainty:

- residual model error
- instrument uncertainty (SNR)
- uncertainty of the aerosol optical depth
- uncertainty due to the missing information of the aerosol type and scale height.
- uncertainty of the surface pressure and temperature
- uncertainty due to the missing information about the true temperature profile
- uncertainty due to the estimation of the surface reflectance and its spectral slope

For the error quantification, these model parameter uncertainties assembled in the error covariance matrix $S_b$ are propagated into the measurement space using the standard error propagation and added to the measurement error covariance matrix $S_c$:

$$S_y = S_b + K_b^T S_b K_b,$$

(8)

where $K_b$ is the parameter Jacobian. The resulting error covariance matrix $S_y$ is then propagated into the state vector space using the Jacobian $K$. The resulting error covariance matrix $\hat{S}$ is a direct measure of uncertainty in TCWV space (Rodgers, 2000):

$$\hat{S} = (K^T S_y^{-1} K)^{-1}.$$

(9)

In the following, it is described how the individual error sources are estimated. As outlined in Sect. 2, the scattering factor $f$ is affected most by the surface reflectance, aerosol height, and aerosol optical thickness. For each of these parameters, a perturbed $f^*$ is calculated from the lookup tables by perturbing the input accordingly. There is no information available about the aerosol scale height and the type (size distribution, absorption, and scattering properties). Consequently, a $f^*$ was calculated assuming an aerosol layer at 6000 m with a thickness of 500 m. Additionally, a $f^*$ was calculated from simulations supposing another aerosol model. These $f^*$ were used to derive perturbed TOA radiances $L_{\text{TOA}}^*$. Finally, the difference, $(\Delta L)^2 = (L_{\text{TOA}} - L_{\text{TOA}}^*)^2$, is added to the measurement error variance $S_x$.

The error due to differences between the simulation and the real temperature (and humidity) profile was evaluated by comparing the atmospheric transmittances derived from a real example radiosonde profile to transmittance using the standard profiles. This effect introduces an error to the pure absorption transmittance of around 2 % (depending on the band). To estimate the uncertainty due to the surface background information, the surface temperature was shifted by 5 K and the surface pressure was perturbed by 20 hPa and subsequently committed to the transmittance forward operator. Again, $(\Delta L)^2$ is calculated and added to $S_b$.

The spectral dependency of the surface reflectance is parametrized with the Normalized Differenced Vegetation Index (for further details see Lindstrot et al., 2012). The uncertainties of the surface reflectance range from 0.5 to 1.2 %. Similar to the approach pictured above, a perturbed TOA radiance is calculated and the resulting deviation is contributed to $S_b$. Finally, the residual model error, that is the difference between measurement and modelled radiance from the last iteration step is added to the measurement covariance matrix that consists of the sensor noise (Table 1).

### 3.4 Validation data sets

The MODIS TCWV retrieval was validated against different ground-based measurements such as MWR data, GNSS water vapour monitoring data, GCOS Upper Air Network (GUAN) radiosonde data, and Aerosol Robotic Network (AERONET) sun photometer data.

#### 3.4.1 ARM microwave radiometer

A data set of ground-based MWR (software version 4.13) of three ARM (Atmospheric Radiation Measurement) sites (north slope of Alaska, NSA, Southern Great Plains, SGP; tropical western Pacific, TWP) for the years between 2002 and 2012 was used for the assessment of the TCWV retrieval.
Figure 3. Normalized frequencies of occurrence for comparisons of TCWV retrieved from the MODIS retrieval using just one absorption band to MWR TCWV data in mm; see text for detailed description.

MWR instruments measure the radiation emitted by the atmospheric water vapour and liquid water at frequencies of 23.8 and 31.4 GHz (Turner et al., 2007). The background of the measurement is the cosmic background temperature. Consequently, it is one of the most accurate methods to determine the TCWV from ground. The uncertainty of the measured TCWV from MWR is expected to be in the range of 0.3 mm (Turner et al., 2003). This data set was used to calculate the correction coefficients (Sect. 4).

3.4.2 Ground-based GNSS

The global 2-hourly GNSS TCWV data set is based on three different resources: the International GNSS Service, US SuomiNet (UCAR/COSMIC) products, and Japanese GEONET data (Wang et al., 2007; NCAR, 2011). Data from 942 stations for the years between 2003 and 2011 were extracted. The uncertainty of these data is not precisely stated by the authors but a similar data set provides an accuracy of 1–2 mm (Gendt et al., 2004).

3.4.3 GUAN radiosonde

A global data set of TCWV derived from GUAN radiosondes, distributed via the Ground Tracking System network and extracted from the DWD (Deutscher Wetterdienst) archive, was compared to the MODIS TCWV retrieval for the period 2003–2005. As radiosondes do not measure the whole vertical column at once, the accumulated TCWV has a relatively high uncertainty which can range between 1 and 10% (Turner et al., 2003).

3.4.4 AERONET sun photometer

A global set of TCWV values from AERONET sun photometer measurements from the years between 2003 and 2014 was used for validation (Direct Sun Algorithm version 2, downloaded from http://aeronet.gsfc.nasa.gov; Bruegge et al., 1992; Reagan et al., 1986). Pérez-Ramírez et al. (2014) stated that the retrieval produces TCWV values with a consistent dry bias of approximately 5–6% and an estimated uncertainty of 12–15%.

Figure 4. Simulated versus measured atmospheric transmittances due to water vapour for each MODIS band using the same data as in Fig. 3; see Sect. 4 for detailed description of the compared variables.

4 Correction

For reasons of clarity and comprehensibility, unless noted otherwise, only data of MODIS from the Aqua platform are used in the following. Nevertheless, the retrieval and the validation were also applied to Terra data. In order to evaluate the performance of the retrieval, TCWV values were compared to ground-based MWR measurements. These data were considered as ground truth and compared to retrieved TCWV from collocated MODIS scenes. In order to assess the behaviour of each band, a one-band retrieval has been established that iteratively fits the simulated radiance to the measured radiance for just one band, using the same architecture and lookup tables as the three-band algorithm. The comparison of MODIS-derived TCWV to MWR data reveals different deviations in each band (Fig. 3). On the one hand, the two data sets correlate linearly and the scattering is very low. On the other hand, there are significant differences: maximal (around 10%) in the case of the retrieval using just band 18 and minimal (around 3%) in case of band 17. When using bands 18 or 19 the retrieval overestimates the TCWV, whereas band 17 has a small dry bias. These deviations add up to a wet bias in the three-band retrieval of −0.6 mm (Fig. 7). Reasons for this could be:

a. a systematic error of the MWR
b. a wrong spectral calibration of the MODIS bands
c. errors in the forward model.

First, the MWR data are very precise: Turner et al. (2003) quantified a measurement uncertainty of 0.3 mm). Second, a bias in the MWR data would introduce the same TCWV shift to every band. Consequently, the MWR error can not explain the different signs.

MODIS features a Spectroradiometric Calibration Assembly (SRCA) that calibrates all bands on orbit and keeps track of all radiometric changes and degradation of the optics (Xiong and Barnes, 2006). Xie et al. (2006) stated an uncertainty of the central wavelength of maximal 0.1 nm for band 17. Nevertheless, we tested the influence of a shifted central wavelength for all absorption bands as follows. A forward operator for the pure absorption was established which is additionally dependent on the central wavelength. The shape of
the relative response functions and all other modules of the TCWV retrieval remained unchanged. The scattering factor was assumed to be independent of the central wavelength. Afterwards, for each band TOA radiances were simulated taking the TCWV values from the MWR validation database and compared to the measured values.

Figures 5 and 6 show the bias (upper row) and the bias-corrected root mean square deviation (RMSD) (lower row) between simulated and measured radiances as a function of the considered central wavelength for each of the 10 detectors of the three absorption bands (represented by 10 black graphs). The dashed vertical lines indicates the nominal central wavelengths of the bands. The outcomes of this study are as follows:

1. The sensitivities to the central wavelength (here the slope of the curves) differ between the bands as they are located in different parts of the spectrum with different absorption strengths.

2. The sensitivities increase with decreasing bandwidth, as expected.

3. The zero of the bias (indicated by the vertical dotted lines) are shifted dissimilarly in different directions with different distance to the original nominal wavelength. These shifts would be necessary to cancel out all systematic differences between simulated and measured radiances.

4. The shifts range from around 0.5 nm at Terra band 18 to 4 nm at Terra band 19. This exceeds by far the accuracy given by Xie et al. (2006).

5. There are different shifts between MODIS Aqua (Fig. 5) and MODIS Terra (Fig. 6) (e.g. Aqua band 19 and Terra band 19). This implies that the specific sensor characteristics differ and that they could be a reason for the deviations between forward model and measurements although SRCA provides a high accuracy.

6. There are differences in the behaviour of individual detectors, especially in Aqua band 18 and Terra band 19. Here, one detector drops out significantly.

Another source of the differences between the simulated and measured radiances could be the forward operator of the retrieval. As presented in Sect. 2, the backbone of the forward operator is the determination of the atmospheric transmission due to water vapour absorption, $T_{\text{noscat}}$, and the scattering factor $f$.

The latter is primarily dependent on the surface reflectance, which is sufficiently constant for the validation...
data set (0.2–0.3). At the same time, the aerosol properties such as AOD, scale height, and aerosol type vary substantially over the year. First, this would introduce not a systematic bias but an increased scattering. Second, the sensitivities of the AOD and scale height in respect to the simulated TCWV do not explain the deviations: Table 3 shows the influence of a doubled AOD, and Table 4 shows the influence of the aerosol scale height on the simulated TCWV for different surface reflectances. The error due to the different aerosol type is maximal around 0.5 % (not shown here). The sensitivities were calculated by deriving the difference of transmittance between an unperturbed and a perturbed AOD and scale height. The difference in transmittance has been transformed into an equivalent TCWV value using a partial derivative of transmittance with respect to TCWV. Additionally, Diedrich et al. (2013) stated the low impact of the AOD on the error of a TCWV retrieval. However, the scattering–absorption interaction can be over- or underestimated in the radiative transfer simulations. The other potential source of error could be the determination of the absorption coefficients. Here, either the database of absorption coefficients or the binning method could be wrong.

In summary, the origin of the deficiency of the retrieval is not exactly known yet. Hence, we introduce correction coefficients in order to adjust the atmospheric transmittance $T_{\text{noscat}}$ in the optical thickness space. The corrected transmittance $T_{\text{corr}}$ is then calculated by

$$T_{\text{corr}} = \exp(a + b \cdot \ln(T_{\text{noscat}})).$$

(10)

The coefficients $a$ and $b$ were obtained by optimizing the difference between the simulated and measured atmospheric transmittance $(T_{\text{corr}} - T_{\text{meas}})$ using the MWR validation data set as a reference (Fig. 4). The uncorrected transmittance $T_{\text{noscat}}$ was calculated by the absorption forward operator, taking viewing geometries from MODIS and the TCWV information from the corresponding MWR measurement. $T_{\text{meas}}$ was derived by

$$T_{\text{meas}}(\lambda) = L_{\text{meas}}(\lambda)/L_0(\lambda),$$

(11)

where $L_{\text{meas}}$ is the measured radiance in the absorption band. $L_0$ is the radiance presuming no absorption of water vapour. It is derived from interpolating the measured radiances (corrected for water vapour) in the window bands onto the position of the absorption band:

$$L_0(\lambda) = [(L_2/T_2 - L_5/T_5)/(\lambda_2 - \lambda_5)] \cdot (\lambda - \lambda_2) + L_2/T_2.$$

(12)

This was done for each band and for MODIS Aqua and MODIS Terra. The derived correction coefficients can be seen in Tables 5 and 6 for Aqua and Terra instruments respectively. In the retrieval, the transmittance is always corrected after Eq. (10). Figure 8 shows the flow of the correction and validation process. As a consequence, the difference of ground-based measurements is reduced to a minimum, as can be seen in the next section.

## 5 Validation

The validation results are shown in Fig. 10. In each plot, the normalized relative frequency of occurrence is shown with high occurrences plotted in red to yellow and low occurrences in dark blue. In the top left corner of each plot, the offset and slope of the linear regression, the bias-corrected root mean square deviation, the bias, the correlation coefficient, and the sample size as number of points are given.

The cloud mask from the MOD05 L2 data, which was extracted from the MOD35 L2 product, was used to filter out cloudy and cloud-contaminated pixels. Only pixels with bit value “100 % clear” were used for the study. MODIS measurements were spatially averaged over an area of 20 × 20 km$^2$ (20 × 20 pixels) to account for e.g. the radiosonde displacement and the time gap between satellite
overpass and ground-based measurements. The location of all considered sites from the different data sets are displayed in Fig. 9.

5.1 ARM MWR

In the top left panel in Fig. 10, the comparison of the corrected retrieval and the MWR is shown where only cases with 100% valid MODIS pixels were considered to exclude the influences of clouds. As expected in Sect. 4, both data sets show almost perfect agreement, although the number of samples is relatively low due to cloud contamination at the NSA and TWP sites. Thus, SGP provides 70% of the number of points in the upper left panel in Fig. 10. Nevertheless, the data set is representative of global observations because dry northern, wet tropical, and mid-latitude conditions are contained.

5.2 Ground-based GNSS

We extracted GNSS TCWV data by taking all collocations with a time difference less than 1 h between measurement and satellite overpass. Here, again, only cases with 100% valid MODIS pixels were considered. The upper right panel in Fig. 10 shows the comparison between the GNSS- and MODIS-derived TCWV values. Although the filtering was very strict, the number of samples is still very high due to the high number of GNSS stations. The bias is −0.8 mm and the RMSD is 1.9 mm, indicating a slight overestimation of MODIS TCWV values.

Figure 11 shows the bias between the two data sets for each station on a world map. Generally, the biases are low (around 1 mm). The majority of stations has a negative bias, meaning that MODIS TCWV values are on average larger than the GNSS values, which corresponds to Fig. 10. Although GNSS stations are not distributed equally over the
Figure 9. Geographic distribution of used validation data.

Figure 10. Normalized frequencies of occurrence for the comparison of the TCWV retrieval to ground-based microwave radiometers (upper left panel) and global ground-based GNSS data (upper right panel), GUAN radiosonde data (lower left panel), and AERONET sun photometer data (lower right panel). For detailed description see Sects. 3.4 and 5.

5.3 GUAN radiosonde

In order to account for the displacement of the radiosonde during its ascent, only cases with a time difference of maximum 2 h between radiosonde and MODIS measurement were considered. In order to account for cloud contamination but also preserve a sufficient high number of data points, only cases with less than 50 % valid MODIS pixels were rejected. This is presumably the reason for the large scatter in the lower left panel of Fig. 10. A bias of $-1.6 \text{ mm}$ and a RMSD of $3.1 \text{ mm}$ are in the range of the radiosonde uncertainty (Turner et al., 2003; Miloshevich et al., 2004).

5.4 AERONET sun photometer

On the lower right panel of Fig. 10 the comparison to the sun photometer measurements is shown. Here, again, only cases with 100 % cloud-free MODIS pixels are taken into account. The agreement between the data sets is very good because the scattering is in the range of the uncertainty of AERONET, although there is a small but significant wet bias. However, this is most likely due to the dry bias in the sun photometer data (Pérez-Ramírez et al., 2014).

5.5 Time dependency

In order to study the temporal constancy of the accuracy of the retrieval, the bias between the retrieved TCWV and GNSS/MWR was calculated as function of time (year) and plotted in Fig. 12. The annual absolute difference between MWR and MODIS (upper panel) is less than 1 mm for both Aqua and Terra and constant over the years, taking into account the scattering represented by the boxes and whiskers. Due to the low number of coincidences, the number of cases per year ranges between 50 and 150.

The GNSS data set provides 10 times more cases (lower panel). Considering the large variance, there is no annual dependency in comparison to GNSS values although the medians tend to be below zero, which is consistent with the negative bias in the scatter plot in Fig. 10. The large area of the whiskers and the large number of outliers in the GNSS data are presumably due to some sources of error in the retrieval of water vapour, such as measurement errors, uncertainties in the mapping functions, and errors due to the assumption of a mean temperature (Wang et al., 2007). These result in deviations that can vary even during 1 day. However, the scattering is not time dependent. Generally, there is no significant difference between Aqua and Terra in both comparisons.
Figure 11. Map of the bias between MODIS TCWV values and GNSS TCWV in mm. Positive values (red circles) imply that TCWV values on that GNSS station are generally larger than the MODIS TCWV values. Blue circles show station where the bias is negative.

5.6 Comparison to MERIS TCWV

Figure 13 shows the comparison of TCWV derived from MODIS and MERIS (Lindstrot et al., 2012) on a regular grid of 0.05° resolution. Each data point is the mean of all valid corresponding sensor pixels. In the left panel, the MODIS overpass over Europe and northern Africa for 2 July 2008 (09:32–09:49 UTC) is presented. On the right-side plot, the corresponding MERIS overpass is shown (09:42–09:59 UTC). In the middle panel the difference between both fields is plotted (MERIS minus MODIS). Only pixels with a valid MERIS and MODIS TCWV value were taken into account. Generally, MERIS has a smaller swath (1150 km) and just covers the west side of the MODIS track. The structures in the TCWV field agree with each other although the location of cloudy pixels is different. However, there are significant differences. MERIS TCWV is systematically higher than MODIS (1 to 3 mm) apart from a small region above the Sahara. Lindstrot et al. (2012) discovered a small wet bias for MERIS in comparison to ground-based TCWV measurements. Whether this or an error in the MODIS retrieval explains the differences will be examined in future studies. Up to now the comparison to ground-based measurements suggests that the MODIS retrieval is not the reason. Furthermore, along-track stripes appear in the difference plot. These features are due to the architecture of the MERIS instrument. It is built out of five individual cameras that have slightly different spectroscopic properties. The central wavelengths of the MERIS channels are viewing angle-dependent. The later indicates that there is still some improvement possible in the MERIS retrieval as well.
6 Summary and outlook

We present a retrieval of TCWV from MODIS measurements for cloud-free land scenes in the near-infrared spectral range. The 1D-Var algorithm is based on a fast forward operator whose basic functionality has been adapted from Lindstrot et al. (2012). Three main sub-procedures derive (1) the surface reflectance, (2) the transmittance due to atmospheric water vapour, and (3) the shortening of the photon path due to atmospheric scattering. A realistic uncertainty estimate is given on a pixel-by-pixel basis where the uncertainties of measurement and forward model are considered respectively.

A validation study against several ground-based reference data sets reveals the high accuracy and precision of the retrieved TCWV values with bias-corrected root mean square deviations between 1 mm (ARM MWR measurements), 2 mm (GNSS and AERONET measurements), and 3.2 mm (GUAN radiosondes). The bias has been reduced to a minimum due to the introduction of correction coefficients for the transmittance calculation within the forward operator (0.1 mm to ARM MWR measurements, −0.8 mm to GNSS measurements, −1.6 mm to GUAN radiosondes). The scattering of the comparison between TCWV values from MODIS and ground-based measurements is in the range of the retrieval uncertainty between 2 and 3 %.

It is intended to use transmission correction not only to increase the accuracy but also to homogenize time series of TCWV over land from different satellites and to use MODIS data as a gap filler between MERIS and OLCI. Due to the fact that the data sets overlap, MERIS and OLCI could be emulated with MODIS in order to intercompare the TCWV time series. A first comparison to MERIS reveals a systematic difference that needs more investigation. L3 data processing and comparison to other space-borne measurements will benefit further studies of the performance of this retrieval.

Acknowledgements. This work was partly performed in the frame of the WadaMo project (DLR – BMWi) and HD(CP)² (BMBF). The authors would like to thank Marc Schröder from DWD for providing the GNSS data.

Edited by: J. Joiner

References

Diedrich, H., Preusker, R., Lindstrot, R., and Fischer, J.: Quantification of uncertainties of water vapour column retrievals...


