**Atmospheric
Chemistry
and Physics**

A multi-instrument comparison of integrated water vapour measurements at a high latitude site

S. A. Buehler[1], S. Östman[1], C. Melsheimer[2], G. Holl[1], S. Eliasson[1], V. O. John[3], T. Blumenstock[4], F. Hase[4], G. Elgered[5], U. Raffalski[6], T. Nasuno[7], M. Satoh[8], M. Milz[1], and J. Mendrok[1]

[1]Department of Computer Science, Electrical and Space Engineering, Luleå University of Technology, Kiruna, Sweden
[2]Institute of Environmental Physics, University of Bremen, Bremen, Germany
[3]Met Office Hadley Centre, Exeter, UK
[4]Karlsruhe Institute of Technology (KIT), Institute for Meteorology and Climate Research (IMK-ASF), Karlsruhe, Germany
[5]Department of Earth and Space Sciences, Chalmers University of Technology, Onsala Space Observatory, Onsala, Sweden
[6]The Swedish Institute of Space Physics, Kiruna, Sweden
[7]Research Institute for Global Change, Japan Agency for Marine-Earth Science and Technology, Yokohama, Japan
[8]Atmosphere and Ocean Research Institute, University of Tokyo, Tokyo, Japan

Correspondence to: S. A. Buehler (sbuehler@ltu.se)

Abstract. We compare measurements of integrated water vapour (IWV) over a subarctic site (Kiruna, Northern Sweden) from five different sensors and retrieval methods: Radiosondes, Global Positioning System (GPS), ground-based Fourier-transform infrared (FTIR) spectrometer, ground-based microwave radiometer, and satellite-based microwave radiometer (AMSU-B). Additionally, we compare also to ERA-Interim model reanalysis data. GPS-based IWV data have the highest temporal coverage and resolution and are chosen as reference data set. All datasets agree reasonably well, but the ground-based microwave instrument only if the data are cloud-filtered. We also address two issues that are general for such intercomparison studies, the impact of different lower altitude limits for the IWV integration, and the impact of representativeness error. We develop methods for correcting for the former, and estimating the random error contribution of the latter. A literature survey reveals that reported systematic differences between different techniques are study-dependent and show no overall consistent pattern. Further improving the absolute accuracy of IWV measurements and providing climate-quality time series therefore remain challenging problems.

1 Introduction

Water vapour is a key component of the climate system: it is highly variable in space and time, and it (1) is a major greenhouse gas, (2) can transport large amounts of latent heat, and (3) is an essential part of the global hydrological cycle. This justifies the great interest in water vapour measurements on all temporal and spatial scales, with in-situ and remote sensing methods.

In this study, we compare integrated water vapour (IWV, also called total column water vapour or precipitable water vapour) measured at Kiruna, Sweden, with five different instruments and methods. The instruments are (1) radiosondes, (2) ground-based Global Positioning System (GPS) measurements, (3) a ground-based Fourier transform infrared spectrometer (FTIR), (4) the ground-based Kiruna Microwave Radiometer (KIMRA), and (5) the satellite-based microwave radiometer AMSU-B. More details on the different datasets can be found in Sect. 2. Furthermore, we compare to ERA-Interim, a reanalysis dataset from the European Centre for Medium-Range Weather Forecasts (ECMWF).

The primary aim of the study is to assess the performance of the different instruments and retrieval methods for IWV at a subarctic site. A long-term data record of IWV would be useful for climate studies. As a step towards establishing

such a data record, we attempt to characterise the systematic and random differences between the different datasets. Although at least a dozen comparison studies have been carried out before (see also Sect. 5), few have focussed on a subarctic site. As the retrieval technologies are different, conclusions valid for (sub)tropical sites cannot be generalised and applied to (sub)arctic sites. Also, most prior studies compare only GPS and radiosonde measurements. Our study focusses on Kiruna, at a latitude of approximately 68° N, roughly 140 km north of the Arctic circle.

Our secondary aims are to study the impact of two issues: the different lower altitude limits for the IWV integration, and the representativeness error, in particular between point measurements and spatially averaged measurements. The method developed here should be of interest for other IWV comparison studies.

This article is partly based on work that was done by Simon Östman for his Master Thesis (Östman, 2010). Sections that are based on text in the thesis are indicated as such.

The article is structured as follows: Sect. 2 introduces the different instruments and datasets. Section 3 then discusses various methodological aspects of the study. Section 4 presents the bulk of the results and Sect. 5 contains their discussion. Finally, Sect. 6 contains a summary and the conclusions.

2   Instruments and datasets[1]

The positions of the various ground based instruments, and the selected locations for the global satellite and model data, are shown in Fig. 1. Table 1 lists the exact positions. We used data from 1996 to 2008, but not all datasets cover this entire period. Figure 2 shows an overview time series of the data themselves. Table 2 gives a summary of key dataset properties. The individual datasets are discussed below.

2.1   "GPS" – ground-based GPS data

The GPS data were acquired at the Esrange Space Center near Kiruna (see Fig. 1 for a map and Table 1 for the exact coordinates). It is operated by SWEPOS, the national network of permanent reference stations for GPS in Sweden (http://swepos.lmv.lm.se/english/). The site is of geodetic quality, meaning that the instrument is mounted on a concrete column fixed to solid bedrock. The data and the corresponding analysis are described in detail by Nilsson and Elgered (2008). Here we present a summary.

Our processed GPS dataset covers the ten year period from November 1996 to November 2006 with continuous measurements (the instrument itself continues to operate). The parameter provided by the GPS processing is the excess propagation path through the atmosphere in the zenith

---

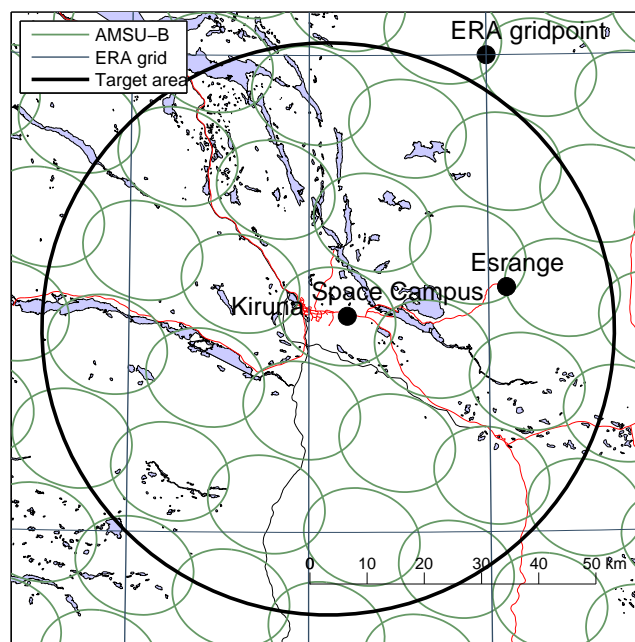[1]The text in this section is a revised and shortened version of Östman (2010, chapter 2).

Fig. 1. A map of the intercomparison area. The map indicates the location of the measurement sites and the ERA-grid. The two measurement sites at Esrange are shown by a single dot. The AMSU-B footprints illustrate a single overpass; the complete set of measurements used in this study do not lie on a grid, but are spread though the target area. The background map data are[©] OpenStreetMap contributors, CC-BY-SA.

direction, also called the Zenith Total Delay (ZTD). Actually measured is the delay for the radio links towards the different GPS satellites, which are seen at different viewing angles. These are converted to equivalent delays in the zenith direction and averaged in 2 h time intervals. So, ZTD represents an average of the time period between 1 h before and 1 h after the nominal measurement time.

The ZTD can be divided into the Zenith Hydrostatic Delay (ZHD) and the Zenith Wet Delay (ZWD). The IWV can then be inferred from the ZWD. Since the standard deviation is based on an uncertainty in the ZTD measurement and since the ZWD is only a small fraction of the total delay, it will be approximately the same for both high and low values of ZWD. This in turn means that the standard deviation for the IWV also will be approximately the same for all measurements. The relative uncertainty will thus be high for low values of IWV, but low for high values of IWV.

Important error sources affecting estimates of the ZTD – and hence also the IWV – are antenna phase centre variations and multipath effects (see e.g. Jarlemark et al., 2010; King and Watson, 2010). A natural effect in this dataset is the sensitivity of the received signal phase to wet snow that in extreme cases can stick to the radome covering the GPS antenna. When such weather conditions occur, site staff regularly clean the instrument. Unlike multipath effects, the

Table 1. Position (latitude, longitue, and altitude) of all datasets. The altitude is the starting altitude of the IWV measurement that we have assumed. Note that our AMSU-B data are areal averages, the position in the table is the center of the averaging area. For ERA-Interim, the selected grid point is the one closest to the GPS. The last column, f_z, gives the correction factor that is applied to this dataset in order to account for the different altitudes, which is derived in Sect. 3.4. Note that the correct GPS receiver altitude is 469 m as stated in Nilsson and Elgered (2008), but 470 m is the value assumed in our analysis. Likewise, the actual ERA-Interim topology altitude is 557 m, but we have assumed 556 m. These small altitude differences have a negligible impact on the results.

Dataset	Location	Latitude	Longitude	Altitude	f_z
GPS	Esrange	67.88° N	21.05° E	470 m	1
Radiosonde	Esrange	67.89° N	21.08° E	335 m	0.953
FTIR	Space Campus	67.84° N	20.41° E	420 m	0.982
Microwave	Space Campus	67.84° N	20.41° E	420 m	0.982
AMSU-B	Kiruna airport	67.82° N	20.33° E	506 m	1.013
ERA-Interim	Selected gridpoint	68.25° N	21.00° E	556 m	1.033

Table 2. Dataset properties. MW = microwave, IR = infrared. The column "Period" shows the time period covered by the dataset in this study. The column "N" shows the total number of measurements used in this study. Column "Restrictions" indicates restrictions for the various datasets, which either follow from the instrument descriptions in Sect. 2, or from our study results (Sects. 4 and 5).

Dataset	Type	Period	N	Restrictions
Ground based at Esrange				
GPS	active MW	1996–2006	41 316	
Radiosonde	in situ	2003–2008	214	
Ground based at Space Campus				
FTIR	passive IR	1996–2008	4151	clear-sky
Microwave	passive MW	2002–2008*	4398	clear-sky
Satellite				
AMSU-B	passive MW	2003–2008	5449	IWV $< 8\,\mathrm{kg\,m^{-2}}$
Model reanalysis				
ERA-Interim	model reanalysis	1979–2012	21 916	

* Divided into two datasets, covering the periods 2002–2005, and 2005–2008.

antenna phase centre variation (PCV) has a static behaviour and can therefore be calibrated using an absolute phase centre correction model, which gives a mean offset of the electrical antenna phase centre compared to the physical antenna reference point, as well as corrections as a function of the elevation angle (Schmid et al., 2007). The data used here were not analysed using the absolute phase centre model, meaning that a bias type of error of the order of up to approximately $1\,\mathrm{kg\,m^{-2}}$ in the IWV may be identified and removed in the future (Ning et al., 2012).

2.2 "Radiosonde" – radiosonde data

Radiosonde profiles used in this study come also from Esrange Space Center. The launch site is close to the GPS receiver, but at significantly lower altitude (see Table 1), because the GPS receiver is on a hill. The implications of different measurement altitudes will be discussed in Sect. 3.4.

Launches are only performed regularly during periods of balloon and sounding rocket campaigns and are therefore limited to short periods. The sondes are all of the Vaisala RS-80, RS-90 or RS-92 types and measure pressure, temperature and relative humidity. Continuous measurements are made during the ascent of the instrument with a two or ten second interval, depending on the setup of the sounding processing system. The following parameters are then available for each measured point in the profile resulting from the flight: flight time [min+sec], pressure [hPa], geopotential height [gpm], temperature [°C], relative humidity [% RH] and dew point [°C].

The total number of radiosonde profiles available for the study is 214 during the time period 2003–2008. Launching campaigns are generally performed from early spring and through the summer months which can be seen in Fig. 2. Radiosonde launches are performed at all times of the day but most often in the early morning (04:00–06:00 UTC). For converting relative humidity to water vapour pressure,
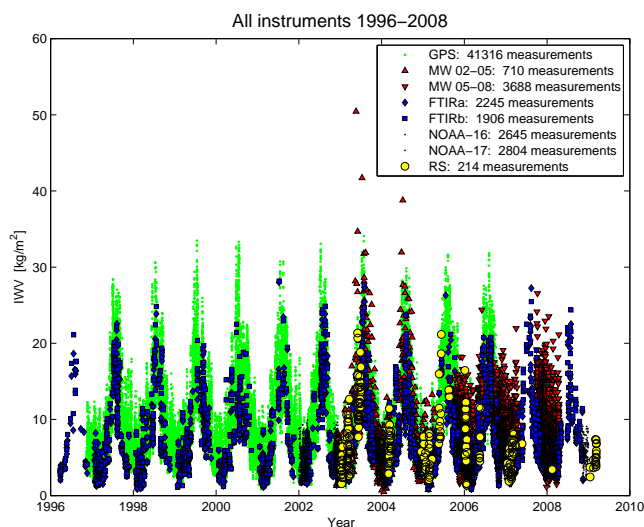
Fig. 2. Overview time series of all integrated water vapour (IWV) data used in this study. Only the ERA-Interim data are not shown, since they would make the figure too crowded.

the equilibrium water vapour pressure according to Sonntag (1994) was used. Vertical integration (with respect to geopotential height) then yields IWV for each radiosonde profile.

Conventional radiosondes do not reliably measure water vapour at high altitudes (close to or above the tropopause), due to the low temperatures and low humidity concentrations there. However, the total column water vapour value is strongly dominated by the high water vapour concentrations near the surface. We therefore here assume that the radiosonde IWV values are accurate.

2.3 "FTIR" – ground-based FTIR data

Atmospheric trace gas measurements using Fourier Transform Infrared (FTIR) spectroscopy have been routinely performed at the Swedish Institute of Space Physics (IRF) since 1996. The location of IRF is at Kiruna Space Campus, approximately 30 km west of Esrange, approximately 8 km from Kiruna town centre (see map in Fig. 1).

Solar absorption spectra are recorded with a Bruker 120 HR spectrometer that is part of a multi-national network of over twenty high resolution FTIR spectrometers called the Network for the Detection of Atmospheric Composition Change (NDACC) (Blumenstock et al., 2006).

When performing Fourier transform infrared spectroscopy, the sun is used as source of radiation (i.e. the instrument is aimed directly at the sun). This means that the technique requires cloud-free conditions, and measurements are limited to times when the sun is above the horizon. Note that the requirement of cloud-free conditions can lead to a sampling dry bias in the data set, but this is not a problem when comparing to other simultaneous measurements, since they will have the same bias.

Spectra are integrated for up to 15 min during noon and 5 min during sunrise and sunset to limit the solar zenith angle variation to 0.2°. The signal to noise ratio of the recorded transmission spectra amounts to several hundreds. The recorded spectra are compared to simulated radiances based on a radiative transfer model which uses atmospheric profiles and spectroscopic line data (HITRAN 2008). The resulting data are total column amounts of the different atmospheric constituents.

As water vapour is highly variable and its total column amount varies by almost 2 orders of magnitude, the retrieval is a demanding task and one can not apply a retrieval based on a single water vapour line. The analysis of the FTIR spectra has been performed with the retrieval code PROF-FIT developed at IMK-ASF (Hase et al., 2004). Schneider and Hase (2009) have shown that by combining several different strong and weak water vapour lines it is possible to achieve high accuracy in the results. Two detectors, one HgCdTe (Mercury-Cadmium-Telluride, MCT) and one InSb (Indium-Antimonide), cover the spectral ranges of 700–1300 cm^{-1} and 1800–5000 cm^{-1}, respectively. This corresponds to wavelength ranges of about 14.3–7.7 µm and 5.5–2.0 µm, respectively.

The FTIR dataset used here consists of the complete measurement series from the IRF Kiruna spectrometer between 1996 and 2008. The data are divided into two sets corresponding to the measurements of the two detectors. We call the sets FTIRa (MCT detector) and FTIRb (InSb detector).

2.4 "Microwave" – ground-based passive microwave data

Since January 2002 IRF Kiruna has been operating their own millimetre wave radiometer called KIMRA, KIruna Microwave RAdiometer (Raffalski et al., 2002). It is in the same building as the FTIR instrument. The instrument was built in cooperation with the Institute for Meteorology and Climate Research at Karlsruhe Institute of Technology (KIT).

The instrument measures thermal emission spectra in the 195–235 GHz range and has the main purpose to monitor stratospheric trace gases. It was originally optimised for ClO measurements. The design also allows the observation of O_3, CO, HNO_3, and N_2O. However, it is now being operated in the "O_3 and CO" mode only because baseline effects interfere with the signatures of the other trace gas measurements.

The instrument allows full coverage of all azimuth and elevation angles above the horizon, since its periscope-like mirror system sticks out through the roof of the IRF building. However, by default the measurement points northward with an elevation angle between 7° and 55°. The integration time for ozone observations is between 20 min and 3 h, for other trace gases it is up to several hours. Data analysis was performed by the millimetre wave group at Karlsruhe Institute of Technology (http://www.irf.se/program/afp/mm/). The IWV data is a by-product of the stratospheric ozone

retrieval, therefore, the water vapour measurements have the same integration time as the ozone measurements.

The actual IWV retrieval procedure is as follows: in order to retrieve ozone profiles the measured spectrum has to be corrected for the varying concentration of tropospheric water vapour, which results in an offset and a scaling of the ozone line. The first step of this correction is a radiative transfer calculation for a standard water vapour profile with the actual temperature and pressure profiles (from the ECMWF analysis). In a second step the water vapour profile is scaled, in order to match the offset shown in the measurement. For the ozone retrieval this offset is subtracted while the stratospheric part of the spectrum is scaled by the tropospheric transmission. IWV is calculated by integrating over the scaled water vapor profile.

The microwave IWV data come in two different sets of which the first covers the period from autumn 2002 till spring 2005 and the second covers autumn 2005 till spring 2008, as shown in Fig. 2. The reason why the two sets do not look similar even though all data come from the same instrument is that details of the retrieval were different: during the first measurement period (2002–2005), out of several measured spectra per day, the one which was best suited for the retrieval of O_3 was selected manually. This means that typically cases with low tropospheric opacity were selected (i.e. low water vapour).

During the second measurement period (2005–2008), several spectra per day were used, hence the larger number of samples and the large scatter. In addition, during the summer months, the radiometer data were not processed, hence the two gaps in the second time series. During the whole measurement series, the 195 GHz line was used for the measurements, except for winter 2007/2008, when parallel observations of CO and O_3 at 231 GHz were carried out.

The instrument used in this study should not be confused with conventional microwave radiometers dedicated to measurements of tropospheric temperature and humidity, which operate at much lower frequency. In their simplest configuration, already described by Westwater (1978), they have two channels, one close to the 22.235 GHz water vapour line, and one at higher frequency, typically around 30 GHz. As explained for example in Rose et al. (2005), the two channels of such instruments can be used to independently measure IWV and the liquid water path (LWP). This is not possible for our instrument, which leads to problems in the presence of clouds, as discussed in Sect. 4.

2.5 "AMSU-B" – AMSU-B satellite data

The Advanced Microwave Sounding Unit-B (AMSU-B) is on the polar orbiting meteorological satellites of the National Oceanic and Atmospheric Administration (NOAA). For the instrument summary in this paragraph we are quoting almost literally from our previous articles using AMSU-B data, of which Buehler et al. (2004) is the earliest and Kot-

tayil et al. (2012) the latest: AMSU-B is a cross-track line scanning, passive, total power microwave radiometer with five channels. Three channels are centred on the strong water vapour line at 183.31 GHz with a sideband spacing of ± 1, ± 3 and ± 7 GHz, respectively. The remaining two channels are window channels centred at 150 and 89 GHz. Each swath consists of 90 samples with a sampling distance of $1.10°$ resulting in a total viewing angle range of $-48.95°$ to $+48.95°$ around nadir (Saunders et al., 1995). The swath width is about 2300 km and the footprint size ranges from $20 \times 16 \, \text{km}^2$ at nadir to $64 \times 52 \, \text{km}^2$ at the most extreme scan angles.

In tropical and midlatitude regions, AMSU-B can be used for humidity profiling, and particularly to measure humidity in the upper troposphere (e.g., Buehler et al., 2008). In polar regions, the atmosphere is too dry for this to work, but AMSU-B data can be used to retrieve the water vapour column with the method described by Melsheimer and Heygster (2008): by using three channels where the surface emissivity is equal but where the water vapour absorption is different, it is possible to derive a relationship between the measured brightness temperatures and the IWV. The method works as long as the atmosphere is not opaque for any of the three channels, i.e. as long as they "see" the surface. Using the three AMSU-B channels at 183 ± 1, 183 ± 3 and 183 ± 7 GHz, IWV values up to about $1.5 \, \text{kg m}^{-2}$ can thus be retrieved. Beyond that, the most water-vapour sensitive channel at 183 ± 1 GHz becomes saturated. Using the channels at 183 ± 3 and 183 ± 7 GHz and at 150 GHz, retrieval of IWV is possible up to about $8 \, \text{kg m}^{-2}$. Using the 89 GHz channel to extend the retrieval range to even higher IWV values requires knowledge about the surface emissivity which has so far only been implemented for sea ice (Melsheimer and Heygster, 2008). Therefore, for higher IWV values, typical for most regions except the polar regions, the retrieval fails except over sea ice.

The AMSU-B IWV data for our study were extracted from the dataset in the same way as for earlier work where we compared AMSU-B upper tropospheric humidity measurements to radiosonde measurements (Buehler et al., 2004; John and Buehler, 2005; Moradi et al., 2010; Kottayil et al., 2012): for each satellite pass that covers the Kiruna area, IWV from all satellite footprints for which the center is within a target area of 50 km radius around Kiruna airport (67.82° N, 20.33° E) were averaged and the standard deviation of this average was calculated (see Fig. 1). The number of footprints that were averaged was up to 35, typically 10 to 20 – depending on how centrally Kiruna was located in the satellite swath. As the algorithm only works up to IWV of about $8 \, \text{kg m}^{-2}$, the period of useful data was restricted to the winter months of November to March. Outside this period, the IWV is above $8 \, \text{kg m}^{-2}$ most of the time. The data processed are from NOAA-16 and NOAA-17, covering the years 2002–2006 and 2002–2008, respectively. Because of generally poorer data quality, data from NOAA-15 were not

used. The minimum number of required IWV values in the target area for calculating the average was set to ten in order to have a meaningful standard deviation. Additionally, this threshold assures that in situations of high variability in the target area (for example because of a meteorological front), this variability results in an increased standard deviation and can be taken into account in the analysis. We analysed the two datasets for NOAA-16 and NOAA-17 separately.

2.6 "ERA-Interim" – ERA-Interim reanalysis data

The ERA-Interim reanalysis dataset (Dee et al., 2011) is produced by the European Centre for Medium-Range Weather Forecasts (ECMWF). It is the latest atmospheric reanalysis by ECMWF and temporal coverage of the data is from 1979 to present. Unlike normal observations, reanalysis data yields spatially complete and coherent records of meteorological variables and IWV is one of the many such variables. The horizontal resolution of ERA-Interim is $0.75° \times 0.75°$ in latitude and longitude. We obtained 6-hourly IWV estimates (00:00, 06:00, 12:00, and 18:00 UTC) for the grid point closest to the GPS station. The coordinates of this grid point are ($21.00°$ E, $68.25°$ N), and the surface altitude from the model topography is 556.81 m. When comparing these data to actual measurements at the same position, there is expected to be a representativeness error due to the limited model resolution. This will be discussed further in Sect. 3.5.

3 Methodology

When trying to compare different IWV datasets for roughly the same location, we are faced with several challenges: firstly, we have to decide which of the datasets to use as reference, at least for the sake of a compact presentation (Sect. 3.1). Secondly, we have to decide which statistical parameters to use for the comparison (Sect. 3.2). Thirdly, we have to define and select temporal matches (Sect. 3.3) and fourthly, deal with different lower altitude limits of the IWV integration (Sect. 3.4). Lastly, we have to address the important issues of representativeness and the impact of horizontal distance and averaging area (Sect. 3.5).

3.1 Reference dataset[2]

In order to present the results in a compact way, we decided to select one of the datasets as reference. The GPS dataset was chosen for this, for the simple practical reason that it has the largest number of coincident data with all other datasets, due to its regular two hourly temporal sampling. However, this does not mean that we automatically regard the GPS dataset as truth. Possible biases of that dataset are also discussed in the results section below.

3.2 Statistical parameters

After temporal matching of the data (see next section), we get for each dataset a time series of pairs of those data with the GPS data. This time series of data pairs is input to our statistical analysis. Some statistical parameters are calculated, and are reported in Sect. 4, and in particular in Table 4.

Firstly, we calculate the absolute and relative difference between the two datasets:

$$\text{diff}(i) = \text{IWV}_X(i) - \text{IWV}_{GPS}(i) , \tag{1}$$

$$\text{reldiff}(i) = \frac{\text{diff}(i)}{\text{IWV}_{GPS}(i)} , \tag{2}$$

where $\text{IWV}_X(i)$ and $\text{IWV}_{GPS}(i)$ are the IWV data of the other dataset and the GPS dataset, respectively. (Note that absolute here means the opposite of relative, not that the absolute value is taken. Both values can be positive or negative.) In most cases the absolute difference is more appropriate than the relative difference to characterize the error between these datasets. This is apparent from the scatterplots shown in Sect. 4, which show random errors that do not scale with IWV value. We therefore mostly base our analysis on the absolute difference, except in the discussion of the lower integration altitude limit in Sect. 3.4. Typically, we report and discuss the mean and standard deviation of the difference, along with the total number of matches and the correlation coefficient.

Secondly, we did a linear regression on the paired data, and report the resulting slope and offset. The regression algorithm used can take into account errors on both datasets where available. It is described in Krystek and Anton (2007). We used the Matlab code by Anton, downloaded from http://www.mathworks.com/matlabcentral/fileexchange/17466.

For most datasets only general error estimates are available. These constant error estimates do not affect the regression result and were therefore not explicitly used in the regression. But the GPS dataset includes an explicit error estimate for each data point, and that was taken into account in the regression. For AMSU-B, the standard deviation of the pixels inside the 50 km radius target area was taken as error estimate in the regression.

3.3 Temporal matching criteria[3]

To be able to compare the different IWV datasets, a method for finding matching measurements is needed. The goal is to find occasions where two instruments measure IWV in the same air mass. Therefore, a matching criterion has to be set to limit the maximum time difference allowed between two individual measurements. This time difference must not be too large because then the two values compared will not represent the same conditions in the atmosphere. On the other

---

[2]This text in this subsection is a revised version of Östman (2010, Sect. 3.3).

[3]This text in this subsection is a revised version of Östman (2010, Sect. 3.1).

hand it must not be too small either as too few matches would then be found.

Besides the purely practical reason of having enough data to work on, there is also a theoretical reason for not making the time window for matches too narrow: the different datasets are not all taken at exactly the same point, and also some are point measurements whereas others are areal averages (these points are discussed further in Sect. 3.5). A good temporal matching criterion is one that is consistent with these spatial scales. The link between the spatial and temporal scales is the wind speed. To keep things simple, we assume a wind speed of $10\,\mathrm{m\,s^{-1}}$ for our sizing arguments below. Although this wind speed can be easily exceeded in individual cases, it is considerably higher than the expected average windspeed.

With this assumed wind speed, an air parcel will move by 36 km per hour. The distance between Esrange and Kiruna is approximately 30 km. The sampling area of the different datasets ranges from practically point measurements (for the radiosonde) to quite large averaging areas (50 km radius circle for AMSU-B, $83 \times 31\,\mathrm{km}^2$ for ERA-Interim).

There is also a horizontal sampling uncertainty, even for the radiosondes, as they drift horizontally during their ascent. Typically, the sondes will reach a height of 5 km in about 15 min. Above this height the water content is so low that it can be ignored for the total column value. At the assumed wind speed of $10\,\mathrm{m\,s^{-1}}$ the sonde will thus travel a horizontal distance of approximately 9 km.

For the ground-based remote sensing instruments (GPS, FTIR, and microwave) there is a similar horizontal sampling uncertainty, since they measure along slant paths of varying elevation angle and horizontal direction. At a height of 5 km, the approximate top of the water vapour column, an elevation angle of 10° above the horizon corresponds to a horizontal distance of 28 km.

To keep things simple, we did not try to set a separate time matching threshold for each pair of measurements, but decided to use a threshold of 1 h for all. So, data with a time difference of less than 1 h are considered a match. This ensures that the typical air mass displacement due to the wind is only a few kilometers.

The matching algorithm works as follows: it looks at two datasets and compares all data point by point. If two measurements, one from each dataset, have time stamps that are within the 1 h time limit from each other, those are considered a matching pair. Here caution has to be taken to get only unique pairs. A data point from one dataset can only be matched to one unique data point from the other set. If there are several matches within the time limit for a certain measurement, only the match with the shortest time difference is chosen.
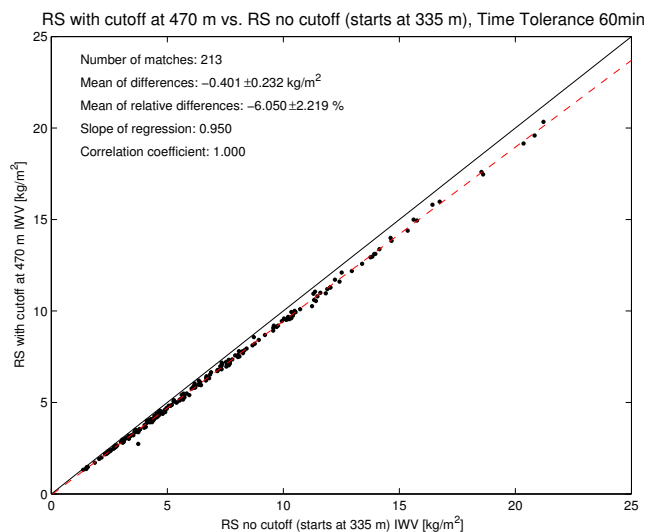

Fig. 3. Scatterplot of radiosonde IWV calculated with a lower integration altitude limit of 470 m versus IWV calculated for the entire radiosonde profile, which starts at an altitude of 335 m. Both "measurements" here are based on the same radiosonde data, only in one case the lowest part of the profile is ignored.

3.4 Impact of lower integration altitude limit

The density of water vapour in the atmosphere decreases approximately exponentially with altitude. When calculating or measuring total column values therefore the lower altitude limit of the integration is very important. If two different sensors are at different altitudes, then the one at the higher altitude will miss the bottom part of the water vapour column and will thus measure a lower IWV value. (Bock et al. (2007) also mention this and develop a correction method for tropical IWV data.)

This effect can be easily studied with the help of the radiosonde data. Since we have the water vapour profile available at high resolution, we can simply start the IWV integration at different altitudes to simulate what a sensor at different altitudes should report as IWV value.

Figure 3 shows this for an assumed sensor altitude of 470 m (the radiosonde measurement itself starts at 335 m). The value of 470 m was chosen because this is the actual altitude of the GPS receiver, which is horizontally very close to the radiosonde launch pad. What the figure shows is a scatterplot of IWV calculated from 470 m upwards (IWV_{470}) versus IWV calculated from 335 m upwards (IWV_{335}), all based on the same radiosonde profile data.

The random error between IWV_{470} and IWV_{335} is small, only approximately 2 % relative error (standard deviation of the relative difference, calculated as described in Sect. 3.2). However, there is a systematic difference, the linear regression line slope a is 0.95. This means that $0.95 \cdot \mathrm{IWV}_{335}$ is a good predictor of IWV_{470}, with a
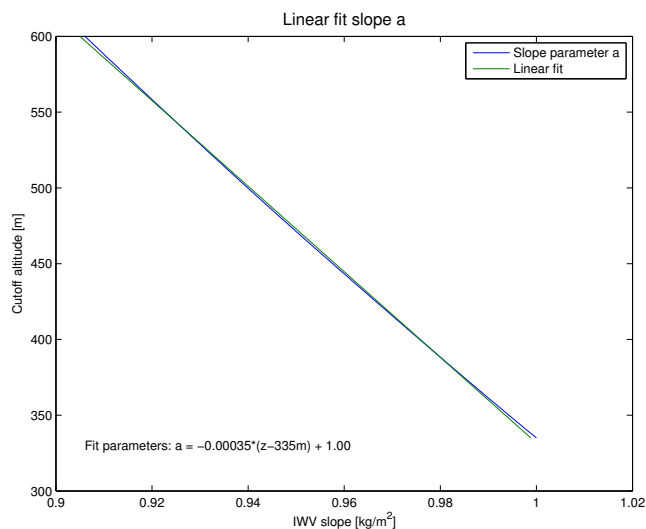
Fig. 4. IWV slope versus lower integration altitude limit, together with a linear fit. As the figure shows, the altitude dependence of the slope can be well approximated by a linear fit.

precision of approximately 2 %. We can use the $a = 0.95$ correction factor to convert between the two measurements.

This can be generalised to other altitude differences. We studied this by keeping IWV_{335} as reference measurement, and doing a linear regression similar to the one in Fig. 3 for a large number of IWV_z, with z ranging from 335 m to 600 m in 1 m steps. The linear regression slopes $a(z)$ for all these different z are shown in Fig. 4. As the figure shows, the slope $a(z)$ itself can be approximated by a linear function of z to very good accuracy. Explicitly, $a(z)$ is given by

$$a(z) = -0.00035 \cdot (z - 335\,\text{m}) + 1. \tag{3}$$

Using the terminology of Bock et al. (2007), we can say that the relative bias $\Delta IWV/IWV$ is -3.5 % per 100 m altitude difference.

Since we use the GPS measurement as the reference, we are converting all other datasets to the altitude of the GPS receiver by multiplying their IWV values with a correction factor f_z that follows directly from the regression slope a in the above discussion. The altitudes that are assumed for the various datasets, and the corresponding correction factors f_z, are listed in Table 1. For the ground-based instruments, the assumed altitude simply is the actual altitude of the instrument. For the radiosonde data, the altitude (335 m) is the one where we start calculating IWV (some profiles start a few meters lower, but then the lowest meters are not used).

For AMSU-B data the reference altitude is the average surface altitude inside the 50 km radius circular target area, which was calculated from ETOPO1 global relief model data (Amante and Eakins, 2009). For ERA-Interim we used the altitude from the model topography, which is 556 m for the se-

lected gridpoint. (The map in Fig. 1 shows both the AMSU-B target area and the ERA-Interim grid.)

In summary, the different datasets natively have different lower integration altitude limits for IWV. To make the data comparable, we correct all measurements to a common reference altitude of 470 m. The random errors introduced by this procedure should be below approximately $0.2\,\text{kg}\,\text{m}^{-2}$ absolute or 2 % relative error.

This method of correction should be generally applicable also to other IWV data, and may be of interest for other IWV comparison studies. However, the caveat is that the actual scaling factors may depend on location (and possibly even on season). For example, Bock et al. (2007) find a value of -4.0 % per 100 m for Africa, a five percent stronger correction than in our case. It is also not obvious for what range of altitude differences our simple correction scheme would work, and also this may be location dependant. For Kiruna our results in Fig. 4 show that the correction works at least for altitude differences up to 250 m.

3.5 Representativeness error and horizontal inhomogeneity

As introduced in Sect. 3.2, two collocated datasets X and Y can be characterized by the mean and standard deviation of their difference. The standard deviation σ_{XY} contains contributions from the random errors of the two individual datasets (σ_X and σ_Y) and a third term, which we call error of representativeness (σ_{RepXY}). If these three errors were uncorrelated, their variances (squares of the standard deviations) would add up according to the error propagation law:

$$\sigma_{XY}^2 = \sigma_X^2 + \sigma_Y^2 + \sigma_{RepXY}^2 \tag{4}$$

If the errors are correlated to some degree, then Eq. (4) has to be modified to contain additional negative terms that account for the correlation. But we assume here that the three errors are uncorrelated, which is probably a good assumption, though very hard to prove.

The terms σ_X and σ_Y represent the random error or noise in the two datasets. For completely different measurement techniques, such as GPS and radiosonde, it is reasonable to assume that their errors are uncorrelated. But this will be violated if both measurements are affected by the same external phenomenon, for example by the presence or absence of clouds.

The representativeness error σ_{RepXY} represents two different problems that make dataset X not perfectly representative for dataset Y: firstly, the measurements may not be perfectly collocated in space and time, and secondly, they may have different sampling characteristics – one may be an instantaneous point measurement and the other an average over some distance in space and time.

Note, that our concept of representativeness error is slightly different from how it is presented for example by O'Carroll et al. (2008). In their three way error analysis, they

Table 3. Temporal and horizontal sampling characteristics of the different datasets. For GPS, FTIR, and the microwave, the temporal sampling simply is the instrument integration time. For the radiosonde it should be noted that, although the sensor gives an instantaneous point measurement of humidity, the IWV calculated from it will be along the flight path of the balloon (both temporally and spatially), so it will not necessarily correspond to the actual instantaneous IWV over some horizontal position. For AMSU-B, the spatial scale comes from our data processing, which is described in Sect. 2.5, not from the native resolution of the instrument, and the time given is the approximate time between the first and last AMSU pixel used for the collocation. For ERA-Interim, the given numbers are the grid resolution. The last column gives the NICAM-based estimate of the error of representativeness $\sigma_{\mathrm{RepXGPS}}$ for the comparison of the respective dataset to the GPS data. For the radiosonde our method is too coarse to provide an estimate, but there still will be a representativeness error.

Dataset	Temporal sampling	Horiz. dist. from GPS	Horizontal area	NICAM $\sigma_{\mathrm{RepXGPS}}$
GPS	2 h	0 km	circle of several km radius	0 per. def.
Radiosonde	2 s	<10 km	irregular column	unknown
FTIR	5–15 min	28 km	slant column	0.65–1.13 kg m^{-2}
Microwave	1–3 h	28 km	slant column	0.65–1.13 kg m^{-2}
AMSU-B	15 s	31 km	circle of 50 km radius	0.66–1.05 kg m^{-2}
ERA-Interim	30 min	41 km	0.75°×0.75° lat/lon	0.82–1.37 kg m^{-2}

assume that each dataset has a representativeness error relative to a common "truth". In contrast to this, in our analysis we define the representativeness error of all datasets relative to the GPS dataset, so the GPS dataset itself has $\sigma_{\mathrm{RepXY}} = 0$ by definition.

Representativeness errors for different datasets may be highly correlated, if they have similar sampling characteristics. For example, if the truth is a point value, then all areal estimates will have highly correlated representativeness errors. The presence of significant representativeness errors for our study leads to the consequence that we cannot carry out a formal multi way error analysis in order to separate the errors of the individual datasets. Instead, we simply present the results relative to the GPS dataset. However, at least we make an attempt to estimate the representativeness errors relative to the GPS, as described below.

We can distinguish three "dimensions" for the sampling characteristics, temporal, vertical, and horizontal. As discussed in Sect. 3.4, vertical sampling is different for the different datasets, but can be corrected to the vertical sampling of the GPS instrument with a rather small random error. We therefore ignore the vertical dimension here. For the temporal and horizontal dimensions the sampling characteristics of the different datasets are summarized in Table 3, which is based on the discussion in Sect. 3.3.

The representativeness error resulting from these sampling characteristics depends on the "length" scale of fluctuations in the given dimension. For the temporal dimension, fluctuations can be due to local processes, particularly convection, and due to advection. Convective activity is usually low in the comparison area, so advection is the more relevant process. As discussed in Sect. 3.3, the ±1 h time window for the matches is chosen such that it corresponds to a spatial scale of only a few kilometers for typical wind speeds. From this we conclude that the largest and most interesting contribution

to the error of representativeness comes from the horizontal sampling characteristics.

We used data from the NICAM model to try to characterize the representativeness error due to the different horizontal sampling. The Nonhydrostatic Icosahedral Atmospheric Model NICAM (Satoh et al., 2008) is a global circulation model with very high resolution. The model run we used has 3.5 km horizontal resolution and 40 vertical levels on a fixed altitude grid. The experiment was started at 00:00 UTC on 15 June 2008, and integrated for 10 days. The details of the analysis of this experiment are given by Nasuno et al. (2012) and Hashino et al. (2012).

IWV was calculated by vertical integration of the model variable specific humidity, starting at 430 m. (We always start at the same altitude, not at the surface, in order to suppress the impact of a varying lower integration limit, making the model comparable to our datasets that are all corrected to the same integration altitude limit. The value of 430 m is the altitude of the closest model level below the GPS altitude of 470 m.)

This gives us a global field of IWV with 3.5 km horizontal resolution. For the analysis below we used a single such field (one snapshot in time), the latitude range 65° to 70° N, and all longitudes. The date of the snapshot is 19 June 2008, close to the summer solstice on the Northern Hemisphere.

Let us explain the estimation procedure with the dataset combination GPS (at Esrange) and FTIR (at Space Campus). The horizontal distance between the two measurements is 28 km (see Table 3). If we had a long time series of NICAM data for both locations we could simply calculate the standard deviation of their difference to estimate the error of representativeness. But for technical reasons we had only a single model field available (the fields are very large, and therefore not easy to store and process).

Instead, we created synthetic NICAM datasets for "Esrange" and "Space Campus" by first selecting random model

points (within the 65–70° N latitude band) for the "Esrange" dataset. For each "Esrange" point we randomly selected a corresponding "Space Campus" data point that is 28 km away in an arbitrary direction. The standard deviation of the difference between the two is our estimate of the error of representativeness and is reported in Table 3.

In order to minimize the impact of topography and surface type variations, we introduced a constraint on the surface altitude (from the model topography). Case A is that it should be below 400 m, case B that it should be zero (ocean areas). We argue that these two cases can be roughly regarded as upper and lower limits of the real representativeness error. The real datasets are influenced by the difference in altitude and by the topography in the surrounding area, but the topography does not change with time. In NICAM case A, all topographic effects enter the random error, whereas in case B no topographic effects enter at all, so both are extreme cases.

The error of representativeness for the other datasets (against the reference GPS) were estimated in a similar fashion. For AMSU, the NICAM "measurement" was offset by 31 km and averaged over a 50 km radius circle. For ERA-Interim it was offset by 41 km and averaged over a $0.75° \times 0.75°$ latitude-longitude area (an area of 83×31 km^2).

Results are given in Table 3. No estimate is given for the GPS-radiosonde pair, since the model resolution is not fine enough to capture the small scale fluctuations that are expected to dominate in that case.

To validate the NICAM estimates, we made an independent test with the AMSU-B dataset. Figure 5 shows in blue the variance of paired data as a function of the distance from Kiruna along the satellite track for the year 2008 and the satellite N17. We assume that the variance is the sum of the pure instrumental error and a representativeness error that can be approximated by a linear function of distance (green fit line). This yields an estimate of the representativeness error of

$$\sigma_{\text{RepD}} = \sqrt{C \cdot D} \qquad (5)$$

where C is a fit constant and D is the along-track distance. For different satellites and years, C varies between 0.010 and 0.018 kg m^{-2} km^{-1}, with a mean value of 0.013 kg m^{-2} km^{-1}. For the Esrange to Space Campus distance, which is relevant for the FTIR-GPS and Microwave-GPS dataset combinations, this yields a representativeness error estimate of 0.60 kg m^{-2}, which is indeed close to the lower NICAM estimate of 0.65 kg m^{-2}. This can be considered as a good consistency, since the AMSU method is likely to somewhat underestimate the true representativeness error, because the data does not contain cases with high IWV.

To summarize, we made rough estimates of the representativeness error between the GPS and the other instruments from NICAM model data. The method used a lot of assumptions, the most important ones being that the error depends on distance and area, but not on direction, and that it depends
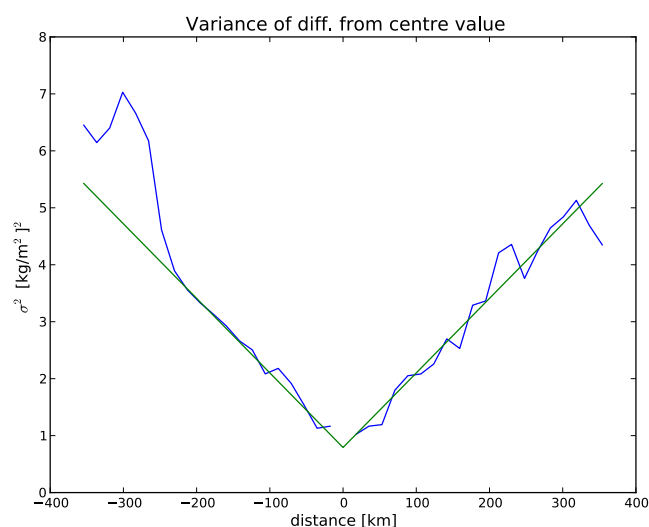

Fig. 5. Variance of paired AMSU-B IWV data as a function of distance along the satellite track for the year 2008 and the satellite N17 (blue) and linear fit (green). Kiruna is at distance 0. The green fit line has the parameters $\sigma^2 = 0.0131|d| + 0.79$.

neither on location nor time (that the variability in the atmosphere is isotropic, homogeneous, and stationary). None of these assumptions are strictly valid, so a rough estimate is all that the method can provide. However, we did validate the NICAM-based method with AMSU data for one case (horizontal distance) and found encouraging consistency.

4 Results

Figures 6 to 8 show the results of the comparison of all other datasets to the GPS measurements. Because of the short interval between GPS measurements, the majority of all measurements from the other instruments will result in a match. The exception are measurements after 2006 when our processed time series for the GPS has ended.

4.1 Radiosondes

Even though there is only a limited number of radiosonde launches, there is a sufficiently high number of matches with the GPS to derive statistics. As seen in the top left plot in Fig. 6, the radiosonde measurements show very good agreement with the GPS measurements. Radiosondes are very slightly wetter at high IWV and drier at low IWV. The slope of the linear regression line is 1.036, radiosondes are on average 0.35 kg m^{-2} wetter at 20 kg m^{-2} IWV and 0.33 kg m^{-2} drier at 1 kg m^{-2}. Random errors (measured by the standard deviation of the difference) are also small, only 0.66 kg m^{-2} (see Table 4). This is the smallest random error of all considered measurement pairs. This is consistent with the fact that the measurement locations for this pair are closer together
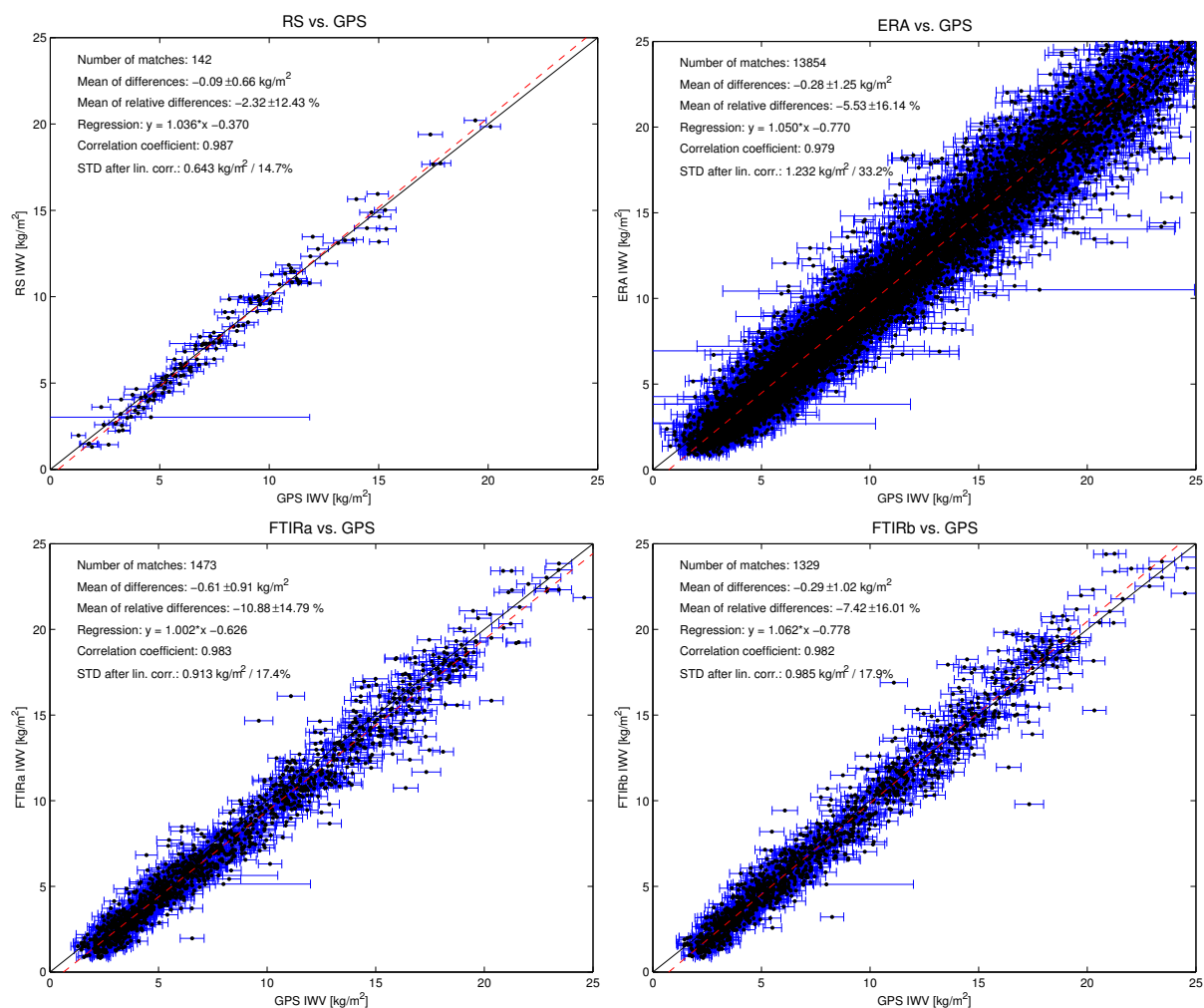
Fig. 6. Scatterplots of other datasets versus GPS: Radiosonde (top left), ERA-Interim (top right), FTIRa (bottom left), and FTIRb (bottom right).

than for other pairs, so we expect a relatively small representativeness error.

4.2 ERA-Interim

The ERA-Interim to GPS comparison is shown in the top right plot of Fig. 6. There is good overall agreement with random errors of approximately $1.25\,\mathrm{kg\,m^{-2}}$. ERA-Interim is drier than the GPS at low IWP values, and slightly moister at high IWV values (above $15\,\mathrm{kg\,m^{-2}}$).

4.3 FTIR

The FTIR measurements span the whole period of the GPS from 1996 to 2006 and therefore also give a large number of matches as seen in the two bottom plots of Fig. 6. As expected, both sets from the two different detectors show a similar amount of matches. Both sets also show very good agreement with the GPS measurements with a standard de-

viation of around $1\,\mathrm{kg\,m^{-2}}$ and a correlation coefficient just above 0.98.

However, FTIRa is consistently drier than the GPS, whereas FTIRb is drier only at low IWV, but moister at high IWV, roughly similar to the radiosonde data. The reason for the differences between the FTIRa and FTIRb datasets likely is in the spectroscopic data, because they use different spectral ranges and therefore different spectral lines (see Sect. 2.3).

Sussmann et al. (2009) have proposed to correct FTIR IWV data with radiosonde data for deriving climate data records. Despite the differences between the two detectors, we argue here that it is better to use the FTIR data as they are, since we do not have proof that the radiosonde data have a lower absolute error.

Table 4. Statistics of all comparisons. *Matches* is the number of matches between the two datasets. *Mean diff.* is the mean value of the differences plus/minus its standard deviation. *Mean rel. diff.* is the same for the relative difference. *Slope* and *Offset* are the slope and offset of the linear regression line, and *Corr.* is the linear (Pearson) correlation coefficient.

Pair	Matches	Mean diff. [kg m^{-2}]	Mean rel. diff.	Slope [kg m^{-2}]	Offset	Corr.
Reference: GPS						
RS-GPS	142	-0.09 ± 0.66	-2.32 ± 12.43	1.04	-0.37	0.987
FTIRa-GPS	1473	-0.61 ± 0.91	-10.88 ± 14.79	1.00	-0.63	0.983
FTIRb-GPS	1329	-0.29 ± 1.02	-7.42 ± 16.01	1.06	-0.78	0.982
MW0205-GPS	640	0.45 ± 2.66	-0.23 ± 27.00	1.21	-1.46	0.931
MW0508-GPS	1385	0.93 ± 1.75	12.93 ± 28.55	1.19	-0.30	0.867
MW0205clear-GPS	62	-0.25 ± 1.17	-4.45 ± 16.47	0.98	-0.10	0.979
MW0508clear-GPS	63	-0.02 ± 0.86	-4.17 ± 18.49	1.12	-0.71	0.967
AMSU-N16−GPS	2169	-0.19 ± 1.05	-2.27 ± 19.58	0.99	-0.15	0.842
AMSU-N17-GPS	1675	-0.06 ± 0.97	-1.62 ± 18.98	1.12	-0.71	0.864
ERA-GPS	13 854	-0.28 ± 1.25	-5.53 ± 16.14	1.05	-0.77	0.979
Reference: FTIRa						
FTIRb-FTIRa	1463	0.30 ± 0.52	3.36 ± 7.28	1.06	-0.11	0.996
MW0205-FTIRa	64	0.30 ± 0.97	3.22 ± 14.19	1.02	0.11	0.985
MW0508-FTIRa	104	0.90 ± 1.08	22.79 ± 29.34	1.09	0.48	0.933
AMSU-N16-FTIRa	171	0.52 ± 0.47	18.43 ± 15.94	1.10	0.19	0.954
AMSU-N17-FTIRa	170	0.55 ± 0.54	19.03 ± 17.48	1.12	0.15	0.950
ERA-FTIRa	597	0.48 ± 1.16	7.60 ± 17.10	1.07	-0.00	0.977
Reference: FTIRb						
MW0205-FTIRb	54	-0.20 ± 0.92	-1.90 ± 12.85	0.96	0.15	0.986
MW0508-FTIRb	89	0.77 ± 0.89	18.33 ± 18.15	1.14	0.15	0.955
AMSU-N16-FTIRb	151	0.50 ± 0.48	18.20 ± 16.87	1.08	0.22	0.953
AMSU-N17-FTIRb	151	0.45 ± 0.56	16.34 ± 17.78	1.07	0.22	0.947
ERA-FTIRb	538	0.20 ± 1.22	4.48 ± 17.24	1.01	0.13	0.974

4.4 Microwave

Of all the comparisons done in this study, the two KIMRA microwave time series (top plots in Fig. 7) show the least agreement with the GPS data. The earlier series of 2002–2005 shows fewer matches than the later series of 2005–2008, but both still have a sufficient amount of matches to do a statistically meaningful comparison. They both show a large wet bias at higher IWV values and also a large random error.

We suspected that the poor agreement is partly due to clouds (and even precipitation) affecting the microwave data. To test this, we sub-selected only those microwave data that have coincident FTIR data. These can be expected to be cloud-free, since the FTIR instrument does only measure under cloud-free conditions. Indeed, this cloud filtering vastly improves the microwave results (Fig. 7, bottom plots).

Our experiment is not strict proof that it is only the clouds that affect the KIMRA measurements. Temperature uncertainties may also play a role. The KIMRA data analysis uses analysis data from ECMWF to estimate the tropospheric temperature profile, and the error implied by that may be correlated with cloudiness, since cloud free conditions are likely to be also meteorologically more stable. We have made no efforts to disentangle these different effects, but suspect that the direct cloud radiative effect is the dominant cause of error.

In summary, we preliminarily conclude that the KIMRA microwave data would be useful for IWV studies, but only if a cloud-flag would be added to it, or would be available from some other source. (Estimating cloudiness from the microwave spectra themselves is not possible for this instrument.) It should be kept in mind here that KIMRA is a stratospheric ozone radiometer, not a dedicated tropospheric water vapour and cloud liquid radiometer, as discussed in Sect. 2.4.

4.5 AMSU-B

The AMSU-B measurements from NOAA-16 and NOAA-17 both have a high number of matches with the GPS data, thus giving a good basis for comparison. As seen in Fig. 8, both instruments show good overall agreement with the GPS and both have a standard deviation around 1 kg m^{-2}. NOAA-16 shows a slight dry bias towards high values whereas
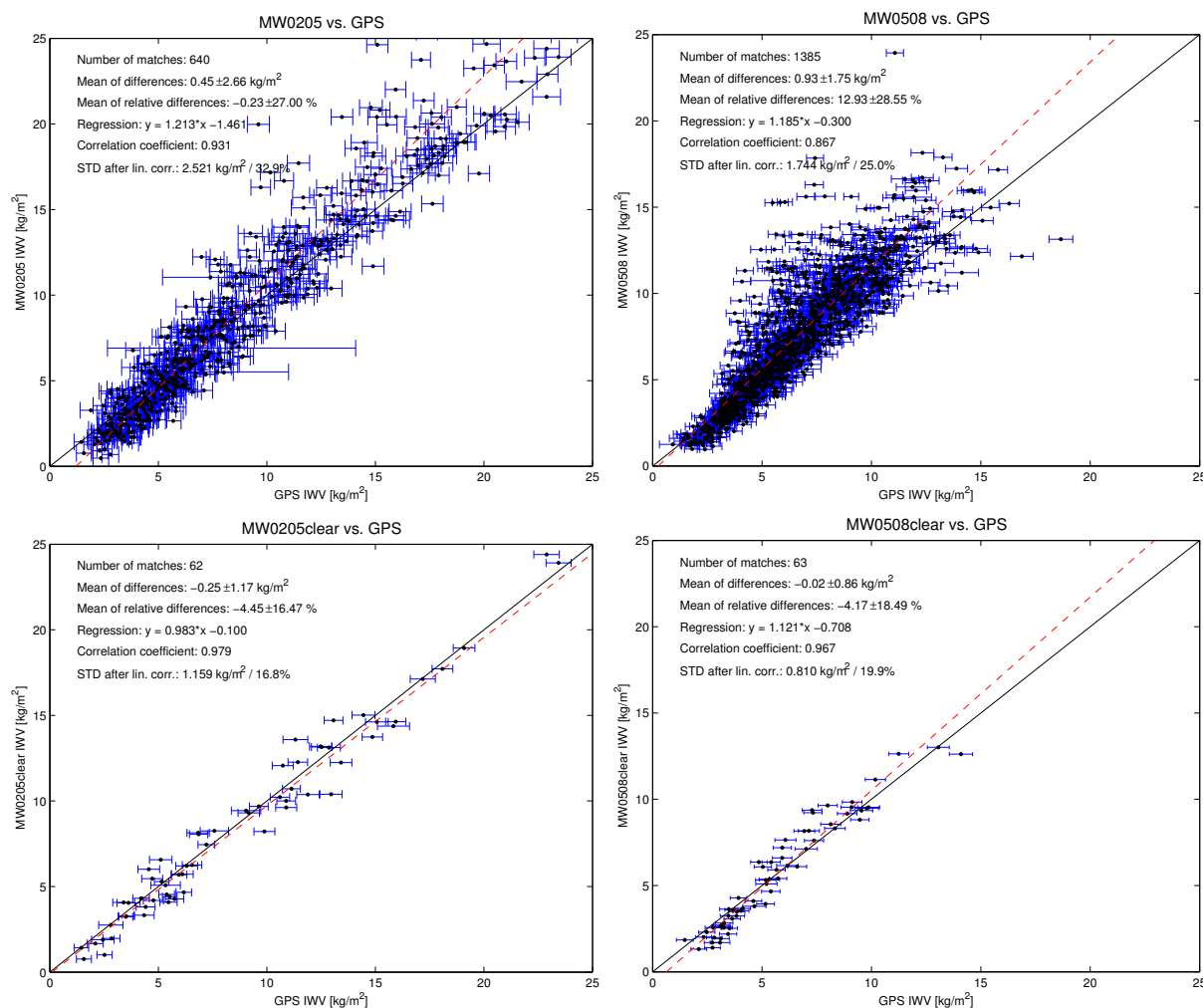
Fig. 7. Scatterplots of other datasets versus GPS: Microwave 2002–2005 (top left), Microwave 2005–2008 (top right), Clear-sky Microwave 2002–2005 (bottom left), and Clear-sky Microwave 2005–2008 (bottom right).

NOAA-17 instead shows a stronger wet bias. Both instruments show a group of outliers with high GPS IWV of 10–15 kg m^{-2} where the AMSU-B IWV is only between 5–10 kg m^{-2}. A possible reason for these can be the upper limit in the algorithm by Melsheimer and Heygster (2008) of about 8 kg m^{-2}. A passing weather system would then result in a large variation of IWV values in the AMSU-B pixels causing the algorithm to only retrieve the lower values found. The large error bars of the outliers also support the possibility of unstable weather conditions at the time of the measurement.

5 Discussion

5.1 Dataset summary

A complete list of the statistical parameters of the comparisons can be found in Table 4. We give the parameters not only for the reference GPS, but also for taking each of the

two FTIR datasets as reference, in order to facilitate comparisons with literature values (see Sect. 5.2). For that purpose it would also have been good to show data with the radiosonde as reference, but the number of matches obtained is then too low to derive valid statistics for all other datasets except the GPS.

A more graphical summary of these data is given in Fig. 9. It shows an overview of the systematic and random differences for all datasets, relative to the GPS. Its left plot shows the difference of all the linear fit lines from the diagonal, and its right plot the observed random errors, together with the NICAM-based estimate of the representativeness error (from Sect. 3.5/Table 3).

The linear fit summary shows that all other datasets are drier than the GPS at low IWV (below 10 kg m^{-2}). This makes us suspect that the Esrange GPS instrument indeed has a moist bias at low IWV values. For high IWV the situation is less clear. Above 15 kg m^{-2}, one of the microwave
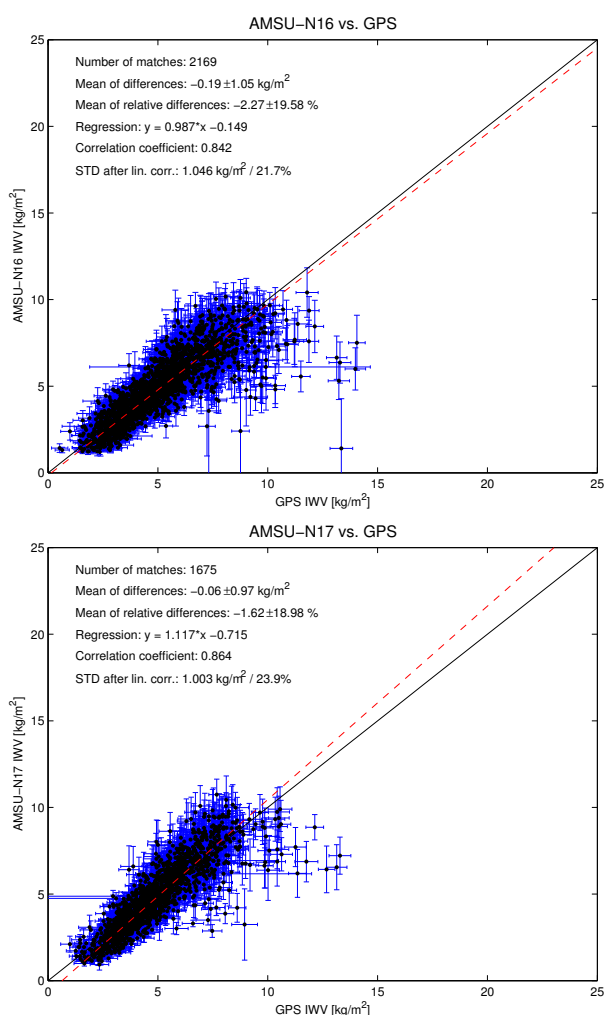
Fig. 8. Scatterplots of other datasets versus GPS: N16 AMSU-B (top) and N17 AMSU-B (bottom).

datasets and FTIRa are drier than the GPS, but radiosondes, ERA-Interim, FTIRb, and the other microwave dataset are moister.

The random errors (right plot in Fig. 9) are all modest (between 0.66 and 1.25 kg m^{-2}, if the cloudy microwave data are neglected). It is likely that a large part of these random errors is due to the error of representativeness. Our estimates of that error are also indicated in the figure.

For FTIR and Microwave the representativeness error should be very similar, since these two instruments are in the same location and have a similar measurement geometry. Also, the microwave data are here sub-selected to have matching FTIR data, in order to remove clouds, so even the atmospheric situations should be similar. The higher random error for the microwave thus is likely to really be due to the measurement itself.

The other dataset with a larger random error (relative to the representativeness error estimate) is ERA-Interim. As ex-

pected, the reanalysis is less precise than a local measurement. However, the difference is not large. Furthermore, the ERA-Interim to GPS comparison is really all-sky, and has more data at very high IWV than any other dataset pair, which most likely also contributes to the larger random errors.

5.2 Comparison to other published results

Table 5 gives a summary of other IWV dataset intercomparison studies that we have found in the literature. It is only a subset of all published studies, where we concentrated on articles that include at least two of our datasets, so that we can compare their results to ours. We also concentrated on large journals, like JGR, ACP, and AMT, in order to keep the number of articles manageable.

The most "popular" datasets appear to be GPS, radiosondes, and sun photometers (the latter unfortunately not included in our study). Particularly interesting for us are the studies by Palm et al. (2010) and Schneider et al. (2010), because they also include an FTIR instrument. Of these two, the Palm et al. study is most similar to our study, both in the location and in the combination of datasets. Below, we discuss the results for different dataset combinations.

5.2.1 Other GPS to radiosonde comparisons

Table 6 shows a summary of the results of other GPS to radiosonde (RS) comparisons. Here we use the radiosonde as the reference, since most other studies have done so. We focus mostly on the linear regression results. The conversion from

$$\text{IWV}_{\text{RS}} = a \cdot \text{IWV}_{\text{GPS}} + b \tag{6}$$

to

$$\text{IWV}_{\text{GPS}} = a' \cdot \text{IWV}_{\text{RS}} + b' \tag{7}$$

is

$$a' = 1/a \qquad \text{and} \qquad b' = -b/a \, . \tag{8}$$

Besides the above reference frame conversion, some results also required IWV unit conversions. We assumed that (Palm et al., 2010):

$$1 \times 10^{21} \, \text{molec cm}^{-2} = 0.3 \, \text{kg m}^{-2} \tag{9}$$

and that

$$\text{IWV[kg m}^{-2}] = \text{ZWD[mm]}/6.5 \tag{10}$$

where ZWD is the zenith wet delay, a humidity unit common in the GPS community. The exact conversion between ZWD and IWV is temperature dependent, with the conversion factor varying between approximately 6.1 and 6.9 (Ning, 2012, Fig. 2.4).

Table 5. An overview of other published IWV comparison studies that include at least two of our datasets. Note that, with the exception of the Palm et al. study, MW for the other studies typically implies a dual channel microwave radiometer designed for IWV and LWP measurements, not an ozone radiometer as in our study.

Study	Location	GPS	RS	FTIR	MW	AMSU	ERA	Sun phot.	VLBI	MODIS	GOME	Sciamachy
This study	Subarctic	X	X	X	X	X	X					
Ning et al. (2012)	Midlatitude[6]	X	X		X		X		X			
Palm et al. (2010)	Arctic	X	X	X	X[9]	X					X	X
Schneider et al. (2010)	Subtropical	X	X	X				X[1]				
Bock et al. (2007)	Tropical[7]	X	X			(X)[2]	X[3]	X				
Sapucci et al. (2007)	Tropical	X	X			(X)[5]		X				
Martin et al. (2006)	Midlatitude	X	X[8]		X[1]							
Bokoye et al. (2003)	Subarctic	X	X				(X)[4]	X				
Li et al. (2003)	Midlatitude	X	X							X		
Niell et al. (2001)	Midlatitude	X	X		X				X			
Ohtani and Naito (2000)	Midlatitude	X	X									
Sierk et al. (1997)	Midlatitude	X			X			X				

[1] multiple kinds, [2] SSM/I, [3] uses ERA-40 and NCEP2, [4] uses GEM data, [5] HSB, [6] also subarctic, [7] also subtropical, [8] different humidity sensor (SRS 400), [9] stratospheric trace gas radiometer, as in our study.
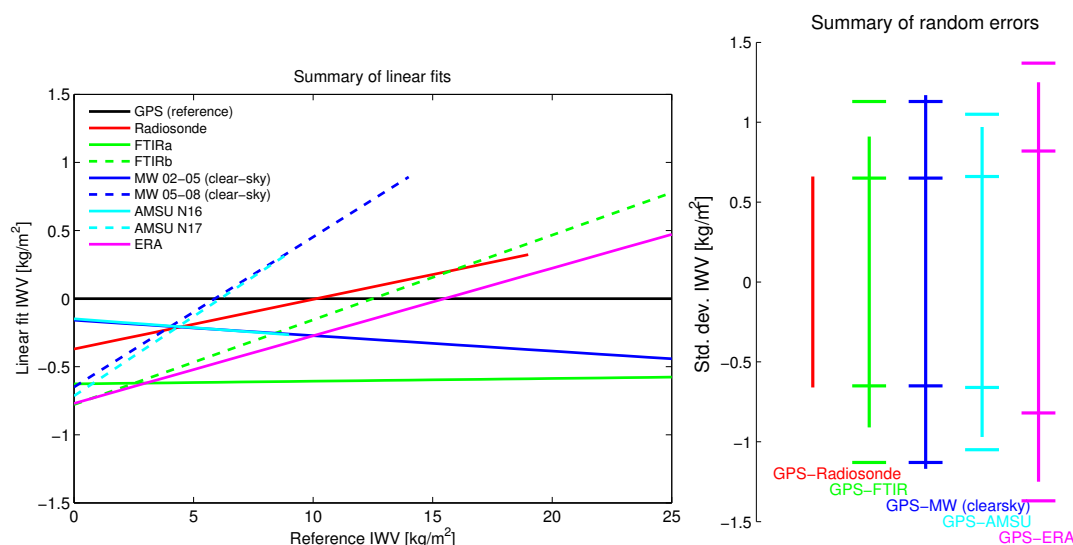

Fig. 9. Systematic and random differences for all datasets relative to the GPS dataset. Left plot: difference of linear regression lines (same data as in Table 4) from the diagonal. Lines are only plotted over the IWV range where we have data. Right plot: standard deviation of the difference to the GPS (vertical bars, same data as in Table 4) and minimum and maximum NICAM estimate of the representativeness error (horizontal lines, same data as in Table 3). For datasets where we have several versions, we somewhat arbitrarily selected one set to represent the technique for the random error plot, namely FTIRa for FTIR, MW 02-05 for microwave, and N17 AMSU-B for AMSU-B.

What can we learn if we examine all studies together? First of all, we note that the systematic differences between GPS and RS vary strongly from study to study. They seem to be particularly large and variable for the two tropical studies (Bock et al., 2007; Sapucci et al., 2007). We speculate that this may be related to poor radiosonde quality. If we disregard the two tropical studies, then the regression slope

still ranges from 0.94 to 1.07 and the offset from -1.5 to $+0.5$ kg m^{-2}. This means that for mid and high latitude conditions the bias between the two techniques should be below $1-2$ kg m^{-2} for any IWV value.

Several factors are likely to affect these observed biases. For the GPS, the exact antenna characteristics play a role, including the characteristics of a possible radome covering

Table 6. GPS versus radiosonde (RS) result summary. Slope and offset of the linear fit $\mathrm{IWV_{GPS}} = \mathrm{slope} \cdot \mathrm{IWV_{RS}} + \mathrm{offset}$ are given where available, otherwise the mean difference (with relative value in parenthesis) is given. STD is the standard deviation of the absolute difference (with standard deviation of relative difference in parenthesis).

Study	Slope	Offset $\mathrm{kg\,m^{-2}}$	Mean diff. $\mathrm{kg\,m^{-2}}$	STD $\mathrm{kg\,m^{-2}}$
This study	0.96	+0.36	+0.09 (+2.32 %)	0.66 (12 %)
Ning et al. (2012)			−0.35 to +0.34	1.23
Palm et al. (2010)	0.94	+0.22		(10–50 %)
Schneider et al. (2010)			−0.66 (−13 %)	1.18 (31 %)
Bock et al. (2007)	0.76–1.14	−0.5 to +17.3		2.6–3.7 (7–16 %)
Sapucci et al. (2007)	0.68–0.95	+5.5 to +17.6		2.5
Martin et al. (2006)	1.01	−0.47		0.96
Bokoye et al. (2003)	1.07	+0.5		1.8
Li et al. (2003)	0.99–1.04	−1.1 to −0.5		0.7–1.2
Niell et al. (2001)	1.06	+0.16		1.5
Ohtani and Naito (2000)	0.95–0.99	approx. −1.5		1.6–3.8

the antenna, and the presence or absence of microwave absorber material around the antenna. (Ning et al. (2011) reported that the addition of absorber material decreased the GPS IWV bias by $1.3\,\mathrm{kg\,m^{-2}}$, and the addition of a radome made a difference of $0.4\,\mathrm{kg\,m^{-2}}$.) How strongly these hardware difference actually affect the IWV data also depends on details of the data analysis, particularly how slant wet delays are mapped to the zenith direction and what satellite elevation angle cutoffs are applied (Ning et al., 2011). For the radiosonde, important factors are the sensor type, launching procedures, and the local time, since some sensors suffer from radiation bias. Taken together, these factors on both the GPS and the radiosonde side can easily explain the observed biases in the datasets.

The last column in Table 6 shows the random error between GPS and RS. Our study is among those with the lowest random error, together with Palm et al. (2010). We speculate that this depends more on the way the studies are set up (temporal and spatial separation between the measurements) than on the measurements themselves. This follows from Sect. 3.5, where we showed that representativeness error plays a large role in these comparisons. In any case, the comparison seems to indicate that our random error value of $0.66\,\mathrm{kg\,m^{-2}}$ appears to be as low as one can get with reasonable effort.

5.2.2 Other results for FTIR

Table 7 shows literature values for the FTIR-GPS comparison from Palm et al. (2010) and Schneider et al. (2010). Our slope and offset is in the opposite direction of Palm et al. (2010), and also our mean difference is in the opposite direction of Schneider et al. (2010). But the discrepancies are moderate, in fact, the discrepancies to these published results are roughly of the same size as the discrepancy between our two datasets FTIRa and FTIRb. Overall, we conclude that

the systematic difference between FTIR and GPS is below $1–2\,\mathrm{kg\,m^{-2}}$ for all IWV values.

Regarding the random errors, ours are slightly higher than those reported by Schneider et al. (2010), and also likely to be slightly higher than those reported by Palm et al. (2010), judging indirectly from their random errors compared to the radiosonde data. We speculate that this is mostly due to our slightly higher representativeness error, due to the horizontal distance between our GPS and FTIR instruments.

The study by Palm et al. (2010) evaluated also a ground-based microwave instrument and the AMSU-B dataset. For the groundbased instrument they, like us, find a higher random error than for the FTIR. However, Palm et al. (2010) did not attribute the larger random error of the microwave to the presence of clouds. Their AMSU-B results also are broadly consistent with ours. In particular, their Fig. 3 shows that the AMSU-B data have outliers for high IWV values close to the maximum IWV range where the algorithm works. This is also clearly seen in our Fig. 8.

5.3 Suitability of instruments for climate research

Climate research requires datasets with high absolute accuracy and, most importantly, long-term stability. Our study demonstrates that none of the participating measurement datasets is automatically suitable for this application.

To start with the radiosondes, their historical data record suffers from frequent and often undocumented sensor changes. The ongoing GRUAN project (Immler et al., 2010) will provide climate-quality radiosonde data, but that does not solve the issue for historical data.

The GPS measurements should in principle be very suitable for climate research, because they can be traced back to frequency and time measurements. But other studies have shown that changes in antenna and algorithm details can

Table 7. Literature results for FTIR versus GPS. Parameters are defined as in Table 6.

Study	Slope	Offset $\mathrm{kg\,m^{-2}}$	Mean diff. $\mathrm{kg\,m^{-2}}$	STD $\mathrm{kg\,m^{-2}}$
This study FTIRa-GPS	1.00	−0.63	−0.61 (−10.88 %)	0.91 (14.79 %)
This study FTIRb-GPS	1.06	−0.78	−0.29 (−7.42 %)	1.02 (16.01 %)
Palm et al. (2010)	0.95	+0.40		
Schneider et al. (2010)			+0.09 (+5.4 %)	0.73 (20 %)

cause jumps even in these data (Ning et al., 2011; Ning and Elgered, 2012).

FTIR and microwave data are suitable if and only if the instruments are kept the same over long time periods. Besides changes in the instruments themselves, changes in the algorithms for the data processing are also problematic. The different alternative sets that we studied (FTIRa/FTIRb and the different microwave datasets) show that processing changes easily introduce systematic differences in the data (compare the two green curves in Fig. 9 for the FTIR data, or the two dark-blue curves for the microwave data. On the other hand, this issue can be solved by reprocessing the raw instrument data at a later time. It is therefore crucial that raw instrument data and auxiliary data are carefully archived.

Lastly, for the AMSU-B data, systematic differences between satellites appear to be significant. They must therefore be corrected before deriving long time series. Such inter-satellite correction efforts are ongoing (John et al., 2012).

Observed biases for some instrument can not necessarily be generalized to other instruments of the same type. This follows directly from the observation that systematic differences between subsets of the same technique (e.g. FTIRa/FTIRb) may exceed differences between different techniques (as demonstrated by Fig. 9). It is also confirmed by the literature survey in the previous section.

6 Summary and conclusions

We have compared six different datasets of IWV for a high-latitude location (near the town Kiruna in Sweden). Five of the datasets are measurement-based, and one is a reanalysis dataset. An overall summary of the comparison is given in Fig. 9. All datasets give a reasonable estimate of IWV. All observed systematic differences are below $\pm 1\,\mathrm{kg\,m^{-2}}$ (the IWV value at this location typically is between 1 and $30\,\mathrm{kg\,m^{-2}}$). The reanalysis does not stick out much compared to the measurements, it has comparable systematic differences to the GPS (which was our reference) but slightly higher random error than all the measurements.

In the case of the KIMRA groundbased microwave data there is the caveat that cloudy data have to be discarded, because KIMRA cannot distinguish the signal of water vapour from cloud liquid. This caveat should not apply to more conventional tropospheric water vapour radiometers.

While the overall fair agreement between the different datasets is encouraging, our attempt to characterise the systematic differences in terms of slope and offset (with the implied aim to correct for them later) has turned out rather discouraging: The slope and offset values seem to depend strongly on the individual instrument (not just the instrument type) and the location. Our literature survey in Sect. 5.2 reveals that the literature is full of reported slopes and offsets for different instrument combinations, but no consistent overall picture emerges, other than that systematic differences typically are below 1–$2\,\mathrm{kg\,m^{-2}}$. It is therefore not obvious how IWV measurement accuracy could be further improved. It is also not obvious which technique is most suitable for recording climate data records from a scientific point of view. So probably practical considerations, such as maintenance costs, will play a large role for this question.

In the article we address some general issues, that should be relevant for any such comparison. These are the lower altitude limit of the IWV integration and the presence of representativeness error. For the former, we find that the lower altitude limit of the IWV integration is important and will introduce a bias if not corrected. (This may account for some of the scatter in the literature values for instrument comparisons.) We developed a new method to correct for the different lower altitude limit of different datasets. The method uses a correction factor that depends linearly on altitude and that is determined from radiosonde data.

For the issue of representativeness error, we attempt to estimate it based on model data, and validate the estimate with AMSU-B data. The method is very rough, but still puts the observed random differences between different datasets into perspective. We find that usually the representativeness error dominates the random error between datasets, so these observed random errors do not say much about the true precision of the involved techniques.

Both of these methods (for altitude limit correction and representativeness error estimate) are generally applicable, but should then be adjusted to the local atmospheric climatology.

References

Amante, C. and Eakins, B.: ETOPO1 1 arc-minute global relief model: procedures, data sources and analysis, Tech. rep., NOAA NESDIS NGDC, 2009.

Blumenstock, T., Kopp, G., Hase, F., Hochschild, G., Mikuteit, S., Raffalski, U., and Ruhnke, R.: Observation of unusual chlorine activation by ground-based infrared and microwave spectroscopy in the late Arctic winter 2000/01, Atmos. Chem. Phys., 6, 897–905, doi:10.5194/acp-6-897-2006, 2006.

Bock, O., Bouin, M.-N., Walpersdorf, A., Lafore, J. P., Janicot, S., Guichard, F., and Agusti-Panareda, A.: Comparison of ground-based GPS precipitable water vapour to independent observations and NWP model reanalyses over Africa, Q. J. R. Meteorol. Soc., 133, 2011–2027, doi:10.1002/qj.185, 2007.

Bokoye, A. I., Royer, A., O'Neill, N. T., Cliche, P., McArthur, L. J. B., Teillet, P. M., Fedosejevs, G., and Thériault, J.-M.: Multisensor analysis of integrated atmospheric water vapor over Canada and Alaska, J. Geophys. Res., 108, 4480, doi:10.1029/2002JD002721, 2003.

Buehler, S. A., Kuvatov, M., John, V. O., Leiterer, U., and Dier, H.: Comparison of Microwave Satellite Humidity Data and Radiosonde Profiles: A Case Study, J. Geophys. Res., 109, D13103, doi:10.1029/2004JD004605, 2004.

Buehler, S. A., Kuvatov, M., John, V. O., Milz, M., Soden, B. J., Jackson, D. L., and Notholt, J.: An Upper Tropospheric Humidity Data Set From Operational Satellite Microwave Data, J. Geophys. Res., 113, D14110, doi:10.1029/2007JD009314, 2008.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hoelm, E. V., Isaksen, L., Kaellberg, P., Koehler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thepaut, J.-N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Q. J. R. Meteorol. Soc., 137, 553–597, doi:10.1002/qj.828, 2011.

Hase, F., Hannigan, J. W., Coffey, M. T., Goldman, A., Höpfner, M., Jones, N. B., Rinsland, C. P., and Wood, S. W.: Intercomparison of retrieval codes used for the analysis of high-resolution, ground-based FTIR measurements, J. Quant. Spectrosc. Radiat. Transfer, 87, 25–52, doi:10.1016/j.jqsrt.2003.12.008, 2004.

Hashino, T., Satoh, M., Hagihara, Y., Kubota, T., Matsui, T., Nasuno, T., and Okamoto, H.: Evaluation of cloud distribution and bulk microphysics simulated by a global cloud-resolving model with combined use of CloudSat and CALIPSO measurements, J. Geophys. Res., in review, 2012.

Immler, F. J., Dykema, J., Gardiner, T., Whiteman, D. N., Thorne, P. W., and Vömel, H.: Reference Quality Upper-Air Measurements: guidance for developing GRUAN data products, Atmos. Meas. Tech., 3, 1217–1231, doi:10.5194/amt-3-1217-2010, 2010.

Jarlemark, P., Emardson, R., Johansson, J., and Elgered, G.: Ground-Based GPS for Validation of Climate Models: the Impact of Satellite Antenna Phase Center Variations, IEEE T. Geosci. Remote, 48, 5487374, doi:10.1109/TGRS.2010.2049114, 2010.

John, V. O. and Buehler, S. A.: Comparison of microwave satellite humidity data and radiosonde profiles: A survey of European stations, Atmos. Chem. Phys., 5, 1843–1853, doi:10.5194/acp-5-1843-2005, 2005

John, V. O., Holl, G., Buehler, S. A., Candy, B., Saunders, R. W., and Parker, D. E.: Understanding inter-satellite biases of microwave humidity sounders using global simultaneous nadir overpasses, J. Geophys. Res., 117, D02305, doi:10.1029/2011JD016349, 2012.

King, M. A. and Watson, C. S.: Long GPS coordinate time series: multipath and geometry effects, J. Geophys. Res., 115, B04403, doi:10.1029/2009JB006543, 2010.

Kottayil, A., Buehler, S. A., John, V. O., Miloshevich, L. M., Milz, M., and Holl, G.: On the importance of Vaisala RS92 radiosonde humidity corrections for a better agreement between measured and modeled satellite radiances, J. Atmos. Oceanic Technol., 29, 248–259, doi:10.1175/JTECH-D-11-00080.1, 2012.

Krystek, M. and Anton, M.: A weighted total least-squares algorithm for fitting a straight line, Meas. Sci. Technol., 18, 3484–3442, doi:10.1088/0957-0233/18/11/025, 2007.

Li, Z., Muller, J.-P., and Cross, P.: Comparison of precipitable water vapor derived from radiosonde, GPS, and Moderate-Resolution

Imaging Spectroradiometer measurements, J. Geophys. Res., 108, 4651, doi:10.1029/2003JD003372, 2003.

Martin, L., Mätzler, C., Hewison, T. J., and Ruffieux, D.: Intercomparison of integrated water vapour measurements, Met. Zeit., 15, 57–64, doi:10.1127/0941-2948/2006/0098, 2006.

Melsheimer, C. and Heygster, G.: Improved Retrieval of Total Water Vapor Over Polar Regions From AMSU-B Microwave Radiometer Data, IEEE T. Geosci. Remote, 46, 2307–2322, doi:10.1109/TGRS.2008.918013, 2008.

Moradi, I., Buehler, S. A., John, V. O., and Eliasson, S.: Comparing upper tropospheric humidity data from microwave satellite instruments and tropical radiosondes, J. Geophys. Res., 115, D24310, doi:10.1029/2010JD013962, 2010.

Nasuno, T., Yamada, H., Yanase, W., and Satoh, M.: Genesis of Typhoon Fengshen (2008) from vortex superposition. Part I: large-scale and synoptic-scale analysis, J. Atmos. Sci., in preparation, 2012.

Nilsson, T. and Elgered, G.: Long-term trends in the atmospheric water vapor content estimated from ground-based GPS data, J. Geophys. Res., 113, D19101, doi:10.1029/2008JD010110, 2008.

Niell, A. E., Coster, A. J., Solheim, F. S., Mendes, V. B., Toor, P. C., Langley, R. B., and Upham, C. A.: Comparison of measurements of atmospheric wet delay by radiosonde, water vapor radiometer, GPS, and VLBI, J. Atmos. Oceanic Technol., 80, 830–850, doi:10.1175/1520-0426(2001)018<0830:COMOAW>2.0.CO;2, 2001.

Ning, T.: GPS Meteorology: With Focus on Climate Applications, Ph.D. thesis, Chalmers University of Technology, 2012.

Ning, T. and Elgered, G.: Trends in the Atmospheric Water Vapor Content From Ground-Based GPS: The Impact of the Elevation Cutoff Angle, IEEE J. Sel. Top. Appl., 5, 744–751, doi:10.1109/JSTARS.2012.2191392, 2012.

Ning, T., Elgered, G., and Johansson, J.: The impact of microwave absorber and radome geometries on GNSS measurements of station coordinates and atmospheric water vapour, Adv. Space. Res., 47, 186–196, 2011.

Ning, T., Haas, R., Elgered, G., and Willén, U.: Multi-technique comparisons of 10 years of wet delay estimates on the west coast of Sweden, J. Geodesy, 86, 565–575, 2012.

O'Carroll, A. G., Eyre, J. R., and Saunders, R. W.: Three-Way Error Analysis between AATSR, AMSR-E, and In Situ Sea Surface Temperature Observations, J. Atmos. Oceanic Technol., 25, 1197–1207, doi:10.1175/2007JTECHO542.1, 2008.

Ohtani, R. and Naito, I.: Comparisons of GPS-derived precipitable water vapors with radiosonde observations in Japan, J. Geophys. Res., 105, 26917–26929, doi:10.1029/2000JD900362, 2000.

Östman, S.: A multi-instrument comparison study of integrated water vapor over Kiruna, Sweden, Master's thesis, Luleå University of Technology, 2010.

Palm, M., Melsheimer, C., Noël, S., Heise, S., Notholt, J., Burrows, J., and Schrems, O.: Integrated water vapor above Ny Ålesund, Spitsbergen: a multi-sensor intercomparison, Atmos. Chem. Phys., 10, 1215–1226, doi:10.5194/acp-10-1215-2010, 2010.

Raffalski, U., Berg, H., Hochschild, G., and Kopp, G.: Continuous ozone measurements over Kiruna during winter/spring 2002: A new millimeter wave radiometer operated at the Swedish Institute of Space Physics, Kiruna, Sweden, in: Stratospheric Ozone 2002, edited by Harris, N., Amanatidis, G., and Levine, J., 369–372, ISBN 92-894-5484-9, download from http://www-imk.fzk.de/asf/mira/Publicat/Dokus/Gotebg02_KirunaIRF.pdf, 2002.

Rose, T., Crewell, S., Löhnert, U., and Simmer, C.: A network suitable microwave radiometer for operational monitoring of the cloudy atmosphere, Atmos. Res., 75, 183–200, doi:10.1016/j.atmosres.2004.12.005, 2005.

Sapucci, L. F., Machado, L. A. T., Monico, J. F. G., and Plana-Fattori, A.: Intercomparison of integrated water vapor estimates from multisensors in the amazonian region, J. Atmos. Oceanic Technol., 24, 1880–1894, doi:10.1175/JTECH2090.1, 2007.

Satoh, M., Matsuno, T., Tomita, H., Miura, H., Nasuno, T., and Iga, S.: Nonhydrostatic icosahedral atmospheric model (NICAM) for global cloud resolving simulations, J. Chem. Phys., 227, 3486–3514, doi:10.1016/j.jcp.2007.02.006, 2008.

Saunders, R. W., Hewison, T. J., Stringer, S. J., and Atkinson, N. C.: The Radiometric Characterization of AMSU-B, IEEE T. Microw. Theory, 43, 760–771, 1995.

Schmid, R., Steigenberger, P., Gendt, G., Ge, M., and Rothacher, M.: Generation of consistent absolute phase-center correction model for GPS receiver and antennas, J. Geodesy, 12, 781–798, 2007.

Schneider, M. and Hase, F.: Ground-based FTIR water vapour profile analyses, Atmos. Meas. Tech., 2, 609–619, doi:10.5194/amt-2-609-2009, 2009.

Schneider, M., Romero, P. M., Hase, F., Blumenstock, T., Cuevas, E., and Ramos, R.: Continuous quality assessment of atmospheric water vapour measurement techniques: FTIR, Cimel, MFRSR, GPS, and Vaisala RS92, Atmos. Meas. Tech., 3, 323–338, doi:10.5194/amt-3-323-2010, 2010.

Sierk, B., Bürki, B., Becker-Ross, H., Neubert, R., Kruse, L. P., and Kahle, H.-G.: Tropospheric water vapor derived from solar spectrometer, radiometer, and GPS measurements, J. Geophys. Res., 102, 22411–22424, 1997.

Sonntag, D.: Advancements in the field of hygrometry, Met. Zeit., 3, 51–66, 1994.

Sussmann, R., Borsdorff, T., Rettinger, M., Camy-Peyret, C., Demoulin, P., Duchatelet, P., Mahieu, E., and Servais, C.: Technical Note: Harmonized retrieval of column-integrated atmospheric water vapor from the FTIR network – first examples for long-term records and station trends, Atmos. Chem. Phys., 9, 8987–8999, doi:10.5194/acp-9-8987-2009, 2009.

Westwater, E. R.: The accuracy of water vapor and cloud liquid determination by dual-frequency ground-based microwave radiometry, Radio Sci., 13, 677–685, doi:10.1029/RS013i004p00677, 1978.