Tropospheric water vapor, convection, and climate
S. C. Sherwood, R Roca, T.M. Weckwerth, N. Andronova

To cite this version:
1. INTRODUCTION

[2] Water vapor arguably lies at the heart of all key terrestrial atmospheric processes. Humidity is essential for the development of disturbed weather, influences (directly and indirectly through cloud formation) the planetary radiative balance, and influences surface fluxes and soil moisture. Water vapor is the only radiatively important atmospheric constituent that is sufficiently short-lived and abundant in the atmosphere so as to be essentially under purely natural control, yet this control endows it with a strong positive feedback on climate changes driven by other influences. The latent heat of water vapor also accounts for roughly half the net radiative cooling of the troposphere which drives convection. Despite its central importance, work to date has not led to a universally accepted picture of the factors controlling water vapor amount, a solid understanding of the mechanisms by which it influences atmospheric processes, or even precise knowledge of its concentrations in many parts of the atmosphere, to say nothing of its trends over time.

[1] Recent progress is reviewed in the understanding of convective interaction with water vapor and changes associated with water vapor in warmer climates. Progress includes new observing techniques (including isotopic methods) that are helping to illuminate moisture-convection interaction, better observed humidity trends, new modeling approaches, and clearer expectations as to the hydrological consequences of increased specific humidity in a warmer climate. A theory appears to be in place to predict humidity in the free troposphere if winds are known at large scales, providing a crucial link between small-scale behavior and large-scale mass and energy constraints. This, along with observations, supports the anticipated water vapor feedback on climate, though key uncertainties remain connected to atmospheric dynamics and the hydrological consequences of a moister atmosphere. More work is called for to understand how circulations on all scales are governed and what role water vapor plays. Suggestions are given for future research.

Convection, which may be (cloud) has moistened the atmosphere for a recent review of atmospheric processes and propositions are specified, rather than uncertainty in how the atmosphere functions.

One of our goals is to try to clear up some of the confusion in the literature on these topics. Competing claims abound: convection moistens (or dries or has no effect on the humidity of) the troposphere, transport of hydrometeors is crucial (or unimportant) in determining humidity, and subtropical relative humidity is low because of vertical (or isentropic) transport of water vapor. Many of these will be noted in this article. Most of the confusion relates to links between water vapor and atmospheric convection. We will argue that this stems largely from ambiguity in the definitions of the mean processes or in the way various hypotheses and propositions are specified, rather than uncertainty in how the atmosphere functions.

Before looking at how humidity is controlled, however, we address the topic of how convective processes are affected locally by humidity variations. The link between these may be appreciated by comparing with the situation for atmospheric lapse rate. The tropical lapse rate is strongly controlled by atmospheric deep convection rather than radiation; the mechanism is that small changes in lapse rate (supposing the near-surface relative humidity does not significantly change) cause large changes in convective heating that rapidly pull the lapse rate back to near-neutrality [e.g., Emanuel et al., 1994]. One may also expect that to the extent that deep convective moisture transports are sensitive to changes in ambient humidity, the humidity profile could be similarly regulated. We will argue that this control is likely to be important but, unlike the case with lapse rate, only in limited regions of the atmosphere.

Since convection is any thermally direct circulation that transports buoyancy upward, the Hadley and Walker circulations are convection, strictly speaking. However, in meteorology the term is typically reserved for turbulent motions at length scales characteristic of precipitation and cumulus cloud production (~100 m to 50 km), and will be used in this way here. A key qualitative distinction is between precipitating or “moist” convection, which may be of variable heights, and nonprecipitating (shallow) convection. See Stevens [2005] for a recent review of atmospheric convection.

Convection is not isolated from the larger-scale circulations, and this makes statements about its global influence (which should be computed relative to a nonconvective world) problematic. It is not obvious what an atmosphere lacking “convection” would or could look like. Convection mixes a local region vertically via updrafts and downdrafts, normally yielding a net upward water vapor transport, and it creates hydrometeors that may independently transport water by falling and reevaporating. But precipitating convection also produces heat that energetically balances mean ascent in the convective region, inevitably making it part of a larger-scale motion that would have transported tracers with or without the smaller-scale effects (Figure 1). Any effort to separate cloud-forming and large-scale motions, or their water transports, is then arbitrary. One avenue is to dictate a truncation scale and define convective eddies and transports as the difference between the true and truncated (large-scale) wind fields and their transports. This framework is helpful in addressing confusion. For example, if “convection” is defined simply as cloud formation, then trivially, it can only dry the atmosphere through any resulting precipitation. If it is defined as convective eddies, things get more interesting. In general both the large-scale and eddy circulation components transport vapor upward, potentially enhancing specific humidity at upper levels and relative humidity at most levels, but at the same time, these motions concentrate water vapor horizontally in a way that encourages loss through precipitation. Ironically, the part of this concentration accomplished by the large-scale flow may make it look like “convection” (cloud) has moistened the atmosphere (i.e., humidity positively correlated with convective activity), but this view ignores what is happening in the rest of the domain where water is being lost and gets causality backward (precipitation is occurring because of high relative humidity). For this reason we do not get far by correlating humidity with other local variables such as surface temperature, nor by focusing only on the convective region itself. We must understand how the scales are related. One may sharpen the question by imagining an atmosphere with the same large-scale flow but no convective eddies and asking whether it is moister or drier than the fully functional atmosphere. But since convection is so inextricably linked to the energetic and dynamic transports within the atmosphere, a more useful question may be how is humidity affected by those aspects of convection that are unconstrained by energy.
and mass conservation (and therefore perhaps not well constrained in a climate model)? Enforcing these constraints has proven a fruitful avenue, as we will show, though insufficient to completely solve the problem.

This review was aided by a recent gathering of scientists at the American Geophysical Union Conference on the Role of Water Vapor in the Climate System held in Kona, Hawaii, in October 2008 [Sherwood et al., 2009], although we cannot discuss here the full range of topics related to atmospheric water vapor. After briefly reviewing the microphysics of water and current water vapor observational capabilities (sections 2 and 3), in section 4 we discuss the (two-way) interaction between water vapor and convection; in section 5, we discuss the large-scale transport and control of humidity; and in section 6, we discuss the interactions of water vapor, hydrology, and climate. Section 7 summarizes where the field stands and notes some future directions that seem promising.

2. WATER FUNDAMENTALS

Removal of a molecular substance from the liquid (or ice) phase to vapor requires the expenditure of energy, the latent heat \( L \) of vaporization (or sublimation). If the condensate has a free surface, individual molecules that have enough energy depart randomly and spontaneously to the vapor phase. The departure rate is highly temperature-dependent. Meanwhile, vapor molecules constantly collide with and stick to the surface at a rate that depends on the vapor pressure. When the two rates exactly balance the vapor pressure is said to be saturated (although “equilibrated” would be more accurate). Saturation vapor pressure \( e^* \) depends only on temperature \( T \), rising according to the Clausius-Clapeyron equation

\[
\frac{de^*}{dT} = \frac{L e^*}{R_v T^2},
\]

where \( R_v \) is the gas constant for water. The relative humidity is the ratio \( e/e^* \).

At the triple point of water \( (T = 0 ^\circ C \text{ and } p = 6.11 \text{ hPa}) \) all three phases are in equilibrium. The rate of increase \( (1/e^*)de^*/dT \) at tropospheric temperatures varies from 5.9% \( ^\circ C^{-1} \) (at 35\(^\circ\)C) to 7% \( ^\circ C^{-1} \) (at 0\(^\circ\)C) over liquid and 7.8% \( ^\circ C^{-1} \) (0\(^\circ\)C) to 17.6% \( ^\circ C^{-1} \) (−85\(^\circ\)C) over ice. Thus, the saturation vapor pressure varies by more than 4 orders of magnitude in the troposphere. The saturation water vapor mixing ratio, or ratio of water to dry air masses, is proportional to the ratio of partial pressures and also varies by 4 orders of magnitude between wet tropical regions and dry winter polar regions.

Water vapor concentrations seldom if ever rise appreciably above saturation with respect to liquid in most of the troposphere. This is because condensation nuclei are sufficiently abundant that supersaturated vapors will always find a condensation site quickly enough to limit supersaturations to at most 1%–2% (which is not detectable by currently available field instruments (see section 3)). However, the same cannot be said of ice saturation, which appears to be exceeded frequently at cold temperatures [Jensen et al., 2005; Kahn et al., 2009]. Also, freezing nuclei are sufficiently rare in the atmosphere that liquid droplets usually must cool well below 0°C to freeze, sometimes even to −35°C or so, where homogeneous freezing (that in the absence of nuclei) rapidly begins [Rosenfeld and Woodley, 2000]. Since saturation vapor pressures over liquid are higher than those over ice at temperatures below zero, water vapor amounts can be supersaturated with respect to ice by up to 60% without exceeding saturation over a liquid. This does not explain supersaturated vapor at temperatures below −40°C, where supersaturations of 10%–30% remain common and sometimes approach 60% or possibly higher. The reason for this is not clear, but one important factor may be the inhibition of vapor deposition by contaminants [e.g., Gao et al., 2004].

Note that phase equilibrium occurs at the conventionally defined saturation vapor pressure governed by (1) only when the condensate has a planar free surface and is made of pure, standard \( ^{1}H_2^{16}O \). Curved surfaces (such as cloud droplets) will evaporate more easily and thus require an equilibrium vapor pressure slightly higher than \( e^* \). Conversely, water contaminated either with solutes or surfactants will evaporate more slowly, thus reaching equilibrium at a lower vapor pressure (sometimes much lower in the case of hazes). In circumstances of significant rates of net phase change (precipitation or evaporation), these effects do not exceed the 1%–2% level and are thus negligible in affecting background humidity, although crucial in determining cloud properties.

Saturation vapor pressure is also dependent on the molecular mass of water molecules: heavier isotopologues \( (^{1}H_2^{18}O, HDO, \text{ and } ^{1}H_2^{17}O) \) have lower volatility because of differences in zero-point energy of vibration. This “vapor pressure isotope effect” results in isotopic fractionation during evaporation (or condensation), with heavier species becoming progressively more concentrated in an evaporating liquid and correspondingly underrepresented in the resulting vapor. Atmospheric water vapor shows particularly strong isotopic effects because (1) the degree of distillation is so high, with water vapor varying by 4 orders of magnitude; (2) vapor pressure effects are stronger at cold temperatures, found in the upper troposphere and high latitudes; and (3) the molecular zero-point energy is determined by its reduced mass and scales with the mass of the substituted atom, a factor of 2 difference in the case of deuterium. Since isotopically heavy molecules make up well below 1% of total water, the above effects are again unimportant for overall humidity or its radiative effects. Observations of isotopic ratios of vapor or precipitation can, however, provide useful diagnostics for physical processes such as evaporation of hydrometeors, mixing of air masses with different condensation histories, identification of water sources, and seasonality of precipitation, especially when compared with model simulations that predict multiple water species on the basis of explicitly simulated multiphase water transport. Indeed, the isotopic composition of ancient precipitation in glacial ice is one of the principal paleothermometers used in determining climate history. Isotopic composition in cave
deposits, fossil plants, and other archives may also be a useful indicator of past hydrological variability, at least in principle. The existence of multiple, distinct isotopologues in water is particularly relevant here as it permits distinction between equilibrium and nonequilibrium effects; this has been used as a proxy of past relative humidity (see section 6).

The strong constraints on humidity imposed by equilibrium thermodynamics are evident in climatological data (Figure 2). Actual and saturation mixing ratios both fall roughly exponentially with height, by 4 orders of magnitude from the tropical boundary layer to the tropopause, in such a way that zonal means of the former remain within 1 order of magnitude of the latter (that is, relative humidity >10%) throughout the troposphere. While this provides a zeroth-order explanation of the atmospheric humidity field, zonal mean mixing ratios do fall to 20% of saturation or less, and relative humidity reaches a few percent or lower in more localized subtropical regions of the troposphere [Spencer and Braswell, 1997]. Thus, large regional variations remain that cannot be explained by the Clausius-Clapeyron equation alone. Many of the key radiative and meteorological influences of water vapor are actually controlled by relative rather than specific humidity, as will be seen sections 4 and 6, so understanding these nonequilibrium variations is crucial.

In understanding the interactions of water vapor and dynamics it is useful to consider quantities that are conserved under phase changes of water [see Emanuel et al., 1994]. Moist static energy (MSE, the sum of moist enthalpy and
gravitational potential energy) is conserved in the limit of negligible kinetic energy and is often useful for analyzing phenomena on longer time scales; its cousin equivalent potential temperature ($\theta_e$, equivalent to moist entropy) is conserved in the limit of negligible irreversible mixing and tends to be useful for shorter time scales. These quantities are traditionally defined so as to be conserved only for liquid-vapor phase changes but can also accommodate freezing. Figure 2 reveals that equivalent potential temperature is smoother than either water vapor field, as a result of the relative simplicity of its sources and sinks (discussed further in sections 4 and 5). Moist static energy (not shown) looks similar.

3. OVERVIEW OF OBSERVATIONAL CAPABILITIES

While a thorough description of water vapor observing techniques and problems is beyond the scope of this review, we introduce in this section the main observing systems that have been used to obtain observational results noted subsequently. We ignore instruments (primarily limb-observing satellites) used primarily to measure the middle or upper atmosphere. We also omit instruments for measuring isotopologues, which are developing rapidly and now include, in addition to traditional in situ or sample-and-return techniques, four satellite sensors and several commercially available ground-based analyzers. We omit these both because of space considerations and because process-relevant results from these instruments are few and (in our judgment) highly preliminary at this time, although rapid progress is possible. A few recent results will be noted in sections 4–6 to give examples of potential; please see the individual papers for more information.

All data sets noted here are affected by issues of calibration accuracy, improvements over time in the quantity and quality of data, and other changes in observation characteristics over time that limit the ability to observe long-term trends reliably (see Elliott [1995] and Kley and Russell [2000] for more detailed discussions of problems in most of the observing systems). Recently, there have been efforts to produce “homogenized” data sets (usually by adding piecewise-constant bias corrections) where it is hoped that long-term changes are more reliable, although this is difficult to verify.

3.1. Station and Shipboard Data

Humidity is routinely measured at thousands of land-based surface observing stations, hundreds of upper air stations worldwide, and a number of ships of opportunity for collecting maritime data. The surface humidity database is of similar scope to that of temperature but has been the subject of significantly less analysis. Surface observations use a variety of techniques, most commonly wet bulb thermometers or moisture-sensitive resistive or capacitive sensors. Errors can arise because of calibration problems, local exposure, or undocumented variations in ship deck height or sensor placement [Elliott and Gaffen, 1991]. Recently, a preliminary homogenization of the global surface data set since 1965 was attempted [Willett et al., 2008], but there are indications of remaining inhomogeneity in the maritime data.

Radiosondes supply the longest record of upper air humidity data, with significant coverage dating back to the International Geophysical Year (1958). Precision of these instruments in controlled conditions is typically a few percent [Elliott and Gaffen, 1991] and has improved somewhat in recent instrument models. These instruments are able to capture vertical structure much better than satellites. Accuracy problems with radiosonde humidity observations have been severe, however, with models of the most commonly used sonde showing dry biases of up to 20% in the middle troposphere even within the last decade [see Miloshevich et al., 2009]. Changes to the network make trend calculation problematic [Parker and Cox, 1995]. Earlier U.S. models did not report relative humidities below 20% at all [Wade, 1994], and older models of all kinds were not reliable below −40°C, which includes most of the upper troposphere. These problems, the lack of strong physical constraints on humidity (other than those described in section 2), and the rapid variation of the moisture field at spatial scales that are short compared to the distance between stations make homogenization a real challenge. Nonetheless, this has recently been attempted by McCarthy et al. [2009]. More accurate balloon-based instruments have been developed for observing the upper troposphere and stratosphere region and have been used in a few locations since the 1950s (see Rosenlof et al. [2001] for a review).

3.2. Satellite

Widely used satellite humidity data date back to the onset of the NOAA polar orbiting program in 1979, although measuring capability has increased greatly since then. The original High Resolution Infrared Sounder (HIRS) downward scanning instrument used several infrared wavelengths to estimate average relative humidity across broad vertical swaths of the troposphere and has proven invaluable for tracking global changes in upper tropospheric humidity [see Buehler et al., 2008], although radiances are strongly affected by upper level cloud cover leading to a dry bias when cloudy scenes are removed [Lanzante and Gahrs, 2000]. Geostationary satellites such as Meteosat carry water vapor imagers similar to HIRS [Schmetz et al., 2002] but often with only a single channel. Improved infrared sounders have been launched more recently, most notably the Atmospheric Infrared Sounder (AIRS) which contains hundreds of channels in the water vapor absorption bands and can resolve vertical layers of a few kilometers, although it is still affected by cloud cover and tends to underestimate wet and dry relative humidity extremes [Fetzer et al., 2008; M. D. Chou et al., 2009].

Microwave radiation is less affected by clouds and thus offers a useful alternative method of moisture sounding from space. The Special Sensor Microwave Imager (SSMI) has been operating since 1988, and the TOPEX Microwave Imager has been operating since 1992, but these imagers can only be used to estimate the column-integrated (total) water vapor (aka precipitable water) since a low-opacity wave-
length is used. Microwave sensors dedicated to humidity observations and observing at more absorptive wavelengths have been flown since the early 1990s, starting with the Special Sensor Microwave Temperature (SSMT/2) and continuing through to Humidity Sounder for Brazil and advanced microwave sounding unit (AMSU) nadir sounders, and are able to detect humidity averaged over several broad layers, especially in the upper troposphere, with significant interference only from thick clouds [Engelen and Stephens, 1999; Susskind et al., 2003]. Similar instruments that observe limb emission, the two Microwave Limb Sounders (MLSs, first flown in 1991), observe moisture above ~350 hPa [Waters et al., 2006; Froidevaux et al., 2006]. None of these sounders have proven as useful for trend analysis as the SSMI and HIRS because of short mission lives and design changes, but they are believed to provide relatively accurate water vapor climatologies and statistics. Techniques employing backscattered sunlight can also be used to sense total column water vapor in a manner somewhat similar to SSMI using Global Ozone Monitoring Experiment (GOME) starting in 1996 and Scanning Imaging Absorption Spectrometer for Atmospheric Chartography (SCIAMACHY) starting in 2002 [Mieruch et al., 2008].

[22] Note that the foregoing passive sounders (other than SSMI and solar techniques) exploit emission that is all or nearly all from water vapor. As such they effectively measure relative, rather than specific humidity: emission depends mainly on the temperature at a given integrated water path, which is determined by relative humidity over some vertical swamp [Simpson, 1927; Soden and Bretherton, 1993]. Retrieval of mixing ratio requires auxiliary temperature information. By contrast, the active sensors described below and in section 3.4 effectively measure water vapor density, which is closely related to mixing ratio. From these, relative humidity retrieval requires auxiliary information.

[23] Most recently, Global Positioning System (GPS) technology has been useful for observing humidity by estimating the traveltime delay of routine GPS signals, which is determined by water vapor amount and temperature integrated along the signal path. This can be used to estimate total column water vapor over suitably equipped surface stations [Wang and Zhang, 2008]. More recently, several satellites have been launched to make occultations through the atmosphere and obtain vertical profiles tomographically up to roughly the −20°C level [Hajj et al., 2004; Wickert et al., 2009]. The GPS technique has the important advantages of being an absolute measurement that does not need an independent calibration and of not being affected at all by clouds or other absorbers (although temperature must be known fairly well to achieve good accuracy). As yet, however, it does not provide either the sampling density or record length characteristic of other techniques, and occultations measure means over paths of 100 km or more which will smear out convective and mesoscale variations.

3.3. Aircraft

[24] Humidity is routinely recorded by research aircraft employed in field studies and operational uses (e.g., hurricane monitoring) using dew point hygrometers mounted on a wing or fuselage. These sensors do not work well when dew point falls below about −20°C−30°C, so a number of more specialized sensors have been developed for upper tropospheric and stratospheric conditions, including frost point hygrometers and lidar absorption and/or scattering instruments.

[25] Routine monitoring of the atmosphere by aircraft did not occur until the advent of the Measurements of Ozone and Water Vapour by Airbus In-Service Aircraft (MOZAIC) program in 1994 [Luo et al., 2007], which deployed a specially modified hygrometer on five Airbus aircraft based in Europe. This provided flight level data along a number of main flight corridors, taken in a consistent manner over many years albeit with less precision (5%–10%) than balloon sensors. The MOZAIC program is now transitioning into the In-Service Aircraft for a Global Observing System project, which is planned to include not only water vapor but cloud, aerosol, and other gas measurements (http://www.iagos.org). The United States and Canada also now have programs to obtain real-time temperature, humidity, pressure, and icing information: Tropospheric Airborne Meteorological Data Reports and Aircraft Meteorological Data Relay, respectively [see Daniels et al., 2006].

3.4. Radar and Lidar

[26] Active lidar and radar sensors are able to remotely sense water vapor profiles or low-level 2-D fields via Raman backscattering and differential absorption lidar (DIAL) or traveltime delay, respectively. Lidars have been deployed only occasionally in field programs [Eichinger et al., 1999; Whitteman et al., 2006; Wulfmeyer et al., 2006], but they have shown robustness in long-term monitoring sites [e.g., Turner et al., 2002; Ferrare et al., 2006]. Radar refractivity retrieval is a relatively new development in which the capacity to estimate variations in signal propagation velocity is added to standard radars [Fabry et al., 1997; Fabry, 2004]. Refractivity measurements are obtained from fixed targets of opportunity. This provides a low-level refractivity map of modest spatial resolution but can serve as a proxy for mesoscale humidity structures.

[27] The passive atmospheric emitted radiance interferometer is capable of retrieving continuous moisture and temperature profiles in clear skies [e.g., Smith et al., 1999; Felte et al., 2003]. This unattended instrument can be used for long-term monitoring applications that are useful to continuously monitor the convective instability of the environment [e.g., Felte and Mecikalski, 2002; Wagner et al., 2008].

4. CONVECTION-VAPOR INTERACTIONS

[28] A full treatment of convection and how it is influenced by its environment is clearly beyond the scope of this review. Here we review some of the large body of work on how water vapor affects convection.

[29] Some relationship between moisture and the formation of any kind of cloud is obvious, since relative humidity must attain 100% somewhere. Difficulties arise because of
the large difference in scales between individual clouds, numerical model grids, and synoptic motions. Clouds are generally found where environmental relative humidity is sufficiently high, long part of the basis for their parameterization in models [e.g., Sundqvist, 1978; Tompkins, 2002]. Measures of convective instability such as convective available potential energy (CAPE) are also sensitive to the water vapor mixing ratio below cloud base, especially in the tropics, where horizontal variations of latent heat often exceed those of enthalpy. To the extent that cumulus updrafts mix with surrounding air, their buoyancy must also become sensitive to humidity above the cloud base. Finally, downdrafts driven by evaporative cooling should be sensitive to relative humidity aloft [Randall, 1980; Emanuel, 1991]. Humidity at a particular location can be changed by horizontal advection, but above the surface it is also easily increased by lifting as evident from the mixing ratio distribution (Figure 2).

30] Much uncertainty remains in how to capture convection in models. For example, general circulation models (GCMs) have difficulty in accurately simulating the diurnal cycle of precipitation, particularly in the warm season over land [e.g., Lee et al., 2008]. This is likely due to deficiencies in the convective parameterization schemes, although “superparameterization”-based models where convection is explicitly calculated also show problems [Zhang et al., 2008], while improvement is evident in some recent parameterizations [Rio et al., 2009]. Many aspects of weather and climate are sensitive to arbitrary variations in model treatment of convection, for example, the structure and location of the Intertropical Convergence Zone [e.g., Bacmeister et al., 2006; Zhang et al., 2007]. Some recent work has shown improvements in prediction due to reanalysis with increased midlevel moisture [e.g., Betts et al., 2009]. Deep convective schemes used in GCMs tend to show less sensitivity of convective heating profiles to moisture above cloud base than suggested by observations and cloud-resolving models [Redelsperger et al., 2002; Derbyshire et al., 2004], which may help explain why simulated convection occurs too early in the day over land [e.g., Yang and Slingo, 2001]. The behavior of humidity during the evolution of convective events in many GCMs is also quite poor, with the lower troposphere often drying out in the model when it moistens in observations [Mapes et al., 2009]. Some preliminary mesoscale model comparisons with observations suggest that the convection-resolving models are producing an excess of moisture in the boundary layer, perhaps because the boundary layer parameterization schemes are mixing too much.

4.1. Convective Initiation and Growth

31] The low- to middle-level water vapor amount and distribution strongly influence the initiation and evolution of convective storms [e.g., Keil et al., 2008]. In particular, the low-level moisture distribution along boundary layer convergence zones is potentially a key to understanding convection initiation along boundaries [Weckwerth and Parsons, 2006]. Low-level boundaries have been observed prior to thunderstorm initiation by surface station wind measurements [e.g., Byers and Braham, 1948], radar fine lines [Wilson and Schreiber, 1986], and thin cumulus cloud lines [Purdom, 1982]. While convergence lines exist frequently, they do not always initiate storms nor do the storms occur uniformly along the convergence lines. Some of the factors controlling convection initiation along boundaries include boundary layer moisture, i.e., low-level moisture variations [e.g., Fabry, 2006]; vertical moisture variations and depth of moisture [e.g., Crook, 1996; Lee et al., 1991; Bluestein et al., 1990]; stability; convergence/updraft strength and depth [e.g., Lee et al., 1991; Ziegler et al., 2007]; horizontal wave patterns [e.g., Carbone, 1982; Murphey et al., 2006]; and the vertical wind shear and its balance with the boundary’s solenoidal circulation [e.g., Rotunno et al., 1988; Wilson et al., 1998]. These factors are believed to play an important role in convective development in nonorographic midlatitude regions but a smaller role in environments that are typically near the threshold for the onset of convection, such as the western Pacific warm pool [e.g., Raymond, 1995] and other tropical regions [e.g., LeMone et al., 1998; Lima and Wilson, 2008].

32] Environmental stability and convergence are the two primary factors influencing convection initiation by boundaries. Static stability is largely impacted by the moisture distribution such that more intense storms occur with greater low-level moisture amounts [e.g., Crook, 1996]. Convergence occurs because of boundaries with strong density differences from the environment, colliding boundaries [Kingsmill, 1995], and intersections between boundaries and horizontal convective rolls [e.g., Atkins et al., 1995; Dailey and Fovell, 1999; Fovell and Dailey, 2001].

33] Convergence lines act to locally deepen the low-level moisture fields [e.g., Bluestein et al., 1990; Ziegler and Rasmussen, 1998]. The cloud fields observed by high-resolution visible satellite imagery further illustrate deep boundary layer moisture [Hane et al., 1987; Mueller et al., 1993; Wilson and Mueller, 1993; May, 1999]. In association with deep low-level moisture is the vertical moisture gradient. Mesoscale modeling results have illustrated that slightly varying the low-level vertical gradients of moisture and temperature will change the strength of the simulated storm from producing no storm to strong convection [Crook, 1996; Lee et al., 1991]. Vertical gradient variations can be achieved with small horizontal changes in the moisture field.

34] Observations have shown that small-scale moisture variations are commonly large enough to be significant for convection [e.g., Weckwerth, 2000; Wakimoto and Murphey, 2008; Buban et al., 2007; Roberts et al., 2008]. Horizontal moisture variations may be caused by small-scale boundary layer organized structures, such as horizontal convective rolls and misocyclones (convective-scale vortices), and have been shown to influence the location of convective development [e.g., Weckwerth et al., 1996; Murphey et al., 2006]. In fact, radar refractivity observations have suggested that enhanced low-level moisture occurs prior to convection initiation [Weckwerth et al., 2005; Fabry, 2006].
The importance of moisture in understanding convection initiation and evolution is further illustrated in simulations showing that the assimilation of GPS integrated water vapor, surface moisture observations, and a few soundings increases the ability of the model to accurately represent the low-level moisture fields [MacDonald et al., 2002]. Assimilation of airborne water vapor DIAL water vapor profiles into a mesoscale model showed significant improvement in quantitative precipitation forecasting (QPF) skill [Wulfmeyer et al., 2006]. Furthermore, high-resolution soil moisture measurements and land surface processes were assimilated with improved QPF skill [Holt et al., 2006].

4.2. Middle to Low Tropospheric Dry Layers and Convection

Large-scale tropical convective systems are strongly influenced by moisture not only near the surface but also aloft. Evidence for this has been obtained by many studies, including Johnson and Lin [1997], who found a strong positive relationship between free tropospheric humidity and strong organized tropical convection in observations; such moisture strongly affects numerically simulated squall lines [e.g., Lucas et al., 2000]. Moisture above the boundary layer is significantly better than CAPE at predicting where convection appears over warm tropical oceans in observations [Sherwood, 1999a], and three-dimensional cloud-resolving simulations of convective–radiative quasi-equilibrium further showed that lower tropospheric moisture was most influential in affecting tropical convection [Tompkins, 2001]. Differences in typical relative humidity between tropical and midlatitude environments helps explain differences in severe storm evolution [Wissmeier and Goler, 2009]. Neelin et al. [2008] argue that convection–humidity relationships can be better understood by analogy to generalized criticality theories, which offer the possibility of elegant constraint on the mean and extremes of convective behavior.

Dry air layers are often found in the low to middle free troposphere (800–400 hPa) of the deep tropics in regions that are normally under the influence of deep convective activity. These have been observed over tropical oceans [Mapes and Zuidema, 1996; Parsons et al., 2000; Jensen and Del Genio, 2006; Cau et al., 2005; Zuidema et al., 2006] as well as over the African monsoon region [Roca et al., 2005]. These dry layers have been shown to originate from the midlatitude jets and are hence coined extratropical dry air “intrusions” [Yoneyama and Parsons, 1999]. Such intrusions have also been reported at higher levels [Wauugh and Polvani, 2000], although their effect on deep convection is more difficult to assess than for the low to middle tropospheric case. These dynamical considerations share a lot in common with the subtropical free troposphere humidity origins extensively presented in section 5. The dry air layers are observed to strongly suppress deep convection [Brown and Bretherton, 1997; Yoneyama, 2003].

Middle to low tropospheric dry layers are thought to inhibit the vertical development of cloud mainly through two pathways: radiation and entrainment. A dry layer will bring about anomalous longwave cooling at its base, which creates a thermal inversion that inhibits cloud vertical development [Mapes and Zuidema, 1996]. Pakula and Stephens [2009] further emphasized the importance of the vertical distribution of radiation in stabilizing the low levels and preventing deep convection occurrence. This distribution is sensitive to that of water vapor, and to details such as continuum effects, even in convectively active cases, supporting a strong role of radiation in explaining the variety of the tropical cloud top distribution [Johnson et al., 1999]. The role of dry air entrainment in reducing cloud buoyancy by evaporating water has been recognized at least since the work of Stommel [1947]; see Jensen and Del Genio [2006] for a recent examination in the global context. A number of studies have investigated the process using cloud-resolving models [e.g., Ridout, 2002; Takemi, 2007]. Redelsperger et al. [2002] detailed the entrainment process at play during a dry layer event during the Tropical Ocean–Global Atmosphere Coupled Ocean–Atmosphere Response Experiment and highlighted the lateral entrainment around the developing clouds as the leading mechanism of buoyancy decrease. The relative role of the two pathways for dry layers to inhibit convection is yet to be resolved, and there is evidence that neither of them may fully explain observed impacts of midtropospheric humidity on convective behavior [Sherwood et al., 2004].

One promising avenue for testing models of convection is through the use of isotopologues of water vapor. Convective models are able to reproduce basic features of the vertical profile of HDO and H$_2$O$^{18}$O and indicate that these are significantly affected by hydrometeor transports [Moyer et al., 1996; Bony et al., 2008], but observations are highly variable and sparse, severely limiting the validation of local process models. One interesting finding however is that the well-known empirical relation between rain rate and isotopic composition of rain water (the so-called “amount effect”) can be explained primarily through the high relative humidity that is normally associated with heavy rain [Risi et al., 2008]. This reduces the reevaporation and hence isotopic fractionation of raindrops and cloud water, an example of how understanding vapor–convection interactions can inform the interpretation of hydrological and paleoclimate data.

4.3. Moisture-Convection Feedbacks and Convective Organization

Convective motions also transport water vapor upward because of the strong vertical gradient, producing positive moisture anomalies aloft [see Zelinka and Hartmann, 2009]. There is growing evidence that the mutual interaction between midlevel relative humidity and convection leads to important feedbacks whereby spatial variations of moist convection cause spatial variations of free tropospheric moisture which, in turn, reinforce those of the moist convection. This aids wind shear and low-level humidity anomalies in helping to organize squall lines [Zipser, 1977; Barnes and Sieckman, 1984; Redelsperger and Lalex, 1988; Lalex and Moncrieff, 1989]. Takemi et al. [2004] found in cloud-resolving model (CRM) sensitivity simulations that squall organization was most sensitive to humidity at the 500 hPa level. Observations
of West African squall lines and dry intrusion occurrences [Roca et al., 2005] tend to confirm the importance of this level in organizing squall lines.

[42] The role of midlevel humidity is complex, and high humidity is not always beneficial for convective growth. For example, high humidity will inhibit the formation of convective downdrafts. Such downdrafts would interfere with the intensification of hurricanes, helping to explain the need for high humidity throughout the column for hurricane intensification [Emanuel, 1995; Zhu et al., 2001], but such downdrafts will produce density currents that in other settings could help trigger nearby convection [see also Wissmeier and Golder, 2009].

[43] Raymond and Torres [1998] suggested that “premoistening” of the free troposphere by congestus could cause the observed “superclustering” of tropical deep convection. Grabowski and Moncrieff [2004] used a cloud-resolving simulation of the entire tropical belt to show that intraseasonal variations, particularly a strong emergent organization of convection reminiscent of the Madden-Julian oscillation (MJO), nearly disappeared when the moisture-convection link was artificially suppressed. This work suggests a strong relationship between humidity, vertical motion, large-scale flow, and moist convection [Grabowski, 2003; Bony and Emanuel, 2005]. The finding is supported by the MJO-like behavior of more idealized models when midlevel moistening by congestus is given an explicit role in modulating deep convection [Khouider and Majda, 2006; Raymond and Fuchs, 2009; Kuang, 2008]. Other studies continue to posit a dominant role for surface fluxes [Sobel et al., 2008, 2009], and both factors probably interact to organize convection.

[44] Convection-vapor interactions may also have identifiable consequences for how convection and water vapor respond to externally forced changes. One convenient example is monsoon circulations, which can involve rapid onset of changes to winds, humidity, and rainfall and often feature interesting intraseasonal breaks or fluctuations (see Bhat [2006], Wu et al. [2009], Bock et al. [2008], and Higgins et al. [2006] for studies in Asia, Australia, Africa, and North America, respectively).

[45] Commonly used convection schemes are probably too insensitive to free tropospheric humidity [Derbyshire et al., 2004]. This may be why traditional GCMs do not simulate the MJO very well [Lin et al., 2006]. One problem may be that shallow, narrow convection entrains much more than deep convection, resulting in an inability of the shallow systems to evolve into deep convective systems and that assumed entrainment rates do not reflect this [Grabowski et al., 2006; Khairoutdinov and Randall, 2006].

4.4. Local Quasi-Equilibrium Ideas

[46] Any quantitative model for the interactions of water vapor, dynamics, and convection must consider the conservation of heat and water substance. If condensed water is removed sufficiently rapidly, then this reduces to conservation of MSE or equivalent potential temperature $\theta_e$ (see section 2). Both quantities are strongly stratified in the extratropics, so that overturning motions there always transport MSE horizontally from the ascending branch to the descending branch, as would happen in dry air. But in the deep tropics MSE reaches a minimum in the lower free troposphere (Figure 2). This means that the magnitude and direction of net horizontal heat transport depends on how deep the overturning circulation is, and can be quite small even with strong circulations due to competing transports of latent heat and dry static energy [Back and Bretherton, 2006].

[47] The tendency of deep motions to export MSE gained from local surface fluxes was exploited by Neelin and Held [1987] to explain the Hadley and Walker circulations in the tropics in a way that elegantly sidestepped dynamical problems. They introduced the concept of “gross moist stability” to denote the effective stratification of the atmosphere to moisture-coupled circulations, where temperature and humidity profiles are slaved to the dynamics by convection and latent heat release is implicit. Various models based on assumed local quasi-equilibrium have emerged to explain aspects of the tropical mean and transient circulation [Emanuel, 2007; Raymond et al., 2009], which must balance horizontal and vertical convergences of moist static energy. Such models entirely neglect the convective initiation problem discussed in section 4.1, in effect assuming that some convection is always present, and by making some closure assumption relating to relative humidity, they can link the flow field to the heating field and obtain steady and/or transient solutions. Clearly, the required quasi-equilibrium assumption will only hold on sufficiently long space and time scales [Brown and Bretherton, 1997]. These models have yielded interesting insights into tropical meteorology; there may be opportunities for carrying these insights over to the development of GCMs (where the most obvious current innovation is “superparameterization” or explicit representation of convection and clouds). A key question left unanswered by these models, to which we now turn, is what controls the relative humidity field itself.

5. Explaining Relative Humidity in the Free Troposphere

[48] As noted in section 1, the intertropical belt, including the deep tropics and the subtropics, is the region responsible for most of the water vapor feedback (see section 6) and where humidity departs most from saturation [Held and Soden, 2000]. Hence, we emphasize this region here, although the discussion applies also to the extratropics to some extent. We do not discuss transport of water vapor into the stratosphere, which was reviewed recently by Fueglistaler et al. [2009]. We also do not address the planetary boundary layer. Regional variations in boundary layer humidity are complicated and subject to many influences such as local soil moisture [e.g., Guerova et al., 2005]. Global variations should be strongly regulated by global energy budget constraints, as discussed briefly in section 6 [see also Schneider et al., 2010].
Figure 3. Schematic of the overturning circulation with emphasis on the mechanism controlling the humidity distribution in the subtropics.

[50] Figure 3 shows a schematic of the large-scale humidity transport in this region. The source of water vapor is surface evaporation, which reaches a maximum over the subtropical oceans in the trade wind regions. The boundary layer water vapor is then transported toward the “meteorological equator” or Intertropical Convergence Zone, where it is transported upward. The upward transport is focused inside convective clouds that occupy a small area at any given time [Riehl and Malkus, 1958] but occur in different places at different times so that the time-averaged flow is upward over broad areas of high climatological rainfall. Upper level detrainment of cloud condensate and saturated air mass provide moisture that is carried away by the large-scale circulation to other regions, including those of subtropical subsidence. In subsidence regions, air is gradually compressed and warmed such that its relative humidity rapidly decreases (a process we refer to as subsidence drying, though this term can also apply in the Eulerian sense to the product \( \frac{w dq}{dz} \), where \( w \) is the upward velocity, when \( w < 0 \)). Dry air originating from the extratropics also mixes with the tropical air there. Ultimately, tropospheric humidity is determined by advection of vapor (on all scales) and by evaporation of advected cloud particles (hydro-meteors, particularly ice crystals). The horizontal flow is considerably more complicated than shown in Figure 3, with unsteady zonal jets and eddies on a variety of length scales, but the result is still persistently dry and moist regions as idealized in Figure 3.

[30] The essential characteristics depicted here (upward transport in relatively narrow elements with broad downward motion through isentropic surfaces) also characterize to a considerable extent the extratropical troposphere. One feature unique to extratropical latitudes is that large sources of isentropic temperature variability dominate in wringing moisture out of the air [Pierrehumbert et al., 2006]. The global moisture circulation can nonetheless be viewed as a single overturning cell in moist potential temperature coordinates resembling a large Hadley cell [Pauluis et al., 2008]. Looked at in this way, air approaching the deep tropics near the surface gains MSE (or \( \theta_e \)), then rises with approximately constant MSE, then returns aloft with MSE gradually falling because of radiative heat loss. This roughly accounts for the distribution shown in Figure 2. The lower MSE can be realized either as lower temperature (at higher latitudes) or, with greater compression of the air, as low relative humidity (in the subtropics).

[51] The poleward transport of humidity by the circulation is critical to maintaining the observed climate because the latent heat transported represents roughly half the total heat transport to high latitudes. Associated with this transport is the arrival of dry air intrusions in the tropics (see section 4.2) and moist tropical air masses at higher latitudes. The bulk of extratropical precipitation originates from equatorward sources of vapor [Dirmeyer and Brubaker, 2007]. Much of the poleward advection occurs in narrowly focused “atmospheric rivers” [Zhu and Newell, 1998], which are responsible for heavy precipitation events in midlatitudes, notably along the west coast of the United States [e.g., Bao et al., 2006; Neiman et al., 2008].

[52] Early efforts to understand the tropospheric specific and relative humidity fields emphasized empirical correlations with factors such as the surface temperature or cloud distributions. The last decade or so of research on the humidity transport in the free troposphere has confirmed that to first order, the water vapor distribution can be quantitatively predicted by taking into account only the large-scale wind and temperature fields without relying upon information on the water condensate in clouds [Yang and Pierrehumbert, 1994; Sherwood, 1996b]. We now review this finding, its limitations, and its implications.

5.1. A Model for the Free Troposphere: Advection-Condensation Paradigm

5.1.1. Description

[53] The advection-condensation (AC) (sometimes called “large-scale control”) paradigm proposes the simplest perspective possible on what determines the water vapor content of the troposphere [Pierrehumbert et al., 2006]. Consider an
air parcel that has reached saturation. From water conservation, the parcel’s water vapor mixing ratio must remain conserved in the absence of sources or sinks because of further condensation, small-scale turbulent mixing with air of different humidity, or evaporation of liquid or ice water. As a result, the mixing ratio of a particular location in the free troposphere may be supposed to be equal to the lowest saturation mixing ratio that air mass has experienced since its departure from the boundary layer, the ultimate source region for moisture, if these other sources are sufficiently small outside the boundary layer.

[54] This simple idea has received a great deal of attention and has seen numerical implementations in both Lagrangian and Eulerian frameworks [Sherwood, 1996b; Salathé and Hartmann, 1997; Pierrehumbert, 1998; Pierrehumbert and Roca, 1998; Dessler and Sherwood, 2000; Galewsky et al., 2005; Hurley and Galewsky, 2010], which appear to have been successful in reproducing observed humidity fields (or those simulated by a complete atmospheric model from which winds were used). The ability of coarse (large-scale) winds to produce fine-scaled distributions of water vapor or other tracers is described quantitatively by the theory of chaotic advection [Pierrehumbert and Yang, 1993; Yang and Pierrehumbert, 1994]. Realistic implementations rely on meteorological analyses of space- and time-varying temperature and wind fields, which are used to compute either Lagrangian back trajectories, starting at a given level and position and computing a trajectory backward in time, or forward evolution of the moisture field by solving the advection equation from an initial condition. The back trajectory approach is largely inherited from stratospheric research [Sutton et al., 1994] and relies on a reverse domain filling approach to estimate moisture fields from the minimum temperatures experienced along trajectories.

[55] In none of these calculations was the role of evaporating hydrometeors explicitly accounted for in the air mass water vapor budget (although they will play an implicit role by altering the winds). In the Lagrangian framework, no diffusion processes are accounted for, and air masses do not exchange vapor. This really means that two processes are neglected: (1) mixing effects of convection (including nonprecipitating convection) or other motions on scales too small to be incorporated into the trajectories and (2) horizontal transport of condensed water. In effect, it is assumed that the net ascent (upward hollow arrow in Figure 1) of vapor in a convective region is sufficient to capture any net role of convective transports (small arrows) in transporting vapor and cloud particles.

[56] The relative humidity $R$ at some target point is simply

$$R_{\text{target}} = \frac{q_{\text{target}}}{q_{\text{target}}^*} = \frac{q}{e^*} = \frac{e_{\text{last}}}{e_{\text{last}}^*} \times \frac{p_{\text{target}}}{p_{\text{last}}},$$

where “last” stands for last saturation, $q$ is the vapor mixing ratio, $q^*$ is the vapor mixing ratio at saturation, $e^*$ is the partial pressure of vapor at saturation (a function of temperature only), and $p$ is the pressure. Hence, the relative humidity from this simplified perspective is nothing but the ratio of the equilibrium water vapor pressure at two temperatures multiplied by a pressure ratio. Figure 4 shows the dependency of the relative humidity as a function of these two temperatures over a range of low to middle tropospheric temperature conditions. At the very dry end of the spectrum, for $R \sim 5\%$, a target temperature of 260 K corresponds to a temperature of saturation of 225 K. A more humid parcel of $R = 10\%$ would correspond to a last saturation warmer by more than 6 K, corresponding to a lowering of the altitude of the last saturation by roughly 1 km. Conversely, warming the target temperature by 6 K, holding $R$ at 5%, would only yield a modest 3 K (0.5 km) warming (lowering). This computation quickly illustrates that nontrivial distributions of relative humidity can result from Clausius Clapeyron nonlinearity even with the simple AC assumptions.

5.1.2. Limitations

[57] Since the real atmosphere features vigorous convective water transport, many are surprised that AC works at all. The convection problem is avoided, however, if the net effect of precipitating (thus, heating) convection is to enforce a near-saturated humidity profile on the spatial scale relevant to AC. Many observational studies suggest that this is indeed the case [Sherwood, 1996b; Bretherton et al., 2004; Holloway and Neelin, 2009], and we have already seen that it is difficult for convection to grow into dry environments (section 4), making it difficult for air to ascend through potential temperature surfaces without being near saturation. This net behavior of deep convection should not be surprising above the freezing level, where large regions become filled with anvil cloud, but it may be unexpected at lower levels that can be strongly affected by unsaturated downdrafts that can reduce relative humidity from its values earlier during the storm evolution. Indeed, relative humidities in convective environments are not quite as high below the freezing level [Sherwood, 1996b] so AC may not work as well in the lower free troposphere.

[58] There are some practical and conceptual difficulties with the advection-condensation model. First, evaporation of cloud particles away from convection and/or cloud-scale mixing in unsaturated environments may not always be negligible. However, to the extent allowed by the other two limitations, these moisture sources can be diagnosed from the errors in the predicted humidity field (see section 5.1.3). Second, the last-saturation event is ill defined (see below). Finally, results depend on the analyzed large-scale winds, which are defined in part on the basis of an imposed truncation scale, and can be corrupted by errors when estimated from data. This last point is not a fundamental problem but limits validation (as do errors in the humidity fields). Investigators have coped with the second and third difficulties by varying details of the implementation.

[59] The last-saturation event is usually established on the basis of large-scale winds and temperatures. The degree of saturation accounting for cloud-scale processes would presumably tend to $<100\%$ when averaged over a typical storm or grid box size, except in very large anvil clouds. A cloud-resolving model would in principle be needed to constrain
the relative humidity actually attained by air in a convective environment undergoing net ascent [e.g., Tompkins and Emanuel, 2000; Pierrehumbert et al., 2006]. On the other hand, evaporation of condensed water as air leaves a convective system could produce an effective “source” relative humidity of more than 100%.

One way to deal with this is to assume a relative humidity other than 100% at the point of last saturation. For example, Sherwood [1996b] assumed a value of 80% throughout the troposphere motivated by radiosonde data (although these data actually suggested values of 90% or higher in the upper troposphere). Subsequent studies usually assumed 100%. Sensitivity studies for humidity in the tropical upper troposphere show that the impact of uncertainty in the last-saturation $R$ is not large and that 100% produces better agreement with data than 80% or 120% [Dessler and Sherwood, 2000]. A variety of model calculations also show weak sensitivity to relative humidity at the last-saturation event, essentially because differences in initial $R$ are reduced at the same rate as $R$ itself under continued subsidence [Sherwood and Meyer, 2006].

The computation of tracer transport from large-scale winds suffers from some methodological limitations. In the Eulerian framework, numerical diffusion smoothes out gradients and extremes in the water vapor field [Hourdin and Armengaud, 1999], although monthly mean fields can be well simulated with appropriate algorithms [Sherwood, 1996b]. In the Lagrangian case, the advective reconstruction does not include any diffusion and yields filamentary structures whose scale decreases monotonically with time of integration, eventually requiring some form of numerical dissipation. The importance of this has received some attention for the stratosphere [e.g., Legras et al., 2003] but not (to our knowledge) for the troposphere. Common strategies that ameliorate this problem are averaging an ensemble of trajectories or examining time-averaged water vapor amounts.

Since temperature varies most strongly in the vertical direction, last-saturation temperatures rely heavily on the fidelity of the vertical component of the wind fields, which is essentially unobserved and in reanalysis is likely to have errors which can spuriously hydrate dry regions or dehydrate moist ones [Sherwood, 1996b; Pierrehumbert et al., 2006]. Reconstructions can also be sensitive to the time sampling rate or dynamical inconsistencies in the reanalysis fields [Stohl et al., 2004; Legras et al., 2005]. To avoid problems with noise in the reanalysis fields, one can use fields estimated from mass or energy balance [Pierrehumbert, 1998; Sardeshmukh, 1993] or inferred via regression analysis from a better observed variable with which it is well correlated [Sherwood, 1996b], which often brings small improvements with respect to the use of reanalysis vertical winds or a fixed cooling rate [Pierrehumbert, 1998]. Recent stratospheric reconstructions are successfully using vertical velocities calculated from the local heating rate and the vertical profile of potential temperature [James et al., 2008] but this is less satisfactory for the troposphere because of the large contribution of convective heat fluxes to the energy

Figure 4. The relative humidity as a function of the temperatures of last saturation and target points. A standard tropical profile is used to perform the computation. Saturation is assumed for ice for temperatures below 0°C and for water otherwise using the Goff-Gratch formula.
budget and the greater uncertainty in local radiative heating rates.

5.1.3. Tests Against Observations and the Debated Role of Condensed Water Transport

Sherwood [1996b] found that relative humidity from 700 to 300 hPa computed from AC agreed with SSMT/2 and radiosonde water vapor observations over tropical oceans to within 10%. Using MLS observations as a reference, Dessler and Sherwood [2000] showed that the monthly averaged reconstruction in the upper troposphere is in agreement with the satellite observations within a few percent both locally and in zonal means. Their implementation tends to slightly overpredict the dry regimes and underpredict the moist ones. Pierrehumbert and Roca [1998], using Meteosat observations, showed a very good agreement of the simulations averaged over a wide vertical region of the midtroposphere (see Figure 5), especially over the driest end of regime for which the differences between the observed and simulated 6.3 μm brightness temperature were <1 K (roughly 10% of relative difference in terms of upper tropospheric relative humidity [Vandeberg et al., 1995]). The instantaneous 500 hPa retrieval from this model was further compared to radiosonde observations over West Africa in dry conditions, and the model was shown to agree with <10% difference for R below 20% (better than the National Centers for Environmental Prediction [NCEP]) but to suffer from a systematic dry bias in the moister conditions [Roca et al., 2005]. In the extratropics, this type of model also performs well compared to AMSU-B satellite observations, in clear as well as cloudy regions, with a general bias of <2 K [Brogniez and Pierrehumbert, 2006]. The implementation of Dessler and Sherwood [2000], but for slight details, successfully reproduces more recent AIRS observations as well [Dessler and Minschwaner, 2007]. Annual distributions at 346 and 547 hPa compared well in the midtroposphere, although better agreement was found for the 346 hPa level.

In general, the above tests are encouraging and reveal first-order accuracy, in that reconstruction discrepancies are within observational error. Unfortunately, nadir sounders and many radiosondes often perform poorly at low relative humidities [Elliot and Gaffen, 1991; Fetzer et al., 2008; M. D. Chou et al., 2009]. Thus, strong statements about the adequacy of the advection-condensation paradigm in describing the real atmosphere cannot yet be made without further tests. Tests in a general circulation model do confirm that the water vapor field calculated by the model GCM can be reasonably well reconstructed using last-saturation-type calculation [Galewsky et al., 2005]. Tests have also shown water vapor to evolve along trajectories in an approximately realistic way in a GCM, even though the trajectories themselves (winds) may depart from those in the real atmosphere [Salathé and Hartmann, 2000]. This result could be model-specific but supports suggestions that difficulties in GCM water transport stem from inadequate resolution of the wind field [Bauer and Del Genio, 2006] and do not necessarily indicate poor physics.

Wright et al. [2009b] show that eliminating all condensate reevaporation in a GCM, but keeping the temperature, wind fields, and convective tendencies the same, caused specific humidity to fall by as much as 30%, with the largest drops in the tropical upper troposphere. The near-uniformity (in the horizontal direction) of percentage reduction in specific humidity reported in this study is consistent with what would be expected based on AC thinking, if water substance remained conserved outside storms but relative humidities in the ascending regions were no longer held near saturation by convective processes. We conclude that hydrometeor transports in the convective region are important in maintaining this near-saturated condition.

Some studies have tried to work backward from errors in AC calculations relative to observations to estimate source or sink strengths due to neglected convective moisture and condensate transports. Sherwood [1996a] found an apparent source of water vapor near 500 hPa associated with cumulus congestus in subsidence regions and evidence of a sink at higher levels. Several studies have compared the evolution of upper tropospheric humidity in air masses depending on their cloud content [Soden, 1998; McCormack et al., 2000]. The most recent and perhaps sophisticated example is Wright et al. [2009a], who analyzed trajectories through convective regions to diagnose sources of moisture from before-after differences in observed humidity (using the AIRS instrument). Like the previous studies, they report that trajectories transiting cloudy regions emerge moister than expected. However, their trajectories were calculated diabatically from clear-sky heating rates, which means they lacked the large-scale ascent in cloudy regions (hollow arrow in Figure 1) that is responsible for bringing air parcels to saturation in the advection-condensation paradigm. A similar omission applies to the earlier studies. It is therefore not surprising that trajectories through clouds emerge too dry, and the results cannot be used to evaluate the advection-condensation model without further analysis. Other studies have pointed out that the apparent moistening of air such as these studies have observed can be attributed to mesoscale lifting of vapor rather than sublimation of cloud ice [Sherwood, 1999b; Soden, 2004; Luo and Rossow, 2004].

Isotopes of water vapor hold potential for constraining models of water vapor, but modeling studies [e.g., Wright et al., 2009a] are few, and observations are limited. Recent measurements of HDO amounts measured in dry, free tropospheric air on Mauna Kea were inconsistent with direct transport from the surface but roughly consistent with an AC calculation [Galewsky et al., 2007]. Global satellite observations of tropospheric water isotopic ratios can, to date, resolve only a very thick layer of the lower to middle troposphere, limiting their utility, though one recent study has suggested some departures from simple AC calculations [Brown et al., 2008]. Work on water isotopes is rapidly expanding, and although much more is needed before robust conclusions can be drawn, these may become an important diagnostic in the future.
Figure 5. (top) Meteosat observed 6.3 μm water vapor equivalent blackbody brightness temperature (WVEBBT) (directly related to relative humidity when clouds are absent) on 15 July 1993. (bottom) Brightness temperature simulated from an advection-condensation calculation (reprinted from Pierrehumbert and Roca [1998]). Brightness temperature in blue regions is affected by clouds, which are not included in the calculation shown in Figure 5 (bottom).
5.2. Simpler Models of Humidity Statistics

[68] The advection-condensation paradigm, while evidently successful, is not a complete theory for predicting free tropospheric water vapor. The general circulation and temperature field or at least key aspects of them must also be predicted. Here we review a few attempts to do this in a simple framework. These are useful in explaining GCM behavior and demonstrate that the paradigm does not fundamentally rest on artificial truncations or gridding of the wind field.

[69] Several investigators have tried to simulate vertical profiles of relative humidity in 1-D radiative-convective equilibrium models. Such simulations have typically shown significant sensitivity of the result to either the choice of convective scheme, vertical resolution, or values of parameters in the scheme [Rennó et al., 1994; Tompkins and Emanuel, 2000; Sun and Lindzen, 1993]. However, the central role of the large-scale flow field in governing relative humidity casts some doubt on the applicability of a 1-D model. This was investigated by Sherwood and Meyer [2006], who showed that a very simple 2-D, steady state model of a diabatically balanced overturning would reproduce both the observed humidity distribution and its sensitivity to a key rain reevaporation parameter as exhibited by a full GCM. Both models became sensitive to this parameter when run in modes emulating a 1-D model but not in more realistic calculations with an overturning circulation and large horizontal gradients of humidity. This suggests that the sensitivity shown by 1-D models is mostly artificial, a conclusion also reached by Ingram [2002] concerning vertical resolution. The lack of horizontal structure in their humidity fields forces rain to fall through highly unsaturated environments, which observations indicate is not typical [e.g., Bretherton et al., 2004].

[70] Folkins et al. [2002] showed that the humidity profile in the tropics could nonetheless be reproduced well above 11 km by a simple 1-D model with the simplest possible treatment of convection. The net outflow from convective systems was constrained globally by the divergence of the clear-sky atmospheric radiative cooling profile, as recognized by many earlier investigators [e.g., Sarachik, 1978]. Folkins et al. simply calculated the humidity that would result from this outflow with no further sources or sinks, which is a relative humidity near saturation below the tropopause and falling quickly below that.

[71] This model fails below 10 km or so, mainly because the humidity becomes dominated by lateral mixing of air out of convective regions by winds that are not part of the net diabatic (overturning) circulation; prior to leaving the convective regions this air has been moistened by falling precipitation or mixing from below by congestus or weaker deep convection [Betts, 1990; Folkins and Martin, 2005]. In the absence of such sources, mixing ratios throughout the subsidence regions would fall to unrealistically dry levels, close to those in the uppermost troposphere as shown explicitly by Pierrehumbert and Roca [1998]. A more realistic picture with vertically uniform lateral mixing successfully reproduces the observed relative humidity profile throughout the free troposphere, which reaches a minimum near 400 hPa and increases gently below that; this profile arises from the vertical variation of the subsidence drying rate [Iwasa et al., 2004; Sherwood and Meyer, 2006]. This picture is still incomplete, as the tropical region is not a closed system but is strongly influenced by mixing to higher latitudes (see section 5.3). No simple model has yet incorporated all of these factors.

[72] The non-Gaussian probability (frequency) distribution (probability density function (pdf)) of tropospheric humidity [John et al., 2006] and the nonlinear dependence of radiative forcing and cloud formation on humidity call for investigations beyond mean values that more fully consider the pdf’s of humidity. Many, though not all, studies have argued that humidity pdf’s are bimodal [see Zhang et al., 2003]. Mapes [2001] argued that this was fundamentally due to the fact that the scale height of mixing is short compared to the vertical length scale of radiative kernels of water; he introduced the concept of a subsidence drying time of the order of days, which is short compared to horizontal advection times (of the order of 1 week) producing very low relative humidities. Sherwood et al. [2006], however, reported an extremely simple power law for the relative humidity pdf in free tropospheric GPS and MLS observations spanning either the tropics or midlatitudes which did not imply bimodality, was similar in both latitude bands, and can be characterized by a single scaling parameter $r$. They also introduced a very simple implementation of the AC paradigm by coupling a uniform subsidence rate to a Poisson statistical representation of the recycling of air through “last-saturation” events, which reproduced the observed scaling behavior. The scaling parameter $r$ was thereby identified as the ratio of the aforementioned subsidence drying time to a horizontal advection time. This calculation suffers from the common weakness (for tropical models) of ignoring interactions with midlatitude dynamics, and it is likely that such a simple result could be produced by other plausible calculations.

[73] Ryoo et al. [2009] recently extended this model to a two-parameter version, which they applied to regional distributions in addition to the global or intertropical one approximating the closed system for which the original calculation would be applicable. The remoistening events were modeled by a Gamma function, defined with an additional $k$ parameter, which is a measure of the randomness of the remoistening events. The larger $k$, the less random the process (up to periodic behavior in the limit of large $k$). The pdf of the relative humidity $R$ is given by

$$P(R) = \frac{k^k R^{k-1}}{G(k)} (\log R)^{k-1},$$

(3)

where $G$ is the Gamma function. In the limit of $k = 1$, the generalized model collapses to that of Sherwood et al. [2006]. The $r$ and $k$ parameters were obtained by fits to pdf’s of AIRS data, yielding impressive fits to the data (Figure 6). At the seasonal scale, over the whole intertropical
Figure 6. Probability density functions for whole tropics (0°–360°E, 30°S–30°N) for AIRS data at 250 hPa. Different colors are for different seasons, and dotted curve is fit to DJF data for generalized model. From Ryoo et al. [2009]. MAM, March, April, and May; SON, September, October, and November.

area or the regional domain (subtropical and tropical), the fit in the UT was nearly perfect. While the latter expression fits the AIRS observations much better than the original one-parameter model did, both show common behavior [Ryoo et al., 2009], and instrument biases are likely to affect the observed pdf’s (see section 3). In both cases, the pdf’s are more realistic in the subtropical UT (200–300 hPa) and in the lower troposphere (850 hPa) than in the midtroposphere around 500–600 hPa. In the UT, over the deep tropics, large $r$ and small $k$ are found, suggesting rapid and random moistening there. Over the dryer subtropical regions, small $r$ and large $k$ suggest slower and less random moistening consistent with a large-scale mixing process modulated by eddies of a preferred time scale. The vertical distribution of the parameters reveals that in the midtroposphere, a minimum of $r$ characterizes the nonconvective columns, $r$ being maximum in the lower troposphere and all through convectively active columns. On the other hand, $k$ is maximum, around 8, in the midtroposphere nonconvective columns and reaches a minimum in the midtroposphere of convectively active regions. Other studies have noted similar variations in the spatial correlation structure of humidity anomalies observed in different regions by aircraft and by AIRS, which may constrain the regional importance of processes on different length scales [Cho et al., 2000; Kahn and Teixeira, 2009].

5.3. Why Are the Subtropics Dry?

The subtropics play a large role in emitting radiation to space because of the relative lack of high clouds and are a region where changes in relative humidity have a large radiative influence [Pierrehumbert, 1995; Held and Soden, 2000]. While low, mixing ratios are not nearly as low as is typical in deep convective outflow regions, indicating other moisture sources. There has been some debate over what the dominant sources are and what determines the resulting dryness.

When cloud microphysics is neglected, the mechanisms that control the water vapor mixing ratio distribution in the subtropical midtroposphere are threefold: subsidence, which brings down dry air from aloft; lateral mixing that brings in moistened air from the deep convective regions at various rates; and net drying by processing of air through the cold extratropics [Pierrehumbert, 1998] (see also Figure 3). In the lower troposphere, some regions also experience direct moistening by shallow cumulus convection [e.g., Sherwood, 1996a]. Suggestively, the $R$ minima coincide closely with the regions of maximum curvature of the isentropic surfaces (see Figure 2), that is, where the temperature is warmest compared to nearby locations on the same isentrope. Galewsky et al. [2005] used a technique based on the transport of a large number of tracers, each of which represents air which has last been saturated in a particular region of the atmosphere, to determine where subtropical air was last saturated during boreal winter. They found that most air arriving near the driest location had last been dehydrated at midlatitudes rather than the upper tropical troposphere, suggesting a dominant role for quasi-isentropic midlatitude eddies in keeping the subtropics dry.

The rate of midlatitude air intrusions varies seasonally and longitudinally. Bates et al. [2001] investigated the variability of the upper tropospheric relative humidity (UTH) field and the role of transient eddy activity at 200 hPa by focusing on the Pacific Ocean. They showed that in boreal spring and winter, midlatitude eddies or Rossby waves can propagate into the subtropics, perturbing the large-scale circulation and hence moistening the otherwise dry UT. The opening/closing of a westerly wave duct, which allows this extratropical mechanism to modify the UTH, is shown to vary from El Niño to La Niña atmospheric state consistently with the satellite UTH record variability. At these altitudes, around 215 hPa, the breaking of Rossby waves at the subtropical tropopause has further been emphasized as an important dynamical process at play in the lateral mixing during the boreal winter [Waugh, 2005]. Using a back trajectory approach, satellite data, and potential vorticity diagnostics, this study suggests that the breaking of Rossby waves results in lowermost stratospheric air intrusions in the subtropical UT that brings in dry air, although less dry than the ambient air, yielding a net increase of relative humidity. Cau et al. [2007] emphasized the link with the extratropics at lower levels and further identified, during winter, four transport mechanisms explaining the relationship between the dry tropospheric air and the geographical origin of air masses: descending air masses associated with extratropical baroclinic systems, the subtropical anticyclone and its variability [e.g., Ryoo et al., 2008], the jet exits and entrances associated with convergence, and stagnation in the upper troposphere. These transport processes operate differently in different regions.

Other studies have found a lesser role for extratropical transports in drying the subtropics. Dessler and Minschwaner [2007] concluded that most air in the sub-
tropics was dehydrated in the tropical upper troposphere. They and Galewsky et al. [2005] both found that the importance of extratropical dry air increased with lower altitudes and on the poleward side of the minimum, as expected from the sloping of isentropes (see Figure 2), which indicates that the tropical upper troposphere is filled with air from below and relatively isolated from higher latitudes. Schneider et al. [2006] found in reanalysis data that the convergence of isentropic moisture fluxes was small in the subtropics and that subsidence drying was instead balanced by cross-isentropic eddy fluxes of moisture, which they interpreted as occurring in local convective clouds. While this is undoubtedly true, it does not contradict the AC paradigm, which predicts only that the large-scale eddy motions alone are enough to perform the convective transport (in fact, the Walker circulation would be among the eddies identified). Likewise, as pointed out by those authors, the small convergence of isentropic transport indicates that moistening from the deep tropics balances drying from higher latitudes but does not imply that either of these alone is unimportant. An interesting question is whether the approximate balances of terms are fortuitous or whether they are dictated by constraints on the system or, alternatively, which types of transport can most easily change the humidity.

An interesting question is whether the approximate balances of terms are fortuitous or whether they are dictated by constraints on the system or, alternatively, which types of transport can most easily change the humidity.

[76] A clue comes from the study of Brogniez et al. [2009], who analyzed the interannual variability of the midtropospheric relative humidity as seen by Meteosat. They focused on the driest tropospheric region found during boreal summer over the eastern Mediterranean Sea. Using an advection-condensation model and NCEP analysis, they showed that over the 20 year length of the satellite record, the variability of this very dry region is governed by the large-scale mixing of tropical and extratropical air masses, ruling out the role of the local temperature variability as a significant contributor to the interannual variability. This discounts cross-isentropic fluxes, which should bring the air toward local saturation, as a controlling driver of changes.

[77] The Galewsky et al. [2005] and Schneider et al. [2006] results can be reconciled by noting that the isentropic mixing ratio contrast between the deep tropics and subtropics, at ~60% relative humidity, is several times larger than that between subtropics and high latitudes, at ~10%–20% relative humidity, since the subtropical relative humidity is so low. Thus, even if most of the air was last saturated at high latitudes, air from the deep tropics could still supply most of the water vapor.

6. WATER VAPOR AND CLIMATE CHANGE

[80] How will water vapor change in a warmer (or colder) climate? This important question is motivated by the many ways humidity influences weather and climate but especially by two specific and attention-getting predictions. First, a strong positive feedback on global warming will result from water vapor unless relative humidity were to decrease rapidly as climate warms [Manabe and Wetherald, 1967; Held and Soden, 2000; Bony et al., 2006]. Second, greater water vapor mixing ratios in warmer climates are expected to cause commensurate increases in peak rain downpours and, indirectly, to slow down the diabatic overturning circulation and increase subtropical dryness [Bosilovich et al., 2005; Held and Soden, 2006]. While we shall focus in this section on these two impacts, it is worth noting in addition that water vapor dominates the infrared radiative cooling profile in the troposphere. It is primarily the increase in upper tropospheric humidity accompanying warming, rather than, say, CO2 increases, that is expected to shift the profile of tropospheric radiative cooling upward and therefore also the vertical extent of convective overturning in the warmer climate [Hartmann and Larson, 2002; Schneider et al., 2010]. Increasing tropopause height is an observed fingerprint of climate change [Santer et al., 2003] and a likely contributor to poleward migration of storm tracks [Lu et al., 2007].

[81] In this section, we first examine recently reported trends in three measures of water vapor: near-surface humidity, column-integrated specific humidity, and upper tropospheric humidity. We then summarize predictions of humidity and its impacts in warmer climates, noting a few more relevant observations.

6.1. Observed Water Vapor Trends

[82] Defining and measuring trends in atmospheric water vapor and hydrological variables is important as these will reflect forced changes so far in the Earth climate system due to the warming climate. In theory, it is expected that a positive trend in tropospheric temperature due to an increase in CO2 should have increased water vapor atmospheric content generally, which, in turn, should have further enhanced the atmospheric greenhouse effect, rainfall extremes, and the tropopause height change as well as having other impacts. In practice, Earth’s climate is buffeted by other factors such as El Niño–Southern Oscillation (ENSO), volcanic activity, solar variability, ozone depletion, and aerosol forcings, each of which may exert global or regional changes that confound any signal related to global temperature. There are also considerable problems in observing long-term trends of water vapor (and precipitation) from available data sets, exceeding the problems for surface temperature for example.

[83] Specific humidity near the surface is increasing according to analyses of the raw data, albeit somewhat more slowly than is predicted at constant R [Dai, 2006]. A new data set corrected for inhomogeneities shows increases globally consistent with constant relative humidity [Willett et al., 2008]. Panepaporation data have suggested decreases in the global hydrological cycle prior to the 1990s but appear to have been dominated by decreases in wind speed rather than changes in relative humidity [Roderick et al., 2007]; for reviews of trends in this and other measures of the hydrological cycle see Huntington [2006] and Roderick et al. [2009]. Over oceans, wind speeds appear to have been increasing fast enough to drive hydrological cycle increases on par with those of the Clausius-Clapeyron relation if changes in near-surface relative humidity and stability are neglected [Wentz et al., 2007]; ship observations since 1981 also show an unexpectedly strong upward trend [Yu and Weller, 2007]. While such increases in the hydrological cycle are at least
quadrapoles those expected on average from the observed warming, they are not necessarily outside the range of decadal variability arising either internally or from aerosol forcing variations [Liepert and Previdi, 2009].

[s8] Trends in total column water vapor have been reported in several studies (summarized in Table 1; rough comparisons of different time periods are reasonable, as underlying tropical surface temperatures have risen fairly steadily at ~0.1°C decade−1 since 1980). Starting from the 1980s, radiosondes [Ross and Elliott, 2001] and SSMI [Wentz and Schabel, 2000; Trenberth et al., 2005; Wentz et al., 2007] showed upward trends in the lower and free troposphere (below 10 km) on regional to global scales. Trenberth et al. [2005] reported a trend over the global oceans of 0.40 ± 0.09 mm (1.3% ± 0.3%) per decade from July 1987 to June 2004. Vonder Haar et al. [2008] examined trends in precipitable water (PW) over the years 1988–1999 in the NASA Water Vapor Project (NVAP) data set [Randel et al., 1996; Simpson et al., 2001], which was constructed by combining the observations from radiosondes, HIRS, and SSMI. They report a decrease of PW at a rate of ~0.29 mm decade−1. However, by subdividing the data into two halves, 1988–1993 and 1994–1999, trends with opposite signs were detected, which calls the significance into question. Regionally, upward trends were seen in the equatorial Pacific and drying in the subtropical regions, in particular in the Southern Hemisphere midlatitude oceans. Caution is necessary in looking at trends in any data product such as NVAP or reanalysis products blending observational platforms that cover different time periods because of their susceptibility to time-varying biases.

[s9] Durre et al. [2009] studied the PW trends from the radiosondes at ~300 stations in the Northern Hemisphere over the period 1973–2006 and found a statistically significant predominantly upward trend of 0.37 mm decade−1. Brown et al. [2007] reported a trend of 0.9 ± 0.06 mm decade−1 for 1992–2005 over oceans from the TOPEX Microwave Radiometer instrument. Regional trends were found to be highly correlated with regional sea surface temperature trends. Mieruch et al. [2008] examined GOME and SCIAMACHY data during 1992–2007, reporting an increasing global trend of 0.19% yr−1 (0.0039 ± 0.0015 g cm−2 yr−1) which was statistically significant when the 1997–1998 ENSO event was removed. They also reported significant regional trend variations: statistically significant increases in the water vapor column in Greenland, east Europe, Siberia, and Oceania and decreases in the northwest United States, Central America, Amazonia, central Africa, and the Arabian Peninsula.

[s6] Trends in free tropospheric relative humidity are clearly important in verifying the existence of a water vapor feedback. Trends from HIRS relative humidity were roughly zero from 1979 to 2003 despite tropospheric warming of roughly half a degree, clearly indicating an increase in specific humidity averaged over the middle and upper troposphere of roughly the magnitude predicted [Soden et al., 2005]. Relative humidity changes were not uniform, however, with increases in the deep tropics and high latitudes and decreases in the subtropics [Bates and Jackson, 2001]. Radiosonde data are of inconsistent quality but were subjected to a homogenization procedure by McCarthy et al. [2009], who report that resulting trends in relative humidity since 1958 from near the surface up to 400 hPa were not distinguishable from zero globally, although regional variations occurred.

[s7] A novel technique to assess trends is the use of laser-based sensors for measuring water vapor isotopologues. Coffey et al. [2006] reported airborne observations of column H216O, H218O, H217O, and HDO above 13 km (dominated by the uppermost troposphere) between 1978 and 2005. They found a long-term increase in Northern Hemisphere H218O of 1.16% ± 0.4% yr−1, roughly consistent with other observations of increasing humidity, but a decrease in HDO of 0.4% yr−1, which bears further investigation.

[s8] Paltridge et al. [2009] examined trends in reanalysis data from the NCEP over 1973–2007, reporting downward trends above 850 hPa in the tropics and southern midlatitudes and at altitudes above 600 hPa in the northern midlatitudes. However, this result had already been reported by Chen et al. [2008], who also noted nearly opposite results in the ERA-40 reanalysis. Numerous studies have concluded that reanalysis data are easily corrupted by time-varying biases and are therefore not useful for trend analysis [see U.S. Climate Change Science Program, 2006]. Each of the principal

### TABLE 1. Reported Trends in Total Column Water Vapor or Precipitable Water and Ocean Surface Hydrological Fluxes

<table>
<thead>
<tr>
<th>Author</th>
<th>Instrument</th>
<th>Region</th>
<th>Period</th>
<th>Change per Decade</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trenberth et al.</td>
<td>SSMI</td>
<td>global PW</td>
<td>1987–2004</td>
<td>0.40 ± 0.09 mm (1.3% ± 0.3%)</td>
</tr>
<tr>
<td>Vonder Haar et al.</td>
<td>NVAP reanalysis</td>
<td>global PW</td>
<td>1988–1999</td>
<td>−0.29 mm</td>
</tr>
<tr>
<td>Durre et al.</td>
<td>radiosondes</td>
<td>Northern Hemisphere PW</td>
<td>1973–2006</td>
<td>0.37 mm</td>
</tr>
<tr>
<td>Brown et al.</td>
<td>TMR</td>
<td>global PW</td>
<td>1992–2005</td>
<td>0.9 ± 0.06 mm</td>
</tr>
<tr>
<td>Mieruch et al.</td>
<td>GOME and SCIAMACHY</td>
<td>global PW</td>
<td>1996–2002</td>
<td>0.39 ± 0.15 mm (1.9% ± 0.7%)</td>
</tr>
<tr>
<td>Wentz et al.</td>
<td>SSMI</td>
<td>tropical PW</td>
<td>1987–2006</td>
<td>0.35 ± 0.11 mm (1.2% ± 0.4%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>tropical P</td>
<td></td>
<td>1.4% ± 0.5%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>tropical E</td>
<td></td>
<td>1.2% ± 0.4%</td>
</tr>
<tr>
<td>Liepert and Previdi</td>
<td>OAFluxb</td>
<td>global E</td>
<td>1987–2004</td>
<td>1.7% ± 0.9%</td>
</tr>
<tr>
<td></td>
<td>HOAPS3</td>
<td></td>
<td></td>
<td>4.7% ± 3.6%</td>
</tr>
</tbody>
</table>

PW, precipitable water; P, precipitation; E, evaporation; SSMI, Special Sensor Microwave Imager; NVAP, NASA Water Vapor Project; TMR, TOPEX Microwave Radiometer; GOME, Global Ozone Monitoring Experiment; SCIAMACHY, Scanning Imaging Absorption Spectrometer for Atmospheric Chartography.

*Objective Analyzed Air-Sea Heat Fluxes.

observational systems (HIRS and radiosondes) that went into these reanalyses shows consistently upward trends.

Thus, all primary data sets support the conclusion that water vapor mixing ratios in the troposphere are increasing at roughly the rate expected from the Clausius–Clapeyron equation. Although a few analyses have found otherwise, these relied on secondary data sets that are less suitable for quantifying trends. Nonetheless, the short length of the records, inhomogeneity issues, and spread of published trends indicates that we cannot yet establish from observed trends the precise climate dependence of water vapor.

6.2. Predictions for Warmer Climates

6.2.1. Water Vapor Feedback and the Upper Troposphere

GCMs have long predicted a strong positive infrared water vapor feedback that is roughly equal to that expected with climate-invariant relative humidity, a behavior anticipated ever since the original study of Arrhenius [1896]. The predicted feedback arises mainly from humidity at upper levels and particularly at low latitudes, where there is relatively little screening by clouds and where temperature and water vapor mixing ratio change the most for a given amount of surface warming [Held and Soden, 2000]. Models do predict regions of modest change in relative humidity [e.g., Wetherald and Manabe, 1980], but these are insufficient to affect the feedback much [Held and Soden, 2000]. A secondary radiative effect of water is caused by its absorption of near-infrared sunlight [Paltridge, 1973], which also contributes to the water vapor feedback [see Trenberth and Fasullo, 2009]. Note that water vapor feedback amplifies not only trends forced externally (e.g., by greenhouse gas increases) but also the seasonal cycle [Wu et al., 2008] and global variations on decadal and longer time scales [Hall and Manabe, 1999].

The advection-condensation paradigm would suggest that GCMs should be able to predict humidity changes in a warmer climate well if they can predict advection well. But whether they can do that is not clear, at least with coarser model resolutions [e.g., Bauer and Del Genio, 2006]. Bretherton et al. [2006] found in a cloud-resolving simulation of an idealized two-dimensional overturning cell that the ascending region narrowed with warming, causing a reduced mean relative humidity over the domain. It is not clear whether this is a robust response for this type of model or why this mechanism does not produce drying in GCMs, but it indicates that circulation changes that are significant from the perspective of relative humidity are not ruled out by any obvious constraints on the system.

Soden and Held [2006] reported that in the GCMs run for the IPCC 2007, water vapor continued to exert the largest feedback. The feedbacks from clouds in the various models were also found to be positive but smaller and much less uniform among the models. The only stabilizing (negative) global feedback came from changes in lapse rate since warming of the atmosphere exceeds that of the surface, which strengthens the climate’s natural Planck stabilization mechanism. The degree of lapse rate change varies somewhat among models, but the impact of this is approximately canceled out by the fact that greater warming aloft also brings a stronger water vapor greenhouse effect, making the two feedbacks highly anticorrelated among GCMs. The net of the two feedbacks is thus very consistent among models and of more modest, though still large, magnitude (not always larger than the cloud feedback). Its magnitude can be determined approximately by what happens to relative humidity as climate warms since changes in this are the way to decouple the water vapor and lapse rate changes (see also section 3.2). The feedback seems to be unaffected by biases in simulated humidity, which vary among models [John and Soden, 2007].

Quantifying GCM feedbacks accurately and consistently has proven tricky. Soden et al. [2008] proposed a method to simplify this based on “radiative kernels,” which describe the differential response of the top of atmosphere radiative fluxes to incremental changes in the feedback variables. This strategy enables one to decompose any feedback into one component that depends on the radiative transfer algorithm and the unperturbed climate state and a second component that arises from the climate response of the feedback variables. They found that the decomposition could not be evaluated directly for the cloud feedback because of strong nonlinearities but can be estimated indirectly from the difference between the full-sky and clear-sky kernels.

Two studies, Lindzen [1990] and Lindzen et al. [2001] (hereinafter referred to as LCH01), arguing against a strong water vapor feedback continue to be influential; LCH01 is worth examining here as a cautionary example. Both studies proposed that cloud microphysical or macrophysical effects will cause relative humidity in the upper troposphere to decrease significantly with warming. Doubt is cast on this by the evident success of the advection-condensation concept discussed in section 5, and several authors disputed various of LCH01’s assumptions [Hartmann and Michelsen, 2002; Lin et al., 2002; Del Genio et al., 2005]. But a perhaps more fundamental and broadly relevant problem, not noted in these critiques, is that the negative correlation between temperature and cloud observations invoked by LCH01 as evidence of a missing feedback was tested against an atmosphere-only GCM, which predicts only one of the two variables involved. In nature the sea surface temperature in the region of the central Pacific Ocean (30°S–30°N, 130°E–170°E) is sensitive to surface heat fluxes and warms under cloud-free conditions [McPhaden, 2002]. Figure 7 compares cloud fraction and near-surface temperature in observations and three coupled GCM simulations over the period 1985–2000. The chosen three models are skillful at reproducing mean radiative fluxes in the tropical latitudinal belt (M1 reproduces well the net, M2 reproduces the net incoming solar, and M3 reproduces the net outgoing longwave radiation [Andronova et al., 2009]). No model closely reproduces the observations, demonstrating the need for further model improvement, but two of the models (M2 and M3) do reproduce a negative correlation between temperature and cloud cover as seen in the observations. The third GCM produces a different relationship, yet all three models predict...
Figure 7. Relation between monthly mean near-surface temperature and cloud fraction, in observations and three coupled climate models (see section 6.2.1), in the region examined by Lindzen et al. [2001]. The observed relationship here is similar to that reported by LCH01, although LCH01 used daily data from 1 month with conditional sampling. The 1985–1998 cloud fraction observations are taken from the International Satellite Cloud Climatology Project data set (http://isccp.giss.nasa.gov/), and the temperature departure is based on the Hadley Centre–Climate Research Unit temperature data set version 3 data [Brohan et al., 2006].

[95] Many studies have examined natural spatial or interannual variability to try to test whether humidity behaves in a way consistent with the Clausius-Clapeyron relation and/or with GCM simulations. Much of the interannual variability in tropics-wide temperature is associated with ENSO. Large regional changes in relative and specific humidity accompany ENSO, including an overall warming of the atmosphere, but with no consistent change in tropical mean relative humidity at least as seen by HIRS [McCarthy and Toumi, 2004], although near the tropopause specific humidities do not increase fast enough to maintain constant relative humidity apparently because of the increased cloud outflow height [Minschwaner and Dessler, 2004]. Some studies have found that GCMs roughly reproduce the observed relationship between humidity and ENSO [Chen et al., 1996], though others show that the predicted regional humidity and cloud changes can be systematically wrong [Sun et al., 2009]. The latter study also finds that a GCM’s feedback on ENSO-like temperature anomalies is not a good predictor of its feedback on global warming, an important caution. Hurley and Galewsky [2010] showed using an advection-condensation calculation that ENSO-related humidity changes are due mainly to changes in where air is dehydrated (because of the altered circulation) rather than changes in the temperature field.

[96] More convincing demonstrations of expected behavior are observed responses to forced variations. These include the observed Clausius-Clapeyron responses to cooling caused by the Pinatubo eruption [Soden et al., 2002] and to the seasonal cycles of hemispheric mean temperature [Namdar and Ramanathan, 1998]. Neither of these is sufficient, however, to nail down the feedback very precisely.

[97] Recently, Dessler et al. [2008] used the radiative kernels from Soden et al. [2008] (see above) to estimate the changes in water vapor associated with interannual temperature variations in 5 years of AIRS observations. They found that (1) over most of the troposphere an increase in the specific humidity accompanied an increase in the globally averaged surface temperature and (2) the globally averaged relative humidity was nearly constant at most altitudes. Such AIRS variations are well reproduced by GCMs [Gettelman and Fu, 2008]. These results further support the expected water vapor feedback, although as with any result based on natural variability (see above), caution is required in extrapolating an empirical relationship to global warming.

6.2.2. Hydrological Changes and the Lower Troposphere

[98] Hydrological changes associated with water vapor are driven mainly by water mass rather than relative humidity and depend more on what happens in the lower troposphere. We focus here on two measures: near-surface humidity and total column water vapor (precipitable water).

[99] Globally, humidity near the surface is tightly constrained by the fact that it must (through its influence on surface evaporation) produce the correct evaporation rate to balance global rainfall, which, in turn, must balance the energy budget of the atmosphere, which is tightly restricted by radiation [Mitchell et al., 1987; Betts and Ridgway, 1989]. A too-high humidity would suppress evaporation until the humidity dropped to the required level. See Lambert and Allen [2009], Takahashi [2009], and Schneider et al. [2010] for further discussion of the relation of global precipitation to global temperature and caveats to the basic argument. Regardless of these, it is difficult to imagine large changes in relative humidity near the surface, although small increases with warming (of the order of 1% K⁻¹) may be expected on the above grounds and are predicted by most GCMs [Richter and Xie, 2008; Liepert and Previdi, 2009], and large decreases in near-surface wind speed could allow decreases in relative humidity [Betts and Ridgway, 1989].

[100] In light of the tight constraint above, the idea of effectively climate-invariant relative humidity in the planetary boundary layer has (so far) been uncontroversial among atmospheric scientists. Most of the column water vapor lies in the boundary layer, so this claim should largely apply to that too. Observed interannual variations [Wentz and Schabel, 2000; Mears et al., 2007] and trends (section 6.1) support this.

[101] Isotopes may offer glimpses into near-surface relative humidity in very different climates. The isotopologues HDO and H₂¹⁸O in ice cores have long been used to estimate Quaternary temperature variations, and it has recently become possible to measure the much less abundant H₂¹⁹O.
Depletion of $\text{H}_2\text{O}$ relative to that of $\text{H}_2\text{O}$ has been proposed as a proxy for near-surface relative humidity in past climates on the assumption that nonequilibrium fractionation during evaporation from the ocean is the source of observed departures from an equilibrium calculation [Landais et al., 2008]. If this is true, much higher relative humidity prevailed over the ocean at the last glacial maximum compared to today, contradicting predictions and recent trends. This is an intriguing suggestion, but many potential complications arise, and the interpretation of such data requires further investigation.

[102] The prediction that peak rain rates will increase with climatic warming is based on the fact that if wind convergence remains the same, more humidity will be imported into a storm [Trenberth, 1999]. This does not take into account possible impacts of higher moisture on the winds and has been challenged by O’Gorman and Schneider [2009], who find that rainfall extremes increase more slowly than column vapor in an idealized GCM because of damping of the baroclinic motions by the increased latent heat release per unit of upward displacement of air. Convective updrafts simulated by CRMs also become slightly weaker in warmer conditions when convective available potential energy and relative humidity are fixed [e.g., Robinson et al., 2008] and tend to decrease in GCMs as well [Held and Soden, 2006]. These effects, and the slowdown of planetary-scale circulations, may seem counterintuitive but can be qualitatively explained by noting that similar motions transport more energy in a more humid atmosphere: if energy transport is constrained not to change much, the motions should generally slow down in a more humid atmosphere. On the other hand, a number of studies have found that precipitation extremes are increasing, often faster than would be suggested by expected increases in column water vapor [Allan and Soden, 2008; Rajeevan et al., 2008; Lenderink and Van Meijgaard, 2008], although not all studies show this [Huntington, 2006]. Thus, while it still seems assured that peak rain rate increases with more column water vapor, quantifying this requires further study.

[103] Similar concerns apply to related predictions that climate zones will become more extreme. Global mean precipitation and evaporation increase at $\sim$2% K$^{-1}$ of global temperature [Allen and Ingram, 2002], but changes in precipitation have much finer spatial structure than those of evaporation. GCM studies [Bosilovich et al., 2005; Held and Soden, 2006; Waliser et al., 2007] show that in a warmer climate the increase in lower tropospheric water vapor produces an increase in horizontal moisture transport. Thus, models predict that with climate warming, the wet regions get wetter, and the dry regions get drier. Though a model consensus exists on this qualitatively, GCM-predicted amounts vary considerably. As with precipitation extremes, the basic physical reasoning begins by neglecting changes in winds [Held and Soden, 2006], yet higher water vapor mixing ratios will undoubtedly influence the circulation through their enhancement of latent heat released per unit cooling, leading to localized effects that can either increase or decrease the precipitation enhancement [C. Chou et al., 2009]. A review of the consequences of this as they are currently understood is given by Schneider et al. [2010]. More work is required to understand how humidity helps determine the general circulation and influences circulations on all scales; existing dynamical theories typically either ignore humidity and flow-dependent heating or employ very simple pseudoadiabatic assumptions.

7. CONCLUSIONS AND OUTSTANDING ISSUES

[104] In this review, we have argued for some significant advances in our ability to characterize water vapor but have also highlighted some continuing challenges and emerging opportunities. While isotopic studies were only lightly touched on in this review, we hope that the examples given will suggest the potential of such observations to help constrain water transport processes and humidity in different climates. Much more work is needed (and is underway) to use models to refine and test favored interpretations of isotopic observations. Although process-relevant conclusions are at present highly preliminary, we have no doubt that a comprehensive review of isotopic studies of atmospheric water will, before long, be warranted.

[105] A key achievement over the last decade or so is to show that the range of humidities observed in the free troposphere, notably the dryness of the subtropics, is quantifiable to first order by a simple advection-condensation theory, described in some detail in section 5. The essence of this concept is that water vapor behaves as a conservative tracer with respect to large-scale (averaged over $\sim$100 km or more) air motions except that it rains out when supersaturated. This model ignores cloud microphysical considerations except to the extent that microphysics indirectly affect the condensational (or radiative) heating and thus the flow field that performs the advection. According to this theory the net impact of convective-scale turbulence and water phase changes on humidity may be essentially captured via the impact of the heating by that convection on the large-scale vertical velocity field, without having to know anything else about these effects. This is a crucial advance as it links the net transport of moisture to those of energy and mass, which are constrained independently by conservation principles. It is therefore too pessimistic to declare that the upward transport of water vapor by convection is a complicated phenomenon that is disconnected from the large-scale behavior, although some such transport undoubtedly occurs. Despite technical limitations, simulations based on advection-condensation agree with a variety of observations to better than 10% accuracy over a wide range of humidity regimes. The simplest models, which include only radiation-driven uniform subsidence after last saturation and idealized large-scale moisture sources, capture the essence not only of the mean relative humidity field but of its full distribution. This shows that the model does not fundamentally depend on the scale truncations often employed in practical computations.

[106] The reason that the advection-condensation model seems to work for humidity is that source regions for free
tropospheric air can be usefully approximated as saturated with water vapor, and subsequent sources or sinks of vapor from condensation or mixing appear to become modest before the air warms much. A similar strategy would generally fail for other passive tracers (ozone, radon, etc.) because of chemical sources or sinks and/or insufficient constraint on initial values; it would fail for energetic quantities (temperature and moist static energy) because of radiative sinks.

Convective transports may not figure explicitly in the advection-condensation model but are clearly important in enforcing the high relative humidities observed in regions of precipitation (sections 4 and 5.1.3) necessary for this model to work. Mutual interactions between convection and moisture throughout the low to middle troposphere now appear to be important in organizing tropical convection. Relatively new technologies enable moisture fields to be characterized well enough to reveal the detailed interplay between low-level moisture and convective initiation and growth and to better forecast convection at short lead times. Much more use should be made of these observations to constrain the ideas underlying simple models/parameterizations of convection. New strategies may be needed to combine the diverse array of water vapor observations available in order to enable the needed understanding of coupling across scales.

The advection-condensation paradigm predicts that several important types of system perturbation would not cause any appreciable change in the free tropospheric relative humidity field if they were not accompanied by any change in the circulation: (1) a uniform warming of the atmosphere, (2) a change in aerosol or other properties affecting droplet growth or freezing in clouds, and (3) emissions of water vapor near the ground. However, the humidity distribution is sensitive to spatial variations in 3-D temperature fields and, potentially, to subtle changes in the circulation which could be caused by any of the above perturbations. These dynamical aspects should be the focus of future studies on humidity. For instance, Pierrehumbert et al. [2006] used a GCM with 4 times the amount of CO$_2$ ($4 \times$ CO$_2$) to compute vapor from air trajectories using either the $4 \times$ CO$_2$ winds and control temperatures and vice versa and revealed that the full pdf of relative humidity remains roughly the same as in the current climate. However, the temperature pattern change alone yielded drier air, while dynamical changes compensated by sampling cold regions less frequently.

The advection-condensation concept, by discounting the role of hydrometeor transports, supports the existence of the long-anticipated feedback by water vapor on global climate. This also continues to be supported by accumulating observational evidence, and we are unaware of any such evidence, capable of withstanding scrutiny, that contradicts it. Water vapor mixing ratios have increased in recent decades near the surface, through the lower troposphere (as quantified by total column water vapor), and in the upper troposphere, each measured by independent observing systems, although trends cannot be determined very precisely (section 6). While this supports the overall feedback, GCMs continue to show biases in means, variance, and correlations of moisture variables (see Figure 7), indicating the need for improvement before predictions generally should be trusted.

Opportunities exist to exploit and test the advection-condensation paradigm further to better constrain the behavior of water vapor in the free troposphere. An enhanced perspective could be gained on the statistics of the remoistening, for example, by exploiting the wealth of satellite cloud data, which has only been attempted in a few studies. Another area needing further research is how to bridge the understanding of the UT/LS and tropical tropopause regions [e.g., Fueglistaler et al., 2009] with the current understanding of the midtroposphere, where idealized studies seem not to provide a fully satisfactory result. For instance, the extension of the analysis of the role of extratropical intrusions resulting from Rossby wave breaking on a cooler isentropic level (∼330 K) would clarify their importance in the tropical midtroposphere. Similarly, incorporating the lateral eddy transport into simple models through idealized GCM simulations or other avenues is needed.

Our basic understanding of the future hydrological cycle changes predicted by the current generation of numerical climate models has improved significantly owing to analyses of the Coupled Model Intercomparison Project 3 archive of model outputs, helping to guide future research. However, many questions remain unanswered, particularly at the regional level, where the simple arguments that have been put forward do not appear sufficient and GCMs behave inconsistently.

Recent rainfall observations reveal some puzzles and challenges. Multiple lines of evidence indicate that severe rainfall amounts are increasing faster than expected, even though column water vapor on average is increasing roughly as expected (section 6). Much more work is needed to explore these extremes (and their climate sensitivity) both theoretically and observationally, particularly with respect to the interaction of water vapor and dynamics. For example, many extreme rainfall events in midlatitudes are associated with “atmospheric rivers” (see section 5), yet we found no studies examining the dynamical mechanisms responsible for focusing the transported tropical moisture in this fascinating manner. Regional trends and interannual variations in humidity also appear in some cases to depart from observations. While we have a working theory for humidity given the circulation, we lack a theory to link cloud microphysical and macrophysical behavior to condensation, heating, and precipitation in the atmosphere. And even if microphysical complications may be ignored, it is not clear how weather extremes (e.g., hourly or daily rainfall) are linked to climate extremes (e.g., the difference in mean rainfall between wet and dry climate zones), though both are expected to intensify in a warmer climate. Studies examining more closely the circumstances that lead to observed rainfall extremes, and attempting to model them quantitatively, appear warranted. We hope that these and the many other remaining mysteries will yield future studies combining the wealth of water vapor and cloud data and conceptual and computational tools that are now available.
ACKNOWLEDGMENTS. We thank William Ingram, Adam Sobel, and Elisabeth Moyer for comments on preliminary versions of the manuscript.

REFERENCES


Arhensius, S. (1896), On the influence of carbonic acid in the air upon the temperature of the ground, Philos. Mag., 41, 237–276.


Gao, R. S., et al. (2004), Evidence that nitric acid increases relative humidity in low latitude cirrus clouds, Science, 303, 516–520.


Jensen, E. J., et al. (2005), Ice supersaturations exceeding 100% at the cold tropical tropopause: Implications for cirrus formation and dehydration, Atmos. Chem. Phys., 5, 851–862.


Mapes, B., J. Bacmeister, M. Khairoutdinov, C. Hannay, and M. Zhao (2009), Virtual field campaigns on deep tropical convection in climate models, *J. Clim.*, 22, 244–257.


Neiman, P. J., F. M. Ralph, G. A. Wick, J. D. Lundquist, and M. D. Dettinger (2008), Meteorological characteristics and overland precipitation impacts of atmospheric rivers affecting the west coast of North America based on eight years of SSM/I satellite observations, *J. Hydrometeorol.*, 9, 22–47.


Pierrehumbert, R. T., H. Brogniez, and R. Roca (2006), On the relative humidity of the atmosphere, in *The Global Circula*-


N. G. Andronova, Department of Atmospheric, Oceanic and Space Sciences, University of Michigan, Ann Arbor, MI 48109, USA.

R. Rocca, Laboratoire de Meteorologie Dynamique, Institut Pierre et Simon Laplace, Place Jussieu, F-75252 Paris CEDEX 05, France.

S. C. Sherwood, Climate Change Research Centre, University of New South Wales, Level 4, Matthews Bldg., Sydney, NSW 2052, Australia. (ssherwood@alum.mit.edu)

T. M. Weckwerth, Earth Observing Laboratory, National Center for Atmospheric Research, Boulder, CO 80307, USA.