PDFs of Tropical Tropospheric Humidity: Measurements and Theory

Ju-Mee Ryoo

Department of Earth and Planetary Sciences, Johns Hopkins University, Baltimore, MD 21218, USA

Takeru Igusa

Department of Civil Engineering, Johns Hopkins University, Baltimore, MD 21218, USA

Darryn W. Waugh

Department of Earth and Planetary Sciences, Johns Hopkins University, Baltimore, MD 21218, USA

Resubmitted to the Journal of Climate on October, 2008

Corresponding author address:

Ju-Mee Ryoo
Department of Earth and Planetary Sciences
Johns Hopkins University
Baltimore, MD 21218, USA.
(Tel) 410-516-8883
(fax) 410-516-7933
Email: jryoo1@jhu.edu
Abstract

The spatial variations in the probability distribution functions (PDFs) of relative humidity (RH) in the tropical and subtropical troposphere are examined using observations from the Atmospheric Infrared Sounder (AIRS) and the Microwave Limb Sounder (MLS) instruments together with a simple statistical model. The model, a generalization of that proposed by Sherwood et al. (2006), assumes the RH is determined by a combination of drying by uniform subsidence and random moistening events, and has two parameters: $r$, the ratio of the drying time by subsidence to the time between moistening events, and $k$, a measure of the randomness of the moistening events. The observations show that the characteristics of the PDFs vary between the tropics and subtropics, within the tropics or subtropics, and with altitude. The model fits the observed PDFs well, and the model parameters concisely characterize variations in the PDFs and provide information on the processes controlling the RH distributions. In tropical convective regions, the model PDFs that match the observations have large $r$ and small $k$, indicating rapid random remoistening, which is consistent with direct remoistening in convection. In contrast, in the non-convective regions there are small $r$ and large $k$, indicating slower, less random remoistening, consistent with remoistening by slower, quasi-horizontal transport. The statistical model derived will be useful for quantifying differences between, or temporal changes in, RH distributions from different datasets or models, and for examining how changes in physical processes could alter the RH distribution.
1 Introduction

Water vapor plays a crucial role in the Earth’s climate system, and the potential for water vapor feedbacks is a major challenge for understanding and predicting climate change. It is therefore important to know the distribution of atmospheric water vapor, and the processes controlling this distribution. As the radiative effect of water vapor is roughly logarithmic in the concentration (e.g., Spencer and Braswell 1997, Held and Soden 2000, Pierrehumbert et al. 2006) it is important to know the full distribution of atmospheric water vapor, and not just the mean and variance.

In recent years there have been several studies that have aimed at addressing these issues by examining the probability distribution functions (PDFs) of observed tropospheric humidity (Soden and Bretherton 1993; Gierens et al. 1999; Spichtinger et al. 2002; Zhang et al. 2003; Sherwood et al. 2006; Luo et al. 2006; Ekström et al. 2007; Read et al. 2007). These studies all showed that the water vapor PDFs are very broad and non-Gaussian, but the characteristics of the PDFs have varied between studies. For example, Soden and Bretherton (1993) noted a lognormal distribution for 200-500 hPa upper-tropospheric humidity, whereas Zhang et al. (2003) and Luo et al. (2006) have reported bimodal PDFs. It is unclear whether the differences in reported PDFs are caused by differences in the instruments making the measurements (including remote versus in-situ measurements), differences in the space-time resolution of the data used, or whether the differences are due to the different regions and time periods considered in the studies.

We investigate some of these issues by examining PDFs for different subregions of the tropics and subtropics (and the whole tropics), using daily measurements from three satellite instruments: the Atmospheric Infrared Sounder (AIRS) instrument on the Aqua satellite (Aumann et al. 2003), and the Microwave Limb Sounder (MLS) instruments on the Upper Atmospheric Research satellite (UARS) (Read et al. 2001), and Aura satellite (Read et al. 2007). We examine how the PDFs vary between regions, and between
We also examine whether the observed PDFs can be reproduced by simple theoretical models. One is the model recently derived by Sherwood et al. (2006). In this model the relative humidity (RH) is assumed to be determined by uniform subsidence and random re-moistening process. These simple assumptions are supported by studies that show tropical humidity can be reproduced using the large-scale to advect a water tracer with no microphysics other than condensation when RH exceeds 100% [e.g., Sherwood 1996, Salathe and Hartman 1997, Pierrehumbert 1998, Dessler and Sherwood 2000, Galewsky et al. 2005]. In the Sherwood et al model the PDF of RH has a simple algebraic form (with exponent related to the ratio of drying to remoistening time). We also consider a generalization of this model that includes an additional parameter, which can be interpreted as a measure of the randomness of the re-moistening events. It will be shown that this generalized model fits the observations better than the model that Sherwood et al. (2006) proposed, and can capture the spatial variations in the PDFs.

The data used in this study are described in the next section. The theoretical model is derived in Section 3. In Section 4, the spatial and vertical variations of AIRS PDFs of RH are examined and compared with the theoretical model. Measurements from other instruments are considered in Section 5, to check the robustness of the results based on AIRS measurements. Finally, conclusions and future work are discussed in Section 6.

2 Data and Methods

2.1 Data

The AIRS data examined are level-2 data retrievals (version 5) which have been binned into a 1° by 1° latitude-longitude grid as in Gettelman et al. (2006). The level 2 data include temperature and water vapor at vertical resolution around 1-2 km and horizontal
resolution of around 50 km. Relative humidity (RH) is computed from AIRS water vapor and temperature retrievals as in Gettelman et al. (2006): Relative humidity over water is calculated for temperatures > \(273K\), relative humidity over ice for temperatures < \(253K\), and a linear combination between these temperatures. The AIRS water vapor is an average for a layer between two pressure levels, and are archived on 28 levels from the surface to the mesosphere. Following the AIRS convention each layer is referenced by the pressure at the bottom of the layer, e.g. RH at 250 hPa corresponds to the RH averaged from 250 to 200 hPa. We examine AIRS RH from 2002 to 2007 on layers with bottoms between 850 hPa and 200 hPa.

The UARS-MLS instrument made upper tropospheric water measurements from September 1991 to July 1997 (with limited coverage after 1994). Measurements were made every 4.1° along an orbit track, with 15 orbits per day, on four pressure surfaces between 147 hPa and 464 hPa (Read et al. 2001). We examine here version 4.9 RH measurements on the 215 hPa surface for northern winters (DJF) 1991/92 to 1993/94. These measurements have approximately 3km vertical resolution, with accuracy and precision of 22% and 10% respectively (Read et al. 2001).

Aura-MLS provides water vapor measurements since July 2004. Water vapor mixing ratios are retrieved from calibrated Aura-MLS observations (Livesey et al. 2006), and relative humidity with respect to ice is computed from water vapor and temperature retrievals as for AIRS (Read et al. 2007). We examine here Aura-MLS version 2.2 RH measurements at 215 hPa for the northern winters 2005/06 and 2006/07.

2.2 Probability Density Functions

Our primary method of analysis in this paper is examination of PDFs. We form and examine PDFs of RH from each of the above datasets. An important issue when calculating and examining PDFs is the space-time scales included in the PDFs. The
characteristics of the PDFs can be sensitive to these scales. This is illustrated in Fig. 1 which shows PDFs of AIRS RH measurements at 400 hPa, for different spatial regions and temporal resolution.

Figure 1(a) compares PDFs for all AIRS data at 400 hPa within the “whole tropics” (30°S-30°N, 0-360°E) for the 2002/03 to 2006/07 northern winters (December to February), using either daily data (solid curve) or monthly-mean data (dashed curve). In both cases the data are on a 1° longitude by 1° latitude grid. There are significant differences between the two PDFs, even though the same measurements where used. The PDF using daily data is much broader with a large peak at low RH and a second, much broader, peak at high RH. In contrast, the PDF of monthly-mean data is much narrower (no very low or very high values) and has a broader peak at moderate RH. In other words, the averaging process in computing monthly-mean data tends to remove extreme values and produce a PDF with a peak close to the long-term average.

The characteristics of the PDFs also depend on the regions considered. This can be seen by comparing the three panels in Fig. 1. Panels (b) and (c) show PDFs using same data source as panel (a) except for two different 10° latitude by 20° longitude sub-regions. Again there are significant differences between PDFs using daily and monthly mean data. There are also, as noted by Ryoo et al. (2008), significant variations between regions. Whereas the PDF for the whole tropics is bimodal, the PDFs for the subregions are unimodal with peaks at high (panel b) or low (panel c) RH. These differences are examined in more detail below.

As mentioned in the Introduction, the logarithmic dependence of water vapor absorption to the water vapor concentration means that it is important to quantify the full variation in RH. Because of this we focus on PDFs of daily data, rather than monthly-mean data which averages out extreme values. Furthermore, as we would also like to link the characteristics of the PDFs to the processes controlling the humidity distribution, and the key processes vary between regions (e.g., Ryoo et al., 2008), we examine the PDFs of
10° latitude by 20° longitude subregions as well as PDFs of the whole tropics.

2.3 Statistical Model

As well as examining the PDFs from the various measurements we also compare these observed PDFs with a statistical model for distributions of RH. This model is a generalization of the model derived in Sherwood et al. (2006) (hereinafter “S06”). As in the S06 model, the generalized model is based on the “time of last saturation” paradigm for tropospheric humidity, in which a parcel’s humidity is equal to the lowest saturation value it has experienced since it has left the boundary layer.

In deriving their model, S06 assumed there is uniform subsidence, and the relative humidity can then be approximated as

$$R = \exp \left( -\frac{t}{\tau_{\text{dry}}} \right),$$

where $t$ is the time since the parcel was last saturated and $\tau_{\text{dry}}$ is the uniform drying time by subsidence. S06 further assumed that “re-moistening” of parcels occurs by random moistening events which are independent of the parcel history (i.e., a Poisson process). The PDF of time between moistening events (time from last saturation) is then

$$P(t) = \exp(-t/\tau_{\text{moist}})/\tau_{\text{moist}},$$

where $\tau_{\text{moist}}$ is the mean time between remoistening events. Combining equations (1) and (2) yields the PDF of the relative humidity $R$:

$$P(R) = rR^{r-1}$$

where $r = \tau_{\text{dry}}/\tau_{\text{moist}}$ is the ratio of drying to moistening time. This distribution is a
special case of the Beta distribution

\[ P_{\text{Beta}} (R) = \frac{R^{\alpha - 1} (1 - R)^{\beta - 1}}{B(\alpha, \beta)} \quad (4) \]

where \( B(\alpha, \beta) \) is the Beta function (Wilks 1995). The PDF in (3) is the Beta distribution with \( \beta = 1 \) and \( \alpha = r \). The corresponding cumulative distribution is \( C(R) = R^r \).

In this paper we consider a two-parameter generalization of the PDF (3). If we were only interested in curve-fitting the PDFs, then the standard statistical procedure would be to use the two-parameter form of the Beta PDF where we do not restrict \( \beta \) to be equal to 1 as in (3). This approach, however, cannot be explained in terms of the physics of the underlying phenomena. A more physically based approach which keeps the concepts of uniform subsidence and random remoistening events is to retain equation (1) to model the relative humidity in terms of the time \( t \) of last saturation, and to generalize the PDF in (2) for the time between events. The natural generalization in this context is to use the gamma PDF, given by

\[ P(t) = \frac{\exp(-kt/\tau_{\text{moist}}) t^{k-1} k^k}{\tau_{\text{moist}}^k \Gamma(k)}, \quad (5) \]

rather than the exponential PDF in (2). This PDF still represents a Poisson process but includes an additional parameter \( k \), which is a measure of the randomness of the remoistening events. For \( k = 1 \) this reduces to the exponential form (2), while larger \( k \) corresponds to less random events. For very large \( k \) (of the order of 100 or larger), the randomness is so small that the events become nearly periodic, with period \( \tau_{\text{moist}} \). For the data considered herein, we find that \( k \) is less than 10, which corresponds to moderate to large randomness in the remoistening events.
The PDF of the relative humidity $R$ is now given by

$$P(R) = \frac{k^{kr} R^{kr-1}}{\Gamma(k)} (-\log R)^{k-1}.$$  \hspace{1cm} (6)

and the CDF is

$$C(R) = 1 - \gamma\left(-\frac{\log R}{\tau_{dry}}\right).$$  \hspace{1cm} (7)

where $\gamma$ is the incomplete Gamma function. Both equations reduce to the original S06 distributions in the limit $k = 1$.

In the following we refer to distributions of RH given by (3) as the S06 model, and distributions given by (6) as the “generalized” model. Below we compare PDFs of the form (6) with the PDFs of the observations discussed in the previous section. However, before this we examine the characteristics of PDFs given by (6).

$P(R)$ for several values of $r$ and $k$ are illustrated in Fig. 2. This shows that a wide range of PDFs can be formed by varying the two parameters. For example, the location of the peak RH varies with $r$: the peak occurs at RH = 0 for $r < 1$ regardless of the value of $k$, whereas the peak is at high RH for $r > 1$. It can be shown that the peak of the PDF occurs at

$$R_{\text{peak}} = \exp\left(-\frac{k-1}{kr-1}\right)$$  \hspace{1cm} (8)

for $k > 1$. This shows that as $r$ increases (for fixed $k$) the peak occurs at larger values, with $P \to 0$ as $r \to 0$ and $P \to 1$ as $r \to \infty$. This differs from the S06 model where the peaks only at RH = 0 (for $r < 1$) or 1 (for $r > 1$). Figure 2 also shows that the width of the PDFs vary with $k$, with narrower distribution for larger $k$.

It is more common to examine the mean and standard deviation of the RH distribution than the full PDFs, so it is of interest to consider the mean and standard deviation of the
above distributions. For distributions given by (6) the mean is

\[ \mu_R = \left( \frac{r}{r + 1/k} \right)^k, \]  

(9)

and the standard deviation is

\[ \sigma_R = k^{k/2}r^{k/2} \sqrt{(kr + 1)^{2k} - (kr + 2)^k k^{k/2} r^k} \]

(10)

The variation of \( \mu_R \) and \( \sigma_R \) with \( r \) and \( k \) is shown in Fig. 3. Both \( \mu_R \) and \( \sigma_R \) depend on \( r \) and \( k \), but \( \mu_R \) depends primarily on \( r \) and is only weakly dependent on \( k \) (Fig. 3(a)), and \( \sigma_R \) is primarily dependent on \( k \) and only weakly dependent on \( r \) (Fig. 3(d)). As a result \( r \) can be estimated from \( \mu_R \) using (9) with, say, \( k = 2 \), and that \( k \) can be estimated from \( \sigma_R \) using (10) with, say, \( r = 2 \).

3 AIRS PDFs

We now examine the PDFs from AIRS RH measurements and compare with the above theoretical distributions. We first investigate AIRS measurement for the 250 hPa layer during northern hemisphere winter (DJF), and then consider other seasons and altitudes.

3.1 250hPa, DJF

The symbols in Fig. 4 show the (a) PDF, and (b) corresponding CDF, of AIRS RH for all northern winter data (2002-2007) within the tropics and subtropics (30°S-30°N, 0-360°E). The observed PDF is broad and asymmetric with a peak around 20% and long tail of moist air. Such broad distributions of upper tropospheric RH have also been observed in data from other satellite instruments, e.g., MLS, Global Positioning Satellite (GPS) (Sherwood et al. 2006) and Odin (Ekström et al. 2007).
Also shown in Fig. 4 are fits to the AIRS data for the S06 and generalized models. The values of $r$ and, in the case of the generalized model, $k$ are found by minimizing the mean square error between model and observed PDFs. We choose this method over other standard statistical techniques, such as the maximum likelihood method, because it is not overly sensitive to the low-probability regions of the PDF. These comparisons show that the generalized model is a better fit to the observed PDF and CDF than the original S06 model. In particular, the generalized model can reproduce the peak of the PDF at RH close to 20%, whereas the peak of the S06 model occurs at RH = 0% (as noted in the discussion of Fig. 1, the peaks of the PDFs from the S06 model can only appear at RH equal to 0% or 100%).

This comparison indicates that inclusion of an additional parameter in the statistical model greatly improves the agreement with the AIRS data. Furthermore evidence of this improved agreement is presented below. Note that S06 examined CDFs from GPS and the two MLS instruments, and not AIRS data. The agreement between the S06 model and GPS PDFs is better than that in 4(b), but the disagreement with MLS PDFs is similar to that for AIRS. The differences between different data sets are examined further in Section 5.

We now consider the PDFs for smaller subregions than the whole tropics. As the differences between distributions is more visible if considering the PDFs rather than the CDFs, in the remainder of the paper we focus on the PDFs of RH, and rather than the CDFs, as considered by S06, but similar results are obtained if CDFs were used (i.e., the best-fit values of $r$ and $k$ are very similar for fits to PDFs or CDFs). Figure 5 shows the PDFs of AIRS RH for six 10° latitude by 20° longitude regions in the tropics (5°S-5°N) and the subtropics (15°N-25°N). As noted by Ryoo et al. (2008), the PDFs vary between regions, both with longitude and between the tropics and subtropics. The location of the peaks of the PDFs varies from around 20% to around 60%, and the width and skewness of the distributions also vary. The fits to the AIRS PDFs for the various subregions for the
S06 \((k = 1)\) and generalized (variable \(k)\) models are also shown in Fig. 5. The generalized model can fit the data for all subregions. This includes not only the peak values but also the range and skewness of the PDFs. Some differences between generalized model and observed PDFs can be seen for high RH, especially in tropical eastern Pacifics \((5^\circ \text{S}-5^\circ \text{N}, 80^\circ \text{W}-100^\circ \text{W})\). However, these are relatively small differences. As for the PDF of the whole tropics, the S06 model cannot match the features of the observed PDFs.

Figure 5 shows that the PDFs of AIRS RH for different tropical or subtropical regions can be represented by the theoretical generalized model. The variations in the PDFs can hence be summarized by variations in \(r\) and \(k\). Figure 6 shows the longitudinal variation of (a, d) \(r\), (b, e) \(k\), and (c, f) error \(\epsilon\) (see below) for the S06 and generalized model fits to PDFs for \(10^\circ\) by \(20^\circ\) regions in the (upper panels) subtropics or (lower panels) tropics. As could be expected from Fig. 5, both \(r\) and \(k\) vary with longitude and latitude.

The value of \(r\) for the tropics is generally larger than in the subtropics, and the longitudinal variation of \(r\) is much larger in the tropics than the subtropics (Fig. 6(a), (d)). The \(r\) for the S06 and generalized models have very similar spatial variations and even quantitative agreement, except in tropical Indian \((\sim 40^\circ \text{E})\), western Pacific \((\sim 120^\circ \text{E})\) and Atlantic \((\sim 50^\circ \text{W})\) oceans. This is somewhat surprising given the different shapes of the PDFs for the S06 and generalized models (e.g., Fig. 5). The similarity in \(r\) between the fits using the S06 and generalized model can be understood in terms of relationships between \(r\) and the mean value \(\mu_R\). As discussed in Section 2, \(r\) is closely related to \(\mu_R\), with only weak sensitivity to \(k\). Hence, for both \(k = 1\) (S06 model) and \(k > 1\) (generalized model) \(r\) will depend primarily on the mean, and not other characteristics, of the RH distributions.

The parameter \(k\) also varies with both longitude and latitude, see Fig. 6(b), (e). In the subtropics \(k\) varies between 2 and 6, while in the tropics \(k\) varies from 3 to 10. In both the tropics and subtropics the longitudes with maximum in \(k\) are the longitudes where \(r\) is a minimum, e.g. in the tropics large \(k\) and small \(r\) occur around \(60^\circ \text{E}\) and \(90^\circ \text{W}\).
The $r$ and $k$ shown above were determined by minimizing the error between the observed and modeled PDF. To estimate the uncertainty in these estimates, a moving blocks bootstrap analysis (Künsch 1989) has been performed (where moving time blocks of data are used to account for correlation in time).

The vertical bars in Fig. 6 show the uncertainty ($\pm 1\sigma$) in the calculated $r$ and $k$. The uncertainty for $r$ is very small in the both subtropics and tropics, and much smaller than the spatial variations in $r$. The uncertainty in $k$ is larger but still less than spatial variations in $k$.

As discussed above the generalized model fits the data better than the S06 model. To quantify this, and the spatial variations in how well the models fit the data, we calculate the root-mean square error $\epsilon$ between the PDFs of the model and data. Figure 6(c) and (f) show that the error for the generalized model is about ten times less than that of S06 model. Also, longitudinal variations in $\epsilon$ for the S06 model are very similar to the variations of $k$ of generalized model. This is because the S06 model is identical to the generalized model when $k = 1$, and as $k$ of the generalized model increases the S06 model deviates from the data, resulting in larger $\epsilon$ when $k$ is larger.

To examine the spatial variations of $r$ and $k$ further, and the relationship with the $\mu_R$ and $\sigma_R$, we compare maps of these fields. Figure 7(a) and (b) show the spatial variations $r$ and $k$. The parameters have similar spatial variations with small (large) $k$ in regions of large (small) $r$. Specifically, there is large $r$ and small $k$ in the tropical western Pacific ($5^\circ S$-$5^\circ N$, $120^\circ E$) and tropical America($5^\circ S$-$5^\circ N$, $60^\circ W$), and small $r$ and large $k$ in the tropical eastern Pacific ($5^\circ S$-$5^\circ N$, $120^\circ W$) and northern subtropical mid-Pacific ($15^\circ N$-$25^\circ N$, $150^\circ E$). The maps of $\mu_R$ and $\sigma_R$ calculated from the AIRS data are shown in Fig. 7(c) and (d). As expected from Fig. 3 and related discussion, there is a strong resemblance between maps of $\mu_R$ and $\sigma_R$ to those of $r$ and $k$, respectively, i.e., there is large $\mu_R$ where $r$ is large and large $\sigma_R$ where $k$ is small.

The variations in $r$ and $k$ could provide insight into the variations in the characteris-
tics of the moistening processes. If the drying time is assumed constant, large $r$ and small $k$ indicates rapid, random remoistening, whereas small $r$ and large $k$ implies slower, more regular moistening processes. There are large $r$ and small $k$, and hence by the above arguments rapid, random remoistening, in the tropical convective regions. In contrast, there are small $r$ and larger $k$ in the dry regions, indicating slower more regular remoistening.

We conjecture that these variations in the remoistening are consistent with our understanding of the physical processes. In the tropical convective regions the moistening is thought to occur by direct rapid moistening by vertical transport in convective systems, whereas lateral mixing by ‘large-scale’ advection plays a larger role in the remoistening the drier tropical and subtropical regions (e.g., Sherwood, 1996, Salathe and Hartmann, 1997, Pierrehumbert 1998, Waugh 2005, Ryoo et al. 2008). This lateral mixing is produced by transient wave activity (Pierrehumbert and Roca, 1998), including Rossby wave breaking along the tropopause (Waugh 2005, Ryoo et al., 2008), and is slower and more regular than convection. The process is not so regular as to be periodic (which would correspond to $k$ of the order of 100), but is considerably less random than processes where $k$ is of order 1.

3.2 Seasonal and Altitudinal Variations

The above analysis considered only northern winter data at 250 hPa. We now examine the seasonal and altitudinal variations of AIRS PDFs, and whether these PDFs are still well fit by the generalized model.

Figure 8 shows PDFs of AIRS data for the whole tropics ($0^\circ$-360$^\circ$E, 30$^\circ$S-30$^\circ$N) for different seasons and at several different altitudes. At all levels there are only weak seasonal variations. There are, however, large variations in the shape of PDFs with altitude. At 400 and 600 hPa the peak occurs at or less than RH = 10%, which is much drier than the peak at 250 hPa, whereas at 850 hPa there is limited dry air and the dry
peak occurs around RH = 40%. In contrast to 250 hPa, the PDFs are bimodal at 400, 600, and 850 hPa, with a second moist peak at RH ≃ 70-80%. Also shown in Fig. 8 are the fits to the data using the generalized model. Because the seasonal variations are small, we only show the fit for DJF data. The model can capture the general characteristics of the vertical variations, in particular the variation in the peak values. As at 250 hPa, the generalized model is a much better fit than the S06 model at the lower levels (not shown). However, the generalized model cannot reproduce the observed bimodal PDFs, and the disagreements between the observed and generalized model is largest when the observed PDFs are most bimodal.

The PDFs shown in Fig. 8 come from the collection of dry and moist RH over the whole tropical regions, which includes moist air from convective region and dry air from non-convective region. When we look at the PDFs for smaller 10° latitude by 20° longitude subregions, most of them are unimodal. For example, in tropical convective regions (e.g. western Pacific (5°S-5°N, 100°E-140°E)), the PDFs have a peak in high RH, while in nonconvective regions like the eastern Pacific (5°S-5°N, 120°W-160°W) the peak of PDF is at low RH. These observed PDFs for the subregions can be well fit by the generalized model. When these different regions are combined, the resulting PDF is simply the average of the PDFs of all subregions. Hence, given that the peaks of the PDFs of convective and nonconvective subregions are at different RH values, bimodality in the PDF of the combined region is expected. There are some subregions at the edge of the tropical convective regions where bimodal PDFs occur (e.g. 5°S-5°N, 40°W-60°W), but the distinction between low values and high values is small (not shown). In addition, this bimodal behavior is tightly related to temporal variations due to the movement of convection into or out of a region, where in this case the PDF is a temporal rather than a spatial average of PDFs with variable locations of peaks.

The bimodal distributions in the middle troposphere are consistent with Zhang et al. (2003), who observed bimodal features in the PDFs of precipitable water using monthly-
mean data. Furthermore, if monthly mean AIRS data are used, rather than daily values, the PDFs for mid-tropospheric RH look similar to the PDFs of precipitable water averaged over 500-300 hPa shown in Zhang et al. (2003), e.g compare Fig. 1(a) with Figure 8 of Zhang et al. (2003). (Difference is due to differences in measurements, regions of interest and time periods. Similarity is much more clear when we compare 3-monthly mean data with them (not shown).)

The vertical variations of $r$, $k$, and $\epsilon$ for PDFs, for the whole tropics and whole year, are shown in Fig. 9. There is a minimum in $r$ in the mid-troposphere, for both the S06 and generalized models. A mid-level minimum in the mean RH from AIRS has already been reported (Gettelmen et al. 2006; Ryoo et al. 2008), and the minimum in $r$ could be expected given the close relationship between $r$ and $\mu_R$. Also, S06 found a minima at the same altitude in their calculations of $r$ from GPS data.

The parameter $k$ also varies in the vertical, with an increase with altitude above 500 hPa. As larger $k$ reflects less random remoistening processes, so this increase of $k$ suggests the moistening processes in the upper troposphere are more affected by less random and relatively slow large-scale process such as subsidence, rather than by rapid moistening by convective updraft from the surface.

The error, $\epsilon$, between model and data are shown in Fig. 9(c). As at 250 hPa, $\epsilon$ for the S06 model is much larger than the generalized model, and the variation of $\epsilon$ for the S06 model is similar to the variation in $k$. The vertical variation of $\epsilon$ for the generalized model differs from that of $r$ and $k$, with local maximuma are those between 300 and 400 hPa and between 600 and 700 hPa. These are altitudes where the PDFs are most bimodal.

Similar vertical variations of $r$ and $k$ occur for sub-regions, although there are variations with longitude, see Fig. 10. As expected from the discussion in Section 2, the vertical and longitudinal variations of $r$ and $k$ are similar to those of $\mu_R$ and $\sigma_R$, respectively (see Figure 1 of Ryoo et al. (2008)). It is interesting to note that in the tropical convective regions ($100^\circ$E-140$^\circ$E, 40$^\circ$W-60$^\circ$W) there are local minima in both $r$ and $k$ at
mid-levels (∼ 400 hPa), whereas in non-convective regions there is a maximum in $k$ at mid-levels.

The vertical variations of $r$ and $k$ in convective regions are consistent with analysis of radiative processes and the energy balance, which show a minimum in convective detrainment at mid-levels (Hartmann and Larson 2002; Folkins et al. 2002, 2007). This analysis indicates that in convective regions the air at and above 200 hPa is comprised mainly of very moist air parcels that have just detrained from convection, whereas around 400 hPa there is a combination of moist air from recent detrainment and very dry air that has subsided from 200 hPa. As a result the remoistening time at mid-levels is longer than aloft, resulting is smaller $r$ (and $\mu_R$). Also, there is larger variability in the moisture and more regularity in remoistening at mid-levels, resulting in a larger $k$.

4 Other Data

Having examined PDFs from AIRS we now consider the PDFs of RH measurements from other instruments to test the robustness of the above results. We first compare with measurements made by UARS-MLS (1992-1994) and Aura-MLS (2005-2007) instruments. The latter overlaps with the AIRS data record enabling a comparison of PDFs for the same time periods. We also compare our results with those shown in S06 for GPS data.

For our analysis of MLS measurements we focus on northern winter (DJF) measurements at 215 hPa (which can be compared with the AIRS 200-250 hPa layer). Figure 11 shows the PDFs of AIRS, UARS-MLS (1992-1994) and Aura-MLS (2005-2007) RH for subregions in the tropics (5°S-5°N) and subtropics (15°N-25°N). Here two different AIRS PDFs are shown. One was formed using all available data and the other using only data sampled at the same locations as Aura MLS. PDFs from different data sets show good agreement except the tropical convective region (5°S-5°N, 120°E-140°E). In this region the AIRS PDFs are narrower with peak around 60%, whereas the UARS-MLS and Aura-
MLS PDFs are broader with low values less than 20% and high values larger than 100% (see also Read et al. 2007). According to Fig. 3(d), this would imply that $k$ from AIRS should be considerably larger than $k$ from both MLS measurements.

Figure 12 compares the longitudinal variation of $r$ and $k$ for the tropics (5°S-5°N) and subtropics (15°N-25°N), for PDFS of AIRS, Aura MLS, and UARS MLS measurements shown in Figure 11. Consider first the parameter $r$. There is good agreement in $r$ from all three datasets even though they cover different years. All three datasets show generally higher values of $r$ in the tropics than the subtropics, with largest values in the tropical convective regions (around 120°E and 60°W), and larger longitudinal variations in the tropics than the subtropics. The largest disagreements between the values of $r$ from the different measurements are in the tropical locations with local maximum in $r$, where $r$ from AIRS is generally larger than from both MLS measurements. This is true even if the same measurement locations are used for the AIRS and Aura-MLS PDFs, indicating that this is a difference in the measurements and not due to differences in the sampling or different years.

The agreement between $k$ from the different datasets is not as good as for $r$. There is qualitative agreement in the longitudinal variations of $k$, but there are quantitative differences. In the tropics, $k$ from AIRS are consistently larger than those from both MLS measurements, even when sampling the same air as Aura-MLS. The largest difference between AIRS and MLS occur in tropical convective regions (see above). There is better agreement between the two MLS datasets, although differences occur when $k$ is larger (with larger $k$ from Aura-MLS). The differences in subtropical $k$ among the datasets is not as consistent as the tropics, but the general tendency is the same, e.g., $k$ is generally larger from AIRS than MLS.

The above comparison focused only on upper tropospheric measurements. As a check on the robustness of the vertical variations we briefly compare our results with the GPS data shown in S06. As discussed above, $r$ from the GPS data show a minimum at the same
height as that from AIRS (see Figure 6 of S06). The values from GPS are smaller than those of AIRS, e.g., for the S06 model the $r$ at 400 hPa is 0.42 from GPS (for Jan 2002 measurements) compared with 0.53 from data (for DJF 2002/03-2006/07 measurements). S06 fit their model to CDFs rather than PDFs and used a different criteria to determine the best fit $r$, but tests show that neither of these cause significant differences in the estimates of $r$. The difference in our estimate $r$ from AIRS and S06 calculations of $r$ from GPS data is thus due to actual differences in the PDFs from the two datasets.

The cause of the above differences between the different data sets are not known, and more research is required to identify these causes. However, even though there are some quantitative differences in the values of $r$ and $k$ for PDFs from different datasets there is overall good agreement in the spatial variations of $r$ and $k$, both horizontally and vertically. This gives us some confidence in general conclusions primarily based on analysis of the AIRS data.

5 Conclusions

Measurements of tropospheric relative humidity (RH) from three different satellite instruments indicate that the probability distribution functions (PDFs) of daily RH are broad and non-Gaussian. This applies not only for PDFs of the whole extended tropical regions (30°S-30°N) but also for PDFs of smaller 10° latitude by 20° longitude subregions. Although the “local” PDFs are all broad, the location of the peak, the skewness, and the width vary between the tropics and subtropics, within the tropics or subtropics, and with altitude.

The observed PDFs for all subregions can be well fit using a simple statistical model that is a generalization of that proposed by Sherwood et al. (2006). This model assumes the RH is determined by a combination of drying by uniform subsidence and random remoistening events, and has two parameters: $r$, the ratio of drying time (via subsidence)
and remoistening time, and \( k \), a measure of randomness of the remoistening.

The parameters \( r \) and \( k \) not only provide a concise way to characterize the RH distributions, but also may provide insight into the processes controlling on the RH distributions. In the tropical convective regions there is large \( r \) and small \( k \) in the upper troposphere, indicating rapid, random remoistening in these regions. In contrast, in dry regions in subtropics and tropical eastern Pacific there is small \( r \) and large \( k \), indicating slower, less random remoistening there.

We conjecture that these variations in the remoistening process are consistent with our understanding of the physical processes in different regions. Previous studies have shown that convection and vertical mixing play the key role in regulating humidity near tropical convective regions, but remositenen in the subtropics comes from lateral advection of moist air from convective regions (e.g., Sherwood 1996, Salathe and Hartman 1997, Pierrehumbert 1998, Dessler and Sherwood 2000, Galewsky et al. 2005, Dessler and Minschwaner 2007). Thus, in tropical convective regions we expect direct rapid remoistening by rapid, random vertical motions, whereas in dry, non-convective regions the remoistening occurs by slower, more regular lateral mixing by large-scale advection. The \( r \) and \( k \) that fit the observed PDFs also vary in the vertical. In the tropics \( r \) and \( k \) both have a mid-level (300-500 hPa) minimum, indicating slower and more regular remoistening in the middle troposphere. This is consistent with a mid-level minimum in convective detrainment and mid-level air being a mixture of recently detrained moist air and very dry air that has subsided from below (Folkins et al. 2002).

Although the satellite data sets considered here show a consistent spatial variation in the PDFs, there are some quantitative differences. For example, the MLS PDFs are generally broader than the AIRS PDFs with higher probability of low RH and high RH in the MLS data (and as a result \( k \) is smaller from MLS data). The magnitude of these differences varies with location, and in some regions the differences are very small (see Figures 10 and 11). The cause of these differences needs to be examined further. It will
also be important to consider other water vapor data sets, in particular those from in-situ measurements. Luo et al. (2007) recently presented PDFs of UT RH from the MOZAIC aircraft program. These PDFs are often bimodal, and appear to differ from the AIRS and MLS PDFs for similar regions and seasons. More analysis is needed to quantify and understand the differences between different data sets.

As discussed above, the spatial variations in $r$ and $k$ appear consistent with our understanding of the physical processes controlling RH distribution. However, this is primarily a qualitative comparison and a more quantitative link between the different physical processes and the parameters $r$ and $k$ is needed. One approach to do this might involve using trajectory-based water vapor simulations. Previous studies have shown that trajectory-based simulations can reproduce upper tropospheric RH observations (Pierrehumbert and Roca 1998; Dessler and Sherwood 2000; Waugh 2005; Dessler and Minschwaner 2007). Analysis of these calculations would enable some of the assumptions used to derive the statistical model to be tested, and would give an opportunity to examine the origin of moisture and the control mechanisms.

There are several potential uses of the statistical model derived here. Given that the model parameters $r$ and $k$ concisely characterize the RH distributions, fitting this model to climate model output may be useful for quantifying differences in RH distributions between climate models and observations. Similarly, the statistical model may also provide a concise way to characterize any temporal changes in simulated RH distributions (e.g., in simulations with increasing greenhouse gases). Finally, the statistical model may also be useful for exploring how changes in physical processes could alter the RH distribution.

Acknowledgments

This work was supported, in part, by a grant from NASA. We thank Andrew Gettelman for providing the gridded AIRS data, and Eric Fetzer, Annmarie Eldering, Bill
Read, and Steve Sherwood for helpful conversations and advice.
References


Zhang, C., B. E. Mapes, and B. J. Soden, 2003: Bimodality in tropical water vapor,

Figure 1: PDFs of AIRS 400hPa RH data for (a) the whole tropics (0°-360°E, 30°S-30°N), (b) a tropical convective subregion (120°-140°E, 5°S-5°N), and (c) a tropical nonconvective region (80°W-100°W, 5°S-5°N). Solid curves show PDFs using daily data, while dashed curves show PDFs from monthly-mean data. All PDFs are formed using data at resolution 1° longitude by 1° latitude.
Figure 2: (a) PDFs of RH for various $r$ and $k$ generated by generalized models: (a) $r = 0.5$, (b) $r = 1$, and (c) $r = 2$ for $k = 1, 2, 4,$ and $10$, respectively.
Figure 3: Plots of (a) mean ($\mu_R$) and (b) standard deviation ($\sigma_R$) versus $r$, and (c) mean ($\mu_R$) and (d) standard deviation ($\sigma_R$) versus $k$, for generalized model.
Figure 4: (a) PDFs of 250hPa RH and whole tropics (0°-360°E, 30°S-30°N) from AIRS data (symbol), and fits by S06 ($k = 1$; dotted) and generalized (variable $k$; solid) models. (b) As in (a) except for CDFs.
Figure 5: As in Fig. 3(a) except for subregions in the subtropics (15°N-25°N, (a) 40°E-60°E, (b) 120°E-140°E, and (c) 80°W-100°W) and tropics (5°S-5°N, (d) 40°E-60°E, (e) 120°E-140°E, and (f) 80°W-100°W).
Figure 6: Longitudinal variation (a, d) $r$, (b, e) $k$, and (c, f) error, $\epsilon$ for S06 (dotted curves) and generalized (solid) models, for subtropics (upper panels) and tropics (lower panels). The vertical bars indicate the one-sigma bounds computed by the moving blocks bootstrap distribution.
Figure 7: Maps of (a) best-fit $r$, (b) best-fit $k$, (c) mean ($\mu_R$), and (d) standard deviation ($\sigma_R$) of 250 hPa AIRS RH.
Figure 8: PDFs for whole tropics (0°-360°E, 30°S-30°N) for AIRS data at 250, 400, 600, and 850 hPa. Different colors are for different seasons, and dotted curve is fit to DJF data for generalized model.
Figure 9: Vertical variation of (a) $r$, (b) $k$, and (c) error, $\epsilon$ for fits to PDFs of whole tropics and whole year, for S06 (dotted) and generalized (solid) models.
Figure 10: The cross-section of longitudinal vs altitudinal variation of (a) $r$, (b) $k$ for the subtropics (15°N-25°N) and tropics (5°S-5°N).
Figure 11: PDFs for 3 subregions in the (a) subtropics (15°N-25°N) and (b) tropics (5°S-5°N) for 250 hPa AIRS, Aura-MLS, and UARS-MLS measurements, respectively. Dashed curves are for AIRS data sampled at Aura MLS locations.
Figure 12: Longitudinal variation (a, c) $r$, and (b, d) $k$ for generalized model fit to AIRS, Aura MLS, and UARS MLS measurements. Dashed curves are for AIRS data sampled at Aura MLS locations.