Intercomparison of atmospheric water vapour measurements in the Canadian high Arctic

Dan Weaver[1], Kimberly Strong[1], Matthias Schneider[2], Penny M. Rowe[3,4], Chris Sioris[5], Kaley A. Walker[1,6], Zen Mariani[7], Taneil Uttal[8], C. Thomas McElroy[5], Holger Vömel[9], Alessio Spassiani[10], James R. Drummond[11]

[1]Department of Physics, University of Toronto, Toronto, Ontario, Canada

[2]Institute of Meteorology and Climate Research (IMK-ASF), Karlsruhe Institute of Technology, Karlsruhe, Germany

[3]NorthWest Research Associates, Redmond, Washington, USA

[4]Department of Physics, Universidad de Santiago de Chile, Santiago, Chile

[5]Department of Earth and Space Science and Engineering, York University, Toronto, Canada

[6]Department of Chemistry, University of Waterloo, Waterloo, Ontario, Canada

[7]Cloud Physics and Severe Weather Research Section, Environment and Climate Change Canada, Toronto, Ontario, Canada

[8]Earth Systems Research Laboratory, NOAA, Boulder, USA

[9]Earth Observing Laboratory, NCAR, Boulder, Colorado, USA

[10]School of Civil Engineering, University of Queensland, Brisbane, Australia

[11]Department of Physics and Atmospheric Science, Dalhousie University, Halifax, Nova Scotia, Canada

Correspondence to: Dan Weaver (dweaver@atmosp.physics.utoronto.ca)

Abstract. Water vapour is a critical component of the Earth system. Techniques to acquire and improve measurements of atmospheric water vapour and its isotopes are under active development. This work presents a detailed intercomparison of water vapour total column measurements taken between 2006 and 2014 at a Canadian high Arctic research site. Instruments include radiosondes, sun photometers, a microwave radiometer, and emission and solar absorption Fourier transform spectrometers (FTSs). Good agreement is observed between all combination of datasets, with correlation coefficients ≥ 0.90 showing high correlations. A variety of biases and calibration issues are revealed and discussed for all instruments.

A new FTS dataset, resulting from the MUSICA (Multi-platform remote Sensing of Isotopologues for investigating the Cycle of Atmospheric water) retrieval technique, is shown to offer accurate measurements of water vapour total columns; however, measurements show a small wet bias of approximately 6%. A new dataset derived from Atmospheric Emitted Radiance Interferometer (AERI) measurements is also shown to provide accurate water vapour measurements, which usefully enables measurements to be taken during day and night (especially valuable during Polar Night).

In addition, limited profile comparisons are conducted using radiosonde and ground-based FTS measurements. Results show MUSICA FTS profiles were within 15% of radiosonde measurements throughout the troposphere.

1   Introduction

Water vapour plays a significant role in the Earth's atmosphere. It is involved in driving atmospheric dynamics (Hwang and Frierson, 2010) and actively impacts atmospheric chemistry (Shindell, 2001). Water vapour has a dominant effect on climate and radiative forcing (Soden et al., 2002; Dessler et al., 2008). Climate change-induced shifts to the global hydrological cycle

5   affect atmospheric transport processes, creating and intensifying droughts and flooding (Trenberth et al., 2013). Understanding the global water cycle has critical value. Yet, our understanding of water vapour's abundances, variability, and transport are incomplete (Stevens and Bony, 2013). Atmospheric models are not able to accurately and precisely represent the water cycle. More observations are key to addressing this deficiency (Trenberth, 2014). Observations of the Arctic region are particularly sparse and important for understanding how the planet's atmosphere is changing (ACIA, 2005).

10   The Arctic region is disproportionately affected by climate change. Temperatures in the Arctic have increased more than elsewhere on the planet. At Eureka, Nunavut (80° N, 86° W), a research site in the Canadian high Arctic, surface temperatures increased by 0.88 ± 0.17 °C per decade between 1972 and 2007 (Lesins et al., 2010). This observed Arctic warming trend is expected to continue (IPCC, 2013). Alongside this warming, the total column of water vapour at Eureka increased by $10 \pm 3\%$ between 1961 and 2007 (Lesins et al., 2010). This aligns with the expectation that water vapour abundances will increase

15   globally as temperatures increase (Soden et al., 2002). In the high Arctic, the impact of this increase in atmospheric water vapour on the radiative balance (and thus climate) is particularly acute (Tobin et al., 1999).

Efforts are underway to improve and expand water vapour measurements. Space-based instruments with water vapour products, such as the Atmospheric Infrared Sounder (Aumann et al. 2003) and the Atmospheric Chemistry Experiment (ACE) (Sheese et al., 2016; Sioris et al., 2010), provide (almost) global measurement coverage. However, obtaining sensitivity to the

20   lower troposphere is challenging for satellites, and detailed studies of specific regions are temporally limited. Balloon-based and ground-based observation networks complement space-based observations. The Global Climate Observing System Reference Upper Air Network (GRUAN) enhances the scientific utility of high-vertical-resolution meteorological balloon measurements; however, GRUAN is geographically limited. Additionally, most measurement sites launch radiosondes only twice daily. The MUlti-platform remote Sensing of Isotopologues for investigating the Cycle of Atmospheric water (MUSICA)

25   project can contribute to this need for frequent high quality water vapour measurements. MUSICA uses existing ground-based Network for the Detection of Atmospheric Composition Change (NDACC) Fourier Transform InfraRed (FTIR) spectrometer observations to produce a precise and accurate measurement of water vapour isotopologues. Techniques for producing datasets from the high quality measurements taken by NDACC FTIRs have been under active development for decades (e.g., Kurylo, 1991). MUSICA uses NDACC spectra from participating sites to produce information about water vapour with reasonable

30   measurement sensitivity throughout the troposphere, as well as information about water vapour isotopologues in the lower and middle troposphere (Schneider et al., 2012; 2015). Multiple measurements are typically taken at FTIR measurement sites each

day. These new MUSICA observations of water vapour provide valuable opportunities for improving our understanding of the climate system and the water cycle.

This study intercompares measurement techniques used at a research facility in the Canadian high Arctic to support investigations into the critical role played by atmospheric water vapour. The primary goal is to assess the accuracy of a new Eureka MUSICA dataset and measurements by an emission FTIR. Intercomparisons are performed between Eureka MUSICA, emission FTIR, radiosonde, sun photometer, and microwave radiometer datasets.

1.1   Measurement site

Eureka, Nunavut is a small research community situated at 80°N in the remote polar desert of Ellesmere Island, Canada. It primarily exists to support Environment and Climate Change Canada's Eureka Weather Station (EWS), but also supports research programs led by universities, government agencies, and other organizations. The most significant of these is the Polar Environment Atmospheric Research Laboratory (PEARL), run by a group of Canadian universities through the Canadian Network for the Detection of Atmospheric Change (CANDAC). Until PEARL opened in 2006, the only information regularly gathered locally about atmospheric water vapour was obtained through the launch of radiosondes at the EWS. The suite of PEARL instruments has expanded the available information about the atmosphere above Eureka substantially. PEARL is strategically located for Arctic studies, as well as the validation of satellite measurements (e.g., Batchelor et al., 2010; Adams et al., 2012).

PEARL consists of multiple facilities. Instruments whose water vapour datasets are used for this study are located at the PEARL Ridge Lab (RL) and the zero-altitude PEARL Auxiliary Laboratory (0PAL). The Ridge Lab is located at 80.05° N, 86.4 °W on top of a ridge at 610 m elevation, 15 km west of Eureka. 0PAL is located in Eureka (79.59° N, 85.56° W) near sea level (10 m a.s.l.), approximately 250 m from the EWS radiosonde launch location. The Ridge Lab and 0PAL sites often experience different local weather conditions (Fogal et al., 2013), which should be considered when comparing measurements taken at the RL and 0PAL.

1.2   Eureka water vapour

Eureka is a challenging site for water vapour measurements. It is an extremely cold and dry environment. Between fall and spring, there are frequent temperature and humidity inversions in the lower troposphere. Open water occurs regionally during summer, but during the rest of the year the region's fjords and sounds are frozen. The surrounding geography is mountainous and variable. Solar-viewing measurements are often made at large solar zenith angles, especially during spring and fall. It is not possible to use atmospheric measurement techniques that require sunlight during Polar Night, which lasts from mid-

October until late-February at Eureka's latitude. These conditions, along with the availability of several instruments located at two different altitudes, offers opportunity to investigate the effectiveness of different measurement techniques.

The Eureka radiosonde dataset has informed weather research for over fifty years, and offers useful information about water vapour abundances and variability near Eureka. Figure 1 shows radiosonde water vapour profiles recorded between August 2006 and December 2015, along with overall and seasonal mean profiles (Eureka radiosonde measurements are described in more detail in Sect. 2.4). The definition of seasons is atypical, with a short spring and long autumn. This reflects the physical character of the annual changes at Eureka (Lesins et al., 2010). The water vapour mixing ratio profiles vary by an order of magnitude between winter and summer. The maximum mixing ratio of water vapour in the atmosphere above Eureka is at the surface during summer months. During the rest of the year, water vapour abundances reach their maximum 1-2 km above the surface.

When analysing the ability of different measurement techniques to capture information about water vapour, its vertical distribution should be considered. Figure 2 illustrates the portion of the total column typically found beneath a given altitude, using Eureka Weather Station radiosonde data from January 2007 to December 2015. 90% of the Eureka water vapour total column is found beneath an altitude of 4.40 km; 50% is found beneath an altitude of 1.60 km. Seasonal analysis gives similar results; however, there are differences in the vertical distribution of water vapour in the lowest few kilometres. The lowest altitudes contain more of the total column during the summer than during the winter. This may be due to the availability of local humidity sources during the summer, when there is open water in the neighbouring fjord. Ground-based instruments without sensitivity to the lowest altitudes may be seasonally biased, and underestimate total columns more in summer than in winter.

Detailed studies of the high Arctic water cycle require more frequent measurements than radiosondes can provide, observations at all altitudes, and precise observations of water vapour isotopes to reveal information about the transport history of water vapour.

One of the PEARL instruments filling the gap in high Arctic measurements is a Bruker IFS 125HR. Measurements made by this high-resolution solar absorption FTIR offer information about the abundances, distributions, and trends for many trace gases, including some vertical information about trace gas profiles (Batchelor et al., 2009; Lindenmaier et al., 2011). Water vapour and δD products are obtained from processing PEARL FTIR spectra using the MUSICA (v2015) Optimal Estimation retrieval technique summarized by Schneider et al. (2016) and Barthlott et al. (2016). This study offers a detailed comparison of the MUSICA dataset to other water vapour measurements made at Eureka, and adds to a collection of other FTIR water vapour intercomparison studies done in a sub-tropical climate (Izaña) by Schneider et al. (2010), an alpine climate (Mt. Zugspitze) by Vogelmann et al. (2011), a sub-Arctic climate (Kiruna) by Buehler et al. (2012), and an Arctic climate (Ny Ålesund) by Palm et al. (2010).

The goals of this study are to describe water vapour abundances near Eureka, Nunavut, assess the agreement between available datasets at Eureka, and examine new FTIR measurements made at PEARL. Section 2 will describe the instruments and datasets available at Eureka. These each involve distinct measurement techniques, whose strengths and weaknesses will be considered. Section 3 will compare the datasets and assess how consistent they are with each other. Particular attention will be given to

5 the new 125HR dataset and the new emission FTIR dataset. Section 4 will discuss the results of the intercomparisons. Section 5 will offer conclusions about the ability of the Eureka datasets contribute to our knowledge of high Arctic water vapour and comment on the implications for future studies.

2   Instrumentation near Eureka

A variety of instruments offer information about water vapour near Eureka. Table 1 summarizes the available datasets, notes

10 how often measurements are taken, and whether total column or profile information is produced. Figure 3 illustrates the atmospheric water vapour measurement time series from each PEARL and EWS instrument. The PEARL Ridge Lab hosts the 125HR solar absorption FTIR and a sun photometer (SPM). 0PAL hosts another SPM and a microwave radiometer (MWR). During the time period examined in this study, there were also emission FTIR instruments that observe downwelling longwave radiation installed at the Ridge Lab and at 0PAL. The Eureka Weather Station launches radiosondes twice per day. A sub-set

15 of radiosonde measurements have been analysed by GRUAN. All datasets reveal the strong seasonal pattern of maximum total column during the warmer summer and minimum total columns during the winter. This section describes these instruments, their water vapour measurement techniques, and uncertainties.

2.1   PEARL Fourier Transform Infrared Spectrometers

There are new water vapour datasets available from three FTIR instruments located at PEARL.

20 ### 2.1.1   Solar absorption FTIR

The Bruker IFS 125HR FTIR Spectrometer (125HR) is a solar absorption FTIR instrument located at the PEARL Ridge Lab. It was installed in July 2006 and joined NDACC thereafter (Batchelor et al., 2009). The 125HR records high-resolution spectra (0.0035 cm^{-1}) during clear-sky conditions while the Sun is above the horizon (late February until late October). Its wide spectral range enables the 125HR to detect many trace gases, including NOy species (Lindenmaier et al., 2011), stratospheric ozone

25 chemistry species (Lindenmaier et al., 2012), and biomass burning products (Viatte et al., 2015). The 125HR can also measure atmospheric water vapour, and is the most northern measurement site of the NDACC, Total Carbon Column Observing Network (TCCON), and MUSICA networks.

The 125HR water vapour dataset used in this study, illustrated by in Fig. 4, was produced using the MUSICA retrieval technique summarized in Schneider et al. (2016) and Barthlott et al. (2016). This process applies an Optimal Estimation technique based on Rodgers (2000), and uses the PROFITT retrieval code of Hase et al. (2004). By using a combination of strong and weak absorption features and a logarithmic scale, the MUSICA retrieval technique seeks to ensure high accuracy and precision across the highly variable abundances of water vapour.

In addition to total columns, vertical information about water vapour is recovered using the MUSICA retrieval technique. Retrievals from PEARL 125HR measurements typically have 2.9 degrees of freedom for signal (DOFS), which represent the number of independent pieces of information acquired about the profile. Figure 5 depicts the rows of a typical MUSICA averaging kernel for Eureka, showing the measurement's sensitivity (the sum of the averaging kernel matrix rows) to information originating at different altitudes. Information about water vapour's vertical distribution is mostly limited to the troposphere.

Significantly, the MUSICA data product also includes total columns of the water vapour isotopologue HDO and its ratio with H_2O, δD, with limited vertical information. This opens up the possibility of investigating the transport history of water vapour (e.g. Schneider et al., 2016), and is a unique contribution to the measurements acquired at PEARL.

2.1.2 MUSICA quality control

MUSICA nominally excludes measurements from spectra recorded at solar zenith angles (SZA) greater than 78.5°. However, this quality control criterion has been relaxed for this study. In addition to filtering out occasional measurements throughout the year, the SZA criterion excludes all measurements obtained between late February (polar sunrise) and late March, as well as between mid-September and late October (polar sunset) due to solar viewing geometries at high latitudes.

Justification for the 78.5° SZA criterion can be seen when examining the SZA dependence of the systematic (e.g., spectroscopic parameters) and statistical errors (e.g. measurement noise). These uncertainty sources are discussed in more detail by Schneider et al. (2012). Figure 6 (a) illustrates that the total error increases rapidly close to 90°. Figure 6 (b) shows that statistical errors begin to increase exponentially for SZAs larger than 78.5°. For the total column comparisons with other instruments, a SZA limit of 85° was applied to the extended MUSICA dataset. Errors increase significantly beyond this point. Including data between 78.5° and 85° improves *N* for a variety of comparisons. Table 3 in Section 3.3.1 shows that different SZA limits have a small impact on observed agreement with other instruments. The limited number of coincidences available with the radiosondes motivates the use of the full extended MUSICA dataset (i.e., no SZA limit) to find a meaningful number of matches for the profile comparisons.

Figure 3 (a) shows both the standard MUSICA dataset with the normal quality control filter applied and the extended MUSICA dataset with this criterion removed.

The sensitivity of MUSICA retrievals to different altitudes is important to consider when performing profile comparisons. This varies from measurement to measurement, and has a seasonal dependency. MUSICA retrievals at Eureka consistently have information throughout the lower and middle troposphere (i.e. sensitivity above 0.9), and often in the upper troposphere. Some retrievals have information in the lower stratosphere, if the sensitivity criterion is relaxed to 0.5. The sensitivity of MUSICA retrievals reaches its highest altitudes at Eureka during March due to the high airmass.

The MUSICA data are additionally quality controlled by requiring reasonable XCO_2 values retrieved from the same spectra that are used for the MUSICA water vapour isotopologue retrievals (for details, see Barthlott et al., 2015). Further quality tests are made by fitting solar line shifts and phase errors of the instrumental line shape (for a summary, see Barthlott et al., 2016).

2.1.3 Emission FTIR

Two related FTIR instruments that observe atmospheric emission at PEARL offer water vapour datasets: the Extended-range Atmospheric Emitted Radiance Interferometer (E-AERI) and the Polar-Atmospheric Emitted Radiance Interferometer (P-AERI). The AERI instruments (Knuteson et al., 2004a,b) measure the downwelling radiation emitted by the atmosphere directly above Eureka continuously, weather permitting, at 1.0 cm^{-1} resolution. E-AERI measurements were calibrated using the standard processing software, while the P-AERI measurements included additional processing (Rowe et al., 2011). The spectral range of the E-AERI is 400 to 3000 cm^{-1}, and that of the P-AERI is 500 to 3000 cm^{-1}. E-AERI measurements were sampled every ~7 minutes. In this work, E-AERI (when located at 0PAL) spectra are averaged over an hour (typically 8 spectra) to reduce noise. P-AERI measurements were sampled every 0.6 to 2 minutes; principal component analysis was used to reduce noise (Antonelli et al., 2004). The sensitivity of downwelling infrared radiance to water vapour is greatest at low altitudes, where water vapour is most abundant. While AERI measurements have been used to retrieve water vapour profiles (e.g., Feltz et al., 2003) retrievals of water vapour using AERI measurements in this work are limited to total column amounts. Measurements of trace gases and radiances taken by these two AERI instruments have been shown to be consistent with each other (Mariani et al., 2012; Mariani et al., 2013).

The P-AERI was installed at 0PAL in March 2006. The E-AERI was installed at the PEARL Ridge Lab in October 2008 (Mariani et al., 2012). After a seven-month overlap period with the E-AERI, the P-AERI was removed. Due to damage incurred to its detectors, the E-AERI did not take measurements between September 2009 and February 2011. Once repaired, the E-AERI was moved to 0PAL, where it remains.

This study will refer to the combined dataset as the AERI; however, its three components will be examined: the P-AERI dataset at 0PAL, the E-AERI dataset at 0PAL, and at the E-AERI dataset while installed at the Ridge Lab. Distinguishing between the two measurement locations is important because of the difference in elevation and because water vapour is most abundant at low altitudes. These datasets are illustrated in Fig. 3 (b).

5 The PEARL AERI water vapour products are new, and are published here for the first time. The retrieval technique applied to the AERI measurements to produce the water vapour dataset is based on Rowe et al. (2008). The water vapour retrieval involves the following steps: 1) An atmospheric profile is created describing the atmospheric state at each radiosonde time, 2) The error budgets for the measured and simulated downwelling radiances are determined, 3) The frequency-dependent uncertainties in measured and simulated downwelling radiances are used to select the best set of frequencies for retrieving water vapour

10 amounts, 4) Clear-sky time periods are identified, 5) Water vapour column amounts are retrieved by finding the scale factor that minimizes the difference between measured and simulated downwelling radiances, where the scale factor scales the first guess water vapour profile. These steps are described in detail below, highlighting differences from Rowe et al. (2008).

First-guess atmospheric profiles are created using height, pressure, temperature, and water vapour amounts from the radiosondes. Above about 10 km, relative humidity is too low to be accurately measured by the radiosonde sensor and thus the

15 water vapour amount is fixed to 5 ppmv. Ozone is determined from ozonesonde flights made at Eureka. Other trace gases, including CO_2 (Conway et al., 2011), N_2O, CH_4, and chlorofluorocarbons (CFC-11, CFC-12, CFC-113) are based on surface measurements made at Alert, Nunavut, or Barrow, Alaska through the NOAA Earth System Research Laboratory (ESRL) Global Monitoring Division. Mixing ratios are assumed to be constant with height. For other trace gases, and for values above the top of the soundings, the subarctic summer and winter standard atmosphere models of McClatchey et al. (1972) are used.

20 The profiles are interpolated onto a layered model atmosphere extending from 0 to 60 km with 60 layers that get progressively closer together moving down from 60 km to the surface.

After developing atmospheric profiles, the error budgets of the measured downwelling radiances are estimated as described in Rowe et al. (2008). Following that, uncertainties in simulated downwelling radiances are determined for monthly-averaged atmospheric profiles. Downwelling radiances are simulated using the line-by-line radiative transfer model (LBLRTM; Clough

25 et al., 2005) for the first-guess atmospheric profile and then with each atmospheric constituent perturbed by its estimated uncertainty. The difference is calculated to give the uncertainty in the simulated radiance due to the uncertainty in the constituent. The uncertainty in temperature is assumed to be 0.5 K and is calculated as a bias in the entire profile. The uncertainty in CO_2 is assumed to be 2%. For other trace gases, including HNO_3, N_2O, CH_4, O_3, CO, C_2H_2, HCN, CCL_4, CFC-11, CFC-12, and CFC-113, an uncertainty of 10% is assumed. In addition, uncertainties of 30% in the continuum of water

30 vapour emission between strong lines due to self-broadening and foreign-broadening are assumed. The combined uncertainties in measured and simulated downwelling radiances are calculated assuming the sources of error are uncorrelated (that is, by

taking the square root of the sum of the squared errors), resulting in uncertainties in units of downwelling radiance (mW*(m^2 sr cm^{-1})$^{-1}$).

To identify the best frequencies for retrieving water vapour, the sensitivity of the downwelling radiance to water vapour is determined by calculating the change in simulated radiance for a given percent change in the first-guess water vapour profile.

5   This value is inverted to give the percent change in water vapour per change in radiance (in units of mW*(m^2 sr cm^{-1})$^{-1}$). Multiplying this by the combined uncertainty in the measured and simulated downwelling radiances removes the radiance units and gives the uncertainty in water vapour as a percentage of the first-guess profile. The 100 frequencies where this uncertainty is the lowest are selected for monthly average atmospheric profiles. Thus the chosen frequencies vary by month. They include frequencies in the microwindows between strong lines from 500 to 600 cm^{-1}, in the atmospheric window between

10   700 and 1100 cm^{-1}, and in the wing of the strong water vapour absorption feature centred at 1600 cm^{-1}, between 1100 and 1400 cm^{-1}.

To identify clear sky time periods for the P-AERI, we make use of cloud property retrievals performed by Cox et al. (2014). For each P-AERI spectrum, if the retrieved cloud optical depth is below 0.6, we assume skies are clear of clouds. For the E-AERI, clear skies are identified as described in Mariani et al. (2013).

15   Water vapour column amounts are retrieved as follows. Optical depths are created (using the LBLRTM) for each model layer using the atmospheric profiles at the radiosonde times. These optical depths depend very weakly on temperature and therefore show little sensitivity to temperature changes between radiosounding times. Two sets of optical depths are created, one for all gases besides water vapour and another for water vapour alone. For each AERI measurement, the non-water gaseous optical depth (τ_g), the water vapour optical depths (τ_w), and the atmospheric temperature (T) are interpolated to the measurement time.

20   Because trace gas amounts are assumed to be constant within each model layer, this is approximately equivalent to interpolating the trace gas amounts to the measurement time and re-computing optical depths. For the layer optical depth due to water vapour (wv), $w_1 \tau_w(wv_1) + w_2 \tau_w(wv_2) \approx \tau_w(w_1 wv_1 + w_2 wv_2)$, where the subscripts 1 and 2 refer to times before and after the time of interest and the w_i are weights for the interpolation. An exception is the self-broadened part of the water-vapour continuum (τ_s), which does not depend linearly on wv, but rather on wv^2; τ_s is therefore excluded from τ_w. τ_s is given by a simple empirical

25   function and can be calculated separately for the interpolated water vapour amount. A forward model was created using the measured downwelling radiances, the first-guess T, wv, τ_g, τ_w and τ_s profiles, the uncertainty described earlier, and a water-vapour scale factor. The forward model scales τ_w and τ_s according to the scale factor, sums τ_g, τ_w and τ_s to get the total optical depth, computes the downwelling radiances in the same manner as the LBLRTM, and reduces the resolution to match that of the AERI. Finally, the forward model calculates the weighted mean difference between the measured and simulated radiances, where the weights are determined from the uncertainties. The value of the scale factor that minimizes the weighted mean

difference is computed from the forward model using golden section search and parabolic interpolation to a tolerance well below the uncertainty level.

2.2 Sun photometer

The PEARL Ridge Lab and 0PAL have both hosted a Cimel SPM. These datasets are illustrated in Fig. 3 (c) and (d), which show the higher total columns measured at 0PAL, due to the lower altitude. SPMs measure solar radiation in eight spectral channels between 340 nm and 1640 nm. These automated sun-viewing radiometers are part of the global Aerosol Robotic Network (AERONET), and contribute data to a global aerosol optical depth (AOD) database (Holben et al., 1998). In this report, the AERONET Level 2.0 data product is used, which has been cloud-screened and quality assured. The Eureka SPMs are calibrated annually, and re-installed in the spring.

SPM data are used to produce total column aerosol optical depth measurements from inversions of spectral direct-sun and sky radiances. For water vapour measurements, a modified Langley plot technique described by Holben et al. (1998) is applied to observations of a spectral window at 940 nm. SPMs make measurements approximately every 3 minutes but are limited to clear sky conditions. Validation studies of the AERONET Cimel SPMs have shown that their data underestimates the water vapour total column by 10% (Alexandrov et al., 2009).

2.3 Microwave radiometer

The MWR at 0PAL, a Radiometrics WVQ-1500, was installed in March 2006 in collaboration with NOAA. The MWR records microwave emissions in five channels between 22-30 GHz with a beam width of 5°. Two of the channels are used to statistically derive the PWV from zenith-pointing measurements (see Liljegren and Lesht, 1996 and Westwater et al., 2001). The measurements and retrieval technique are applied a few times per minute. This technique enables the MWR to observe water vapour all day and night in most conditions, and to capture short-term variability. 30-minute averages have been calculated for use in this study. This MWR timeseries is shown in Fig. 3 (e).

The MWR dataset's quality control has not removed all problematic measurements. Large spikes into high unphysical PWV values, which occurred during precipitation events, have been filtered out manually. There has not been maintenance and calibration of the MWR, on account of its remote location, since a 2008 visit.

2.4 Radiosondes

Radiosondes are widely used to measure temperature, pressure, and atmospheric water vapour (relative humidity) at meteorological stations around the world. Eureka radiosonde measurements are taken by an instrument payload lofted into the

atmosphere by a hydrogen-filled balloon, launched twice daily (11:15 and 23:15 UT) from the Eureka Weather Station. Occasionally, radiosondes are launched at other times of the day for campaign-related reasons. Typically, these balloons (and the measurements) reach the middle of the stratosphere (30 to 33 km). In this work, if a radiosonde did not reach 15 km, its data was filtered out. This ensures the calculation of the total column using an integrated profile is justifiable. Indeed, the 2930

5   radiosondes which reached a minimum altitude of 30 km between 2006 and 2015 show that 99.4 ± 1.2% of water vapour above Eureka is located below 15 km.

The Vaisala RS-92 radiosonde model currently used by the Eureka Weather Station has been subject to extensive testing and validation. Relative humidity is measured using a thin-film capacitor. This design has been shown to work well at cold temperatures (below -70° C) and low abundances (below 5 ppm), but its sensitivity to water vapour is limited at low pressures

10   (Miloshevich et al., 2009). Moreover, the reported relative humidity values are given in whole numbers, whereas the measured values include two decimal place precision. This introduces a 0.5% relative humidity round-off error (Miloshevich et al., 2009). The RS-92 Vaisala radiosonde model is also known to have a dry radiation bias due to solar heating of the sensor (Vömel et al., 2007). Measurements taken during the AIRS Water Vapour Experiment (AWEX) campaign at the Atmospheric Radiation Measurement Southern Great Plains site in 2004 showed a daytime dry bias of 6-8% when compared with a co-located

15   microwave radiometer (Miloshevich et al., 2009). This error depends on the solar radiation intensity, which is a function of the SZA and sensor orientation. The lack of a protective cover introduces a second error source, which partially offsets the radiative heating: forced cooling occurs as the sensor rises with the balloon. This effect depends strongly on pressure, and thus decreases with altitude.

Radiosonde relative humidity measurements were converted to mixing ratio, which were integrated and converted to

20   precipitable water vapour for total column comparisons.

2.4.1   GRUAN

A subset of the radiosonde data has been analysed using the technique developed by GRUAN. GRUAN aims to provide a traceable reference standard (Immler et al., 2010), motivating its inclusion in this study. GRUAN data processing for the RS-92 instrument is described by Dirksen et al. (2014), and accounts for known biases in radiosonde measurements.

25   Another advantage of the GRUAN processing is that it recovers GPS location information about radiosonde flights, enabling an estimation of typical flight paths and distances from Eureka. The input for this re-analysis of the radiosonde measurements required raw data files, which were not available for all sonde flights. This is illustrated in Fig. 3 (f), which shows both the radiosonde and GRUAN timeseries. Based on the available raw radiosonde files processed by GRUAN (gaps in the raw file record result in a smaller number of measurements than the standard radiosonde product over the same time period),

radiosondes launched from the Eureka Weather Station typically stay close to Eureka. Radiosondes reached a mean maximum (horizontal) distance from the Eureka of 64.1 ± 47.3 km and 81% of flights stayed within 100 km of the Ridge Lab.

2.5 Summary of Eureka water vapour datasets

This section has presented the water vapour datasets from several Eureka ground-based instruments, radiosondes, and two satellite instruments with Eureka-coincident measurements. Table 1 summarized these datasets and notes how often measurements are taken, and whether total column or profile information is produced. Table 2 summarizes estimated accuracies for each water vapour product.

3 Comparison of water vapour measurements

3.1 Method

Two approaches have been taken in comparing the available datasets. The 125HR, P-AERI, E-AERI, SPMs, MWR, and radiosondes produce total column measurements. Coincident total column measurements have been compared using difference and correlation plots. Radiosonde and 125HR profiles have been similarly compared.

No instrument is used as a common reference. A full accounting of the biases between every combination of instruments is presented that shows how each dataset relates to the others.

When comparing profiles with the 125HR, radiosonde profiles were smoothed by the 125HR averaging kernel and using the MUSICA a priori profile. The procedure for smoothing followed Rodgers and Connor (2003). This accounts for altitude sensitivity differences between the instruments. Before smoothing, the 125HR a priori profile is used to fill any gaps in the comparison profile (i.e., altitudes above the upper limit of radiosonde measurements). After smoothing, altitudes for which there were no original data were removed.

Specific altitudes have been selected to compare the radiosonde and 125HR profiles, representative of different regions of the troposphere.

In this study, both total column and profile values are compared using absolute and percent differences. These are calculated using:

$$difference = X - Y \tag{1}$$

$$\% \, difference = \frac{(X - Y)}{\frac{X + Y}{2}} * 100\%. \tag{2}$$

Percent differences are considered with respect to the average of the two measurements to avoid taking one of them as a reference.

Profile differences shown are the mean of individual coincident profile differences, following Eq. (1) and Eq. (2), rather than the differences between the instruments' mean profiles.

3.2   Coincidence criteria

A two-hour temporal coincidence criterion was applied for all comparisons. If multiple coincidences were found within this interval, only the closest pair was kept. Each pair of coincident measurements is thus independent of others contributing to the overall assessment of different measurement techniques. This method often results in a substantially smaller time difference between measurements than is otherwise permitted by the criterion. All comparisons were also performed using all possible coincidences within this criterion. While significantly increasing the number of matches, the observed agreement between instruments was very similar.

The wide time criterion was chosen to ensure sufficient matches were found for a reasonable study, especially for comparisons involving the radiosondes. Radiosondes are launched twice a day at 6:15 and 18:15 local time. This generally does not align with measurements that require sunlight (i.e. SPMs, 125HR), especially during spring and fall. Figure 7 illustrates the trade-off between the mean percent difference (and scatter, the mean standard deviation) between the 125HR and radiosonde and SPM measurements. The scatter shows how consistently different the instruments are at each temporal coincidence criterion.

Other instruments show similar patterns with an initial increase in the number of coincident pairs levelling off for a larger temporal coincidence criterion. Since only the closest pair is kept, the benefit to a larger temporal criterion is much less when the comparison instrument offers a high temporal density of data (e.g. the 125HR and SPM-RL).

3.3   Precipitable water vapour comparisons

In this study, water vapour total columns are compared using PWV units. This unit represents the height of the layer of water which would result from the condensation of the entire total column to standard temperature and pressure. (It may be useful to note that mm of PWV are equivalent to [kg m^{-2}].) PWV values near the equator, where warm atmospheric temperatures can hold much greater quantities of water vapour, can be as large as 50 mm. In Eureka, as shown by Fig. 3, water vapour columns rarely exceed 20 mm PWV during summer and columns are frequently below 2 mm PWV during winter.

3.3.1 Radiosondes

The accuracy of the Eureka radiosonde dataset is usefully characterized by comparison to the 2377 radiosonde measurements processed by GRUAN. The radiosonde and GRUAN total columns agree closely, with a mean difference (RS - GRUAN) of -4.6 ± 3.8% (R = 1.00). Differences reveal that the Eureka radiosonde water vapour total columns have a small dry bias relative to GRUAN. The magnitude of this bias varies seasonally, with radiosonde columns dry biased by up to around 1 mm PWV during the summer (0.6 mm PWV or 5.0% on average). During winter, agreement is very close; the differences in the columns are 0.1 mm PWV (or 4.2%) on average. This is seen clearly in the profile differences as well, as shown in Fig. 8.

3.3.2 125HR

The PEARL Ridge Lab's 125HR water vapour measurements show good agreement with those of other Ridge Lab instruments. The percent differences between the 125HR and instruments located at the Ridge Lab or with sufficient profile information to create a total column from the altitude of the Ridge Lab (610 m) are illustrated by Fig. 9.

The Ridge Lab SPM shares the same location and the same solar-viewing measurement geometry as the 125HR. Despite measuring the same airmass at the same time (the mean difference in measurement time between the 125HR and SPM-RL is 5.3 minutes), Fig. 9 (a) shows that SPM-RL measurements are consistently smaller than the 125HR. The MUSICA product is consistently wet biased (measures more water vapour) with respect to the SPM. The difference between the instruments varies seasonally, and is largest in the summer. Seasonal variations in the difference between FTIR and sun-photometer/radiometers have also been reported by Schneider et al. (2010) for a subtropical site.

During the period of time when the E-AERI was installed at the Ridge Lab, its measurements compared favourably with the 125HR. This is shown in Fig. 9 (b). The mean percent difference in coincident measurements was -0.4 ± 4.4 mm PWV (5.1 ± 5.1%). This is the closest agreement of all comparisons with the 125HR in this study. It is consistent with the result of the GRUAN comparison, showing a small wet bias in the Eureka MUSICA product.

When comparing the 125HR and radiosondes, radiosonde profiles were smoothed with the 125HR averaging kernels and the total column calculated down to the PEARL Ridge Lab altitude of 610 m. Only 64 coincidences were found. As shown in Fig. 9 (c), the mean percent difference is 11.4 ± 13.2%, with no clear seasonality. This difference is larger than the expected accuracy of the measurements. The MUSICA product is wet biased with respect to the radiosondes, beyond the small dry bias of the radiosondes.

Comparisons with the GRUAN radiosonde products are limited to only 10 coincident measurements. The mean percent difference, shown in Fig. 9 (d), was 1.5 ± 9.5%. Extending the coincidence criterion to allow for measurements within 3 hours instead of 2 hours produces 93 coincidences. In this case, the difference between the 125HR and GRUAN is 6.1 ± 10.6%.

Increasing the coincidence criterion to four hours yields a very similar result for 171 matches. (If the radiosonde comparison's criterion was also extended to 3 hours, the agreement with the 125HR would become $10.0 \pm 12.8\%$, which is consistent with the 2 hour coincidence criterion result.) This comparison shows that the MUSICA product (v2015) is very likely wet biased by about 5% relative to the GRUAN product.

Comparisons between the Ridge Lab and 0PAL instruments show substantially larger total columns observed from 0PAL's altitude, as expected. Nonetheless, correlation coefficients (R) for all comparisons involving the 125HR are 0.96 or better.

Table 3 shows that these comparisons (shown in Fig. 9), made with the extended MUSICA dataset with a relaxed 85° SZA limit applied, are consistent with the agreement that would be found with other Eureka instruments if the standard MUSICA dataset was used.

3.3.3   AERI

As illustrated in Fig. 10, comparisons between the E-AERI while it was located at 0PAL and other 0PAL instruments show close agreement. Since AERI measurements occur throughout the day and night, and during Polar Night, many coincidences are found with radiosondes and the MWR. Comparisons with the E-AERI showed agreement of $6.7 \pm 6.2\%$ (N = 196) with radiosondes, $2.0 \pm 6.2\%$ (N = 166) with GRUAN, and $3.4 \pm 5.2\%$ (N = 130) with the SPM-0PAL. Despite being co-located, agreement between the E-AERI and MWR were the least favourable in this study, $41.7 \pm 89.4\%$ over 1191 matches.

The P-AERI showed agreement of $5.6 \pm 6.6\%$ (N = 517) with the radiosondes, $4.7 \pm 5.0\%$ (N = 2662) with the SPM-0PAL, and $-4.7 \pm 15.1\%$ (N = 6975) with the MWR.

Comparisons between the E-AERI while it was installed at the PEARL Ridge Lab and Ridge Lab instruments (125HR and SPM-RL) show good agreement as well ($5.1 \pm 5.1\%$ with 176 coincidences for X = E-AERI-RL and Y = 125HR) and $6.7 \pm 4.4\%$ with 867 coincidences, (X = E-AERI-RL and Y = SPM-RL). This is illustrated in Fig. 11.

Correlation plots illustrating the agreement between the E-AERI at the Ridge Lab and the 125HR and SPM-RL are illustrated in Fig. 12, along with correlation plots illustrating the agreement between the E-AERI while located at 0PAL and the radiosonde datasets.

3.3.4   Summary of PWV comparisons

Comparisons were conducted between all combinations of the instruments involved in this study. Not every combination of instruments was shown in detail in this section, which has focused on comparisons between the new FTIR datasets (125HR and AERIs) and instruments located at the same altitude. A complete set of correlation plots for available instrument dataset

combination is given in Fig. 13. Each row and column of correlation plots in Fig. 13 illustrates how well a specific Eureka instrument agrees with the other instruments. Biases are significantly influenced by whether comparison instruments are located at the RL or 0PAL. For example, the 125HR shows a high ("wet") bias with respect to the E-AERI and SPM at the RL but a low ("dry") bias with respect to the E-AERI, SPM, and other instruments at 0PAL. Presenting all combinations of instruments in this manner allows the biases between the instruments to be observed as they relate to one-another.

The results of all instrument comparisons are summarized in Table 4 (125HR, SPM-0PAL, E-AERI-0PAL, P-AERI-0PAL, MWR, GRUAN, RS, E-AERI-RL, and SPM-RL), which includes the number of coincidences found, mean absolute and percent differences, correlation coefficients, and correlation plot slopes. High resolution versions of Fig. 13 and Table 4 are included in the supplementary materials as Fig. S1 and Table S1 respectively for easier readability.

3.4 Profile comparisons

3.4.1 125HR and radiosondes

Comparison between the 125HR and radiosondes smoothed by 125HR averaging kernels in the troposphere shows agreement within 14%; agreement above 6 km is better than 24 ppmv and 5%. If the comparison is extended to 15 km, agreement is within 2 ppmv and 6%. However, the better agreement at higher altitudes might be partly due to the fact that there the 125HR remote sensing system is less sensitive.

Between 2 and 5 km, the magnitude of the observed difference between 125HR and radiosonde measurements is slightly larger than the assumed accuracy of the measurement techniques (which is about 10% for the FTIR, see Fig. 5 of Schneider et al. (2012) and 3-8% for the radiosonde, depending on the season, based on GRUAN comparisons). The 125HR shows consistently higher water vapour abundances, as illustrated in Fig. 14. If all coincidences are used (rather than unique pairs of profiles), 250 coincident profiles are found showing the 125HR to have a similar wet bias of 6 to 14% throughout the troposphere.

Examining specific altitudes representative of the lower, middle and upper troposphere reveals that the agreement is worst between the 125HR and radiosondes at 3 km (-14.2 ± 25.4%) and best in the upper and lower troposphere (4.2% ± 20.3% at 1.01 km; -0.5 ± 38.0% at 7.98 km). As illustrated in Fig. 15, this agreement does not show a clear seasonal dependence; however, this might be due to the lack of sufficient coincidences throughout the year.

The agreement between the 125HR and radiosondes is also illustrated in Fig. 16, which shows good correlation (R values above 0.88) between the measurements at all altitudes.

Increasing the temporal coincidence criterion for the 125HR-radiosonde comparison to 3 hours resulted in 381 coincident measurements (compared to 84 for 2 hours), and results showed similar agreement.

3.5 Summary of profile comparisons

Comparisons between the two instruments with water vapour profile information were somewhat limited by the relatively small number of coincident measurements. Table 5 provides a summary of comparisons between the 125HR and the radiosonde measurements. The standard deviations of all altitudes examined in this comparison show there can be large

5 variability in the agreement between profiles.

4 Discussion

4.1 Total column comparisons

Total column comparisons between Eureka water vapour datasets exhibited good overall agreement. All comparisons showed R values of 0.96 or greater, indicating excellent correlation between Eureka water vapour measurements taken by several

10 different instruments. The E-AERI-0PAL and SPM-RL comparison is an exception; it had an R value of 0.90. However, this comparison only included 56 coincident measurements across a limited number of summer days, different measurement geometries, and different measurement locations.

As expected, the 600 m altitude difference between the two measurement sites leads to higher total PWV observed by instruments at 0PAL when compared to those at the Ridge Lab. This is illustrated by the comparison between the SPM at

15 0PAL and the SPM at the Ridge Lab in Fig. 17, which shows a mean percent difference of 33.6 ± 9.5% (X = SPM-0PAL and Y = SPM-RL in Eq. 2) using a very large number (N = 37,306) of closely coincident measurements (mean coincidence time of 4.25 minutes). The magnitude and variability of the difference between identical instruments at the two sites a short distance away (15 km) demonstrates the significant influence of altitude and meteorological variability on water vapour measurements. This provides useful context for other comparisons of instruments at the Ridge Lab and 0PAL sites. Instruments that sample

20 different air masses because of their line-of-sight, location, measurement timing and length, or sensitivity will show differences because water vapour has high variability over short timespans, altitudes, and distances.

4.1.1 125HR

The Eureka MUSICA product shows close agreement with other instruments' measured water vapour total columns; however, it consistently shows a wet bias. The agreement between the PEARL 125HR and (smoothed) Eureka radiosondes, for example,

25 is 11 ± 13%, which is a larger difference than expected from the accuracy of the instruments. Given that the radiosondes appear to have a small dry bias of 4 to 5% based on GRUAN comparisons, the radiosonde measurements suggest the 125HR has a wet bias of around 6%. Indeed, comparing the 125HR with the GRUAN data, using 3 hours as the time criterion, the bias is

6.1 ± 10.6%. Furthermore, comparisons between the 125HR and other instruments support this wet bias observation. Agreement between the 125HR and the E-AERI (while co-located at the Ridge Lab) is 5.1 ± 5.1%, which is similar to the bias exhibited by the radiosondes and GRUAN comparison. Agreement between the 125HR and the Ridge Lab SPM is 15.1 ± 6.5%. This wet bias is larger than other comparisons; however, this result is consistent with other comparisons in the context

5   of SPMs' known underestimation of water vapour by up to 10%. Differences between the 125HR and SPM-RL are usefully informative since the instruments share a solar-viewing measurement geometry, are co-located (within meters), and have closely coincident measurement times.

4.1.2   Differences between MUSICA retrieval v2012 and v2015 results at Eureka

The previous MUSICA retrievals (v2012) for the Eureka 125HR showed closer agreement with other instruments than the

10 current (final) MUSICA retrieval version (v2015) used in this study. Agreement between the Eureka v2012 MUSICA retrieval and radiosondes (2.5 ± 8.2%), E-AERI-RL (-0.9 ± 4.8%), and SPM-RL (8.6 ± 6.2%) are within expected instrument accuracies and biases (e.g., the dry bias of the radiosondes and SPMs). Moreover, the v2012 comparison results are consistent with previous MUSICA intercomparison studies (e.g. Schneider et al., 2010). A version of Table 4, which summarizes the water vapour total column comparisons, using the MUSICA v2012 retrieval instead of v2015 is available in the supplementary

15 materials as Table S2.

Two major changes in the v2015 MUSICA retrieval, aimed at improving the network-wide consistency of the MUSICA FTIR products from sites around the world, may have contributed to this change in observed agreement with other high Arctic measurements. First, the a priori profile used for v2015 retrievals was a global average of water vapour, replacing v2012's site-specific a priori profiles based on local radiosonde measurements. For Eureka retrievals, this meant the v2015 a priori

20 profile had more water vapour than is present at Eureka except during summer. Second, the v2015 retrieval replaced two of v2012's spectral windows containing strong absorption features (which had saturated at high-humidity sites such as Wollongong, Australia), with one spectral window having weaker absorption lines (Barthlott et al., 2016). It is unlikely that the strong absorption features used in v2012 saturate in high Arctic spectra, and likely that they contributed useful information to the retrieval. While MUSICA's change in spectral lines enables consistency across FTIRs located at sites with very different

25 humidity conditions worldwide, the v2015 spectral windows and global a priori might not be ideal for high-latitude, low-humidity conditions. Nevertheless, an Arctic bias of a few percent is still a good value for a dataset that is optimised for global consistency.

The difference between the v2015 and v2012 Eureka MUSICA products, illustrated in Fig. 18, is 5.9 ± 0.8% overall. Differences are greatest during summer and follow a seasonal pattern. Figure 19, a correlation plot for both datasets, shows

30 that the difference increases linearly as humidity increases. This difference is very similar in magnitude to the observed bias

relative to other Eureka instruments, and suggests changes to the MUSICA retrieval (the selected spectral windows and the usage of a global uniform a priori profile) may have slightly decreased the accuracy at the extreme Arctic site of Eureka.

4.1.3 AERI instruments

Close agreement with measurements taken by other co-located Eureka instruments confirms the accuracy of the new AERI datasets. The P-AERI (located at 0PAL) showed good agreement with the co-located SPM-0PAL (4.5 ± 5.0%), RS (5.6 ± 6.6%), and MWR (-4.7 ± 15.1%). The small wet bias observed with the SPM and RS may be due to the dry bias of those instruments. The E-AERI, while located at 0PAL, showed very similar results in comparisons to the SPM-0PAL and radiosonde datasets. While located at the Ridge Lab, the E-AERI showed similar agreement with the SPM-RL. However, while it was at 0PAL, the E-AERI showed the least agreement of all instrument comparisons with respect to the MWR (41.7 ± 89.4%). The significant difference between the E-AERI and P-AERI comparisons with the MWR is due to the different time period of overlap with the MWR. The comparisons presented in this work revealed that the MWR developed a significant calibration problem affecting its water vapour measurements after the winter of 2010. This is discussed in the next sub-section. Thus, the disagreement between the E-AERI and MWR is not an indication of a bias in the E-AERI water vapour product, which otherwise agrees very well with other measurements.

4.1.4 Microwave radiometer

The MWR observes water vapour year-round; however, its data agrees least with other instruments and that agreement often has relatively large scatter.

Agreement between the radiosondes and the MWR (approximately co-located and with similar viewing geometries) is close from mid-2006 until the fall of 2010, but then worsens dramatically thereafter. Prior to 2010 the MWR agreed with the radiosondes within 2.9 ± 15.3%, which is consistent with the small RS dry bias. After 2010 the agreement was -13.2 ± 140.0%. The differences between the MWR and radiosondes is shown in Fig. 20 (a). Close inspection of the MWR timeseries reveals that after the winter of 2010-2011, MWR water vapour total columns regularly reach negative values in low-humidity conditions.

The accuracy problem of the MWR dataset is likely caused by the limited calibration and maintenance the instrument has received since installation. Poor agreement with the MWR is interpreted to indicate a problem with the MWR data, rather than of other instruments. For example, the P-AERI shows close agreement with the MWR; however, it was only installed at 0PAL from 2006 to 2009. The E-AERI, a very similar instrument, shows poor agreement with the MWR. This is shown in Fig. 20 (b). The E-AERI, however, was installed at 0PAL in 2011, after the MWR appears to have suffered calibration problems.

4.1.5 Radiosondes

Comparisons to the GRUAN radiosonde product shows that the standard radiosonde dataset has a dry bias, particularly in the Eureka summer. This is consistent with other radiosonde studies (e.g. Vomel et al., 2007; Miloshevich et al., 2009). This dry bias helps explain why the radiosonde and GRUAN measurements don't show identical agreement with other instruments.

5   Comparisons between GRUAN and the 125HR, SPM-0PAL, and E-AERI-0PAL show similar or slightly better agreement (-4.7 ± 4.9%, 0.55 ± 6.39%, -2.0 ± 6.2%) than comparisons between the standard radiosonde data and those same instruments (-11.6 ± 12.4%, -2.07 ± 6.7%, -6.7 ± 6.2%).

4.1.6 Sun photometers

Nearly all comparisons between PEARL SPMs and co-located instruments suggest the SPMs have a dry bias. The exception
10   is the comparison between the SPM at 0PAL and the GRUAN dataset.

4.2 Profile comparisons

Profile comparison results show moderate agreement between the measurements relative to the total column comparisons.

Agreement between the 125HR and Eureka radiosondes of better than 15% throughout the troposphere confirms that the MUSICA product provides reliable information about tropospheric water vapour. This is not as close as the 10% accuracy
15   reported by Schneider et al. (2016); however, this comparison was limited by the small number of coincidences and involved two measurement sites with different meteorological conditions. Observed differences are likely impacted by sampling of different air. While radiosondes sample air close to Eureka throughout their profile, the 125HR's solar-viewing geometry samples air hundreds of km south of Eureka (even within the troposphere) due to the large SZA of high-latitude measurements. In addition, the observed 125HR dry bias at most altitudes with respect to the radiosonde profiles is consistent with the observed
20   dry bias in the total column comparisons. Within the context of a small radiosonde dry bias and a small 125HR wet bias, the agreement between coincident 125HR and radiosonde profiles is within the expected accuracies of the instruments.

The wet bias observed in v2015 relative to the v2012 MUSICA total columns at Eureka also exists in the profiles. Comparisons between the v2012 Eureka product and the radiosondes were within 10% at all altitudes.

5 Conclusions

25   This study compared high Arctic water vapour measurements taken by several different instruments located at Eureka, Nunavut. This large site-wide intercomparison has confirmed the value and reliability of new measurements (e.g., the Eureka

MUSICA 125HR and AERI products), and provided a detailed accounting of the comparability of measurements from a variety of commonly-used atmospheric monitoring instruments.

The accuracy of the MUSICA dataset derived from Eureka 125HR spectra is supported by comparisons with coincident measurements taken at Eureka. The MUSICA v2015 product shows close agreement with other instruments; however, it shows a small wet bias, which was not observed in comparisons using the previous MUSICA v2012 retrieval. Changes to the MUSICA retrieval intended to balance the needs of globally distributed FTIR sites appears to have created a small wet bias at this extremely dry high-latitude site. This underscores the challenge in assuring consistency across global observation networks as well as the performance of measurement techniques operating across a wide range of conditions.

This result affirms the conclusions of previous intercomparison studies at other MUSICA sites that 125HR measurements can yield reliable and precise information about atmospheric water vapour total columns. Moreover, these results also support the use of 125HR measurements taken beyond MUSICA's standard 78.5° SZA limit. Comparisons in this study included SZAs up to 85° without sacrificing the consistency with other instruments' measurements. Relaxing this constraint is useful for Polar sites. Observed differences between the 125HR and other instruments are consistent with well-understood measurement technique biases and differences in observation geometry. 125HR profiles showed reasonable agreement with coincident profiles from radiosondes.

The new AERI datasets presented in this study showed close agreement with other Eureka instruments. AERI measurements thus offer reliable continuous observations of atmospheric water vapour total columns without reliance on sunlight. Adding this capability to existing water vapour observations is especially valuable at Eureka because there is no sunlight between mid-October and late-February, and for parts of the day during spring and fall. 125HR and AERI retrievals thus offer a reliable, accurate, and frequent source of information about atmospheric water vapour at Eureka.

Radiosonde measurements are often used a reference because of their reliable and well-understood character, as well as their high vertical resolution. The GRUAN product resulting from processing the raw radiosondes data reveals that the Eureka radiosondes measurements are affected by a small dry bias, which is largest in the summer. This should be taken into account when using them as a reference for climatological and atmospheric investigations.

PEARL's MWR stands out as a potentially problematic dataset. Reasonable correlations and agreement between the MWR and other datasets were found; however, under low-PWV conditions MWR measurements after fall 2010 show significant disagreement with other co-located instruments. The comparisons conducted in this work demonstrate the need to continuously validate data, especially data resulting from experiments in remote locations where maintenance and calibration opportunities are difficult and infrequent but necessary. Despite the observed negative bias in MWR water vapour measurements during low-PWV times of the year after fall 2010, good agreement was observed during summer months throughout the timeseries.

The moderate number of coincident measurements with radiosondes presented in this study suggest the 125HR offers accurate information about water vapour abundances in the troposphere. The assertion of Schneider et al. (2016) that MUSICA retrievals offer approximately 10% accuracy is affirmed; however, a small wet bias of 5 to 6% is observed at Eureka.

No single instrument is capable of capturing information about atmospheric water vapour at all times. The widespread agreement across the suite of Eureka instruments with different observation strengths offers a valuable capability to collectively measure water vapour abundances and variability in the Canadian high Arctic. The agreement seen between total column measurements suggest that a unified water vapour product at Eureka could be explored. This appears especially promising with a reliable AERI product available to fill in the Polar Night measurement gap left by the SPMs and 125HR, as well as the large temporal gaps (i.e. 12 hours) between radiosonde measurements.

A follow on study of Eureka radiosonde and MUSICA water vapour profiles compared to a suite of satellite datasets is planned for the future.

Acknowledgements.

CANDAC/PEARL funding partners are the Arctic Research Infrastructure Fund, Atlantic Innovation Fund/Nova Scotia Research Innovation Trust, Canadian Foundation for Climate and Atmospheric Science, Canada Foundation for Innovation, Canadian Space Agency (CSA), Environment and Climate Change Canada (ECCC), Government of Canada International

5 Polar Year, Natural Sciences and Engineering Research Council (NSERC), Ontario Innovation Trust, Ontario Research Fund, Indian and Northern Affairs Canada, and the Polar Continental Shelf Program. Spring visits to PEARL were made as part of the Canadian Arctic ACE Validation Campaigns supported by CSA, ECCC, NSERC, and the Northern Student Training Program. This work also received funding from the NSERC CREATE Training Program in Arctic Atmospheric Science, the NSERC-supported PAHA project, and the CSA-supported CAFTON project.

10 The authors would like to thank PEARL Site Manager Pierre Fogal and the CANDAC operators for logistical and operational support at Eureka; ECCC for providing the radiosonde data; AERONET/AEROCAN for sun photometer data; Rodica Lindenmaier, Rebecca Batchelor, and Joseph Mendonca for 125HR measurements; Mareile Wolff for early E-AERI measurements; and CANDAC Data Manager Yan Tsehtik. Chris Cox at NOAA and Patrick Sheese at the University of Toronto offered helpful comments.

15 The authors wish to thank the staff at ECCC's Eureka Weather Station for logistical and on-site support.

MUSICA has been funded by the European Research Council under the Eureopean Community's Seventh Framework Programme (FP7/2007-2013) / ERC Grant agreement number 256961.

The NOAA SEARCH program provided funding for the installation and operation of the P-AERI at Eureka from 2006 to 2009. Von P. Walden provided calibrated quality-controlled P-AERI data.

20 P.M.R. acknowledges funding from FONDECYT Regular 1161460 and from the National Science Foundation under PLR 1543236

6 References

ACIA: Arctic Climate Impact Assessment. Cambridge University Press, 1042p, 2005.

Alexandrov, M.D., Schmid, B., Turner, D.D., Cairns, B., Oinas, V., Lacis, A.A., Butman, S.I., Westwater, E.R., Amirnov, A., and Eilers, J.: Columnar water vapor retrievals from multifilter rotating shadowband radiometer data, J. Geophys. Res., 114, D02306, 2009.

Antonelli, P.: A principal component noise filter for high spectral resolution infrared measurements, J. Geophys. Res., 109 (D23), D23102, 2004.

Aumann, H. H., Chahine, M.T., Gautier, C., Goldberg, M., Kalnay, E., McMillin, L., Revercomb, H., Rosenkranz, P.W., Smith, W.L., Staelin, D., Strow, L., and Susskind, J.: AIRS/AMSU/HSB on the aqua mission: Design, science objectives, data products, and processing systems, IEEE Trans. Geosci. Remote Sens., 41, 253-264, doi:10.1109/TGRS.2002.808356, 2003.

Barthlott, S., Schneider, M., Hase, F., Wiegele, A., Christner, E., González, Y., Blumenstock, T., Dohe, S., García, O. E., Sepúlveda, E., Strong, K., Mendonca, J., Weaver, D., Palm, M., Deutscher, N. M., Warneke, T., Notholt, J., Lejeune, B., Mahieu, E., Jones, N., Griffith, D. W. T., Velazco, V. A., Smale, D., Robinson, J., Kivi, R., Heikkinen, P., and Raffalski, U.: Using XCO$_2$ retrievals for assessing the long-term consistency of NDACC/FTIR data sets, Atmos. Meas. Tech., 8, 1555-1573, doi:10.5194/amt-8-1555-2015, 2015.

Barthlott, S., Schneider, M., Hase, F., Blumenstock, T., Kiel, M., Dubravica, D., García, O. E., Sepúlveda, E., Mengistu Tsidu, G., Takele Kenea, S., Grutter, M., Plaza, E. F., Stremme, W., Strong, K., Weaver, D., Palm, M., Warneke, T., Notholt, J., Mahieu, E., Servais, C., Jones, N., Griffith, D. W. T., Smale, D., and Robinson, J.: Tropospheric water vapour isotopologue data (H$_2^{16}$O, H$_2^{18}$O and HD^{16}O) as obtained from NDACC/FTIR solar absorption spectra, Earth Syst. Sci. Data Discuss., doi:10.5194/essd-2016-9, 2016.

Batchelor, R.L., Strong, K., Lindenmaier, R., Mittermeier, R.L., Fast, H., Drummond, J.R., and Fogal, P.F.: A new Bruker IFS 125HR FTIR spectrometer for the Polar Environment Atmospheric Research Laboratory at Eureka, Canada - measurements and comparison with the existing Bomem DA8 spectrometer. J. Atmos. Oceanic Tech., 26 (7), 1328-1340, 2009.

Batchelor, R., Kolonjari, F., Lindenmaier, R., Mittermeier, R., Daffer, W., Fast, H, Manney, G., Strong, K., and Walker, K.: Four Fourier transform spectrometers and the Arctic polar vortex: instrument intercomparison and ACE-FTS validation at Eureka, Atmos. Meas. Tech., 5, 51-66, 2010.

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., and Mendrok, J.: A multi-instrument comparison of integrated water vapour measurements at a high latitude site. Atmos. Meas. Tech., 12, 10925-10943, 2012.

Clough, S.A., Shephard, M.W., Mlawer, E.J., Delamere, J.S., Iacono, M.J., Cady-Pereira, K., Boukabara, S., Brown, P.D.: Atmospheric radiative transfer modeling: a summary of the AER codes. J. Quant. Spectrosc. Radiat. Transfer, 91, 233-244, 2005.

Conway, T. J., Lang, P. M., and Masarie, K.A.: Atmospheric carbon dioxide dry air mole fractions from the NOAA ESRL Carbon Cycle Cooperative Global Air Sampling Network, 1968–2010, cited 2011.

Cox, C.J., Turner, D.D., Rowe, P.M., Walden, V.P.: Cloud microphysical properties retrieved from downwelling infrared radiance measurements made at Eureka, Nunavut, Canada (2006-2009, J. Appl. Meteor., Climatol., 53, 772-789, 2014.

Dessler, A. E., Zhang, Z., and Yang, P.: Water vapor climate feedback inferred from climate fluctuations, 2003-2008. Geophys. Res. Lett., 35, L20704, 2008.

Dirksen, R.J., Dommer, M., Immler, F.J., Hurst, D.F., Kivi, R., and Vömel, H.: Reference quality upper-air measurements: GRUAN data processing for the Vaisala RS92 radiosonde. Atmos. Meas. Tech., 7, 4463-4490, 2014.

Feltz, W., Smith, W. L., Howell, H. B., Knuteson, R. O., Woolf, H., and Revercomb, H.: Near-continuous profiling of temperature, moisture, and atmospheric stability using the Atmospheric Emitted Radiance Interferometer (AERI). J. Appl. Meteorol., 42(5), 584–597, 2003.

Fogal, P., LeBlanc, L.M., and Drummond, J.: The Polar Environment Atmospheric Research Laboratory (PEARL): Sounding the Atmosphere at 80° North. Arctic, 66 (3), 377-386, 2013.

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

Holben, B., Eck, T., Slutsker, I., Tanre, D., Buis, J., Setzer, A., Vermote, E., Reagan, J.A., Kaufman, Y.J., Nakajima, T., Lavenu, F., Jankowiak, I., and Smirnov, A.: AERONET - a federated instrument network and data archive for aerosol characterization. Remote Sens. Environ., 66, 1-16, 1998.

Hwang, Y.-T., and Frierson, D.M.W.: Increasing atmospheric poleward energy transport with global warming. Geophys. Res. Lett., 37, 24, L24807, 2010.

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, 2010.

Knuteson, R.O., Revercomb, H.E., Best, F.A., Ciganovich, N.C., Dedecker, R.G., Dirkx, T.P., Ellington, S.C., Feltz, W.F., Garcia, R.K., Howell, H.B., Smith, W.L., Short, J.F., and Tobin, D.C.: Atmospheric Emitted Radiance Interferometer. Part I: Instrument Design. J. Atmos. Oceanic Tech., 21, 1763-1789, 2004a.

Knuteson, RawlinsR.O., Revercomb, H.E., Best, F.A., Ciganovich, N.C., Dedecker, R.G., Dirkx, T.P., Ellington, S.C., Feltz, W.F., Garcia, R.K., Howell, H.B., Smith, W.L., Short, J.F., and Tobin, D.C.: Atmospheric Emitted Radiance Interferometer. Part II: Instrument Performance. J. Atmos. Oceanic Tech., 21, 1777-1776, 2004b.

Kurylo, M. J.: Network for the detection of stratospheric change (NDSC). Proceedings of SPIE. The International Society for Optical Engineering, Remote Sensing of Atmospheric Chemistry, 1491, 168-174, 1991.

Lesins, G., Duck, T. J., and Drummond, J. R.: Climate Trends at Eureka in the Canadian High Arctic, Atmos.-Ocean, 48 (2), 59–80, 2010.

Liljegren, J. C., and Lesht, B.M.: Measurements of integrated water vapor and cloud liquid water from microwave radiometers at the DOE ARM cloud and radiation testbed in the U.S. Southern Great Plains, paper presented at International Geoscience and Remote Sensing Symposium, IEEE, Lincoln, Nebr., doi:10.1109/IGARSS.1996.516767, 1996.

Lindenmaier, R., Strong, K., Batchelor, R.L., Bernath, P., Chabrillat, S.H., Chipperfield, M., Daffer, W.H., Drummond, J.R., Feng, W., Jonsson, A.I., Kolonjari, F., Manney, G.L., McLinden, C.A., Menard, R., and Walker, K.A.: A study of the Arctic NOy budget above Eureka, Canada. J. Geophys. Res., 116, D23302, 2011.

Lindenmaier, R., Strong, K., Batchelor, R.L., Chipperfield, M.P., Daffer, W.H., Drummond, J.R., Duck, T.J., Fast, H., Feng, W., Fogal, P.F., Kolonjari, F., Manney, G.L., Manson, A., Meek, C., Mittermaier, R.L., Nott, G.J., Perro, C., and Walker, K.A.: Unusually low O_3, HCl, and HNO_3 column measurements at Eureka, Canada during spring 2011. Atmos. Chem. Phys., 12, 3821-3835, 2012.

Mariani, Z., Strong, K., Wolff, M., Rowe, P., Walden, V., Fogal, P.F., Duck, T., Lesins, G., Turner, D.S., Cox, C., Eloranta, E., Drummond, J.R., Roy, C., Turner, D.D., Hudak, D., and Lindenmaier, I.A.: Infrared measurements in the Arctic using two Atmospheric Emitted Radiance Interferometers. Atmos. Meas. Tech., 5, 329-344, 2012.

Mariani, Z., Strong, K., Palm, M., Lindenmaier, R., Adams, C., Zhao, X., Savastiouk, V., McElroy, C.T., Goutail, F., Drummond, J.R.: Year-round retrievals of trace gases in the Arctic using the Extended-range Atmospheric Emitted Radiance Interferometer. Atmos. Meas. Tech., 6, 1549-1565, 2013.

McClatchey, R. A., Fenn, R. W., Selby, J. E. A., Volz, F. E., and Garing, J. S.: Optical properties of the atmosphere. 3rd ed. Tech. Rep. AFCRL-72-0497, Air Force Geophysical Laboratories, 108 pp, 1972.

Miloshevich, L., Paukkunen, A., and Oltmans, S.J.: Development and Validation of a time-lag correction for Vaisala radiosonde humidity measurements, J. Atmos. Oceanic Tech., 21, 1305-1327, 2004.

Miloshevich, L., Vömel, H., Whiteman, D., and Leblanc, T.: Accuracy assessment and correction of Vaisala RS92 radiosonde water vapor measurements. J. Geophys. Res., 114, D11305, 2009.

5   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, 2010.

Rodgers, C., and Connor, B.: Intercomparison of remote sounding instruments. J. Geophys. Res., 108, 4116, 2003.

Rowe, P. M., Miloshevich, L. M., Turner, D. S., and Walden, V. P.: Dry Bias in Vaisala RS90 Radiosonde Humidity Profiles over Antarctica. J. Atmos. Oceanic Tech., 25 (9), 1529–1541, 2008.

10   Rowe, P.M., Neshyba, S.P., and Walden, V.P.: Responsivity-based criterion for accurate calibration of FTIR emission spectra: theoretical development and bandwidth estimation, Opt. Express, 19, 5451-5463, 2011.

Schneider, M., Romero, P., Hase, F., Blumenstock, R., 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, 2010.

15   Schneider, M., Gonzalez, Y., Barthlott, S., Hase, F., Yoshimura, K., Garcia, O.E., Sepulveda, E., Gisi, M., Kohlhepp, R., Dohe, S., Blumenstock, T., Strong, K., Weaver, D., Palm, M., Deutscher, N., Warneke, T., Notholt, J., Lejeune, B., Demoulin, P., Jones, N., Griffith, D.W.T., Smale, D., Robinson, J., Sherlock, V., and Johnston, P.V.: Ground based remote sensing of tropospheric water vapour isotopologues within the project MUSICA. Atmos. Meas. Tech., 5, 3007-3027, 2012.

Schneider, M., González, Y., Dyroff, C., Christner, E., Wiegele, A., Barthlott, S., García, O. E., Sepúlveda, E., Hase,
20   F., Andrey, J., Blumenstock, T., Guirado, C., Ramos, R., and Rodríguez, S.: Empirical validation and proof of added value of MUSICA's tropospheric δD remote sensing products. Atmos. Meas. Tech., 8, 483-503, 2015.

Schneider, M., Wiegele, A., Barthlott, S., González, Y., Christner, E., Dyroff, C., García, O. E., Hase, F., Blumenstock, T., Sepúlveda, E., Mengistu Tsidu, G., Takele Kenea, S., Rodríguez, S., and Andrey, J.: Accomplishments of the MUSICA project to provide accurate, long-term, global and high-resolution observations of tropospheric $\{H_2O, \delta D\}$ pairs – a review, Atmos.
25   Meas. Tech., 9, 2845-2875, doi:10.5194/amt-9-2845-2016, 2016.

Sheese, P.E., Walker, K.A., Boone, C.D., Bernath, P.F., Froidevaux, L., Funke, B., Raspollini, P., von Clarmann, T.: ACE-FTS ozone, water vapour, nitrous oxide, nitric acid, and carbon monoxide profile comparisons with MIPAS and MLS. J. Quant. Spectrosc. Radiat. Transfer, 2016.

Shindell, D.T.: Climate and ozone response to increased stratospheric water vapor. Geophys. Res. Lett., 28, 1551-1554, 2001.

Sioris, C.E., Zou, J., McElroy, C.T., McLinden, C.A., and Vömel, H.: High vertical resolution water vapour profiles in the upper troposphere and lower stratosphere retrieved from MAESTRO solar occultation spectra. Adv. Space Res., 46, 642-650, 2010.

5   Stevens, B. and Bony, S.: What are climate models missing? Science, 340, 1053-1054, 2013.

Trenberth, B.J., Wetherald, R.T., Stenchikov, G.L., Robock, A.: Global cooling after the eruption of Mount Pinatubo: a test of climate feedback by water vapor. Science, 296, 5568, 727-730, 2002.

Tobin, D. C., et al., Downwelling spectral radiance observations at the SHEBA ice station: Water vapor continuum measurements from 17 to 26μm, J. Geophys. Res., 104 (D2), 2081–2092, 1999.

10   Trenberth, K.E., Dai, A., van der Schrier, G., Jones, D.P., Barichivich, J., Briffa, K.R., and Sheffield, J.: Global warming and changes in drought. Nature Climate Change, 4, 17-22, 2013.

Trenberth, K. E.: Challenges for Observing and Modeling the Global Water Cycle, in Remote Sensing of the Terrestrial Water Cycle (eds V. Lakshmi, D. Alsdorf, M. Anderson, S. Biancamaria, M. Cosh, J. Entin, G. Huffman, W. Kustas, P. van Oevelen, T. Painter, J. Parajka, M. Rodell and C. Rüdiger), John Wiley and Sons, Inc, Hoboken, NJ, 2014.

15   Westwater, E.R., Han, Y., Shupe, M.D., Matrosov, S.Y.: Analysis of integrated cloud liquid and precipitable water vapor retrievals from microwave radiometers during the Surface Heat Budget of the Arctic Ocean project, J. Geophys. Res., 106, 32019-32030, 2001.

Viatte, C., Strong, Hannigan, K., J., Nussbaumer, E., Emmons, L., Conway, S., Paton-Walsh, C., Hartley, J., Benmergui, J., and Lin, J.: Identifying fire plumes in the Arctic with tropospheric FTIR measurements and transport models, Atmos. Chem. Phys, 15, 2227-2246, 2015.

Vogelmann, H., Sussman, R., Trickl, T., and Borsdorff, T.: Intercomparison of atmospheric water vapor soundings from the differential absorption lidar (DIAL) and the solar FTIR system on Mt. Zugspitze. Atmos. Meas. Tech., 4, 835-841, 2011.

Vömel, H., Selkirk, H., Miloshevich, L., Valverde-Canossa, J., Valdes, J., Kyrö, E., Kivi, R., Stolz, W., Peng, G. and Diaz, J.: Radiation dry bias of the Vaisala RS92 humidity sensor. J. Atmos. Oceanic Tech., 24, 953-963, 2007.

Table 1: Eureka water vapour datasets.

Location	Instrument	Period	Measurement duration	Measurement frequency	Number	Measurement types available	
						Total column	Profile
Ridge Lab	125HR MUSICA extended	Aug 2006 – Oct 2014	5 minutes	1-12 per day between late February and mid-October	3364 (2713 with SZA>85°)	✓	✓
	125HR MUSICA standard	August 2006 – August 2014	5 minutes	1-12 per day between late March and mid-September	1889	✓	✓
	Sun photometer	March 2007 – Sept. 2015	1 minute	every 15 minutes	89,867	✓	
	E-AERI	Oct 2008 - Sept 2009	7 minutes	continuously	8321	✓	
0PAL	Sun photometer	April 2007 - Sept. 2013	1 minute	every 15 minutes	63,010	✓	
	Microwave radiometer	Aug 2006 - June 2013	30 minute averages	continuously	105,058	✓	
	P-AERI	March 2006 - April 2009	7 minute	continuously	267,957	✓	
	E-AERI	Feb 2011 - July 2013	Hourly averages	continuously	1353	✓	
Eureka Weather Station	Radiosonde standard	Aug. 2006 – Dec. 2015	2 hours	twice per day	6872	✓	✓
	Radiosonde GRUAN	Sept 2008 - March 2013	2 hours	twice per day, but with gaps	2384	✓	✓

Table 2: Estimated accuracies of water vapour column retrievals for Eureka instruments.

	Instrument	Accuracy	Reference
Total column	MUSICA extended	1.80% (σ = 0.61%)	Calculated from error of each datapoint
	MUSICA standard	~1%	Schneider et al. (2012)
	Sun photometer	~10%	Alexandrov et al. (2009)
	Microwave radiometer	~20%	Westwater et al. (2001)
	Radiosonde	~15%	Miloshevich et al. (2009)
	GRUAN	5.70% (σ = 1.6%)	Calculated
	E-AERI	3-11% varies monthly	Rowe et al. (2008)
	P-AERI	3-11% varies monthly	Rowe et al. (2008)
Profile	MUSICA	~10%	Schneider et al. (2012)
	Radiosonde	~15%	Miloshevich et al. (2009)

Table 3: Summary of total column comparison of 125HR with different SZA limits, quality control (QC), and coincidences (pairs vs. all) with co-located Eureka instruments. Standard error of the mean is noted in brackets next to the standard deviation. GRUAN and RS total columns are the result of integrating the smoothed profiles from the Ridge Lab altitude (610 m).

	125HR standard MUSICA (78.5° SZA limit)	125HR extended MUSICA (85° SZA limit; QC)	125HR extended MUSICA (no SZA limit or QC)	125HR extended MUSICA (85° SZA limit; QC, all coincidences)
(a) **E-AERI (RL)**	N = 168 R = 0.99 m = 0.93 0.45 ± 0.43 mm (0.03) 5.61 ± 4.64% (0.36)	N = 181 R = 0.99 m = 0.93 0.41 ± 0.45 mm (0.03) 5.10 ± 5.06% (0.38)	N = 182 R = 0.99 m = 0.93 0.41 ± 0.44 mm (0.03) 5.14 ± 4.90% (0.36)	N = 4577 R = 0.99 m = 0.93 0.42 ± 0.51 mm (0.01) 5.18 ± 6.16% (0.09)
(b) **SPM (RL)**	N = 1273 R = 0.99 m = 0.84 1.05 ± 0.65 mm (0.02) 15.07 ± 6.51% (0.18)	N = 1297 R = 0.99 m = 0.84 1.04 ± 0.65 mm (0.02) 15.11 ± 6.49% (0.18)	N = 1374 R = 0.99 m = 0.84 1.05 ± 0.69 mm (0.02) 15.18 ± 6.97% (0.19)	N = 49200 R = 0.98 m = 0.85 1.05 ± 0.77 mm (0.00) 15.02 ± 8.97% (0.04)
(c) **RS (EWS)**	N = 54 R = 0.94 m = 0.95 0.78 ± 0.93 mm (0.13) 12.08 ± 12.80% (1.74)	N = 64 R = 0.96 m = 0.91 0.70 ± 0.95 mm (0.12) 11.37 ± 13.17% (1.65)	N = 65 R = 0.96 m = 0.92 0.61 ± 0.88 mm (0.11) 9.56 ± 14.63% (1.81)	N = 186 R = 0.95 m = 0.90 0.83 ± 0.95 mm (0.07) 12.16 ± 12.64% (0.93)
(d) **GRUAN (EWS)**	N = 6 R = 0.99 m = 0.97 0.42 ± 0.49 mm (0.20) 6.08 ± 5.68% (2.32)	N = 10 R = 0.99 m = 0.94 0.21 ± 0.42 mm (0.13) 1.48 ± 9.46% (2.99)	N = 10 R = 0.99 m = 0.95 0.17 ± 0.48 mm (0.15) 1.23 ± 10.02% (3.17)	N = 19 R = 0.99 m = 0.99 0.18 ± 0.42 mm (0.10) 2.12 ± 7.86 % (1.80)

Table 4: Results of intercomparison of Eureka PWV measurements, including number of coincidences (N), correlation coefficient (R), slope of the correlation line (m), mean difference ($X - Y$) and standard deviation of the difference in mm, mean percent difference ($2 * (X-Y)/(X+Y) * 100\%$) and standard deviation of the percent difference. The standard error of the mean (SE) is noted in brackets alongside the standard deviations. The x-axis instruments along the top are "X" in the difference equations; the y-axis instruments along the top are "Y" in the difference equations. Comparisons between 125HR and radiosonde and GRUAN datasets involve smoothing the radiosonde profile using MUSICA averaging kernels, and interpolation to the 125HR retrieval grid prior to integrating the total column. A high resolution version of this table can be found in the Supplementary Materials as Table S1.

	125HR (RL)	SPM (0PAL)	E-AERI (0PAL)	P-AERI (0PAL)	MWR (0PAL)	GRUAN (EWS)	RS (EWS)	E-AERI (RL)	SPM (RL)	
125HR (RL)		N = 917 R = 0.98 m = 0.89 1.57 ± 0.72 mm (0.02) 18.46 ± 8.42 % (0.28)	N = 27 R = 0.99 m = 0.80 1.08 ± 1.41 mm (0.27) 11.75 ± 11.42 % (2.20)	N = 247 R = 0.99 m = 0.83 1.01 ± 0.91 mm (0.06) 16.36 ± 8.03 % (0.51)	N = 1920 R = 0.99 m = 0.92 0.64 ± 0.63 mm (0.01) 11.60 ± 20.60 % (0.47)	N = 9 R = 0.99 m = 1.03 -0.31 ± 0.42 mm (0.14) -4.92 ± 1.64 % (1.56)	N = 63 R = 0.96 m = 0.99 -0.69 ± 0.89 mm (0.11) -11.65 ± 12.36 % (1.56)	N = 176 R = 0.99 m = 1.06 -0.42 ± 0.45 mm (0.03) -5.14 ± 5.09 % (0.38)	N = 1274 R = 0.99 m = 1.16 -1.05 ± 0.64 mm (0.02) -15.14 ± 6.30 % (0.18)	**125HR (RL)**
SPM (0PAL)	N = 954 R = 0.98 m = 1.07 -1.57 ± 0.73 mm (0.02) -18.50 ± 8.60 % (0.28)		N = 130 R = 0.98 m = 0.97 0.40 ± 0.65 mm (0.06) 3.35 ± 5.16 % (0.45)	N = 2662 R = 0.98 m = 0.90 0.54 ± 0.66 mm (0.01) 4.74 ± 4.96 % (0.10)	N = 9510 R = 0.99 m = 1.01 -0.67 ± 0.60 mm (0.01) -7.67 ± 7.34 % (0.08)	N = 265 R = 0.99 m = 0.93 -0.18 ± 0.67 mm (0.04) -0.55 ± 6.39 % (0.39)	N = 692 R = 0.99 m = 0.94 -0.10 ± 0.65 mm (0.02) -2.07 ± 6.67 % (0.25)	N = 776 R = 0.96 m = 1.01 -2.29 ± 0.69 mm (0.02) -24.89 ± 7.79 % (0.28)	N = 37,175 R = 0.97 m = 1.24 -2.91 ± 1.05 mm (0.01) -33.58 ± 9.51 % (0.05)	**SPM (0PAL)**
E-AERI (0PAL)	N = 24 R = 0.99 m = 1.20 -0.82 ± 1.28 mm (0.26) -10.28 ± 10.48 % (2.14)	N = 121 R = 0.98 m = 0.97 -0.39 ± 0.64 mm (0.06) -3.32 ± 5.18 % (0.47)		N = 0	N = 1191 R = 0.99 m = 1.02 -0.55 ± 0.51 mm (0.01) -41.66 ± 89.37 % (2.59)	N = 167 R = 1.00 m = 0.99 -0.02 ± 0.24 mm (0.02) -2.46 ± 5.93 % (0.46)	N = 196 R = 1.00 m = 1.01 -0.16 ± 0.22 mm (0.02) -6.70 ± 6.20 % (0.44)		N = 55 R = 0.90 m = 1.08 -3.94 ± 0.59 mm (0.08) -33.49 ± 4.79 % (0.65)	**E-AERI (0PAL)**
P-AERI (0PAL)	N = 263 R = 0.99 m = 1.18 -0.99 ± 0.86 mm (0.05) -16.34 ± 7.44 % (0.46)	N = 2804 R = 0.98 m = 1.07 -0.54 ± 0.66 mm (0.01) -4.76 ± 4.95 % (0.09)	N = 0		N = 7981 R = 0.99 m = 1.09 -0.03 ± 0.50 mm (0.01) 3.96 ± 15.16 % (0.17)	N = 113 R = 0.99 m = 0.94 -0.08 ± 0.24 mm (0.02) -4.89 ± 15.33 % (1.44)	N = 517 R = 1.00 m = 1.01 -0.16 ± 0.28 mm (0.01) -5.65 ± 6.62 % (0.29)	N = 2556 R = 0.99 m = 1.11 -0.27 ± 0.11 mm (0.00) -16.99 ± 5.35 % (0.11)	N = 2822 R = 0.97 m = 1.29 -2.83 ± 1.26 mm (0.02) -31.34 ± 8.58 % (0.16)	**P-AERI (0PAL)**
MWR (0PAL)	N = 1905 R = 0.99 m = 1.07 -0.64 ± 0.59 mm (0.01) -11.53 ± 20.07 % (0.46)	N = 8837 R = 0.99 m = 0.96 0.68 ± 0.60 mm (0.01) 7.75 ± 7.32 % (0.08)	N = 1191 R = 0.99 m = 0.96 0.55 ± 0.51 mm (0.01) 41.66 ± 89.37 % (2.59)	N = 6975 R = 0.99 m = 0.91 0.00 ± 0.49 mm (0.01) -4.66 ± 15.11 % (0.18)		N = 2057 R = 0.99 m = 0.93 0.46 ± 0.78 mm (0.02) 17.29 ± 52.57 % (1.16)	N = 4343 R = 0.99 m = 0.94 0.22 ± 0.59 mm (0.01) 5.19 ± 100.12 % (1.52)	N = 2443 R = 0.99 m = 1.15 -0.71 ± 0.71 mm (0.01) -19.94 ± 14.48 % (0.29)	N = 10,007 R = 0.98 m = 1.23 -1.93 ± 0.98 mm (0.01) -25.50 ± 9.03 % (0.09)	**MWR (0PAL)**
GRUAN (EWS)	N = 10 R = 0.99 m = 0.94 0.21 ± 0.42 mm (0.13) 1.48 ± 9.46 % (2.99)	N = 265 R = 0.99 m = 1.05 -0.18 ± 0.67 mm (0.04) 0.55 ± 6.39 % (0.39)	N = 166 R = 1.00 m = 1.04 -0.04 ± 0.27 mm (0.27) 2.03 ± 6.16 % (0.48)	N = 113 R = 0.99 m = 1.04 0.08 ± 0.24 mm (0.02) 4.89 ± 15.33 % (1.44)	N = 2055 R = 0.99 m = 1.09 -0.53 ± 0.68 mm (0.01) -17.77 ± 52.34 % (1.15)		N = 2377 R = 1.00 m = 1.05 -0.25 ± 0.25 mm (0.01) -4.57 ± 3.77 % (0.08)	N = 99 R = 0.98 m = 1.32 -0.27 ± 0.44 mm (0.04) -11.46 ± 14.91 % (1.50)	N = 250 R = 0.99 m = 1.36 -2.91 ± 1.47 mm (0.09) -33.65 ± 8.85 % (0.56)	**GRUAN (EWS)**
RS (EWS)	N = 64 R = 0.96 m = 0.91 0.70 ± 0.95 mm (0.12) 11.37 ± 13.17 % (1.65)	N = 692 R = 0.99 m = 1.03 0.10 ± 0.65 mm (0.02) 2.07 ± 6.67 % (0.25)	N = 196 R = 1.00 m = 0.99 0.16 ± 0.22 mm (0.02) 6.70 ± 6.20 % (0.44)	N = 517 R = 1.00 m = 0.99 0.16 ± 0.28 mm (0.01) 5.65 ± 6.62 % (0.29)	N = 4278 R = 0.99 m = 1.00 -0.22 ± 0.59 mm (0.01) -4.78 ± 100.01 % (1.53)	N = 2377 R = 1.00 m = 0.95 0.25 ± 0.25 mm (0.01) 4.57 ± 3.77 % (0.08)		N = 157 R = 0.99 m = 1.27 -0.66 ± 1.00 mm (0.08) -10.64 ± 11.37 % (0.91)	N = 1036 R = 0.99 m = 1.34 -2.41 ± 1.30 mm (0.04) -29.29 ± 8.81 % (0.27)	**RS (EWS)**
E-AERI (RL)	N = 181 R = 0.99 m = 0.93 0.41 ± 0.45 mm (0.03) 5.10 ± 5.06 % (0.38)	N = 786 R = 0.96 m = 0.91 2.29 ± 0.69 mm (0.02) 24.90 ± 7.77 % (0.28)		N = 2331 R = 0.99 m = 0.88 0.27 ± 0.10 mm (0.00) 17.09 ± 5.38 % (0.11)	N = 2569 R = 0.99 m = 0.86 0.72 ± 0.70 mm (0.01) 19.95 ± 14.48 % (0.28)	N = 99 R = 0.98 m = 0.73 0.27 ± 0.44 mm (0.04) 11.46 ± 14.91 % (1.50)	N = 157 R = 0.99 m = 0.78 0.66 ± 1.00 mm (0.08) 10.64 ± 11.37 % (0.91)		N = 889 R = 0.99 m = 1.12 -0.53 ± 0.48 mm (0.02) -6.69 ± 4.40 % (0.15)	**E-AERI (RL)**
SPM (RL)	N = 1297 R = 0.99 m = 0.84 1.04 ± 0.65 mm (0.02) 15.11 ± 6.49 % (0.18)	N = 37,306 R = 0.97 m = 0.76 2.91 ± 1.05 mm (0.01) 33.58 ± 9.51 % (0.05)	N = 56 R = 0.90 m = 0.75 3.95 ± 0.58 mm (0.08) 33.56 ± 4.77 % (0.64)	N = 2617 R = 0.97 m = 0.73 2.86 ± 1.25 mm (0.02) 31.31 ± 8.52 % (0.17)	N = 10,732 R = 0.98 m = 0.79 1.94 ± 0.98 mm (0.01) 25.42 ± 9.01 % (0.09)	N = 250 R = 0.99 m = 0.71 2.91 ± 1.47 mm (0.09) 33.65 ± 8.85 % (0.56)	N = 1036 R = 0.99 m = 0.72 2.41 ± 1.30 mm (0.04) 29.29 ± 8.81 % (0.27)	N = 867 R = 0.99 m = 0.88 0.54 ± 0.48 mm (0.02) 6.73 ± 4.41 % (0.15)		**SPM (RL)**
	125HR (RL)	**SPM (0PAL)**	**E-AERI (0PAL)**	**P-AERI (0PAL)**	**MWR (0PAL)**	**GRUAN (EWS)**	**RS (EWS)**	**E-AERI (RL)**	**SPM (RL)**	

Table 5: Summary of water vapour profile comparisons at representative altitudes using coincidence criteria of 2 hours. Comparisons are calculated as 125HR (instrument X from Eq. (2)) on the vertical column minus radiosondes (instrument Y) along the top.

		Radiosondes		
	Altitude	difference $\pm$ σ [ppmv]	difference $\pm$ σ [%]	N
125HR	3.0 km	208.4 $\pm$ 466.3	14.2 $\pm$ 25.5	84
	4.9 km	65.7 $\pm$ 227.5	8.9 $\pm$ 36.5	84
	6.4 km	21.3 $\pm$ 121.5	2.8 $\pm$ 39.4	84
	7.9 km	5.8 $\pm$ 48.6	-0.5 $\pm$ 38.0	84
	9.8 km	0.5 $\pm$ 9.4	38.8 $\pm$ 23.9	84
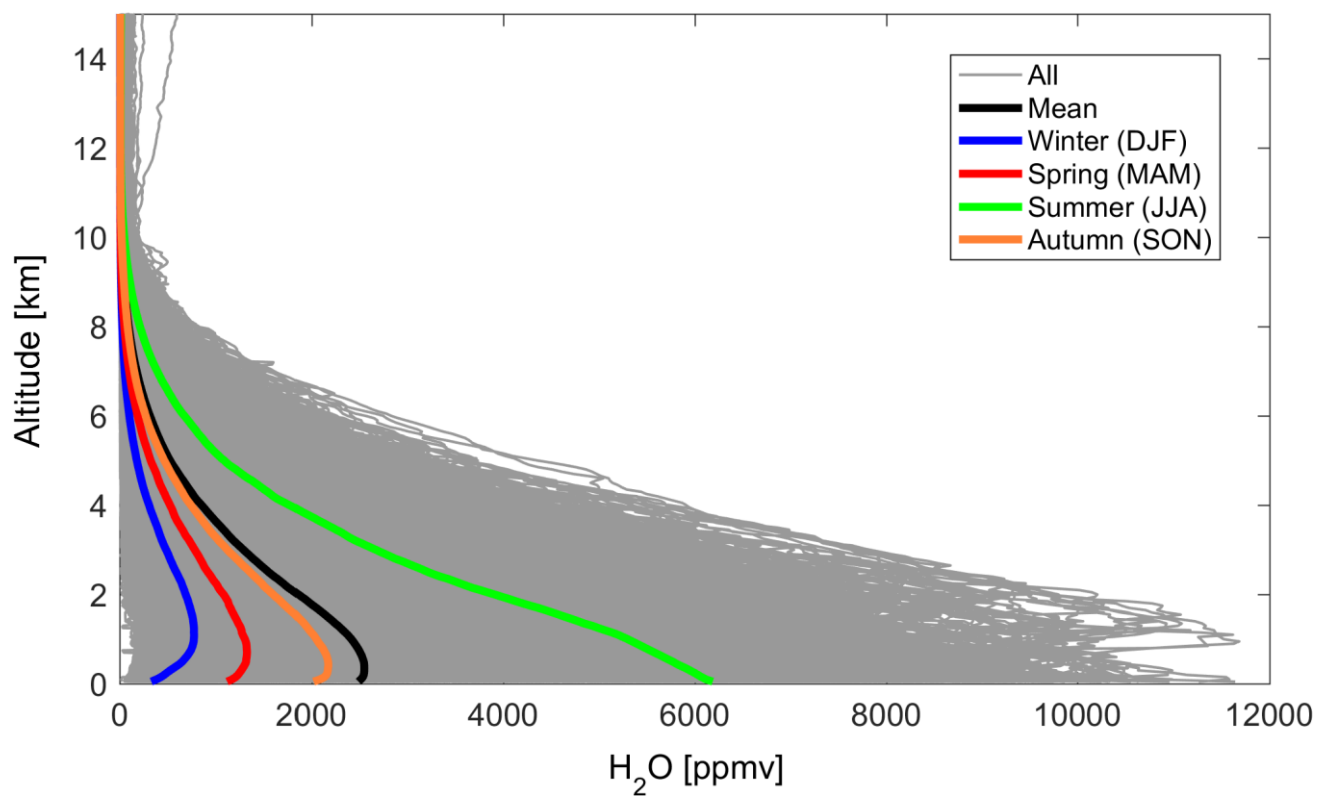
Figure 1: Eureka radiosonde water vapour profiles (Aug. 2006 to Dec. 2015).
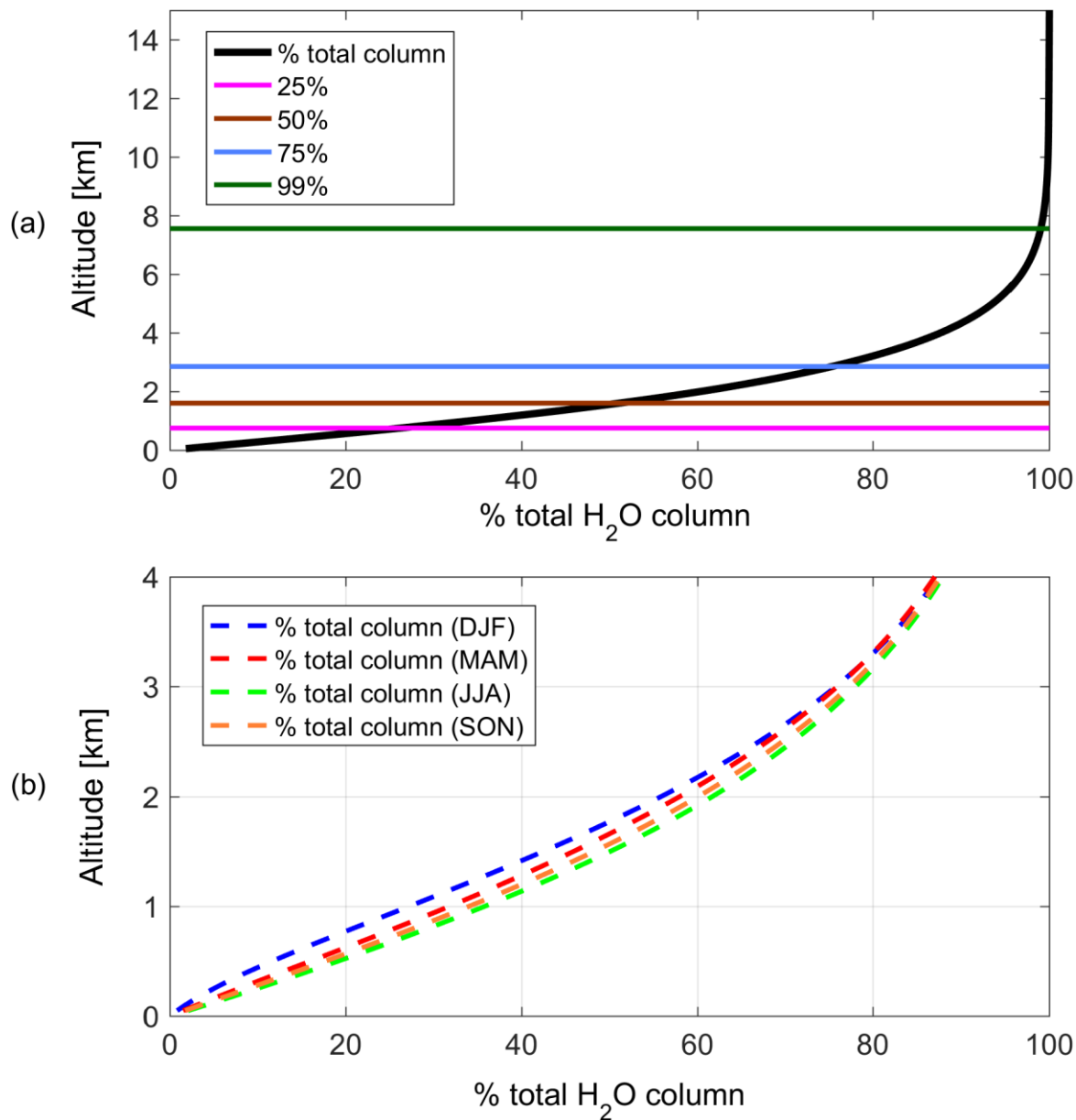
Figure 2: Distribution of water vapour above Eureka from radiosonde profiles (Aug. 2006 to Dec. 2015). The black line illustrates how much of the total column is beneath a given altitude. One-fifth of the water vapour total column is typically located beneath 610 m, the altitude of the PEARL Ridge Lab. The horizontal lines show the altitudes at which each quartile is passed. (a) altitudes up to 15 km. (b) the mean distribution of the water vapour total column up in the lower troposphere by season.
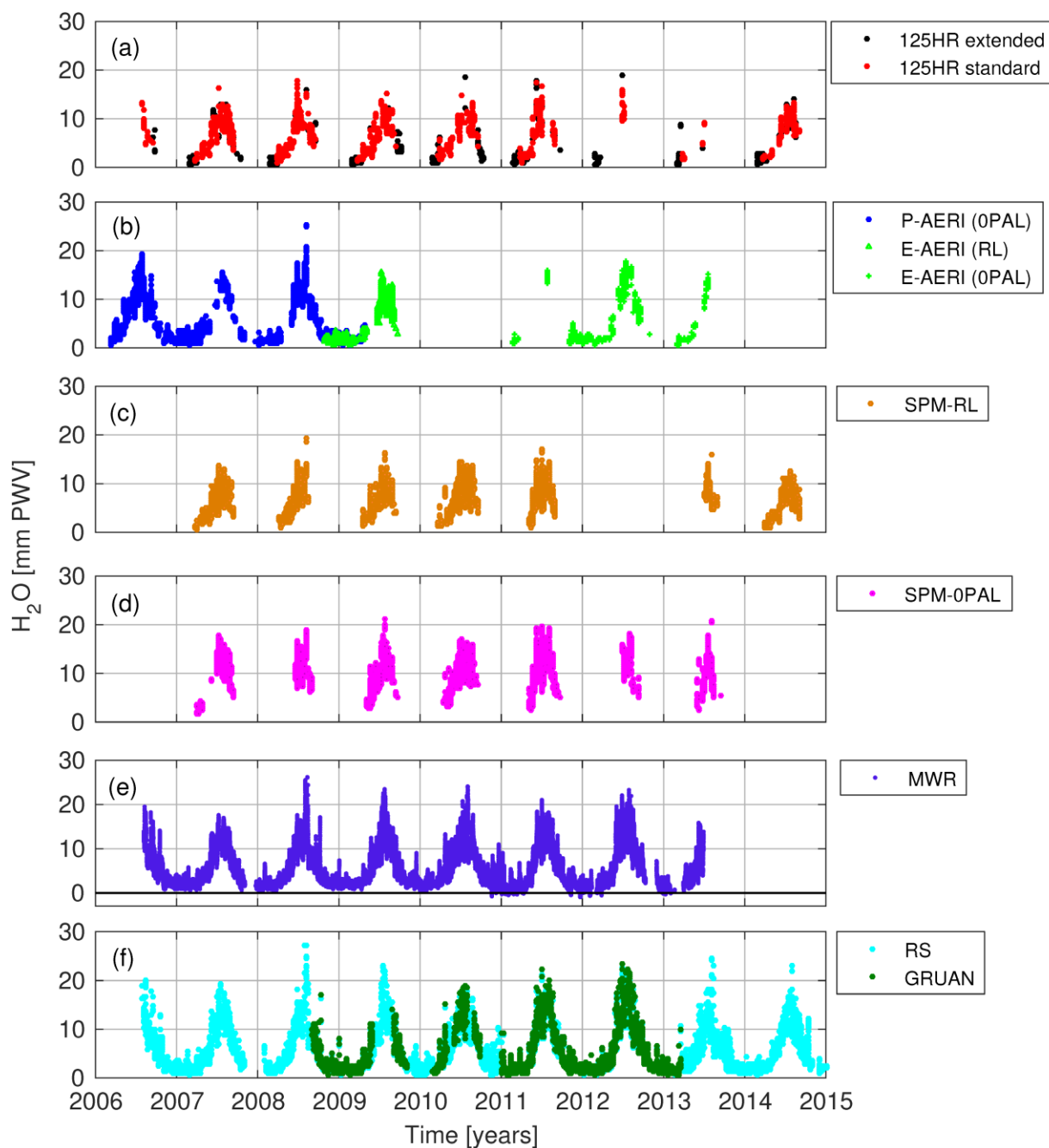
Figure 3: Total column precipitable water vapour (PWV) at Eureka. (a) 125HR water vapour. The extended timeseries has relaxed the MUSICA SZA quality control criterion of the standard MUSICA timeseries. (b) P-AERI and E-AERI datasets, with location of the E-AERI noted. (c) the Ridge Lab's SPM dataset. (d) 0PAL's SPM dataset. (e) MWR dataset. (f) Eureka Weather Station radiosonde dataset, showing the standard data product as well as the GRUAN product.
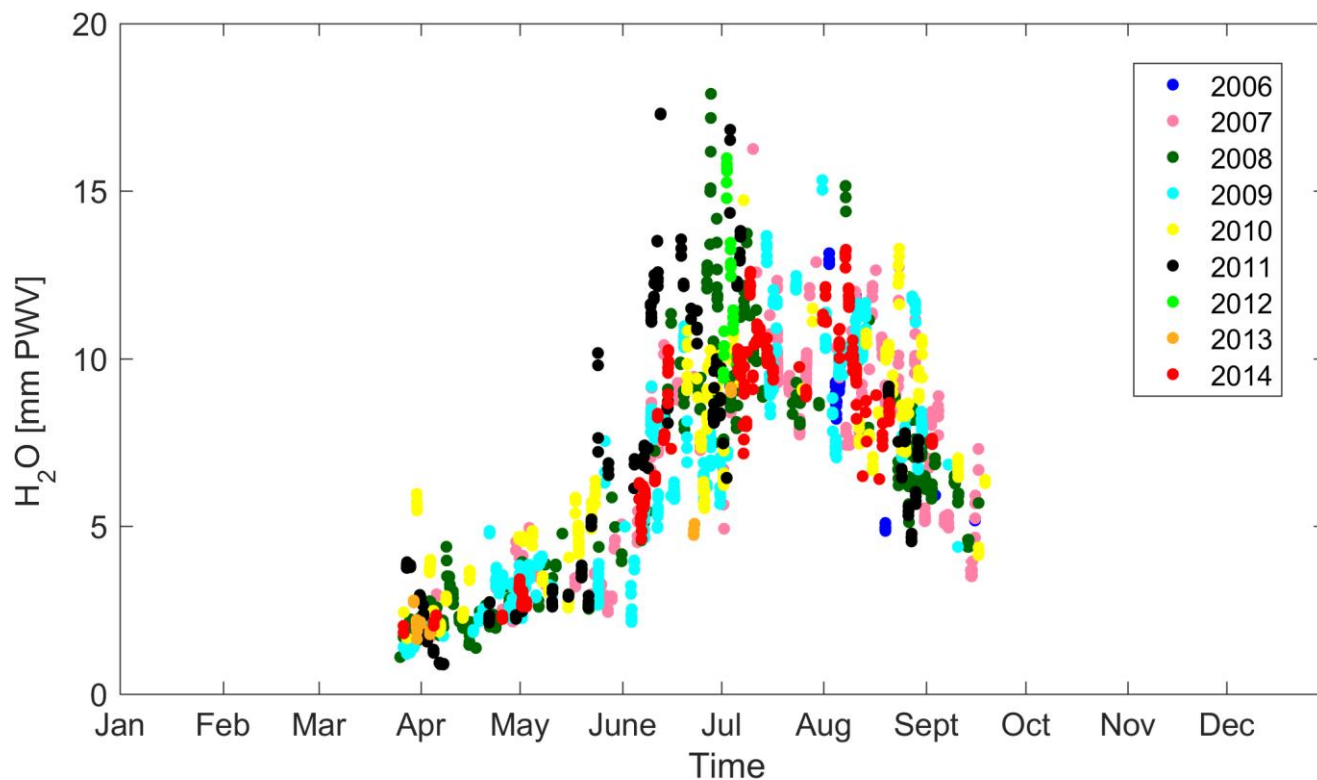
Figure 4: MUSICA water vapour dataset from the Eureka 125HR measurements. The lack of data between mid-September and late-March is caused by the MUSICA quality control filter's solar zenith angle (SZA) criterion (<78.5°) and polar night (mid-October until mid-February).
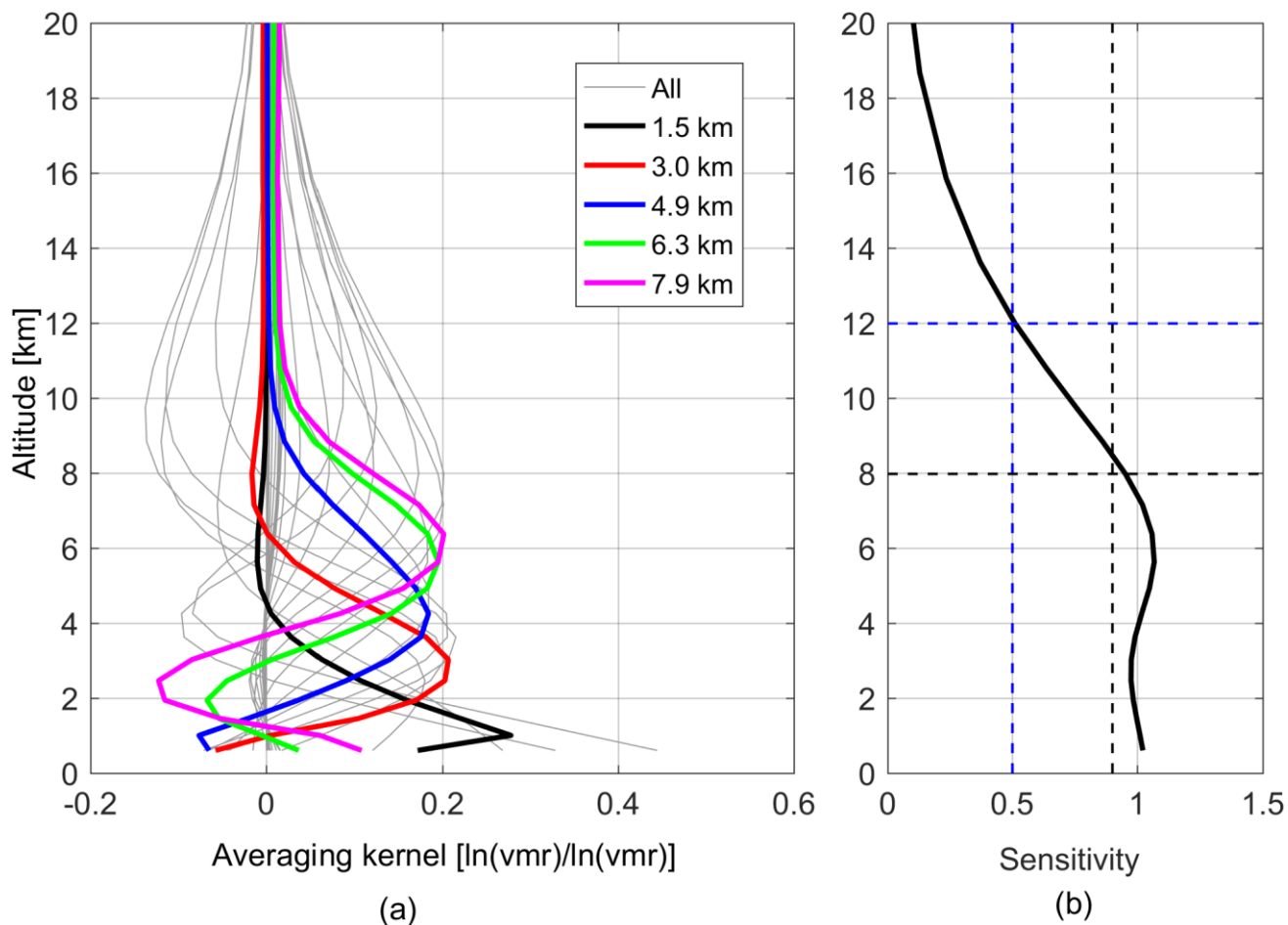
Figure 5: (a) example of a typical MUSICA averaging kernel for a Eureka 125HR measurement, taken on August 3, 2006. Five altitudes are highlighted in color to illustrate the measurement's ability to distinguish between different parts of the troposphere. (b) the retrieval's sensitivity, with two thresholds for measurement information noted by dashed lines. The highest altitude where sensitivity is above 0.9 (8.0 km) and 0.5 (12.0 km) is shown using blue and black dashed lines, respectively.
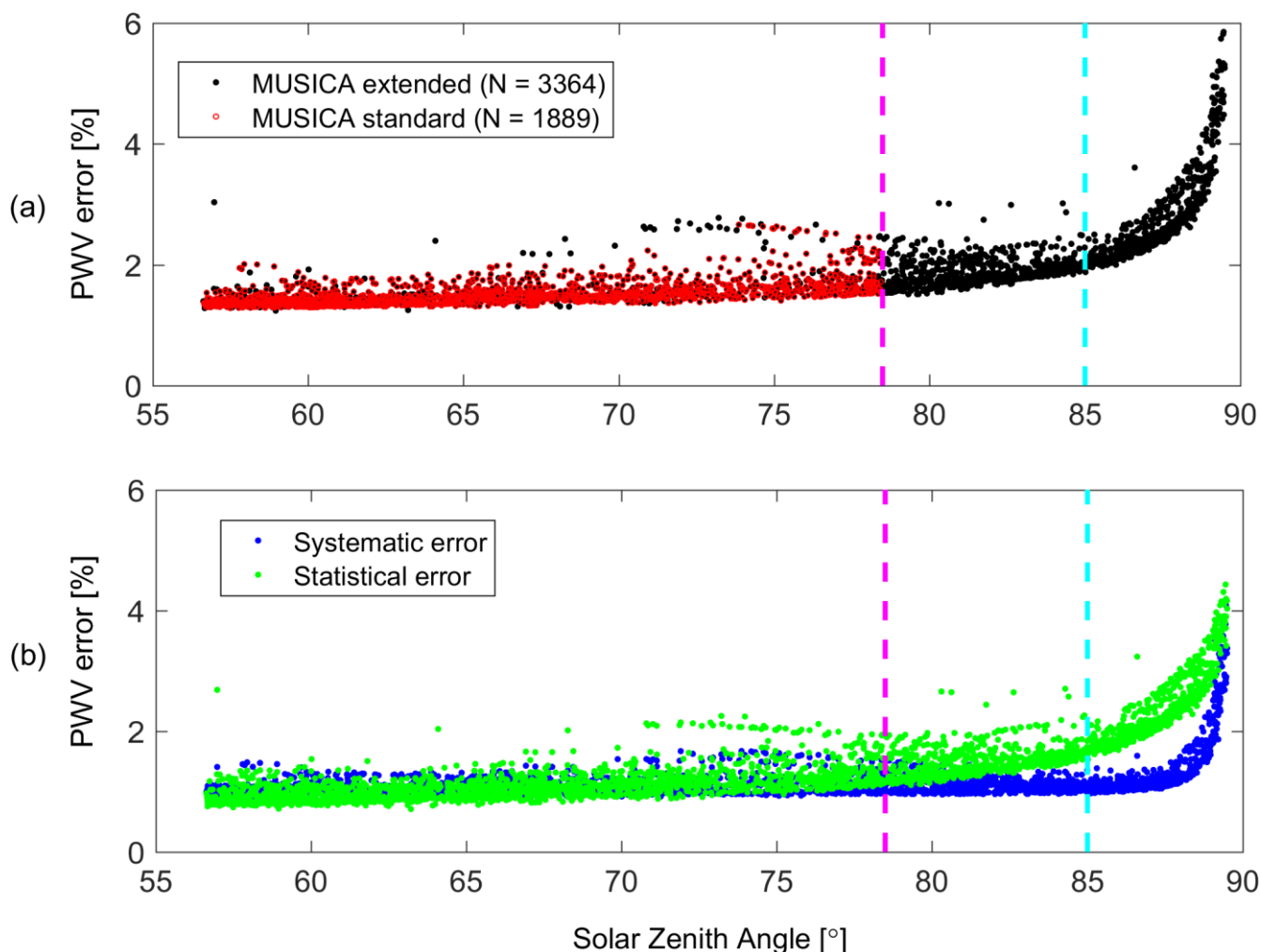
Figure 6: Eureka MUSICA water vapour total error vs. SZA. The dashed magenta line denotes the 78.5° SZA criterion which is part of the standard MUSICA quality control. The dashed cyan line denotes the 85° SZA threshold used in this study for water vapour total column comparisons. (a) the total error. (b) the statistical and systematic errors. The MUSICA extended dataset includes data usually filtered out by MUSICA's quality control criteria.
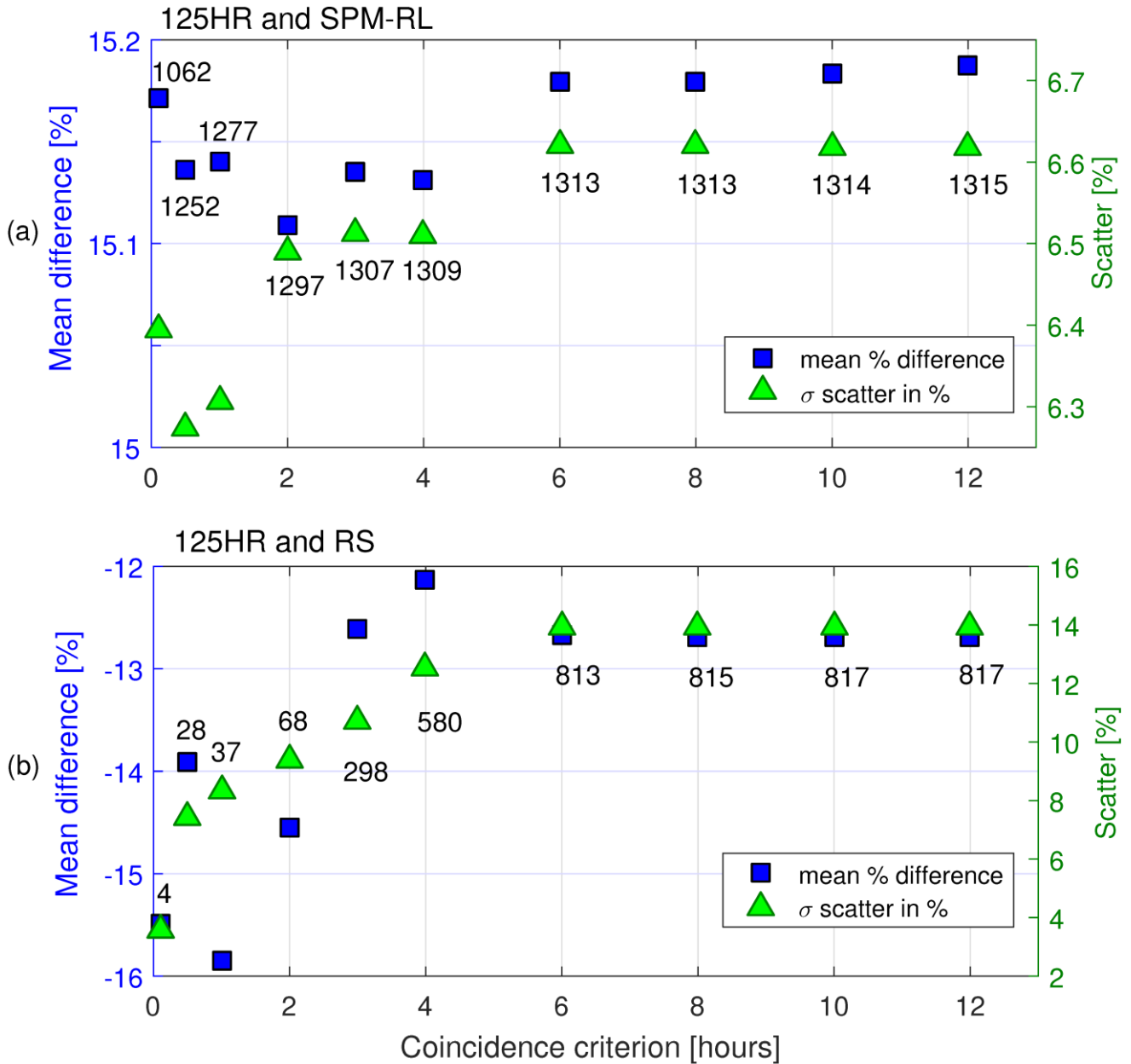
Figure 7: The variation in mean percent difference and scatter (standard deviation) between the water vapour total column of the 125HR and other instruments as a function of the temporal coincidence criterion. (a) 125HR compared to the SPM at the Ridge Lab (125HR – SPM); (b) 125HR and the (smoothed) radiosondes (125HR – RS). The resulting number of coincidences at each temporal coincidence criterion is labelled next to each data pair.
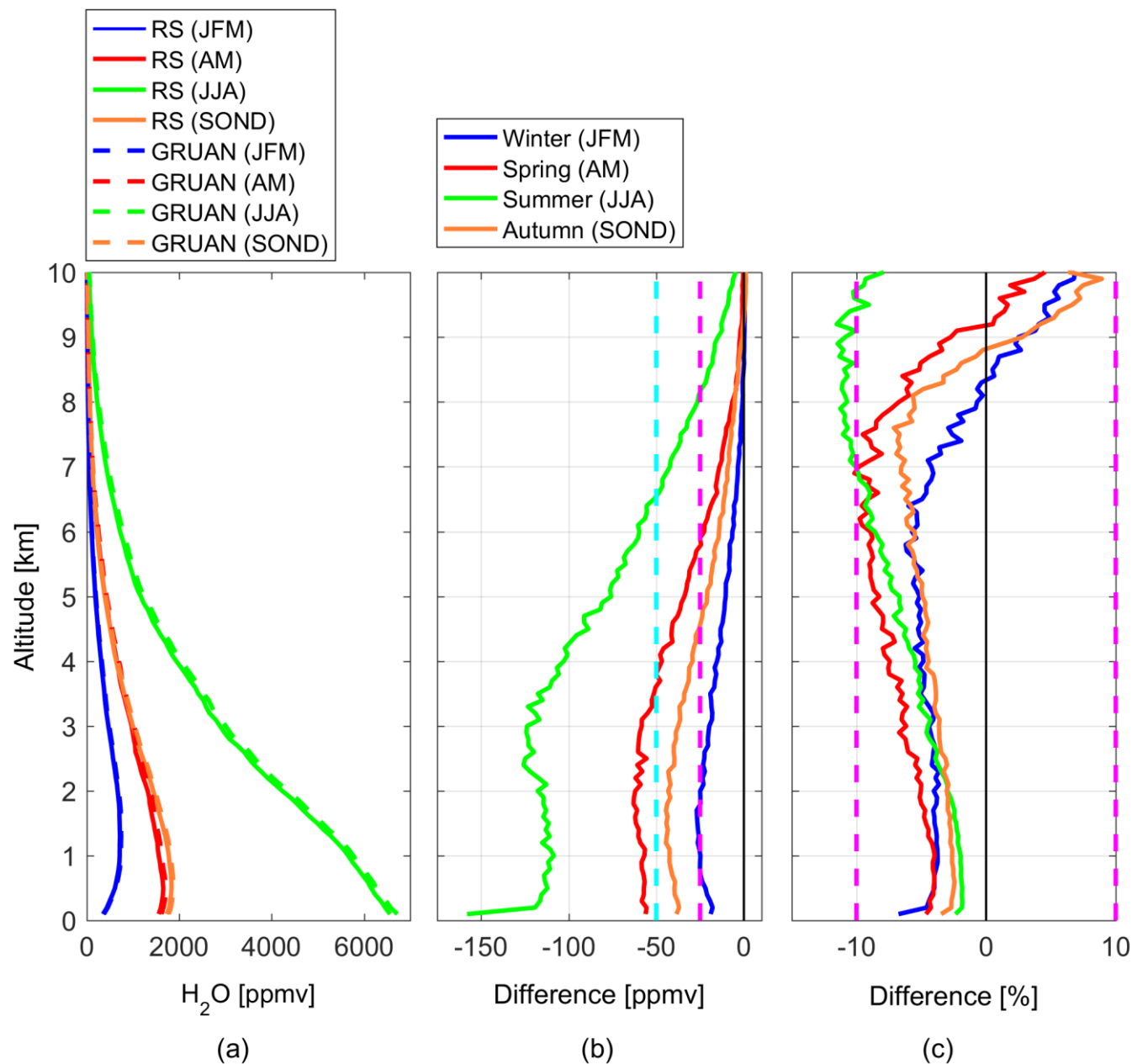
Figure 8: (a) Coincident radiosonde and GRUAN profiles by season. (b) Differences between the profiles. (c) Percent differences. (X = RS and Y = GRUAN in Eq. (1) and (2).)
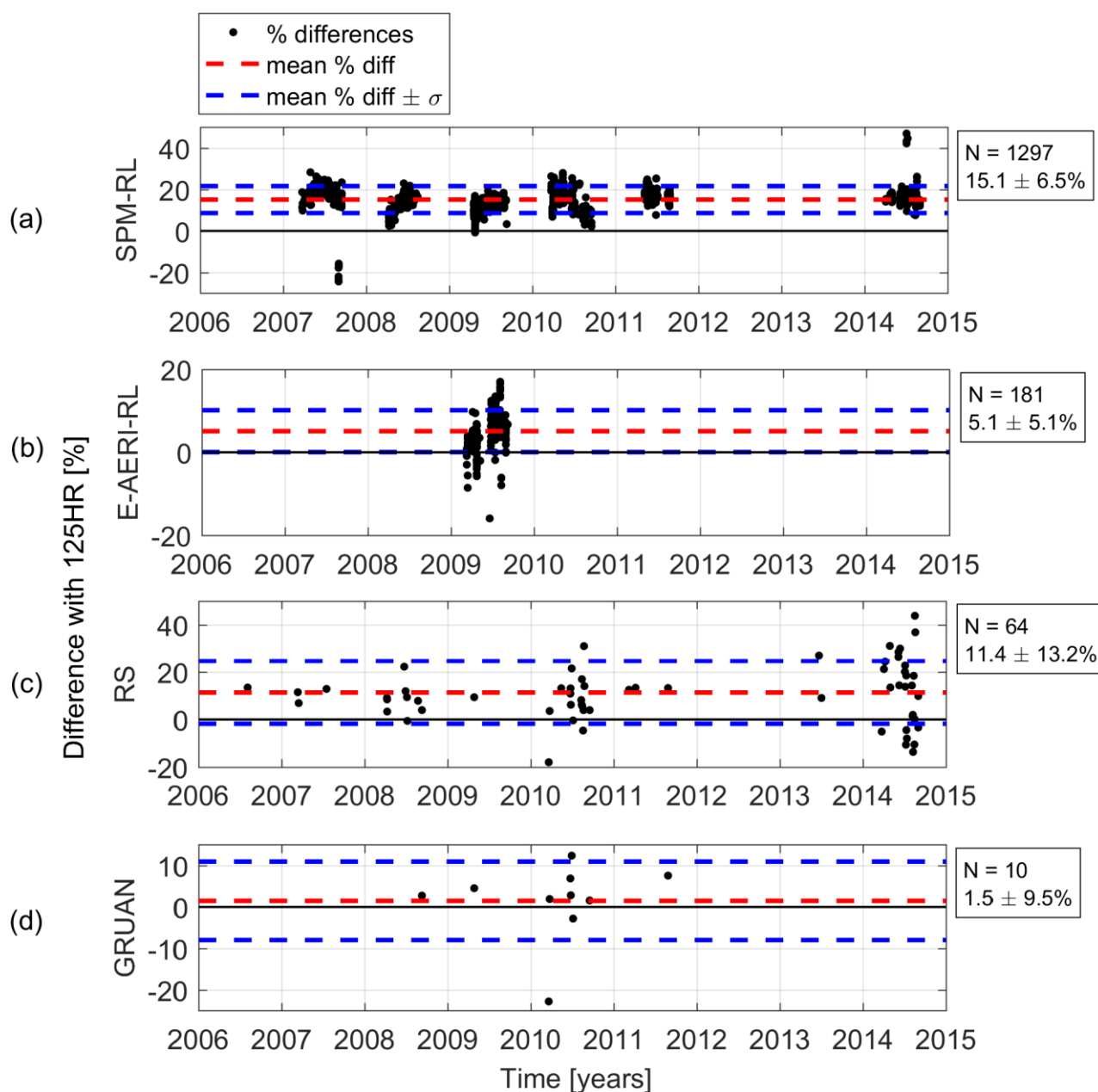
Figure 9: Comparisons of the 125HR water vapour with PEARL Ridge Lab and Eureka Weather Station instruments. (In Eq. 2, X = 125HR, Y = comparison instrument). The red lines denote the mean difference; while blue lines denote one standard deviation above and below the mean difference. The number of comparison pairs, as well as the mean percent difference and standard deviation are noted for each comparison.
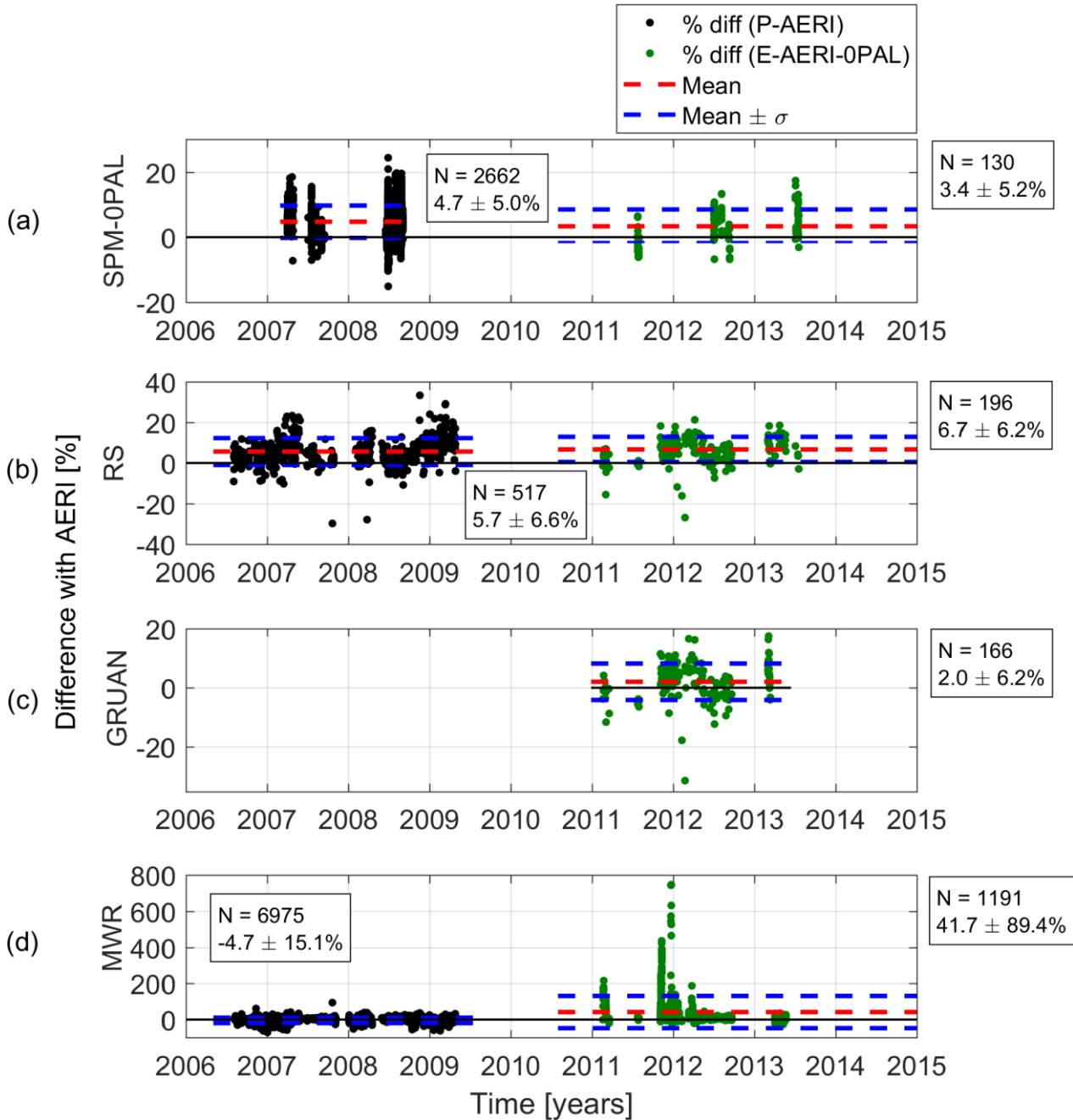
Figure 10: Same as Fig. 10 but for 0PAL-based AERI percent difference comparisons with 0PAL and EWS instruments. (Percent differences calculated using Eq. (2) where X = P-AERI or X= E-AERI-0PAL). The y-axis scale of (d) cuts off extreme outliers (which reach nearly to 1200%) to ensure readability.
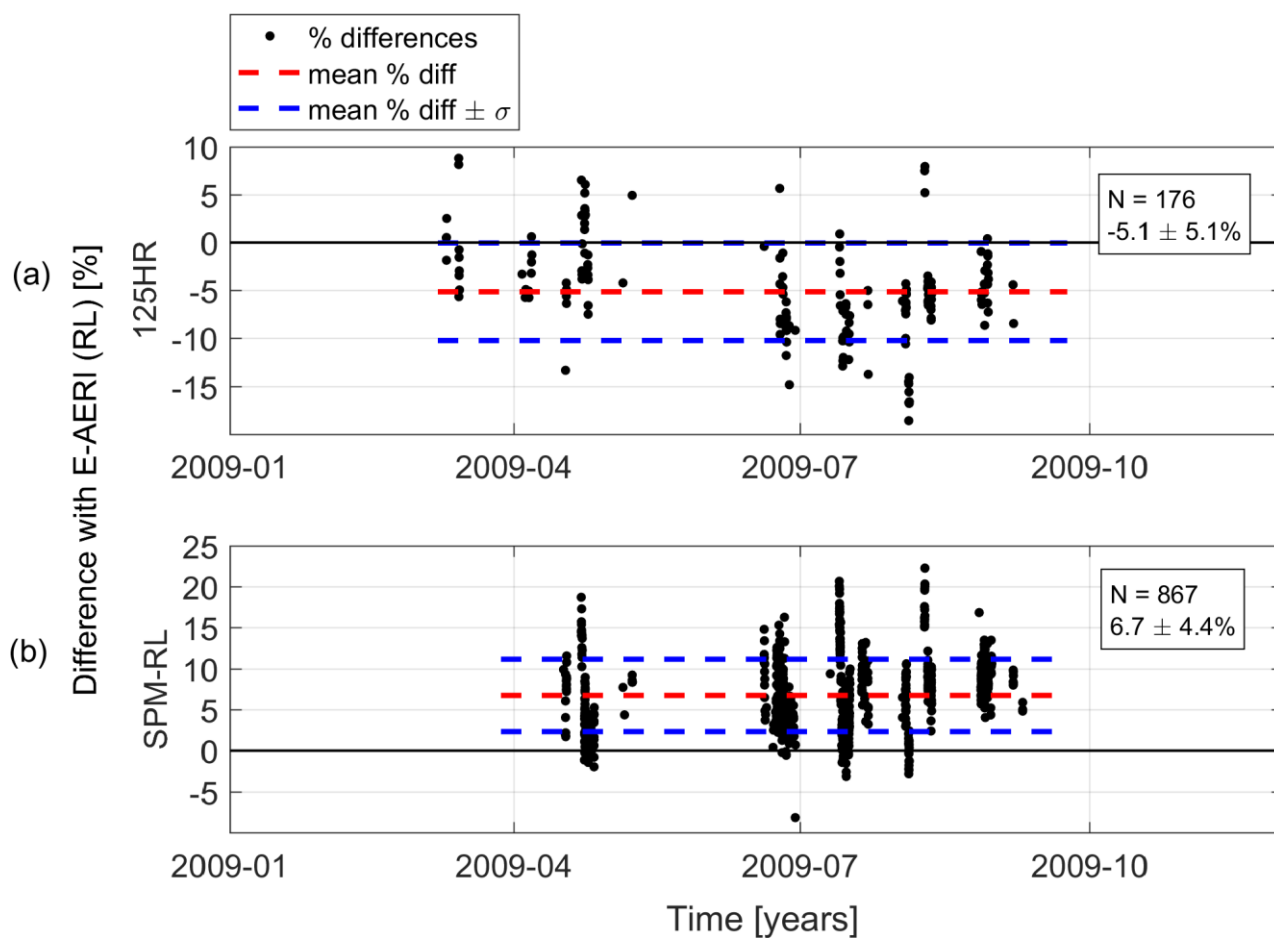
Figure 11: Same as Fig. 10 but for Ridge Lab-based comparisons between the E-AERI and the 125HR and SPM-RL. (Percent differences calculated using Eq. (2) where X = E-AERI-RL).
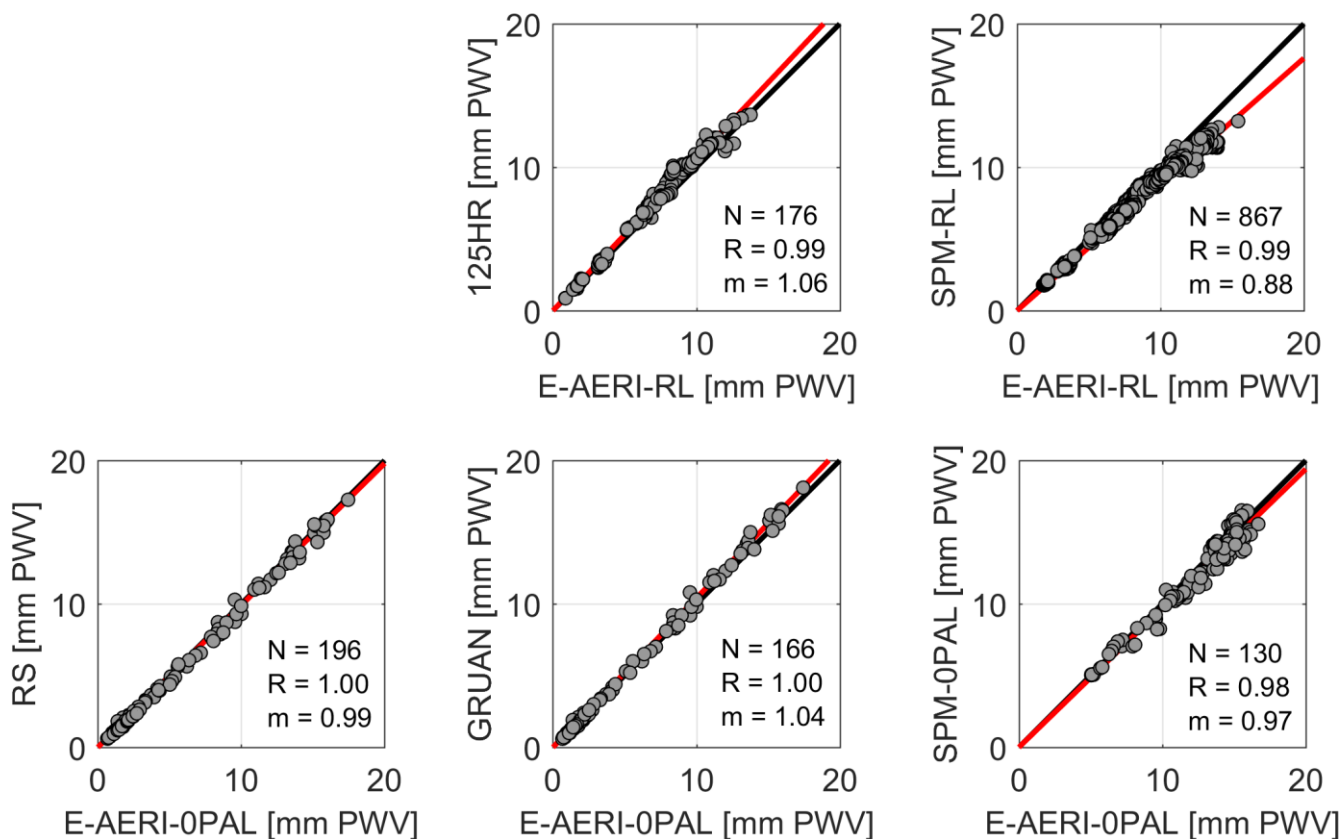
Figure 12: E-AERI-RL and E-AERI-0PAL water vapour correlation plots with the 125HR, SPMs, RS and GRUAN. The number of coincident measurement pairs (*N*), the *R* value, as well as the slope of the best fit line (*m*) is given on each plot.
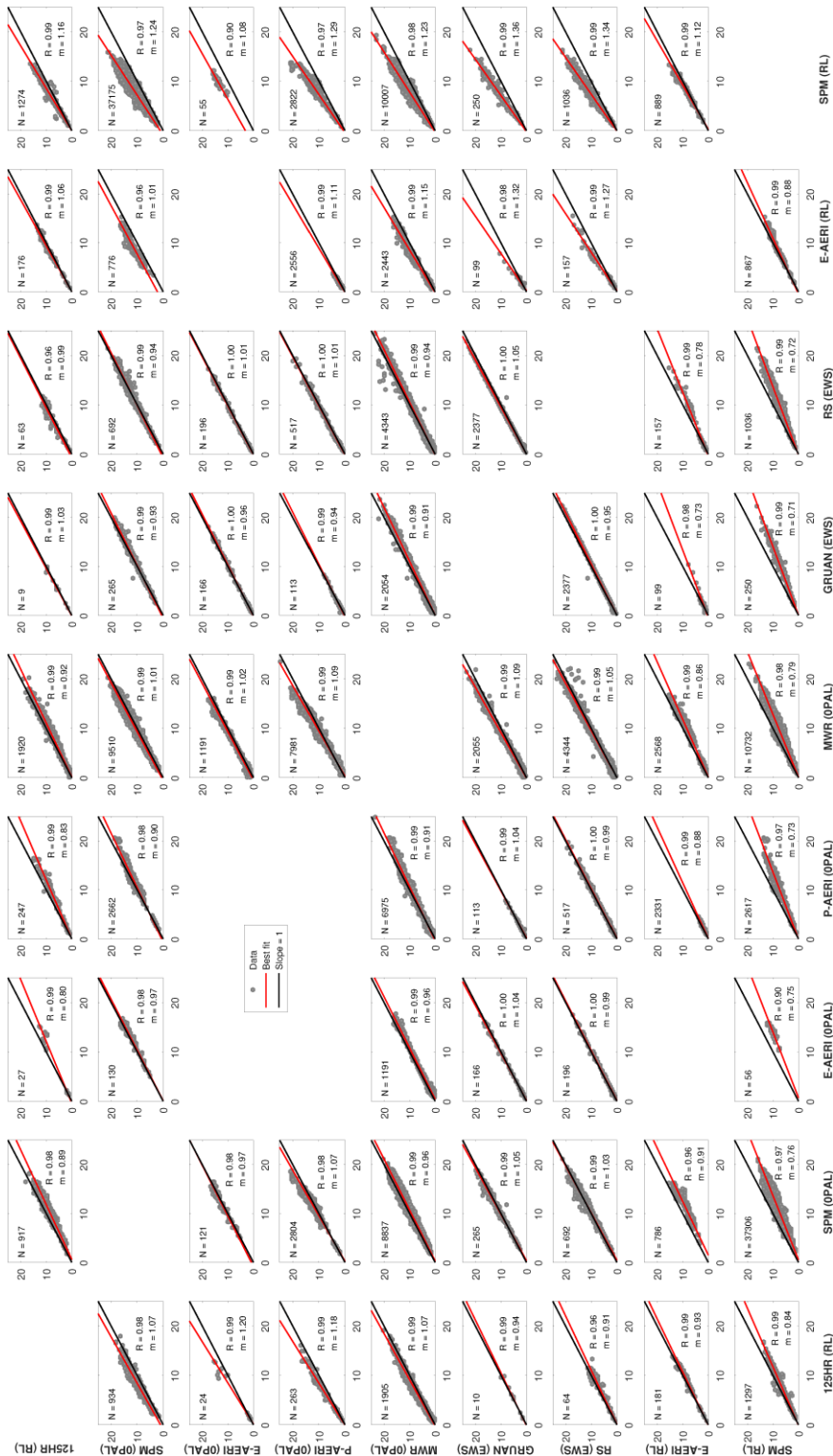
Figure 13: Water vapour total column (mm PWV) correlations between Eureka instruments. Data used for the radiosonde and GRUAN comparisons with the 125HR have been smoothed with the MUSICA averaging kernels. Blank spaces indicate that no coincidences were available for that pair of instruments. A high resolution version of this figure can be found in the Supplementary Materials as Figure S1.
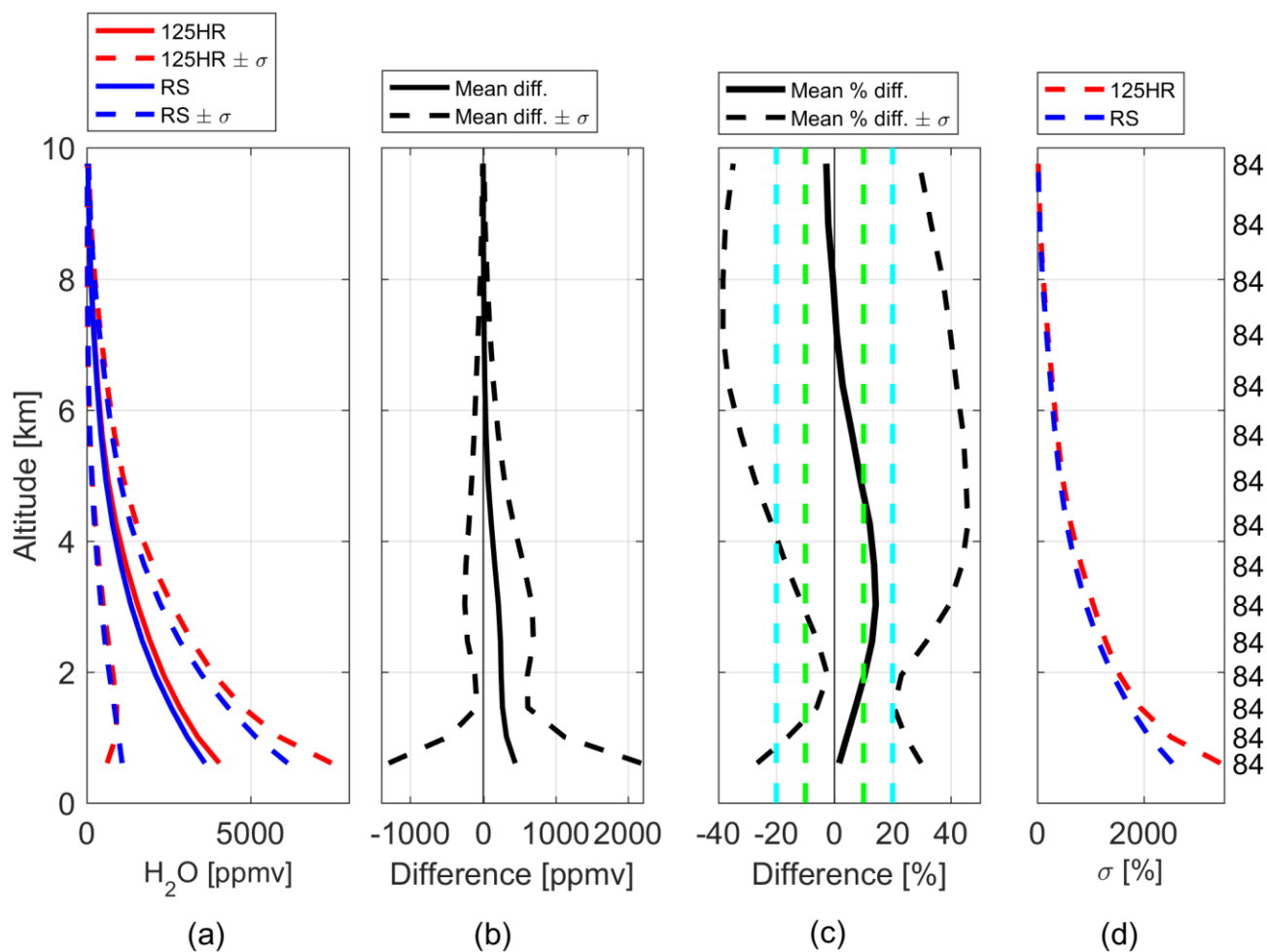
Figure 14: 125HR and RS profile comparisons, showing altitudes up to 10 km. (a) the mean of each instruments' coincident profiles; (b) the mean of the absolute differences (the vertical green dotted lines denote ±25 ppmv, the vertical cyan dotted lines denote ±50 ppmv); (c) the mean percent differences (the vertical green dotted lines denote ±10%, the vertical cyan dotted lines denote ± 20%); (d) the standard deviation of the coincident profiles. The numbers on the right-hand side indicate how many pairs of profiles contributed to each altitude level.
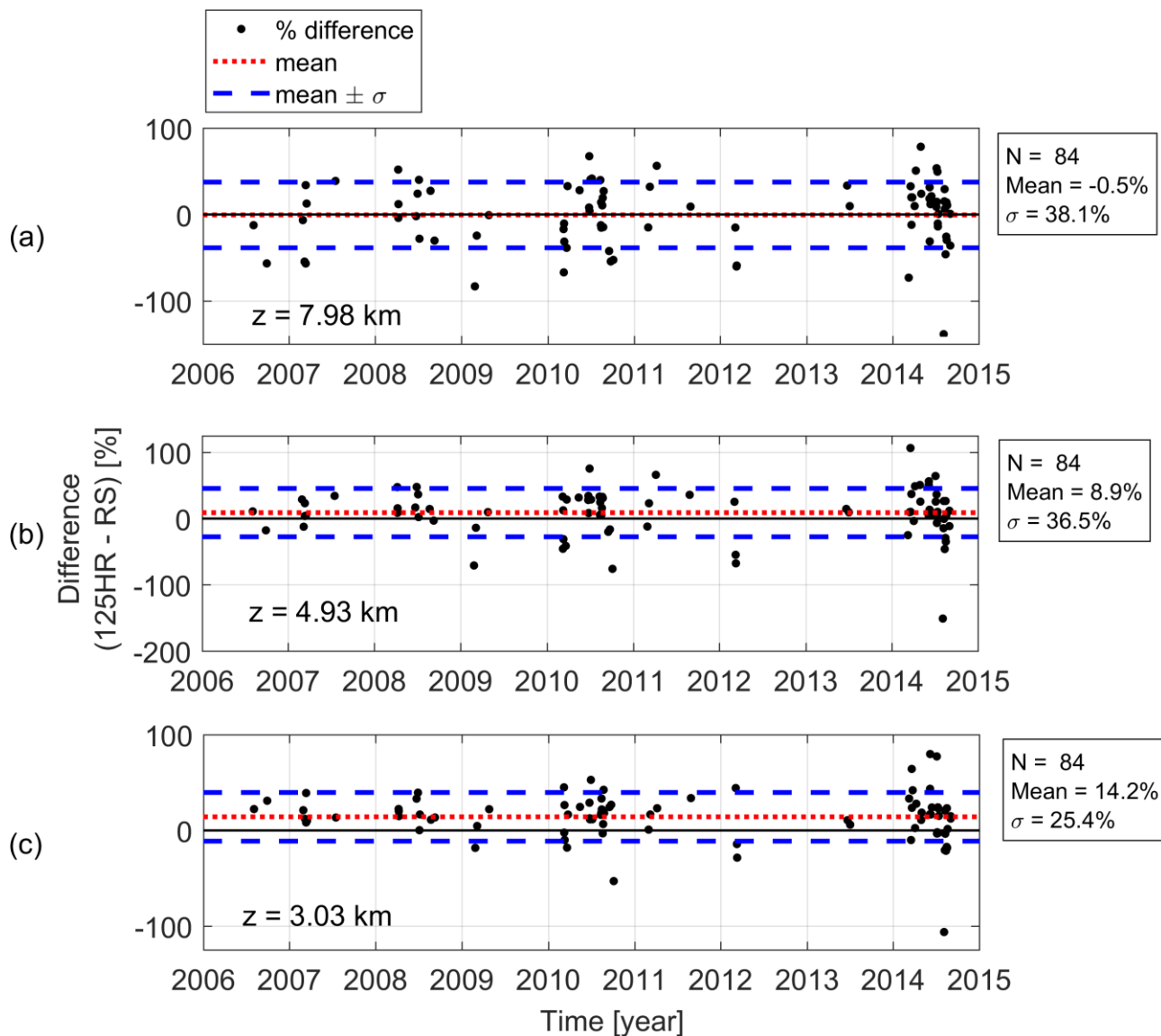
Figure 15: Percent difference between 125HR and radiosonde measurements at representative altitudes.
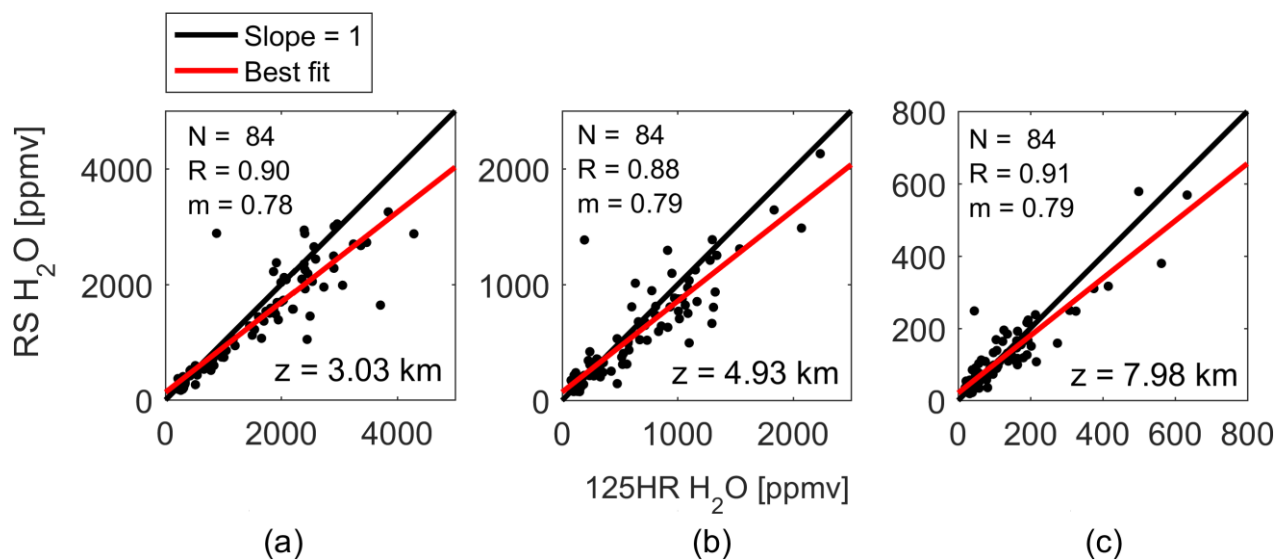
Figure 16: Correlation plots for 125HR and radiosonde water vapour measurements at the three altitude levels indicated.
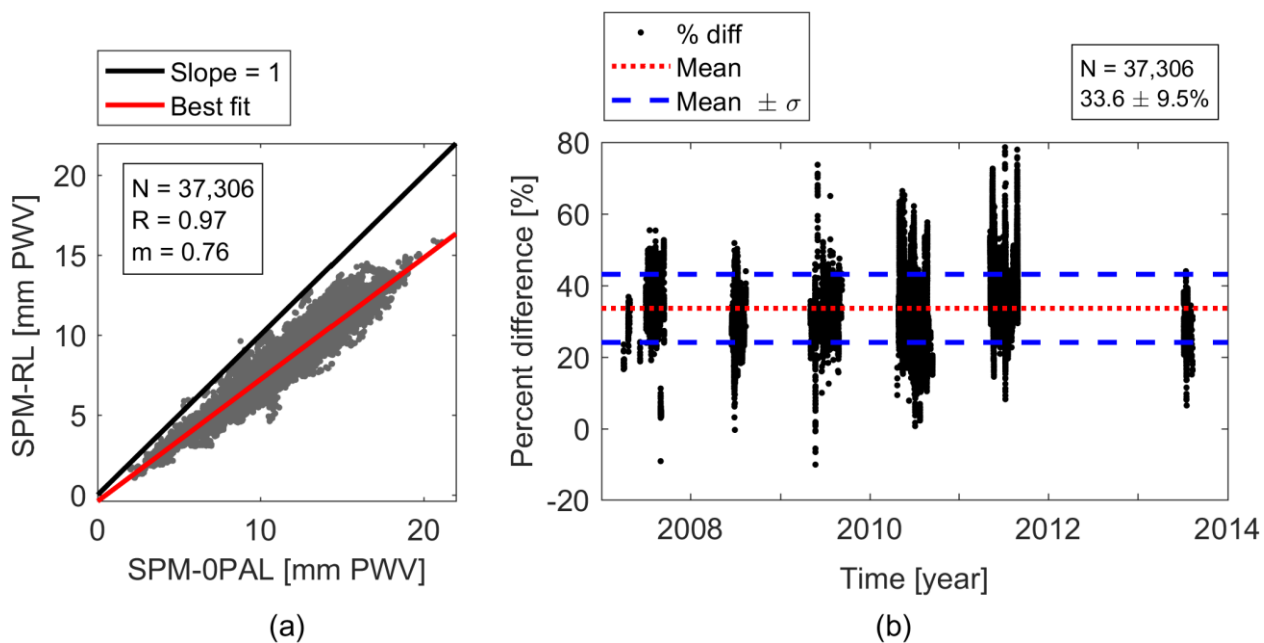
Figure 17: Comparisons between the water vapour measured by SPMs located at the Ridge Lab and at 0PAL. (a) the correlation plot between the two instruments; (b) the percent differences (SPM-RL minus SPM-0PAL).
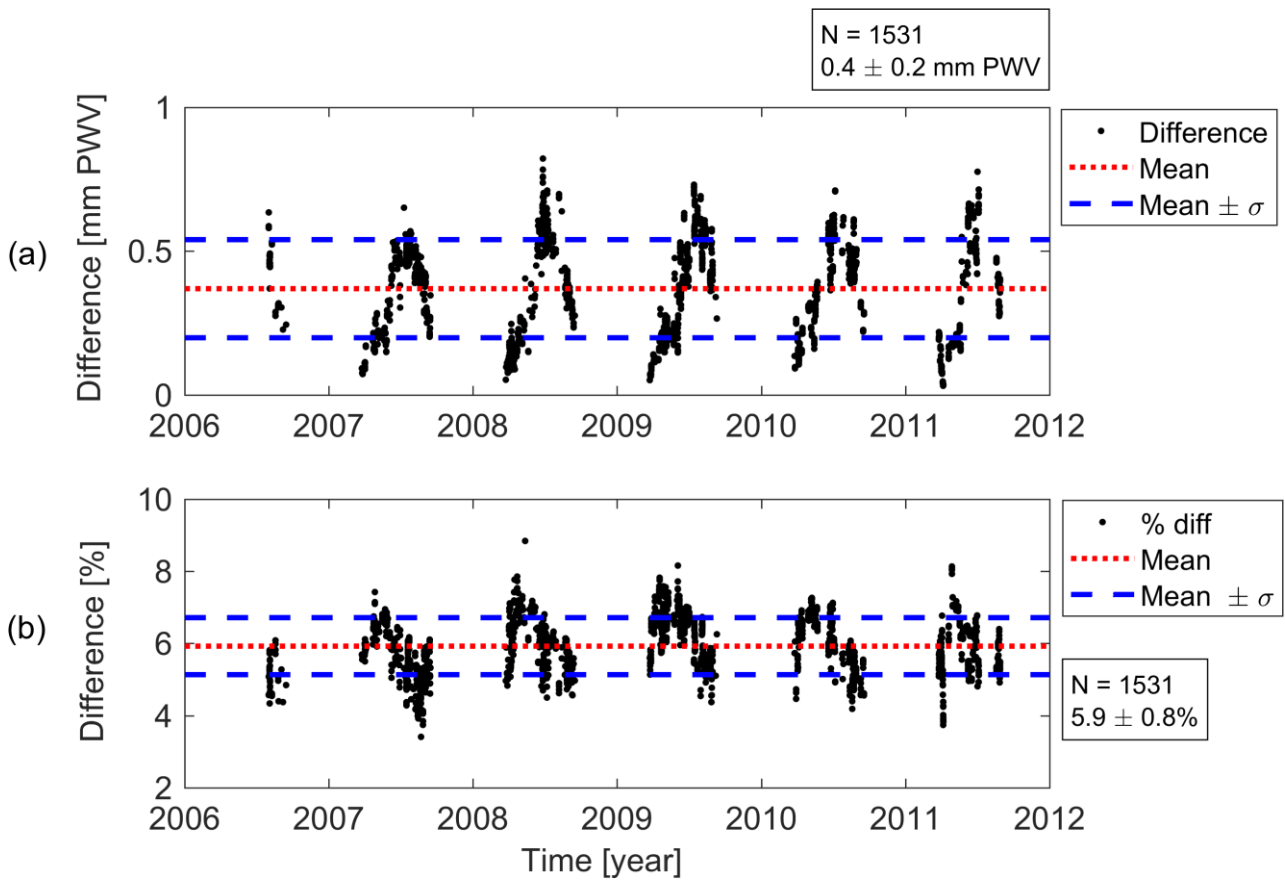
Figure 18: Difference between the current Eureka MUSICA product (v2015) and the previous Eureka MUSICA product (v2012). (a) differences and (b) percent differences, where X = v2015 and Y = v2012 in Eq. (1) and (2).
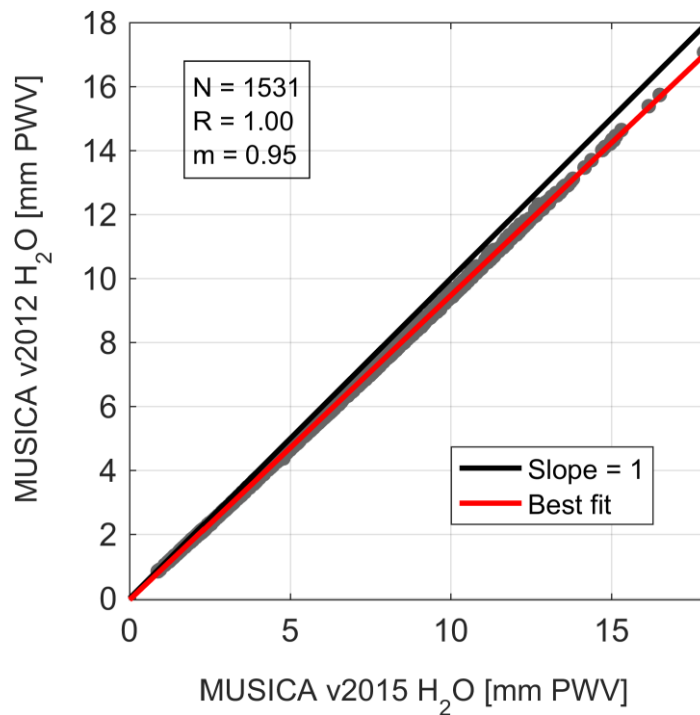
Figure 19: MUSICA v2015 vs. MUSICA v2012 water vapour at Eureka (PEARL Ridge Lab).
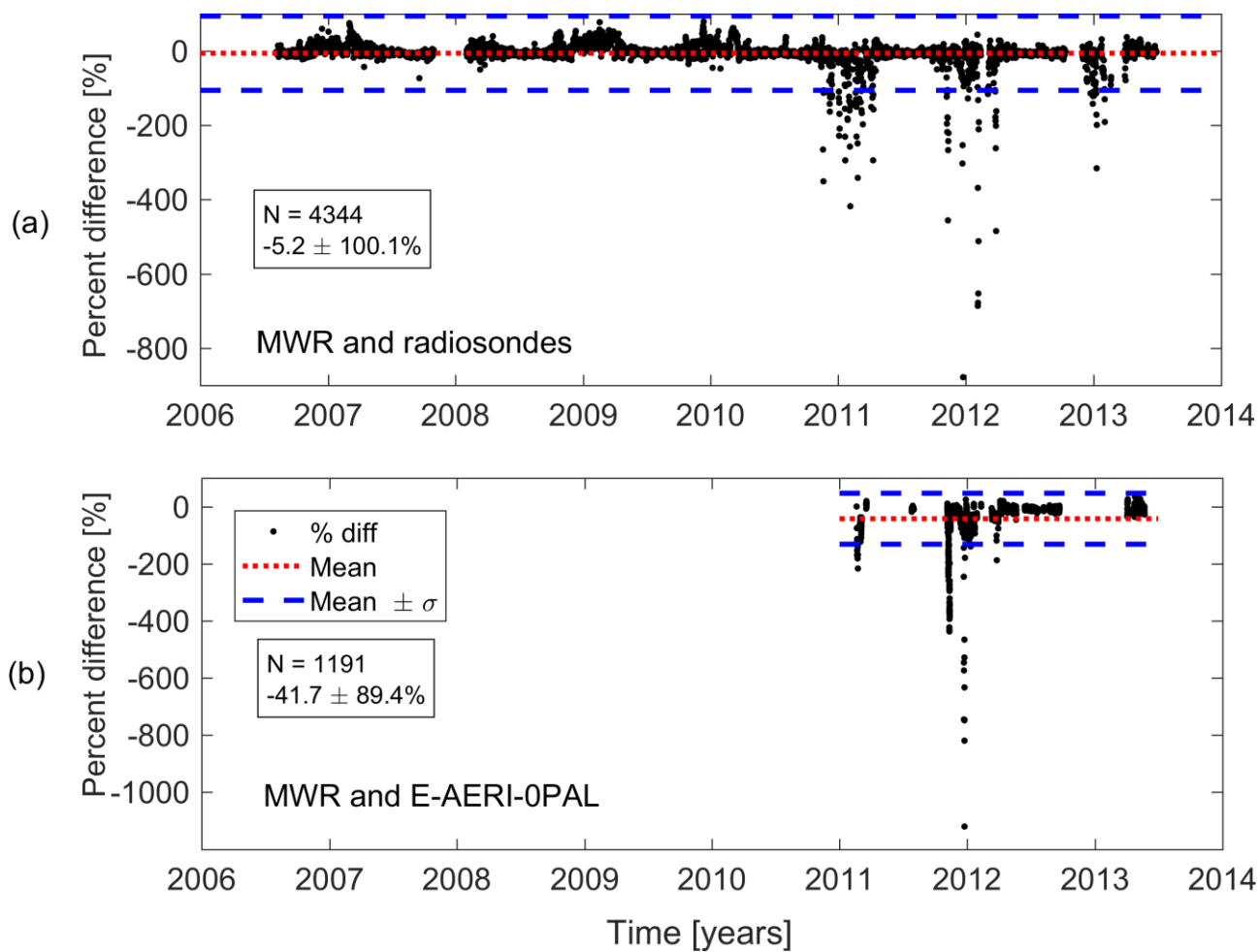
Figure 20: Comparisons between the MWR and co-located E-AERI and radiosondes. (a) the percent difference comparison between the MWR and E-AERI while located at 0PAL. (b) the percent differences between the MWR and radiosondes. The change in agreement from November 2010 onwards is clearly evident.