GEWEX Water Vapor Assessment: Validation of AIRS Tropospheric Humidity Profiles with Characterised Radiosonde Soundings

Tim Trent[1,2], Marc Schröder[3], and John Remedios[1,2]

[1]Earth Observation Science, Department of Physics and Astronomy, University of Leicester, University Road, Leicester, LE1 7RH, UK.
[2]National Centre for Earth Observation, Department of Physics and Astronomy, University of Leicester, University Road, Leicester, LE1 7RH, UK.
[3]Satellite-Based Climate Monitoring, Deutscher Wetterdienst/Frankfurter Strasse 135, 63067 Offenbach, Germany

Key Points:

- Altitudes below 250 hPa corrected operational radiosondes show similar uncertainty performance to GRUAN.
- AIRS tropospheric water vapor biases relative to GRUAN are within 6±0.3% ppmv below 300 hPa.
- Estimated collocation uncertainty in the northern hemisphere shown to reduce to below 1% ppmv for calculated yearly biases

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1029/2018JD028930

Corresponding author: T. Trent, tjt11@le.ac.uk

Abstract

The tropospheric water vapor profile record from the Atmospheric Infrared Sounder (AIRS) now spans over a decade, making it a valuable resource for climate studies. To be considered as a Climate Data Record (CDR) it is key that the ultimate performance of these observations are understood. The GEWEX water vapor assessment (G-VAP) has been tasked with characterising the current state of the art in water vapor products currently available for climate analysis. Within the scope of this exercise, water vapor profiles from AIRS have been assessed using collocated characterised *in situ* measurements of tropospheric water vapor between 2007 and 2012. We first show how previously published methods for correcting radiosondes can be applied to global records, which show high correlations to GCOS Reference Upper-Air Network (GRUAN) performance at pressures between the surface and 250 hPa. We go further and show the first comparison of uncertainties from both the newly created Characterised Radiosonde Measurement (CRM) and GRUAN data sets. Global estimates of AIRS water vapor profile (wet/dry) biases relative to GRUAN and CRM are within 6±0.3 % ppmv and 15±0.1 % ppmv below 300 hPa respectively. The CRM record allows latitudinal analysis for the first time, which when examined shows sensitivity to changes in absolute concentration due to large scale circulation in the International Tropical Convergence Zone (ITCZ). This paper advances the use of state-of-the-art *in situ* records for characterising absolute performance, recognising that long term stability needs further research.

1 Introduction

Water vapor is an important greenhouse gas within the atmosphere which influences (directly and indirectly) the radiative balance of the Earth as well as surface and soil moisture fluxes. It is sufficiently abundant and short-lived that it is essentially under natural control (Sherwood, Roca, Weckwerth, & Andronova, 2010). With a predominant capacity for positive feedback (≈ 2 W m^{-2} K^{-1}, Dessler, Zhang, and Yang (2008)), water vapor acts as the largest amplification mechanism for anthropogenic climate change compared to radiative forcing from greenhouse gases (Chung, Soden, Sohn, & Shi, 2014). This makes water vapor a critical variable for climate studies (Held & Soden, 2000; Trenberth, Fasullo, & Smith, 2005). The role of tropospheric water vapor in atmospheric processes extends over a wide range of spatial and temporal scales, from the global climate to micrometeorology (Bevis et al., 1992).

In order to use satellite observations to capture these varying scales, assessment of their performance against characterised *in situ* measurements (with well-defined uncertainties) is of high importance. The new generation of infrared (IR) sounders that have replaced the long standing High-resolution Infrared Radiation Sounder (HIRS) series of instruments and are able to resolve vertical structure of water vapor in the troposphere. In this lowest region of the Earth's atmosphere the absolute concentrations of water vapor vary by over four orders of magnitude. Biases and uncertainties in profile measurements near the surface can therefore exceed the actual concentrations in the upper troposphere. Observations of water vapor profiles in the lower stratosphere and troposphere are considered an Essential Climate Variable (ECV) by the Global Climate Observing System (GCOS). For a satellite-based water vapor ECV, GCOS prescribes target requirements of 5% measurement uncertainty and a stability of 0.3% per decade. Therefore, validation of any satellite humidity profile requires *in situ* measurements that are independent of the satellite retrieval, a detailed uncertainty budget, evidenced traceability to SI and follow agreed community best practices to assess their application for climate studies. These fiducial reference measurements are key to assess the application of AIRS water vapor for climate studies.

The GEWEX Data and Assessments Panel (GDAP) has initiated the GEWEX water vapor assessment (G-VAP), which has the major purpose to characterise the current

state-of-the-art in satellite water vapor products. The overall goal is to conduct consistent evaluations and inter-comparisons in order to point out strengths, differences and limitations of long-term satellite data records, in particular with respect to stability. Besides the large amount of available satellite data records such data records are updated frequently and it will be a future element of G-VAP to reassess the change in quality between the different versions. Through this effort the selection process of suitable water vapor data products by the general (climate analysis) community and in particular by GDAP is supported. Further details are available at www.gewex-vap.org and Schröder et al. (2017).

Launched aboard the Aqua platform on the 4^{th} of May 2002, the Atmospheric Infrared Sounder (AIRS), with its companion instrument the Advanced Microwave Sounding Unit (AMSU-A) (Aumann et al., 2003) is the first of a new generation of advanced IR nadir sounders providing hyper-spectral radiances for numerical weather prediction (NWP) (Le Marshall et al., 2006), and research purposes. The overall objectives of AIRS are to observe the global water and energy cycles, climate variation and trends, and the response of the climate system to increased greenhouse gases (Chahine et al., 2006).

AIRS provides humidity profiles with a good vertical resolution through the troposphere with a mean resolution of 2.7 $\pm$ 1 km between 600-1000 hPa, 2.8 $\pm$ 0.7 km between 600-300 hPa and 3.14 $\pm$ 1.4 km between 300-100 hPa (Maddy & Barnet, 2008). The increased sensitivity of AIRS to the vertical distribution of water vapor is an important step in observing capability because concentrations of atmospheric water vapor are mainly controlled by temperature rather than anthropogenic emission (Myhre et al., 2013). With retrievals of atmospheric temperature from AIRS showing an agreement to *in situ* observations within $\approx$0.5 K (Feltz et al., 2017), understanding the absolute performance of AIRS humidity retrievals is key for improving our understanding of radiative forcing and feedback effects from water vapor.

Since 2011 (Immler & Sommer, 2011), the GCOS Reference Upper-Air Network (GRUAN) has provided highly characterised 'climate quality' observations of temperature and humidity in order to meet the criteria required for reference measurements (Dirksen et al., 2014; Immler et al., 2010). The network is not intended to be globally complete but will sample major climate regions. GRUAN currently contains sixteen sites (at the time of this study) with the intention to grow to several dozen in number. All sites are attached to an institution that has relevant expertise to manage them and adhere to the GCOS climate monitoring principles (Seidel et al., 2009). Like their operational counterparts, GRUAN radiosondes are launched primarily at the routine times of 00:00 hrs and 12:00 hrs UTC, with some sites performing additional launches at six to twelve hourly spacing. This has the drawback that when collocating upper-atmosphere soundings with satellite overpasses (such as Aqua) with GRUAN, larger spatial/temporal criteria have to be employed compared to operational stations in order to produce meaningful statistics. However, while operational radiosondes produce more collocations, these lower-quality soundings contain uncorrected/accounted sources of bias and uncertainty.

In this study we attempt to bridge this gap by applying documented corrections to operational soundings at sites that have been identified as using Vaisala RS92 radiosondes. The methodology used here also allows for the calculation of the associated uncertainty with the bias correction estimate. These new bias-corrected *in situ* profiles are referred to as 'characterised' rather fiducial as there are a number of assumptions that have to be applied to the operational archive. Therefore, these characterised observations can be thought of as complimentary to GRUAN rather than a direct substitution. Assessment of these corrections are performed though comparisons of common soundings i.e. the same profiles reported separately through GRUAN and Characterised Radiosonde Measurements (CRM) archives. Finally, the primary objective of producing CRM is to assess AIRS water vapor profile retrievals within the context of G-VAP.

This paper is structured as follows; section 2 introduces the AIRS humidity data set and the radiosonde archives from GRUAN and the UK Met Office (UKMO). In section 3 we describe the generation of CRM using an adapted and extended correction algorithm that has been applied to the low resolution radiosondes presented in section 2. Section 4 outlines the collocation and comparison methodology to assess AIRS humidity biases. Global analysis of AIRS comparisons to GRUAN and CRM are presented in section 5.

2 Data Description

This section provides an overview of all the data used in this study.

2.1 AIRS L2 Humidity Profiles

The AIRS Water vapor profile retrievals used in this study come from the version 6 level 2 (L2) support product (E. Olsen et al., 2013; E. T. Olsen et al., 2007). The support product represents the real skill level of the AIRS retrieval (Wong et al., 2015), have 4 times denser vertical resolution (Gettelman et al., 2004) and include the first guess profiles and averaging kernels needed for profile validation (Rodgers & Connor, 2003). The AIRS retrieval process uses a combination of infrared (IR) and microwave (MW) radiances to provide estimates of all-sky (up to 80% cloud fraction) temperature, water vapor, trace gases, skin temperature, cloud and outgoing long wave (OLR) radiation properties within the 45 km AMSU (nadir) instantaneous field of view (IFOV) commonly known to as the 'golf ball'. From hereon in it shall be referred to as the AIRS field of regard (FoR).

Introduced in the version 5 release of the L2 product (E. T. Olsen, Fishbein, Lee, & Manning, 2011), the trapezoid functions ($\mathbf{F}$) are used map the compressed layer averaging kernel ($\mathbf{A}$) on to the AIRS full retrieval grid. Associated with the averaging kernel and degrees of freedom of the retrieved profile products, these trapezoid functions also define the upper and lower bounds for which the retrieved layer quantities of water vapor, atmospheric temperature (T) and trace gases are defined. With this in mind, another factor to consider is that the retrieval state can only be perturbed as a superposition of the trapezoid during an iteration. The faces of each trapezoid equal 0.5 and drop off to zero linearly in ln(p) with the exception where they terminate at the top of atmosphere (TOA) or the surface. Here the trapezoid face extends without dropping off, this is often referred to as "halftop/halfbot". The number of trapezoids are dependent on the parameter being retrieved (e.g. H_2O=11), and therefore limits the independent structure that can be resolved. For further information on the trapezoids, layers and levels in the AIRS L2 products please see E. T. Olsen et al. (2011). Therefore, to interpolate between the AIRS 100 level retrieval grid and the 11 H_2O layers the layer averaging kernel $\mathbf{A}$ is mapped by multiplying it with the trapezoid function and its pseudo-inverse ($\mathbf{F}'$):

$$\tilde{\mathbf{A}} = \mathbf{F}\mathbf{A}\mathbf{F}'. \tag{1}$$

Where $\tilde{\mathbf{A}}$ is the level averaging kernel. The result of mapping between the 11 layer and 100 level grids is shown in Figures 1a & 1b respectively. The averaging kernel also provides information content within the profile by summing the diagonal elements of the matrix. This is referred to as the degrees-of-freedom (DOF) of signal, and is the same value regardless of whether $\mathbf{A}$ or $\tilde{\mathbf{A}}$ is used in the calculation. Figure 1c presents the mean AIRS H_2O DOF global distribution for 2007. The mean vertical distribution of the averaging kernel diagonal elements is illustrated in Figure 1d, where the cross section was calculated from nadir retrievals from 2007.

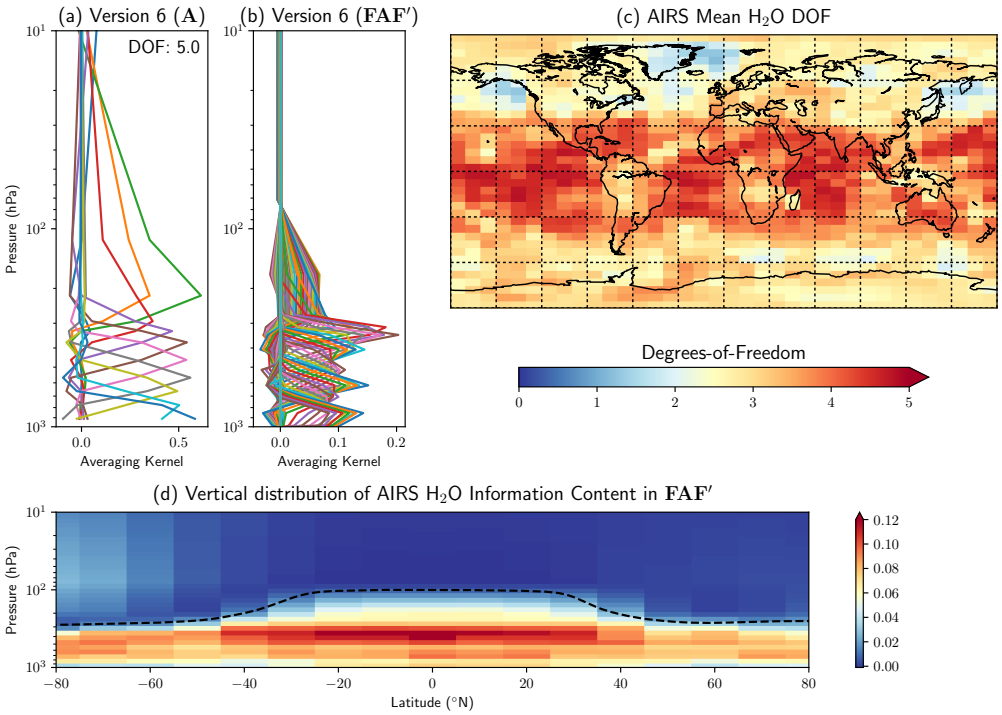


Figure 1. Example of an AIRS version 6 averaging kernel (**A**) over the Southern Great Plains (SGP) site from the 24^{th} of June 2009. This example demonstrates the employment of the trapezoid functions (**F**) which map the **A** from the 11 stored layers (1a), to the AIRS full 100 level retrieval grid (1b). The total information content of the AIRS H_2O profiles are represented by the degrees-of-freedom (DOF). Average global DOF distributions for AIRS (2007) are shown in (1c), with the vertical distribution (unitless) as a function of latitude (1d). Black dashed line in (1d) is a climatological cold point tropopause calculated from AIRS L2 data. AIRS mean DOF values for the version 5 humidity L2 product ranged between 4.46 in the tropics to 2.89 at the poles (Maddy & Barnet, 2008).

2.2 Radiosondes

This study employs two different radiosonde datasets to investigate the performance of AIRS H_2O profiles. The first is based on operational radiosonde soundings from more than 900 global upper-air stations that were extracted from the UKMO data archive held at Centre for Environmental and Data Analysis (CEDA). This dataset consists of vertical profiles of temperature, dew-point temperature, wind speed and wind direction from the surface to altitudes generally between 20-30 km. While upper-air data can be report up to four times a day (00hrs, 06hrs, 12hrs & 18hrs UTC), the actual number of ascents varies widely between countries and stations. In the UKMO archive approximately two-thirds of stations report (at least) twice daily at 00:00 hrs and 12:00 hrs UTC. The UKMO receives upper air station data at standard resolution which measurements are reported at standard and significant pressure levels. Standard pressure levels are defined at 1000, 925, 850, 700, 500, 400, 300, 250, 200, 150, 100, 70, 50, 30, 20 and 10 hPa. Significant pressure levels are added to improve the vertical description of the measurement profile, these can occur for example at levels in the profile where an inversion occurs or where there are significant changes in humidity not present in the standard levels. These additional levels are also intended to allow for profiles to be linearly interpolated such that

the temperature profile does not deviate from the observed value by more than 1 K below, and 2 K above 300 hPa (UKMO, 2006). This data record form the basis for the Characterised Radiosonde Measurement (CRM) Archive that is described in section 3.

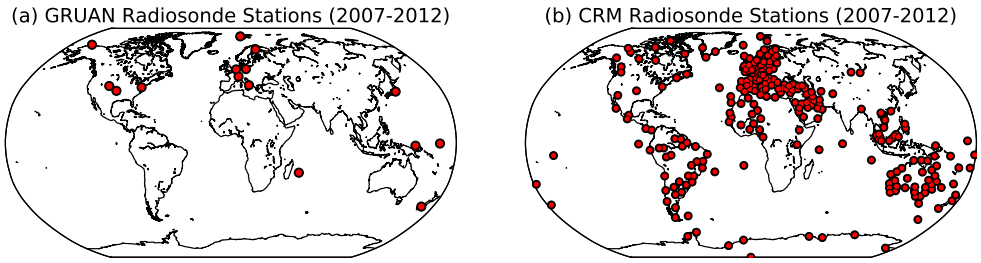


Figure 2. Global distribution of radiosonde sites used in this study from the GRUAN (a) and CRM (b) networks. Matches between AIRS and the respective radiosonde launches are performed for the time period January 2007 to December 2012.

The second source of radiosonde measurements come from the GCOS Reference Upper-Air Network (GRUAN) (Dirksen et al., 2014; Immler et al., 2010) archive. GRUAN atmospheric soundings are reported on time intervals of 2 seconds during flight from the surface into the Upper Troposphere/Lower Stratosphere (UTLS) rather than the set pressure grid of the operational radiosonde archives. The higher vertical resolution of GRUAN data captures changes in humidity and temperature which can be missed or underrepresented by standard and significant pressure levels of operational records. The scope of GRUAN is to provide long-term fiducial measurements that can be used for calibration/validation exercises, the study of atmospheric processes and determining trends. Certified GRUAN sites not only undergo an annual review, but are also subject to periodic auditing of their measurement programs to ensure all sites continue to meet GRUAN practice standards. The global distribution of GRUAN sites used in this study are shown in Figure 2a.

3 Characterising Operational Radiosondes

Presented here is a description of the corrections applied to the identified RS92 radiosondes from the UKMO operational archive. The technique is based around correction techniques developed in Miloshevich, Paukkunen, Vömel, and Oltmans (2004); Miloshevich, Vömel, Whiteman, and Leblanc (2009); Miloshevich et al. (2006), and have been adapted for lower resolution operational profiles. This approach provides a statistical estimate of the profile uncertainty which is crucial for understanding and assessing satellite retrieved profile performance. The global location of UKMO sites used are shown in Figure 2b. The original CRM record (Trent, 2015) used Murphy and Koop (2005) to calculate vapour pressures. However, for this study CRM was recalculated using Hyland and Wexler (1983) to be consistent with GRUAN.

3.1 Radiosonde Correction and Uncertainty Estimation

Published correction methods (Miloshevich et al., 2004, 2009, 2006) are performed on a standardised vertical temporal grid (6 second) representing the ascent of the balloon due to the time-dependent components to the sensor response. Global operational radiosonde databases, however only report on a subset of pressure levels which exclude

the relevant time information. Therefore, to interpolate onto a vertical time grid (or time profile) the ascent rate of the balloon needs to be estimated.

The method employed for this study assumes a constant rate of ascent and determines an uncertainty that is propagated through the correction method. To estimate the constant rate of ascent observations from the GRUAN data archive were used as they include ascent rate values. At the time of calculation the data record included 14,414 soundings, with over 80 million individual measurements. Examination of the ascent rate data from GRUAN observations yields a mean value of 4.73 ms^{-1} ± 0.86 ms^{-1} (Trent, 2015). Next the time interval (δt) for adjacent pressure levels is calculated by converting the difference between levels P1 and P2 from hPa into m (δz). This δz value is then divided by the ascent rate v_z to yield δt. The time profile is then the cumulative sum of the δt values. Once the low resolution time profile has been created, the maximum (time) value is then used to create the appropriate number of levels for the 6 second vertical grid. The ascent time information is then used to interpolate the low resolution temperature, relative humidity (RH) and pressure profiles onto the 6 second grid.

This method was tested on soundings from the UKMO Cambourne site to investigate the effects of interpolation on the lower resolution profiles.The Cambourne site was chosen as; i) it was possible to obtain both low and high (temporal) resolution data products, ii) RS92 model radiosondes are used, iii) not part of the GRUAN network and iv) the site experiences a large variability of weather originating from the Atlantic ocean due to its location of the site in the south-west of the UK. This last point was the most significant reason for choosing Cambourne as the factors that have the biggest effect on the ascent rate are atmospheric density and the coefficient of drag outside factor retaining to the actual balloon itself. To ensure that a linear interpolation routine was sufficient for this task, five other algorithms; i) cubic spline, ii) quadratic, iii) least squares quadratic, iv) Akima (1970) and v) Akima (1991) were used on a test set of profiles measured at Cambourne from 2009 (see Trent (2015) for further discussion).

The low resolution profiles were interpolated using each algorithm, corrected and then compared back to the correct high resolution profile. The linear interpolation method performed as well, and in some cases, better than the other algorithms. In general differences are less than 0.5% RH below 300 hPa, with an increase to between 1.2 - 1.5% RH nearer 200 hPa. This increase is due to the lower sampling in the operational sounding which can fail to capture the dehydration in the UTLS accurately. Therefore, the limiting factor for the profile interpolation is the residual structure in the low resolution profile.

The newly interpolated RH profile then has a time-lag correction applied to account for the response delay of the radiosonde sensors (Miloshevich et al., 2004):

$$\mathrm{RH_c} = \frac{\mathrm{RH} - (\mathrm{RH}(\mathrm{t_0}) \times \mathrm{X})}{1 - \mathrm{X}}, \tag{2}$$

where $\mathrm{X} = e^{-\frac{\Delta \mathrm{t}}{\tau}}$ and $\Delta \mathrm{t} = \mathrm{t} - \mathrm{t_0}$. Vaisala laboratory measurements of sensor time-lag (τ) which are given as a function of Temperature (T) are then used to calculate the corrected humidity ($\mathrm{RH_c}$) from the measured humidity (RH). By fitting a simple exponential this gives the relationship:

$$\tau(\mathrm{t}) = \alpha \times e^{\mathrm{A+BT}}, \tag{3}$$

where A and B are constants (-0.7399 and -0.07718 respectively) and the scale factor α (0.8) corresponds to approximately two standard deviations below the mean. This is an attempt to tackle this problem conservatively, meaning that 5% of sensors will be over corrected and 95% will be slightly under corrected. This approach simplifies the method

outlined in (Miloshevich et al., 2004). Next the interpolated profile is smoothed using a derivative based function (Miloshevich et al., 2004) that compensates for noise and humidity gradient amplification during the time-lag correction. The algorithm minimises the local third derivative which maximises 'smoothness' by constraining all the data points to within a specified tolerance of the original data. Therefore, vertical detail is retained while having a known maximum uncertainty in the smoothing operation. Further information can be found in Appendix B of Miloshevich et al. (2004). Finally the empirical mean bias correction is applied (Miloshevich et al., 2009), which corrects for a dry bias created by daytime solar heating of the radiosondes sensors.

$$\mathrm{RH_{corr}} = \mathrm{G(P, RH)} \times \mathrm{RH}_{TLAG}, \tag{4}$$

where RH_{TLAG} is the smoothed, time-lag corrected RH profile ($\mathrm{RH_c}$). The correction factor $\mathrm{G(P, RH)}$ is generally:

$$\mathrm{G(P, RH)} = \frac{100}{(\mathrm{F(P, RH)} + 100)}, \tag{5}$$

where $\mathrm{F(P, RH)}$ is polynomial function fit to the RH profile:

$$\mathrm{F(P)} = \sum_{i=0}^{\mathrm{N}} \mathrm{a}_i \times \mathrm{P}^i, \tag{6}$$

here a and P are pre-computed coefficients and N is the order of the fit. This step accounts for mean calibration bias and for solar radiation error in daytime soundings, and they are valid from the surface to about 18 km altitude (100 hPa daytime).

Once corrections are applied to an individual sounding there are several sources of uncertainty that need to be accounted for. These are described as either a fraction of the measured RH profile (%) or in absolute RH units (% RH). The night-time residual bias uncertainty for RS92 corrections is primarily due to the statistical uncertainty in the mean cryogenic frost point hygrometer (CFH) (±2%) as well as an offset in RH uncertainty (±0.5% RH). Comparisons to microwave radiometer (MWR) also showed a calibration (production) batch-dependent variability of ±2% relative to the long-term mean. As this additional bias uncertainty is not representative of the majority of cases, therefore an "expected value" is adopted which when combined with other components gives an uncertainty estimate of bias of ±(3% + 0.5% RH). The time-dependence of uncertainty variability is a diurnal component of the total estimate and therefore applicable to daytime corrections. However, there is greater statistical uncertainty pertaining to this estimate which leads to a bias uncertainty of ±(4% + 0.5% RH). In addition to this, studies of an unpublished DWD data set of pre-launch ground checks at 100% RH for 2005-2008 highlighted a moist bias of 3% compared to the Miloshevich study (Miloshevich et al., 2009). This was attributed to a change in the Vaisala calibration at the beginning of 2006 and represents a moist bias uncertainty of 1% resulting in an updated night-time uncertainty estimate of ±(4% + 0.5% RH) and a daytime uncertainty estimate of ±(5% + 0.5% RH). With the addition of the ascent rate uncertainty the final uncertainty estimates are ±(8% + 0.46% RH) and ±(9% + 0.46% RH) for night-time and daytime respectively.

3.2 Comparison of GRUAN and CRM RH Profiles

Before inference can be drawn from collocated AIRS and radiosonde profiles it is important to see how the different correction approaches compare. Common soundings from CRM and GRUAN are first identified before being sampled on the AIRS L2 standard pressure levels. For a level to be sampled from either CRM or GRUAN the radiosonde

Table 1. Regression coefficients from comparisons of CRM and GRUAN RH profiles on the lowest 12 AIRS standard pressure levels. Coefficients are calculated using orthogonal distance regression, where β is the gradient and α is the intercept. The root mean square error (RMSE) and intercept are given in units of relative humidity.

Pressure (hPa)	β	α (%)	R	RMSE (%)
100	0.91±0.01	1.11±0.02	0.94	1.3
150	0.93±0.01	1.20±0.03	0.89	2.3
200	0.98±0.01	1.45±0.08	0.94	6.5
250	1.00±0.01	1.68±0.10	0.96	6.7
300	1.00±0.01	2.03±0.13	0.97	6.0
400	1.02±0.00	2.98±0.09	0.99	5.2
500	1.04±0.00	2.89±0.07	0.99	5.2
600	1.06±0.01	3.36±0.21	0.97	8.8
700	1.04±0.01	2.92±0.15	0.99	6.3
850	1.02±0.01	2.43±0.23	0.98	5.9
925	1.01±0.00	2.67±0.18	0.99	4.4
1000	1.01±0.01	2.44±0.60	0.92	7.9

pressure level must be within 1 hPa of the AIRS grid. The collated radiosonde profiles are then analysed at each pressure level using orthogonal distance regression (ODR), as it accounts for uncertainties in both variables (Boggs and Rogers (1990)). The slope (β), intercept (α), correlation (R) and root mean square error (RMSE) are calculated for each level and presented in Table 1. From these results three key points stand out:

1. in the mid-to-lower troposphere (altitudes below 500 hPa) CRM has an RH dependent bias of less than ±6% RH. The UTLS has the largest RH dependent bias of 9% at 100 hPa.
2. CRM has a mean offset bias relative to GRUAN of 2.72% between 1000-300 hPa, which reduces to 1.34% at altitudes above 300 hPa.
3. Higher RMSE values are found generally between 600-200 hPa indicating lower precision in the correction algorithm.

In addition to a direct comparison of corrected RH values it is also possible to examine the uncertainty distributions from CRM and GRUAN (Figure 3). Probability density functions (PDFs) for both GRUAN and CRM were calculated using the Kernel Density Estimation (KDE) method (Parzen, 1962; Rosenblatt, 1956):

$$\mathrm{f(U_{RH})} = \frac{1}{\mathrm{nh}} \sum_{i=1}^{\mathrm{n}} \mathrm{K}\left(\frac{\mathrm{U_{RH}} - \mathrm{U_{RH_i}}}{\mathrm{h}} \right), \tag{7}$$

where $\mathrm{f(U_{RH})}$ is the PDF for either CRM or GRUAN, $\mathrm{U_{RH}}$ is the respective radiosonde uncertainty, n is the number of data points, K is the kernel function and h is the bandwidth smoothing parameter. The KDE method is used here as it allows for the PDF to be estimated and compared without being effected by any discontinuities in either distribution. For this study, a Gaussian kernel density function is used and the bandwidth is calculated using Scott's Rule (Scott, 2015). Once calculated for both GRUAN and CRM, the two PDFs ($\mathrm{f(U_{CRM})}$ & $\mathrm{f}(U_{\mathrm{GRUAN}})$ respectively) are then compared by calculating the Hellinger distance, $\mathrm{D_H}$ (Hellinger, 1909):

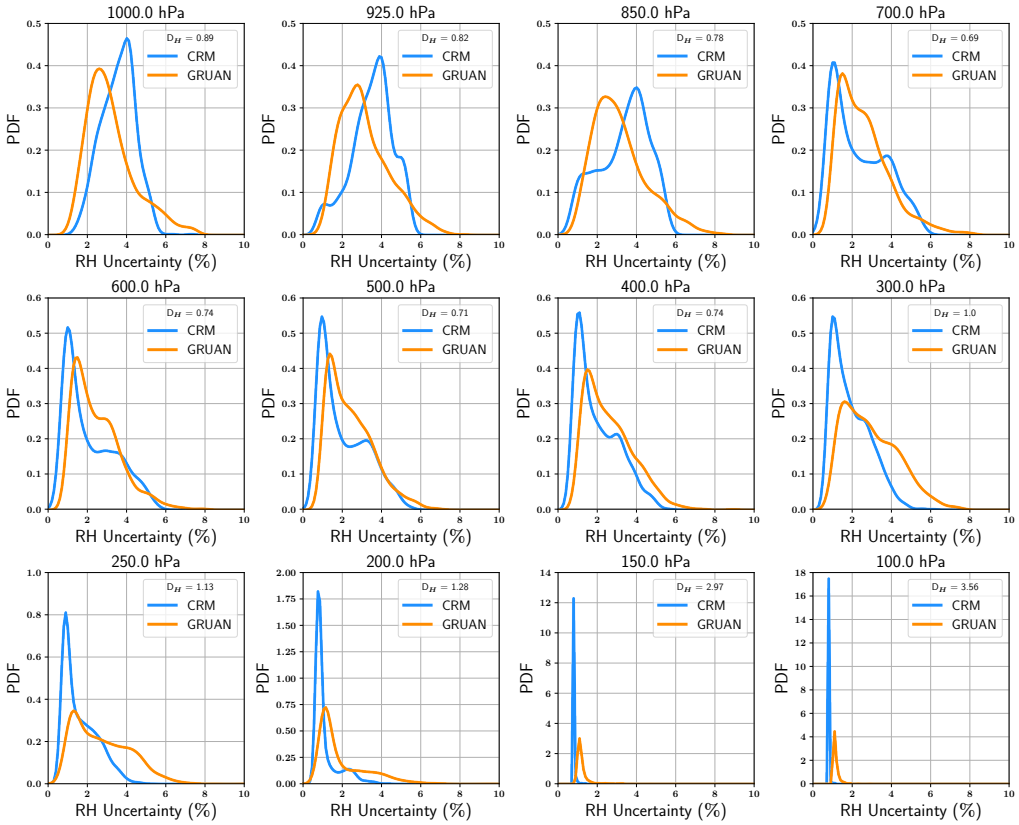


Figure 3. PDF distribution of RH uncertainties for CRM (blue) and GRUAN (orange) for the lowest 12 standard AIRS pressure levels. Data is taken from common soundings found in both data bases that cover 4 sites. The Hellinger distance (D_H), a measure of the similarity between the 2 PDFs is also displayed. At 250 hPa the 2 distributions start to diverge.

$$D_H = \frac{1}{\sqrt{2}} \sqrt{\sum_{i=1}^{N} \left(f(U_{CRM})_i - f(U_{GRUAN})_i \right)^2}. \tag{8}$$

The Hellinger distance (Hellinger, 1909) is a metric used to measure the difference between PDF distributions, and is a probabilistic analogue of the Euclidean distance. A $\sqrt{2}$ is included to allow the distance values to range between 0 and 1. Where values exceed 1, the two distributions are completely different or non-comparable.

From examination of Figure 3, RH uncertainty distributions of CRM and GRUAN visually look reasonably comparable. In the lowest 3 levels (1000-850 hPa) there are distinct offsets between the separate PDF modal values with CRM being consistently higher than GRUAN. Between 700-400 hPa the bi-modal distribution in both uncertainty PDFs is similar as the Hellinger distances here are at their lowest. At 300 hPa CRM uncertainties begin to display a reduction in RH uncertainties between 4-8% relative to GRUAN. Here the Hellinger distance changes from ≈0.7 at the mid-troposphere to 1.0. At pressures 250-100 hPa the distributions diverge further and become distinctly different with values exceeding 1 (1.13, 1.28, 2.97 and 3.56 respectively). Here in the UTLS GRUAN displays a consistent higher RH uncertainty values in the trailing edge of the PDF.

4 Selection of AIRS Water Vapor Profiles for Validation

Having discussed water vapor profiles in terms of RH so far, the analysis now switches to absolute concentrations. AIRS and radiosonde water vapor profiles are first converted mixing ratios, with the approach outlined in Nalli et al. (2013) used for AIRS. The volume mixing ratio (VMR) is then calculated by multiplying the mixing ratio by the ratio of molecular weight of water to dry air. Due the large range of water vapor VMR values found in the troposphere results are reported as scaled fractions of the absolute concentration, or % ppmv.

4.1 Collocation and Convolution of Water Vapor Profiles

The approach used by this study considers an initial broad collocation criteria, with match-ups being removed through a secondary filtering process. For AIRS, a water vapor profile is considered collocated with a CRM or GRUAN sounding if the satellite measurement falls within ±3 hours and 100 km of the radiosonde launch. This approach yields 1500-3000 monthly collocations with CRM, and 100-300 collocations with GRUAN after 2009. With the use of broad criteria, any mismatch introduced in the collocation will impact on the performance of individual comparisons (Sun, Reale, Seidel, & Hunt, 2010; Sun et al., 2017). The impact of selected collocation temporal and spatial is in itself a large topic and goes beyond the scope of this study. We do however investigate and discuss the impact of our match-up window on the collocation uncertainty in section 5.3.

Profiles retrieved from AIRS (or any other remote sensing instrument) are not simple measurements or representations of the atmospheric state rather a best estimate to which a smoothing function has been applied (Rodgers & Connor, 2003). The averaging kernel is used to smooth (or convolve) the radiosonde profile to the vertical resolution of AIRS:

$$\mathbf{x}_{est} = \mathbf{x}_o + \tilde{\mathbf{A}}(\mathbf{x}_t - \mathbf{x}_o), \tag{9}$$

where $\mathbf{x}_o$ is the AIRS first guess profile, $\tilde{\mathbf{A}}$ is the averaging kernel that has been interpolated onto the 100 level retrieval grid using the trapezoid functions, $\mathbf{x}_t$ is high resolution reference profile on the AIRS 100 level grid, and $\mathbf{x}_{est}$ is the convolved reference profile. Because changes in column density of water vapor in the thermal infrared (TIR) have greater linearity in log space relative to any absolute change, equation 9 is expressed as (Maddy & Barnet, 2008):

$$\ln(\mathbf{x}_{est}) = \ln(\mathbf{x}_o) + \tilde{\mathbf{A}} \times \ln\left(\frac{\mathbf{x}_t}{\mathbf{x}_o}\right). \tag{10}$$

Once the matched radiosonde profile has been convolved three quality flags from the AIRS product are applied to the the radiosonde and AIRS water vapor profile. The first two flags define the upper and lower bounds of the retrieved AIRS profile. The PGood flag removes levels from upper atmosphere which do not meet quality control criteria, while the nGoodSup index flag defines the lowest surface level in the AIRS retrieval grid to be used. The Qual_H2O flag describes the overall quality of the water vapor field, where the 'Best' retrievals have a Qual_H2O of 0, 'Good' retrievals have a Qual_H2O of 1 and 'Do not use' for Qual_H2O values of 2 (E. Olsen et al., 2013; E. T. Olsen et al., 2007). Individual values are then masked in both profiles at levels where:

1. The pressure index is greater than the PGood flag.
2. Pressure levels below the nGoodSup index.
3. Qual_H2O is equal to 2.
4. AIRS retrieval uncertainty is greater than 25% ppmv.

5. Radiosonde uncertainty is greater than 20% ppmv .
6. AIRS profile values below 20 ppmv. E. J. Fetzer et al. (2008) demonstrated the low end sensitivity limit of AIRS to absolute amounts of water vapor to be around 15-20 ppmv in the UTLS. Values below this threshold are masked from the analysis.
7. Cloud fraction inside AIRS FoR is above 80%. Here we use a secondary independent estimate of cloud fraction from the Moderate Resolution Imaging Spectroradiometer (MODIS) which is collocated on-board the Aqua platform with AIRS.

This process on average removes 82% of the collocations. However, the initial collection of match-ups yields large numbers e.g. 1,301,655 of pairs at CRM sites. Therefore, after filtering large numbers of matched pairs still remain.

5 Results

This section presents the results of the AIRS match-ups to the 2 characterised radiosonde databases. The first set of results are shown as 'global' values or the total result over all sites. CRM results are further filtered to only include available GRUAN sites to investigate the effects of differing global sampling. The global sampling of CRM is further used to understand AIRS water vapor biases as a function of latitude before finally providing the first estimates of collocation uncertainty between these match-ups. All results are reported as % of absolute concentrations of water vapor in ppmv.

The bias (Δq) between AIRS and the radiosonde water vapor profiles is defined as:

$$\Delta \mathrm{q} = 100 \times \left(\frac{(\mathrm{q_A} - \mathrm{q_R})}{\mathrm{q_R}} \right), \tag{11}$$

where $\mathrm{q_A}$ and $\mathrm{q_R}$ are the AIRS and convolved radiosonde water vapor profile measurement respectively. Next the modified z score (Iglewicz & Hoaglin, 1993) provides a final filter that is applied to the data in order to identify outliers:

$$\mathrm{z} = \frac{0.6745 \times (\Delta \mathrm{q} - \mu_{\Delta \mathrm{q}})}{\sigma_{\Delta \mathrm{q}}}, \tag{12}$$

where $\mu_{\Delta \mathrm{q}}$ is the median AIRS layer difference and $\sigma_{\Delta \mathrm{q}}$ is the median absolute difference (MAD) of the AIRS layer differences. If the minimum calculated z value is greater than the theoretical maximum value $\left(\frac{\mathrm{N}-1}{\sqrt{\mathrm{N}}}\right)$, where N is the number of data points, then those points are considered non-Gaussian and a default bad data flag is assigned to that layer. All points whose |z| score is greater than 3.5 (recommendation by Iglewicz and Hoaglin (1993)) are removed. Both the measurements contribute a uncertainty for each point in the profile, so the uncertainty of the bias $\left(\mathrm{U}(\delta \mathrm{q})\right)$ is calculated using classical error propagation methods such that:

$$\mathrm{U}(\delta \mathrm{q}) = \left[\left(\frac{\sqrt{\mathrm{U}(\mathrm{q_A})^2 + \mathrm{U}(\mathrm{q_R})^2}}{\mathrm{q_A} - \mathrm{q_R}} \right)^2 + \left(\frac{\mathrm{U}(\mathrm{q_R})}{\mathrm{q_R}} \right)^2 \right]^{\frac{1}{2}} \times |\Delta \mathrm{q}|, \tag{13}$$

where $\mathrm{U}(\mathrm{q_A})$ and $\mathrm{U}(\mathrm{q_R})$ are the AIRS retrieved profile (Susskind, Blaisdell, & Iredell, 2014) and characterised radiosonde profile uncertainty (Immler et al. (2010) in the case of GRUAN) respectively. Finally any propagated uncertainty is calculated as the linear combination of the average bias uncertainty and standard error:

$$\mathrm{U}(\Delta \mathrm{q}) = \left[\left(\frac{1}{\mathrm{N}} \left(\sum_{i=1}^{\mathrm{N}} \mathrm{U}(\delta \mathrm{q})_i^2 \right)^{\frac{1}{2}} \right)^2 + \left(\frac{\sigma_{\Delta \mathrm{q}}}{\sqrt{\mathrm{N}}} \right)^2 \right]^{\frac{1}{2}}. \quad (14)$$

5.1 Global Results

Table 2 presents the statistics from AIRS comparisons averaged over all available CRM and GRUAN sites along with a subset of CRM sites. Median biases are calculated on layers between adjacent levels in the AIRS standard L2 up to 250 hPa. This level was chosen as above this altitude the CRM and GRUAN uncertainties are non-comparable as evidenced by Figure 3. Results are split into three match-up scenarios:

1. AIRS biases relative to all collocated GRUAN soundings (AIRS-GRUAN). These results indicate that both AIRS and GRUAN agree within 1.5 % ppmv for 4 lowest layers. This suggests i) AIRS data is suitable for climate records globally, ii) GRUAN data is very good for absolute bias.
2. AIRS biases relative to all collocated CRM soundings (AIRS-CRM). While biases to CRM are higher than those relative to GRUAN they are generally less than ±15 % ppmv, peaking between 600-500 hPa. While AIRS biases relative to CRM appear higher, the global distribution of CRM stations allows assessment of latitudinal biases and the spread of AIRS data (GRUAN limits precision of comparison).
3. AIRS biases at collocated CRM stations that are common to both CRM and GRUAN (AIRS-CRM$_G$). It is important to note that the AIRS-GRUAN and AIRS-CRM$_G$ do not contain exactly the same collocations due to differences in the 2 records i.e. common stations not soundings. However, what it does provide is 'GRUAN-like' sampling of CRM to assess the impact of differences in global sampling.

All-sky biases were also examined as a function either very dry (XLow) or very wet (XHigh) atmospheres. The definition comes from Roman et al. (2016), where XLow and XHigh refer to TCWV thresholds of less than 5 kg m$^-$2 and greater than 50 kg m$^-$2 respectively. For collocations in very dry atmospheres AIRS shows a strong dry bias between 700-300 hPa relative to GRUAN. A similar pattern is seen for CRM (& CRM$_G$) matches though this is relative to the general all-sky bias and manifests as a reduction in the wet bias magnitude. For collocated AIRS profiles at both CRM and GRUAN sites in very wet atmospheres performance is within ±5 % ppmv in the mid-to-lower troposphere. However, for GRUAN sites the uncertainty increases by an order of magnitude which can be attributed to the lower number of matched profiles.

Results from these three scenarios are also split in to clear-sky matches and all-sky diurnal components in Table 2. Clear-sky biases consistently display relatively wetter biases compared to their all-sky, with the largest disparities seen in the mid-troposphere. E. J. Fetzer, Lambrigtsen, Eldering, Aumann, and Chahine (2006) demonstrated that the the AIRS infrared radiances only improved TCWV estimates by a few percent, therefore the first AMSU first guess profile will still be dominating in clear-sky scenes. All three scenarios see drier daytime biases for the lowest two surface layers. In the mid-troposphere (700-300 hPa) AIRS is generally wet-biased relative the radiosonde measurements from both records. These biases are stronger for CRM comparisons, however this is expected as CRM has lower precision relative to GRUAN in this region (Table 1).

Finally, it is important to note the 250-300 hPa layer results for AIRS-CRM which has the strongest bias from results presented in Table 2. Comparisons to CRM show an order of magnitude difference from AIRS-CRM$_G$ results and GRUAN. This layer at mid-to-high latitudes is sensitive to the tropopause where water vapor values will fall to ≈5

Table 2. AIRS median biases, defined as a percentage of the absolute H_2O concentration (% ppmv) for global comparisons at CRM and GRUAN radiosonde sites. Layers are defined by the AIRS standard L2 pressure grid. For each layer the associated uncertainty is also reported in % ppmv. Reported biases are split into six categories, i) all-sky matches (All), ii) clear-sky where cloud fraction is <1% (Clear), iii) Daytime matches (DAY), iv) night-time matches (NGT), v) extreme low TCWV (XLow) and vi) extreme high TCWV (XHigh). The thresholds for XLow and XHigh are taken from Roman et al. (2016) and are TCWV < 5 kg m^2 and TCWV > 50 kg m^2 respectively. Biases for CRM sites that are common to both CRM and GRUAN records are shown and are denoted by the subscript 'G' (CRM_G). Note CRM_G has no XHigh values as sites are restricted to mid-to-high latitudes in the northern hemisphere. Biases are reported up to 250 hPa as CRM and GRUAN uncertainties are no longer compatible above that level (Figure 3). Also included is the percentage of data points that were rejected during the filtering stage after collocation. The initial number of matched pairs are 82,138 with GRUAN, 1,301,655 with CRM and 62,366 with CRM_G sites. Missing data is represented by a dash (—).

		Pressure Layer (hPa)							
		250-300	300-400	400-500	500-600	600-700	700-850	850-925	925-1000
AIRS-GRUAN (% ppmv)	All	0.24±0.43	-4.92±0.34	-2.93±0.30	-5.70±0.31	1.03±0.29	1.06±0.22	1.43±0.15	-0.66±0.16
	Clear	-0.83±0.98	-3.61±0.75	-1.26±0.71	-2.63±0.75	6.33±0.84	3.86±0.61	1.45±0.4	-0.38±0.48
	DAY	2.64±0.63	-2.01±0.47	-2.35±0.39	-5.77±0.39	3.51±0.41	0.91±0.35	-1.47±0.26	-3.81±0.24
	NGT	-1.69±0.60	-8.08±0.48	-3.76±0.48	-5.42±0.49	-1.74±0.41	1.18±0.28	2.92±0.19	2.06±0.22
	XLow	-0.11±0.84	-4.79±0.61	-6.59±0.45	-13.25±0.46	-7.65±0.41	-1.88±0.36	-0.75±0.27	-0.6±0.26
	XHigh	-4.72±4.29	-7.37±4.3	-3.92±3.95	4.25±3.31	6.5±2.85	3.18±2.3	-2.2±1.8	—
	Rejected (%)	82.63	81.47	81.69	82.26	80.9	81.9	73.68	76.1
AIRS-CRM (% ppmv)	All	-21.74±0.19	0.39±0.08	9.74±0.08	14.92±0.09	14.19±0.08	5.48±0.05	-0.51±0.04	-2.55±0.03
	Clear	-15.22±0.45	5.14±0.22	14.47±0.23	21.49±0.26	22.43±0.25	9.82±0.17	0.63±0.11	-1.66±0.09
	DAY	-24.61±0.24	0.74±0.10	11.73±0.1	17.57±0.11	16.39±0.1	5.41±0.07	-2.35±0.05	-4.42±0.04
	NGT	-16.23±0.29	-0.29±0.15	4.9±0.15	8.03±0.16	9.25±0.14	5.60±0.09	1.84±0.06	1.40±0.05
	XLow	-6.28±0.24	1.34±0.13	1.35±0.12	2.22±0.14	3.36±0.13	2.15±0.1	-0.52±0.08	-1.73±0.06
	XHigh	-21.49±1.0	-5.18±0.53	4.67±0.49	0.9±0.42	-1.65±0.34	-3.4±0.29	-4.86±0.21	-5.01±0.17
	Rejected (%)	94.26	85.67	79.28	78.22	76.71	79.23	71.85	58.5
AIRS-CRM_G (% ppmv)	All	0.13±0.67	5.75±0.41	7.72±0.43	11.08±0.49	17.08±0.52	10.50±0.43	4.10±0.31	0.90±0.21
	Clear	-7.2±1.72	5.85±1.11	16.69±1.27	28.14±1.49	38.63±1.65	17.38±1.17	5.52±0.79	2.13±0.58
	DAY	0.37±0.73	6.46±0.45	8.43±0.47	11.86±0.53	18.96±0.61	10.76±0.55	1.91±0.40	-1.57±0.23
	NGT	-1.73±1.60	1.95±0.97	3.85±1.05	6.23±1.24	10.10±0.96	9.78±0.71	7.10±0.51	8.32±0.43
	XLow	3.15±0.85	6.15±0.5	4.59±0.48	5.72±0.54	9.46±0.6	8.29±0.54	4.15±0.4	3.46±0.3
	XHigh	—	—	—	—	—	—	—	—
	Rejected (%)	96.67	88.49	86.49	86.55	87.62	91.23	88.56	76.7

manuscript submitted to *JGR: Atmospheres*

ppmv, well below AIRS sensitivity to absolute concentration of 15-20 ppmv (E. J. Fetzer et al., 2008).

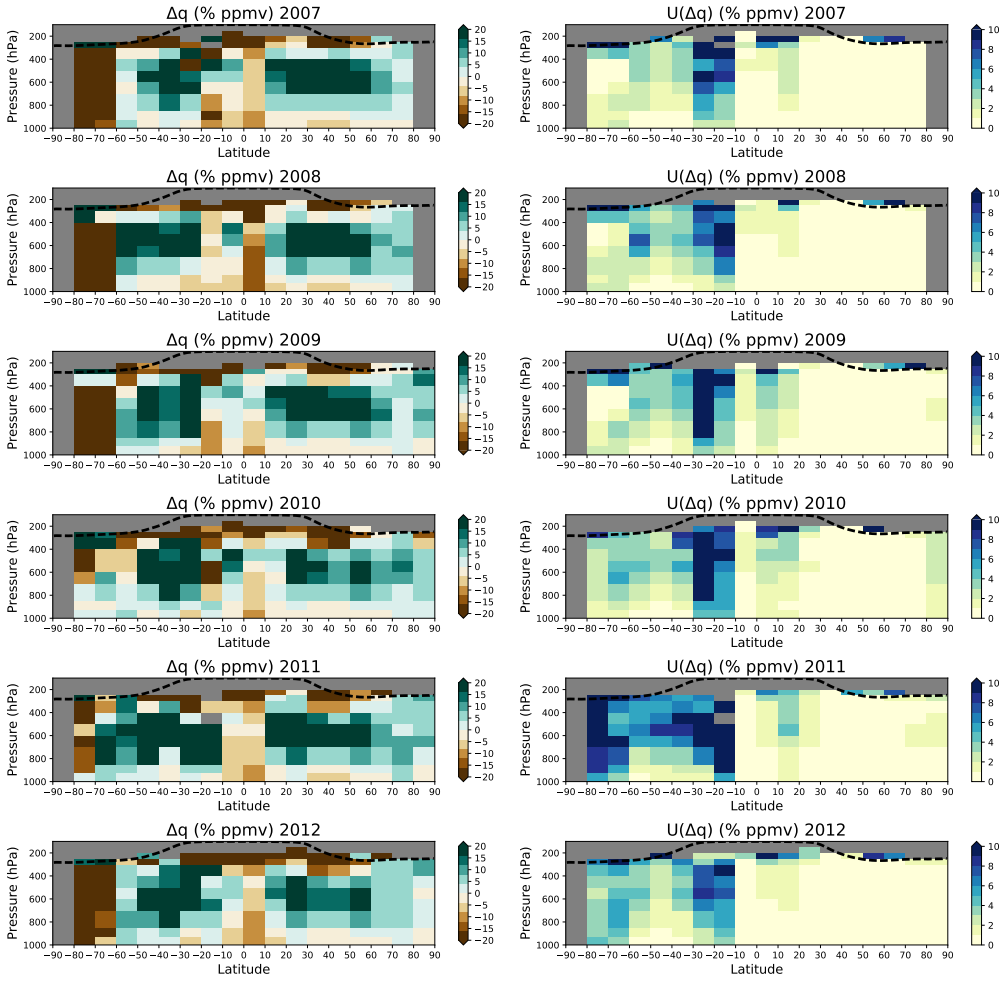


Figure 4. Latitudinal distribution of AIRS yearly median bias (Δq) and the median bias uncertainty (U(Δq)) relative to CRM in % ppmv. The black dashed line is a climatological cold point tropopause calculated from AIRS L2 data.

5.2 Latitudinal dependence

Figure 4 shows median biases for AIRS relative to CRM as a function of latitude, altitude and year with the associated uncertainty. Here the relative global sampling of CRM can be used to examine spatial dependence of AIRS biases. Collocations are first grouped into yearly results before binned by latitude every 10 degrees between 90° south to 90° north for each atmospheric layer. Outside the tropics the lack of results in UTLS are due to radiosonde profiles not going higher than 200 hPa. This is because of poor sampling (especially in the southern hemisphere), and the filtering steps applied to the collocated profiles. While in the tropics there are some data for 200 hPa, section 3.2 demonstrated low confidence in results at pressure levels less than 250 hPa. Examination of the biases in this way highlight sensitivity in 4 key areas:

1. Deep convective regions. AIRS biases as a function of latitude show a general sensitivity to large scale atmospheric circulation. All years exhibit a predominant dry bias between 0° and 10° north that extends from the surface to the upper troposphere. The lowest range of dry biases are seen in 2010 where the column values are between -8.85±1.14 % ppmv and -21.92±7.49 % ppmv (surface to 250 hPa). These values are representative of all other years except for 2008 where dry biases at this latitude are higher, ranging between -12.28±0.54 % ppmv and -49.31±1.47 % ppmv in the upper troposphere.
2. Hadley cell regions. Situated either side of the observed dry bias in the tropics are mid-latitudinal regions of wet bias situated in the centre of the Hadley cell region. Although these areas appear to be clearly defined in the northern hemisphere the highest values are observed in the south. The 'wettest' biases of 48.08±8.7 % ppmv is observed in 2011 at 500-600 hPa between 30°S to 40°S, though all years see regions where biases exceed 30 % ppmv.
3. The southern hemisphere below 60°S. At high latitudes in the southern hemisphere AIRS is extremely dry-biased compared to CRM. Sampling in this region catches stations on the Antarctic coast, South America and parts of Australia. These dry biases are predominately higher than other observed values and typically range between -25 % ppmv and -95 % ppmv. While these biases are high, for the southern hemisphere the uncertainty on the bias (with the exception of 2011) is below 6% ppmv while on average below 3% ppmv. This region performs better than the mean uncertainty for latitudes between 30° and 10°S. Below 70°S in atmospheric layers sensitive to the UTLS the strong dry bias changes to a high wet bias.
4. The UTLS. The final notable region involves the 250 to 300 hPa layer that is sensitive to the UTLS, between ±60°. At these altitudes, depending on latitude there is a varying amount of stratospheric information included in the tropospheric retrievals. Another key aspect to also consider is the lower confidence in the corrected CRM radiosonde profiles at these altitudes. In conjunction with the sensitivity of AIRS to low concentrations of water vapor (E. J. Fetzer et al., 2008) make this region challenging.

While comparing one year to another one final aspect to consider is sampling. Figure 5 shows the fractional difference of sampling within the 10° bins relative to 2007. In the majority of bins collocation numbers to CRM are lower than those in 2007. There are a few exceptions, in particular in 2008 there are almost twice as many collocations than in 2007 between 10°N to 20°N. The manifestation of variable sampling can be seen how the uncertainty in the median bias reduces. The right-hand-side of Figure 4 shows these uncertainties distributed as a function of latitude and pressure. The first thing that is noticeable is that there is a hemispherical split in the uncertainty as the higher number of northern hemisphere station results in a majority of uncertainty values less than 1 % ppmv. The second noticeable feature is a persistent region of high uncertainty values between 30°S and 10°. In 2008 there is significant higher sampling relative to 2007 where the uncertainties vary from 0.92 to 2.85 % ppmv, compared to to 2011 where sampling is significantly lower. Here the range in uncertainty varies between 1.44 and 4.74 % ppmv. When considering GCOS-200 (GCOS, 2016) targets, consistent global sampling will be key to ensure the robust assessment of long-term stability in satellite records.

5.3 Collocation Uncertainty

A key advantage of using operational radiosonde networks is that they generate large numbers of matches relative to collocations performed at climate networks like GRUAN. With these larger numbers there is a higher confidence that the true bias is within the uncertainty of the reported median bias. However, the calculated uncertainty only partially describes the calculated bias. As the collocation framework is imperfect the uncertainty arising from the match up process also needs to be assessed. Characterising

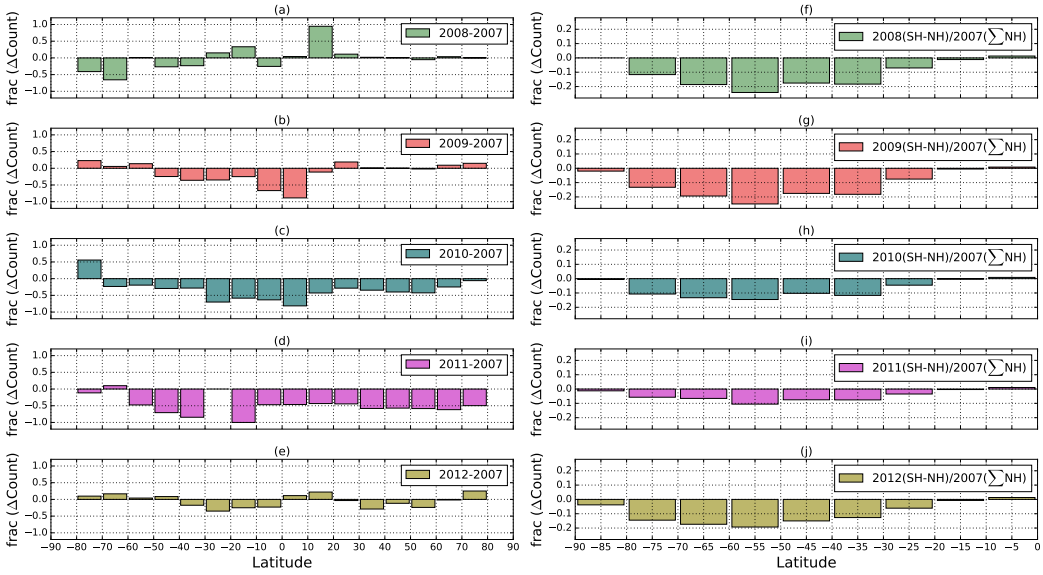


Figure 5. Fractional difference in CRM collocations relative to 2007 as a function of latitude (a-e). Bars represent the number of collocations before filtering in each 10° bin relative to 2007. The lower number of collocation during 2011 are disproportionately higher in the southern hemisphere. The right-hand-side (f-j) shows the fractional difference between sampling in the southern and northern hemispheres normalised by the total northern hemisphere sampling for 2007. Data again is binned in 10° latitudinal intervals from 0° to 90°S. Again 2011 is highlighted, as the lower number in total global soundings results smaller relative differences between the north and south.

the collocation uncertainty is highly non-trivial and a focus for large European projects like Fidelity and uncertainty in climate data records from Earth Observations (FIDUCEO) and the Gap Analysis for Integrated Atmospheric ECV CLImate Monitoring (GAIA-CLIM). In this study we employ the consistency test from Immler et al. (2010), which is also used in the NOAA Products Validation System (NPROVS) (Reale, Sun, Tilley, & Pettey, 2012) to estimate the collocation uncertainty (σ):

$$|\mathrm{m}_1 - \mathrm{m}_2| < \mathrm{k} \cdot \sqrt{\sigma^2 + {\mathrm{u}_1}^2 + {\mathrm{u}_2}^2}, \tag{15}$$

where m_1 and u_1 are the AIRS retrieved water vapor and retrieval uncertainty respectively, and m_2 and u_2 are the radiosonde measured water vapor and measurement uncertainty respectively. The coverage factor, k is the interval about the mean value as a multiple of the standard uncertainty (Immler et al., 2010). It should be noted that any error in the estimations of u_1 and/or u_2 would impact the collocation uncertainty, i.e., all unknowns, all unclassified and inconsistencies would impact the collocation uncertainty.

Cases are considered consistent where k is less or equal to 1 and in agreement when k is less or equal to 2. Therefore, to calculate σ three assumptions are applied to the data:

1. The median bias is a robust estimate of the theoretical bias due to large number of collocations, which are in the order of 10^4 to 10^5 matches for the majority of bins.

2. The median bias is also agrees to some degree within the uncertainty, i.e. k ¡ 2.
3. The uncertainties u_1 and u_2 correctly and fully describe the uncertainties of the two data records.

Firstly equation 15 can be rewritten as:

$$\delta \mathrm{m} = \mathrm{k}\left(\sigma^2 + \mathrm{U}_{\delta \mathrm{m}}^2\right)^{\frac{1}{2}}, \tag{16}$$

where $\delta \mathrm{m} = |\mathrm{m}_1 - \mathrm{m}_2|$ and $\mathrm{U}_{\delta \mathrm{m}} = \left({\mathrm{u}_1}^2 + {\mathrm{u}_2}^2\right)^{\frac{1}{2}}$. Next equation 16 is then rearranged to make σ the subject:

$$\sigma = \left[\left(\frac{\delta \mathrm{m}}{\mathrm{k}}\right)^2 - \mathrm{U}_{\delta \mathrm{m}}^2\right]^{\frac{1}{2}}. \tag{17}$$

The collocation criteria for this study uses a basic set of parameters to match AIRS retrievals with CRM soundings. These can lead to large uncertainties in the collocation especially in regions of high variability, or by where the collocation is sampling completely different areas. Through averaging large numbers of matches the collocation uncertainty will reduce in the same manner as the uncertainty on the bias. In this study we then make the assumption that for the yearly statistics to be valid estimates for AIRS performance then they must also be either consistent (k = 1) or within agreement (k = 2) (Immler et al., 2010). Therefore, by substituting the values of 1 and 2 for k, σ can then be calculated. These results are shown in Figure 6 with the calculated collocation uncertainty assuming k is equal to 1 on the left-hand-side and for when k is equal to 2 on the right-hand-side.

Under both these assumptions collocation uncertainty estimates in the northern hemisphere reduce to below 2 % ppmv within the troposphere (at altitudes below 250 hPa), with most values below 1 % ppmv. The higher region of bias uncertainty seen between 30°S to 10°S (Figure 4) is mirrored in the collocation uncertainty, especially so for the $k = 2$ hypothesis. All years see collocation uncertainties exceeds 20 % ppmv in this region, with the highest value seen in 2011 (70 % ppmv). The southern hemisphere is noticeably different in all years for both assumptions of k. The lower number of stations (2) below the equator introduces structure into the collocation estimates with the higher values seen below -60° and above -30°. Between these latitudinal zones higher values are seen above 800-700 hPa, whereas nearer the surface values are close to their northern hemisphere counterpart. This lower surface collocation uncertainty is concentrated in the zone where longitudinally there is a higher density of stations. On the whole, collocation uncertainty in the southern hemisphere halves when the assumption on k is relaxed from 1 to 2. The biggest impact on the collocation uncertainty is the sampling. The UTLS and 2009 Antarctic regions see the highest collocation values between 32 and 47 % ppmv, with UTLS uncertainties not restricted to a single hemisphere. For 2011 the majority of all regions exhibit lower numbers of collocations (Figure 6), though this is not evenly distributed. For the UTLS confidence in CRM corrections falls relative to GRUAN, therefore some of the collocation uncertainty could also be attributed to unknown additional sources of bias not characterised during the CRM corrections.

A further example of how the collocation uncertainty is effected by sampling is shown in Figure 7. For this experiment the temporal and spatial criteria for match-up selection are halved to 50 km and ±1.5 hours respectively, and σ_{k1} & σ_{k2} recalculated. These new values are then subtracted from those shown in Figure 6, with the result represented by Figure 7. The stricter match-up criteria reduces the number of collocations to ≈10% (and lower) of the original values. This has the immediate impact in the region between

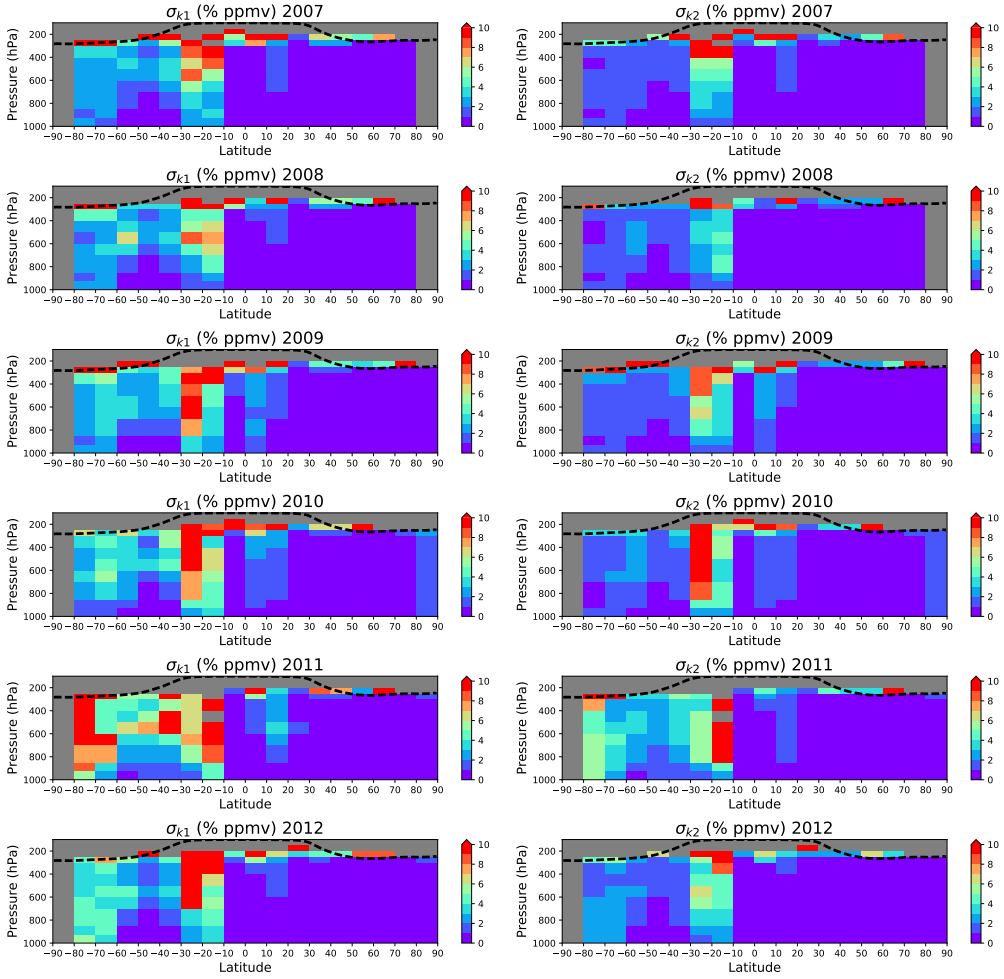


Figure 6. Yearly latitudinal distribution of collocation uncertainty (σ) between AIRS and CRM water vapor profiles in % ppmv for two k statistic values. Plots on the left hand side of the figure are the collocation uncertainties for the assumption that k = 1 (σ_{k1}), while the plots on the right assume k = 2 (σ_{k2}). Like Figure 4 the black dashed lines are the climatological cold point tropopause.

60°S and 20°N, where the sampling density is low. With fewer match-ups there is either i) a significant increase in σ_k due to a reduction in signal-to-noise, or ii) a higher frequency in missing data relative to the 100 km and ±3 hour window. In the Northern Hemisphere, the lower signal-to-noise causes σ_k to increase by less than 1 % ppmv on average for both k hypothesises. For match-ups below 60°S, there is a average reduction in σ_k between 1-2 % ppmv for most years. Collocations in this region come from Antarctica stations, which are mainly sited near the coast (Figure 2). This tighter match-up criteria acts to remove the majority of AIRS soundings over the Southern Ocean, which were inflating the collocation criteria.

Ensuring that both the reference measurement, which is made over 90-120 minutes, and the almost instantaneous satellite observation are representative of one another is the main challenge when validating water vapor profile retrievals using radiosondes. With upper-atmosphere soundings launched at synoptic times (e.g. 00:00, 06:00, 12:00 & 18:00 hrs) finding large numbers of coincident satellite overpasses requires broad criteria. These

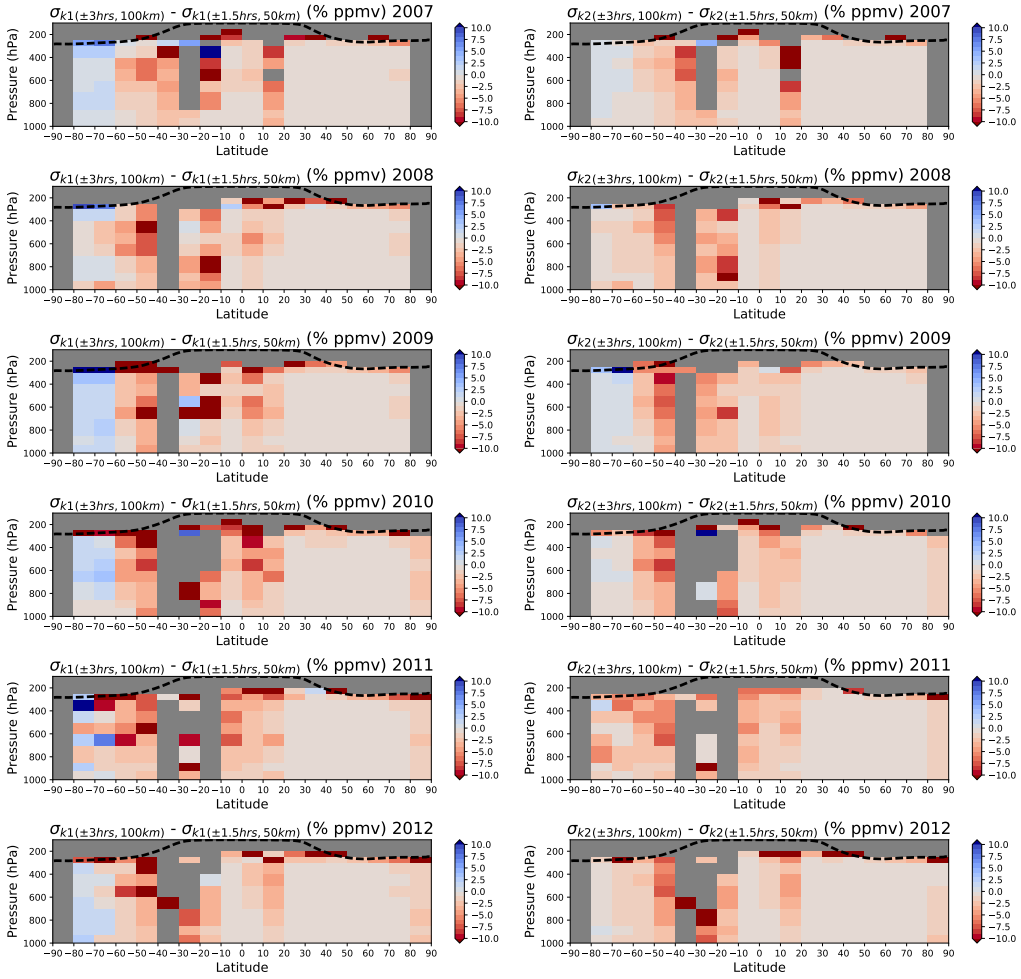


Figure 7. Difference between collocation uncertainties shown in Figure 6 and their recalculated values when the collocation criteria are tightened to match-ups within 50 km and ±1.5 hours of launch. Reduction in the number of cases found for the stricter criteria results in higher frequency of missing data (grey regions), especially between 40°S and 20°S.

results demonstrate that we can minimise the impact of σ_k if take large averages are used, which is fine when investigating yearly to decadal biases or stability. To understand performance on a sounding-by-sounding basis then capturing and characterising the full representativeness of each collocation is needed. This has advantages, especially when collocating with Southern Hemisphere sites. The complexity of this approach goes beyond the scope of this study but remains a topic of interest to G-VAP.

6 Discussion and Conclusions

With over a decade of measurements, AIRS can provide the first Climate Data Record (CDR) of high-resolution humidity profiles from the new generation of hyper-spectral IR sounders. The provision of averaging kernels and other retrieval statistical outputs in the AIRS L2 support product has allowed us to assess the performance of AIRS water vapor in a mathematically and physically sound way not yet demonstrated in other studies. For this study we have used 2 characterised radiosonde data sets. While both

are accurate, one has higher performance (low bias) and the other provides greater latitudinal distribution. Unique to this study we use the uncertainties to help characterise the performance of AIRS water vapor biases, and apply the AIRS averaging kernels to the radiosonde profiles. Commonly used in trace gas validation, the application of averaging kernels in the convolution of a reference profile allows for like-for-like comparisons of the collocated AIRS water vapor observations.

In an initial study by Hagan et al. (2004), balloon soundings and aircraft measurements were used to look at AIRS version 3 retrieved water between 500-100 hPa. These first results found that AIRS had RMSE values of 25% or better for closely matched observations (within 1 hour). This result was a significant improvement over the beta validation estimate of 50.5% (E. Fetzer et al., 2003) which used operational radiosondes. Studies using AIRS version 4 water vapor retrievals found profile biases were better than ±15% when compared to a 2-year study with operational radiosondes (Divakarla et al., 2006). However, comparisons at dedicated sites from the study by Tobin et al. (2006) showed that profile biases were within ±5% below 400 hPa, with an increasing dry bias between 400-200 hPa of -10%. These biases were reported to be largely independent of cloud fraction. Our analysis of AIRS version 6 water vapor profiles with 6 years of GRUAN radiosondes shows improved performance relative to these studies. In all-sky conditions biases are below ±1.5 % ppmv between the surface and 600 hPa, while in clear-sky conditions (cloud fraction <1%) there is an observed increased wet bias relative to GRUAN. This behaviour is also observed with matches at CRM sites, though it is more pronounced with clear-sky wet biases exceeding 20 % ppmv. It could be thought that because CRM has denser sampling of very wet atmospheres that this would be responsible for the increased wet bias. However, analysis of AIRS performance in extreme wet and dry atmospheres in Table 2 does not suggest this as the sole reason as:

1. GRUAN shows AIRS to have a small dry bias in the mid-to-lower troposphere in very dry water vapor columns and dry bias that increases in altitude for wet water vapor columns.
2. Comparisons to CRM in the same region exhibit drier biases for the lower troposphere for both extremes, with dry biases in wet atmospheres and wet biases in dry atmospheres.
3. For the upper troposphere, CRM shows AIRS to be predominately wet-biased for both high and low TCWV thresholds, while the inverse is true for GRUAN.

Therefore, this is suggestive that regions where TCWV amounts are between 5-50 kg m^{-2} the differences in regime variability influence the larger observed bias values. The diurnal split for all scenarios in Table 2 confirms this with daytime matches showing wetter biases compared to night-time matches. Another point to consider here is that CRM also under-samples the vertical H_2O profile relative to GRUAN, which could also be introducing a bias into the comparisons.

A recent study by Wong et al. (2015) used collocated cloud properties from MODIS (Aqua) to assess cloud-induced uncertainties. This study demonstrated that AIRS has an increasing dry bias (-15 to -40%) as a function of cloud optical depth in the lower troposphere and that the largest dry biases (< -30%) occur in either high or low thick clouds. These dry biases occur around the same altitude as the cloud top pressure. They also showed that the AIRS physical retrieval actually reduced dry biases in the first guess in the presence of thin cloud. Further comparisons against reanalysis (Hearty et al., 2014) also attributed large dry biases (< -30%) in AIRS version 5 retrievals over mid-latitude storm tracks and deep convective clouds. Over stratus regions however, the authors found wet bias reaching up to 20%. Yue et al. (2013) confirm this in their study where they show that the total biases are dominated by cloud state-dependent sampling e.g. in deep convective regions. In the version 6 we still see sampling bias in convective regions, with strong dry biases located within the 'convective pump' straddled (north and south) by

a mid-tropospheric wet-biased regions. In this region we found AIRS is wet-biased by 1.07±3.42 and 49.27±19.90 % ppmv to the descending air mass around the Hadley cells, while being dry-biased to the ascending air mass at the equator where the bias can exceed 80 % ppmv. Understanding these biases over climate time scales is vital for studies into water vapor feedback. In particular the largest contribution to the water vapor feedback stems from the tropical free troposphere, and the OLR is highly sensitive to changes at the dry end of the humidity distribution e.g. in regions of descending air masses around the Hadley cells.

The assessment of satellite retrieved water vapor profiles for climate purposes poses a real challenge. The example of GCOS-200 (GCOS, 2016) target requirements (5% measurement uncertainty and a stability 0.3% per decade) are still optimal requirements as they have yet to be fully demonstrated. Individual profiles from satellite or radiosonde measurements can have relatively high associated uncertainties in addition to the representativeness of the collocation. In this study we use averaging over time scales of one and six years to reduce random uncertainty and collocation components. We assumed that averaging over large numbers of matches allows the collocation uncertainty to be considered consistent (Immler et al., 2010), even when we use broad collocation criteria. However, it has been shown by Pougatchev et al. (2009) that collocation criteria of 30 minutes and 25 km or better are needed in order to obtain differences between radiosonde and IASI data within GCOS requirements. The implication of this would greatly reduce the number of available collocations for analysis, with the match-ups in the southern hemisphere all but removed.

A new approach to estimate the collocation uncertainty was introduced by Kinzel et al. (2016) and refined in Liman, Schröder, Fennig, Andersson, and Hollmann (2018). They estimate the collocation uncertainty as a function of TCWV, near-surface specific humidity, near-surface wind speed and sea surface temperature (SST). The global mean collocation uncertainty is approximately 0.5 g kg^{-1} (ranging from 0.4 to 0.7 g kg^{-1}, Liman et al. (2018)) for the Hamburg Ocean Atmosphere Parameters and Fluxes from Satellite data (HOAPS, Andersson et al. (2017)). In this study the collocation uncertainty was estimated on yearly errors and thus shows smaller values. Assuming default climatological surface humidity values (Remedios et al., 2007) mean AIRS collocation uncertainties can be summarised as:

1. Tropics (±20°): A surface humidity of 27,250 ppmv results in collocation uncertainties between 241.15 to 313.83 ppmv (0.15 to 0.20 g kg^{-1}, or 0.88 to 1.15 % ppmv).
2. Mid-latitudes (20°-60°): A surface humidity of 11,660 ppmv results in collocation uncertainties between 62.11 1 to 71.69 ppmv (0.04 to 0.04 g kg^{-1} , or 0.53 to 0.61 % ppmv).
3. Polar summer (≥60°): A surface humidity 7788 ppmv results in collocation uncertainties between 86.77 to 113.19 ppmv (0.05 to 0.07 g kg^{-1}, or 1.11 to 1.45 % ppmv).
4. Polar winter (≥60°): A surface humidity of 2103 ppmv results in collocation uncertainties between 23.43 to 30.56 ppmv (0.01 to 0.02 g kg^{-1}, or 1.11 to 1.45 % ppmv).

These results are the average over all six years of AIRS collocations with CRM (Figure 6), and for both consistency hypothesises (i.e k = 1 or k =2). By halving the match-up criteria we have shown that the majority of these estimates would increase, with the exception polar values for the Southern Hemisphere (Figure 7). Therefore, we assume that the consistency of the collocation uncertainty (representativeness) is affected by the collocation process which in turn will have an impact on our bias estimates. As more sites are added to the GRUAN inventory providing greater global coverage it should be possible to examine this aspect in further detail. Within the GAIA-CLIM project they showed that they could minimise the uncertainty in the collocation by using the 300 hPa

spatio-temporal information (Verhoelst et al., 2017). Operational radiosonde records such as CRM, the Analyzed Radiosonde Archive (ARSA) and the GCOS Upper-Air Network (GUAN) do not contain this information. Understanding the representativeness of collocations with these records will improve validation efforts for climate analysis, where the satellite record exists outside of the GRUAN time line.

When considering historical records, observed issues in stability and differences to the ensemble mean can be attributed to changes in the observation system (Schröder et al., 2016). When profiles are considered this attribution is still valid frequently, but less systematic. First results from the G-VAP characterisation of the long-term profile records in terms of bias have been published (Schröder et al., 2017). They show that regional maxima in systematic differences differ when different pressure levels are considered, with a focus on standard deviation and stability. Associated analysis will be further enhanced in future G-VAP activities. The work presented here can serve as an excellent case study to guide the characterisation in terms of bias during future studies within assessments of water vapor records. Note that regions with maxima in inter-comparison results generally coincide with regions of sparse ground-based and *in-situ* data availability.

An important pre-requisite for such an effort is the availability of high quality reference data with uncertainty information. Here we have reprocessed global radiosonde data for the period January 2007-December 2012, and demonstrated feasibility for a reprocessing of RS92 radiosonde archives. We also demonstrate the benefits of characterised measurements from GRUAN over the operational RS92 records, through the bias uncertainty estimates. A stable and bias corrected multi-station radiosonde archive is of high value for the validation of satellite-based water vapor products. Therefore, reducing these uncertainties is key for the assessment of any climate record; with the production of a stable, characterised, global, retroactive radiosonde archive a real challenge. Further development and reprocessing of radiosonde data is needed to allow for satellite records spanning from the present back to the late 1970s to be exploited for climate analysis.

The value of systematic differences can be enhanced if the total difference can be broken down into individual components. In the characterisation of the RS92 archive we also compare the vertical distribution of CRM uncertainty with those from GRUAN. One apparent difference is the inability of CRM to fully capture the random uncertainty component which could allow for collocated profiles that should be excluded during the consistency test to be included. A large variety of such structural uncertainties can contribute to the total bias (see Kummerow, Schulz, and Bojkov (2011) for a brief discussion).

The AIRS version 6 water vapor product is an example of data which is based on the combination of hyper-spectral and microwave observations. Similar retrievals exist for IASI and The Cross-track Infrared Sounder (CrIS), though they have not been reprocessed consistently until now. Incorporating observations from companion instruments such as AMSU and MODIS to characterise elements such as cloud fraction or surface moisture content would aid in constraining AIRS retrievals and further improving performance. As IASI and CrIS fly with similar companion instruments this approach could be applied across the board in the construction of a consistent water vapor CDR. With EUMETSAT EPS-SG IASI observations will be sustained until beyond 2030, with a similar commitment from NOAA with CrIS through the Joint Polar Satellite System (JPSS) program. Thus, we may speak of a high potential for climate monitoring using hyper-spectral data in a forward looking sense. The challenge is to maintain the radiosonde systems and asses stability with both GRUAN and CRM, once GRUAN reaches decadal capacity across all sites.

Acknowledgments

Dr Tim Trent & Prof. John Remedios would like to acknowledge the funding from Natural Environment Research Council through Natural Centre for Earth Observation, con-

tract number PR140015. Dr Marc Schröder acknowledges the financial support of the EUMETSAT member states through CM SAF. The authors would also like to thank Edward Olsen for all his help in understanding the AIRS L2 support product. AIRS data can be downloaded from https://airs.jpl.nasa.gov/data/get_data. Information on RS92 sites is taken from 2007 release of WMO catalogue of radiosondes and upper-air wind system information available from ftp://ftp.wmo.int/wmo-ddbs/Radiosondes.xls. The UKMO radiosonde archive is available from the Centre for Environmental Data Archival (CEDA) at http://www.ceda.ac.uk/. The GRUAN RS92-GDP data archive can be accessed via the project home page https://www.gruan.org/.

References

Akima, H. (1970). A new method of interpolation and smooth curve fitting based on local procedures. *Journal of the ACM (JACM)*, *17*(4), 589–602.

Akima, H. (1991). A method of univariate interpolation that has the accuracy of a third-degree polynomial. *ACM Transactions on Mathematical Software (TOMS)*, *17*(3), 341–366.

Andersson, A., Graw, K., Schröder, M., Fennig, K., Liman, J., Bakan, S., ... Klepp, C. (2017). Hamburg ocean atmosphere parameters and fluxes from satellite data - hoaps 4.0. Retrieved from `https://doi.org/10.5676/EUM_SAF_CM/HOAPS/V002` doi: 10.5676/EUM_SAF_CM/HOAPS/V002

Aumann, H. H., Chahine, M. T., Gautier, C., Goldberg, M. D., Kalnay, E., McMillin, L. M., ... others (2003). Airs/amsu/hsb on the aqua mission: Design, science objectives, data products, and processing systems. *Geoscience and Remote Sensing, IEEE Transactions on*, *41*(2), 253–264.

Bevis, M., Businger, S., HERRING, T., Rocken, C., ANTHES, R., & WARE, R. (1992). Gps meteorology- remote sensing of atmospheric water vapor using the global positioning system. *Journal of Geophysical Research*, *97*(D14), 15787–15801.

Boggs, P. T., & Rogers, J. E. (1990). Orthogonal distance regression. *Contemporary Mathematics*, *112*, 183–194.

Chahine, M. T., Pagano, T. S., Aumann, H. H., Atlas, R., Barnet, C., Blaisdell, J., ... others (2006). Airs improving weather forecasting and providing new data on greenhouse gases.

Chung, E.-S., Soden, B., Sohn, B., & Shi, L. (2014). Upper-tropospheric moistening in response to anthropogenic warming. *Proceedings of the National Academy of Sciences*, *111*(32), 11636–11641.

Dessler, A., Zhang, Z., & Yang, P. (2008). Water-vapor climate feedback inferred from climate fluctuations, 2003–2008. *Geophysical Research Letters*, *35*(20).

Dirksen, R., Sommer, M., Immler, F., Hurst, D., Kivi, R., & Vömel, H. (2014). Reference quality upper-air measurements: Gruan data processing for the vaisala rs92 radiosonde. *Atmospheric Measurement Techniques*, *7*(12), 4463–4490.

Divakarla, M. G., Barnet, C. D., Goldberg, M. D., McMillin, L. M., Maddy, E., Wolf, W., ... Liu, X. (2006). Validation of atmospheric infrared sounder temperature and water vapor retrievals with matched radiosonde measurements and forecasts. *Journal of Geophysical Research: Atmospheres (1984–2012)*, *111*(D9).

Feltz, M., Borg, L., Knuteson, R., Tobin, D., Revercomb, H., & Gambacorta, A. (2017). Assessment of noaa nucaps upper air temperature profiles using cosmic gps radio occultation and arm radiosondes. *Journal of Geophysical Research: Atmospheres*, *122*(17), 9130–9153.

Fetzer, E., Eldering, A., Fishbein, E. F., Hearty, T., Irion, W. F., & Kahn, B. (2003). *Validation of airs/amsu/hsb core products for data release version 3.0.* National Aeronautics and Space Administration, Jet Propulsion Laboratory.

Fetzer, E. J., Lambrigtsen, B. H., Eldering, A., Aumann, H. H., & Chahine, M. T. (2006). Biases in total precipitable water vapor climatologies from atmospheric infrared sounder and advanced microwave scanning radiometer. *Journal of Geophysical Research: Atmospheres*, *111*(D9).

Fetzer, E. J., Read, W. G., Waliser, D., Kahn, B. H., Tian, B., Vmel, H., ... Dang, V. (2008). Comparison of upper tropospheric water vapor observations from the microwave limb sounder and atmospheric infrared sounder. *Journal of Geophysical Research: Atmospheres*, *113*(D22), n/a–n/a. Retrieved from `http://dx.doi.org/10.1029/2008JD010000` (D22110) doi: 10.1029/2008JD010000

GCOS. (2016). The global observing system for climate: Implementation needs. Retrieved from `https://library.wmo.int/doc_num.php?explnum_id=3417`

Gettelman, A., Weinstock, E., Fetzer, E., Irion, F., Eldering, A., Richard, E., ... others (2004). Validation of aqua satellite data in the upper troposphere and lower stratosphere with in situ aircraft instruments. *Geophysical research letters*, *31*(22).

Hagan, D. E., Webster, C. R., Farmer, C. B., May, R. D., Herman, R. L., Weinstock, E. M., ... Newman, P. A. (2004). Validating airs upper atmosphere water vapor retrievals using aircraft and balloon in situ measurements. *Geophysical Research Letters*, *31*(21), n/a–n/a. Retrieved from `http://dx.doi.org/10.1029/2004GL020302` doi: 10.1029/2004GL020302

Hearty, T. J., Savtchenko, A., Tian, B., Fetzer, E., Yung, Y. L., Theobald, M., ... Won, Y.-I. (2014). Estimating sampling biases and measurement uncertainties of airs/amsu-a temperature and water vapor observations using merra reanalysis. *Journal of Geophysical Research: Atmospheres*, *119*(6), 2725–2741.

Held, I. M., & Soden, B. J. (2000). Water vapor feedback and global warming 1. *Annual Review of Energy and the Environment*, *25*(1), 441–475.

Hellinger, E. (1909). Neue begründung der theorie quadratischer formen von unendlichvielen veränderlichen. *Journal für die reine und angewandte Mathematik*, *136*, 210–271.

Hyland, R., & Wexler, A. (1983). Formulations for thermodymnamic properties of the saturated phases of h_2o from 173.15 k to 473.15 k. *Ashrae transactions, A*, *2*, 500–513.

Iglewicz, B., & Hoaglin, D. C. (1993). *How to detect and handle outliers* (Vol. 16). Asq Press.

Immler, F., Dykema, J., Gardiner, T., Whiteman, D., Thorne, P., & Vömel, H. (2010). Reference quality upper-air measurements: guidance for developing gruan data products. *Atmospheric Measurement Techniques*, *3*(5), 1217–1231.

Immler, F., & Sommer, M. (2011). *Brief description of the rs92 gruan data product (rs92-gdp). gruan tech* (Tech. Rep.). Document GRUAN-TD-4, 17 pp.[Available online at http://www. gruan. org.].

Kinzel, J., Fennig, K., Schröder, M., Andersson, A., Bumke, K., & Hollmann, R. (2016). Decomposition of random errors inherent to hoaps-3.2 near-surface humidity estimates using multiple triple collocation analysis. *Journal of Atmospheric and Oceanic Technology*, *33*(7), 1455–1471.

Kummerow, C., Schulz, J., & Bojkov, B. (2011). Gewex/esa data user element globvapour workshop on long-term water vapor data sets. *GEWEX Newsletter*, *21*(2).

Le Marshall, J., Jung, J., Derber, J., Chahine, M., Treadon, R., Lord, S., ... others (2006). Improving global analysis and forecasting with airs. *Bulletin of the American Meteorological Society*, *87*(7), 891–894.

Liman, J., Schröder, M., Fennig, K., Andersson, A., & Hollmann, R. (2018). Uncertainty characterization of hoaps 3.3 latent heat-flux-related parameters. *Atmospheric Measurement Techniques*, *11*(3), 1793.

Maddy, E. S., & Barnet, C. D. (2008). Vertical resolution estimates in version 5 of

airs operational retrievals. *Geoscience and Remote Sensing, IEEE Transactions on*, *46*(8), 2375–2384.

Miloshevich, L. M., Paukkunen, A., Vömel, H., & Oltmans, S. J. (2004). Development and validation of a time-lag correction for vaisala radiosonde humidity measurements. *Journal of Atmospheric and Oceanic Technology*, *21*(9), 1305–1327.

Miloshevich, L. M., Vömel, H., Whiteman, D. N., & Leblanc, T. (2009). Accuracy assessment and correction of vaisala rs92 radiosonde water vapor measurements. *Journal of Geophysical Research: Atmospheres (1984–2012)*, *114*(D11).

Miloshevich, L. M., Vömel, H., Whiteman, D. N., Lesht, B. M., Schmidlin, F., & Russo, F. (2006). Absolute accuracy of water vapor measurements from six operational radiosonde types launched during awex-g and implications for airs validation. *Journal of Geophysical Research: Atmospheres (1984–2012)*, *111*(D9).

Murphy, D., & Koop, T. (2005). Review of the vapour pressures of ice and supercooled water for atmospheric applications. *Quarterly Journal of the Royal Meteorological Society*, *131*(608), 1539–1565.

Myhre, G., Shindell, D., Bréon, F., Collins, W., Fuglestvedt, J., Huang, J., ... others (2013). Anthropogenic and natural radiative forcing. *Climate change*, *423*.

Nalli, N. R., Barnet, C. D., Reale, A., Tobin, D., Gambacorta, A., Maddy, E. S., ... others (2013). Validation of satellite sounder environmental data records: Application to the cross-track infrared microwave sounder suite. *Journal of Geophysical Research: Atmospheres*, *118*(24).

Olsen, E., Fetzer, E., Hulley, G., Manning, E., Blaisdell, J., Iredell, L., ... others (2013). Airs/amsu/hsb version 6 level 2 product user guide. *NASA-JPL Tech Rep*, 134.

Olsen, E. T., Fishbein, E., Granger, S., Lee, S., Manning, E., Weiler, M., ... Susskind, J. (2007). Airs/amsu/hsb version 5 data release user guide. *Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California*.

Olsen, E. T., Fishbein, E., Lee, S., & Manning, E. (2011). Airs/amsu/hsb version 5 level 2 product levels, layers and trapezoids.

Parzen, E. (1962). On estimation of a probability density function and mode. *The annals of mathematical statistics*, *33*(3), 1065–1076.

Pougatchev, N., August, T., Calbet, X., Hultberg, T., Oduleye, O., Schlüssel, P., ... Bingham, G. (2009). Iasi temperature and water vapor retrievals–error assessment and validation. *Atmospheric Chemistry and Physics*, *9*(17), 6453–6458.

Reale, T., Sun, B., Tilley, F. H., & Pettey, M. (2012). The noaa products validation system (nprovs). *Journal of Atmospheric and Oceanic Technology*, *29*(5), 629–645.

Remedios, J. J., Leigh, R. J., Waterfall, A. M., Moore, D. P., Sembhi, H., Parkes, I., ... Hauglustaine, D. (2007). Mipas reference atmospheres and comparisons to v4.61/v4.62 mipas level 2 geophysical data sets. *Atmospheric Chemistry and Physics Discussions*, *7*, 9973–10017. Retrieved from `https://www.atmos-chem-phys-discuss.net/7/9973/2007/` doi: 10.5194/acpd-7-9973-2007

Rodgers, C. D., & Connor, B. J. (2003). Intercomparison of remote sounding instruments. *Journal of Geophysical Research: Atmospheres (1984–2012)*, *108*(D3).

Roman, J., Knuteson, R., August, T., Hultberg, T., Ackerman, S., & Revercomb, H. (2016). A global assessment of nasa airs v6 and eumetsat iasi v6 precipitable water vapor using ground-based gps suominet stations. *Journal of Geophysical Research: Atmospheres*, *121*(15), 8925–8948.

Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density func-

tion. *The Annals of Mathematical Statistics*, 832–837.

Schröder, M., Lockhoff, M., Forsythe, J. M., Cronk, H. Q., Haar, T. H. V., & Bennartz, R. (2016). The gewex water vapor assessment: Results from intercomparison, trend, and homogeneity analysis of total column water vapor. *Journal of Applied Meteorology and Climatology*, *55*(7), 1633-1649. Retrieved from https://doi.org/10.1175/JAMC-D-15-0304.1 doi: 10.1175/JAMC-D-15-0304.1

Schröder, M., Lockhoff, M., Shi, L., August, T., Bennartz, R., Borbas, E., ... Yang, Q. (2017). Gewex water vapor assessment (gvap). Retrieved from https://www.wcrp-climate.org/resources/wcrp-publications

Scott, D. W. (2015). *Multivariate density estimation: theory, practice, and visualization*. John Wiley & Sons.

Seidel, D. J., Berger, F. H., Immler, F., Sommer, M., Vömel, H., Diamond, H. J., ... others (2009). Reference upper-air observations for climate: Rationale, progress, and plans. *Bulletin of the American Meteorological Society*, *90*(3), 361–369.

Sherwood, S., Roca, R., Weckwerth, T., & Andronova, N. (2010). Tropospheric water vapor, convection, and climate. *Reviews of Geophysics*, *48*(2).

Sun, B., Reale, A., Seidel, D. J., & Hunt, D. C. (2010). Comparing radiosonde and cosmic atmospheric profile data to quantify differences among radiosonde types and the effects of imperfect collocation on comparison statistics. *Journal of Geophysical Research: Atmospheres*, *115*(D23).

Sun, B., Reale, A., Tilley, F. H., Pettey, M. E., Nalli, N. R., & Barnet, C. D. (2017). Assessment of nucaps s-npp cris/atms sounding products using reference and conventional radiosonde observations. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *10*(6), 2499–2509.

Susskind, J., Blaisdell, J. M., & Iredell, L. (2014). Improved methodology for surface and atmospheric soundings, error estimates, and quality control procedures: the atmospheric infrared sounder science team version-6 retrieval algorithm. *Journal of Applied Remote Sensing*, *8*(1), 084994.

Tobin, D. C., Revercomb, H. E., Knuteson, R. O., Lesht, B. M., Strow, L. L., Hannon, S. E., ... Cress, T. S. (2006). Atmospheric radiation measurement site atmospheric state best estimates for atmospheric infrared sounder temperature and water vapor retrieval validation. *Journal of Geophysical Research: Atmospheres (1984–2012)*, *111*(D9).

Trenberth, K., Fasullo, J., & Smith, L. (2005). Trends and variability in column-integrated atmospheric water vapor. *Climate Dynamics*, *24*(7-8), 741-758. Retrieved from http://dx.doi.org/10.1007/s00382-005-0017-4 doi: 10.1007/s00382-005-0017-4

Trent, T. J. (2015). *Climate and variability of water vapour in the troposphere* (Unpublished doctoral dissertation). Department of Physics and Astronomy.

UKMO. (2006). *Met office global radiosonde data. ncas british atmospheric data centre.* http://catalogue.ceda.ac.uk/uuid/f2afaf808b61394b78bd342ff068c8cd. (Accessed: 2017-06-03)

Verhoelst, T., Lambert, J.-C., Fasso, A., Madonna, F., Gardiner, T., Green, P., ... Kivi, R. (2017, 10). *Library of (1) smoothing/sampling error estimates for key atmospheric composition measurement systems, and (2) smoothing sampling error estimates for key data comparisons* (Tech. Rep.). Royal Belgian Institute for Space Aeronomy (BIRA-IASB), Brussels, Belgium.

Wong, S., Fetzer, E. J., Schreier, M., Manipon, G., Fishbein, E. F., Kahn, B. H., ... Irion, F. W. (2015). Cloud-induced uncertainties in airs and ecmwf temperature and specific humidity. *Journal of Geophysical Research: Atmospheres*, *120*(5), 1880–1901.

Yue, Q., Fetzer, E. J., Kahn, B. H., Wong, S., Manipon, G., Guillaume, A., & Wilson, B. (2013). Cloud-state-dependent sampling in airs observations based on

cloudsat cloud classification. *Journal of Climate*, *26*(21), 8357–8377.

Figure 1.

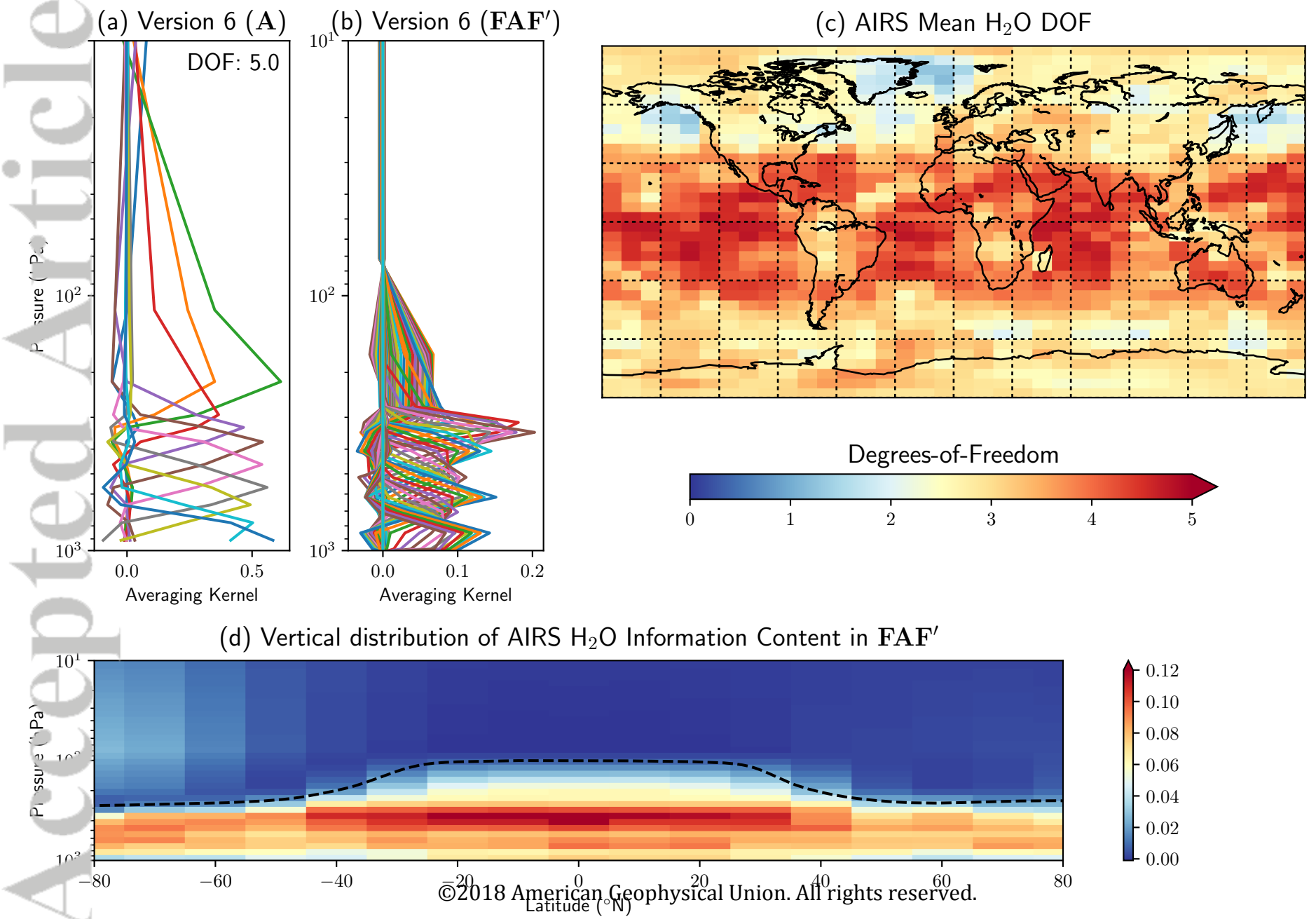


Figure 2.

(a) GRUAN Radiosonde Stations (2007-2012)

(b) CRM Radiosonde Stations (2007-2012)

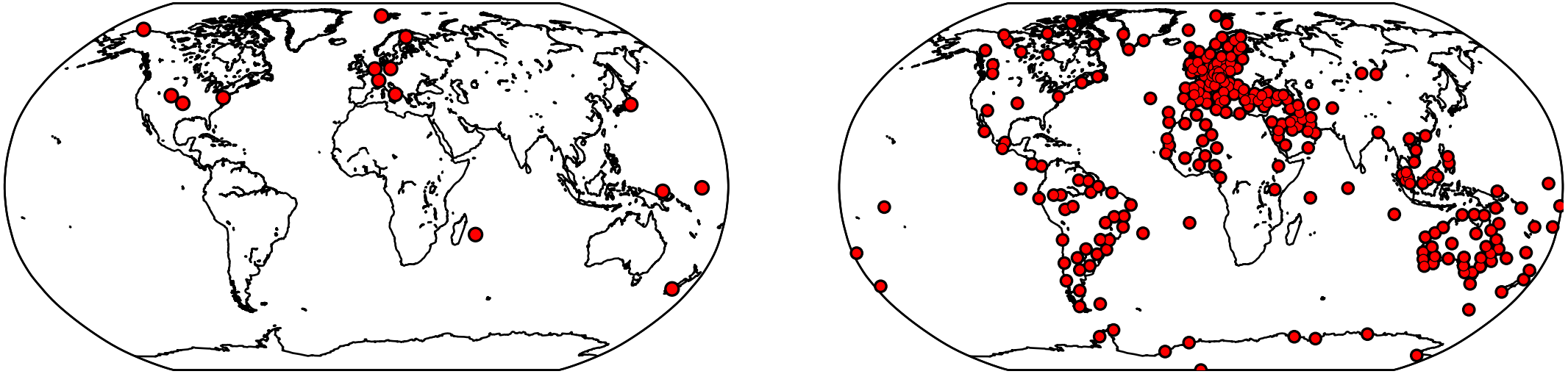
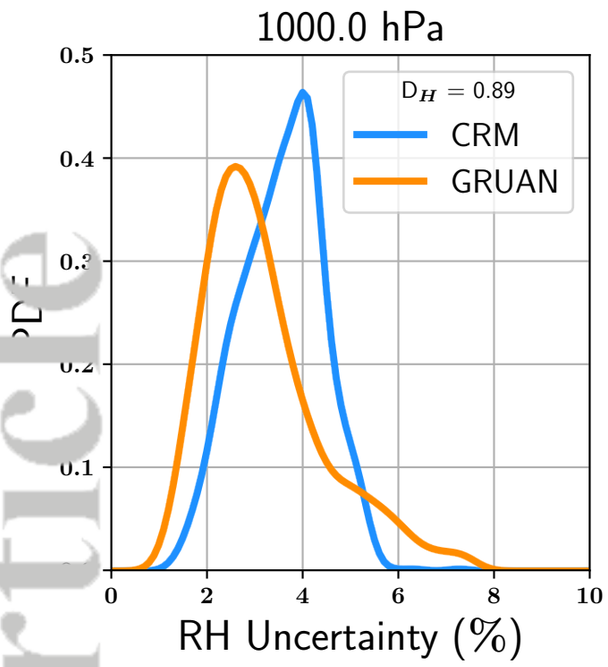
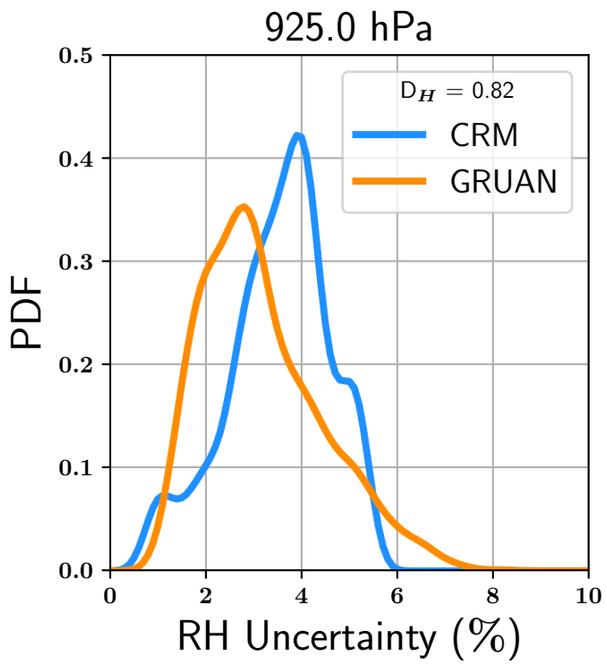
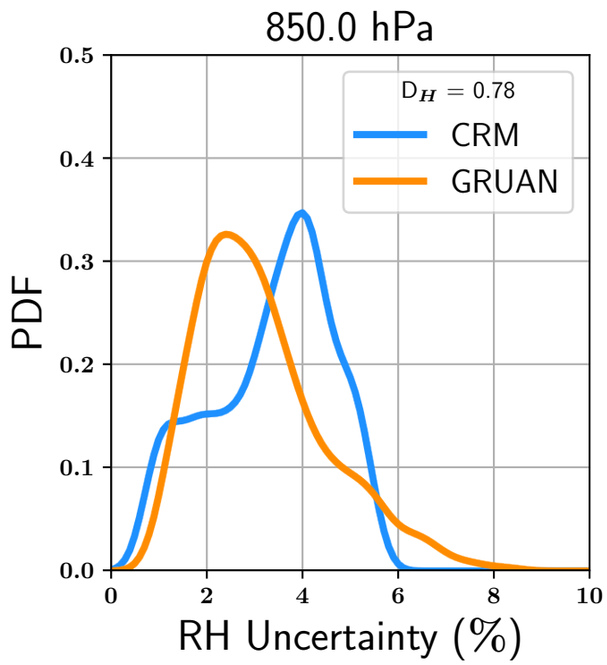
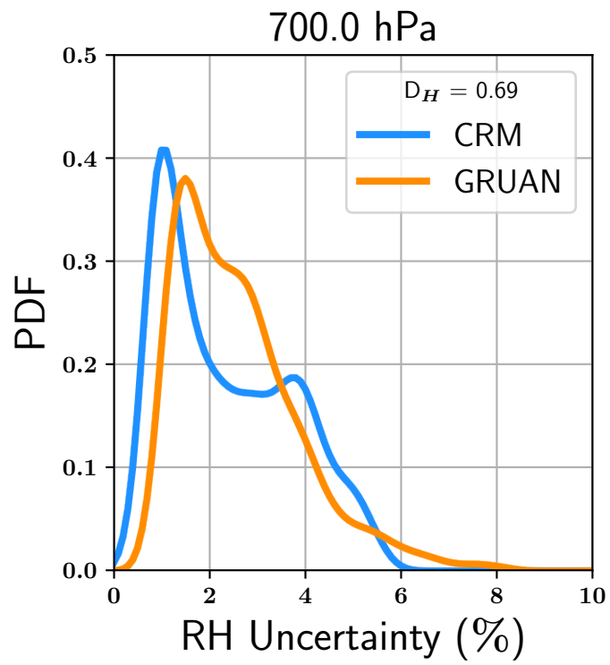


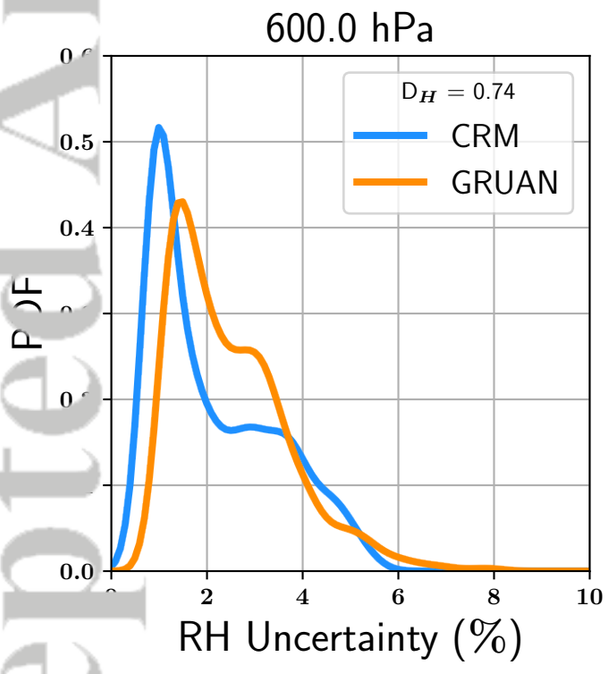
Figure 3.

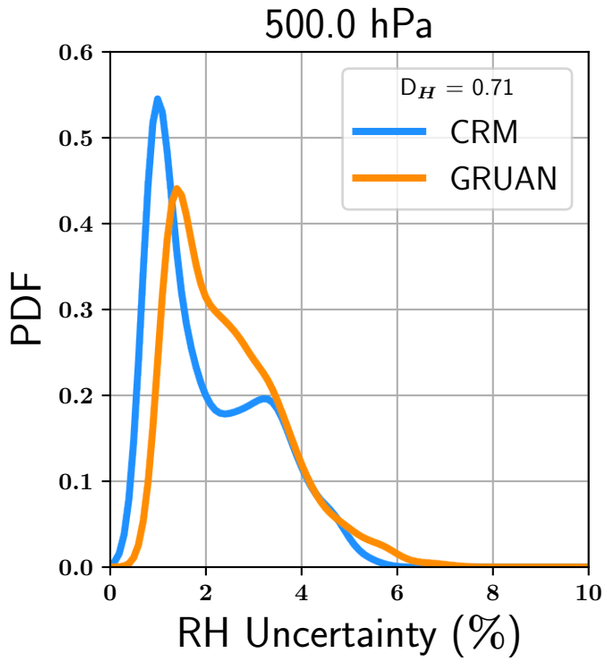


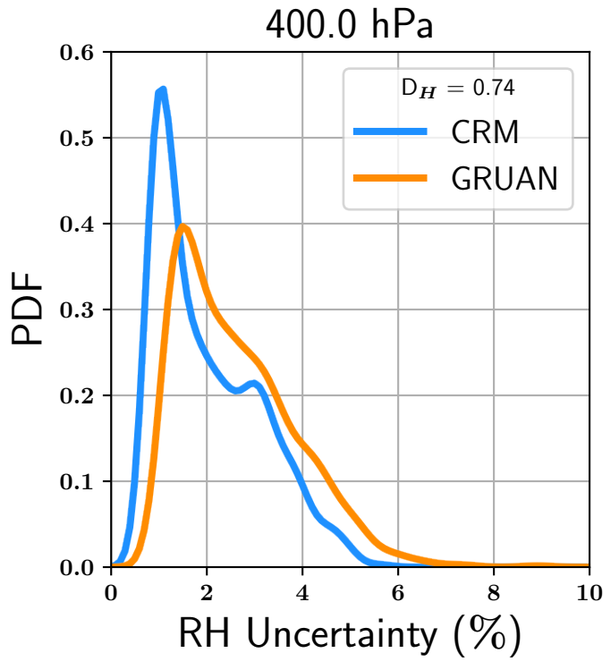


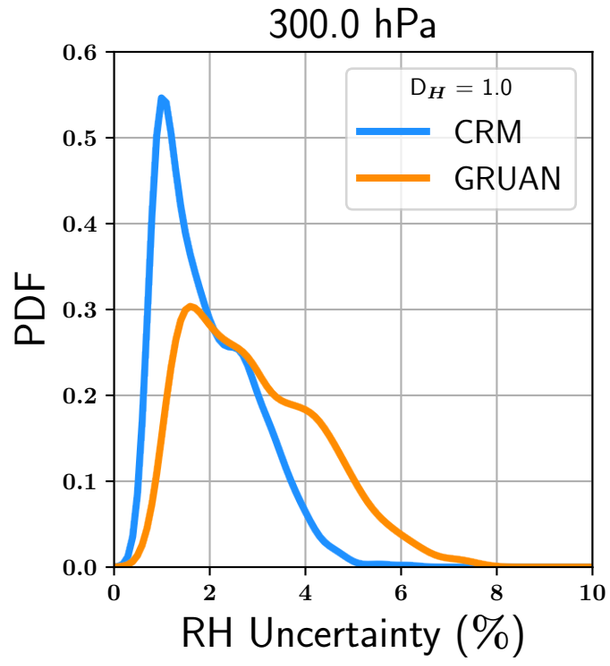


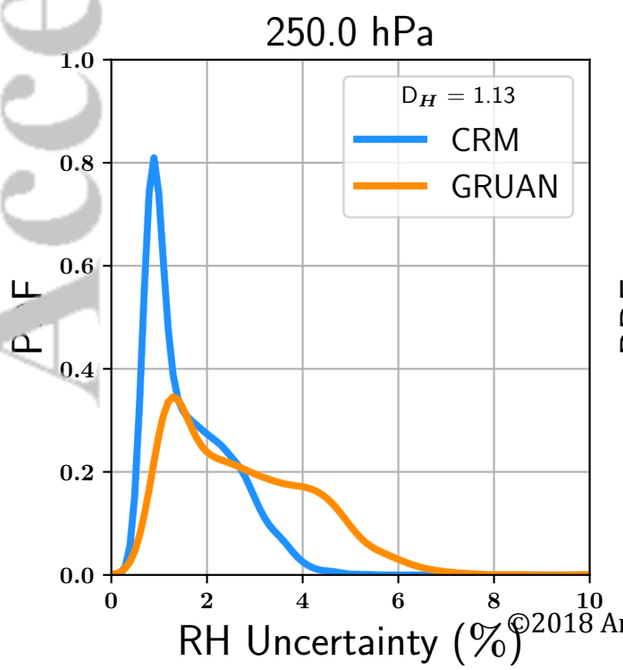


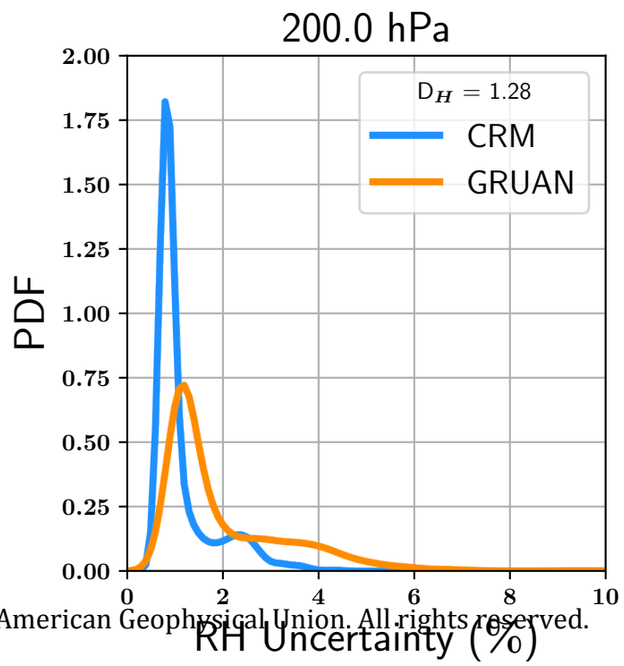


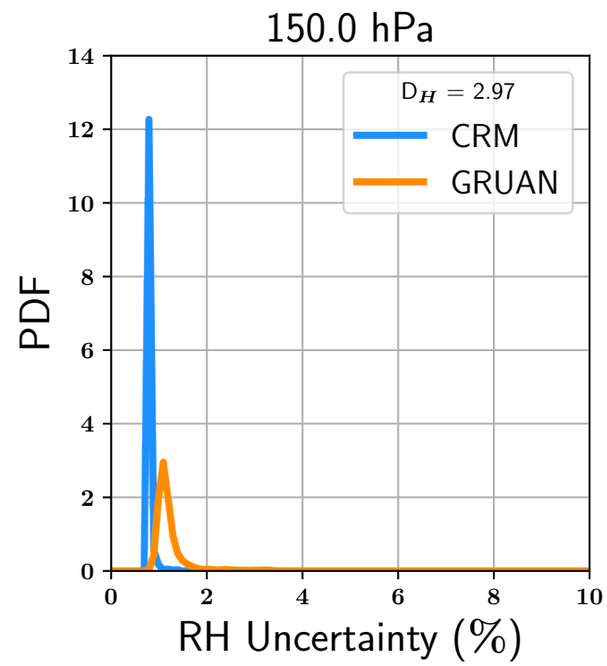












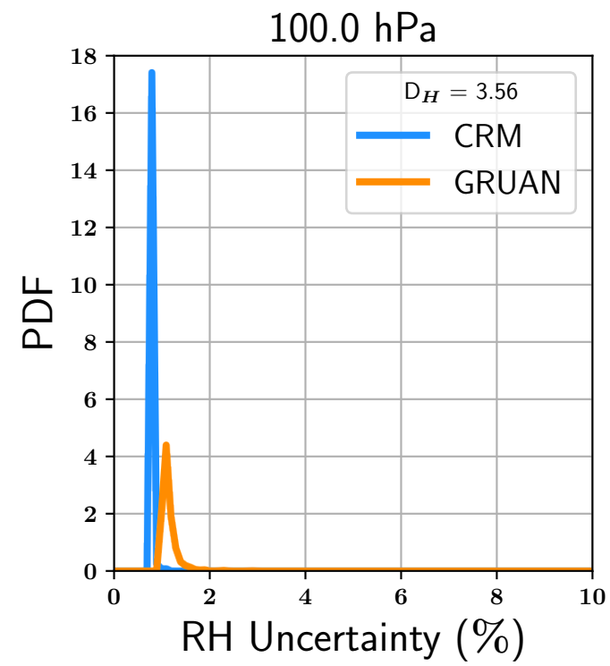
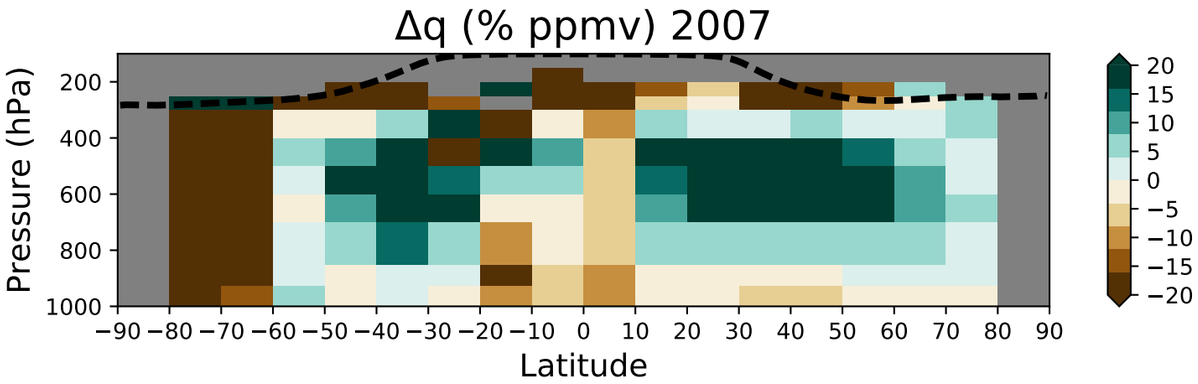
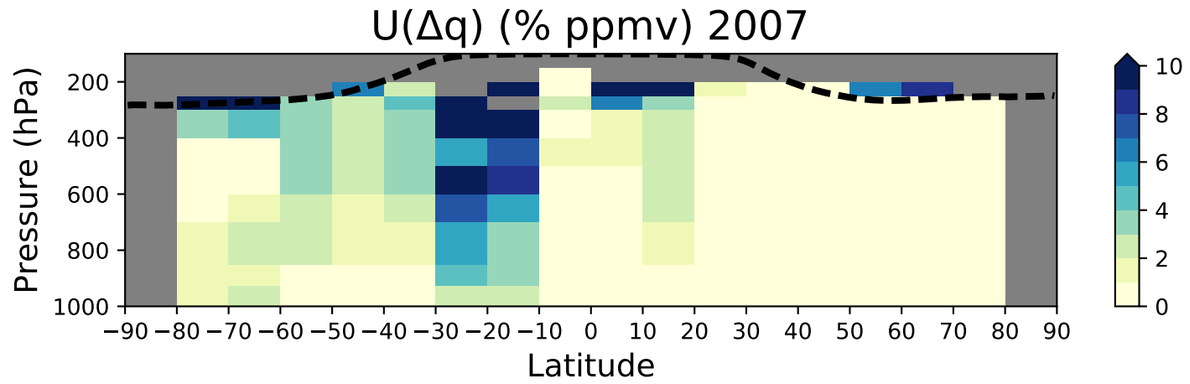
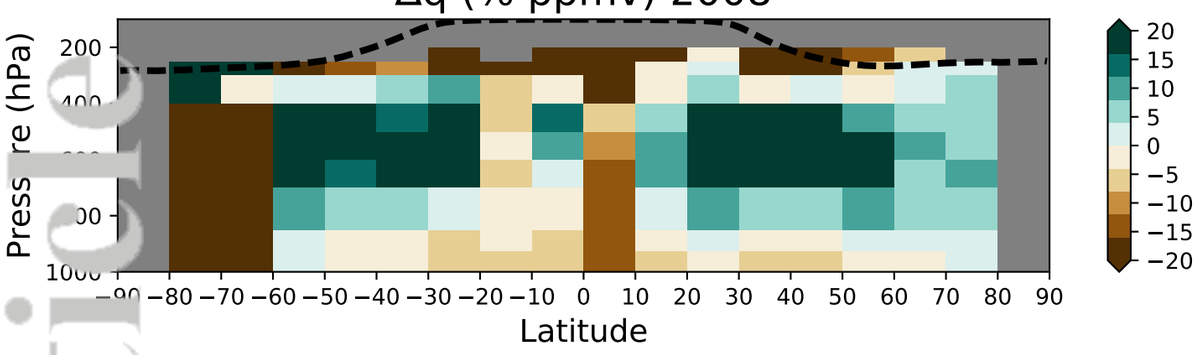
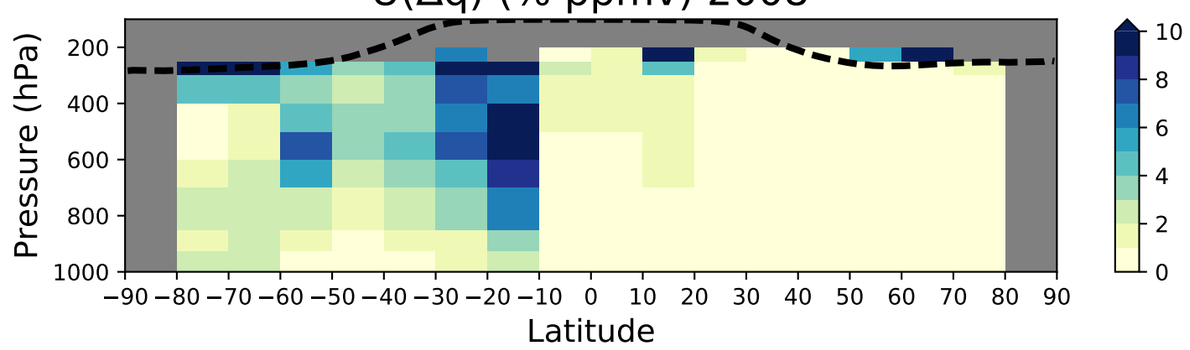


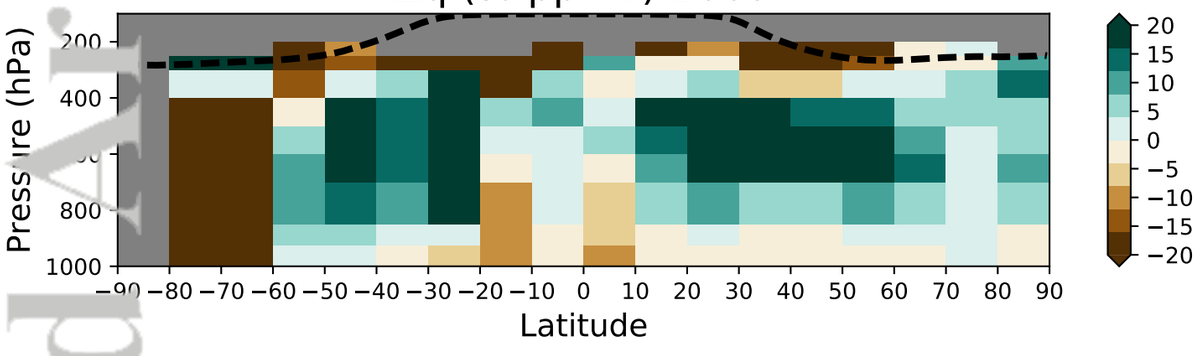
Figure 4.

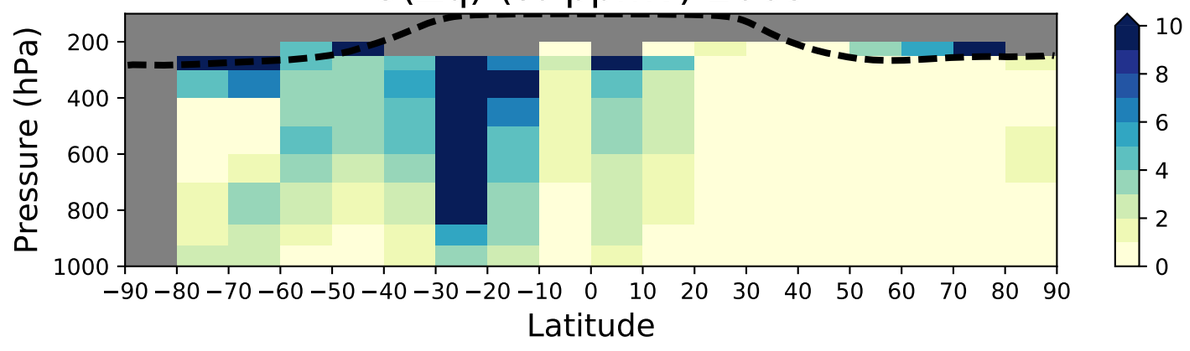


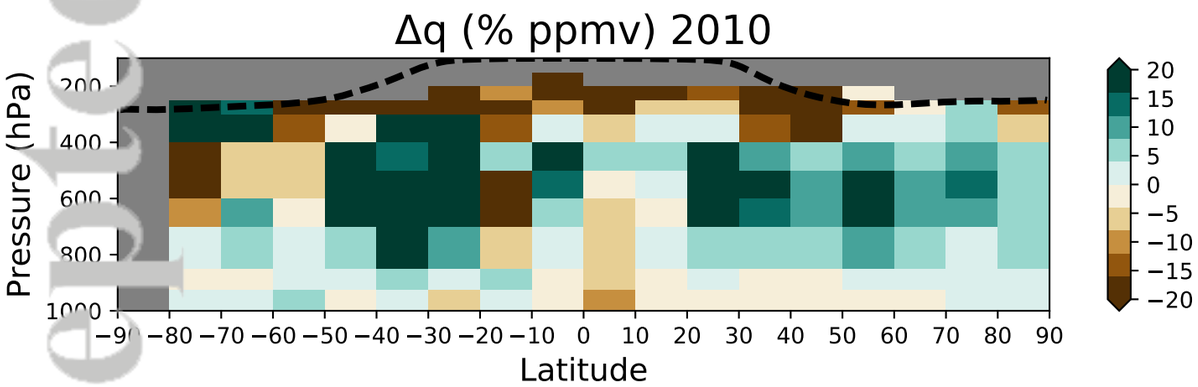


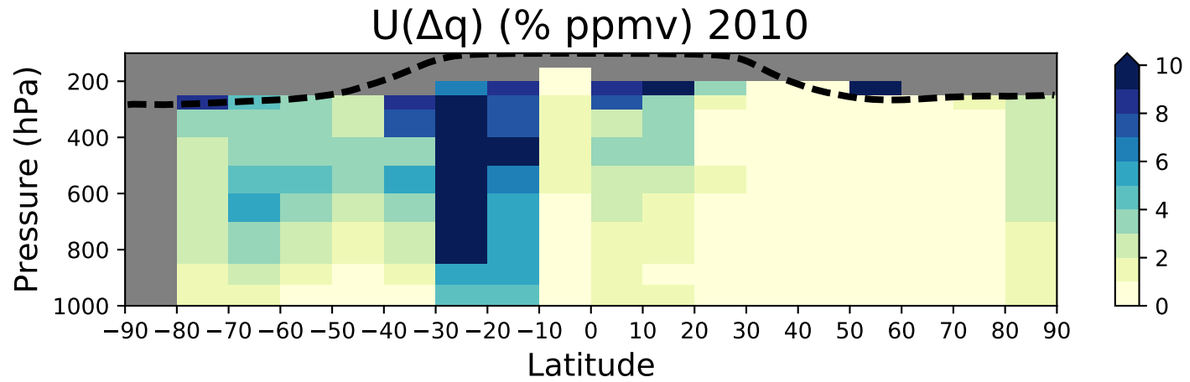


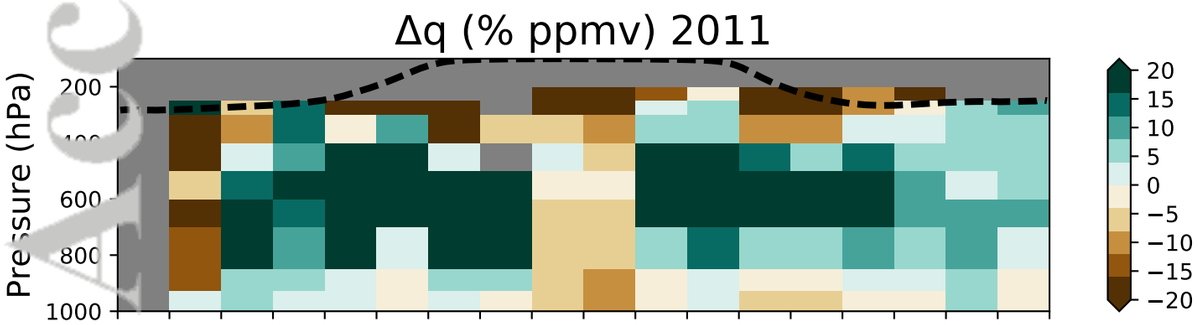


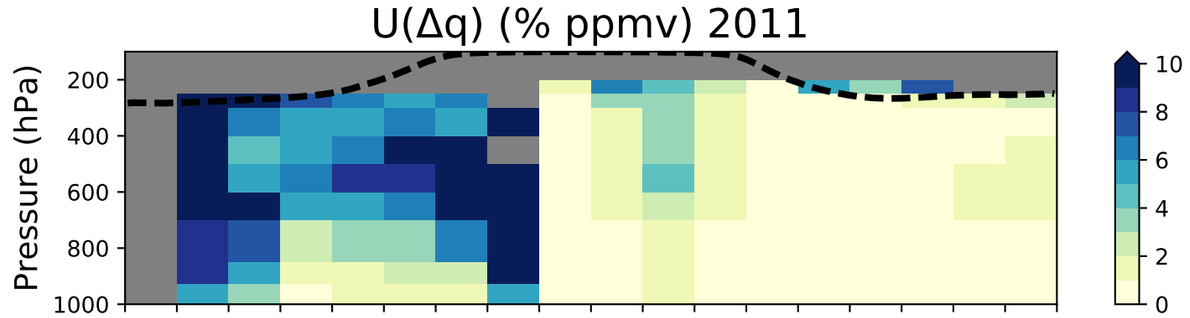


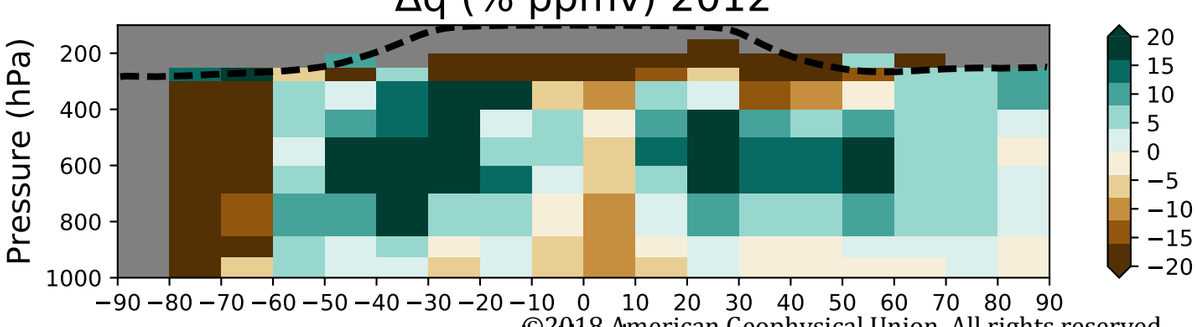












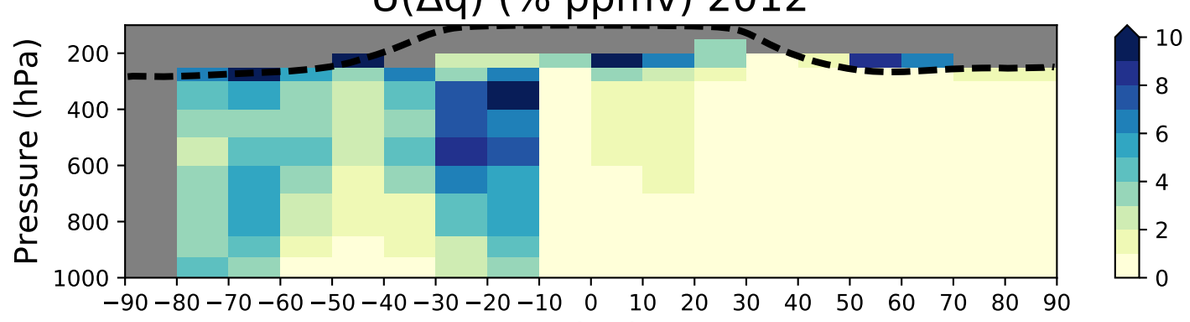


Figure 5.

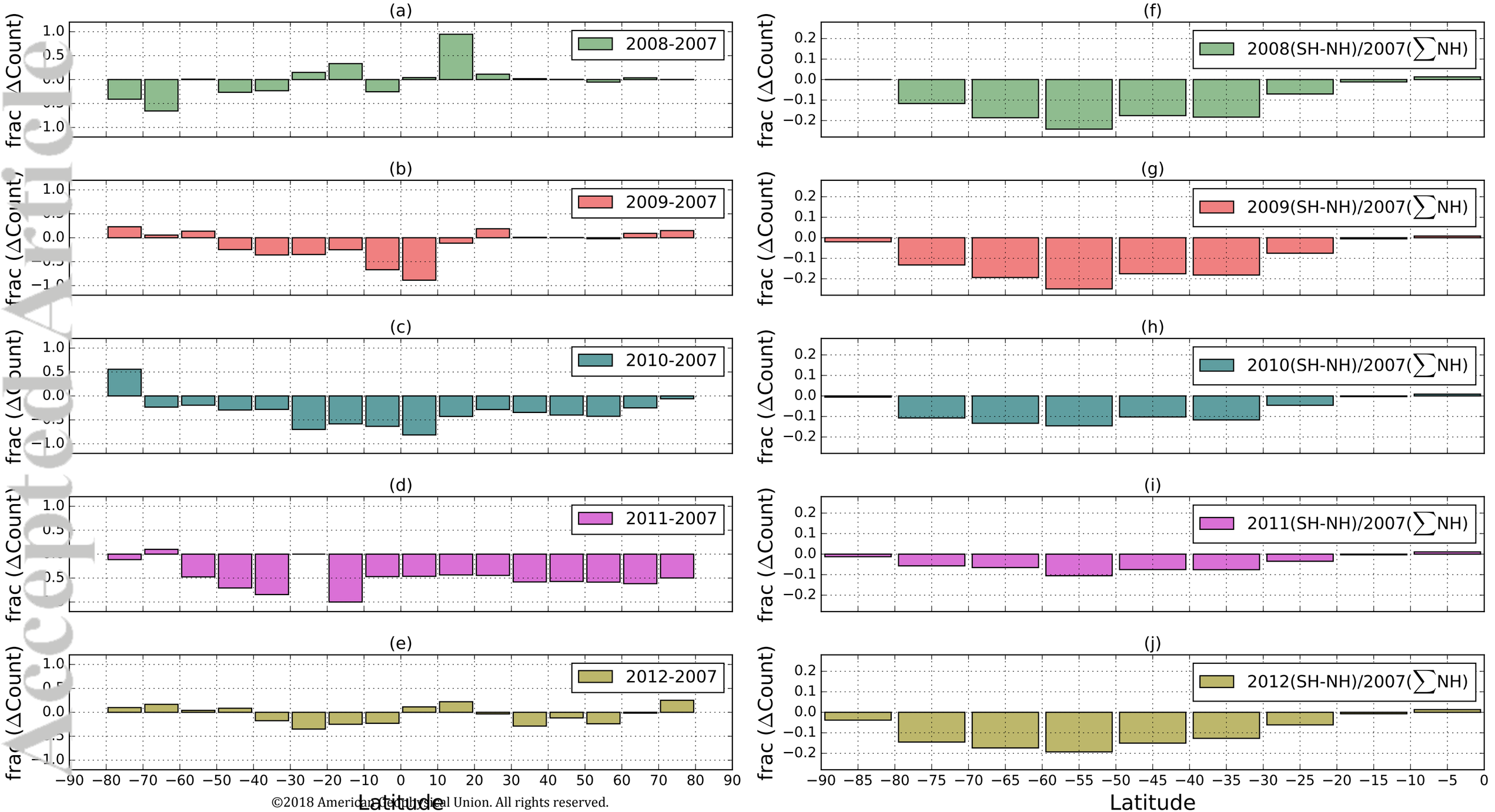
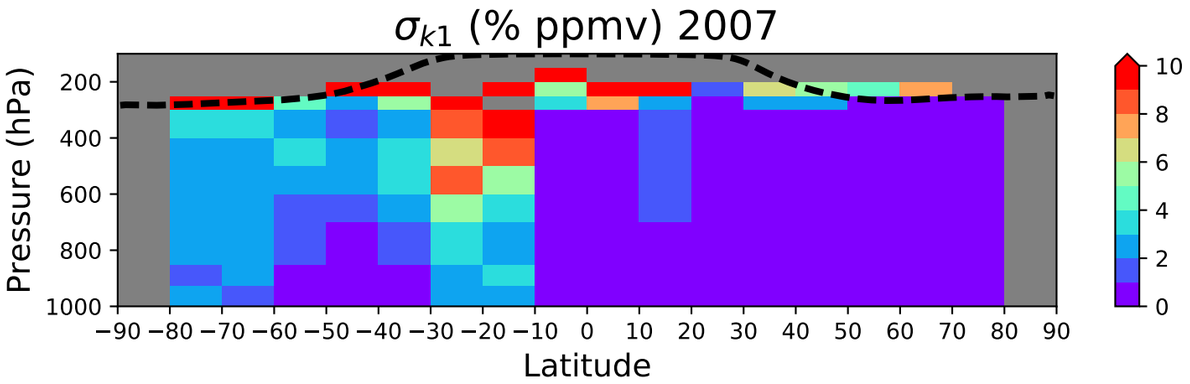
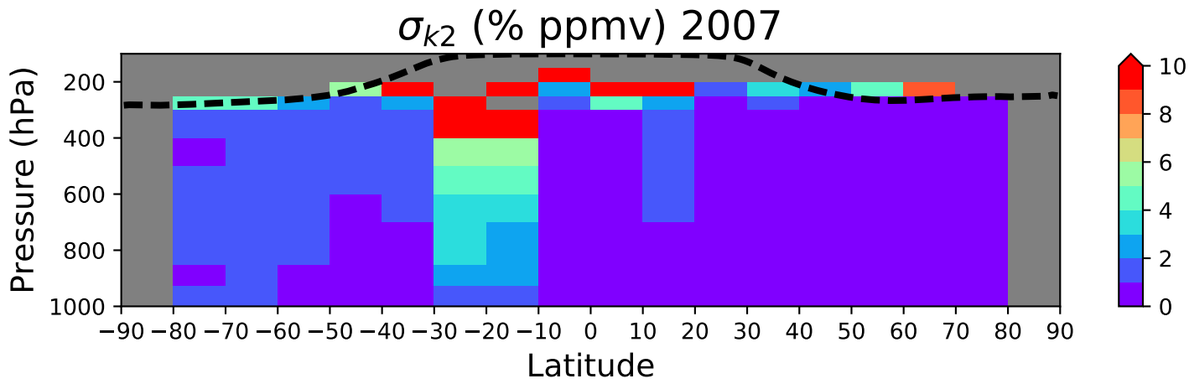
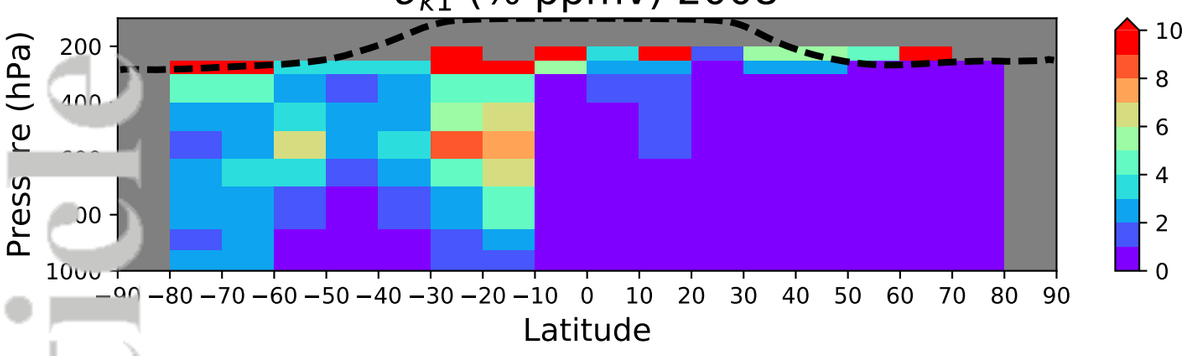
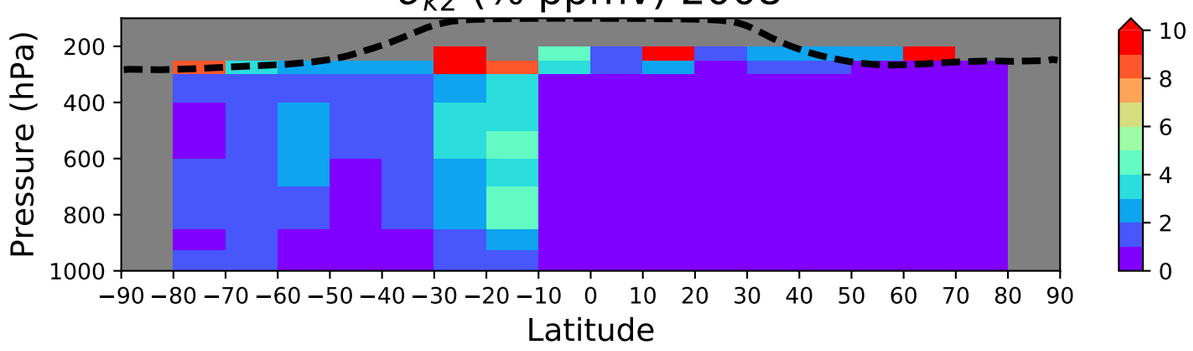


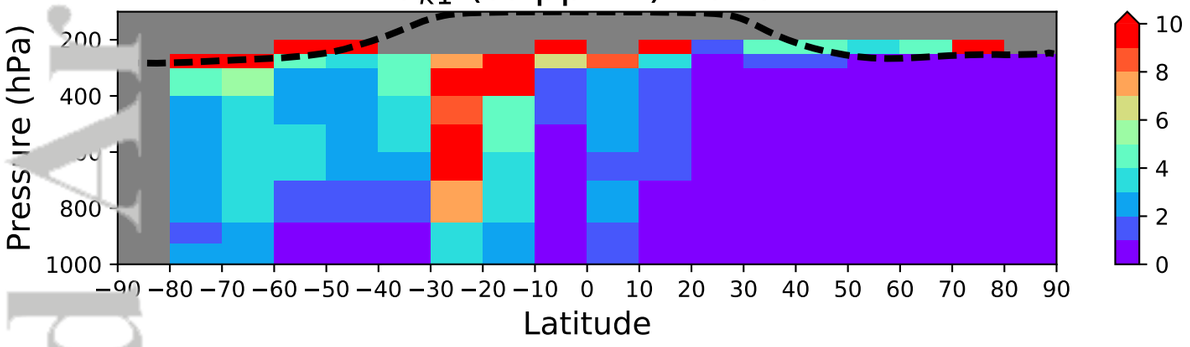
Figure 6.

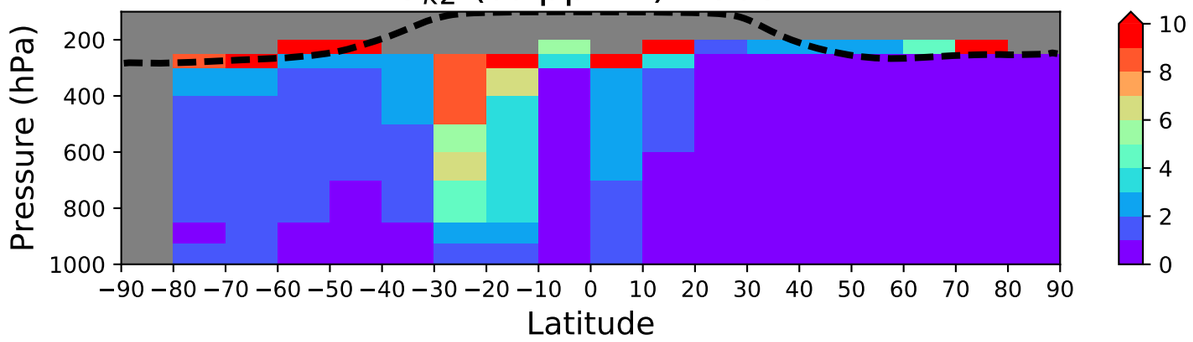


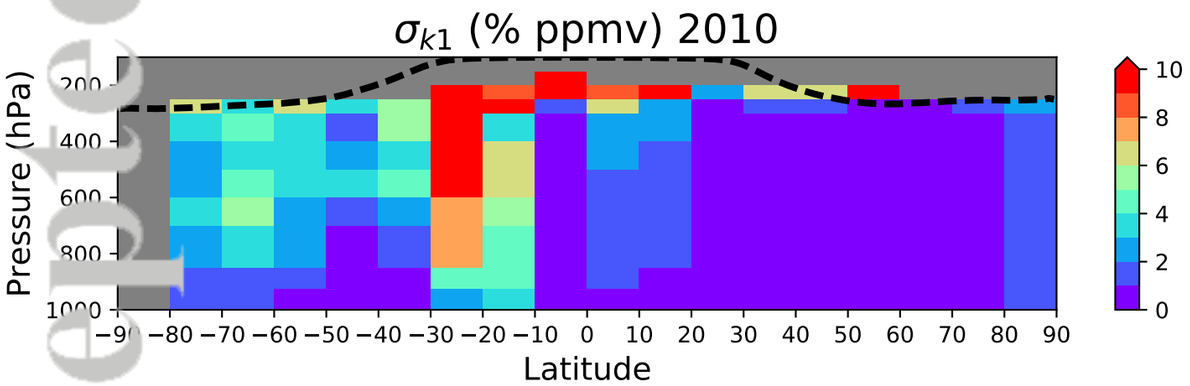


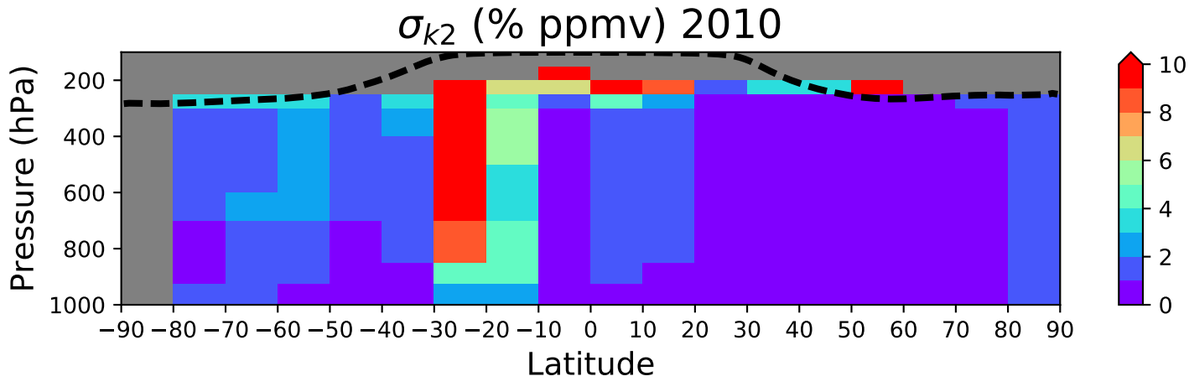


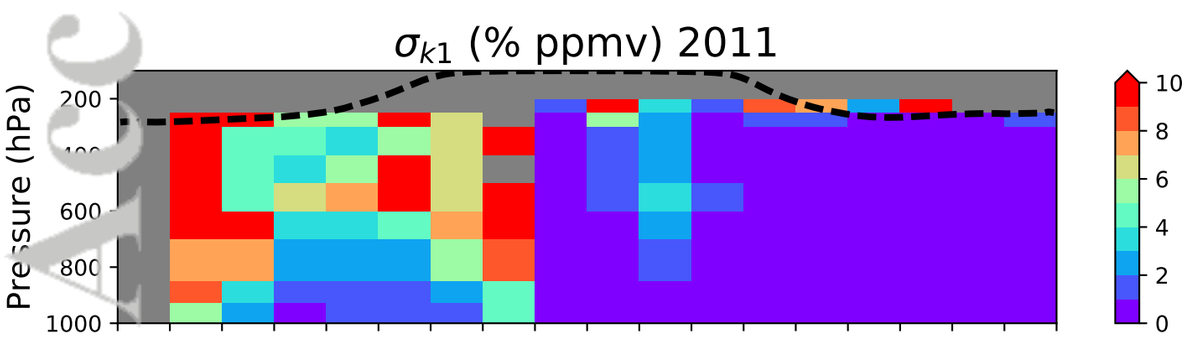


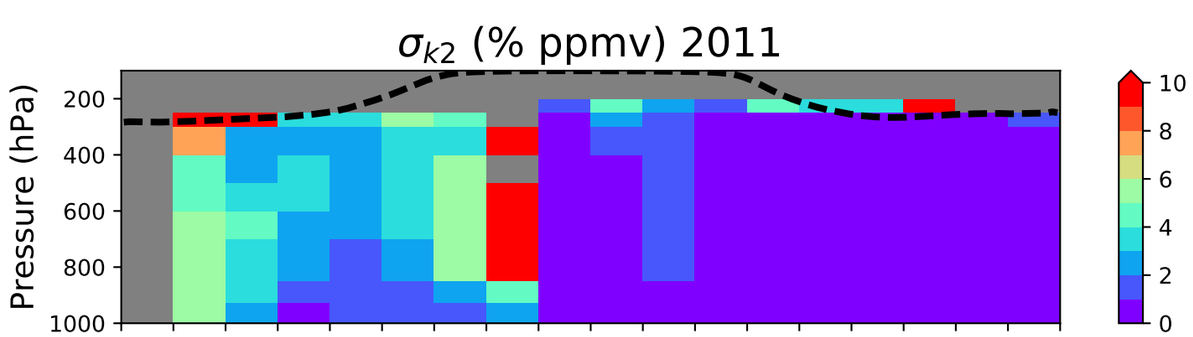


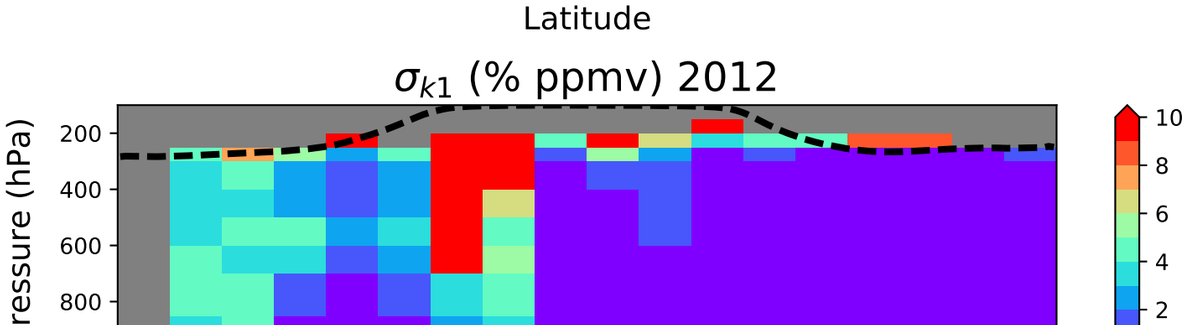












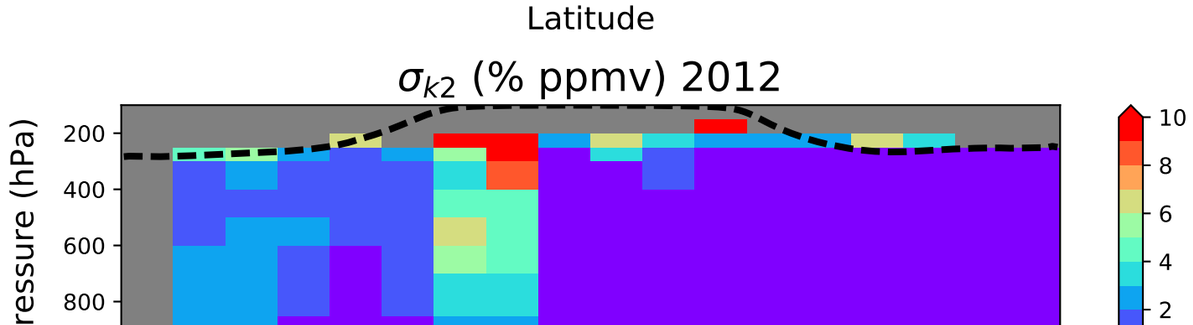
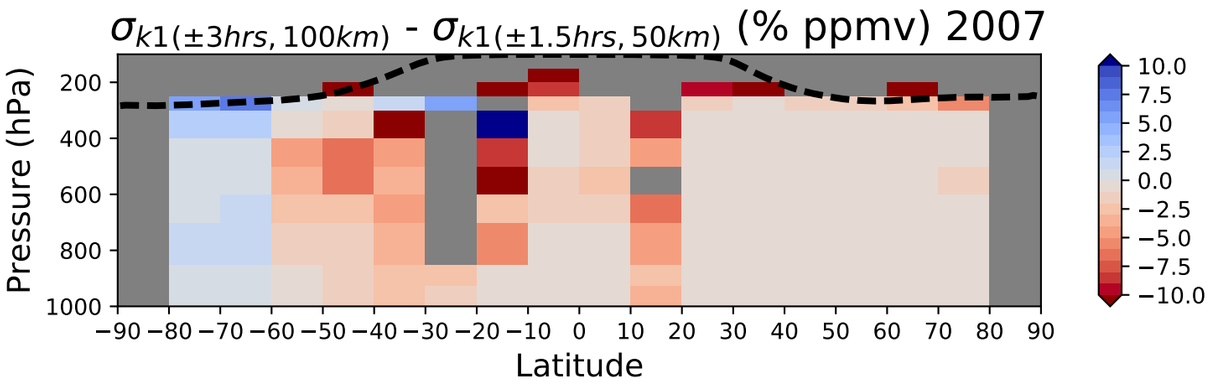
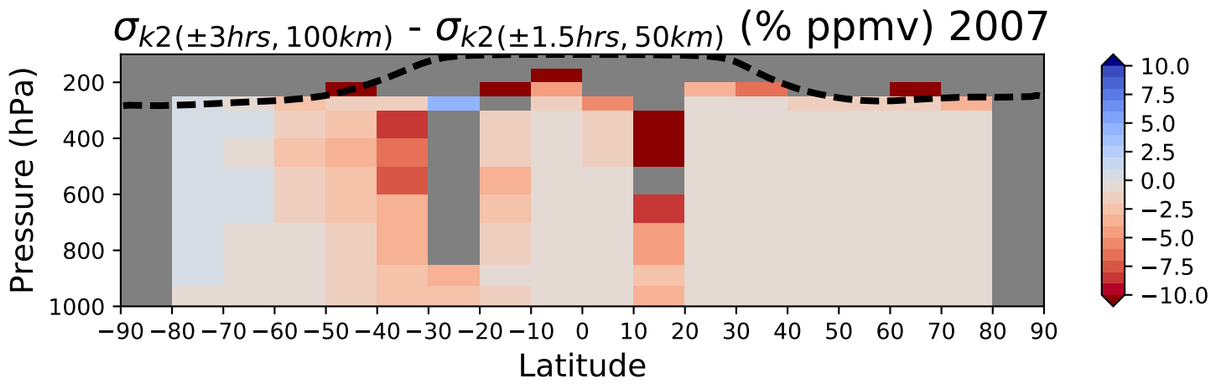
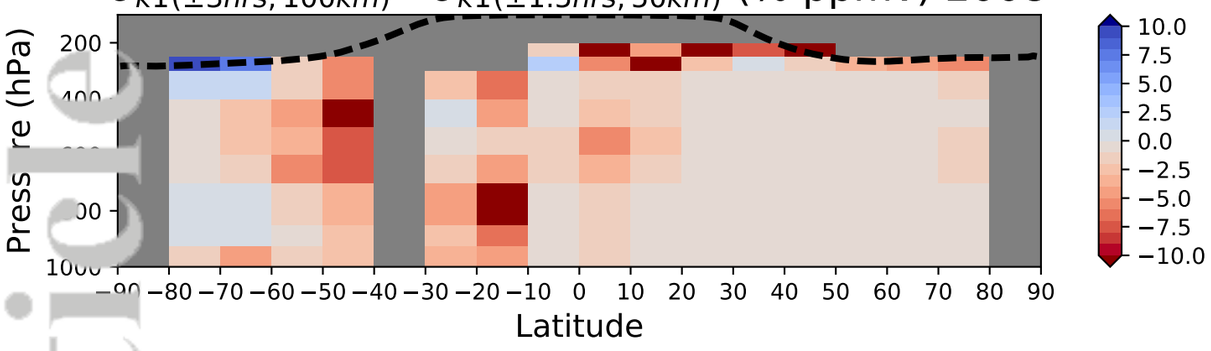
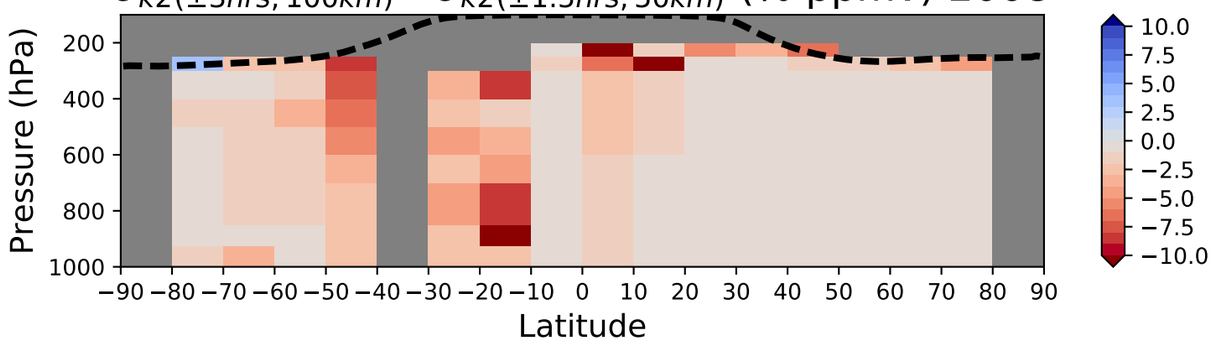


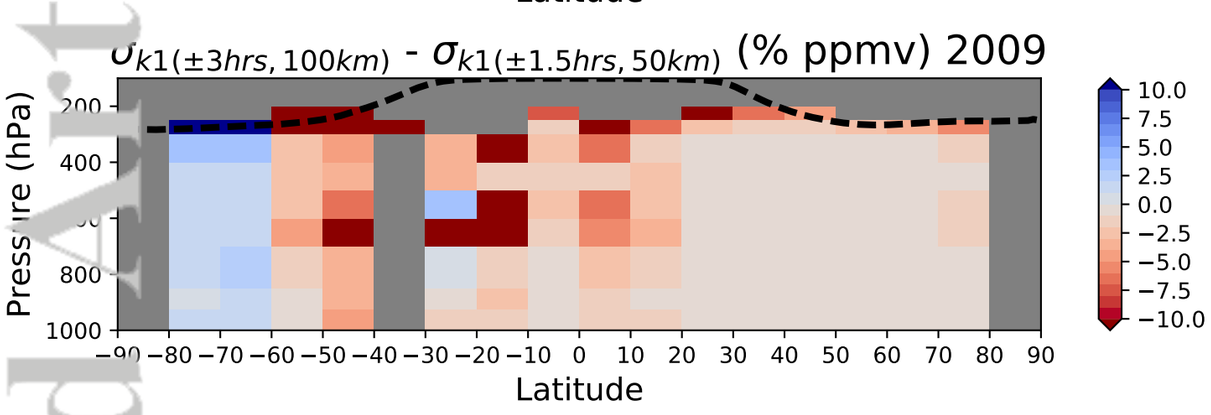
Figure 7.

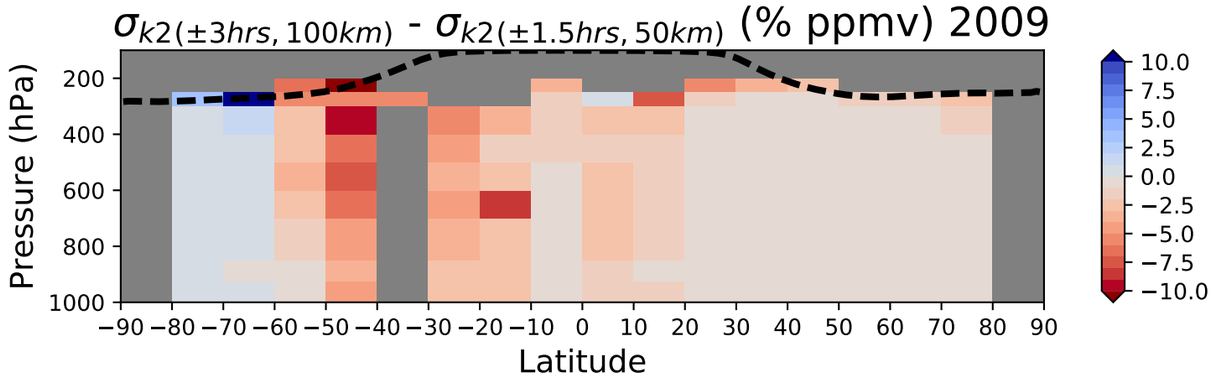


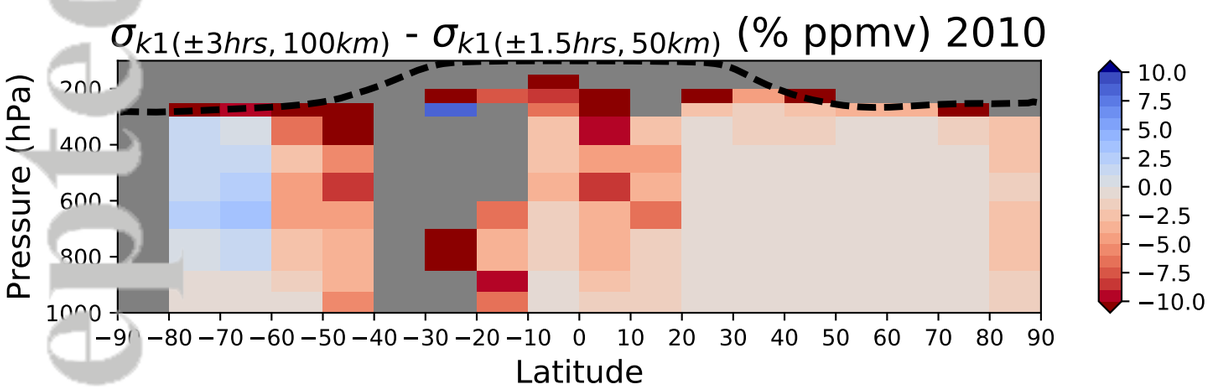


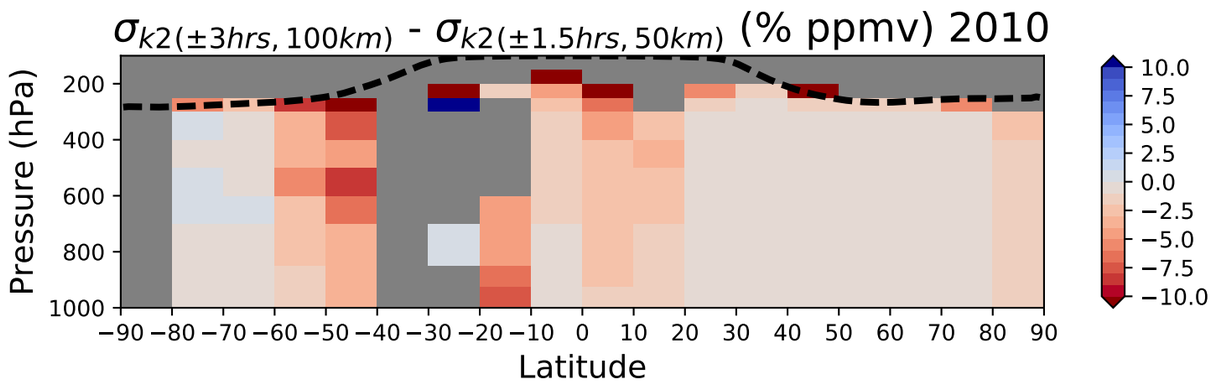


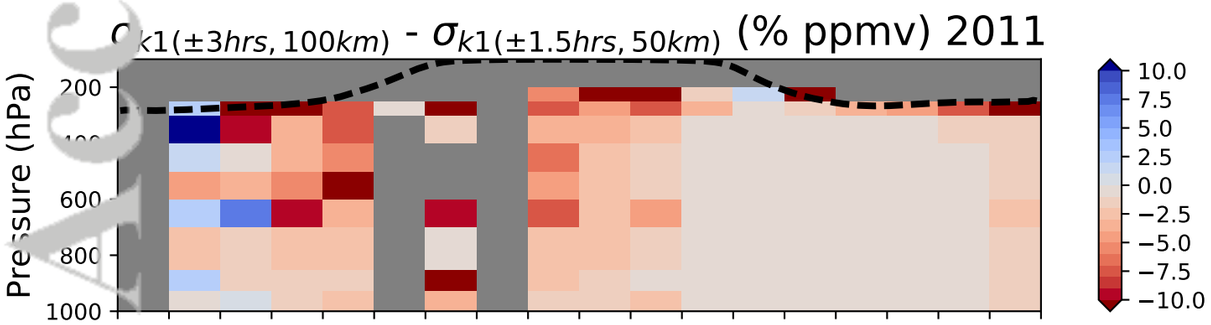


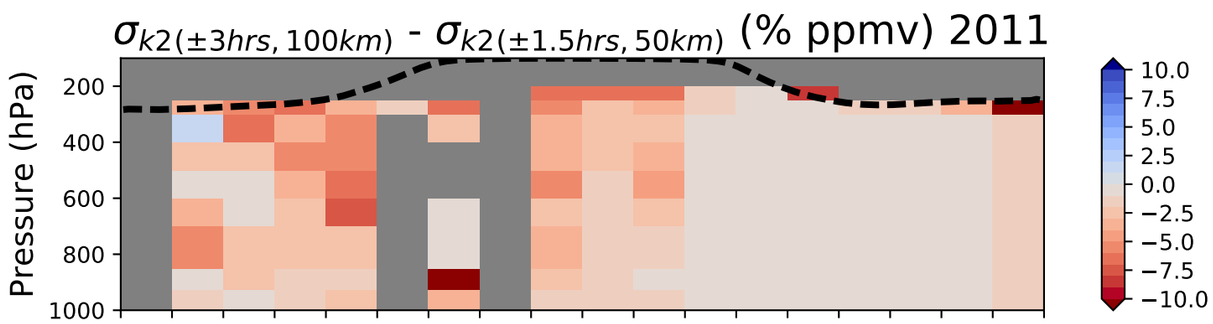


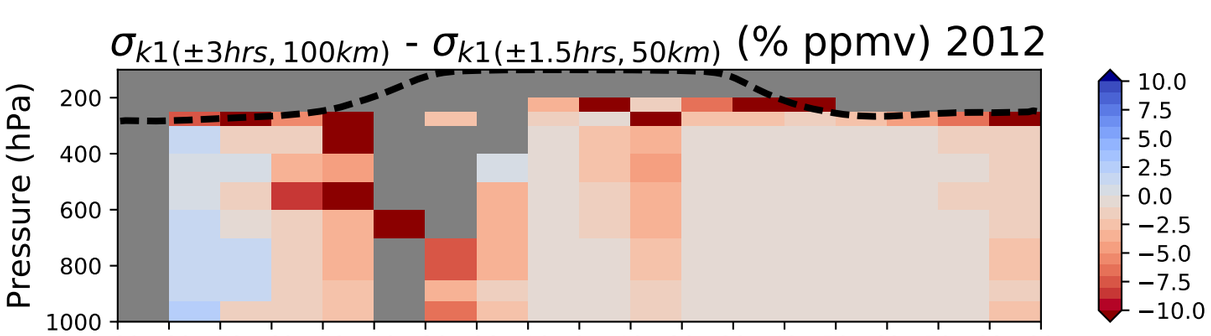


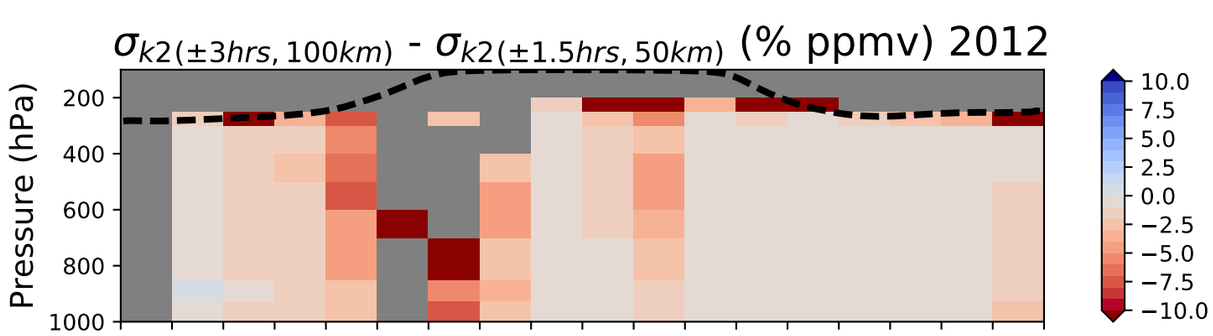


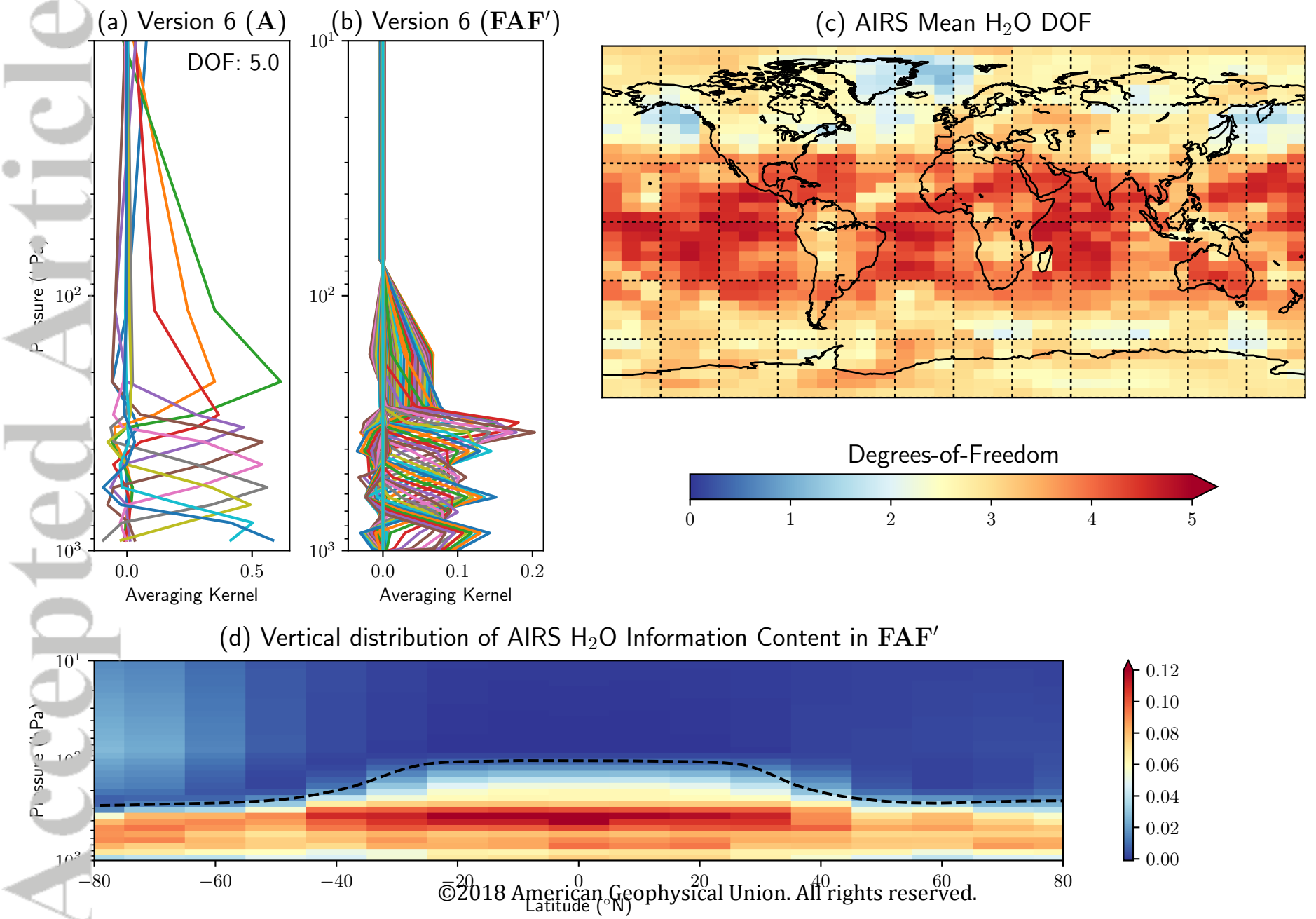












(a) GRUAN Radiosonde Stations (2007-2012)

(b) CRM Radiosonde Stations (2007-2012)