JOURNAL OF GEOPHYSICAL RESEARCH, VOL. ???, XXXX, DOI:10.1029/,

Multi-model Assessment of the Upper Troposphere and Lower Stratosphere: Tropics and Global Trends

A. Gettelman[1], M. I. Hegglin[2], S-W. Son[3], J. Kim[3], M. Fujiwara[4], T. Birner[5], S. Kremser[6], M. Rex[7], J. A. Añel[8], H. Akiyoshi[10], J. Austin[9], S. Bekki[17], P. Braesike[23], C. Brühl[15], N. Butchart[22], M. Chipperfield[21], M. Dameris[14], S. Dhomse[21], H. Garny [14], S. C. Hardiman[22], P. Jöckel[15], D. E. Kinnison[1], J. F. Lamarque[1], E. Mancini[20], M. Marchand[17], M. Michou[13], O. Morgenstern[18], S. Pawson [16], G. Pitari[20], D. Plummer[11], J. A. Pyle[23], E. Rozanov [19], J. Scinocca[12], T. G. Shepherd[2], K. Shibata [24], D. Smale[18], H. Teyssèdre[13], W. Tian[21]

[1] National Center for Atmospheric Research, Boulder, CO, USA
[2] Department of Physics, University of Toronto, Toronto, ON, Canada
[3] McGill University, Montreal, QE, Canada
[4] Hokkaido University, Sapporo, Japan
[5] Department of Atmospheric Science, Colorado State University, Fort Collins, Colorado, USA.
[6] Freie Universitaet Berlin, Berlin, Germany
[7] Alfred Wegener Institut, Potsdam, Germany
[8] EPhysLab, Universidade de Vigo, Ourense, Spain
[9] NOAA Geophyiscal Fluid Dynamics Laboratory, Princeton, NJ, USA
[10] National Institute for Environmental Studies, Tsukuba, Japan
[11] Environment Canada, Toronto, Ontario, Canada
[12] Canadian Center for Climate Modeling and Analysis, Victoria, BC, Canada
[13] GAME/CNRM, Météo-France, CNRS, Tolouse, France
[14] Institut für Physik der Atmosphäre, Deutsches Zentrum für Luft- und Raumfahrt, Oberpfaffenhofen, Germany
[15] Max Planck Institut für Chemie, Mainz, Germany
[16] NASA Goddard Space Flight Center, Greenbelt, MD
[17] LATMOS-Institut Pierre-Simone Laplace, UVSQ, UMPC, CNRS/INSU, France
[18] National Institute for Water and Atmosphere, Lauder, New Zealand
[19] Physikalisch-Meteorologisches Observatorium Davos, Davos, Switzerland
[20] Universita degli Studi de L'Aquila, L'Aquila, Italy
[21] University of Leeds, Leeds, UK
[22] United Kingdom Met Office, Exeter, UK
[23] Cambridge University, Cambridge, UK
[24] Meteorological Research Institute, Tsukuba, Japan

Submitted to JGR, 30 November 2009, Revised 19 April 2010

0148-0227/10/$9.00

Abstract. The performance of 18 coupled Chemistry Climate Models (CCMs) in the Tropical Tropopause Layer (TTL) is evaluated using qualitative and quantitative diagnostics. Trends in tropopause quantities in the tropics and the extra-tropical Upper Troposphere and Lower Stratosphere (UTLS) are analyzed. A quantitative grading methodology for evaluating CCMs is extended to include variability and used to develop four different grades for tropical tropopause temperature and pressure, water vapor and ozone. Four of the 18 models and the multi-model mean meet quantitative and qualitative standards for reproducing key processes in the TTL. Several diagnostics are performed on a subset of the models analyzing the Tropopause Inversion Layer (TIL), Lagrangian cold point and TTL transit time. Historical decreases in tropical tropopause pressure and decreases in water vapor are simulated, lending confidence to future projections. The models simulate continued decreases in tropopause pressure in the 21st century, along with $\sim$1K increases per century in cold point tropopause temperature and 0.5-1ppmv per century increases in water vapor above the tropical tropopause. TTL water vapor increases below the cold point. In two models, these trends are associated with 35% increases in TTL cloud fraction. These changes indicate significant perturbations to TTL processes, specifically to deep convective heating and humidity transport. Ozone in the extra-tropical lowermost stratosphere has significant and hemispheric asymmetric trends. O_3 is projected to increase by nearly 30% due to ozone recovery in the Southern Hemisphere (SH) and due to enhancements in the stratospheric circulation. These UTLS ozone trends may have significant effects in the TTL and the troposphere.

1. Introduction

The upper troposphere/lower stratosphere (UTLS) plays a key role in radiative forcing of the climate system and chemistry-climate coupling (see *Shepherd* [2007] for a recent review). The tropical tropopause layer (TTL) sets the boundary condition for air entering the stratosphere [*Brewer*, 1949]. Since the tropical tropopause is itself not a transport barrier, it has come to be thought of as a layer of finite depth. We here regard the TTL as being synonymous with the tropical UTLS for the purpose of model validation. The TTL is the region in the tropics within which air has characteristics of both the troposphere and the stratosphere. Representing the TTL region accurately in global models is critical for being able to simulate the future of the TTL and the effects of TTL processes on climate and chemistry.

The TTL is the layer in the tropics between the level of main convective outflow and the cold point tropopause (CPT), about 12–19km [*Gettelman and Forster*, 2002]. The TTL has also been defined by *Fueglistaler et al.* [2009] as a shallower layer between the level of zero clear sky radiative heating and the CPT (15–19km). We will use the deeper definition of the TTL here because we seek to understand not only the stratosphere, but the tropospheric processes that contribute to TTL structure (see below). The TTL is maintained by the interaction of convective transport, convectively generated waves, radiation, cloud microphysics and the large-scale stratospheric circulation. The TTL is the source region for most air entering the stratosphere, and therefore the TTL sets the chemical boundary conditions of the stratosphere. Clouds in the TTL, both thin cirrus clouds and convective anvils, have a significant impact on the radiation balance and hence tropospheric climate [*Corti et al.*, 2006].

In this study we present quantitative evaluations of coupled Chemistry Climate Models (CCMs) in the TTL. We also present key historical trends in the TTL for model evaluation, and key future projections in the TTL and the extra-tropical lowermost stratosphere (LMS) that may affect the TTL by rapid quasi-isentropic transport. This study builds on earlier work by *Gettelman and Birner* [2007], who analyzed 2 models and *Gettelman et al.* [2009], who analyzed trends for 11 CCMs. Here we extend these works by performing a more quantitative set of model diagnostics using 18 updated models and analyze trends for the future. These CCMs were run for the CCM Validation 2 (CCMVal-2) project experiments as input to the 2010 World Meteorological Organization (WMO)/United Nations Environment Programme (UNEP) assessment of stratospheric ozone depletion. A companion paper on the extra-tropical UTLS by *Hegglin et al.* [2010] also includes an assessment of model performance.

Section 2 describes the diagnostics and models, Section 3 describes comparison data sets. Section 4 presents results of historical runs, Section 5 presents results of trends and conclusions are presented in Section 6.

2. Models, Diagnostics and Grading

The TTL is the source of most stratospheric air, and water vapor in the stratosphere is regulated by tropopause temperatures [*Brewer*, 1949]. Hence the correct representation of the TTL critically depends on a correct representation of tropical tropopause temperature and water vapor. Diagnostics will also focus on variability in the TTL, for examining large scale and long-term variability in tropopause temperature. The different diagnostics are used to grade model skill. Quantitative grades are applied to some of the diagnostics. These quantitative diagnostics can be used as metrics of model performance.

2.1. Models and Experiments

The models and simulations used in this study are part of the CCM Validation round 2 (CCMVal-2) inter-comparison project. All of the models are coupled CCMs. A CCM is a General Circulation Model (GCM) of the atmosphere that includes prognostic chemical species that are used in the dynamics and thermodynamic equations of the model. Most importantly, chemically active ozone and water vapor are used in the GCM radiative heating equation. CCMVal-1 models have been documented by *Eyring et al.* [2006] and results reported in *World Meteorological Organization* [2007]. The performance of these models in the TTL has been examined by *Gettelman et al.* [2009]. Here we perform quantitative analyses on a new set of models. The list of models and basic references are presented in Table 1.

Further information on the attributes of each model is available in the references in Table 1, or in *Morgenstern et al.* [2010], a comprehensive description of the models. Salient features of the models are noted here. CMAM is coupled to an ocean model, while the other models use specified Sea-Surface Temperatures (from observations or another coupled model run for the future). Many of the models share a common heritage. E39CA, EMAC and (NIWA-) SOCOL are all based on the European Center Hamburg (ECHAM) GCM. UMETRAC, UMSLIMCAT and UMUKCA models are based on the Unified Model (UM). However, UMUKCA and EMAC are based on newer versions of their respective model. WACCM and CAM3.5 share the heritage of the NCAR Community Atmosphere Model version 3.5. All models have an inorganic chemistry scheme including chlorine and bromine (except for E39CA) chemistry. Only three models (CAM3.5, EMAC and ULAQ) have a comprehensive description of tropospheric chemistry. As indicated in

Table 1, most models have 6–9 layers in the UTLS, corresponding to a vertical resolution of about 1km. EMAC and E39CA have higher vertical resolution in this region (12 and 15 levels). ULAQ and SOCOL have lower vertical resolution (3–5 levels). For most models the horizontal resolution is ∼200–300km. ULAQ is significantly lower than this. The CCMVal-2 models include a larger set than CCMVal-1 (14 v. 11 models) and there are now 13 models with simulations to 2100 (v. 2 models in CCMVal-1). More importantly, there are 4 new models, and one discontinued. There are numerous changes to each model (see *Morgenstern et al.* [2010]), and these points are discussed as they are relevant for the results.

Model simulations analyzed comprise two types of runs, as specified by *Eyring et al.* [2008]. The first are 'historical runs' from 1960–2005, with specified boundary conditions for the sea surface temperature (SST), and specified concentrations of greenhouse gases and halogens, known as 'REF-B1'. Runs for the future from 1960–2100 are called 'REF-B2' and use emissions scenarios and SST fields as discussed in *Eyring et al.* [2008].

2.2. Quantitative Diagnostics

The list of diagnostics used in this study is shown in Table 2 and described in more detail below (and in each section). Diagnostics 1–4 have quantitative grades applied. Table 2 also indicates the data source(s) used for evaluation and grading. Some diagnostics (especially 6 and 7) required special outputs, often instantaneous output, and were not performed for all models. Monthly mean output is supplied on CCMVal-2 levels (see Figure 5).

Diagnostic 1: The Temperature of the Cold Point Tropopause (TCPT): It is critical that models reproduce the amplitude and phase of the annual cycle of TCPT as this regulates water vapor and total hydrogen in the stratosphere. Because of the non-linearity of the

Clausius-Clapeyron equation regulating water vapor saturation vapor mixing ratios, the annual cycle is more important than the mean value over the year. This is a simplified diagnostic of the true 'Lagrangian Cold Point' which we can examine in only a few models and which is not quantitative (see below). One measure of uncertainty is the grading of re-analysis systems compared to each other (ideally all 'observations' should have a perfect grade of 1), which gives a sense of the variation between analysis models.

Diagnostic 2: Tropopause Pressure: The pressure of the lapse rate tropopause (PTP) provides a basic measure of whether the tropopause is in the right location and how it varies over the annual cycle and response to inter-annual forcing. Responses to major forced events (ENSO and volcanoes are included in historical runs) should resemble observations. Anomalies of lapse rate tropopause pressure have been shown to be more robust than TCPT in observations and models [*Gettelman et al.*, 2009]. Simulated PTP anomalies can be compared to re-analysis systems. As described below, the grading for this diagnostic includes the correlation with inter-annual anomalies and the mean values from re-analysis systems in similar coordinates.

Diagnostic 3: Water vapor above the Cold Point Tropopause (CPT). In conjunction with TCPT, the water vapor concentration above the CPT at 80hPa is the dominant term in the total hydrogen budget of the stratosphere. This budget is important for radiation and chemistry (for example, Polar Stratospheric Cloud formation). Models should simulate appropriately the water vapor concentration in the lower tropical stratosphere, and its annual cycle.

Diagnostic 4: Ozone in the TTL is affected by both transport and chemistry. TTL ozone is an important indicator of TTL processes, as well as another baseline indicator of the entry of air into the lower stratosphere. It can be a proxy for the entry of short lived species into the stratosphere (for which we do not have sufficient observations for CCM validation). Models should represent the vertical structure of ozone and its annual cycle. Ozone is also radiatively important in the TTL, and thus critical for a correct representation of the TTL thermal structure. Since ozone is chemically produced in the TTL by various processes, it is also an integrated measure of TTL chemistry processes and TTL transport time. Differences in ozone may be due to different chemical processes (for example NOx production by lightning), which may or may not be present in a given model.

The following diagnostics do not include quantitative grades but provide a more detailed process-level view of model solutions. In most cases they required more detailed output than provided by most models, but they provide more insight into TTL processes.

Diagnostic 5: Correlations between 80hPa H_2O mixing ratio and TCPT. H_2O at 80 hPa and TCPT can be compared by translating TCPT into water vapor using the saturation vapor mixing ratio (Q_{SAT}), a function of temperature and pressure. There should be a correlation between 80hPa H_2O and TCPT. This can also be expressed as the saturation vapor mixing ratio of the TCPT (Q_{SAT}(TCPT)) and the ratio H_2O /Q_{SAT}(TCPT) should reflect the integral of physical mixing processes and dehydration.

Diagnostic 6: Tropopause Inversion Layer. The Tropopause Inversion Layer (TIL) is a layer of increased static stability that occurs just above the tropopause [*Birner*, 2006]. The TIL provides an integrated look at the dynamical structure of the TTL in the vertical. It not only shows the separation between the stratosphere and troposphere, but also provides insights into the correct dynamical results of convection in the upper troposphere, and transport and dynamics in the lower stratosphere. The static stability structure is sensitive to the radiative balance of the TTL, and hence transport of H_2O and O_3, as well as large-scale dynamics.

Diagnostic 7: TTL transport pathways and residence time. The transport time through the TTL is a complex diagnostic reflecting a mix of transport processes, including large-scale advection and mixing, as well as rapid convective motion in the vertical. Representing the transport time and pathways through the TTL is critically important for calculating the minimum temperature experienced by a parcel (which regulates water vapor). It is possible to alter stratospheric water vapor by changing transport pathways but not changing the mean temperature. Transport time is also critical for short lived species, whose lifetimes are less than a small multiple of the transport time. Several studies have attempted to assess the transport time, and here we will use Lagrangian trajectory studies to estimate transport times from a subset of models and compare them to observations.

2.3. Grading

Grades are used to obtain quantitative information on model behavior for some diagnostics. Mean values of a certain quantity or the amplitude and phase of a seasonal cycle can be used as a grade. Here, quantitative grades are defined following *Douglass et al.* [1999] and *Waugh and Eyring* [2008], with extensions to look at variability. Grades are based on defining monthly means after spatial averaging. *Douglass et al.* [1999] define a grade based on monthly mean differences:

$$g_m = max(0, 1 - \frac{1}{n}\sum_{i=1}^{n}\frac{|\mu_{iobs} - \mu_{imod}|}{n_g \sigma_{iobs}}) \quad (1)$$

Here, μ_i is a monthly mean quantity for month i from either a model (mod) or observations (obs) and $n = 12$. n_g a scaling factor representing a number of standard deviations (σ). σ_i is calculated for each month (i). If a model is more than n_g standard deviations from the observations, then $g_m = 0$. We set n_g=3 (3σ threshold) for temperature and water vapor following *Waugh and Eyring* [2008]. Because tropopause pressure is estimated from a set of coarse resolution standard levels, variability in the observations (also interpolated to these levels) is very low. So we set the 3σ threshold ($n_g \sigma_{obs}$) in Equation 1 to 10 hPa for tropopause pressure (reflecting an uncertainty of one CCMVal-2 level).

We also define a grade based on correlated variability where μ' are anomalies from a mean quantity and $\mathcal{C}$ is the linear correlation coefficient.

$$g_c = (\mathcal{C}(\mu'_{mod}, \mu'_{obs}) + 1)/2 \qquad (2)$$

For analysis here the correlation is taken on annual mean values, and thus reflects correlations of inter-annual variability between a model and observations.

We can also define a diagnostic based on the magnitude of the monthly variance of a quantity:

$$g_v = \max(0, 1 - \frac{1}{n} \sum_{i=1}^{n} \frac{|\sigma_{iobs} - \sigma_{imod}|}{n_g \sigma_{iobs}}) \qquad (3)$$

Where σ is calculated each month (i) and $n = 12$.

A single grade is then the linear combination: $G_{sum} = (g_m + g_c + g_v)/3$. The composite grade is designed to better represent uncertainty and forced variability. This partly (but not completely or rigorously) addresses shortcomings in the application of grades recently identified by *Grewe and Sausen* [2009].

We have evaluated grades using several different measures of σ_{obs} and μ_{obs} from different reanalysis systems or estimated from σ_{obs} and μ_{obs} estimated from an ensemble of re-analysis systems. While the quantitative grades do change, the relative grades between models and the spread are robust across the different methods examined. For clarity, we will report grades against one set of observations, and grade other observational data sets against that in each quantitative model summary figure to estimate the spread in grades from the observations. We also examine the multi-model mean, calculated by summing model outputs to generate a multi-model μ_{mod}. Quantitative grades for individual components are reported. The goal of applying grades is to quantitatively determine model deficiencies with sufficient detail to understand where and why models perform or do not perform well.

3. Observations and Analyses

High quality measurements in the TTL and the global UTLS for the use of model validation are challenging to obtain. In-situ instruments on balloons or aircraft are challenged by the low pressure and low temperature conditions. Remote sensing techniques used to observe the stratosphere are challenged by saturation of the measured radiances in the UTLS in many commonly used wavelengths. Additional difficulties arise from the small vertical and horizontal length scales found in the chemical and dynamical fields in the UTLS – the result of the large dynamical variability in the tropopause region. Here an overview is given of the observational data sets used for the model-measurement comparisons in the UTLS in order to provide critical information about their accuracy, precision, and potential sampling issues.

3.1. Balloon data

A variety of balloon data sources are available and used in these analyses. The global radiosonde network provides a comprehensive view of the thermal structure of the UTLS. High vertical resolution radiosondes have provided a wealth of information about the TTL structure. However, inhomogeneities in radiosonde records over time often make use of raw records problematic for trend analysis, and care must be taken when trends are analyzed [*Seidel and Randel*, 2006].

3.2. Satellite data: HALOE

Recently, satellite instruments have achieved the technological maturity to remotely sound the UTLS from space, offering an unprecedented temporal and spatial coverage of this region. Here we use water vapor observations from the Halogen Occultation Experiment (HALOE) on the UARS satellite [*Russell III et al.*, 1993]. HALOE H_2O observations have been extensively validated (e.g. *SPARC* [2000]). HALOE validation and a 13 year record (1992-2004) gives us high confidence in HALOE performance. More recent satellite measurements have not been thoroughly validated in the UTLS.

3.3. NIWA Ozone Data Set

For comparisons of simulated ozone, we use the National Institute for Water and Atmosphere (NIWA) Ozone data set described by *Hassler et al.* [2008]. The data set is a 4D reconstruction (latitude, longitude, altitude and time) using satellite and ozonesonde measurements. The current version as noted by *Hassler et al.* [2008] does not correct for known data artifacts, and may not be suitable for trends. Here we use the data base for climatological comparisons.

3.4. Meteorological Analyses

Operational meteorological analyses are produced on a daily basis by weather forecast centers. These analyses (or 're-analyses' if they are produced by consistent forecast models over time) are very valuable for model comparison, since they provide complete fields that are closely tied to observations, but with similar space scales and statistics as global models. Here we use analyses from the National Centers for Environmental Prediction and National Center for Atmospheric Research (NCEP) described by *Kalnay et al.* [1996], the NCEP and Department of Energy (NCEP2) described by *Kanamitsu et al.* [2002], the Japanese Re-Analysis (JRA) described by *Onogi et al.* [2007], the European Centre for Medium Range Weather Forecasts (ECMWF) 40 year re-analysis (ERA40) described by *Uppala et al.* [2005] and 'Interim' analysis (ERAI) described by *Uppala et al.* [2008]. For information on the different re-analyses (ERA40, NCEP, JRA) the reader is referred to *Randel et al.* [2002] and their references. A few distinct caveats common to re-analyses have to be noted. Because of the inhomogeneity of input data, specifically the introduction of significant assimilation of satellite observations starting in the late 1970's, estimating trends from re-analysis systems is difficult, and in general not scientifically justified across the late-1970's. Trend analysis since the late-1970's does usually have utility. We will use these data to estimate 'observed' trends in the UTLS. Second, re-analysis systems can have systemic biases. Perhaps most notable as an example is a significant warm bias to NCEP/NCAR reanalysis tropopause temperatures, caused by the selection of assimilated data used [*Pawson and Fiorino*, 1998]. Thus the re-analyses need to be treated with some caution. For comparison purposes with temperature and the tropopause, we will use the

ERA40 reanalysis, because of its high quality and a relatively long (20 year) record for comparison.

4. Results

In this section we present results of quantitative diagnostics (1–4 in Table 2) and their grades first. We then discuss diagnostics that are not quantitative (5) or calculated on a sub-set of models (6–7). The latter diagnostics are useful for looking in more detail at the thermal structure and transport in the TTL.

4.1. Cold Point Tropopause Temperature

The annual cycle of tropical TCPT for 18 CCMs is illustrated in Figure 1 using the REF-B1 CCMVal-2 model fields. Also shown in addition to the models are several re-analysis systems (ERA40, NCEP, NCEP2, JRA25, ERAI). All re-analyses use monthly means interpolated to CCMVal-2 standard levels (noted on Figure 5), so that the models and re-analysis systems are on the same temporal and vertical grid. TCTP is the cold point temperature on these standard levels, with no further interpolation. The gray region is 3σ from the ERA40 re-analyses. In general almost all models are able to reproduce the annual cycle. There are significant offsets between the models, but the monthly averages of 9 models are clustered within 3σ of the mean of ERA40, as seen in Figure 1 and in the quantitative grades (g_m) in Figure 2. The multi-model mean is very close to ERA40 and ERAI, closer than other analysis systems. These results are also better than CCMVal-1 models reported by *Gettelman et al.* [2009] due to the reduction of outliers, and addition of new or revised models that are closer to observations. Note that there is general quantitative agreement between the re-analyses, with 'grades' (compared to

ERA40) ranging from 0.6-0.8 (Figure 2). Lower g_m scores are largely due to mean monthly offsets (Equation 1). The amplitude and phase of the annual cycle are in good agreement between most observation systems and models. Note that NCEP and NCEP2 have a known warm TCPT bias [*Pawson and Fiorino*, 1998] that causes the g_m score to be zero when compared to ERA40.

Most models do not show strong long-term trends in TCPT, as indicated in Figure 3. The mean model trend is not significantly different from zero. NCEP and NCEP2 re-analyses show strong cooling, which is not seen in the ERA40, JRA25 or ERAI analyses (noted by *Zhou et al.* [2001]). ERA40 and ERAI also do not have trends significant at the 99% level. Note that these 'observed' trends may differ from other reported cooling trends reported from radiosondes [*Gettelman and Forster*, 2002; *Seidel and Randel*, 2006] because of limited sampling from selected radiosonde stations and the gridding and interpolation to the CCMVal-2 standard set of vertical levels. The lack of agreement among re-analyses highlights the uncertainty in long-term variability of the TCPT.

Inter-annual variability is also illustrated in Figure 3, and used for estimating correlation grades (g_c). Most models and re-analysis systems show warming of TCPT in 1991, associated with the eruption of Mt. Pinatubo. Some models have a warming that is much too large (CNRM-ACM, SOCOL, Niwa-SOCOL, MRI). This is factored into the grades for variability (g_v) as described in Equation 3 and illustrated in Figure 2. In CNRM-ACM, the warming is due to excessive heating by volcanic aerosols. Other modes of tropical variability, such as the El Nino Southern Oscillation (ENSO) or the Quasi-Biennial Oscillation (QBO) affect the tropical tropopause [*Zhou et al.*, 2001], but the effects are not clearly seen in the low vertical resolution analysis, and with many CCMs that do not have a QBO. Inter-annual anomalies are not correlated between models and re-analyses, or between re-analyses themselves.

4.2. Lapse Rate Tropopause Pressure

The pressure of the lapse rate tropopause (PTP) has been shown to be a more robust diagnostic than TCPT [*Gettelman et al.*, 2009]. PTP is more sensitive to increasing thickness below, and TCPT is a more confined vertical response. It is easier to get the bulk thickness (latent heat release) right in a model than TCPT details. This can be seen in a high (0.9 or 1.) correlation g_c among most re-analysis systems compared to ERA40 (Figure 4). Grades for 18 models are calculated based on the annual cycle (g_m), variance about monthly means (g_v) and inter-annual anomalies (g_c). The meridional structure of tropopause pressure from models and analysis systems is shown in Figure 5. The models all broadly reproduce the observed tropopause structure. There are some differences in the pressure of the tropical tropopause, which all analysis systems place near the 100 hPa level (when interpolated to CCMVal-2 levels, which are the horizontal lines in Figure 5). Several models shift the tropopause up or down by a level. There are large differences however in the diagnosed tropopause at high latitudes.

Long-term changes in PTP from 20°S-20°N are shown in Figure 6. There is good agreement between inter-annual anomalies of most of the models, as well as trends in PTP. The simulated variability in models is higher than in the observations. Most models and analysis systems show decreases in PTP associated with volcanic events (Agung 1963, El Chichon 1983, Mt. Pinatubo 1991), though the model variability is larger. In particular it is too large for CNRM-ACM, which jumps 2 levels (90 to 115hPa). The anomalies for CNRM-ACM are also evident in TCPT. PTP grades indicate a high degree of consistency among the analysis systems as noted above. CCMVal models can broadly reproduce trends and variability, but with too much variance.

4.3. Ozone

The annual cycle of tropical (20S–20N) ozone at 100 hPa is illustrated in Figure 7 from 18 models. The annual cycle of ozone near the tropical tropopause reflects a combination of: (1) chemical production (ozone is produced in the TTL at a rate of a few parts per billion per day), (2) vertical transport of ascending air, and (3) mixing with stratospheric air from higher latitudes that contains more ozone. Air with higher ozone is likely to have either (a) ascended more slowly or (b) mixed with more high-latitude air. Air with lower ozone is due to rapid transport in deep convection from the marine boundary layer. The seasonal cycle reflects these processes (chemical production and transport). Ozone is compared to the combined and processed NIWA observational data set [*Hassler et al.*, 2008] and grades based on the annual cycle and variance for this data set. Most models reproduce the phase of the annual cycle of ozone correctly in the tropics. Two models (UMSLIMCAT and CNRM-ACM) have a significantly different annual cycle of ozone (Figure 7). Many models have lower amplitude (and mean), while ULAQ, UMUKCA-METO and UMUKCA-UCAM have higher amplitude (and mean), indicating perhaps slow transport times in the TTL.

The spread of model O_3 values is reflected in many $g_m = 0$ grades (Figure 8). The CCM spread is as large as in the CCMVal-1 models (*Gettelman et al.* [2009], figure 8) with some models as similar outliers (e.g.: ULAQ). Note that the 3 models with tropospheric chemistry (CAM3.5, EMAC and ULAQ) do not have consistently better performance: ULAQ is high, and CAM3.5 and EMAC are low, and all have relatively low total (G_{sum})

grades. The higher altitude (lower pressure) tropopause in CAM3.5 and EMAC would tend to lower 100hPa O_3.

4.4. Water Vapor

Water vapor in the lower stratosphere is critical for the chemistry and climate of the stratosphere, affecting both stratospheric chemistry by regulating total hydrogen as well as affecting UTLS temperatures through the radiative impact of water vapor [*SPARC*, 2000]. Thus reproducing the transport of water vapor through the tropical tropopause is a critical requirement of CCMs in the TTL. Representing the appropriate relationships between cold point temperature and water vapor is also critical, as it requires the appropriate representation of processes that regulate water vapor, at least at the large scale.

Figure 9 presents the annual cycle of water vapor from 16 CCMs and HALOE in the lower stratosphere just above the TTL and the cold point (80 hPa). UMUKCA models fix water vapor in the stratosphere and are not shown. As pointed out by *Mote et al.* [1996], this is the entry point or 'recording head' of the stratospheric 'tape recorder' circulation. The transport associated with this circulation is discussed in *Eyring et al.* [2006]. Here we focus on the entry point. Most models are able to reproduce the annual cycle of water vapor with a minimum in NH spring and a maximum in NH fall and winter. There is a wide spread in the 'entry' value of water vapor at this level: from 2-6 ppmv, with observations from HALOE closer to 3–4 ppmv. The spread results in 5 models with $g_m = 0$. (Figure 10). The uncertainties in HALOE observations are discussed in detail in *SPARC* [2000], and are less than ±20% at this level. The shading indicates 3σ inter-annual variability, but is similar to this 20% range. These results are slightly better than CCMVal-1 models [*Gettelman et al.*, 2009] due to a tighter temperature range (Figure 1).

The multi-model mean does indicate that most models shift the water vapor minimum at 80hPa 1–2 months too early, though the multi-model mean water vapor mixing ratio is very similar to HALOE. The annual cycle is virtually absent in UMETRAC, CNRM-ACM and CCSRNIES.

4.5. Saturation at the Cold Point

Another method of examining the dehydration process is to look at the relationship between TCPT and water vapor just above the cold point (80hPa). This is a broad way of understanding integrated TTL transport and dehydration in the absence of data for off-line Lagrangian cold point calculations as in Section 4.7. TCPT regulates H_2O [*Brewer*, 1949], so the relationship can be analyzed by looking at the ratio of water vapor to the saturation vapor mixing ratio at the cold point (QSAT(CPT)). For example, minimum ERA40 TCPT (Figure 1) is about 192K, which corresponds at 80hPa to a QSAT of 5.5ppmv. Figure 11 is an update of this relationship shown in *Gettelman et al.* [2009] for 16 models.

Note that the UMUKCA models have very high cold point temperatures (consistent with high ozone at 100hPa as a result of slow transport times), so their water vapor was fixed (and they are not shown). The results indicate that most of the models cluster similarly to the observations (H_2O from HALOE and TCPT from ERA40) near a line that would imply 70% saturation with constant temperatures and transport (which is not the case, hence water is less than implied by TCPT). *Gettelman et al.* [2009] present results for 90hPa where the atmosphere is slightly drier and results are closer to a 0:0.6 line. The spread of the models is similar between CCMVal-1 and CCMVal-2. Three models are near the 1:1 line. MRI is high due to permitted ice-supersaturation. However, 3 models

(CNRM-ACM, CCSRNIES and UMETRAC) have significantly more lower stratospheric H_2O than would seem to be justified by their TCPT. This indicates potential problems in fundamental transport, variability and/or condensation processes in the TTL. This is also clear from Figure 9 and H_2O grades (Figure 10).

4.6. Tropical Tropopause Inversion Layer

Recent studies using high-resolution radiosonde data have revealed the presence of a temperature inversion layer, typically a few kilometers deep, located right above the tropopause [*Birner et al.*, 2002; *Birner*, 2006; *Bell and Geller*, 2008]. This Tropopause Inversion Layer (TIL) is also characterized by a sharp and strong buoyancy frequency maximum. The buoyancy frequency (also called the Brunt-Väisälä frequency) is defined as $N^2 = \frac{g}{\theta}\frac{d\theta}{dz}$. The presence of the TIL has been further confirmed by Global Positioning System (GPS) Radio Occultation (RO) data [*Randel et al.*, 2007; *Grise et al.*, 2009]; these independent measurements have shown that the TIL is present almost everywhere from the deep tropics to the pole in both hemispheres (Figures 12 a and d) with a minimum value in winter hemisphere polar regions. Although the formation and maintenance mechanisms of the TIL remain to be determined, its presence has potentially important implications for the cross-tropopause exchange of passive tracers/water vapor and for the dynamical coupling between stratosphere and troposphere, and has recently been receiving significant attention.

The zonal-mean structure of the TIL, simulated by REF-B1 integrations for 9 models (listed in Figure 13) with available instantaneous data, is examined and compared with observations. The observed TIL is derived from the GPS-RO data set of the Constellation

Observing System for Meteorology, Ionosphere, and Climate (COSMIC) mission from April 2006-April 2009 with about 2500-3000 soundings per day.

All analyses are performed on the log-p coordinate with tropopause pressure (p_{TP}) as a reference level: i.e. $z = -Hln(p/p_{TP})$ where H is a scale height of 8 km. Note that the conventional log-p coordinate uses surface pressure as a reference level. At each model grid point (or COSMIC profile) tropopause pressure is first computed on the native model or GPS-RO vertical grid using the WMO definition of lapse-rate tropopause. The instantaneous fields of interest, such as temperature and N^2, are then interpolated onto the tropopause-based z coordinate using a log-p linear interpolation, and are averaged over longitudes for DJF and JJA. Resulting seasonally-averaged fields in each model are finally interpolated onto 5-degree interval latitudes to construct multi-model mean fields. The COSMIC data are also binned into 5-degree intervals in latitudes. The observed TIL is computed using both data at full (or raw) levels and data only at CCMVal-2 standard levels (Figure 5). Degraded observations allow a more direct comparison of the simulated TIL with observations.

The analysis results and the average of 9 models are summarized in Figure 12 in terms of N^2. As shown in Figures 12a and d, sharp maxima of N^2, located just above the tropopause ($z = 0$), are distinct. They are generally stronger in the summer hemisphere than in the winter hemisphere, but have little hemispheric difference: i.e. the N^2 distribution in the NH summer is quantitatively similar to the one in the SH summer. These findings are consistent with previous work [*Randel et al.*, 2007; *Grise et al.*, 2009].

Figures 12b and e show the N^2 distribution for degraded GPS data. Maximum values of N^2 are lower. In addition, their locations are somewhat higher than those in the raw data. The effect is small in the tropics and larger at high latitudes. This strong sensitivity is not surprising as both tropopause pressure and temperature, which directly affect the sharpness of the TIL [*Bell and Geller*, 2008], are underestimated in coarse resolution GPS data.

The above results suggest that the CCMVal-2 models may not be able to reproduce a quantitative structure of the observed TIL, simply because of coarse resolution in the vertical. Data to perform the TIL analysis was not available for the two highest vertical resolution models (E39CA and EMAC). The simulated TIL (Figures 12c,f) is generally weaker and broader than observed using full resolution GPS RO data (Figures 12a,d). Simulations do look more like estimates from observations using CCMVal-2 vertical resolution (Figures 12b,e). Analysis of higher vertical resolution runs from WACCM with 300m vertical resolution in the UTLS (WACCM-hires) does indicate that at higher vertical resolution this model has an increased peak N^2 near the tropopause in better agreement with GPS RO observations.

Figure 13 illustrates profiles of N^2 from GPS observations and simulations in the tropics for 2 seasons from 9 models and WACCM-hires. The CCMVal-2 models underestimate N^2 in the troposphere and misplace the tropical TIL. Simulated N^2 in the tropical lower stratosphere is also much larger than observed by GPS RO, even at degraded resolution. The difference from observations might be caused by less adiabatic cooling associated with weak upwelling. Note that WACCM-hires has a larger peak N^2 and sharper gradient and closer to the tropopause than the standard resolution model. In addition, two of the lower vertical resolution models analyzed (CCSRNIES, SOCOL, see Table 1) also have very broad TIL structures.

It should be emphasized that, although the quantitative structure of the TIL is somewhat underestimated, the CCMVal-2 models successfully reproduce the qualitative structure of the TIL including its seasonality. In fact, the models' simulated TIL is more realistic than one derived from re-analysis data, especially in the extra-tropics [*Birner et al.*, 2006]. This may be because the re-analysis systems are ingesting data that may cause degradation to the structure, either through error covariances or coarse vertical resolution associated with assimilated data. Further discussion of the TIL in the extra-tropics can be found in *Hegglin et al.* [2010].

4.7. Transport in the TTL

Lagrangian trajectory studies are established tools for studying transport processes in the tropical tropopause and in particular transport from the troposphere to the stratosphere (e.g. *Hatsushika and Yamazaki* [2003]; *Bonazzola and Haynes* [2004]; *Fueglistaler et al.* [2004]). Stratospheric water vapor is strongly correlated with the Lagrangian Cold Point [*Fueglistaler and Haynes*, 2005]. We analyze the minimum temperature (T_{min}) and TTL residence time of two CCMVal-2 models, CMAM and E39CA, and compare them to ERA40 trajectories following the methodology of *Kremser et al.* [2009]. These models provided the necessary instantaneous 6-hourly fields of temperature, winds and heating rates needed to perform the calculation. Two sets of T_{min} calculations were performed using ERA40. A 'standard' calculation used 3D winds and a diabatic calculation used vertical winds based on heating rates following *Wohltmann and Rex* [2008]. The latter set of calculations using diabatic calculations is referred to as the 'reference' calculation.

The trajectories were analyzed to determine the geographical distribution of points where individual air masses encounter their minimum temperature and thus minimum water vapor mixing ratio (referred to as dehydration points) during their ascent through the TTL into the stratosphere. In addition, the residence times of air parcels in the TTL were derived.

For all years analyzed, both CCMs have a warm bias of the temperatures in the dehydration points of about 6 K (E39CA) and 8 K (CMAM) in NH winter and about 2 K (E39CA) and 4 K (CMAM) in NH summer compared to the ERA40 reference calculation. This is not the same as the temperature bias in the models (Figure 1). The Eulerian mean tropical T is about 3K low for E39CA and 1K high for CMAM. Thus the overall degree of dehydration simulated during transport of air into the stratosphere could be significantly too low, a known shortcoming of simulations with CCMs [*Eyring et al.*, 2006]. The reasons for the warm bias are probably deficiencies in transport, given differences from the model Eulerian TCPT.

Figure 14 shows that the overall geographical distribution of dehydration points in the simulation based on ERA40 data are fairly well reproduced by both CCMs in NH winter 1995–1996 (December–February, DJF). This suggests that the geographical distribution of dehydration points in winter is fairly robust. A closer look at the figure reveals that in E39CA the region of the main water vapor flux is shifted eastwards compared to ERA40 and the model shows excessive water vapor transport through warm regions over Africa. CMAM compares very well with the reference calculations and if anything only slightly overestimates the water vapor transport over the warm regions of South America. These overestimates in warm regions however are sufficient to create a significant warm bias to the Lagrangian cold point estimates.

In NH summer (June–August, JJA) 1996 the reference calculations show that the water vapor transport into the stratosphere is clearly dominated by the Indian monsoon and downwind regions (not shown), similar to *Fueglistaler and Haynes* [2005]. This is largely reproduced by CMAM, which also reproduces the location of this feature nicely. But the water vapor flux through the warm regions over Africa is overestimated. In E39CA the impact of the Indian monsoon is not well reproduced and dehydration in NH summer 1996 occurs mostly over the central Pacific rather than over India and the westernmost Pacific. The differences indicate deficiencies in TTL transport. This is different than the Eulerian transport discussed in Section 4.5.

The residence times in the upper part of the TTL (θ=385–395K) were derived from the trajectory calculations to examine the time scales of transport processes through the TTL, the key parameter for chemical transformation of air before it gets into the stratosphere. The average residence time in this layer in ERA40 diabatic calculations is about 9 days (DJF) and 12 days (JJA). These times are cut in half (faster transport) if the 'standard' winds are used. CMAM trajectories remain about 11 days (DJF) and 10 days (JJA) in the TTL, but with a long tail to the distribution for long residence times up to 30 days. E39CA residence times are 6 days in both seasons, with a similar distribution to ERA40. Thus the models do not discriminate residence time seasonally as well as ERA40.

5. Trends

The CCMVal-2 'historical' (past) and 'future' model runs provide a unique multi-model ensemble to examine trends in the UTLS. UTLS trends for CCMVal-1 models, and for Intergovernmental Panel on Climate Change (IPCC) 4th Assessment Report (AR4) models, have recently been analyzed by *Gettelman et al.* [2009], *Son et al.* [2009b] and *Son* *et al.* [2009a]. Historical trends have also been presented for REF-B1 historical simulations in the context of validating the models against observations (Figures 3 and 6). Here we further discuss historical trends and present some basic results of future trends in the UTLS from CCMVal-2 models. We present key trends from the simulations in the tropical UTLS, and in the extra-tropical LMS (below the tropical tropopause that may impact the TTL. For the latter we focus only on tropopause pressure and O_3. More details on extra-tropical diagnostics are in the companion paper by *Hegglin et al.* [2010].

Future runs were processed using zonal mean data. As noted by *Son et al.* [2009b] and *Gettelman et al.* [2009], the use of zonal mean temperatures does not significantly affect values or trends of derived tropopause parameters. We have further validated this by using four models to calculate PTP and TCPT using both 2D zonal monthly mean and 3D monthly mean temperatures (CMAM, CCSRNIES, MRI and SOCOL). Results indicate that there is less than a ±10% difference in the magnitude of the trends, and no change in significance.

5.1. Tropical Tropopause Trends

Tropical PTP in the models over the historical period is well constrained. Historical trends are similar to analysis systems, and indicate a decrease in pressure (Figure 6) in REF-B1 simulations. The robustness of the tropopause pressure grade was also noted for CCMVal-1 models by *Gettelman et al.* [2009]. Almost all models have historical trends that are close to observations and highly significant. Over 1980–1999, analyses have trends of -0.4 hPa/decade, and models are slightly higher (-0.3 to -0.9 hPa/decade). The four 'best' models (CMAM, E39CA, GEOSCCM, WACCM: see Section 6) have a mean trend of -0.6 hPa/decade. Inter-annual variability is highly correlated with observations, and generally small. Model absolute values of pressure vary, with many close to the observations, but several models are a standard level (10–15hPa) above or below. There are also generally larger decreases in pressure in the sub-tropics where the tropopause gradients are large. This implies a meridional shift in the tropopause. Future trends (from REF-B2 runs) are illustrated in Figure 15. Note that for the multiple ensembles for WACCM (3) and CMAM (2) the future trends are quantitatively the same for different ensemble members or the single model ensemble mean. There are some large differences in trends in the models. CMAM, UMSLIMCAT, UMUKCA-METO and CNRM-ACM have future trends that are larger (-10–15 hPa per century) than other models (-5 hPa/century). The multi-model mean is about -7hPa per century. In the vase of CMAM, this looks to be due to a large increase in the simulated future Brewer-Dobson Circulation [*McLandress et al.*, 2010].

Historical tropical cold point temperature trends are illustrated for the REF-B1 runs in Figure 3. Models do not show the cooling over the last 25 years seen in NCEP and NCEP2. However, an analysis of the distribution of the historical trends in space indicates coherent patterns of warming and cooling: in general the patterns represent alterations to the equatorial Kelvin wave and Rossby wave patterns induced by the change in strength of an equatorial heat source [*Gill*, 1980]. The heat source variations are changes in convection. However, different models put these patterns in different locations in the tropics. For the subset of models with cloud variables, historical trends indicate cooling in the western Pacific, and increases in clouds there. Some models indicate cooling in different regions. The overall picture is one of cooling in some regions balancing warming, for little net historical trend. This indicates that TCPT patterns respond to changes in tropical deep convection. The confidence in analysis systems might be limited by the sparse input data used for constraining the analysis models in the tropics.

TCPT future trends (from REF-B2 runs) are illustrated in Figure 16. Most models (including the best performing ones) show a slow increase in minimum temperature of 0.5–1.0 K per century. Several models (ULAQ, UMUKCA-METO) have larger future trends. As seen in Figure 9, the future temperature trends will have implications for future water vapor trends, and do have implications for future cloud trends as well.

5.2. TTL Water Vapor Trends

There exist no consistent observations of historical water vapor trends over long periods of time. There are indications of long term increases in water vapor from a variety of records [*SPARC*, 2000], and an increase in water vapor in the 1990s observed by HALOE, followed by a step change decrease after 2000. The overall historical trend in HALOE H_2O from 1992–2004 is negative (-0.05ppmv yr^{-1}) and significant at the 99% level. Almost all models also simulate a negative H_2O trend over this period, with the multi-model mean -0.03ppmv yr^{-1}. If one model with high variance (CNRM-ACM) is excluded from the multi-model mean, the trend is significant at the 99% level.

The long-term observed increase is broadly consistent with increases in methane in the latter half of the 20th century. Recent changes in water vapor (since 1992) are broadly consistent with changes in the tropical tropopause temperature (see Section 4.4 and *Randel et al.* [2006]). The changes in TCPT are partially related to changes in tropical upwelling induced by SST anomalies [*Rosenlof and Reid*, 2008]. Thus CCMs can translate surface forcing into lower stratospheric water vapor changes.

Future changes in water vapor just above the cold point are illustrated in Figure 17. Also illustrated in Figure 17 are multiple ensembles from WACCM (3) and CMAM (2), confirming that their future trends are different from each other, but consistent across the same model ensemble members. Models generally indicate that water vapor in the lower stratosphere will increase. Most model future trends are from 0.5–1.0 ppmv per century, or nearly 25%. These future trends are affected very little by methane oxidation at 80hPa, so that is unlikely to be a cause of these future trends. This is consistent with the magnitude of future TCPT trends, and future temperature trends of 0.5–1K per century at 193K translate into a 0.5–1ppmv per century increase in water vapor. Models with larger future temperature trends, or a stronger correlation between water vapor and temperature, indicate larger future increases in water vapor. This is true for example of ULAQ and CMAM (large T increase) as well as MRI, CNRM-ACM and CCSRNIES (strong dependence of H_2O on T). SOCOL indicates a large change in water vapor, without a large change in temperature. Note that UMUKCA models (fixed water vapor) and GEOSCCM (output problem with water vapor) are not included in the analysis of REF-B2. Future water vapor trends are also illustrated in Figure 18, indicating larger water vapor trends in the upper tropical troposphere at the convective outflow level near 200hPa.

5.3. Tropopause Relative Trends

Radiatively active tracers such as H_2O and O_3 exhibit large gradients across the tropopause. The radiative response to changes in these tracers is therefore expected to be highly sensitive to the detailed structure of the trends of H_2O and O_3 in the global UTLS [*Randel et al.*, 2007]. Generally, one expects the trends in absolute (e.g. pressure) coordinates to be affected by tropopause height trends. Therefore we show two sets of future trends, in absolute coordinates as well as in tropopause-based coordinates to highlight the sensitivity of trends to the tropopause. Trends are calculated based on the zonal monthly mean output with respect to the tropopause obtained from the zonal monthly mean temperature data.

Figure 18 shows multi-model ensemble of annual mean trends of O_3 (top) and H_2O (bottom) for the period 1960–2100 based on the 9 REF-B2 models with data from 1960–2100. Models included are: CAM3.5, CCSRNIES, CMAM, LMDZ-repro, MRI, SOCOL, ULAQ, UMSLIMCAT, and WACCM. The left panels show future trends in conventional (absolute) coordinates whereas the right panels show future trends in tropopause-based coordinates. The latter are obtained by first calculating the decadal shift in tropopause pressure followed by shifting the decadal changes of the respective field (O_3 or H_2O) to a reference tropopause pressure. The shift in the tropopause is shown on the left panels. Here, the average over the period 1960-1980 is used as reference state.

Future O_3 trends are negative (-2% decade^{-1}) in conventional coordinates in the tropical lower stratosphere. Decreasing O_3 is consistent with a strengthening of tropical upwelling (an enhancement of the BDC). Moderate increases of around 0.5-1.5% decade^{-1} are found throughout the upper troposphere and in the extra-tropical lower stratosphere. These results are consistent with *Hegglin and Shepherd* [2009] and *Li et al.* [2009] in the tropics and mid-latitudes, but differ in the SH polar regions. In tropopause-based coordinates however the future trends are strongly positive above the tropopause in both the tropics and extra-tropics (4-5% decade^{-1}). In the tropics the sign is reversed between conventional and tropopause based coordinates. Ozone decreases due to faster upwelling which results from an enhanced BDC. Thus O_3 decreases at any given pressure level. This may be a direct result of higher tropical SST [*Deckert and Dameris*, 2008].

But the gradient of ozone around the tropopause increases as the tropopause moves to higher altitudes, so relative to the tropopause, O_3 increases. This future trend is larger than the decrease at fixed altitude/pressure due to the strengthened BDC. In the extra-tropical lower stratosphere both contributions are positive (increasing BDC increases ozone) and are therefore amplified in tropopause-based coordinates.

H_2O exhibits strong positive future trends in the upper troposphere from the realistic upper troposphere (UT) base state. The base state has high humidity in tropical convective outflow regions and low humidity in down-welling branches of the Hadley and Walker circulations [*Gettelman and Birner*, 2007]. In the tropical UT maximum future trends of 7–8% decade^{-1} are found around 200 hPa. These future trends are likely due to increases in surface to middle tropospheric temperature associated with anthropogenic greenhouse gas induced warming. In conventional coordinates one also finds rather strong positive changes throughout the extra-tropical LMS of between 3-5% decade^{-1}. However, these changes in the LMS are in part caused by the future upward tropopause trend: in tropopause-based coordinates the strong positive trend in H_2O is largely confined to the upper troposphere whereas stratospheric H_2O shows moderate changes of around 2% decade^{-1} throughout the global lower stratosphere.

Increases in H_2O coincide with significant increases in cloud frequency of occurrence. Only a few models provided 3D TTL cloud fields for REF-B1: CAM3.5, LMDZrepro and WACCM. For all three models, the historical trend in fractional cloud coverage (cloudiness) averaged from 200-100hPa over 1960-2005 was significant at +0.0015/decade (absolute). With an average cloud fraction of 0.05, this represents 3%/decade increase in TTL cloudiness. Unfortunately, no observations of clouds exist for a similar period with such precision, and existing determinations of cloud fractions in the TTL vary strongly with instrument sensitivity. For future scenarios, results were available for 2 models (CAM3.5 and 3 WACCM realizations). CAM3.5 and WACCM are essentially versions of the same underlying tropospheric GCM, so these should be considered for clouds as 4 realizations of a similar model. Future trends in TTL cloudiness are significant at the 99% level and similar to REF-B1, +0.0012/decade (absolute), 2.5%/decade, or 25% over the 21st century (35% over the 1960–2100 period). Future trends in cloudiness are driven not by future temperature trends (since the local temperature is increasing), but by increases in water vapor of 4–9% decade^{-1} (Figure 18), modulated (reduced) by increasing temperature.

5.4. Extra-Tropical Tropopause Trends

Trends in extra-tropical tropopause pressure for future scenarios are shown as anomalies over the south (Figure 19 left panel) and north (in Figure 19, right panel) polar caps for REF-B2 simulations from 1960–2100. Multiple ensembles are shown for WACCM and CMAM. As in the tropics, PTP is expected to decrease in both hemispheres. The magnitude of the overall future trends (-20 hPa per century) are not quantitatively different between hemispheres over the 21st century. However, it is clear that there are differences in future polar tropopause pressure trends between the hemispheres: the trends in the SH polar regions are not steady, but are larger from 1960-2000 and lower (flatter) from 2000-2050. As noted by *Son et al.* [2009b] in comparing IPCC AR4 models with and without ozone depletion, these differences are due to the effects of ozone depletion (1960-2000) and recovery (2000-2050).

Quantitative trends were examined in 3 different periods, broadly characterized by ozone loss (1960–2000), ozone recovery (2001–2050), and steady ozone (2051–2099). SH tropopause pressure decreases more strongly during the ozone loss period (-0.5 hPa/yr), is flat or increases during ozone recovery, and decreases slightly during steady ozone period (-0.2 hPa/yr). Throughout all these periods there are changes in anthropogenic greenhouse gas concentrations, climate and surface temperature. In the NH, by contrast, future trends are similar in all periods and slightly negative (-0.2 hPa/yr).

5.5. Extra-Tropical Ozone Trends

Figure 18 indicates changes in ozone in the extra-tropical LMS in the 21st century. Figure 20 indicates the time-series of O_3 anomalies for the SH (left) and NH (right) averaged over the LMS (40–60 latitude, 200–100hPa). Trends are similar if different averaging domains are used. Future O_3 trends in the SH are strongly influenced by anthropogenic O_3 depletion and recovery and are not monotonic. NH future O_3 trends however are broadly monotonic in the 21st century. Since most CCMVal-2 models do not include tropospheric ozone chemistry, and those that do (CAM3.5) do not simulate different trends, these future trends must be due to changes in transport, either from decreases in isentropic transport from the tropics (reduced fraction of tropical air) or enhanced descent in the BDC. Overlaid on this trend is likely a moderate ozone depletion and recovery effect, especially evident in the SH. For the NH region in Figure 20, these future trends of +2% decade^{-1} indicate an increase of nearly +30% (0.1 ppmv) by the end of the 21st century from present (year 2000) conditions. The change is most significant and large right above the tropopause (Figure 18).

6. Summary and Conclusions

6.1. Quantitative Diagnostics and Discussion

Figure 21 includes the grading obtained for four diagnostics and provides an overall assessment of how well the models performed in the TTL. There are 4 models that score at least 0.5 on all 4 diagnostics and have consistent transport and trends: CMAM, E39CA, GEOSCCM and WACCM. The multi model mean scores highly on all the quantitative diagnostics. There are 5 more models that have 3 of 4 grades above 0.5 (AMTRAC, CAM3.5, MRI, UMETRAC, ULAQ). These thresholds are quantitatively arbitrary, but every model below this threshold has a significant deficiency in the TTL noted in the paper, and none of the highest scoring models have any obvious deficiencies in the formulation of TTL processes (e.g.: H_2O above the TCPT is appropriate for TCPT) though they may still have biases (e.g: individual grade components like $g_m = 0$). Models with obvious deficiencies score significantly lower on specific grades or components of grades. The addition of components for variance and correlation allows further insight into processes. We have not investigated the statistical significance of these grades, discussed by [*Grewe and Sausen*, 2009], and leave that as a subject for future work.

6.2. Qualitative Discussion

TCPT: The annual cycle of tropical cold point temperatures are reproduced by most models, as is the amplitude and timing of the annual cycle. There remain some significant biases between models. The UMUKCA model temperatures are too high, and CNRM-ACM and CCSRNIES temperatures are too low. CNRM-ACM has too large a response to volcanic perturbations, and SOCOL and Niwa-SOCOL are also high in this regard. Most models do not have strong trends in TCPT over the historical period. Re-analysis systems also disagree regarding estimated TCPT trends over the satellite period (since 1980).

PTP: Most models plase the tropical tropopause pressure at the right level (about 100 hPa). The UMUKCA models have higher (120hPa) PTP, which may be a reason for their tropopause temperature warm bias. The high PTP in UMUKCA models may be a function of a slightly different vertical structure in the tropopause region, and a slower BDC. CNRM-ACM, CCSRNIES, the SOCOL models and EMAC have lower tropopause pressures. Most models have historical trends in tropopause pressure consistent with observations. Again, CNRM-ACM has too large a response to volcanic events. In general model variance is higher than observed inter-annual variance of tropopause pressure. Trends are consistent between models and analysis systems and variability is highly correlated.

Tropical Ozone: The annual cycle in 100hPa ozone is generally well reproduced with high JJA summer ozone. There are some differences in the absolute value of ozone. The UMUKCA models and ULAQ have significantly higher O_3 at 100hPa than observed. CNRM-ACM and UMSLIMCAT have the wrong annual cycle. Models with tropospheric chemistry (CAM3.5, EMAC, ULAQ) do not appear to perform significantly better. The multi-model mean is a good estimate of the observations.

Tropical Water Vapor: UMETRAC, CNRM-ACM, ULAQ and MRI are too wet at 80hPa, and several models (LMDZrepro, EMAC, CMAM) are too dry, with water vapor below 3ppmv. The annual cycle is not as well produced, with many models shifted relative to HALOE observations by 1–2 months. The models generally reproduce the observed decrease in 80hPa H_2O from 1992–2004. With respect to the Cold Point Temperature and

Water Vapor correlation, there are 3 models (CCSRNIES, CNRM-ACM and UMETRAC) that are clear outliers: there appears to be more water vapor than the temperatures would permit if transport were occurring similarly to observations. UMUKCA models prescribe TTL water vapor.

Tropopause Inversion Layer: Models are able to simulate a TIL. The TIL resembles observations on a similar coarse vertical resolution, but extends deeper vertically than high vertical resolution observations. The maximum value of N^2 is found at higher altitude than observed. Higher vertical resolution does improve model simulations. Models reproduce the annual cycle in TIL structure, with the tropical TIL slightly stronger during DJF and the extra-tropical TIL stronger in the summer hemisphere.

Lagrangian Cold Point: Two models examined broadly reproduce the distribution of Lagrangian minimum temperatures (T_{min}) in analysis systems. However, T_{min} is higher than the ERA40 reference calculation, due to differences in transport location. Consistent with a high T_{min}, H_2O is high in one model (E39CA) but not in the other (CMAM). Further work with more models is needed to better understand these differences.

There is a spread of residence times in the two models, mirroring spread in analysis systems using different vertical advection. It is likely that model residence times are a stringent test of the model vertical advection schemes and schemes that are too diffusive will have short residence times.

6.3. Conclusions

The results of this analysis indicate that there is a spread in performance among models in the TTL relative to observations, and there are some (4) models with quantitatively better results relative to observations, but half of the models (9 of 18) perform well on most (3 of 4) grades. The multi model mean generally is a very good representation of the TTL. Quantitative grades including variability confirm the qualitative view of models. Further work to make the grading of models more rigorous is desired.

The tropical tropopause pressure and CPT exhibit significant biases between models, although the seasonal cycles are generally reasonable. This finding implies a wide range of tropical LS H_2O values. However, the spread of CPT values is smaller than for CCMVal-1 models [*Gettelman et al.*, 2009], indicating improvement in overall model performance. The amplitude and phase of the annual cycle is improved and all models monthly anomalies of TCPT are within 3σ of the observations.

Critically, many models and the multi-model mean can now broadly reproduce recently observed decreases in lower stratospheric water vapor, likely related to SST variability. Thus models can translate SST forcing into changes in lower stratospheric H_2O.

Comparison of the TCPT with H_2O reveals simulated transport behavior different from observations where models have higher water vapor concentrations above the cold point than implied by the saturation value of TCPT. The observed mean ratio of 80hPa water vapor to the saturation value at the cold point minimum temperature is about 0.65–0.7, and most models reproduce this ratio, yielding increased confidence in TTL transport.

Lagrangian cold points in the two models examined have a reasonable distribution but suffer from temperature biases, and the TIL depth is generally too deep and slightly shifted from observations. The representation of the TIL appears to be a function of vertical resolution. Degraded resolution observations are more similar to models, and a higher vertical resolution model (δz=300m in the TTL) has gradients in stability that better resemble observations. Hence higher vertical resolution seems to improve the representation of stability in the TTL.

Simulations indicate significant impacts of stratospheric O_3 depletion on historical and future trends in extra-tropical tropopause pressure and on historical and future O_3 trends in the extra-tropical LMS. NH and SH future trends are very different, and SH trends are not monotonic due to O_3 depletion and recovery. Ozone depletion strengthens the trends in the SH, and recovery weakens the trends. This is consistent with other recent analyses with CCMVal-1 models [*Son et al.*, 2009b]. Extra-tropical LMS O_3 trends may impact O_3 concentrations in the TTL through quasi-isentropic transport. Extra-tropical PTP trends are indicators of shifts in the sub-tropical jets and circulation that may impact the tropics, for example by increasing the width of the tropical belt [*Seidel et al.*, 2008].

The projected O_3 increase in the NH extra-tropical LMS is nearly 30% by the end of the 21st century. This is not due to tropospheric chemistry, but most likely is due to increased down-welling from an enhanced BDC and the effects of ozone recovery, also noted by *Hegglin and Shepherd* [2009] and *Li et al.* [2009]. These significant changes might affect the tropopause structure, and radiative forcing calculated at the tropopause, as well as the stratosphere-troposphere exchange of ozone and upper tropospheric ozone. Understanding the mechanisms for this increase using CCMs with tropospheric chemistry is a critical future endeavor [*Hegglin and Shepherd*, 2009; *Stevenson*, 2009].

Future increases in tropical ozone with respect to the tropopause also strongly imply changes to TTL transport that might affect short lived species (for example, those containing bromine). Future CCM simulations should include a suite of short lived compounds to better evaluate TTL transport and chemistry.

Simulations show good historical fidelity with observed trends and anomalies in PTP. Models do not reproduce historical TCPT trends, but these are uncertain from re-analyses. Models project decreases in tropical PTP in the 21st century. Simulated quantitative trends in PTP are similar to trends found by *Gettelman et al.* [2009] with a small subset of CCMVal-1 models run to 2100. The quantitative values quoted are for those 4 models with high quantitative grades, yielding a higher confidence in these results than in earlier analyses.

Models reproduce recent decreases in H_2O seen in re-analyses and HALOE observations. This yields confidence in future trends. Increasing H_2O in the tropical lower stratosphere is associated with increasing TCPT and decreasing PTP. Changes over 2000–2100 are significant nearly +1K in TCPT and +1ppmv of water vapor, representing a 20–30% increase. There remains some spread in reported model results, but most outliers for trends occur due to noted model deficiencies that are traceable to low performance in some diagnostics.

However, there is little spatial coherence across models in the structure of historical or future trends in water vapor (and temperature), except to tie them to the parametrized process of deep cumulus convection. There are large future increases in water vapor in the lower region of the TTL near 200hPa. Consistent with this picture, there are significant increases in TTL cloudiness (35% over the 1960-2100 period) in the one family of models with cloud fields to 2100. Thus improving confidence in convective parametrization and its effect on tropical atmospheric dynamics and thermodynamics is critical for improving confidence in predictions of the future state of the TTL, both for transport into the stratosphere and radiative effects on surface climate.

What has changed since CCMVal-1 [*Gettelman et al.*, 2009]? First, there are many more models for analysis, so the multi-model mean is more significant. Second, the spread of TCPT has narrowed. Third, historical runs now simulate modest recent decreases in lower stratosphere H_2O, as do observations. This yields increasing confidence in future trends in TCPT and H_2O. Fourth, we have a much more detailed picture from a limited subset of models of the thermal structure of the TTL (TIL) and the transport through the TTL in simulations. There are still deficiencies in many models in TCPT and TTL transport, but quantitative assessment indicates at least half the models are performing acceptably in the TTL.

The strongest overall recommendations for improving the representation of the TTL in CCMs are: (1) improving vertical resolution and (2) addition of tropospheric chemistry and short lived species. Additionally, making available limited high frequency output (for trajectory studies) would improve the level of possible process-based analysis.

7. Acknowledgments

The National Center for Atmospheric Research is sponsored by the United States National Science Foundation. We acknowledge the modeling groups for making their simulations available for this analysis, the Chemistry-Climate Model Validation (CCMVal) Activity for WCRP's (World Climate Research Programme) SPARC (Stratospheric Processes and their Role in Climate) project for organizing and coordinating the model data analysis activity, and the British Atmospheric Data Center (BADC) for collecting and archiving the CCMVal model output. CCSRNIES research was supported by the Global Environmental Research Fund of the Ministry of the Environment of Japan (A-071). CCSRNIES and MRI simulations were made with the supercomputer at the National Institute for Environmental Studies, Japan. European contributions were supported by the European Union Integrated Project SCOUT-O3. WACCM-hires simulations were performed at the Centro de Supercomputacion de Galicia. Thanks to Darryn Waugh for the use of code for producing tables.

References

Akiyoshi, H., et al., A CCM simulation of the breakup of the antarctic polar vortex in the years 1980–2004 under the CCMVal scenarios, *J. Geophys. Res.*, *114*(D03103), doi:10.1029/2007JD009261, 2009.

Austin, J., and N. Butchart, Coupled chemistry-climate model simulation for the period 1980 to 2020: ozone depletion and the start of ozone recovery, *Q. J. R. Meteorol. Soc.*, *129*, 3,225–3,249, 2003.

Austin, J., and R. J. Wilson, Sensitivity of polar ozone to sea surface temperatures and chemistry, *subitted to J. Geophys. Res.*, 2009.

Bell, S. W., and M. A. Geller, Tropopause inversion layer: Seasonal and latitudinal variations and representation in standard radiosonde data and global models, *J. Geophys. Res.*, *113*(D05109), doi:10.1029/2007JD009022, 2008.

Birner, T., Fine-scale structure of the extratropical tropopause region, *J. Geophys. Res.*, *111*, doi:10.1029/2005JD006301, 2006.

Birner, T., A. Dornbrack, and U. Schumann, How sharp is the tropopause at midlatitudes?, *Geophys. Res. Lett.*, *29*(14), doi:1029/2002GL015142, 2002.

Birner, T., D. Sankey, and T. G. Shepherd, The tropopause inversion layer in models and analyses, *Geophys. Res. Lett.*, *33*, L14,804, doi:10.1029/2006GL026,549, 2006.

Bonazzola, M., and P. H. Haynes, A trajectory-based study of the tropopause region, *J. Geophys. Res.*, *109*(D20112), doi:10.1029/2003JD004356, 2004.

Brewer, A. W., Evidence for a world circulation provided by the measurements of helium and water vapor distribution in the stratosphere, *Q. J. R. Meteorol. Soc.*, *75*, 351–363, 1949.

Corti, T., B. P. Luo, Q. Fu, H. Vömel, and T. Peter, The impact of cirrus clouds on tropical troposphere-to-stratosphere transport, *Atmos. Chem. Phys.*, *6*, 1725–1747, 2006.

de Grandpré, J., S. R. Beagley, V. I. Fomichev, E. Griffioen, J. C. McConnell, A. S. Medvedev, and T. G. Shepherd, Ozone climatology using interactive chemistry: results from the Canadian Middle Atmosphere Model, *J. Geophys. Res.*, *105*, 26,475–26,491, 2000.

Deckert, R., and M. Dameris, Higher tropical ssts strengthen the tropical upwelling via deep convection, *Geophys. Res. Lett.*, *35*(L10813), doi:10.1029/2008GL033719, 2008.

Déqué, M., Frequency of precipitation and temperature extremes over france in an anthropogenic scenario: model results and statistical correction according to observed values, *Global and Planetary Change*, *57*, 16–26, 2007.

Douglass, A. R., M. J. Prather, T. M. Hall, S. E. Strahan, P. J. Rasch, L. C. Sparling, L. Coy, and J. M. Rodriguez, Choosing meteorological input for the global modeling initiative assessment of high-speed aircraft, *J. Geophys. Res.*, *104*, 27,545–27,564, 1999.

Egorova, T., E. Rozanov, V. Zubov, E. Manzini, W. Schmutz, and T. Peter, Chemistry-climate model SOCOL: a validation of the present-day climatology, *Atmos. Chem. Phys.*, *5*, 1557–1576, 2005.

Eyring, V., et al., Assessment of temperature, trace species, and ozone in chemistry-climate model simulations of the recent past, *J. Geophys. Res.*, *111*(D22308), doi:10.1029/2006JD007327, 2006.

Eyring, V., et al., Multi-model projections of stratospheric ozone in the 21st century, *J. Geophys. Res.*, *112*, doi:10.1029/2006JD008332, 2007.

Eyring, V., et al., Overview of the new CCMVal reference and sensitivity simulations in support of upcoming ozone and climate assessments and the planned SPARC CCMVal assessment, *SPARC Newsletter*, (30), 20–26, 2008.

Fueglistaler, S., and P. H. Haynes, Control of interannual and longer-term variability of stratospheric water vapor, *J. Geophys. Res.*, *110*(D08107), doi:10.1029/2005JD006019, 2005.

Fueglistaler, S., H. Wernli, and T. Peter, Tropical troposphere-to-stratosphere transport inferred from trajectory calculations, *J. Geophys. Res.*, *109*(D03108), doi:10.1029/2003JD004069, 2004.

Fueglistaler, S., A. E. Dessler, T. J. Dunkerton, I. Folkins, Q. Fu, and P. W. Mote, The tropical tropopause layer, *Rev. Geophys.*, *47*(1004), doi:10.1029/2008RG000267, 2009.

Garcia, R. R., D. Marsh, D. Kinnison, B. A. Boville, and F. Sassi, Simulations of secular trends in the middle atmosphere, 1950-2003, *J. Geophys. Res.*, 2007.

Garny, H., M. Dameris, and A. Stenke, Impact of prescribed ssts on climatologies and long-term trends in CCM simulations, *Atmos. Chem. Phys.*, *9*, 6017–6031, 2009.

Gettelman, A., and T. Birner, Insights on tropical tropopause layer processes using global models, *J. Geophys. Res.*, *112*(D23104), doi:10.1029/2007JD008945, 2007.

Gettelman, A., and P. M. F. Forster, A climatology of the tropical tropopause layer, *J. Met. Soc. Japan*, *80*(4B), 911–924, 2002.

Gettelman, A., et al., The tropical tropopause 1960–2100, *Atmos. Chem. Phys.*, *9*, 1621–1637, 2009.

Gill, A. E., Some simple solutions for heat–induced tropical circulation, *Q. J. R. Meteorol. Soc.*, *106*, 447–462, 1980.

Grewe, V., and R. Sausen, Comment on "quantitative performance metrics for stratospheric-resolving chemistry-climate models" by waugh and eyring, *Atmos. Chem. Phys. Disc.*, *9*, 14,141–14,164, 2009.

Grise, K. M., D. W. J. Thompson, and T. Birner, A global survey of static stability, *submitted to J. Climate*, 2009.

Hassler, B., G. E. Bodeker, and M. Dameris, Technical note: A new global database of trace gases and aerosols from multiple sources of high vertical resolution measurements, *Atmos. Chem. Phys.*, *8*, 5403–5421, 2008.

Hatsushika, H., and K. Yamazaki, Stratospheric drain over Indonesia and dehydration within the tropical tropopause layer diagnosed by air parcel trajectories, *J. Geophys. Res.*, *108*(D19), doi:10.1029/2002/JD002986, 2003.

Hegglin, M. I., and T. G. Shepherd, Large climate-induced changes in uv index and stratosphere-to-troposphere ozone flux, *Nature Geosci.*, *2*, 687–691, 2009.

Hegglin, M. I., et al., Multi-model assessment of the upper troposphere and lower stratosphere: Extra-tropics, *submitted to J. Geophys. Res.*, 2010.

Hein, R., et al., Results of an interactively coupled atmospheric chemistry-general circulation model: Comparison with observations, *Ann. Geophysicae*, *19*(435-457), 2001.

Jöckel, P., et al., The atmospheric chemistry general circulation model ECHAM/MESSy1: consistent simulation of ozone from the surface to the mesosphere, *Atmos. Chem. Phys.*, *6*(5067-5104), 2006.

Jourdain, L., S. Bekki, F. Lott, and F. Lefevre, The coupled chemistry model LMDz Reprobus: description of a transient simulation of the period 1980-1999, *Ann. Geophysicae*, *6*, 1391–1413, 2008.

Kalnay, E., et al., The NCEP/NCAR 40-year reanalysis project, *Bull. Am. Meteorol. Soc.*, *77*(3), 437–471, 1996.

Kanamitsu, M., W. Ebisuzaki, J. Wollen, S.-K. Yang, J. J. Hnilo, M. Fiorino, and G. L. Potter, NCEP-DEO AMIP II reanalysis (R-2), *Bull. Am. Meteorol. Soc.*, pp. 1631–1643, 2002.

Kremser, S., I. Wohltmann, M. Rex, U. Langematz, M. Dameris, and M. Kunze, Water vapor transport in the tropical tropopause region in coupled chemistry-climate models and era40 reanalysis data, *Atmos. Chem. Phys.*, *9*(8), 2679–2694, 2009.

Lamarque, J. F., D. E. Kinnison, P. G. Hess, and F. M. Vitt, Simulated lower stratospheric trends between 1970 and 2005: Identifying the role of climate and composition changes, *J. Geophys. Res.*, *113*(D12301), doi:10.1029/2007JD009277, 2008.

Li, F., R. S. Stolarski, and P. A. Newman, Stratospheric ozone in the post-cfc era, *Atmos. Chem. Phys.*, *9*, 2207–2213, 2009.

McLandress, C. A., I. Jonsson, D. A. Plummer, M. C. R. a J. F. Scinocca, and T. G. Shepherd, Separating the dynamical effects of climate change and ozone depletion: Part 1. southern hemisphere stratosphere, *submitted to J. Climate*, 2010.

Morgenstern, O., et al., The world avoided by the montreal protoccol, *Geophys. Res. Lett.*, *35* (L16811), doi:10.1029/2008GL034590, 2008.

Morgenstern, O., et al., Evaluation of the new UKCA climate-composition model. part 1: The stratosphere, *Geosci. Model Dev.*, *1*, 43–57, 2009.

Morgenstern, O., et al., Review of the formulation of present-generation stratospheric chemistry-climate models and associated external forcings, *in press, J. Geophys. Res.*, 2010.

Mote, P. W., et al., An atmospheric tape recorder: The imprint of tropical tropopause temperatures on stratospheric water vapor, *J. Geophys. Res.*, *101* (D2), 3989–4006, 1996.

Onogi, K., et al., The JRA-25 reanalysis, *J. Meteor. Soc. Japan*, *85*, 369–432, 2007.

Pawson, S., and M. Fiorino, A comparison of reanalyses in the tropical stratosphere. Part 1: thermal structure and the annual cycle, *Climate Dynamics*, *14*, 631–644, 1998.

Pawson, S., R. S. Stolarsky, A. R. Douglass, P. A. Newman, J. E. Nielsen, S. M. Frith, and M. L. Gupta, Goddard earth observing system chemistry-climate model simulations of stratospheric ozone-temperature coupling between 1950 and 2005, *J. Geophys. Res.*, *113* (D12103), doi:10.1029/2007JD009511, 2008.

Pitari, G., E. Mancini, V. Rizzi, and D. Shindell, Impact of future climate and emission changes on stratospheric aerosols and ozone, *J. Atmos. Sci.*, *59*, 414–440, 2002.

Randel, W. J., M. L. Chanin, and C. Michaut, SPARC intercomparison of middle atmosphere climatologies, *Tech. Rep. WCRP 116, WMO/TD - 1142, SPARC Report 3*, SPARC, World Climate Research Program, 2002.

Randel, W. J., F. Wu, H. Vomel, G. E. Nedoluha, and P. F. Forster, Decreases in stratospheric water vapor since 2001: links to changes in the tropical tropopause and the brewer-dobson circulation, *J. Geophys. Res.*, *111*(D12), doi:10.1029/2005JD006744, 2006.

Randel, W. J., M. Park, F. Wu, and N. Livesey, A large annual cycle in ozone above the tropical tropopause linked to the brewer-dobson circulation, *J. Atmos. Sci.*, *64*(12), 4479–4488, 2007.

Rosenlof, K. H., and G. C. Reid, Trends in the temperature and water vapor content of the tropical lower stratosphere: Sea surface connection, *J. Geophys. Res.*, *113*(D06107), 2008.

Russell III, J. M., et al., The halogen occultation experiment, *J. Geophys. Res.*, *98*(D6), 10,777–10,797, 1993.

Schraner, M., et al., Technical note: Chemistry-climate model SOCOL: version 2.0 with improved transport and chemistry/microphysics schemes, *acp*, *8*, 5957–5974, 2008.

Scinocca, J. F., N. A. McFarlane, M. Lazare, J. Li, and D. Plummer, Technical note: The CCCma third generation AGCM and its extension into the middle atmosphere, *Atmos. Chem. Phys.*, *8*, 7055–7074, 2008.

Seidel, D. J., and W. J. Randel, Variability and trends in the global tropopause estimated from radiosonde data, *J. Geophys. Res.*, *111*(D21101), doi:10.1029/2006JD007363, 2006.

Seidel, D. J., Q. Fu, W. J. Randel, and T. Reichler, Widening of the tropical belt in a changing climate, *Naure Geosci.*, *1*, 21–24, doi:10.1038/ngeo.2007.38, 2008.

Shepherd, T. G., Transport in the middle atmosphere, *J. Met. Soc. Japan*, *85B*, 165–191, 2007.

Shibata, K., and M. Deushi, Long-term variations and trends in the simulation of the middle atmosphere 1980-2004 by the chemistry-climate model of the meteorological research institute, *Ann. Geophysicae*, *26*, 1299–1326, 2008a.

Shibata, K., and M. Deushi, Simulation of the stratospheric circulation and ozone during the recent past (1980-2004) with the MRI chemistry=climate model, *Tech. Rep. CGER Supercomputer Mongraph Report Vol 13*, National Institute for Environmental Studies, 2008b.

Son, S. W., N. F. Tandon, L. M. Polvani, and D. W. Waugh, Ozone hole and southern hemisphere climate change, *Geophys. Res. Lett.*, *36*(L15705), doi: 10.1029/2009/GL038671, 2009a.

Son, S. W., et al., The impact of stratospheric ozone recovery on tropopause height trends, *J. Clim.*, *22*, 429–445, doi:10.1175/2008JCLI12215.1, 2009b.

SPARC, *Assessment of Water Vapor in the Upper Troposphere and Lower Stratosphere*, WMO/TD-1043, Stratospheric Processes and Their Role In Climate, World Meteorological Organization, Paris, 2000.

Stenke, A., M. Dameris, V. Grewe, and H. Garny, Implications of lagrangian transport for coupled chemistry-climate simulations, *Atmos. Chem. Phys.*, *9*, 5489–5504, 2009.

Stevenson, D. S., Putting the wind up ozone, *Nature Geosci.*, *2*(677-679), 2009.

Teyssèdre, H., et al., A new topospheric and stratospheric chemistry and transport model MOCAGE-climat for multi-year studies: evaluation of the present-day climatology and sensitivity to surface processes, *Atmos. Chem. Phys.*, *7*, 5815–5860, 2007.

Tian, W., and M. P. Chipperfield, A new coupled chemistry-climate model for the stratosphere: The importance of coupling for future o3-climate predictions, *Q. J. R. Meteorol.*

Soc., *131*, 281–304, 2005.

Tian, W., M. P. Chipperfield, L. J. Gray, and J. M. Zawodny, Quasi-biennial oscillation and tracer distributions in a coupled chemistry-climate model, *J. Geophys. Res.*, *111*(D20301), doi:10.1029/2005JD006871, 2006.

Uppala, S., D. Dee, S. Kobauashi, P. Berrisford, and A. Simmons, Towards a climate data assimilation system: status update of era-interim, *ECMWF Newsletter*, (115), 12–18, 2008.

Uppala, S., et al., The era-40 re-analysis, *Q. J. R. Meteorol. Soc.*, *131*, 2961–3012, 2005.

Waugh, D. W., and V. Eyring, Quantitative performance metrics for stratosphere-resolving models, *Atmos. Chem. Phys.*, *8*, 5699–5713, 2008.

Wohltmann, I., and M. Rex, Improvement of vertical and residual velocities in pressure or hybrid sigma-pressure coordinates in analysis data in the stratosphere, *Atmos. Chem. Phys.*, *8*(2), 265–272, 2008.

World Meteorological Organization, *Scientific Assessment of Ozone Depletion: 2006*, WMO Report 50, World Meteorological Organization, Geneva, 2007.

Zhou, X., M. A. Geller, and M. Zhang, The cooling trend of the tropical cold point tropopause temperatures and its implications, *J. Geophys. Res.*, *106*(D2), 1511–1522, 2001.

Table 1. Description of models used in this study. Horizontal resolution is in degrees of latitude (longitudes are 20-50% larger), and truncation is in parentheses if the model is not on a latitude-longitude grid. (T for triangular, R for rhomboidal) TTL levels are the number of levels between 300 and 100 hPa.

	Name	Horiz. Res.	TTL Lvs	References
1	AMTRAC3	2		*Austin and Wilson* [2009]
2	CAM3.5	2	7	*Lamarque et al.* [2008]
3	CCSRNIES	2.8(T42)	6	*Akiyoshi et al.* [2009]
4	CMAM	3.75 (T31)	7	*Scinocca et al.* [2008]; *de Grandpré et al.* [2000]
5	CNRM-ACM	(T63)	8	*Déqué* [2007]; *Teyssèdre et al.* [2007]
6	E39CA	3.75(T30)	15	*Stenke et al.* [2009]; *Garny et al.* [2009]; *Hein et al.* [2001]
7	EMAC	2.8(T42)	12	*Jöckel et al.* [2006]
8	GEOSCCM	2	7	*Pawson et al.* [2008]
9	LMDZrepro	2.5	8	*Jourdain et al.* [2008]
10	MRI	2.8(T42)	6	*Shibata and Deushi* [2008a, b]
11	SOCOL	3.75 (T30)	5	*Schraner et al.* [2008]; *Egorova et al.* [2005]
12	Niwa-SOCOL	3.75 (T30)	5	See SOCOL
13	ULAQ	11.5 (R6)	3	*Pitari et al.* [2002]; *Eyring et al.* [2006, 2007]
14	UMETRAC	2.5	9	*Austin and Butchart* [2003]
15	UMSLIMCAT	2.5	9	*Tian and Chipperfield* [2005]; *Tian et al.* [2006]
16	UMUKCA-METO	2.5	7	*Morgenstern et al.* [2008, 2009]
17	UMUKCA-UCAM	2.5	7	See UMUKCA-METO
18	WACCM	2	7	*Garcia et al.* [2007]

Table 2. Diagnostics used in this study. Monthly means are used for analysis, except for instantaneous data noted by a superscript 'i' in the table. Monthly means are on CCMVal-2 standard levels (shown in Figure 5) and instantaneous data is on model levels. Data sets are described in more detail in the text.

Diagnostic	Variables	# Models	Data
1 TCPT	T	18	Re-analyses
2 PTP	T	18	Re-analyses
3 O3	O_3	18	NIWA
4 H2O	H_2O	16	HALOE
5 QSAT(TCPT)	H_2O	16	Re-analyses, HALOE
6 TIL	T^i	9	GPS
7 Transport	T^i,U^i,V^i	2	ERA40

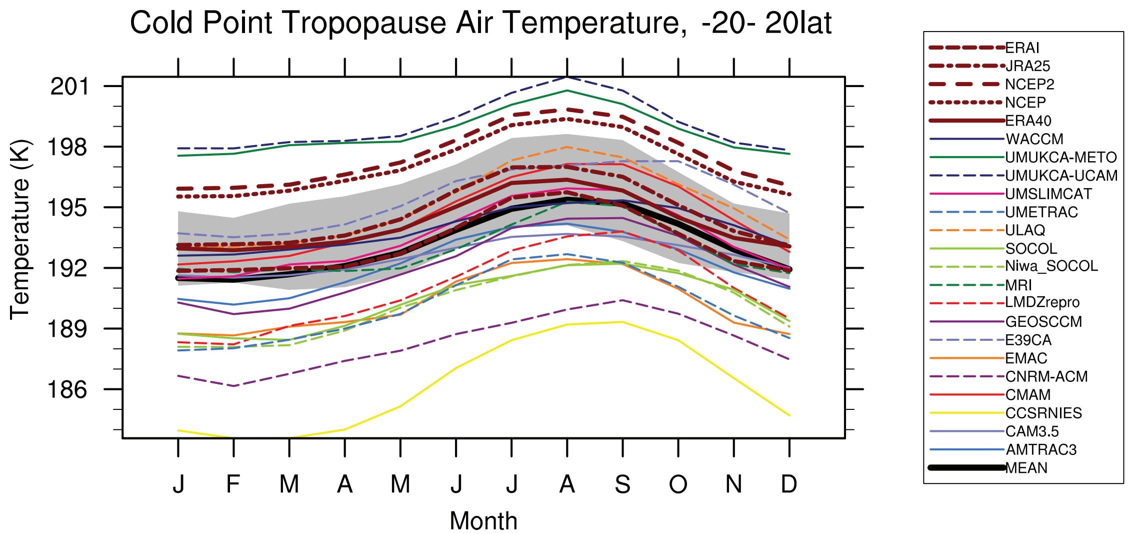

Figure 1. Annual cycle of tropical (20S-20N) cold point tropopause temperature (TCPT) from models and observations. Output and observations are from the period 1980-1999. Gray shaded region is 3σ variability from ERA40 analyses. Reanalysis systems in brown with different line styles: ERA40 (solid), ERAI (short dash), JRA25 (dot dash), NCEP (dotted), NCEP2 (long dashed). The multi-model mean (MEAN) is the thick black line.

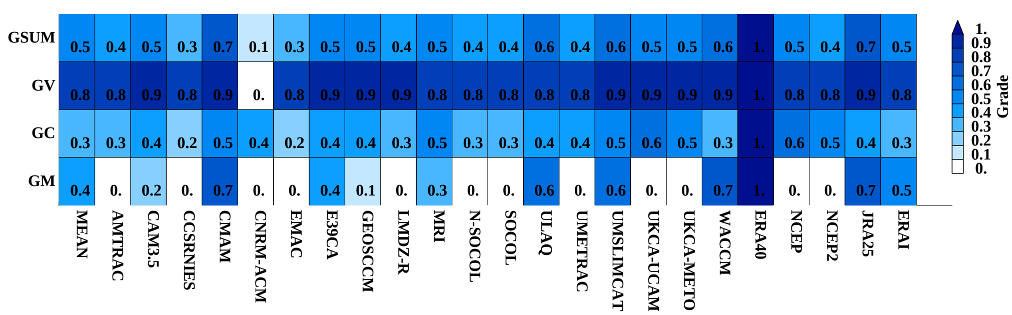

Figure 2. Quantitative diagnostic summary of Cold Point Tropopause Temperature (TCTP) for mean (GM), correlation (GC), variance (GV) and the average (GSUM).

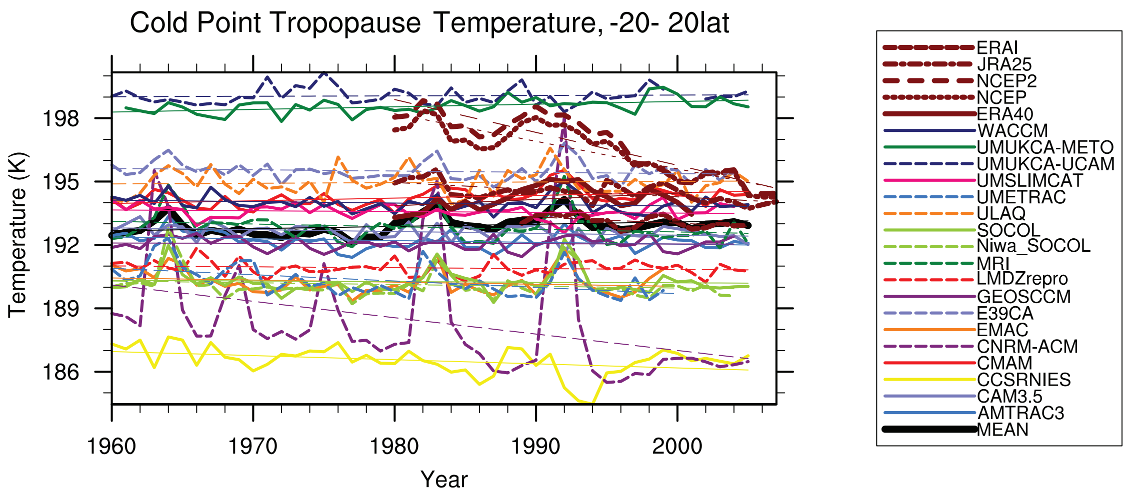

Figure 3. Cold point tropopause temperature time series for 20S-20N from models and re-analyses for 1960-2007. Thin lines are linear fits. The multi-model mean (MEAN) is the thick black line.

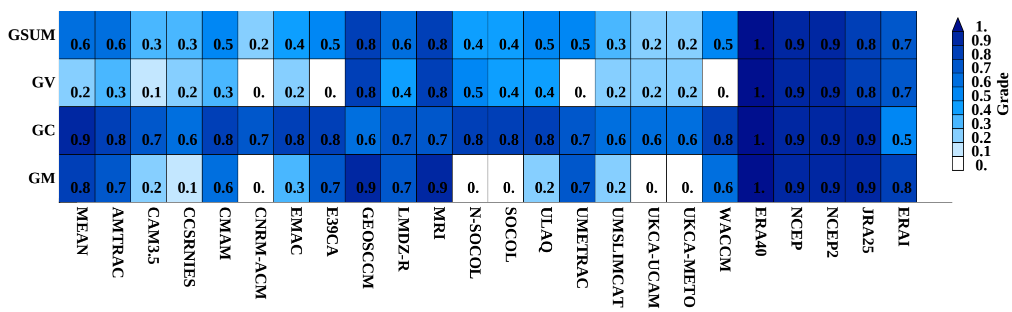

Figure 4. Quantitative grades summary of Lapse Rate Tropopause Pressure for mean (GM), correlation (GC), variance (GV) and the average (GSUM).

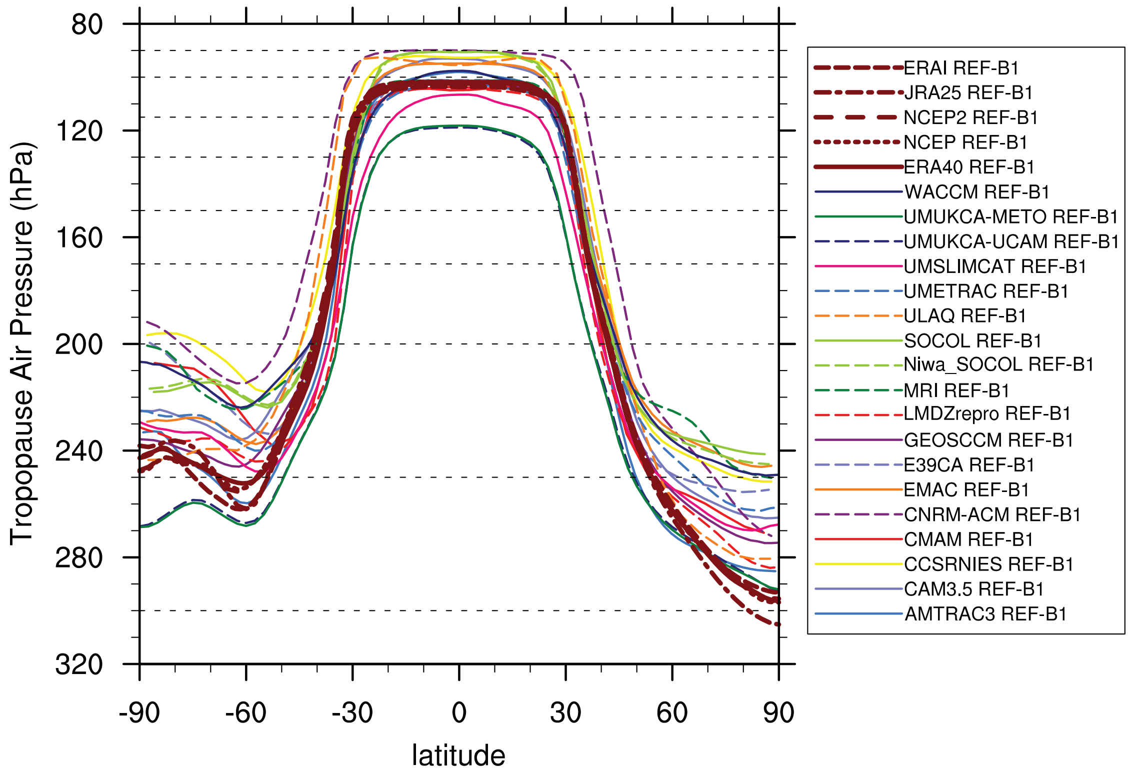

Figure 5. REF-B1 lapse rate tropopause pressure (PTP) annual zonal mean for 1980–1999 from models and analysis systems. Dotted lines represent CCMVal-2 vertical level structure in the UTLS, with levels at 400, 300, 250, 200, 170, 150, 130, 115, 100, 90, 80, 70, 50 hPa.

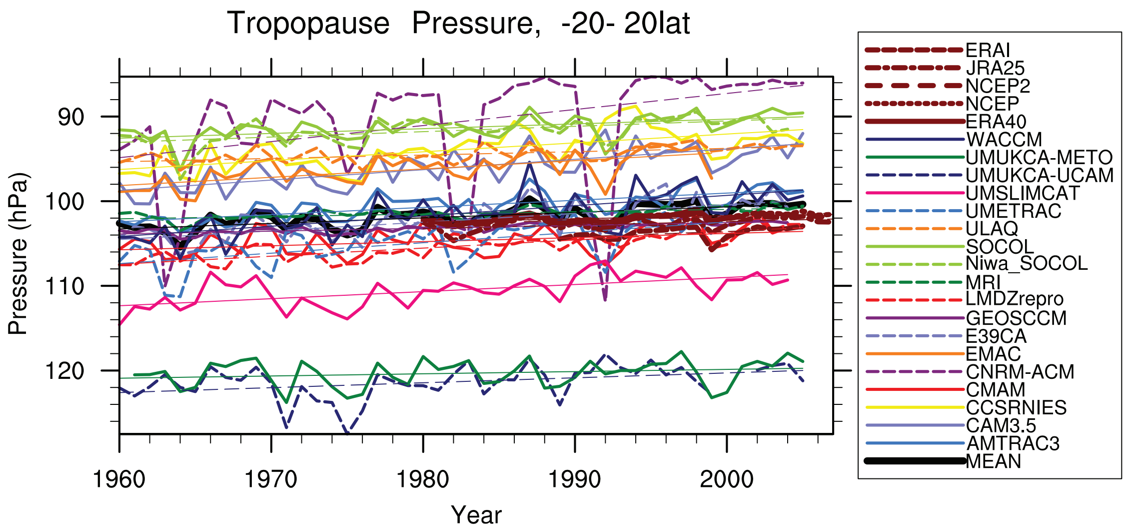

Figure 6. Lapse Rate Tropopause Pressure (PTP) time series for 20S-20N from models and 4 re-analyses for 1960-2007. Thin lines are linear fits. The multi-model mean (MEAN) is the thick black line.

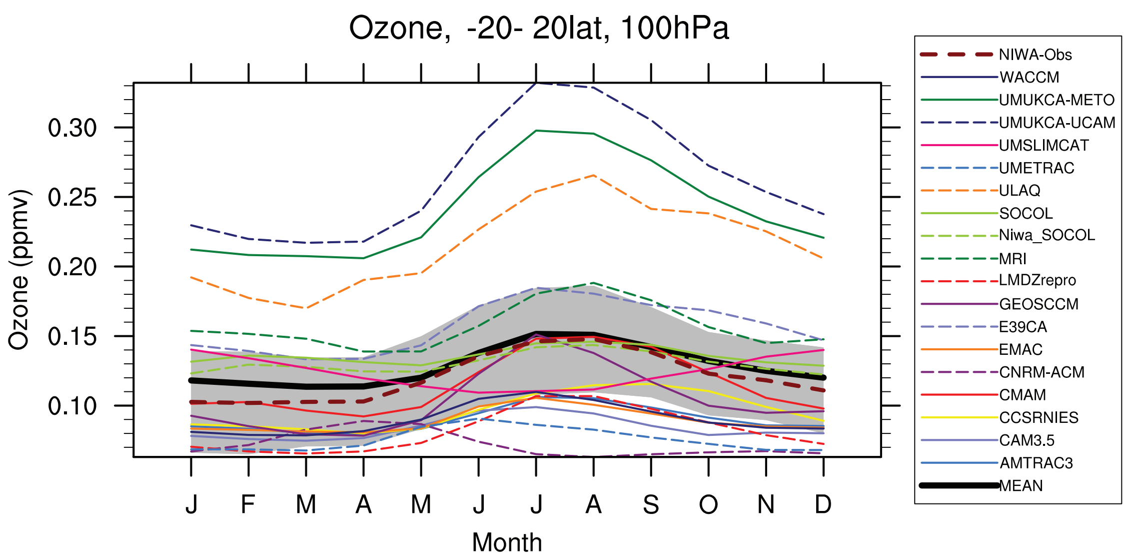

Figure 7. Annual cycle of tropical (20S-20N) 100hPa ozone mixing ratio from models and observations. Output and observations are from the period 1980-2005. Gray shaded region is 3σ variability from NIWA observational data set (dashed brown line). The multi-model mean (MEAN) is the thick black line.

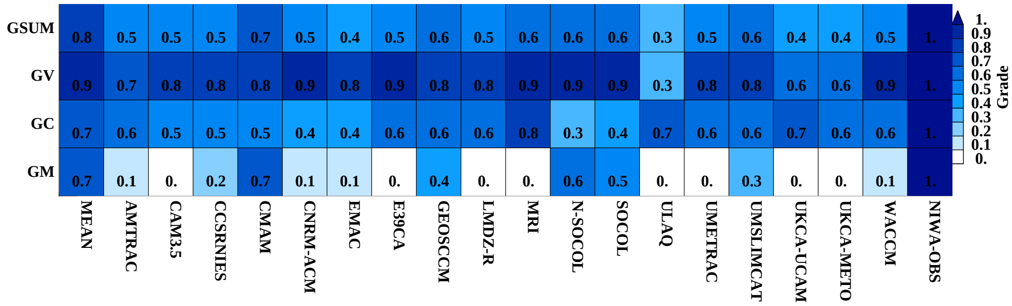

Figure 8. Quantitative diagnostics summary of 100hPa Ozone mixing ratio for mean (GM), correlation (GC), variance (GV) and the average (GSUM).

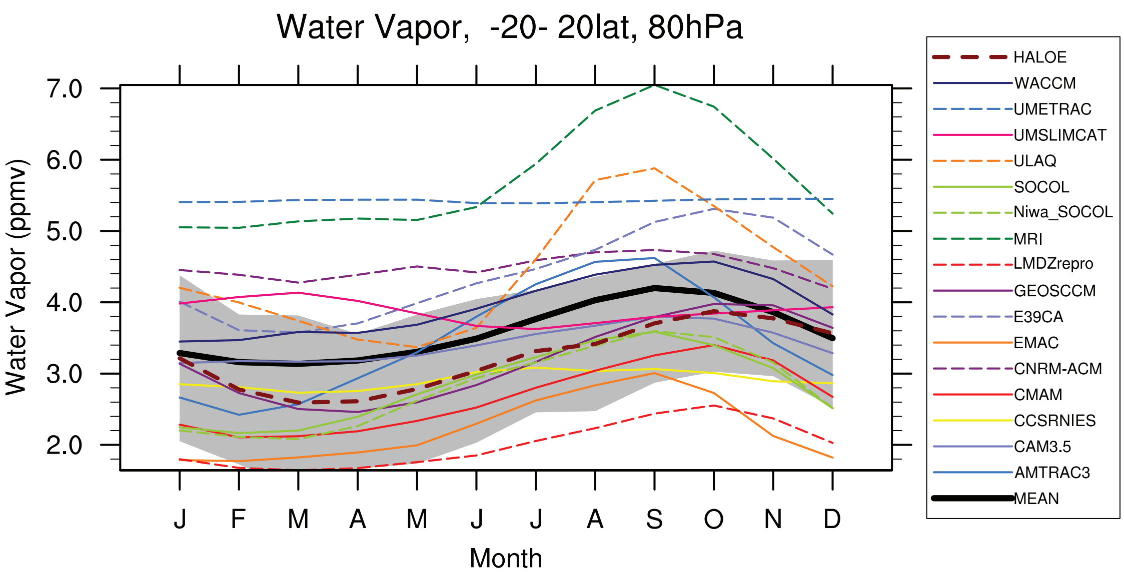

Figure 9. Annual cycle of tropical (20S–20N) water vapor at 80 hPa from models and observations. Output from the period 1992–2004. Gray shaded region is 3σ variability from HALOE observations over 1992–2004 (thick brown dashed line). The multi-model mean (MEAN) is the thick black line.

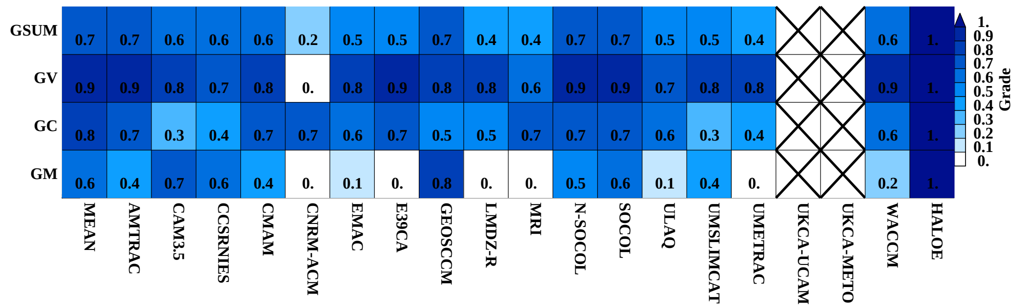

Figure 10. Quantitative diagnostics summary of 80hPa water vapor mixing ratio for mean (GM), correlation (GC), variance (GV) and the average (GSUM).

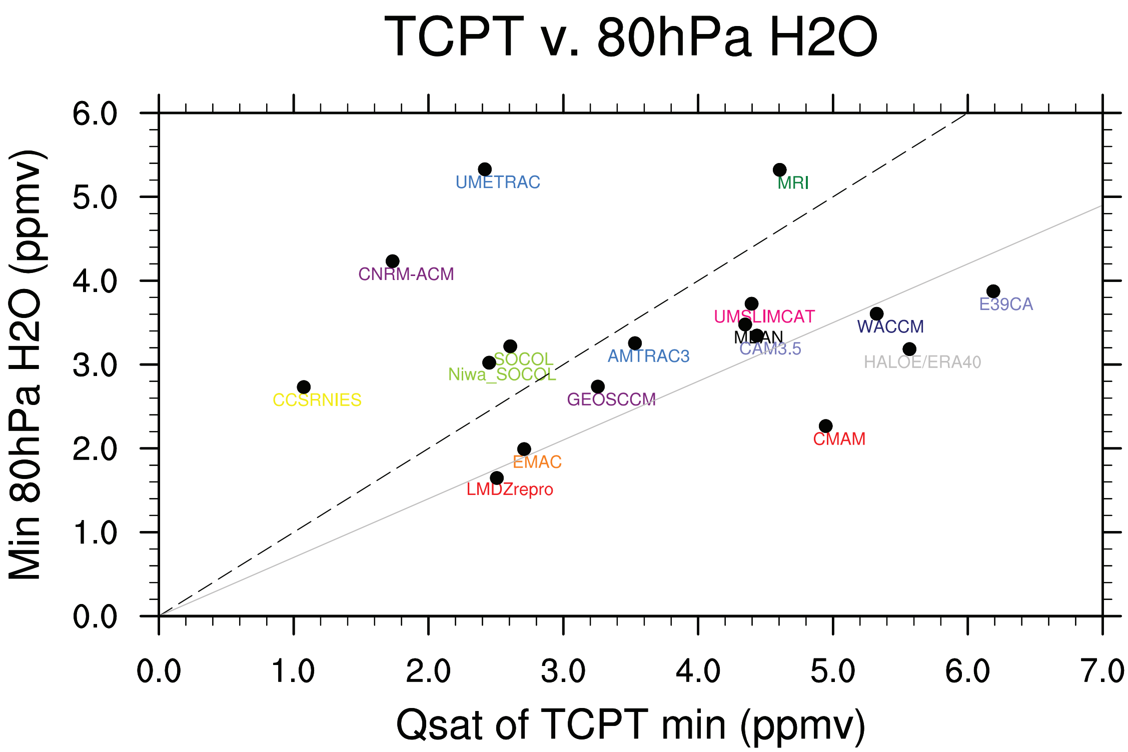

Figure 11. Correlation of minimum monthly mean water vapor with saturation vapor mixing ratio (QSAT) of the minimum monthly mean TCPT from CCMVal-2 models (1980-1999), HALOE and ERA40 observations (HALOE over 1992-2005) and multi-model mean (MEAN-black). The black dashed line is the 1:1 line, indicating 100% saturation. The gray line is the 0.7:1 line, indicating 70% saturation.

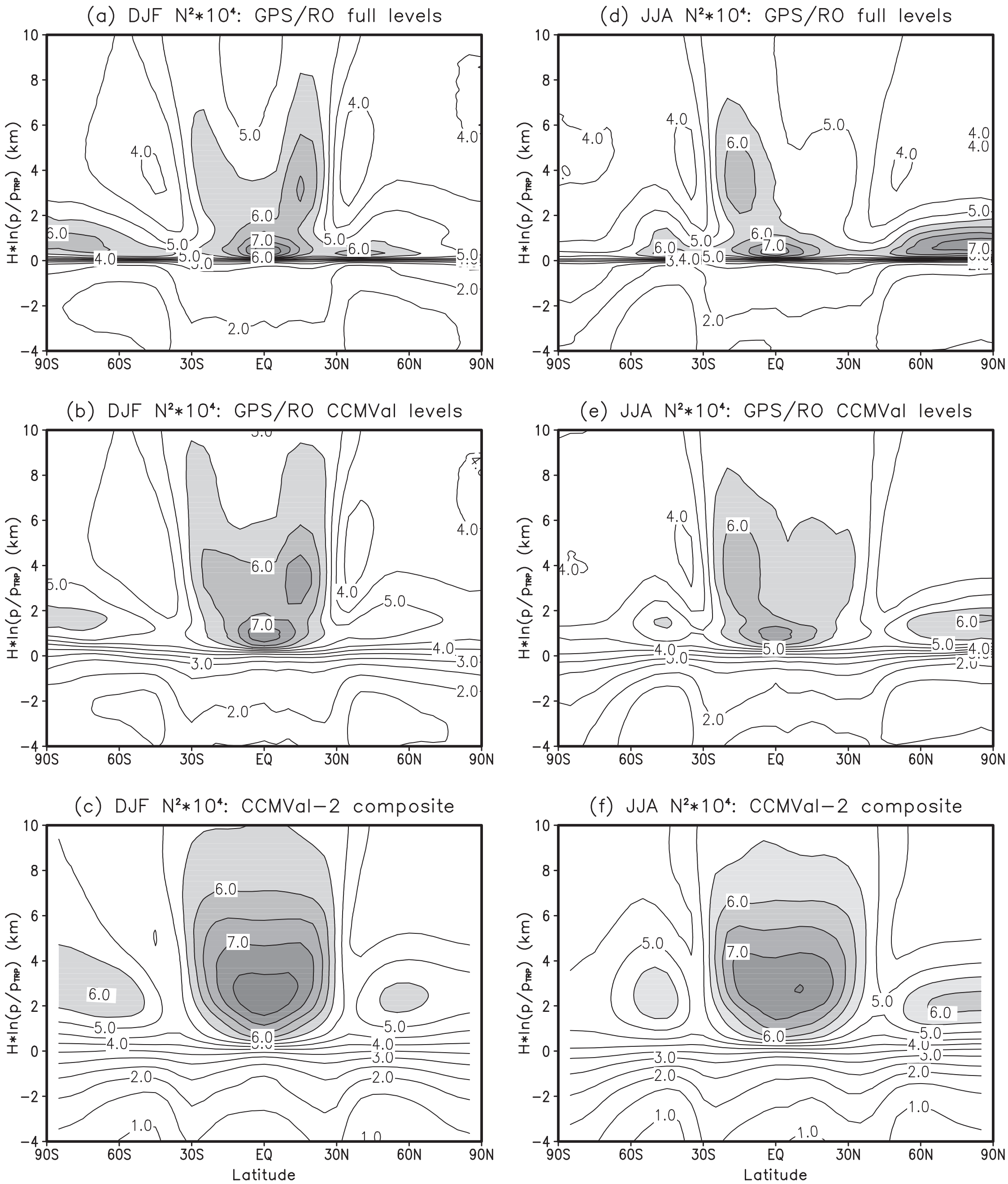

Figure 12. Zonally-averaged N^2 as a function of latitudes and log-p height on the tropopause based coordinate: (a,d) COSMIC GPS RO data, (b,e) COSMIC GPS RO data using only CCMVal-2 standard pressure levels, and (c,f) composite of REF-B1 integrations from 9 Models. Two seasons are shown separately: (a,b,c) DJF and (d,e,f) JJA. Contour intervals are $0.5 \mathrm{x} 10^{-4} s^{-1}$. Values greater than or equal to $5.5 \mathrm{x} 10^{-4} s^{-1}$ are shaded. y=0 denotes the location of the tropopause.

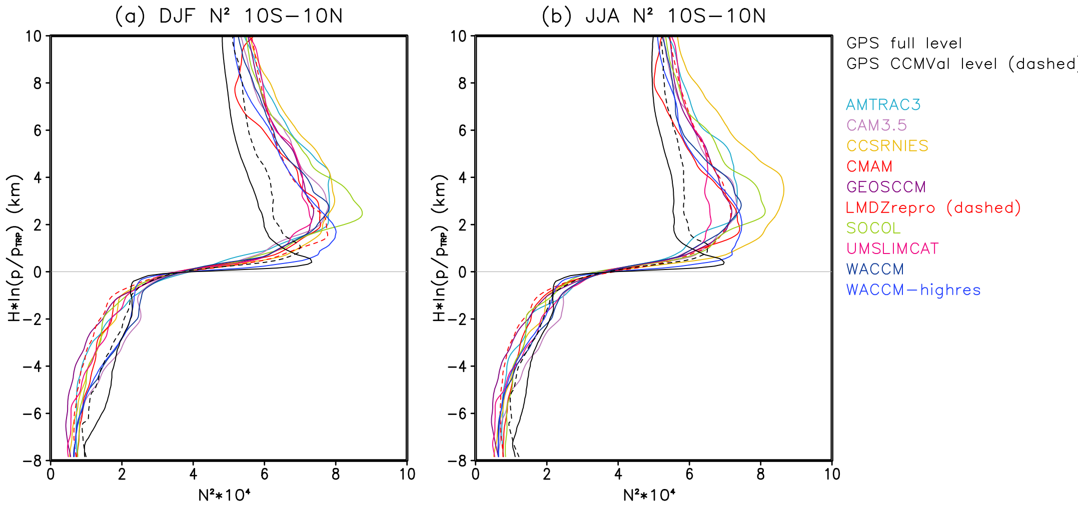

Figure 13. Vertical profiles of N^2 in each model and GPS RO observations in the tropics.

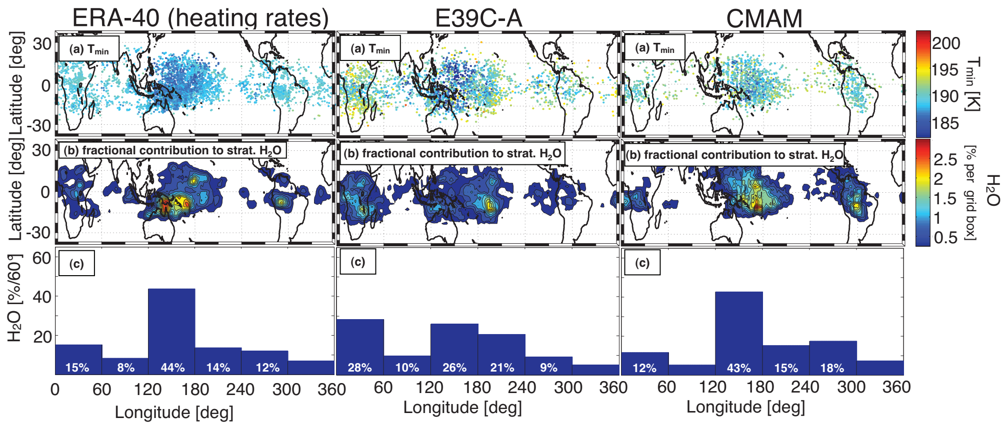

Figure 14. NH winter 1995-1996. The scatter plots (panel a) show the geographical distribution of the dehydration points for ERA40 (left), E39CA (middle), and CMAM (right). Color code in (a) shows the minimum temperatures experienced by the trajectories. Panel (b) illustrates the fractional contribution to stratospheric water vapor from different geographical areas, expressed as percentage contribution per individual 10 x 5 grid boxes. Panel (c) shows longitudinal distribution of the water vapor entry value, i.e. the value from (b) integrated over latitude (30°N-30°S) per 60° longitude.

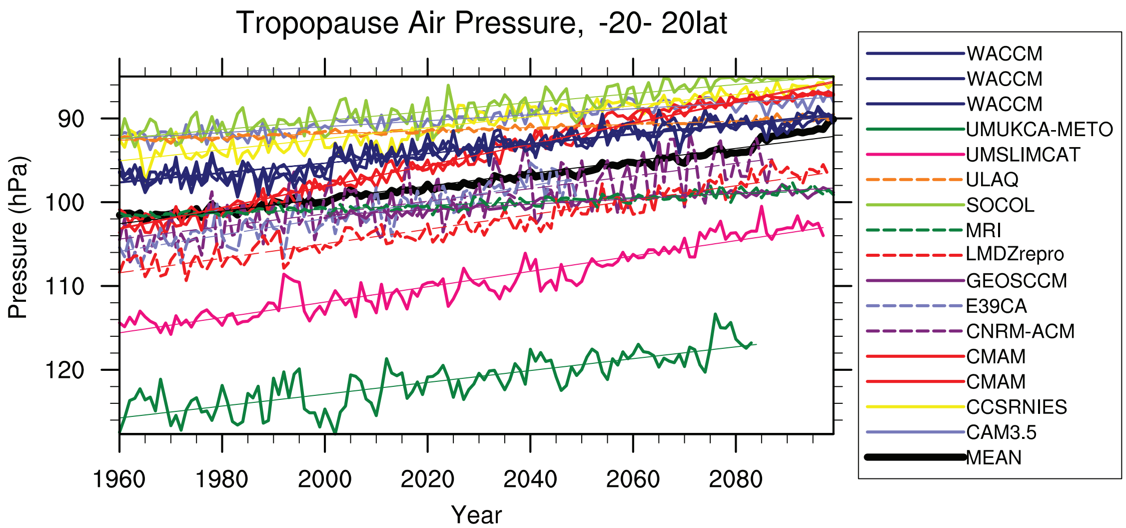

Figure 15. Lapse Rate Tropopause Pressure time series from 20S-20N for future REF-B2 scenarios. Thin lines are linear fits. Multi-model mean (MEAN) is the thick black line.

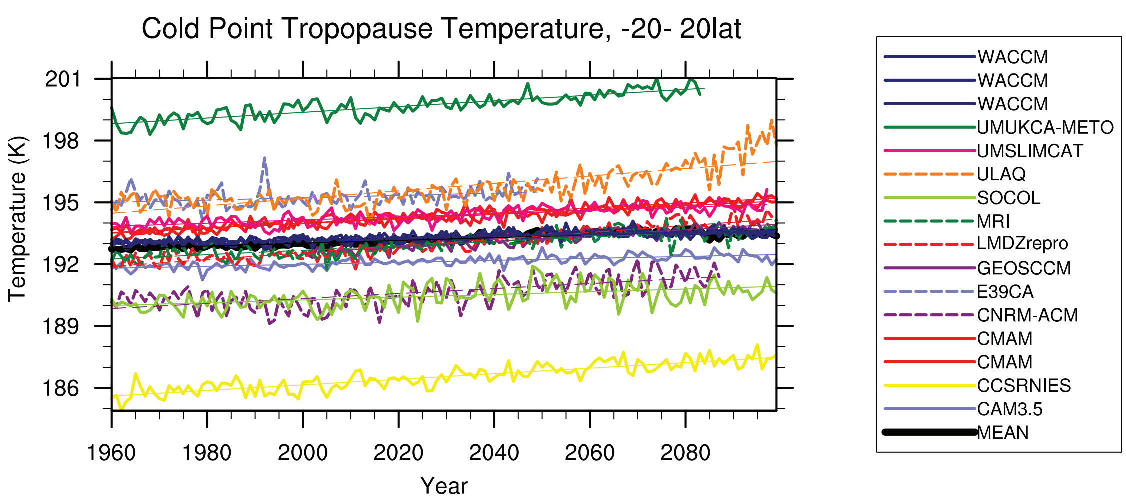

Figure 16. Cold Point Temperature time series from 20S–20N for future REF-B2 scenarios. Thin lines are linear fits. Multi-model mean (MEAN) is the thick black line.

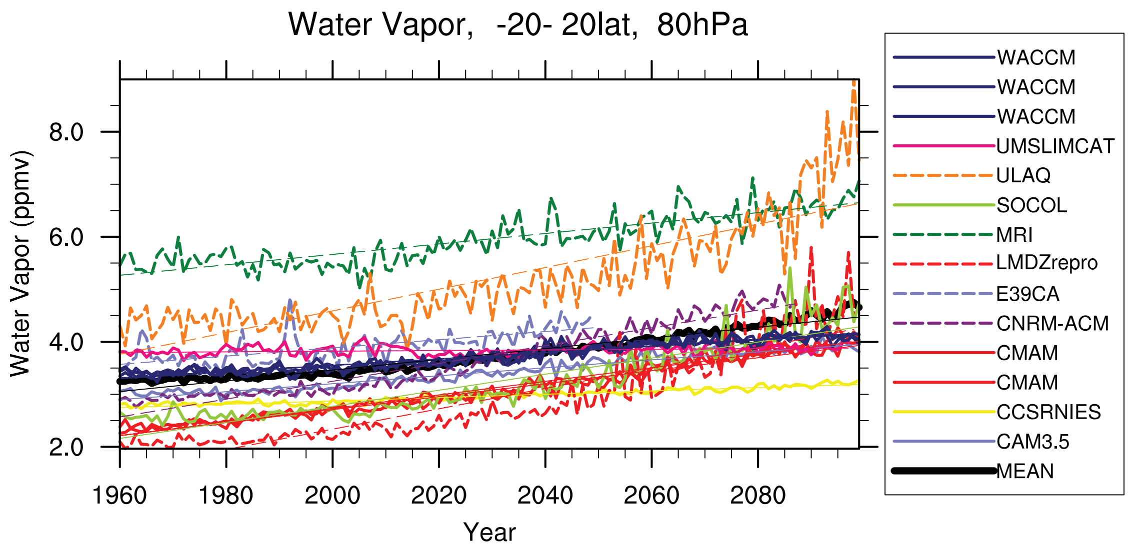

Figure 17. 80 hPa Water Vapor time series from 20S-20N for future REF-B2 scenarios. Thin lines are linear fits. Multi-model mean (MEAN) is the thick black line.

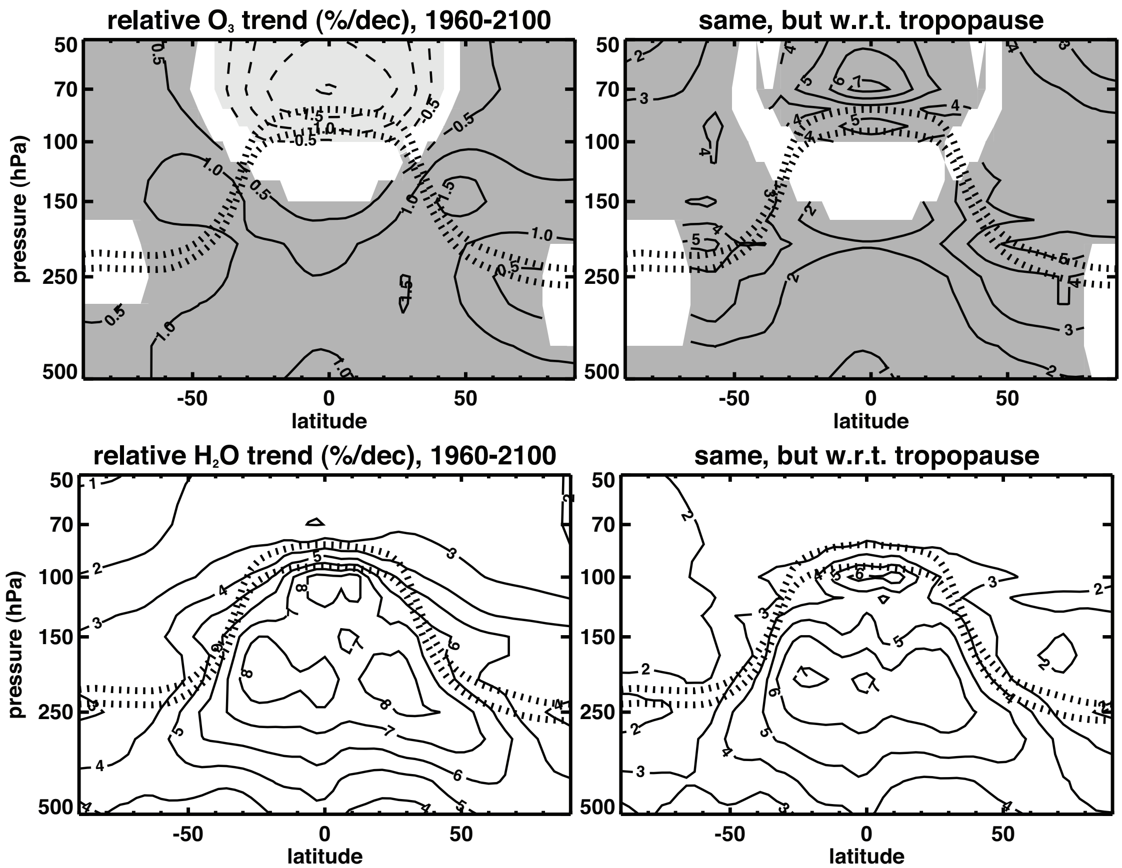

Figure 18. Multi-model mean trends in O_3 (upper panels) and H_2O (lower panels) in pressure (left panels) and tropopause coordinates (right panels). Shading indicates the 95% significance level. For H_2O, the calculated trends are significant at the 95% level. Dotted lines in each panel denote the tropopause with the lower line corresponding to the reference period [1960-1980] and the upper line corresponding to the year 2100.

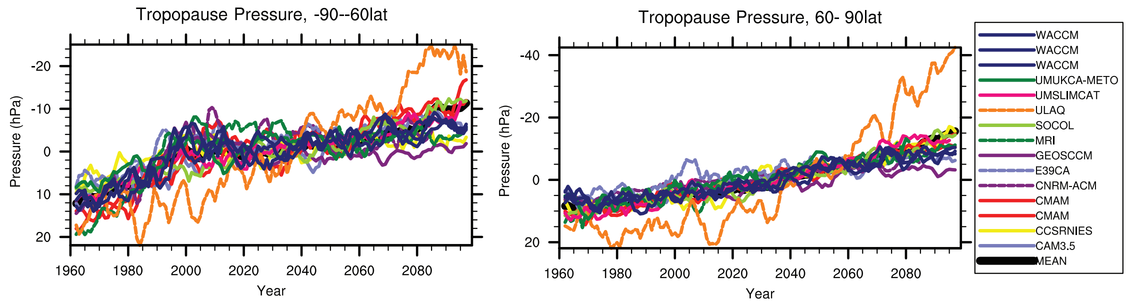

Figure 19. Northern and Southern Hemisphere extra-tropical tropopause pressure time series from 90S–60S (left) and 60N–90N (right) for future REF-B2 scenarios. Multi-model mean (MEAN) is the thick black line.

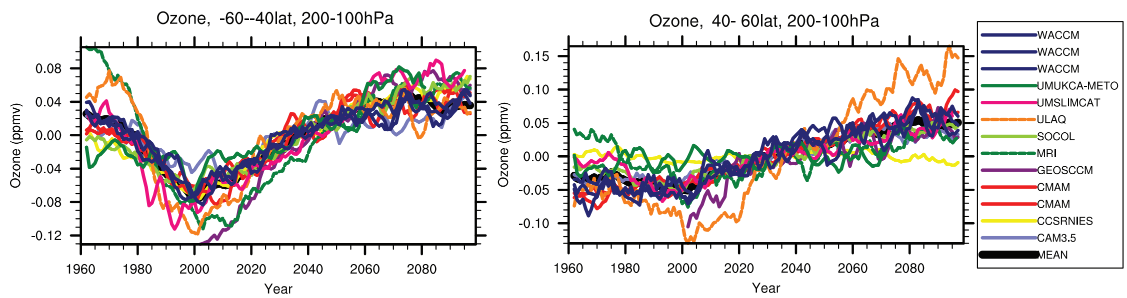

Figure 20. Ozone trends in the (A) Southern and (B) Northern extra-tropical lowermost stratosphere (40–60 latitude, 200–100hPa). Multi-model mean (MEAN) is the thick black line.

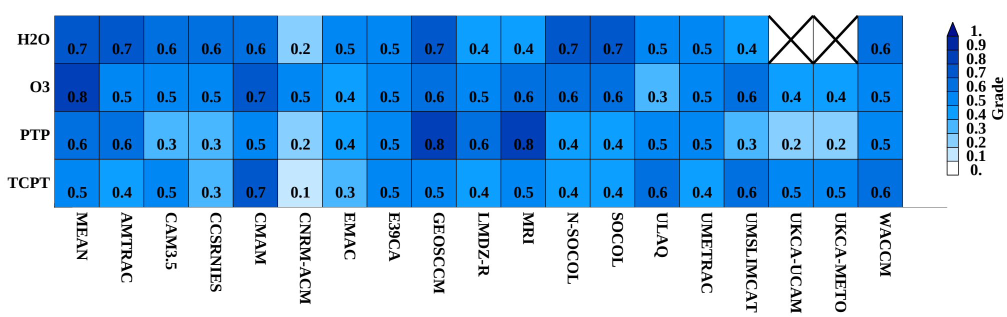

Figure 21. Quantitative grades summary (G_{sum}) for 4 diagnostics: Water Vapor (H2O), Ozone (O3), Tropopause Pressure (PTP) and Tropopause Temperature (TCPT).