Water vapor satellite products in the European Arctic: an inter-comparison against GNSS data

Javier Vaquero-Martínez[a,b,*], Manuel Antón[a,b], Roberto Román[c], Victoria E. Cachorro[c], Huiqun Wang[d], Gonzalo González Abad[d], Christoph Ritter[e,1]

[a] *Departamento de Física, Universidad de Extremadura, Badajoz (Spain)*
[b] *Instituto Universitario de Investigación del Agua, Cambio Climático y Sostenibilidad (IACYS), Universidad de Extremadura, Badajoz (Spain)*
[c] *Grupo de Óptica Atmosférica, Universidad de Valladolid, Valladolid (Spain)*
[d] *Smithsonian Astrophysical Observatory, Cambridge, Massachusetts (United States)*
[e] *Alfred-Wegener-Institut; Telegrafenberg A45, 14473 Potsdam (Germany)*
[f] *Institut für Geophysik und Meteorologie, Universität zu Köln, Pohligstr. 3, 50969 Cologne (Germany)*

Abstract

The European Arctic is a region of high interest for climate change. Water vapor plays a fundamental role in global warming; therefore, high-quality water vapor monitoring is essential for assimilation in forecast simulations. The seven analyzed instruments on-board satellite platforms are: Atmospheric Infrared Sounder (AIRS), Global Ozone Monitoring Instrument 2 (GOME-2), Moderate-Resolution Imaging Spectroradiometer (MODIS), Ozone Monitoring Instrument (OMI), SCanning Imaging Absorption Spectrometer for Atmospheric Carthography (SCIAMACHY) and Polarization and Directionality of the Earth's Reflectances (POLDER). The GNSS data from Ny-Alesund are matched to satellite observations of IWV in a 30-minute temporal window, and 100-km radius. Then, statistics and the distribution of satellite-ground differences under different conditions are studied. The correlation coefficient (R^2) with ground-based measurements is about 0.7 for all products except OMI (R^2=0.5), and MODIS NIR and POLDER (R^2=0.3). OMI shows high bias and variability compared to the rest of products. RMSE values are of the order of 3mm for all satellites, except OMI (7mm) and POLDER (5mm). Bias (MBE) is negligible for

*Corresponding author
Email address: javier_vm@unex.es (Christoph Ritter)

AIRS, close to +1.6mm for GOME-2 and MODIS IR, +0.8mm for MODIS NIR, +5.9mm for OMI, -2.7mm for POLDER and -1.2mm for SCIAMCHY. All satellite products tend to overestimate small IWV values and underestimate large IWV values. Variability also increases with IWV. An underestimation of the satellite products and an increase on the variability is generally observed for large Solar Zenith Angle (SZA) values. Under cloudy conditions, underestimation and variability are increased. Seasonal behavior is driven by the typical cloud cover (CC), SZA, and IWV values. In summer, it is typical to find conditions with large IWV, small SZA and large CC values. Therefore, in summer months satellite products are more biased (either positively or negatively) and with more variability, but in relative terms they are less biased and exhibit less variability.

Keywords: Satellite; Water vapor; GNSS; European Arctic; Ny-Alesund

---

Highlights

- NIR products do not seem adequate for IWV retrieval in the European Arctic.
- VIS and IR products show a fair agreement with GNSS ($R^2 = 0.7$).
- General tendency to overestimation (underestimation) for low (high) IWV.
- GOME-2, AIRS and MODIS-NIR products show dependence with SZA.
- Quality control to avoid cloudy scenes is of the highest importance.

1. Introduction

Among the atmospheric components, water vapor is considered a key one in the climate system: it plays an important role in many processes, such as hydrological cycle, infrared absorption, or energy transportation (Myhre et al., 2013). Changes in the hydrological cycle can affect strongly the living environment of people and overall animals in the Arctic. For instance, it is thought that global

warming will change the hydrological cycle of the Arctic, causing more frequent and more intense rainfall (Bengtsson et al., 2011). Also, increased evaporation and atmospheric moisture amplifies polar warming by reinforcing water vapor and cloud feedbacks (Bintanja et al., 2020; Bogerd et al., 2020). Water vapor presents a positive feedback in the climate system (Colman, 2003, 2015).

Integrated water vapor (IWV), also known as precipitable water vapor (PWV) or total column water vapor (TCWV), is a magnitude equivalent to condensing all the water vapor in the atmospheric vertical column and measuring the height that the liquid water would reach in a vessel of the same cross section as the column (Román et al., 2015). Its units are superficial density (typically g mm^{-2}) or length (typically mm).

Water vapor has been declared an essential climate variable by the Global Climate Observing System (GCOS, 2010), highlighting the importance of high-resolution long time series in order to detect both local and global trends.

Water vapor usually presents a high variability in the spatial and temporal domains. Therefore, it is crucial to use different measurement techniques that provide representative data to have a proper representation of the spatial variability and temporal cycles, as well as redundant information to assess the quality of water vapor measurements. The techniques to gather atmospheric water vapor information are varied. To list some, ground-based techniques include: radiosondes, microwave radiometers, Lidar, photometers, ground-based GNSS. Among the ground-based techniques, radiosondes have traditionally been used as a reference to test other techniques, since radio-sounding is considered a very reliable technique (Antón et al., 2015; du Piesanie et al., 2013; Ohtani & Naito, 2000). However, it has a poor temporal resolution of generally one to four measurements a day, as well as poor spatial resolution (limited number of stations around the world). Also, radiosonde data suffer, in occasions, from systematic observational errors, as well as inhomogeneity and instability (Gaffen, 1994; Wang et al., 2003). As a result, these errors could cause regional biases if only radiosondes were used in the satellite data validation process (Wang & Zhang, 2008, 2009).

In addition, Global Navigation Satellite System (GNSS) meteorology, which is another technique to derive IWV (Bevis et al., 1992) with a high time resolution, has been tested against radio-sounding showing very good precision and accuracy. For instance, in Vaquero-Martínez et al. (2019), radiosonde IWV data were used to validate GNSS measurements at Global Climate Observing System (GCOS) Reference Upper-Air Network (GRUAN) sites, resulting in bias and standard deviations within the submilimeter regime and determination coefficients (R^2) above 0.98. Moreover, Buehler et al. (2012) showed that measurements differ, between GPS and other instruments, like radiosondes or Fourier-Transform infrared spectrometer, in less than 1 mm at high latitude sites. Among its many advantages, it must be noted that ground-based GNSS measurements of IWV are available for all weather conditions, with a high temporal resolution (from 5 min to 2 h), accuracy better than 3 mm and long-term stability (Wang & Zhang, 2008). For these reasons, GNSS data are being used in the last decades as a reference to validate other instruments (Köpken, 2001; Prasad & Singh, 2009; Rama Varma Raja et al., 2008; Román et al., 2015; Vaquero-Martínez et al., 2017b,a, 2018; Guerova et al., 2016; Carbajal Henken et al., 2020; Sakai et al., 2019).

Nevertheless, GNSS still needs a ground-based station to work, resulting in an unequal distribution of stations, and, hence, the inability to cover the oceans. Therefore, the use of remote sensing from satellite sensors is fundamental for many applications (weather forecasts, climate studies, and so on), being capable of providing a snapshot of water vapor worldwide. The available satellite remote sensing techniques using different spectral domains are promising, providing global coverage. Yet, most of these techinques are limited (daytime, only clear skys, and so on).

Particularly, high latitude regions are generally a challenge for satellite-based remote sensing techniques. For instance, the high reflectivity and large solar zenith angles are problematic for near-infrared (NIR) and visible (VIS) sensors (Bedka et al., 2010). The drier conditions due to low temperatures also make necessary to have high precision in the measurements. Furthermore, it is well

known that the Arctic is a hotspot of warming. During the last three decades its climate has suffer dramatic changes. As an example, the annual near-surface temperature increase in the region has been more than twice as large as the global average (IPCC, 2014). Therefore, the Arctic is an area of paramount importance to climate change.

This paper aims to quantify the uncertainties on satellite water vapor products using GNSS water vapor data-set as a common reference. The products validated are those from the following instruments: Atmospheric Infrared Sounder (AIRS), Global Ozone Monitoring Instrument 2 (GOME-2), Moderate-Resolution Imaging Spectroradiometer (MODIS), Ozone Monitoring Instrument (OMI), SCanning Imaging Absorption SpectroMeter for Atmospheric CHartographY (SCIAMACHY) and Polarization and Directionality of the Earth's Reflectances (POLDER). All of them have been previously validated: GOME-2 (Antón et al., 2015; Grossi et al., 2015; Kalakoski et al., 2016; Noël et al., 2008; Román et al., 2015), MODIS (Bennouna et al., 2013; Chang et al., 2015; Gao & Li, 2008; Li et al., 2003; Ningombam et al., 2016; Prasad & Singh, 2009; Vaquero-Martínez et al., 2017a), AIRS (Hagan et al., 2004; Milstein & Blackwell, 2016; Rama Varma Raja et al., 2008), OMI (Wang et al., 2019) ,SCIAMACHY (du Piesanie et al., 2013; Noël et al., 2005; Schrijver et al., 2009), POLDER (Vesperini et al., 1999). However, not many validation exercises have been approached in an inter-sensor comparison (Alraddawi et al., 2018; Fionda et al., 2019; Pałm et al., 2010; Thomas et al., 2011; Vaquero-Martínez et al., 2018). This work is the first, to our knowledge, to address level 2 products (not time-averaged) of a wide variety of satellite sensors in the Arctic.

Particularly, this work focuses on the station at Ny-Alesund, which is located in Svalbard, one of the nothernmost archipielagos in the Arctic. Figure 1 shows a topographic map of Svalbard. It is an area of interest in the European Arctic, since its specific synoptic regime brings more moisture from the lower latitudes compared to the rest of the region (Mewes & Jacobi, 2019), and belongs to the region with the highest increasing temperature trend in the Arctic (Susskind et al., 2019). Ny-Alesund is located in the west coast of Svalbard, and the village

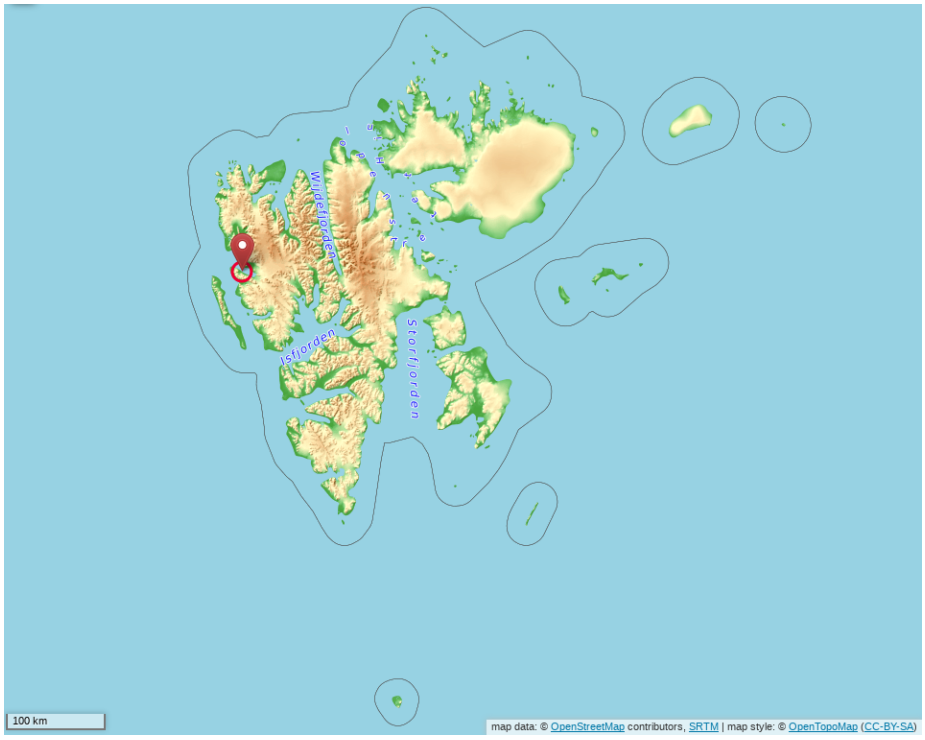

Figure 1: Topographic map of Svalbard, with Ny-Alesund location marked. Source: OpenTopoMaps

is an international center for various research activities. Its high latitudinal position implies polar night condition between 24 October and 18 February, and polar day between 18 April and 24 August, respectively. It is set in a fjord surrounded by glaciers, moraines, rivers, mountains and a typical tundra system. Its special orography makes Ny-Alesund not be a very representative location for the general Arctic (Maturilli et al., 2013), but also for this reason it stands as an excellent test bench for satellite validation.

This paper aims to compare satellite water vapor products with GNSS measurements at Ny-Alesund station, which is the only European Arctic station that counts with a long GNSS water vapor data series, starting in 2010. Although there are other satellite validation exercises at high latitude regions in literature (i.e., Alraddawi et al., 2018; Bedka et al., 2010; Pałm et al., 2010), this is the only one, to our knowledge, that validates instantaneous measurements and compares several IWV products against the same ground-based reference. The paper is structured as follows: after this Introduction, Section 2 describes the data used and explains the methodology applied: how the matching between data values was made, as well as the statistical approach followed for the analysis of satellite product performance. Results are drawn in Section 3 and, Section4 discuss the results. Finally, conclusions can be found in Section 5.

2. Materials and Methods

2.1. Data

2.1.1. Ground-based data

GNSS ground-based stations allow the retrieval of the IWV, through the method proposed by Bevis et al. (1992). The GNSS station receives information from the constellation of GNSS satellites through microwaves that travel with the velocity of light. Therefore, the time needed to reach the receiver can be used to estimate the distance to the satellites, and by triangulation, the position of the receiver. However, there are a number of corrections to be done, due to clock errors, relativistic effects, and the presence of atmosphere.

Particularly, the troposphere causes a delay in the signal, which is known as Slant Tropospheric Delay (STD). This quantity is converted to the Zenith Tropospheric Delay (ZTD) through the mapping functions (Boehm et al., 2006a,b; Niell, 2000). ZTD can be divided in two components, a Zenith Hydrostatic Delay (ZHD) and a Zenith Wet Delay (ZWD). ZHD can be obtained by a simple model (Saastamoinen, 1972) if atmospheric pressure at the station level is known. Thus, ZWD can be obtained from subtracting ZHD from ZTD. IWV can be retrieved from ZWD values if the water vapor weighted mean temperature (T_m, see Davis et al. (1985)) is known. The value of T_m allows to calculate the parameter κ for the conversion between ZWD and IWV, as Eq. (1) shows.

$$\mathrm{IWV} = \kappa \mathrm{ZWD} \tag{1}$$

The ground-based station used in this work is named "NYA2". Data for this station are available from 2010 to the present. GNSS data are processed by Potsdam GFZ, using the GFZ EPOS8 software. This product provides both the ZTD and the IWV, both in real time and post-processing. In this work, post-processing IWV products are used. GFZ EPOS.P8 software processing is based on least squares adjustment using a sliding window approach and makes use of the International Earth Rotation and Reference System Service (IERS) standards. The Precise Point Positioning (PPP) strategy is used, in which each station is processed separately, fixing the high quality GNSS orbits and clocks. The processing is done with a 24 hour data window, elevation cut off angle of 7°, sampling rate of GNSS data of 2.5 minutes. The ZTD solutions are retrieved with a resolution of 15 minutes, tropospheric east and north gradients with hourly resolution, and GMF mapping functions. This data-set have been validated against radiosondes (Vaquero-Martínez et al., 2019), with a strong agreement (mean differences of -0.4 ± 0.7 mm and $R^2 = 0.983$) and small dependencies between both instruments.

Moreover, human observations of cloud cover (CC) at the study station were recorded by Norwegian Meteorological Institute. These CC observations were

Table 1: Summary of characteristics of the different products in this study. It includes instrument, wavelength range (Wl. range), pixel size (km), passing frequency (Pasing freq.) and Period.

Instrument	wl. range	Pixel size (km)	Passing freq.	Period
AIRS	IR (hyper-spectral)	13.5	~100 min	2010-2017
GOME-2	614.0-683.2 nm	80 x 40	~100 min	2010-2017
MODIS IR	6535-8700 nm	5 x 5	~100 min	2010-2017
MODIS NIR	869-940 nm	5 x 5	~100 min	2010-2017
OMI	426.0-468.5 nm	13 x 24	~100 min	2010-2017
POLDER	910 and 865 nm	50 x 50	~60 min	2010-2013
SCIAMACHY	around 700 nm	32 x 215	~100 min	2010-2012

taken every day at 00, 06, 12, 18 h UTC and they were expressed in oktas of sky covered by clouds, which consist of whole numbers from 0 (cloud-free sky) to 8 (overcast sky). These measurements are interpolated linearly to the time of measurement of the satellite in this work. The station used in this work is NY-ÅLESUND (station number 99910), in the period 2010-2017.

2.1.2. Satellite data

Satellite data is retrieved from the following sensors: AIRS, GOME-2, MODIS, OMI, POLDER and SCIAMACHY. Table 1 summarizes the main characteristics of these sensors and the IWV products retrieved from them.

AIRS. AIRS (Aumann et al., 2003) is an instrument on-board NASA's Aqua satellite platform. It is a hyper-spectral infrared sounder. It overpasses the Arctic zone every 1 hour and 40 minutes approximately.

The data used in this work is the L2 standard atmospheric and surface product version 6 (AIRS2RET.006). The retrieval (Susskind et al., 2003, 2006) method involves microwave radiometry and infrared measurements of brightness from Earth's surface and atmosphere. Several parameters are estimated at the same time and then improved in an iterative process. For the first guess, the

microwave product is used. The hyper-spectral features of AIRS allows the use of a large number of spectral channels (66 channels) that include several ones belonging to high water vapor absorption bands and others outside these bands. AIRS data used in this work are from the period 2010-2017.

GOME-2. GOME-2 (Callies et al., 2000) is a medium-resolution UV-VIS-NIR spectrometer. It is specifically designed with the objective of measuring ozone, both vertical profile and total atmospheric content, but other atmospheric compounds are also retrieved, like sulfur dioxide, nitrogen dioxide, bromine oxide, and water vapor. GOME-2 is mounted on two satellite platforms, MetOp-A and MetOp-B, with default scan widths of 960 km and 1920 km, with a ground pixel of 40 km $\times$ 40 km.

The IWV product of GOME-2 used in this work is derived from the GOME Data Processor (GDP, version 4.8) which is retrieved by the German Aerospace Center, Remote Sensing Institute (DLR-IMF) in the framework of the EUMETSAT satellite Application Facility on Atmospheric Chemistry Monitoring (O3 M SAF) (Grossi et al., 2015).

The retrieval method is based on Differential Optical Absorption Spectrography (DOAS), which is thoroughly explained in Wagner et al. (2006) and Wagner et al. (2003). The algorithm consists of three basic steps. First, DOAS fitting of water vapor, O_2 and O_4. In this step, the use of three types of vegetation spectra improves the broadband filtering, and a correction for the ring effect is applied. Then, the non-linear absorption correction is applied, since the water vapor absorption is not linear with IWV. For this second step a mathematical convolution of the water vapor absorption spectrum with the instrument slit function is used to obtain the correction factors. The effect is more pronounced for large values of slant column density (SCD) of water vapor. In the third and last step, the vertical column density (VCD) is calculated. The SCD must be converted to VCD to make the quantity geometry-independent, allowing the conversion to IWV. In order to perform this correction, the SCD is divided by a convenient Air Mass Factor (AMF), which is derived from oxygen absorption,

so the actual AMF used is for oxygen. Although oxygen's and water vapor's AMFs are expected to be similar, oxygen's is expected to be larger than water vapor's. Thus, to correction factors from a look-up table are applied. The look-up table considers several variables, like Solar Zenith Angle (SZA), line of sight angle, surface albedo and relative azimuth. The look-up table is calculated using radiative transfer models.

The wavelengths involved are between 614 and 683 nm. The spectral resolution is about 0.54 nm. One of the main advantages of this approach is that it does not requires a priori information or external calibration sources.

MODIS. MODIS (King et al., 1992) is a spectroradiometer on-board two satellites, namely Aqua and Terra. Aqua passes through the equator from south to north in the afternoon, while Terra does it from north to south in the morning. MODIS measures in 36 spectral bands, which go from 0.4 to 15 μm. Level 2 moisture profiles and IWV are derived with a spatial resolution of 1 km $\times$ 1 km, but averaged to a 5 km $\times$ 5 km resolution.

The IWV measurements of MODIS are included in the MOD05_L2 and MYD05_L2 collection 6. However, IWV is obtained in the MOD07 and MYD07 products and added to MOD05 and MYD05 for convenience.

In this work, two products are analyzed, NIR and IR products. These products are 05_L2 version 6.1. The NIR product (Gao & Kaufman, 1992) uses 2-channel and 3-channel rationing techniques and look-up tables with values of these ratios and total water vapor matching these values. The look-up tables are generated using radiative transfer models. Once this is obtained, the value is converted to IWV taking into account the geometry of the viewing and solar angles. This product has an "Unobstructed Field of View (FOV) Quality Flag", and the data with this flag equal to 0 (meaning "Confident Cloudy") are discarded for this work.

The IR product (Seemann et al., 2003, 2006) deals with a synthetic regression and a subsequent non-linear physical retrieval used to improve the fit in an iterative way.

OMI. OMI (Ahmad et al., 2003; Levelt et al., 2006) is a UV/VIS imaging spectrometer on-board Aura satellite. Aura has a sun-synchronous polar orbit, sampling the whole planet daily at 13:30 local time (LT). The nominal pixel size of OMI is 13 km $\times$ 24 km at nadir. OMI channel covers $\simeq 350 - 500$ nm, at a spectral resolution of $\simeq 0.5$ nm.

The retrieval algorithm is the Smithsonian Astrophysical Observatory (SAO) Version 4, and it has two steps. First, the SCD is retrieved using a spectral fitting algorithm. Then, the conversion to vertical column density (VCD) is calculated as the ratio between VCD and AMF. The conversion to IWV is then straightforward. For the spectral fitting, some parameters are used along with water vapor, such as wavelength shift, under-sampling, closure polynomials, interfering molecules and Raman scattering. The starting and ending wavelengths used were optimized by reconciling several factors (Wang et al., 2019). Data with OMI's cloud fraction over 0.25 are filtered out.

POLDER. POLDER (Deschamps et al., 1994) is an optical imaging radiometer, first launched in 1996 on-board ADEOS I satellite. However, new generations were launched on-board ADEOS II and finally on-board PARASOL micro-satellite, which was shutdown in 2013. In this work POLDER / PARASOL data were used.

The POLDER product used in this work is version 1.01 RB2. Water vapor measurements are based on the ratio of reflected radiances at 910 and 865 nm. This partially reduces the variation of surface reflectance with wavelength (although it can be affect by aerosol load), providing water vapor through an empirical equation (Vesperini et al., 1999). The spatial resolution is 50 km $\times$ 50 km.

SCIAMACHY. SCIAMACHY (Bovensmann et al., 1999) is a spectrometer on-board the Environmental Satellite (Envisat). The operational period of the instrument went from March 2002 to April 2012, and the pixel size is 60 km $\times$ 30 km.

The product used in this work is offline level 1B version 1.1 (SCI_NL___1PWDPA). The algorithm used to retrieve IWV from SCIAMACHY is based on the method

known as Air Mass Corrected Differential Absorption Spectroscopy (AMC-DOAS, Noël et al. (2004)). The algorithm uses bands around 700 nm. Although it is based on the DOAS approach, it is modified to account for saturation effects and uses O_2 and H_2O absorption features to derive a correction for the AMF (Air Mass Correction, AMC). AMC represents how similar are atmospheric conditions to the conditions considered in the model calculations. The closer the AMC to the unit, the better the match. Therefore, AMC also contains information about the IWV measurement quality. In order to asses the quality of the product, SCIAMACHY data are filtered following these criteria: SZA below 88°, AMC greater than 0.8. Clouds are not specifically filtered, but the AMC constraint eliminates most of the cloudy scenes.

2.2. Methodology

2.2.1. Matching criteria

In order to perform the inter-comparison, it is necessary to match the GNSS ground-based data with the satellite data, pairwise. There are two main criteria, one spatial and another temporal:

- Spatially, the closest pixel to the station is taken, provided that the distance between the pixel center and the station is less than 100 km. Figure 2 shows that there are no significant differences with this distance across the different products performance. In the case of OMI, the pixels in a $0.25° \times 0.25°$ area around the site are selected and averaged (weighted average is used so that the higher quality data get more weight)

- Temporally, the GNSS measurement chosen to match each satellite closest pixel is the one whose difference in time with the satellite passing is the smallest, and this differences must be less than 30 minutes.

For quality selection, only information from the satellite products have been used (see section 2.1 for details).

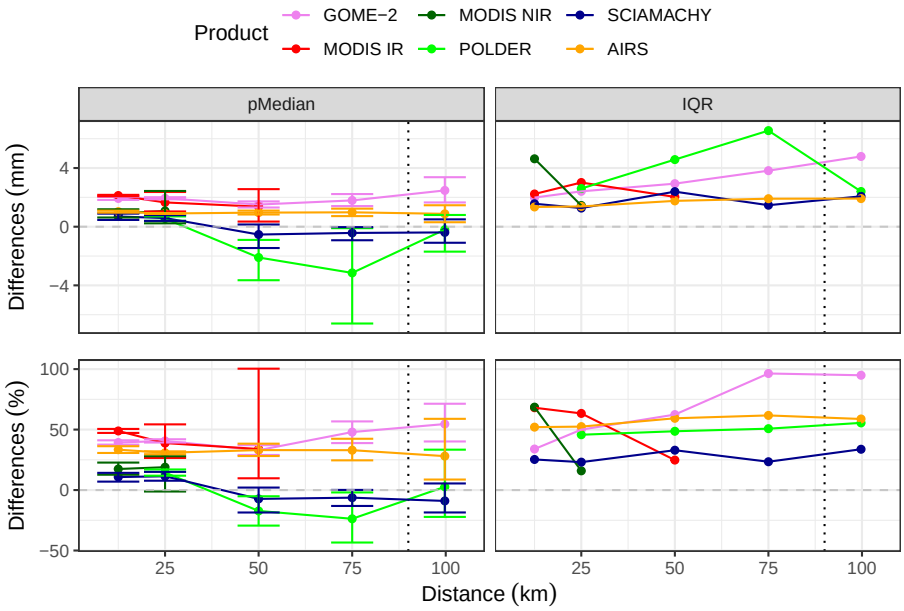

Figure 2: Dependence of differences on distance. Pseudomedian (pMedian) is shown with error bars representing the 95 % confidence interval in the Wilcoxon test (note that in some cases the interval is very small, so the error bars are not seen)

2.2.2. Statistical analysis

The analysis performed to analyze the measurements has two parts. First part deals with obtaining general statistics to characterize the relationship between satellite and GNSS measurements. Apart from regression analysis, the distribution of the differences is studied. The differences are studied as absolute differences and relative differences. Eq. (2) and Eq. (3) show the definition of absolute and relative differences, respectively.

$$\delta_i = w_i^{\mathrm{sat}} - w_i^{\mathrm{GNSS}} \tag{2}$$

$$\delta_i = 100 \cdot \frac{w_i^{\mathrm{sat}} - w_i^{\mathrm{GNSS}}}{w_i^{\mathrm{GNSS}}} \tag{3}$$

The index i represents a point in time for which a GNSS measurement and satellite measurement are matched, while w is the IWV measured by either GNSS or satellite (sat). The reason for studying both distributions is that on the one hand, relative differences are useful to measure quality of a measurement. On the other hand, small IWV values can produce a small absolute differences but also a large relative difference. In these cases (which are usual in Ny-Alesund), it would be more appropriate the study of the absolute difference. Additionally, the absolute differences allow to analyse the shape of the distribution (which follow typically a normal distribution). Thus, studying both distributions give a more complete picture of the features of the comparison. For instance, in Berezin et al. (2016), a comparison between radiosonde and microwave radiometer reported relative differences reaching 50%, which was below 1 mm.

Some indices are used to characterize the distributions of differences and relative differences, like the (Relative) Mean Bias Error, (R)MBE, which is defined as the mean of (relative) differences; (Relative) Mean Absolute Bias Error, (R)MABE, which is the mean of the absolute value of (relative) differences; and Inter-Quartile Range (IQR), which is the range in which the central half of the (relative) differences is found. (R)MBE indicates the accuracy of the measure-

ments, that is to say, if satellite data generally overestimate (positive values) or underestimate (negative values) the GNSS data. (R)MABE and IQR show the precision, or the width of the range in which the satellite-GNSS differences can be found.

The second part of the analysis is focused on studying the dependence of differences on IWV, SZA, seasonality and cloud cover (CC). Data is classified into bins of data of similar value of the study variable (i.e. IWV and SZA). Particularly, 8 bins are used, and the intervals used are those that make an equal (or similar) number of cases for each bin. Then, the precision and accuracy of the satellite products in each bin are calculated through two indices. These indices typically are Mean Bias Error (MBE, mean of the differences) and Standard Deviation (SD) of differences. However, as Section 3 exhibits, the differences do not follow the normal distribution, so instead another two indices will be used. For accuracy (overestimation/underestimation) , pseudo-median (Wilcoxon, 1945, median of midpoint of every pair of elements of the data) and for precision, IQR. These are more suitable for non-normal and non-symmetric distributions. Thus, the evolution of accuracy and precision along the study variable can be examined in depth, and conclusions can be drawn on the dependence of performance of the study variable.

3. Results

3.1. Site climatology

Some aspects of the climatology of the Ny-Ålesund station are studied, as this will help understand the reasons why the satellite products may fail at high latitudes. Figure 3 shows that IWV is the largest in the months of July, August and September, while the smallest values of monthly IWV are registered at January, February, and March. Regarding sky conditions, it is observed that the months with low CC are scarce, and accumulated in the season between December and April. This implies that, generally, the months with more IWV

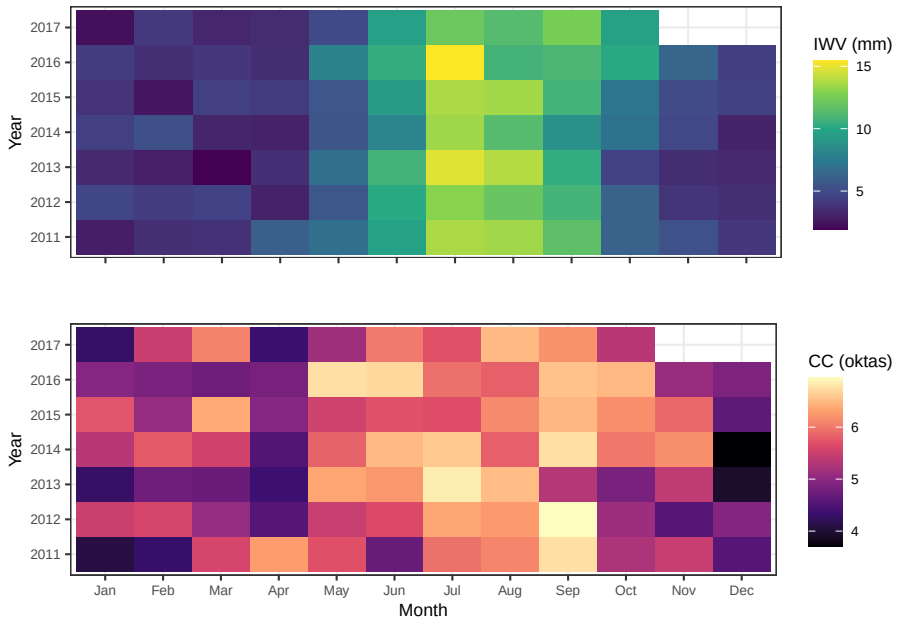

Figure 3: Daily means of IWV and CC in Ny-Ålesund Station

are also the ones with larger CC, gathered around the warm season (June-September), which is coincident with more hours of sunshine and smaller SZA.

3.2. General statistics

Some general statistics are given in Table 2. It must be noted that the different products, due to their different nature, and because they are from different missions, have different time-spans and different number of data. For instance, NIR and VIS products are not available in winter because of the lack of sun-light. Also, POLDER and SCIAMACHY have a short time-span, since POLDER/PARASOL was shutdown in 2013 and SCIAMACHY in 2012. Therefore, a small study, with two versions, have been conducted, using (a) only months in which all products have data avaiailable, and (b) only months in which all products except PARASOL and SCIAMACHY have data available. Then, similar tables as Table 2 have been produced (Tables S1 and S2 in Supplementary Material) with very similar results to those from the aforemen-

tioned Table. In the version (a), shown in Table S1, MODIS-NIR showed worse statistics (i.e., R^2=0.12), due to the small number of data-points ($N = 221$). Therefore, Table 2 is considered valid and representative for this study.

The statistics generally agree with those in Thomas et al. (2011), where comparisons between GNSS and satellite instruments (AIRS, MODIS NIR) were performed in Antarctic sites (for AIRS, R values between 0.70 and 0.95; while for MODIS NIR, R values are between 0.69 and 0.90). The central values of the differences, given by MBE, RMBE, Median and RMedian, show positive values except for SCIAMACHY (VIS) and POLDER (NIR). OMI IWV has the strongest overestimation ($\simeq$ 5 mm and 63 %) and presents higher variability (IQR) than most products. This product also has less data available than the rest (296 data-points). Regarding linearity, the linear correlation coefficients show an adequate agreement between satellite and GNSS IWV ($R^2 > 0.5$) for those product using VIS or IR spectral lines, while for NIR products the linear correlation is very poor. It has been observed (see section 3.3.3) that cloudy scenes cause a clear underestimation of IWV for NIR measurements. The number of data is also very different across the different products, due to two main reasons: (1) algorithms relying on solar radiation (VIS or NIR) cannot measure in winter or night (lack of sunshine), and (2) some satellites have not been operational for the whole 2010-2017 period (POLDER, SCIAMACHY). The regression lines corresponding to the linear fit between satellite and GNSS IWV have positive intercept for all the products in this work, being of the order of a few mm. The slopes are generally close to the unity ($0.7 - 1.0$), except for NIR products, which have a very low slope ($0.3 - 0.5$), and OMI, with a rather high slope (1.26). AIRS statistics reported in Bedka et al. (2010) for station in Barrow (71.323°N, 156.616°W) have better performance (slope = 0.91, intercept = 0.017, R^2 = 0.94) when compared against a ground-based microwave radiometer. This can be likely due to differences both in the instruments used as ground truth, and differences in methodology (i.e. a more relaxed quality control in the current work), as well as differences in the climatology of both arctic sites, Barrow and Ny-Alesund. Particularly,

Table 2: Summary statistics of the IWV satellite products with respect to IWV from ground-based GNSS receiver. The columns are: product; type of radiation, ToR; (relative) mean biased error, (R)MBE; (relative) median, (R)Median; relative inter-quartile range, (R)IQR; (relative) mean absolute biased error, (R)MABE; root mean square error (RMSE), intercept of the linear regression; slope of the linear regression; correlation coefficient, R^2; number of data, N. Relative quantities are in percent, physical quantities are in mm, intercept is in mm, slope and R^2 have no dimensions.

Product	ToR	RMBE	RMedian	RMABE	RIQR	MBE	Median	MABE	IQR	RMSE	intercept	slope	R^2	N
AIRS	IR	19.88	5.95	37.79	49.38	-0.07	0.32	1.71	2.45	2.36	1.06	0.74	0.73	17165
GOME-2	VIS	28.53	19.89	36.25	46.08	1.58	1.18	2.45	2.89	3.49	1.52	1.01	0.70	13395
MODIS IR	IR	56.10	32.08	63.65	63.46	1.67	1.85	2.32	2.58	2.93	2.60	0.86	0.67	6819
MODIS NIR	NIR	28.40	6.01	47.86	66.24	0.81	0.55	2.93	4.70	3.78	3.11	0.50	0.29	722
OMI	VIS	63.54	59.66	65.87	57.76	5.87	5.76	6.10	5.82	7.32	3.27	1.26	0.54	296
POLDER	NIR	-14.17	-16.84	35.40	56.75	-2.67	-1.40	3.66	6.00	5.18	2.56	0.31	0.22	2372
SCIAMACHY	VIS	-10.36	-8.41	23.45	39.24	-1.23	-0.59	2.04	2.87	2.92	1.11	0.73	0.68	885

the presence of clouds causes the well-known shielding effect (Kokhanovsky & Rozanov, 2008). This effect causes the satellite instrument to be sensitive only to the water vapor above the clouds, which results in IWV underestimation.

Moreover, Alraddawi et al. (2018) reported a comparison between satellite monthly IWV products (AIRS, SCIAMACHY, MODIS) against GNSS in the Arctic, with better results in Ny-Alesund (i.e. better correlation coefficients, lower bias). For instance, in that work, AIRS monthly IWV is reported to have R^2=0.98, and -0.1 mm, while SCIAMACHY shows $R^2 = 0.94$ and bias 1.5 mm, and MODIS (Terra only, and only bright land and ocean sun-glint) NIR exhibited $R^2 = 0.92$ and bias 0.4 mm. The use of monthly means could result in a reduction of noise in the comparison. However, Pałm et al. (2010) studied the performance of several instruments, GOME-2 and SCIAMACHY among them, against GNSS IWV, with similar results to those reported in Table 2. For instance, for GOME-2, R^2 was 0.77 and slope 0.72, while for SCIAMACHY R^2 was found to be 0.72 and slope 0.73.

In order to check if these statistics could be different for day or night measurements, we show in Table 3 IR products (the only ones with day and night data available) statistics for day and night separately. Day and night measure-

Table 3: Summary statistics of the IWV satellite products with respect to IWV from ground-based GNSS receiver. The columns are: product; moment (day or night); (relative) mean biased error, (R)MBE; (relative) median, (R)Median; relative inter-quartile range, (R)IQR; (relative) mean absolute biased error, (R)MABE; root mean square error (RMSE), intercept of the linear regression; slope of the linear regression; correlation coefficient, R^2; number of data, N. Relative quantities are in percent, physical quantities are in mm, intercept is in mm, slope and R^2 have no dimensions.

Product	Moment	RMBE	RMedian	RMABE	RIQR	MBE	Median	MABE	IQR	RMSE	intercept	slope	R^2	N
AIRS	day	7.42	0.04	28.27	40.61	-0.43	0.00	1.95	2.96	2.69	1.17	0.74	0.76	9818
AIRS	night	36.54	17.07	50.51	61.94	0.40	0.60	1.39	1.87	1.83	1.09	0.70	0.66	7347
MODIS IR	day	44.56	31.17	48.97	52.07	1.93	1.98	2.40	2.28	3.05	2.78	0.88	0.69	4728
MODIS IR	night	82.21	37.80	96.83	112.09	1.10	1.34	2.15	3.19	2.65	3.57	0.43	0.31	2091

ments are very similar, except for MDOIS IR. The strong differences observed in MODIS IR between day and night data could be related to the low temperatures that would make IR radiation smaller and therefore more difficult for the satellite to retrieve the IWV.

In order to visualize the distributions of differences and relative differences, Figure 4 shows that the peaks of the distributions are close to the zero line, except for OMI. While NIR products show some tendency to underestimation (negative differences, probably due to the presence of clouds), OMI IWV tends to be greater than GNSS IWV. OMI's overestimation can be due to the fact that the process of optimization for SCD estimation is done for overall. Therefore more region-specific optimization could help in adjusting this wet bias.

Relative differences can reach very large values, 4270 % (MODIS IR), while other products have their maximum under one hundred ten percent (POLDER). Minima (underestimations) are more modest, all of them being closer to zero than -100%, which is probably due to the fact that IWV cannot have negative values.

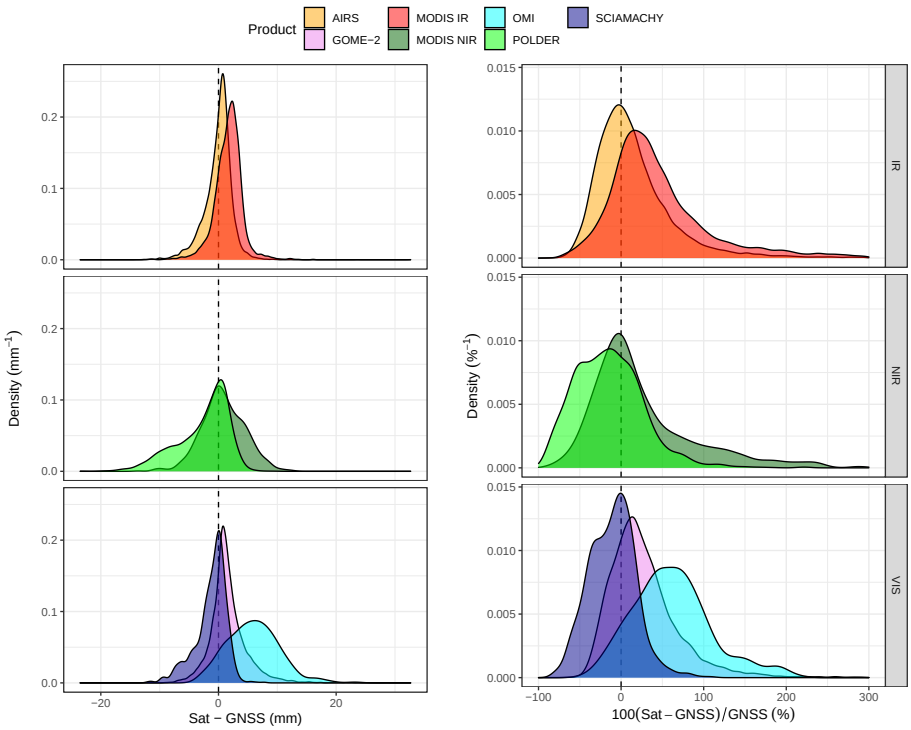

Figure 4: Density plots of regular and relative differences. Note that relative differences have been trimmed at 300 to make the plot more visual

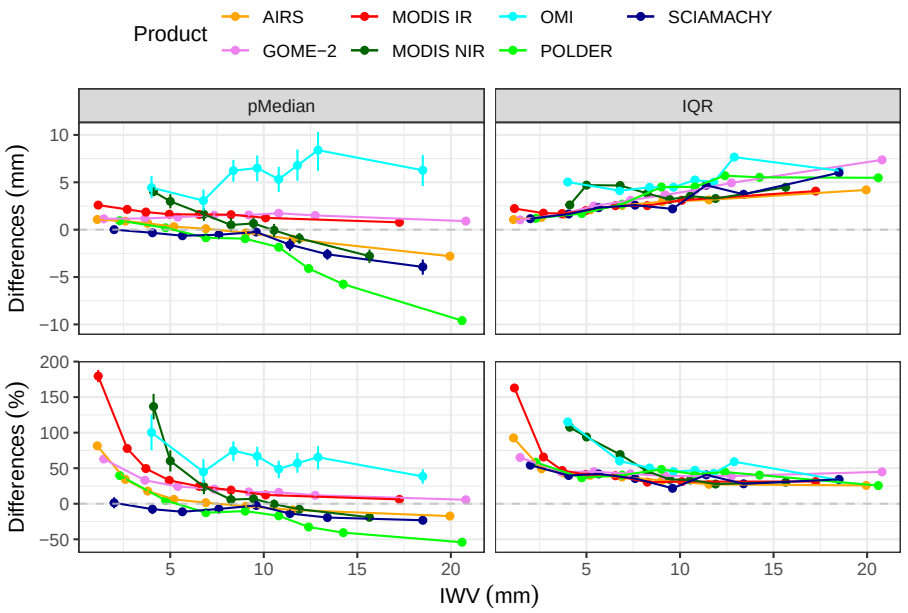

Figure 5: Dependence of differences on IWV values. Pseudomedian (pMedian) is shown with error bars representing the 95 % confidence interval in the Wilcoxon test (note that in some cases the interval is very small, so the error bars are not seen).

3.3. Dependencies with other variables

3.3.1. Dependence with IWV

IWV values can influence the performance of the measurements due to, for instance, saturation effects. Figure 5 shows the pseudomedian (left) and IQR (right) for both the differences (up) and relative differences (bottom). The relation between IWV and pseudomedian of differences is approximately a straight line. NIR products underestimate more as IWV increases, and the same happens (more weakly) to SCIAMACHY and AIRS, while GOME-2 and MODIS IR have a rather constant overestimation ($\sim$ 2 mm). OMI shows a high bias around 6 mm with some fluctuations. IQR of differences also shows a linear relation with IWV, increasing with this variable. The greater values of IQR are from MODIS NIR and OMI ($\simeq$ 7 mm).

The evolution of relative differences pseudomedian with IWV is similar for

all satellite products: small IWV values have the highest relative differences, then decreasing and stabilizing at low relative differences (sometimes falling below zero). This has been observed in previous works in other regions (Román et al., 2015; Vaquero-Martínez et al., 2018). SCIAMACHY shows a rather constant value between 0 % and 25 %, while OMI also shows value around 75 %. POLDER reach −50 % at large IWV. Relative IQR also decreases with IWV, but stabilizing at values between 25 and 50 %. Again, MODIS NIR and OMI have larger relative IQR than the rest. AIRS shows a rather constant value between 25 and 50 %.

3.3.2. Dependence with SZA

Dealing with large SZA (i.e. the sun is very low, close to the station horizon) situations is difficult for those retrieval algorithms that rely on sun's radiation reflected by the Earth and atmospheric components, because of the decrease in the amount of radiation per unit of area. Therefore, Figure 6 shows the pseudomedian and IQR of both differences and relative differences in bins of SZA for each product.

The overestimation of $\sim$ 6.2mm (62%) is evident for OMI, with a slight decrease for SZA $> 68°$. POLDER shows a rather constant pseudomedian of −2.3 mm ($\sim$ −17%), while MODIS NIR exhibits some overestimation for low SZA, and underestimation for SZA $> 70°$. AIRS pseudomedian goes from negative values (−0.6 mm or −2.8 %) at low SZA to positive values (0.6 mm or 26%) at high SZA. GOME-2 exhibits increasing values of pseudomedian with SZA, from 0.9 mm (10%) to 2.1 mm (60%). MODIS IR shows decreasing values of pseudomedian between 2.1 mm (28 %) and 0.9 mm (23 %) up to 90°. It must be noticed that in this case the relative pseudomedian values are fluctuating and do not show this decreasing tendency (with values between 23 % and 43%). This difference between relative differences and differences in millimeters could be caused by the IWV daily and annual cycle For higher SZA (night), MODIS IR increases to 1.5 mm (70 %).

Moreover, IQR values are large for NIR and OMI products ($>$ 5 mm,

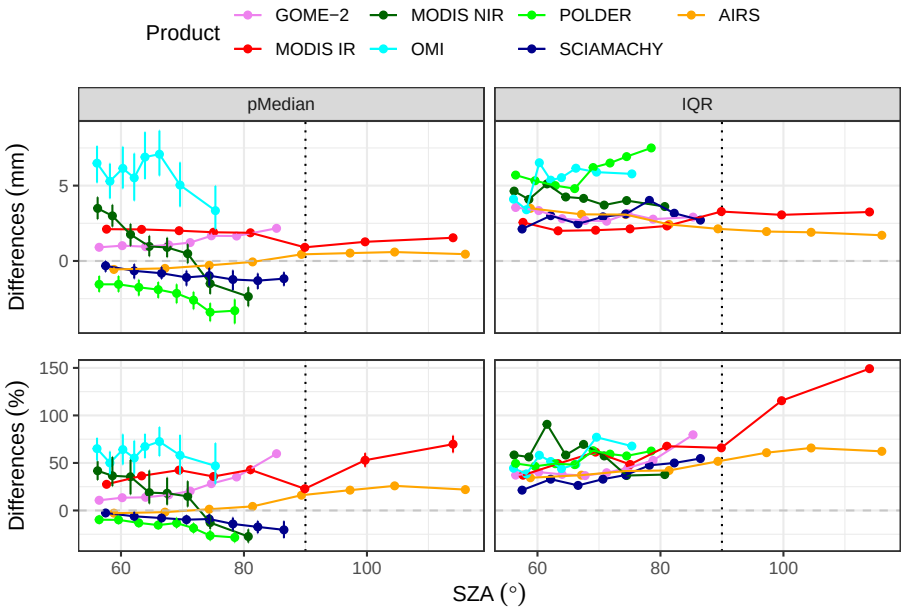

Figure 6: Dependence of differences on SZA values. Pseudomedian (pMedian) is shown with error bars representing the 95 % confidence interval in the Wilcoxon test (note that in some cases the interval is very small, so the error bars are not seen)

> 50 %), while generally lower for the rest, being over 50 % only for high values of SZA. MODIS IR relative IQR increases strongly with SZA at night measurements, reaching 150 %. However, MODIS NIR relative IQR exhibits a peak at 70°. This may be related to the diurnal and annual cycle of water vapor and therefore different values of IWV at different SZA values. Specifically, low values of IWV are associated with large SZA (winter) and vice versa. AIRS exhibits a decreasing tendency with SZA for IQR, ranging from 1.7 mm to 3.5 mm, while its relative IQR increases with SZA from 34 % to 66 %.

3.3.3. Cloud Dependence

Clouds can affect strongly the performance of satellite measurement. As mentioned above, they are responsible for the shielding effect (Kokhanovsky & Rozanov, 2008). In Figure 7, pseudomedian (left) and IQR (right) of the differences (top) and relative differences (bottom) are plotted against the CC in bins. The increase of CC results, specially over 6 oktas, in an increase of the underestimation of IWV (sometimes this is compensed by an overestimation on low CC situations). POLDER reaches underestimations of −5 mm, which indicates that this product needs some quality control to avoid retrievals under very cloudy conditions. NIR and VIS products show higher IQR of differences ($\sim$ 5 mm) than the rest of products (which yield similar results, around 2.5 mm). It must be noticed that some of the algorithms filter out cloudy scenes with different thresholds.

Pseudomedian of relative differences shows a rather decreasing linear trend. POLDER reaches −40 %. GOME-2 and MODIS IR exhibit similar results, while OMI shows its typical overestimation in this station. Overestimation under low CC scenes can be explained because in such scenes IWV is typically small, resulting in overestimation due to small IWV (see section 3.3.1). SCIAMACHY shows the lowest values of relative IQR (below 32 %), and MODIS products show higher relative IQR at low CC. Results reported by Román et al. (2015) show similar behavior with CC for GOME-2 in the Iberian Peninsula.

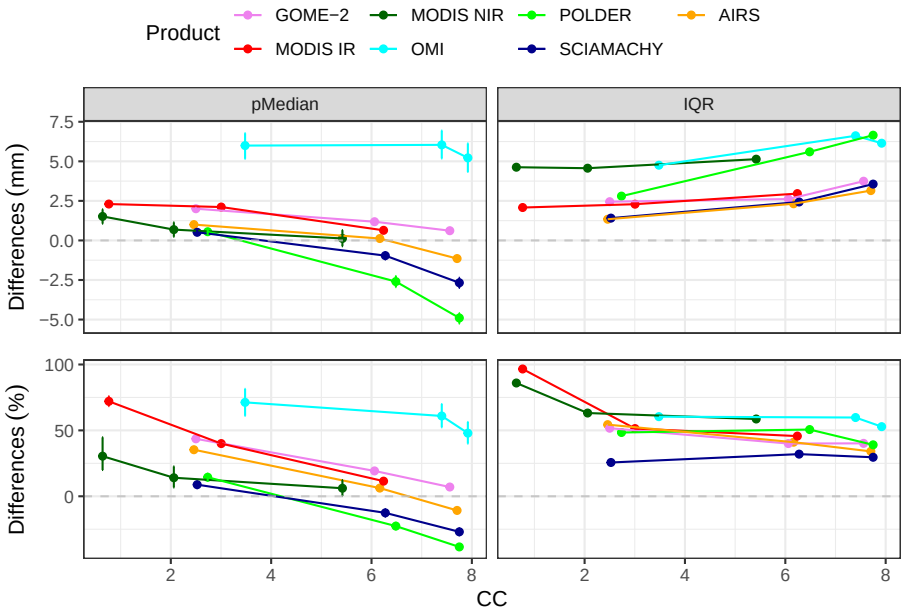

Figure 7: Dependence of differences on Cloud Cover values. Pseudomedian (pMedian) is shown with error bars representing the 95 % confidence interval in the Wilcoxon test (note that in some cases the interval is very small, so the error bars are not seen)

3.3.4. Seasonal dependence

The meteorological conditions change on season, which can cause a seasonal cycle in the performance of satellite measurements. Figure 8 shows the evolution of pseudomedian and IQR along the months of the year. NIR and VIS algorithms do not cover winter months (November, December, January) because there is not any sunlight at the station during these months. AIRS shows the best values of pseudomedian (closer to zero), with some underestimation in summer and overestimation in winter. Regarding the rest of products, pseudomedian is generally worst (positive for VIS and IR products, and negative for NIR and SCIAMACHY) in summer months, while relative pseudomedian is generally better (closer to zero). Something similar happens with IQR and relative IQR: the first increases in summer and decreases in winter, while the latter increases in winter and is smaller in summer. This is probably due to the presence of larger IWV during summer than during winter months. POLDER and OMI generally have the larger IQR and relative IQR values of all satellites. However, MODIS IR shows a very large relative IQR in December and January, reaching 175 %, although in these months IQR is about 3 mm. It must be noted that values of IWV in these months are typically below 5 mm. These results are similar to those reported by Alraddawi et al. (2018).

4. Discussion

This work aims to shed light on the performance of water vapor satellite products the station Ny-Alesund, which belongs to the European Arctic. This Norwegian station has been selected for having a long-term, high quality GNSS IWV time-series available to be used as reference values.

This study includes several IWV satellite products, which use different regions of the electromagnetic spectrum. GOME-2, SCIAMACHY and OMI use VIS bands; MODIS IR and AIRS, IR bands, and MODIS NIR and POLDER, NIR bands.

Correlation coefficient R^2 was found to be below 0.3 for the NIR products

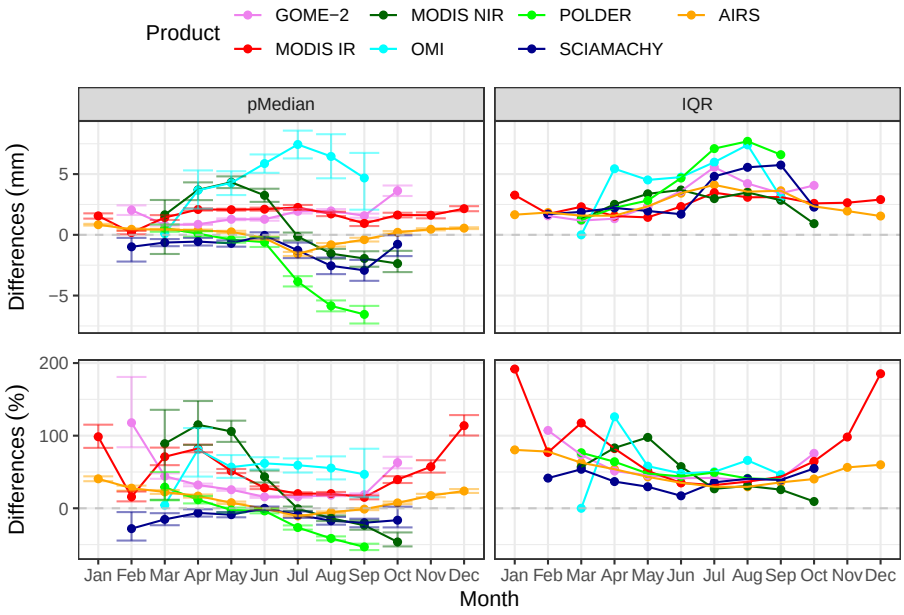

Figure 8: Seasonal dependence of differences. Pseudomedian (pMedian) is shown with error bars representing the 95 % confidence interval in the Wilcoxon test (note that in some cases the interval is very small, so the error bars are not seen)

in this work, which indicates that these two products are not fitted to provide good quality IWV data, at least for L2 measurements (measurements not averaged in time or space). Particularly, POLDER showed notable underestimation under cloudy-sky conditions. For MODIS NIR, many situations of heavy cloud coverage are filtered out thanks to its Unobstructed Field of View Quality Flag, which allows to certain improvement in the correlation. On the other hand, POLDER product bias (pseudomedian) does not have a negligible dependence on SZA, while MODIS NIR tends to underestimation for large SZA. Seasonally, these products are more dry-biased during the warm season, and there is some wet-bias during the cold season. This is expected because of the presence of clouds and large IWV. Therefore, we do not recommend these products for climate studies or weather forecasts, unless the data are treated or filtered to avoid the worse scenarios.

VIS products exhibit a notable correlation with GNSS in Ny-Alesund ($R^2 \sim 0.7$), except for OMI ($R^2 = 0.54$). OMI showed a substantial bias (MBE) of almost 6 mm (probably due to the use of global optimization of the slant column retrieval) and similar variability (IQR), while for GOME-2 and SCIAMACHY these were about $2-3$ mm. The pseudomedian shows a weak dependence with IWV (some increase could be observed in OMI), and relative pseudomedian had the typical behavior observed in other satellite IWV validations. IQR increased with IWV for the three satellite products. Large SZA generally worsens VIS products, increasing underestimation for SCIAMACHY and overestimation for GOME-2, while OMI's typical overestimation is reduced. The seasonal performance analysis showed that during summer overestimation/underestimation is increased, but decreased in relative terms. IQR also increases in summer, but in relative terms decreases in summer. The presence of clouds increases overestimation or decreases underestimation, and increases IQR. From the VIS products studied in this work, we consider that OMI does not meet the proper quality for climate studies or weather forecasts, although the algorithm could be tuned to solve the problems OMI IWV product exhibits. The rest of VIS satellite products can be used in climate studies and weather forecasts.

IR products have shown similar correlation to VIS products ($R^2 \sim 0.7$). AIRS bias is negligible, while MODIS IR is around 1.6 mm. MODIS IR pseudomedian changes weakly with IWV (always overestimating around 1.6 mm), while AIRS pseudomedian slightly decreases, going from overestimation at small IWV, to underestimation at large IWV. IQR increases in both cases linearly with IWV, up to 5 mm. Regarding SZA, both IR product increase their tendency to overestimation in relative terms as SZA increases, while in absolute terms, MODIS IR has a rather constant value below 2.5 mm and AIRS slightly increases with SZA, always close to the zero. IQR is rather constant for both products ($\sim$ 2.5 mm), but relative IQR increases with SZA. The seasonal behavior is similar to the other products, with MODIS IR overestimation and AIRS underestimation in summer. AIRS shows generally better statistical values than MODIS IR. It must also be noted that, being the only products that can make measurements at night, these two products can measure IWV in November, December and January. The behavior with different CC conditions is similar to the rest of products, although they are less affected by very cloudy skies. Hence, we consider that IR products can be used to study the climate of the European Arctic or in weather forecasts regarding this region.

5. Conclusions

The European Arctic region, where Ny-Alesund station is set, is a challenge for remote sensing satellite observations. Nevertheless, IR and VIS satellite retrievals can provide good quality data for this fundamental region in the climate system. OMI (VIS), however, presented a high wet bias (5-6 mm) which should be address in next versions of the algorithm. VIS products were specially affected by large SZA values. IR products exhibit some dependences with IWV and SZA, although it seems to be less affected by very cloudy scenes.

This study suggests that the quality of the NIR products used (MODIS NIR and POLDER) needs to be improved to study this region, although more research is needed to investigate whether this is a local effect at Ny-Alesund or

it is general in the region. Particularly, quality control for cloudy-scenes and large SZA values could be useful.

In any case, it is still necessary to improve the algorithms of all satellite products to solve the dependencies that this work has found.

6. Acknowldegments

This work was partially supported by Junta de Extremadura and FEDER funds IB18092. JVM thanks Junta de Extremadura and European Social Funds for the predoctoral grant PD18029. VEC and RR are grateful to the Spanish Ministry of Science, Innovation and Universities for the support through the ePOLAAR project (RTI2018-097864-B-I00).

Authors thank the Norwegian Meteorological institute (MET Norway) for providing cloud cover information; GFZ Helmholtz Center Postdam for the GNSS IWV product, and the science and support teams of GOME-2, MODIS, POLDER, AIRS, OMI and SCIAMACHY for retriving and making available their respective products. Authors also aknowldge the R Core Team (2016) packages tidyverse 1.2.1 (Wickham, 2017), lubridate 1.7.1 (Grolemund & Wickham, 2011), ggthemes 3.4.2 (Arnold, 2018), ggpubr 0.2.4 (Kassambara, 2019), RColorBrewer 1.1-2 (Neuwirth, 2014), kableExtra 1.1.0 (Zhu, 2019). Finally, authors would like to thank the comments of the anonymous reviewers, which helped improve the quality of this work.

References

Ahmad, S. P., Levelt, P. F., Bhartia, P. K., Hilsenrath, E., Leppelmeier, G. W., & Johnson, J. E. (2003). Atmospheric products from the ozone monitoring instrument (OMI). In W. L. Barnes (Ed.), *Optical Science and Technology, SPIE's 48th Annual Meeting* (p. 619). San Diego, California, USA. doi:10.1117/12.506042.

Alraddawi, D., Sarkissian, A., Keckhut, P., Bock, O., Noël, S., Bekki, S., Irbah, A., Meftah, M., & Claud, C. (2018). Comparison of total water

vapour content in the Arctic derived from GNSS, AIRS, MODIS and SCIAMACHY. *Atmospheric Measurement Techniques*, *11*, 2949–2965. doi:10.5194/amt-11-2949-2018.

Antón, M., Loyola, D., Román, R., & Vömel, H. (2015). Validation of GOME-2/MetOp-A total water vapour column using reference radiosonde data from the GRUAN network. *Atmospheric Measurement Techniques*, *8*, 1135–1145. doi:10.5194/amt-8-1135-2015.

Arnold, J. B. (2018). *ggthemes: Extra Themes, Scales and Geoms for 'ggplot2'*. URL: https://CRAN.R-project.org/package=ggthemes r package version 3.4.2.

Aumann, H. H., Chahine, M. T., Gautier, C., Goldberg, M. D., Kalnay, E., McMillin, L. M., Revercomb, H., Rosenkranz, P. W., Smith, W. L., Staelin, D. H., Strow, L. L., & Susskind, J. (2003). AIRS / AMSU / HSB on the Aqua Mission : Design , Science Objectives , Data Products , and Processing Systems. *Processing*, *41*, 253–264.

Bedka, S., Knuteson, R., Revercomb, H., Tobin, D., & Turner, D. (2010). An assessment of the absolute accuracy of the Atmospheric Infrared Sounder v5 precipitable water vapor product at tropical, midlatitude, and arctic ground-truth sites: September 2002 through August 2008. *Journal of Geophysical Research*, *115*. doi:10.1029/2009JD013139.

Bengtsson, L., Hodges, K. I., Koumoutsaris, S., Zahn, M., & Keenlyside, N. (2011). The changing atmospheric water cycle in Polar Regions in a warmer climate. *Tellus A: Dynamic Meteorology and Oceanography*, *63*, 907–920. doi:10.1111/j.1600-0870.2011.00534.x.

Bennouna, Y. S., Torres, B., Cachorro, V. E., Ortiz de Galisteo, J. P., & Toledano, C. (2013). The evaluation of the integrated water vapour annual cycle over the Iberian Peninsula from EOS-MODIS against different ground-based techniques. *Quarterly Journal of the Royal Meteorological Society*, *139*, 1935–1956. doi:10.1002/qj.2080.

Berezin, I. A., Timofeyev, Y. M., Virolainen, Y. A., & Volkova, K. A. (2016). Comparison of ground-based microwave measurements of precipitable water vapor with radiosounding data. *Atmospheric and Oceanic Optics*, *29*, 274–281. doi:10.1134/S1024856016030040.

Bevis, M., Businger, S., Herring, T. A., Rocken, C., Anthes, R. A., & Ware, R. H. (1992). GPS Meteorology: Remote Sensing of Atmospheric Water Vapor Using the Global Positioning System. *Journal of Geophysical Research*, *97*, 15787–15801. doi:10.1029/92JD01517.

Bintanja, R., van der Wiel, K., van der Linden, E. C., Reusen, J., Bogerd, L., Krikken, F., & Selten, F. M. (2020). Strong future increases in Arctic precipitation variability linked to poleward moisture transport. *Science Advances*, *6*, eaax6869. doi:10.1126/sciadv.aax6869.

Boehm, J., Niell, A., Tregoning, P., & Schuh, H. (2006a). Global Mapping Function (GMF): A new empirical mapping function based on numerical weather model data. *Geophysical Research Letters*, *33*, 3–6. doi:10.1029/2005GL025546.

Boehm, J., Werl, B., & Schuh, H. (2006b). Troposphere mapping functions for GPS and very long baseline interferometry from European Centre for Medium-Range Weather Forecasts operational analysis data. *Journal of Geophysical Research: Solid Earth*, *111*, 1–9. doi:10.1029/2005JB003629.

Bogerd, L., Linden, E. C., Krikken, F., & Bintanja, R. (2020). Climate State Dependence of Arctic Precipitation Variability. *Journal of Geophysical Research: Atmospheres*, *125*. doi:10.1029/2019JD031772.

Bovensmann, H., Burrows, J. P., Buchwitz, M., Frerick, J., Noël, S., Rozanov, V. V., Chance, K. V., & a. P. H. Goede (1999). SCIAMACHY: Mission Objectives and Measurement Modes. *Journal of the Atmospheric Sciences*, *56*, 127–150. doi:10.1175/1520-0469(1999)056<0127:SMOAMM>2.0.CO;2.

Buehler, S. A., Östman, S., Melsheimer, C., Holl, G., Eliasson, S., John, V. O., Blumenstock, T., Hase, F., Elgered, G., Raffalski, U., Nasuno, T., Satoh, M., Milz, M., & Mendrok, J. (2012). A multi-instrument comparison of integrated water vapour measurements at a high latitude site. *Atmospheric Chemistry and Physics*, *12*, 10925–10943. doi:10.5194/acp-12-10925-2012.

Callies, J., Corpaccioli, E., Eisinger, M., a. Hahne, & a. Lefebvre (2000). GOME-2 - Metop's second-generation sensor for operational ozone monitoring. *ESA Bulletin-European Space Agency*, *102*, 28–36.

Carbajal Henken, C., Dirks, L., Steinke, S., Diedrich, H., August, T., & Crewell, S. (2020). Assessment of Sampling Effects on Various Satellite-Derived Integrated Water Vapor Datasets Using GPS Measurements in Germany as Reference. *Remote Sensing*, *12*, 1170. doi:10.3390/rs12071170.

Chang, L., Gao, G., Jin, S., He, X., Xiao, R., & Guo, L. (2015). Calibration and Evaluation of Precipitable Water Vapor From MODIS Infrared Observations at Night. *IEEE Transactions on Geoscience and Remote Sensing*, *53*, 2612–2620. doi:10.1109/TGRS.2014.2363089.

Colman, R. (2003). A comparison of climate feedbacks in general circulation models. *Climate Dynamics*, *20*, 865–873. doi:10.1007/s00382-003-0310-z.

Colman, R. A. (2015). Climate radiative feedbacks and adjustments at the Earth's surface. *Journal of Geophysical Research: Atmospheres*, *120*, 3173–3182. doi:10.1002/2014JD022896.

Davis, J. L., Herring, T. A., Shapiro, I. I., Rogers, A. E. E., & Elgered, G. (1985). Geodesy by radio interferometry: Effects of atmospheric modeling errors on estimates of baseline length. *Radio Science*, *20*, 1593–1607. doi:10.1029/RS020i006p01593.

Deschamps, P.-Y., Breon, F.-M., Leroy, M., Podaire, A., Bricaud, A., Buriez, J.-C., & Seze, G. (1994). The POLDER mission: Instrument characteristics and

scientific objectives. *IEEE Transactions on Geoscience and Remote Sensing*, *32*, 598–615. doi:10.1109/36.297978.

du Piesanie, A., Piters, A. J. M., Aben, I., Schrijver, H., Wang, P., & Noël, S. (2013). Validation of two independent retrievals of SCIAMACHY water vapour columns using radiosonde data. *Atmospheric Measurement Techniques*, *6*, 2925–2940. doi:10.5194/amt-6-2925-2013.

Fionda, E., Cadeddu, M., Mattioli, V., & Pacione, R. (2019). Intercomparison of Integrated Water Vapor Measurements at High Latitudes from Co-Located and Near-Located Instruments. *Remote Sensing*, *11*, 2130. doi:10.3390/rs11182130.

Gaffen, D. J. (1994). Temporal inhomogeneities in radiosonde temperature records. *Journal of Geophysical Research*, *99*, 3667. doi:10.1029/93JD03179.

Gao, B.-C., & Kaufman, Y. J. (1992). The MODIS Near-IR Water Vapor Algorithm Product ID : MOD05 - Total Precipitable Water. *Algorithm Technical Background Document*, (pp. 1–25).

Gao, B.-C., & Li, R.-R. (2008). The Time Series of Terra and Aqua MODIS Near-IR Water Vapor Products. In *IGARSS 2008 - 2008 IEEE International Geoscience and Remote Sensing Symposium* (pp. 186—- 189). IEEE volume 3. doi:10.1109/IGARSS.2008.4779314.

GCOS (2010). *Implementation Plan For the Global Observing System for Climate in Support of the UNFCCC*. Technical Report GCOS-138.

Grolemund, G., & Wickham, H. (2011). Dates and times made easy with lubridate. *Journal of Statistical Software*, *40*, 1–25. URL: http://www.jstatsoft.org/v40/i03/.

Grossi, M., Valks, P., Loyola, D., Aberle, B., Slijkhuis, S., Wagner, T., Beirle, S., & Lang, R. (2015). Total column water vapour measurements from GOME-2 MetOp-A and MetOp-B. *Atmospheric Measurement Techniques*, *8*, 1111–1133. doi:10.5194/amt-8-1111-2015.

Guerova, G., Jones, J., Douša, J., Dick, G., de Haan, S., Pottiaux, E., Bock, O., Pacione, R., Elgered, G., Vedel, H., & Bender, M. (2016). Review of the state of the art and future prospects of the ground-based GNSS meteorology in Europe. *Atmospheric Measurement Techniques*, *9*, 5385–5406. doi:10.5194/amt-9-5385-2016.

Hagan, D. E., Webster, C. R., Farmer, C. B., May, R. D., Herman, R. L., Weinstock, E. M., Christensen, L. E., Lait, L. R., & Newman, P. A. (2004). Validating AIRS upper atmosphere water vapor retrievals using aircraft and balloon in situ measurements. *Geophysical Research Letters*, *31*, n/a–n/a. doi:10.1029/2004GL020302.

IPCC (Ed.) (2014). *Climate Change 2013 - The Physical Science Basis: Working Group I Contribution to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge: Cambridge University Press. doi:10.1017/CBO9781107415324.

Kalakoski, N., Kujanpää, J., Sofieva, V., Tamminen, J., Grossi, M., & Valks, P. (2016). Validation of GOME-2/Metop total column water vapour with ground-based and in situ measurements. *Atmospheric Measurement Techniques*, *9*, 1533–1544. doi:10.5194/amt-9-1533-2016.

Kassambara, A. (2019). *ggpubr: 'ggplot2' Based Publication Ready Plots*. URL: https://CRAN.R-project.org/package=ggpubr r package version 0.2.4.

King, M., Kaufman, Y., Menzel, W., & Tanre, D. (1992). Remote sensing of cloud, aerosol, and water vapor properties from the moderate resolution imaging spectrometer (MODIS). *IEEE Transactions on Geoscience and Remote Sensing*, *30*, 2–27. doi:10.1109/36.124212.

Kokhanovsky, A., & Rozanov, V. (2008). The uncertainties of satellite DOAS total ozone retrieval for a cloudy sky. *Atmospheric Research*, *87*, 27–36. doi:10.1016/j.atmosres.2007.04.006.

Köpken, C. (2001). Validation of Integrated Water Vapor from Numerical Models Using Ground-Based GPS, SSM/I, and Water Vapor Radiometer Measurements. *Journal of Applied Meteorology*, *40*, 1105–1117. doi:10.1175/1520-0450(2001)040<1105:VOIWVF>2.0.CO;2.

Levelt, P., van den Oord, G., Dobber, M., Malkki, A., Huib Visser, Johan de Vries, Stammes, P., Lundell, J., & Saari, H. (2006). The ozone monitoring instrument. *IEEE Transactions on Geoscience and Remote Sensing*, *44*, 1093–1101. doi:10.1109/TGRS.2006.872333.

Li, Z., Muller, J.-P., & Cross, P. (2003). Comparison of precipitable water vapor derived from radiosonde, GPS, and Moderate-Resolution Imaging Spectroradiometer measurements. *Journal of Geophysical Research*, *108*, 4651. doi:10.1029/2003JD003372.

Maturilli, M., Herber, A., & König-Langlo, G. (2013). Climatology and time series of surface meteorology in Ny-Ålesund, Svalbard. *Earth System Science Data*, *5*, 155–163. doi:10.5194/essd-5-155-2013.

Mewes, D., & Jacobi, C. (2019). Heat transport pathways into the Arctic and their connections to surface air temperatures. *Atmospheric Chemistry and Physics*, *19*, 3927–3937. doi:10.5194/acp-19-3927-2019.

Milstein, A. B., & Blackwell, W. J. (2016). Neural network temperature and moisture retrieval algorithm validation for AIRS/AMSU and CrIS/ATMS. *Journal of Geophysical Research: Atmospheres*, *121*, 1414–1430. doi:10.1002/2015JD024008.

Myhre, G., Shindell, D., Bréon, F.-M., Collins, W., Fuglestvedt, J., Huang, J., Koch, D., Lamarque, J.-F., Lee, D., Mendoza, B., Nakajima, T., Robock, A., Stephens, G., Takemura, T., & Zhang, H. (2013). Anthropogenic and Natural Radiative Forcing. In *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (pp. 659–740). IPCC. (Ipcc ed.).

Neuwirth, E. (2014). *RColorBrewer: ColorBrewer Palettes*. URL: https://CRAN.R-project.org/package=RColorBrewer r package version 1.1-2.

Niell, A. E. (2000). Improved atmospheric mapping functions for VLBI and GPS. *Earth, Planets and Space*, *52*, 699–702. doi:10.1186/BF03352267.

Ningombam, S. S., Jade, S., Shrungeshwara, T., & Song, H.-J. (2016). Validation of water vapor retrieval from Moderate Resolution Imaging Spectroradiometer (MODIS) in near infrared channels using GPS data over IAO-Hanle, in the trans-Himalayan region. *Journal of Atmospheric and Solar-Terrestrial Physics*, *137*, 76–85. doi:10.1016/j.jastp.2015.11.019.

Noël, S., Buchwitz, M., Bovensmann, H., & Burrows, J. P. (2005). Validation of SCIAMACHY AMC-DOAS water vapour columns. *Atmospheric Chemistry and Physics*, *5*, 1835–1841. doi:10.5194/acp-5-1835-2005.

Noël, S., Buchwitz, M., & Burrows, J. P. (2004). First retrieval of global water vapour column amounts from SCIAMACHY measurements. *Atmospheric Chemistry and Physics*, *4*, 111–125. doi:10.5194/acp-4-111-2004.

Noël, S., Mieruch, S., Bovensmann, H., & Burrows, J. P. (2008). Preliminary results of GOME-2 water vapour retrievals and first applications in polar regions. *Atmospheric Chemistry and Physics*, *8*, 1519–1529. doi:10.5194/acp-8-1519-2008.

Ohtani, R., & Naito, I. (2000). Comparisons of GPS-derived precipitable water vapors with radiosonde observations in Japan. *Journal of Geophysical Research: Atmospheres*, *105*, 26917–26929. doi:10.1029/2000JD900362.

Pałm, M., Melsheimer, C., Noel, S., Heise, S., Notholt, J., Burrows, J., & Schrems, O. (2010). Integrated water vapor above Ny A˚ lesund, Spitsbergen: A multi-sensor intercomparison. *Atmos. Chem. Phys.*, (p. 12).

Prasad, A. K., & Singh, R. P. (2009). Validation of MODIS Terra, AIRS, NCEP/DOE AMIP-II Reanalysis-2, and AERONET Sun photometer derived integrated precipitable water vapor using ground-based GPS receivers

over India. *Journal of Geophysical Research*, *114*, D05107. doi:10.1029/2008JD011230.

R Core Team (2016). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing Vienna, Austria. URL: https://www.R-project.org/.

Rama Varma Raja, M. K., Gutman, S. I., Yoe, J. G., McMillin, L. M., & Zhao, J. (2008). The Validation of AIRS Retrievals of Integrated Precipitable Water Vapor Using Measurements from a Network of Ground-Based GPS Receivers over the Contiguous United States. *Journal of Atmospheric and Oceanic Technology*, *25*, 416–428. doi:10.1175/2007JTECHA889.1.

Román, R., Antón, M., Cachorro, V., Loyola, D., Ortiz de Galisteo, J., de Frutos, A., & Romero-Campos, P. (2015). Comparison of total water vapor column from GOME-2 on MetOp-A against ground-based GPS measurements at the Iberian Peninsula. *Science of The Total Environment*, *533*, 317–328. doi:10.1016/j.scitotenv.2015.06.124.

Saastamoinen, J. (1972). Atmospheric Correction for the Troposphere and Stratosphere in Radio Ranging Satellites. In S. W. Henriksen, A. Mancini, & B. H. Chovitz (Eds.), *Geophysical Monograph Series* (pp. 247–251). Washington, D. C.: American Geophysical Union. doi:10.1029/GM015p0247.

Sakai, T., Nagai, T., Izumi, T., Yoshida, S., & Shoji, Y. (2019). Automated compact mobile Raman lidar for water vapor measurement: Instrument description and validation by comparison with radiosonde, GNSS, and high-resolution objective analysis. *Atmospheric Measurement Techniques*, *12*, 313–326. doi:10.5194/amt-12-313-2019.

Schrijver, H., Gloudemans, A. M. S., Frankenberg, C., & Aben, I. (2009). Water vapour total columns from SCIAMACHY spectra in the 2.36 Mm window. *Atmospheric Measurement Techniques*, *2*, 561–571. doi:10.5194/amt-2-561-2009.

Seemann, S. W., Borbas, E. E., Li, J., Menzel, W. P., & Gumley, L. E. (2006). Technical Report October Cooperative Institute for Meteorological Satellite Studies, University of Wisconsin-Madison.

Seemann, S. W., J. Li, W. P. M., & L. E. Gumley, . (2003). Operational retrieval of atmospheric temperature, moisture, and ozone from MODIS infrared radiances. *Journal of Applied Meteorology*, *42*, 1072–1091.

Susskind, J., Barnet, C., & Blaisdell, J. (2003). Retrieval of atmospheric and surface parameters from AIRS/AMSU/HSB data in the presence of clouds. *IEEE Transactions on Geoscience and Remote Sensing*, *41*, 390–409. doi:10.1109/TGRS.2002.808236.

Susskind, J., Barnet, C., Blaisdell, J., Iredell, L., Keita, F., Kouvaris, L., Molnar, G., & Chahine, M. (2006). Accuracy of geophysical parameters derived from Atmospheric Infrared Sounder/Advanced Microwave Sounding Unit as a function of fractional cloud cover. *Journal of Geophysical Research*, *111*. doi:10.1029/2005JD006272.

Susskind, J., Schmidt, G. A., Lee, J. N., & Iredell, L. (2019). Recent global warming as confirmed by AIRS. *Environmental Research Letters*, *14*, 044030. doi:10.1088/1748-9326/aafd4e.

Thomas, I. D., King, M. A., Clarke, P. J., & Penna, N. T. (2011). Precipitable water vapor estimates from homogeneously reprocessed GPS data: An intertechnique comparison in Antarctica. *Journal of Geophysical Research*, *116*. doi:10.1029/2010JD013889.

Vaquero-Martínez, J., Antón, M., Ortiz de Galisteo, J. P., Cachorro, V. E., Álvarez-Zapatero, P., Román, R., Loyola, D., Costa, M. J., Wang, H., Abad, G. G., & Noël, S. (2018). Inter-comparison of integrated water vapor from satellite instruments using reference GPS data at the Iberian Peninsula. *Remote Sensing of Environment*, *204*, 729–740. doi:10.1016/j.rse.2017.09.028.

Vaquero-Martínez, J., Antón, M., Ortiz de Galisteo, J. P., Cachorro, V. E., Costa, M. J., Román, R., & Bennouna, Y. S. (2017a). Validation of MODIS integrated water vapor product against reference GPS data at the Iberian Peninsula. *International Journal of Applied Earth Observation and Geoinformation*, *63*, 214–221. doi:10.1016/j.jag.2017.07.008.

Vaquero-Martínez, J., Antón, M., Ortiz de Galisteo, J. P., Cachorro, V. E., Wang, H., González Abad, G., Román, R., & Costa, M. J. (2017b). Validation of integrated water vapor from OMI satellite instrument against reference GPS data at the Iberian Peninsula. *Science of The Total Environment*, *580*, 857–864. doi:10.1016/j.scitotenv.2016.12.032.

Vaquero-Martínez, J., Antón, M., Ortiz de Galisteo, J. P., Román, R., Cachorro, V. E., & Mateos, D. (2019). Comparison of integrated water vapor from GNSS and radiosounding at four GRUAN stations. *Science of The Total Environment*, *648*, 1639–1648. doi:10.1016/j.scitotenv.2018.08.192.

Vesperini, M., Breon, F.-M., & Tanre, D. (1999). Atmospheric water vapor content from spaceborne POLDER measurements. *IEEE Transactions on Geoscience and Remote Sensing*, *37*, 1613–1619. doi:10.1109/36.763275.

Wagner, T., Beirle, S., Grzegorski, M., & Platt, U. (2006). Global trends (1996-2003) of total column precipitable water observed by Global Ozone Monitoring Experiment (GOME) on ERS-2 and their relation to near-surface temperature. *Journal of Geophysical Research*, *111*, D12102. doi:10.1029/2005JD006523.

Wagner, T., Heland, J., Zöger, M., & Platt, U. (2003). A fast h2o total column density product from gome: Validation with in-situ aircraft measurements. *Atmospheric Chemistry and Physics Discussions*, *3*, 323–353. doi:10.5194/acpd-3-323-2003.

Wang, H., Souri, A. H., Gonzalez Abad, G., Liu, X., & Chance, K. (2019). OMI Total Column Water Vapor Version 4 Validation and Applications. *At-*

mospheric Measurement Techniques Discussions, (pp. 1–34). doi:10.5194/amt-2019-89.

Wang, J., Carlson, D. J., Parsons, D. B., Hock, T. F., Lauritsen, D., Cole, H. L., Beierle, K., & Chamberlain, E. (2003). Performance of operational radiosonde humidity sensors in direct comparison with a chilled mirror dew-point hygrometer and its climate implication: PERFORMANCE OF OPERATIONAL RADIOSONDE HUMIDITY SENSORS. *Geophysical Research Letters*, *30*. doi:10.1029/2003GL016985.

Wang, J., & Zhang, L. (2008). Systematic Errors in Global Radiosonde Precipitable Water Data from Comparisons with Ground-Based GPS Measurements. *Journal of Climate*, *21*, 2218–2238. doi:10.1175/2007JCLI1944.1.

Wang, J., & Zhang, L. (2009). Climate applications of a global, 2-hourly atmospheric precipitable water dataset derived from IGS tropospheric products. *Journal of Geodesy*, *83*, 209–217. doi:10.1007/s00190-008-0238-5.

Wickham, H. (2017). *tidyverse: Easily Install and Load the 'Tidyverse'*. URL: https://CRAN.R-project.org/package=tidyverse r package version 1.2.1.

Wilcoxon, F. (1945). Individual Comparisons by Ranking Methods. *Biometrics Bulletin*, *1*, 80. doi:10.2307/3001968.

Zhu, H. (2019). *kableExtra: Construct Complex Table with 'kable' and Pipe Syntax*. URL: https://CRAN.R-project.org/package=kableExtra r package version 1.1.0.